



Inteligência Artificial

Inteligência Artificial

Ruy Flávio de Oliveira

© 2018 por Editora e Distribuidora Educacional S.A.

Todos os direitos reservados. Nenhuma parte desta publicação poderá ser reproduzida ou transmitida de qualquer modo ou por qualquer outro meio, eletrônico ou mecânico, incluindo fotocópia, gravação ou qualquer outro tipo de sistema de armazenamento e transmissão de informação, sem prévia autorização, por escrito, da Editora e Distribuidora Educacional S.A.

Presidente

Rodrigo Galindo

Vice-Presidente Acadêmico de Graduação e de Educação Básica

Mário Ghio Júnior

Conselho Acadêmico

Ana Lucia Jankovic Barduchi

Camila Cardoso Rotella

Danielly Nunes Andrade Noé

Grasiele Aparecida Lourenço

Isabel Cristina Chagas Barbin

Lidiane Cristina Vivaldini Olo

Thatiane Cristina dos Santos de Carvalho Ribeiro

Revisão Técnica

Alessandra Cristina Santos Akkari

Hugo Tanzarella Teixeira

Vanessa Cadan Scheffer

Editorial

Camila Cardoso Rotella (Diretora)

Lidiane Cristina Vivaldini Olo (Gerente)

Elmir Carvalho da Silva (Coordenador)

Leticia Bento Pieroni (Coordenadora)

Renata Jéssica Galdino (Coordenadora)

Dados Internacionais de Catalogação na Publicação (CIP)

Oliveira, Ruy Flávio de
O48i Inteligência artificial / Ruy Flávio de Oliveira. – Londrina :
Editora e Distribuidora Educacional S.A., 2018.
224 p.

ISBN 978-85-522-1141-9

1. Redes neurais. 2. Prolog. 3. Cognição. I. Oliveira, Ruy
Flávio de. II. Título.

CDD 006.3

Thamiris Mantovani CRB-8/9491

2018

Editora e Distribuidora Educacional S.A.
Avenida Paris, 675 – Parque Residencial João Piza
CEP: 86041-100 – Londrina – PR
e-mail: editora.educacional@kroton.com.br
Homepage: <http://www.kroton.com.br/>

Sumário

| | |
|--|------------|
| Unidade 1 Fundamentos da inteligência artificial | 7 |
| Seção 1.1 - Introdução à inteligência artificial | 9 |
| Seção 1.2 - Metodologias de inteligência artificial | 26 |
| Seção 1.3 - Inteligência artificial na prática | 44 |
| Unidade 2 Resolução de problemas em inteligência artificial | 63 |
| Seção 2.1 - Tipos e representação de problemas | 65 |
| Seção 2.2 - Buscas com inteligência artificial | 80 |
| Seção 2.3 - Soluções via inteligência artificial | 92 |
| Unidade 3 Lógica nebulosa | 111 |
| Seção 3.1 - Introdução à lógica nebulosa | 113 |
| Seção 3.2 - Conjuntos fuzzy | 128 |
| Seção 3.3 - Elementos de lógica nebulosa | 145 |
| Unidade 4 Redes neurais artificiais | 163 |
| Seção 4.1 - Introdução às redes neurais | 165 |
| Seção 4.2 - Redes neurais e aprendizado | 182 |
| Seção 4.3 - Elementos de Redes Neurais | 203 |

Palavras do autor

Imagine um parente seu, cardíaco, contando em uma festa de fim de ano que baixou um aplicativo em seu celular, o qual identifica momentos em que a arritmia — condição crônica desse parente — está atingindo níveis fora do padrão. Ele demonstra para você o tal aplicativo colocando o dedo sobre o flash do aparelho (que liga automaticamente quando o aplicativo é inicializado), e em poucos segundos o diagnóstico aparece na tela. Ele explica que a câmera capta as variações de cor quando o dedo pulsa, um efeito amplificado pelo flash, e que o aplicativo trabalha com inteligência artificial. Você ficaria tranquilo e sentiria que seu parente está tendo um bom atendimento quanto à sua arritmia? Ou rapidamente diria a ele que não deve se medicar ou adotar condutas com base em um software para smartphone?

Pois é, se você escolheu a segunda opção, saiba que errou. Esse aplicativo ainda não existe, mas toda a tecnologia já está disponível — inclusive o software por trás da solução — e os resultados obtidos são mais precisos que os gerados por um painel de especialistas, como nos mostram Rajpurkar et al. (2017). Isso mesmo: um simples aplicativo de celular consegue resultados melhores do que os de um painel de especialistas no diagnóstico de certas doenças cardíacas.

E você sabe qual é a grande ferramenta tecnológica que alicerça essa solução? A inteligência artificial. A cada dia que passa, esse campo de estudos se expande e produz resultados mais importantes e úteis para todos. E é exatamente por conta dessa importância que estamos iniciando os estudos em inteligência artificial.

Na Unidade 1 você vai conhecer os fundamentos da inteligência artificial, com ênfase em seus conceitos e propósitos.

Na Unidade 2 você vai entender como ocorre a resolução de problemas por meio das várias técnicas de inteligência artificial. Após finalizar as Unidades 1 e 2, você vai conhecer, compreender e ser capaz de resolver problemas em inteligência artificial utilizando as metodologias apresentadas.

Na Unidade 3 você vai conhecer a Lógica Nebulosa e suas aplicações práticas.

Na Unidade 4, por fim, você vai conhecer as Redes Neurais, e vai entender como essa nova tecnologia está revolucionando não só a inteligência artificial, mas como também tem o potencial de revolucionar tudo à sua volta. Após finalizar as Unidades 3 e 4, você vai conhecer, compreender e saber utilizar os principais conceitos relacionados à Lógica Nebulosa e a redes neurais artificiais.

Em suma: essa é uma disciplina em que você vai entrar em contato com muitas ferramentas relevantes para seu futuro profissional. Aproveite bastante, faça todos os exercícios e atividades complementares, e prepare-se para aprender muitas coisas interessantes.

Fundamentos da inteligência artificial

Convite ao estudo

Olá!

Veja você como são as coisas: a imaginação do ser humano parece não ter limites. Assim que novas possibilidades tecnológicas surgem, alguém invariavelmente imagina como essas possibilidades podem alterar radicalmente nossas vidas, de maneira imprevisível. Tem sido assim desde que Júlio Verne nos levou ao redor da Lua e ao centro da Terra. O mesmo fenômeno pode ser observado desde que os primeiros computadores passaram a funcionar, ainda na década de 1940: contos sobre como estas máquinas — que na época não passavam de gigantescas e barulhentas calculadoras — ganhariam consciência e tomariam suas próprias decisões, nem sempre em benefício de nós, seus criadores. E a característica (ainda inexistente naquela época) que possibilita esse relacionamento supostamente de igual para igual entre o computador e o homem não é a velocidade nem a capacidade de processamento, mas sim algo que para nós seres humanos é absolutamente natural: nossa inteligência. E para que os computadores possam nos desafiar de alguma maneira, precisarão exibir um comportamento que possa ser considerado “inteligente”. É aí que começamos nossos estudos: definindo o que é a inteligência artificial e contextualizando seus conceitos.

Ao conhecer os conceitos de Inteligência Artificial, você compreenderá um pouco mais sobre a resolução de problemas que pode ser alcançada por meio dessa tecnologia, utilizando para tanto as metodologias que serão apresentadas.

Uma grande empresa nacional está em busca de diversificar seus investimentos no Brasil. Tendo em sua composição experientes investidores no mercado financeiro, chegou

ao conhecimento dessa empresa um artigo da revista *The Economist* informando que a inteligência artificial, quando bem utilizada, pode melhorar os ganhos nos investimentos. O responsável pelo setor ficou curioso, mas como não conhece nada acerca dessa tal "inteligência artificial", decidiu procurar sua empresa, que desenvolve soluções nessa área para, quem sabe, iniciar projetos utilizando a tecnologia em seus negócios. Para tanto, primeiramente, a empresa cliente precisa conhecer um pouco mais acerca do assunto, e conta com seu apoio.

Será que a inteligência artificial pode mesmo beneficiar essa empresa cliente? Como tornar isso possível?

Na primeira seção desta unidade você vai conhecer os conceitos básicos da inteligência artificial, como no caso de sua definição e principais características.

Na segunda seção você vai conhecer as metodologias fundamentais da inteligência artificial.

Na terceira seção você vai conhecer algumas das principais materializações práticas das técnicas de inteligência artificial, por meio de soluções históricas que surgiram nesse contexto.

Prepare-se: um mundo novo de conhecimento e de possibilidades profissionais está prestes a se abrir para você.

Seção 1.1

Introdução à inteligência artificial

Diálogo aberto

Um dos conceitos mais fascinantes para a ciência é o de inteligência. Desde o princípio da Humanidade, primeiramente os filósofos e depois os cientistas buscam entender o que é ser inteligente, e como essa inteligência se manifesta nos seres vivos. Recentemente, com a evolução dos sistemas computacionais digitais introduzidos a partir do século XX, os estudos da inteligência passaram a questionar se os computadores podem, de alguma forma, exibir comportamento que possamos classificar como “inteligentes”. Para estudar a fundo esse assunto, algumas definições são fundamentais: o que é inteligência? O que permite que uma entidade ou sistema possa ser classificado como inteligente? Quais as características, possibilidades e limitações de um sistema computacional dito inteligente? É o que veremos na presente seção.

Nessa primeira etapa da consultoria a ser prestada à empresa de investimentos, que quer conhecer um pouco mais acerca da inteligência artificial para melhorar o desempenho de seus investimentos, você deve apresentar ao investidor um estudo sobre o que é inteligência artificial, com exemplos. Além disso, o estudo deve pontuar como essa área teve início e evoluiu até os dias de hoje. Para finalizar, você deve mostrar o que temos hoje, à nossa disposição, em termos de estado da arte em inteligência artificial. Apresente seus resultados sob forma de um relatório para a empresa de investimentos, considerando que se trata de indivíduos esclarecidos, mas não especialistas em assuntos de TI. Como apresentar o conceito de inteligência artificial de maneira que faça sentido para alguém que não é da área, tornando-o compreensível e palatável?

Nessa presente seção você vai conhecer os principais conceitos do campo de estudos da inteligência artificial, vai saber como tem sido a evolução histórica da área nas últimas décadas, além de conhecer alguns exemplos de como a inteligência artificial está sendo usada nos dias de hoje.

Bons estudos!

Você já parou para pensar o quanto o conceito de inteligência é importante para todos nós? Já percebeu que nossa própria espécie é definida com base em nossa capacidade de pensar e tomar decisões inteligentes? Pois é, cada um dos mais de sete bilhões de seres humanos em nosso planeta pertence à espécie *Homo sapiens*, ou seja, como nos mostram Russell e Norvig (2013), pertencemos ao gênero humano (*Homo*) e à espécie que pode ser considerada “sábia” (*sapiens*). Sim, você tem razão ao pensar que algumas pessoas não são nada sábias. Mas isso não muda o fato de que somos a única espécie do planeta capaz de apresentar sabedoria, isto é, conhecimento baseado em um longo processo de aprendizado, capacitando o indivíduo a tomar decisões acertadas e a realizar análises precisas sobre as situações que enfrenta, sempre com base em seu conhecimento, sua capacidade de raciocínio e sua experiência (TEJEDA, CHERUBIN, 2016).

Definição e Fundamentos de Inteligência Artificial

Há muitas definições acerca do que é inteligência artificial, e parece não haver consenso entre os vários autores que estudam o assunto. Isso porque há muitas maneiras de se definir o que é inteligência sob o ponto de vista humano, e os pesquisadores desse campo de estudo também apresentam várias definições (LIMA, PINHEIRO, SANTOS, 2014).

Uma definição de inteligência artificial mais geral e que nos aproxima de um entendimento intuitivo sobre o assunto é a que nos oferecem Lima, Pinheiro e Santos (2014), que afirmam que a I.A. é o conjunto de ações que, se fossem realizadas por um ser humano, seriam consideradas inteligentes. É uma definição que nos chama a comparar as ações dos computadores com as nossas próprias para definirmos se estamos diante de um comportamento inteligente ou não. Em que pese ser uma definição útil, não é de todo precisa, uma vez que há algumas ações que o computador consegue realizar que não traduzem adequadamente um comportamento inteligente:

- Um computador é muito mais rápido que o ser humano para fazer cálculos matemáticos, mas essa rapidez é mecânica, e mesmo pessoas com severas deficiências mentais em

alguns casos exibem essa capacidade, como Jourdain (1997) nos mostra nos casos de savantismo (síndrome do idiota-prodígio).

- Um computador consegue guardar e recobrar grandes quantidades de dados — ordens de grandeza mais que um ser humano — mas isso também não é sinal claro de inteligência (JOURDAIN, 1997).

Rosa (2011) apresenta uma variação mais proveitosa dessa definição incompleta quando afirma que inteligência artificial é a capacidade de os computadores realizarem tarefas que os seres humanos inerentemente realizam melhor até o momento. Em outras palavras: quando um computador consegue fazer uma tarefa que até então apenas os seres humanos (e aqui nos referimos a seres humanos comuns, sem deficiências ou habilidades especiais) conseguiam fazer, este computador exhibe traços de inteligência, e por ser uma entidade fabricada — construída por seres humanos em vez de ter surgido espontaneamente na natureza — é artificial.

Russell e Norvig (2013) afirmam que a inteligência artificial é a capacidade dos sistemas cibernéticos (formados por computadores, softwares, sensores e atuadores) de imitar funções cognitivas dos seres humanos, funções estas que podemos resumir na resolução de problemas por meio do aprendizado apoiado na percepção. É importante analisarmos os três termos chave da definição anterior:

- **Resolução de Problemas** – Este termo é o de compreensão mais comum, e pode ser entendido como a capacidade de realizar tarefas julgadas corretas por algum critério de avaliação. No caso de um problema matemático qualquer, por exemplo, a resolução pode ser compreendida como a resposta correta ao problema, atingida por meio das regras matemáticas envolvidas no assunto em questão. Um computador consegue, com algoritmos e técnicas de programação, resolver problemas. A maior parte das resoluções possíveis ocorre por meio da execução mecânica de procedimentos, o que não denota inteligência. Isso significa que este primeiro termo chave não é o item principal desta definição, uma vez que até uma calculadora feita com engrenagens — a exemplo da Pascalina de Blaise Pascal, criada em 1649 (CHAPMAN, 1942) — consegue resolver problemas.

A resolução de problemas, em se tratando de inteligência artificial, é necessária, mas não é suficiente.

- **Aprendizado** – No caso da inteligência artificial o aprendizado é a característica dos algoritmos que têm a capacidade de melhorar seu desempenho por meio da experiência (RUSSELL; NORVIG, 2013). O aprendizado de máquina, como é chamado (do inglês *machine learning*), consiste em um algoritmo ajustar seus parâmetros internos com dados oriundos de execuções anteriores, atingindo melhores resultados a cada execução.
- **Percepção** – Para nós, seres humanos, percepção é uma definição que se refere a dois tipos de fenômenos: no âmbito **físico** é a capacidade de coletar dados acerca do ambiente a nossa volta, por meio de nossos cinco sensores naturais (a temperatura de um objeto ou sua cor, por exemplo); no âmbito **mental**, é a capacidade de deduzir as situações que ocorrem a nossa volta com a análise das informações que chegam até nós (se uma pessoa está triste ou feliz com base em seu semblante, sua postura, suas palavras e sua entonação, por exemplo). No caso dos sistemas cibernéticos, a percepção é baseada em sensores que coletam objetivamente dados acerca do problema em questão. Russell e Norvig (2013) argumentam que esta percepção, no caso dos sistemas cibernéticos, vai desde a leitura de dados por um software (leitura lógica de dados, no caso, armazenados em arquivos), até sensores físicos, isto é, que interagem com o ambiente, como no caso de acelerômetros (medidores de aceleração direcional), sensores de som, termômetros, medidores de intensidade luminosa e afins.



Assimile

Inteligência artificial é a capacidade de os computadores realizarem tarefas que os seres humanos inerentemente realizam melhor, até o momento.

Fundamentos da Inteligência Artificial

Além da combinação desses três elementos, dois deles centrais à inteligência artificial, e o terceiro (percepção) pertinente ao conceito, há outras características inerentes ao

estudo da inteligência artificial, entre as quais podemos citar como principais (FRANCO, 2014):

- **Representação do Conhecimento** – Não é difícil perceber que para a resolução de problemas reais, em muitos casos, um sistema cibernético de inteligência artificial necessitará de informações genéricas acerca do contexto desse problema. Russell e Norvig (2013) afirmam que a representação do conhecimento é o processo de identificar informações genéricas e as representar de uma forma que o computador as reconheça e seja capaz de processá-las. Dado um determinado assunto, o sistema de inteligência artificial deverá ser capaz de identificar e representar adequadamente vários aspectos sobre o assunto, entre eles:
 - o Objetos.
 - o Propriedades.
 - o Categorias.
 - o Relações.
 - o Situações em que os objetos podem se encontrar, propriedades que podem apresentar, relações que podem ocorrer.
 - o Eventos que podem ocorrer.

Toda a representação de um determinado assunto (cujos exemplos englobam os *bullets* apresentados) constituem uma ontologia. Os sistemas de inteligência artificial precisam da definição formal e completa de todos os aspectos da ontologia, de forma que consigam processar as informações e resolver os problemas que nesta ontologia possam ser representados. A Figura 1.1 a seguir apresenta uma ontologia simplificada:

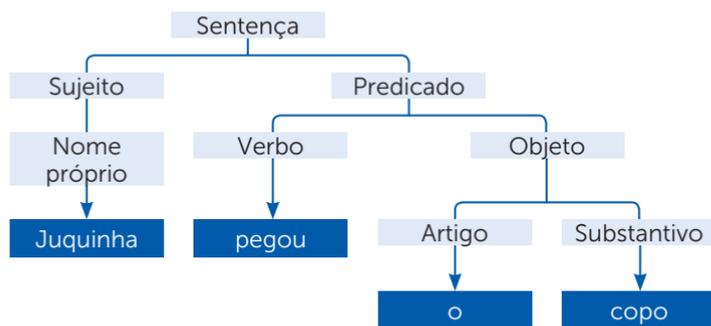
Figura 1.1 | Ontologia simplificada



Fonte: adaptada de Russell e Norvig (2013).

- **Processamento de Linguagem Natural** – Um sistema de inteligência artificial precisa ter a capacidade de interagir com o mundo real, e para tanto, a capacidade de interagir com seres humanos utilizando não as linguagens formais e não ambíguas dos computadores, mas também a linguagem humana – com todas as suas idiossincrasias e imperfeições – é fundamental. Um sistema de inteligência artificial que consiga processar adequadamente a linguagem natural tem a capacidade de absorver informações diretamente de conversas com seres humanos e também de material multimídia (vídeos, textos falados), bem como da mídia escrita (livros, artigos e textos em geral). Para tanto, o processamento de linguagem natural deve ser capaz de dividir um texto em suas partes fundamentais sob o ponto de vista gramatical e assim identificá-lhe todos os elementos (RUSSELL; NORVIG, 2013). A Figura 1.2 a seguir mostra a divisão sintática de uma sentença simples:

Figura 1.2 | Divisão sintática de uma sentença simples



Fonte: adaptada de Russell e Norvig (2013).

- **Criatividade** – A questão da criatividade vem propondo desafios enormes à psicologia e, mais recentemente, à inteligência artificial. A própria definição e delimitação do que é criatividade encontra obstáculos cognitivos em sua definição. Classicamente, Newell, Shaw e Simon (1963), em seu artigo seminal sobre o assunto, apresentam quatro características que, se estiverem presentes, apontam para uma solução criativa para uma questão qualquer:

- **A resposta é inovadora e útil** – uma resposta que utilize elementos existentes de maneira nova é inovadora. Contudo, esta resposta deve ter relevância dentro da questão tratada, não podendo derivar de seu propósito central.
- **A resposta requer que rejeitemos formas anteriores de se pensar sobre a questão** – A criatividade artificial se apresenta quando um sistema de inteligência artificial — livre dos hábitos e da mecânica a que nós, seres humanos, muitas vezes nos prendemos, e busca maneiras diferentes de resolver a questão. Neste caso não ocorre tanto uma rejeição em si das formas corriqueiras de se pensar, mas não havendo familiaridade com estas formas mais comuns, o sistema de inteligência artificial explora novas formas com a mesma intensidade, sem predileções.
- **A resposta surge de um processo intenso de motivação e persistência** – Neste caso a inteligência artificial não apresenta nem motivação nem persistência ativa, mas sim cumpre com o programa a ela designado. Ocorre que a inteligência artificial, uma vez tendo iniciado sua tarefa, a executará sem “preguiça” até que seus programadores definam que o resultado está adequado. Como veremos mais à frente, uma rede neural gera resultados cada vez melhores a cada iteração, podendo chegar a resultados surpreendentes após algumas horas ou alguns dias de treinamento.
- **A resposta surge do esclarecimento de uma questão que anteriormente era pouco formalizada e/ou delimitada** – Um sistema de inteligência artificial busca identificar — à sua maneira — o ponto central (ou os pontos centrais) de um problema e resolvê-lo. Ocorre que em assim procedendo, criará uma formalização própria do problema, e as soluções que daí surgirem poderão ser encaradas como criativas sob o ponto de vista do problema inicial, mais vago e menos formalizado.

Nos dias de hoje ainda podemos usar os quatro critérios apresentados para identificar traços de criatividade em um sistema de inteligência artificial.



Uma rede neural consegue, nos dias de hoje e após apenas seis horas de treinamento, produzir música que lembra o estilo barroco. Em que pese o programa não entender nada de música, é possível descrever o resultado como sendo criativo? Ouça o resultado no seguinte vídeo <https://www.youtube.com/watch?v=SacogDL_4JU&t=300s> e chegue às suas próprias conclusões. CARYKH. **Computer evolves to generate baroque music**. Acesso em 17 mar. 2018.

Histórico da Inteligência Artificial

McCorduck (2004) nos mostra que a ideia de seres artificiais inteligentes e antropomórficos é quase tão antiga quanto a humanidade. A autora nos aponta o poema Pigmalião, de Ovídio, no qual o autor — o escultor Pigmalião — esculpe uma estátua de marfim tão bela e perfeita que se apaixona por ela. Em súplica a Afrodite por um amor que seja à semelhança de sua estátua, se surpreende quando ela própria ganha vida e corresponde ao seu amor. Ainda à época de Ovídio, McCorduck (2004) nos relata que Heron de Alexandria construía homens mecânicos, a que batizava de “autômatos”. Os autômatos de Heron pouco mais eram que manequins com mecanismos internos que lhes permitia alguma movimentação com base em configurações realizadas pelo próprio inventor. Sob o ponto de vista da lógica, algum tempo antes de Ovídio, o filósofo Aristóteles definia as regras do silogismo, que permitem o desenvolvimento de raciocínios matematicamente exatos, sem contradições ou ambiguidades, quando desenvolvidos dentro das regras expressas. O silogismo aristotélico vigora até hoje na lógica booleana, absolutamente necessária aos computadores que permeiam nossa sociedade.

Saindo da antiguidade e chegando à era moderna, podemos perceber que a história da inteligência artificial baseada em computadores digitais quase que se confunde com a história da própria computação. Logo após do surgimento do primeiro computador digital, em 1941, criado pelo engenheiro alemão Konrad Zuse (MCCORDUCK, 2004), a corrida pela criação de máquinas cada vez mais potentes para auxiliar no esforço de guerra (especificamente para uso durante a Segunda Guerra Mundial) estava em curso.

McCorduck (2004) nos mostra que em 1943 Warren Sturgis McCulloch e Walter Pitts publicam o artigo seminal intitulado *A Logical Calculus of the Ideas Immanent in Nervous Activity* (em português, Um cálculo lógico das ideias imanentes à atividade nervosa), que seria a base para toda a teoria de redes neurais que desenvolvemos nos dias de hoje (e que serão estudadas em mais detalhes na Unidade 4 deste livro).

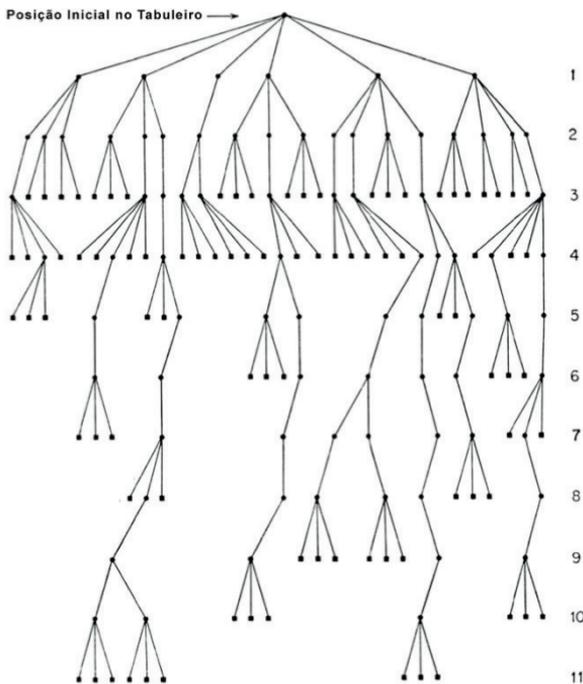
Em 1948, John von Neumann afirma, como resposta a um questionamento surgido durante uma de suas palestras: "Você insiste que há algo que uma máquina não pode fazer. Se você me disser precisamente o que é que uma máquina não pode fazer, então eu sempre posso fazer uma máquina que fará exatamente isso!" De fato, a afirmação de Von Neumann se alicerça sobre a tese de Church-Turing, que afirma que qualquer procedimento formalizável pode ser realizado por uma máquina de Turing, ou seja, por um computador genérico (MCCORDUCK, 2004).

Pouco tempo depois, em 1950, o próprio Alan Turing, prevendo o desenvolvimento do campo da inteligência artificial, propõe o Teste de Turing para mensurar o equivalente de inteligência de um computador (e que veremos na segunda seção dessa nossa disciplina de inteligência artificial).

Nilsson (2010) nos conta que em 1952, dois pesquisadores atingiram o mesmo resultado surpreendente como precursores da inteligência artificial: Arthur Samuel, da IBM, e Christopher Strachey, da Universidade de Oxford implementaram, independentemente, versões do jogo de damas. Em ambos os casos o resultado foi que o jogo de computador conseguia ser bem-sucedido contra bons jogadores amadores. As implementações de Samuel e Strachey são significativas porque demonstram as primeiras instâncias em que um computador foi usado para desempenhar uma atividade patentemente humana, e já começou fazendo isso — mesmo com as severas limitações dos computadores do início da década de 1950 — de maneira rápida e eficiente. Samuel iria ainda mais longe, e em 1955 melhoraria seu jogo, inserindo elementos que permitiam ao computador utilizar informações de jogos anteriores, e assim melhorar seu desempenho. Esta é a primeira instância de aprendizado de máquina que temos registrada na história dos computadores digitais.

Ainda sobre os programas de Samuel e Strachey é importante destacar que são as primeiras instâncias de aplicação de métodos heurísticos na solução de problemas. A solução dada em ambos os casos constituía em montar uma árvore de busca com as possibilidades de jogadas, aplicando um método de atribuição de valor para cada uma delas em função da vantagem que atribuem ao jogador. O programa sempre tenta se antecipar às possibilidades, estabelecendo uma árvore com vários níveis de profundidade, buscando o caminho de maior valor. A Figura 1.3 a seguir representa a árvore de decisão de um jogo de damas a partir de seu ponto inicial. O jogador que começa a partida tem seis possibilidades (seis posições) por onde começar, com a árvore sendo calculada com base nos valores das posições. Situações de baixo valor são descartadas e não há continuidade dos cálculos, que se limitam aos ramos de maior valor.

Figura 1.3 | Árvore de busca e decisão do jogo de damas



Fonte: adaptada de Nilsson (2010).

Em 1956, No Dartmouth College, os pesquisadores John McCarthy, Marvin Minsky, Nathan Rochester e Claude Shannon organizaram a primeira conferência sobre inteligência artificial. Aliás, o termo "inteligência artificial" foi criado para este evento, que deu início aos estudos e pesquisas específicas sobre o assunto. Muitos foram os estudos e artigos que surgiram desse evento, e o campo começou a florescer a partir de seus resultados, ao longo das décadas de 1950 e 1960.

Na década de 1960, mais precisamente em 1963, Leonard Uhr e Charles Vossler publicaram o artigo *A Pattern Recognition Program That Generates, Evaluates, and Adjusts Its Own Operators* (em português Um programa de reconhecimento de padrões que avalia e ajusta seus próprios operadores), descrevendo um programa capaz de aprender usando mecanismos de aquisição de informações, de modificação de seus próprios parâmetros e de adaptação às condições a ele apresentadas, configurando-se em um programa capaz do que passou a ser chamado de *machine learning*, ou aprendizado de máquina.

Em 1965, o pesquisador Joseph Weizenbaum, do MIT, criou o programa Eliza, capaz de compreender sintaticamente um conjunto de sentenças digitadas e de dar respostas, comunicando-se com quem trocasse mensagens. É um dos exemplos mais clássicos de linguagem natural sendo "compreendida" pelo computador.

Em 1972, o pesquisador Alan Comerauer desenvolveu a linguagem Prolog, voltada para a resolução de problemas por meio de aprendizado e interpretação lógica de sentenças.

Em 1974, o pesquisador Paul Werbos descreveu seu algoritmo de retropropagação, que seria fundamental uma década depois no desenvolvimento das redes neurais, e até os dias de hoje são utilizadas em processos de aprendizado de máquina.

Em 1979, o programa BKG, desenvolvido por Hans Berliner, venceu o então campeão mundial de gamão, configurando a primeira vitória expressiva de um computador sobre um humano. Dezoito anos depois, em 1997, o computador Deep Blue, da IBM, derrotaria Garry Kasparov — então campeão mundial do jogo de xadrez —, e mais dezenove anos depois, em 2016, o programa AlphaGo, do Google, derrotaria o campeão mundial do jogo de tabuleiro Go, Lee Sedol.

Na década de 1980, os trabalhos com redes neurais explodiram, culminando em 1989 com a criação do ALVINN (*Autonomous Land Vehicle In a Neural Network*, em português, “automóvel autônomo em uma rede neural”). Uma vez atribuído um destino ao ALVINN, ele conseguia chegar ao seu destino — andando a lentos 10km/h — sem ajuda de um motorista.



Exemplificando

O site mistupid.com disponibiliza o “Jogo da Velha Imbatível”, que utiliza o algoritmo chamado minimax para aprender as regras e melhores estratégias do jogo, tornando-o imbatível. Experimente: você só vai conseguir dois resultados: perder ou empatar, em 100% dos casos. MISTUPID. **Stupid Game: Unbeatable Tic-Tac-Toe**. Disponível em <<http://mistupid.com/games/tictactoe.htm>>. Acesso em: 19 mar. 2018.

Na década de 1990 vimos as pesquisas em inteligência artificial saírem das universidades e passarem a ser exploradas pelas empresas. O caso mais notório surgiu em 1998 e tem um nome conhecidíssimo: Google. A empresa baseada em Mountain View, na Califórnia, especializa-se em aplicar inteligência artificial em tudo o que faz, inclusive na máquina de busca que arremessou a empresa ao sucesso mundial que é hoje (KELLY, 2016).

Os assistentes pessoais Siri (da Apple) Alexa (da Amazon), Cortana (da Microsoft), OK Google (da Google) e Bixby (da Samsung) são exemplos claros de como o treinamento de redes neurais — a expressão mais atual da inteligência artificial — pode beneficiar o público em geral. A cada dia que passa esses assistentes compreendem com maior clareza o que lhes falamos e podem nos auxiliar com maior precisão (KELLY, 2017).

Soluções Avançadas em Inteligência Artificial

Todos esses avanços históricos citados no item anterior nos trazem aos dias de hoje, em que podemos perceber a inteligência artificial começando a fazer parte — ainda que sutilmente — de nossas vidas. A inteligência artificial, segundo Kelly (2017), está iniciando uma revolução que colocará a tecnologia no mesmo patamar da eletricidade no início do século XX. Naquele momento um século atrás, a disponibilização da eletricidade nas tomadas

das residências provocou a eletrificação de inúmeros processos antes manuais. Inventores do mundo todo passaram a utilizar a eletricidade para desempenhar papéis que antes dependiam do esforço humano. A geladeira, as batedeiras e liquidificadores, os aquecedores, os barbeadores elétricos e uma miríade de outros aparelhos foram inventados para fazer com eletricidade —ou seja, de maneira automática — tarefas que antes dependiam de musculatura humana. Kelly (2017) vê o mesmo ocorrendo a partir de agora com a inteligência artificial: empresas como a Google disponibilizarão sistemas de IA por meio das “tomadas” de nossa internet residencial, e inventores do mundo todo vão usar essa tecnologia para facilitar nossa vida de maneiras até o momento impensadas.

Um exemplo disso são os automóveis autodirigidos, ou auto pilotados, em fase de desenvolvimento. Dotados de redes neurais de última geração (que veremos com mais detalhes na Unidade 4), estes automóveis tendem a provocar uma revolução no trânsito, reduzindo drasticamente os acidentes e os congestionamentos.



Pesquise mais

O pesquisador João Fernandes Teixeira publicou em 1990, na coleção Primeiros Passos, o livro “O que é Inteligência Artificial”, contando de maneira simples e didática do que se trata este campo de estudos. Em que pese o livro estar datado em muitos aspectos, a história do desenvolvimento da inteligência artificial em seus primórdios está ali contada com detalhes interessantes, entre as páginas 5 e 10. Vale a leitura.

TEIXEIRA, J. F. **O que é inteligência artificial**, São Paulo: Brasiliense, 1990.

A história do desenvolvimento da inteligência artificial é fascinante. Como a própria história da informática, o campo evoluiu de um conjunto de desejos e teorias para uma área desenvolvida e com vastas realizações, preservando ainda inúmeras possibilidades de desenvolvimento futuro.

Nas próximas seções vamos explorar um pouco mais do desenvolvimento que a inteligência artificial nos disponibiliza nos dias de hoje.

Até lá!

Para apresentar seu relatório com os dados solicitados pela empresa que quer compreender um pouco mais o que é inteligência artificial, você pode compor um documento com os seguintes itens, para os quais poderá encontrar os dados e informações necessárias na presente seção e em suas pesquisas sobre o assunto:

- **Cabeçalho do Relatório**
 - Título do Relatório.
 - Data.
 - Autor/Responsável pelo Relatório.
 - Destinatário do Relatório (nome do executivo da empresa de investimento que solicitou o documento).
 - Número da versão e da revisão do presente documento.
- **Definição de Inteligência Artificial**
 - Aqui você deve procurar construir a definição a partir do conceito de inteligência sob o ponto de vista humano, e chegar à definição de inteligência artificial como a conhecemos nos dias de hoje.
 - Com dados e informações coletados de diversas fontes, você poderá apresentar um conjunto de definições e apontar as vantagens e deficiências de cada uma, explicação para o cliente a que conclusão chegou.
- **Histórico da Inteligência Artificial**
 - A história da inteligência artificial é rica e cheia de detalhes importantes e interessantes. Procure encontrar pontos relevantes nessa história, de maneira a mostrar ao cliente o quanto o assunto tem evoluído e o quanto pode auxiliar no ponto em que estamos atingindo nos dias de hoje.
- **Conclusão**
 - Com base nas informações do relatório, apresente uma conclusão acerca do estado atual da inteligência artificial e de como esta poderá evoluir e auxiliar tanto o cliente como as pessoas em geral.

Pronto! Agora que você já mostrou conhecer o assunto e já apresentou como a inteligência artificial se desenvolveu nessas

últimas décadas, a empresa está entusiasmada com as possibilidades e com os resultados que pode conseguir.

Parabéns!

Avançando na prática

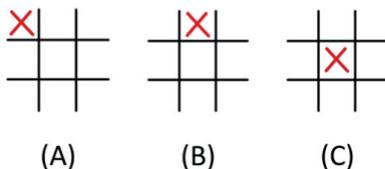
A árvore de decisão do jogo da velha

Descrição da situação-problema

Um dos jogos mais simples para a inteligência artificial é o jogo da velha. Nesse jogo antigo, os jogadores se alternam em preencher um quadro de 3 x 3, buscando formar linhas, colunas ou diagonais com 3 símbolos iguais. Quem conseguir, vence, e se ninguém conseguir, o jogo empata.

Apesar de o quadro em branco permitir 9 posições iniciais, por conta da simetria, podemos reduzir estas posições iniciais a apenas 3, representadas pelos itens A, B e C, na Figura 1.4, a seguir:

Figura 1.4 | As 3 jogadas iniciais possíveis no jogo da velha



Fonte: elaborada pelo autor.

Cada uma dessas jogadas iniciais gera um conjunto de possibilidades (muitas delas também simétricas), e assim podemos construir a árvore de decisão do jogo da velha.

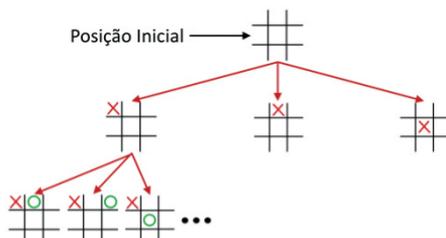
Sua tarefa é construir essa árvore de decisão, eliminando as posições simétricas. Como é essa árvore? Como ficará após sua construção?

Resolução da situação-problema

Para criar a árvore de decisão do jogo da velha, comece pela posição inicial e vá criando ramos, um para cada possibilidade de jogada (sempre descontando as posições simétricas). A cada ramo, pense nas possibilidades de jogada futura, e vá abrindo novos níveis

à medida que forem necessários. A Figura 1.5, a seguir, mostra o nível inicial, o primeiro nível (as três jogadas iniciais possíveis) e o início do segundo nível:

Figura 1.5 | O início da árvore de decisão do Jogo da Velha



Fonte: elaborada pelo autor.

Continue a preencher os ramos e a abrir novos níveis até obter a árvore completa. Não se esqueça de marcar os caminhos mais desejáveis (e os menos desejáveis também) para que consiga saber quais jogadas fazer ao longo do jogo.

Faça valer a pena

1. A inteligência artificial é a capacidade dos sistemas cibernéticos (formados por computadores, softwares, sensores e atuadores) de imitar funções cognitivas dos seres humanos, funções estas que podemos resumir em algumas características fundamentais para o embasamento do conceito em si.

Assinale a alternativa que contém as características fundamentais que devem estar presentes em um sistema computacional para que possa ser caracterizado, de fato, como uma inteligência artificial.

- a) resolução de problemas – criatividade – percepção.
- b) resolução de problemas – aprendizado – criatividade.
- c) sensibilidade – nobreza – ética.
- d) resolução de problemas – aprendizado – percepção.
- e) aprendizado – percepção – criatividade.

2. Observe a afirmação a seguir:

Uma inteligência artificial deve ter a capacidade de utilizar dados e resultados anteriores para melhorar seus parâmetros de funcionamento e, assim, produzir resultados cada vez melhores em execuções futuras de suas funções.

Assinale a alternativa que melhor sintetiza a capacidade descrita na afirmação anterior.

- a) Criatividade.
- b) Processamento de Linguagem Natural.
- c) Representação do Conhecimento.
- d) Resolução de Problemas.
- e) Aprendizado.

3. Em 1965 o pesquisador Joseph Weizenbaum, do MIT (Massachusetts Institute of Technology), criou o programa ELIZA, capaz de compreender sintaticamente um conjunto de sentenças digitadas e de dar respostas, comunicando-se com quem trocasse mensagens com ele.

Assinale a alternativa que sintetiza melhor a capacidade exibida pelo programa ELIZA.

- a) Representação do Conhecimento.
- b) Processamento de Linguagem Natural.
- c) Empatia para com o interlocutor.
- d) Aprendizado baseado em experiência.
- e) Raciocínio aeroespacial.

Seção 1.2

Metodologias de inteligência artificial

Diálogo aberto

Você já parou para pensar em como a inteligência artificial auxilia nossas vidas? Pode até parecer um assunto distante de nosso dia a dia, mas quando observamos mais de perto, podemos encontrar vários exemplos de como temos a vida facilitada por sistemas de IA. Um exemplo simples, mas bastante interessante, é o processo de recomendações da Netflix, que usa de um sistema inteligente baseado em métodos estatísticos para recomendar filmes e séries para os assinantes do serviço. A cada programação a que assistimos ou que escolhemos e desistimos por não ser o que esperamos, o sistema adiciona os dados coletados por nosso comportamento e ajusta seus parâmetros internos, aumentando as chances de que sua próxima sugestão nos seja mais agradável. E mais: este mesmo mecanismo é utilizado para que a Netflix produza novos filmes e séries. Um exemplo disso é a série *Stranger Things*, que foi produzida porque a Netflix identificou interesse de seus assinantes pelo período da década de 1980, por séries de mistério e teorias da conspiração e também a predileção por histórias com crianças como personagens principais. E quem identificou todas essas tendências? O sistema de IA da Netflix, baseado em aprendizado estatístico.

Mas será que esse é o único tipo de solução de IA? Na presente sessão, veremos que não.

Nessa segunda etapa da consultoria que sua organização vai prestar à empresa que quer conhecer um pouco mais acerca da inteligência artificial para melhorar o desempenho de seus investimentos, você deve identificar a inteligência artificial mostrando ao investidor que se trata de um campo amplo de estudos, com várias técnicas e tecnologias sendo desenvolvidas e aplicadas. Deve, também, mostrar os dois lados da discussão sobre inteligência artificial, de acordo com o modelo do quarto chinês, ampliando para o cliente o conceito de "inteligência". O que é inteligência, e o que caracteriza a matéria em questão como sendo artificial?

O que a faz diferente de outros tipos de inteligência? Quais são suas classificações?

Na presente sessão veremos as subáreas da inteligência artificial, suas perspectivas e aplicações, entenderemos o que é e para que serve o Teste de Turing e a analogia do Quarto Chinês, e conheceremos alguns dos esforços iniciais em linguagem natural com os programas ELIZA e PARRY, bem como dos avanços dos *chatterbots* (robôs de conversação).

Bons estudos!

Não pode faltar

Depois de entendermos como a inteligência artificial deixou a mitologia e os sonhos da antiguidade, passou pelos primeiros progressos na aurora do computador, na década de 1940, e chegou até os dias de hoje, é hora de compreendermos um pouco mais sobre como se classificam as soluções de IA nos dias de hoje, bem como sobre alguns dos conceitos mais básicos que embasam filosoficamente os estudos de IA.

Subáreas da Inteligência Artificial

É importante observar que, como afirma Nilsson (2010), não há um paradigma unificado em IA, com vários enfoques sendo explorados na atualidade. Isso porque os sistemas de computação e informação são versáteis e as combinações possíveis de hardware e software permitem que arquiteturas e soluções cada vez mais criativas, potentes e versáteis sejam criadas.

Uma das maneiras de se dividir os estudos de IA é proposta tanto por Rosa (2011) quanto por Russell e Norvig (2013), classificando os estudos de IA em quatro grandes campos de pesquisa:

- A IA que pensa como ser humano.
- A IA que pensa racionalmente.
- A IA que age como ser humano.
- A IA que age racionalmente.

Observe que as quatro divisões visam a estudar o pensamento e a ação (que pode ser sintetizada na tomada de decisão e na ação

física, quando apropriado). Em ambos os casos, opõem-se duas ideias: a centralização em como nós, seres humanos, fazemos (pensamos e agimos) e em como é racional fazer (o que nem sempre é nossa especialidade).

Já Shapiro (2016) detalha três enfoques para os estudos de IA, a saber:

- **Psicologia Computacional** – As técnicas são usadas para criar sistemas que exibem o comportamento humano.
- **Filosofia Computacional** – As técnicas são usadas para criar uma mente computacional, que consiga pensar por si mesma.
- **Ciência Computacional** – As técnicas são usadas para criar um sistema computacional, que consiga agir por si mesmo.

Segundo Shapiro (2016), em uma máquina ou computador, o comportamento humano, aliado à capacidade de pensar e à capacidade de agir, juntos, formam o que conhecemos por “Inteligência Artificial”.

Em que pese ambas as abordagens apresentadas indicarem formas proveitosas para dividirmos e estudarmos a IA, adotaremos uma forma mais pragmática, mais atenta às técnicas para categorizarmos os estudos da área. Por essa mecânica temos três campos de estudo, compreendendo três formas diferentes e complementares de encararmos as pesquisas em IA, que são:

- **Enfoque Simbólico** – O enfoque simbólico dos estudos de IA foi o principal curso de ação dessa ciência desde o seu início, ainda na década de 1940, até a década de 1980, segundo Haugeland (1989). Foi este próprio autor quem cunhou para o enfoque simbólico o apelido de “a boa e velha inteligência artificial”. O enfoque simbólico baseia-se no fato de que nós, seres humanos, manipulamos símbolos e reagimos a eles. Não só no nível mais baixo das palavras, mas no próprio campo dos conceitos, nós trabalhamos e reagimos a símbolos que nos são apresentados, ou seja, o que depreendemos mentalmente de nossas observações.
- o **Principal Ferramenta** – Manipuladores de linguagem natural, como no caso da linguagem de programação PROLOG, que permite a criação de lógicas próprias a partir de entradas simbólicas e de manipulações desses símbolos.

- **Principais Exemplos** – Os sistemas especialistas, desenvolvidos sobre bases de conhecimento são os exemplos clássicos. Em especial, os sistemas de apoio ao diagnóstico médico são ferramentas desenvolvidas com essa técnica. O DiagnosisPro (disponível em: <https://diagnosispro.com.cutestat.com/>. Acesso em: 26 abr. 2018) é um sistema especialista em uso há mais de duas décadas, auxiliando como uma segunda opinião em diagnósticos médicos.
- **Enfoque Sub Simbólico** – Em que pese a mente humana ser capaz de processar símbolos (conceitos, palavras, etc.), o cérebro humano funciona em um patamar que lida com elementos muito mais simples, naquilo que Nilsson (1998) convencionou chamar de patamar sub simbólico. De fato, nosso cérebro lida com sinais elétricos e catalisadores químicos e, de maneira semelhante, o computador lida com sinais elétricos que codifica como “0” e “1”. Os elementos sub simbólicos são agrupados de maneira a formarem o que compreendemos como símbolos. O sistema computacional – em especial o software residente nesse sistema –, trabalhando com os elementos de um sistema sub simbólico, tem a capacidade de os agrupar e deles extrair os significados mais complexos, que são as respostas aos problemas que se propõem a resolver.
 - **Principal Ferramenta** – A principal ferramenta do enfoque sub simbólico é a rede neural, que emula o funcionamento dos neurônios em um cérebro humano, imitando as conexões entre neurônios e fortalecendo ou enfraquecendo estas conexões à medida que o conjunto se aproxima ou se afasta das respostas esperadas.
 - **Principais Exemplos** – Rede Neural LSTM (*Long Short Term Memory*), criada e disponibilizada gratuitamente pelo pesquisador Andrej Karpathy, da Universidade de Stanford.
- **Enfoque Estatístico** – Russell e Norvig (2013) afirmam que a partir da década de 1990 várias ferramentas matemáticas passaram a ser usadas em processos computacionais, dado que os avanços em hardware e software já permitiam sua adoção. As ferramentas de processamento estatístico

passaram a ser alvo de pesquisas em IA, e algumas soluções interessantes começaram a surgir. Estas ferramentas têm a desvantagem de serem focadas em problemas específicos, não havendo flexibilidade para sua adaptação. Contudo, não há como negar que os problemas a que elas se propõem resolver, o fazem com rapidez e precisão.

- o **Principal Ferramenta** – As ferramentas de *data mining* (mineração de dados) contam com o apoio de mecanismos estatísticos para aumentarem a precisão de seus resultados.
- o **Principais Exemplos** – O mecanismo (proprietário) de *data mining* da Netflix é um exemplo de como a IA, com base em processos estatísticos, pode aprender as preferências de seus usuários e sugerir-lhes novos programas a assistir, bem como sugerir aos executivos da empresa quais novos programas produzir.



Assimile

Os *chatterbots*, ou programas de conversação com o usuário por meio de linguagem natural, são exemplos de IA simbólica, enquanto as redes neurais são exemplos de IA sub simbólica.

Aplicações e Perspectivas da Inteligência Artificial

Kelly (2017) nos mostra que a inteligência artificial nos cerca cada dia mais, disponibilizando a capacidade de aprendizado e adaptação a mecanismos, dispositivos e sistemas que anteriormente eram completamente incapazes de se adaptar às nossas necessidades. De fato, Kelly (2017) afirma que a IA se encontra nos dias de hoje na mesma situação que a energia elétrica se encontrava no início do século XX. Segundo o autor, a invenção da lâmpada elétrica por Thomas Edison e a padronização da corrente alternada por Nikola Tesla impulsionaram e facilitaram a disponibilização de energia elétrica nos lares e nas empresas. Essa facilidade de acesso incentivou inventores do mundo todo a criarem dispositivos que realizavam tarefas antes impossíveis, ou que dependiam da musculatura humana e animal para serem desempenhadas. O refrigerador, a batedeira, a máquina de lavar e um sem-número

de novas invenções vieram a facilitar a vida de pessoas e negócios no mundo todo. Pois segundo Kelly, estamos à beira da mesma revolução, agora com respeito à IA: a disponibilidade globalizada da Internet, aliada à explosão na capacidade de processamento e barateamento do hardware, bem como aos avanços nas técnicas de software, deixam a IA ao alcance de todos nós. Já começam a surgir os primeiros resultados dessa revolução, com os assistentes pessoais (Siri, Cortana, Alexa, OK Google, etc.), mas esse é apenas o começo.

Leça (2017) aponta seis grandes áreas em que a IA já começa a ser aplicada, com resultados importantes e promissores:

- 1. Agricultura** – Aqui a IA se vale da internet das coisas (conexão de dispositivos simples à Internet por meio de endereçamento IP) e da análise preditiva (técnica estatística que permite a previsão probabilística de eventos com base na análise dos dados pregressos disponíveis) para permitir o planejamento e a melhoria do desempenho na produção agrícola
- 2. Transporte e mobilidade** – Aqui não há como deixar de mencionar os carros autônomos, atualmente em teste por várias montadoras e centros de pesquisa no mundo todo. Esta tecnologia pode ser aplicada tanto aos veículos de passeio quanto ao transporte rodoviário de cargas, bem como ao transporte ferroviário. No que diz respeito à geolocalização, a IA já é usada por sistemas pessoais de GPS, disponíveis em smartphones, com base em sistemas de apoio de IA, indicando o melhor caminho aos usuários, com base em cálculos feitos em tempo real sobre dados colhidos dos próprios usuários do sistema.
- 3. Saúde e diagnósticos** – Desde os tempos dos antigos sistemas especialistas da década de 1990, o setor de saúde vem abraçando as tecnologias que auxiliam nos diagnósticos, e a IA está sempre no centro dessas tecnologias. Um exemplo interessante nos fornece Upbin (2013), apresentando o caso do computador da IBM chamado “Watson” (em homenagem ao primeiro CEO da empresa), adotado pelo centro de pesquisas em câncer Memorial Sloan-Kettering, em Nova York, que vem auxiliando no diagnóstico de câncer. O autor afirma ainda que o executivo do projeto criador do Watson, Manoj Saxena,

apresenta como dado que 90% dos enfermeiros que utilizam o Watson seguem suas sugestões.

4. **Cadeia de Fornecimento (Supply Chain)** – Da mesma maneira que a IA pode auxiliar no planejamento da produção agrícola, aumentando sua eficiência, a IA também pode ser aplicada nos processos de cadeia de fornecimento em uma empresa ou indústria qualquer, melhorando a precisão das manutenções preditivas, da previsão de necessidade de insumos e da automatização da logística, sempre tão crucial à indústria.
5. **Manufatura** – A robotização de linhas de montagem já não é novidade há muito tempo; máquinas e robôs vêm realizando trabalhos repetitivos desde a Revolução Industrial. Contudo, a capacidade de aprender e se adaptar a novas condições de atuação atribui nova gama de possibilidades aos robôs da manufatura.
6. **Assistentes Virtuais** – Os assistentes virtuais pessoais estão disponíveis a qualquer um que tenha um smartphone e queira saber se há algum restaurante chinês nas imediações, o que está passando no cinema, ou se vai chover amanhã. Estes sistemas dependem pesadamente de IA para compreender nossa fala e nossa intenção, treinando-se a acertar a cada vez que com eles interagimos. Mas não para por aí: os *chatterbots* (ou *chatbots*, como também são chamados) vem sendo utilizados em atendimentos on-line e atendimentos telefônicos via URA (Unidade de Resposta Audível) em serviços de atendimento telefônico.

Esses são apenas alguns dos principais exemplos do uso de IA para melhorar nossas vidas. Leça (2017) afirma ainda que o mercado de IA para máquinas inteligentes está em alta, e deve movimentar por volta de US\$15,2 bilhões em 2019.



Refleta

Onde, no seu dia a dia, você consegue enxergar o uso de inteligência artificial? Sua vida seria mais simplificada ou complicada sem o apoio dessas instâncias de IA?

Teste de Turing, Quarto Chinês

Em seu artigo seminal sobre inteligência artificial, Turing (1950) propõe um teste simples para avaliar se uma máquina exhibe comportamento inteligente. Este teste é conhecido, desde então, como Teste de Turing, e vem sendo utilizado periodicamente para avaliar o estado do desenvolvimento da inteligência artificial.

O Teste de Turing é baseado em uma antiga brincadeira de salão, como o próprio Turing (1950) explica: é o “Jogo da Imitação”— que aliás foi o título de filme de 2014 que conta parte da história de Alan Turing, em especial seu trabalho como criptologista na Segunda Guerra Mundial. O Jogo da Imitação consiste em um homem e uma mulher, separados em uma sala, com um árbitro enviando-lhes perguntas e recebendo as respostas — ambos por meio de mensagens escritas ou pelo teclado/tela de um sistema computacional. O objetivo do árbitro é determinar, depois de um número finito de perguntas, quem é o homem e quem é a mulher, enquanto ambos buscam confundi-lo sem mentir.

Com base nesse antigo jogo, Turing (1950) propõe uma variação: em vez de um homem e uma mulher, o árbitro conversa com um ser humano e com uma máquina, e seu objetivo é definir quem é quem. O próprio autor simplifica o jogo, opondo um árbitro e uma entidade desconhecida, com o objetivo de identificar se essa entidade é um ser humano ou não.

Caso o árbitro não consiga identificar o ser humano na primeira variação, ou identificar se seu colóquio é com um computador ou com um ser humano na versão simplificada, Turing afirma que a máquina no primeiro caso e a entidade (seja um ser humano ou uma máquina) no segundo caso, são entidades inteligentes. O artigo de Turing tem sido discutido desde sua publicação, e direcionou alguns dos avanços em IA que temos à disposição nos dias de hoje (e que veremos mais à frente dessa mesma seção).

Russel e Norvig (2013) afirmam que o Teste de Turing é útil para identificar uma forma **operacional** de inteligência, isto é, não se preocupa com o mecanismo utilizado, mas sim com seu resultado.

O Teste de Turing ainda é usado nos dias de hoje para avaliar a capacidade dos sistemas computacionais de imitar o comportamento humano. Exemplo disso é o Prêmio Loebner, um concurso anual

realizado no Reino Unido, que premia os sistemas cibernéticos capazes de confundir árbitros não-especialistas (KEEDWELL, 2014).

Os programas construídos para tentar passar no Teste de Turing dependem da manipulação de linguagem natural (o idioma em que o teste é conduzido) e são considerados sistemas simbólicos (como já definimos anteriormente nessa mesma seção). Essa manipulação simbólica deu origem à principal crítica ao Teste de Turing: o argumento do Quarto Chinês, proposto como experimento mental por Searle (1980).

O argumento do Quarto Chinês, segundo Searle (1980), é o seguinte: imagine que seja possível construir um programa de computador que consiga passar no Teste de Turing mantendo conversações em chinês. Um árbitro chinês é incapaz de distinguir o comportamento desse programa do comportamento de uma pessoa que efetivamente fala chinês. Agora imagine que no lugar do programa, nós criamos um quarto com uma pequena janela por onde podem entrar e sair papéis escritos. Nessa sala colocamos um indivíduo com papel, lápis, borrachas e gabinetes para armazenar seus escritos e resultados. Esse indivíduo seguirá as mesmas instruções do programa que passou no Teste de Turing em chinês, isto é, receberá textos escritos em caracteres chineses, seguirá as regras do programa e gerará respostas, que escreverá em papel e devolverá pela janela. Este indivíduo — desde que siga exatamente os comandos do programa que passou no Teste de Turing — também passará no mesmo teste, isto é: convencerá o árbitro que é uma entidade inteligente no idioma chinês. O detalhe é que o resultado será o mesmo ainda que o indivíduo não entenda absolutamente nada de chinês.

Este argumento opõe o que Searle (1980) chama de **IA Forte** e **IA Fraca**:

- Uma máquina (hardware mais software) que *literalmente* compreenda chinês é uma IA Forte. Nesse caso, existe o que o autor chama de **intencionalidade**.
- Uma máquina que simule compreender chinês (como no caso do Quarto Chinês) é uma IA Fraca. Nesse caso, falta **intencionalidade**.

Em resposta a Searle (1980), Rey (1986) apresenta o que chama de **Argumento Sistêmico**, afirmando que o indivíduo no quarto chinês é apenas a CPU do sistema. Em que pese esse indivíduo não

saber chinês, o sistema como um todo — formado pelo quarto, a janela, as regras do programa, o material de escrita, e tudo mais — *sabe* chinês. O indivíduo é apenas uma peça do todo, e o todo (o sistema completo) efetivamente sabe chinês.

Rey (1986) separa o conhecimento do Quarto Chinês dos componentes desse quarto da mesma maneira que a psicologia e a neurociência separam a mente do cérebro e seus componentes celulares, químicos e elétricos.



Pesquise mais

O primeiro programa a passar oficialmente no Teste de Turing se chama “Eugene Goostman”, e imita o comportamento de um garoto de 13 anos. O feito ocorreu em 2014, e você pode ler mais a respeito no artigo a seguir: ROHR, Altieres. Computador convence juizes de que é o garoto de 13 anos em ‘teste de Turing’. **G1**, 2014. Disponível em: <<http://g1.globo.com/tecnologia/noticia/2014/06/computador-convence-juizes-que-e-garoto-de-13-anos-em-teste-de-turing.html>>. Acesso em: 31 mar. 2018).

ELIZA, PARRY, Chatterbots

O Teste de Turing deu início a uma corrida para a criação de máquinas (hardware mais software) que conseguissem confundir árbitros. Whitby (1996) cita a criação do programa ELIZA, em 1966 e do programa PARRY, em 1972, como dois dos principais resultados gerados pelo Teste de Turing.

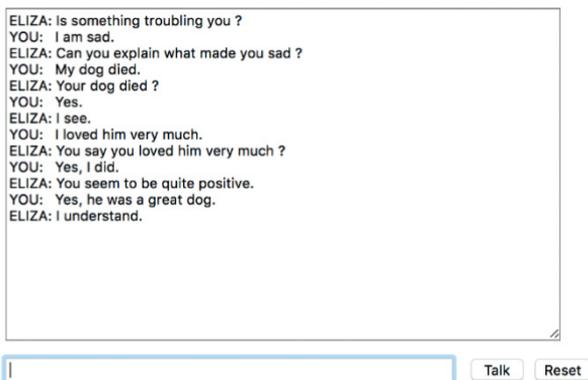
Russell e Norvig (2013) apontam o ELIZA como sendo um programa de computador criado em 1966 pelo pesquisador Joseph Weizenbaum, do MIT.

O programa utiliza um mecanismo rudimentar de reconhecimento de padrões que permite ao programa identificar alguns elementos gramaticais de uma sentença e responder com base nesse elemento gramatical. O programa em si não sabe o que está acontecendo, não consegue interpretar as sentenças, mas, conseguindo identificar (ainda que de forma imprecisa) sujeito, verbo, predicado e alguns elementos como objetos e advérbios, pode compor uma resposta coerente.

A maneira encontrada por Weizenbaum para fazer isso foi determinando que ELIZA é uma psiquiatra durante uma sessão de

análise. A função do psiquiatra durante uma sessão é questionar o paciente, permitindo que este se expanda sobre as dificuldades psicológicas que enfrenta e, de maneira limitada, ELIZA consegue fazer exatamente isso. A Figura 1.6 a seguir exemplifica um diálogo real entre um ser humano (o autor) e o programa ELIZA:

Figura 1.6 | Exemplo de diálogo com o programa ELIZA



Fonte: elaborada pelo autor.

O diálogo representado na Figura 1.6 foi gerado por uma versão do ELIZA hospedada no site Masswerk, e o programa em si pode ser testado no endereço <<http://www.masswerk.at/elizabot/>>. O programa funciona para o idioma original para que foi criado, o inglês. O diálogo apresentado pode ser assim traduzido:

ELIZA: Tem alguma coisa te incomodando?

VOCÊ: estou triste.

ELIZA: Você pode explicar o que te deixou triste?

VOCÊ: Meu cachorro morreu.

ELIZA: Seu cachorro morreu?

VOCÊ: Sim.

ELIZA: Entendo.

VOCÊ: Eu o amava muito.

ELIZA: Você diz que o amava muito?

VOCÊ: Sim, eu amava.

ELIZA: Você parece ter bastante certeza.

VOCÊ: Sim, ele era um ótimo cachorro.

ELIZA: Eu compreendo.

Algumas características do ELIZA merecem especial atenção:

- O programa não aprende, ou seja, não armazena informações para adicionar ao seu algoritmo de processamento.
- O programa apenas identifica o papel gramatical de alguns termos chave.
- O programa contém um conjunto limitado de respostas pré-programadas que adicionam elementos identificados para incitar a continuidade do diálogo.
- O programa não compreende o teor do diálogo em si, apenas imita esta compreensão.

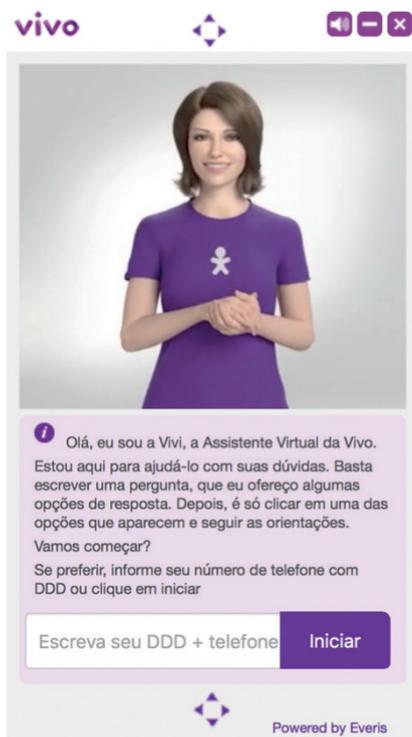
Ainda que diante de tantas limitações, o ELIZA representou a primeira instância de um programa que se aproxima da capacidade de conversar em linguagem natural. Seus rudimentos de processamento simbólico de termos gramaticais têm sido usados desde então em versões mais robustas e eficientes de programas de bate-papo automatizado.

Como evolução dos conceitos primeiramente codificados no ELIZA, Russell e Norvig (2013) apresentam o PARRY, criado em 1972 pelo psiquiatra Kenneth Colby, da Universidade de Stanford. Diferentemente da personalidade de um psiquiatra exibida pelo ELIZA, o programa PARRY exibe a personalidade de um paciente esquizofrênico. Após a criação do programa, ainda em 1972, PARRY foi submetido a dois grupos de psiquiatras, um deles interagindo diretamente com o programa e com pacientes reais por meio de teclado e tela, e outro lendo as transcrições desses diálogos. O resultado foi que em apenas 48% dos casos os psiquiatras foram capazes de identificar quem era o paciente real e quem era o PARRY (COLBY et al., 1972).

Os programas ELIZA e PARRY têm sido fundamentais no desenvolvimento de programas de manipulação simbólica de linguagem natural, e historicamente foram a base que deu origem a inúmeros projetos de *chatterbots* (ou *chatbots*) em uso hoje em dia. Diferentemente dos programas que visam vencer o Teste de Turing, nos dias de hoje os *chatterbots* são usados inclusive no atendimento ao cliente. Um exemplo brasileiro desse tipo uso

para as IA de tratamento simbólico de linguagem natural é a Vivi, a atendente eletrônica dos serviços de suporte da empresa VIVO, que dá apoio ao cliente na hora de encontrar soluções referentes a problemas que enfrenta com os serviços oferecidos pela empresa. A Figura 1.7 ilustra a tela inicial do programa Vivi:

Figura 1.7 | Tela inicial da atendente eletrônica Vivi, da VIVO



Fonte: <<http://vivo.com.br>>. Acesso em 31 mar. 2018.



Exemplificando

O *Mitsuku Chatbot* é um premiado programa de conversa em linguagem natural que já venceu o prêmio Loebner em 2013, 2016 e 2017. O programa consegue manter conversações em inglês sobre vários assuntos, sendo praticamente indistinguível de um ser humano. Você pode acessar o *Mitsuku Chatbot* e testá-lo. WORSWICK, Steve. **Mitsuku Chatbot**. Disponível em: <<http://www.square-bear.co.uk/mitsuku/turing/>>. Acesso em: 31 mar. 2018. Se tiver dificuldades com

o idioma inglês, pode traduzir suas falas para o inglês e as falas de Mitsuku para o português usando o tradutor do Google. Disponível em: <<https://translate.google.com.br>>. Acesso em: 31 mar. 2018.

Agora que entendemos um pouco mais os tipos de inteligência artificial, suas aplicações no mundo real e as principais decorrências do Teste de Turing, estamos prontos para algumas atividades extras e para a próxima fase do nosso estudo sobre inteligência artificial.

Até a próxima!

Sem medo de errar

Com o objetivo de auxiliar a empresa de investimentos sobre a validade da inteligência artificial como ferramenta para aumentar os ganhos, você deve, nessa etapa, apresentar à empresa como é feita a classificação dos tipos de IA. Para isso, pontue:

- Não há um paradigma unificado no que concerne a essa classificação.
- Das várias classificações em existência, aquela que busca identificar a forma de funcionamento das IAs é bastante útil, pois também auxilia a identificar seus usos.
- As principais categorias são:
 - o Enfoque Simbólico – A mente humana funciona dessa maneira: atribuindo significado a símbolos e manipulando estes símbolos. Os sistemas especialistas e os chatbots são exemplos desse tipo de IA.
 - o Enfoque Sub simbólico – O cérebro humano (que é diferente da mente humana) funciona dessa forma, manipulando sinais bem menos significativos que os símbolos com que estamos acostumados. As redes neurais são exemplos desse tipo de IA.
 - o Enfoque Estatístico – Uso de ferramentas estatísticas como elementos de aprendizagem para as IAs que funcionam nesse modelo. Os sistemas de previsão e sugestão operam com base nesse mecanismo.

Você também deve, em seu documento, pontificar o que John Searle (1980) chamou de IA Forte e IA Fraca por meio de seu argumento do Quarto Chinês, ou seja, pode-se argumentar que a imitação da inteligência — em que há apenas processamento mecânico, mas não há intencionalidade — é inferior à inteligência dita “real”, em que há intencionalidade.

Por fim, seu documento deve chegar a uma conclusão para o cliente: será que a IA Fraca é menos efetiva no que concerne a investimentos financeiros do que a IA Forte? O ponto é mostrar, por meio de todos os exemplos disponíveis de IA nos dias de hoje, que IA Forte é uma ideia ainda a ser atingida, pois todos os exemplos de IA são análogos ao Quarto Chinês: todos emulam inteligência, mas nenhum exibe as características de intencionalidade que ainda é inerente apenas ao ser humano. Mesmo assim, os exemplos de IA Fraca já permitem excelentes análises e bons resultados quanto aos investimentos, e já pode ser usada para esse propósito.

Avançando na prática

Os *chatbots* brasileiros

Descrição da situação-problema

Estudando o mercado de atendimento, um investidor decidiu que gostaria de patrocinar a criação de uma empresa de suporte por meio de inteligência artificial, uma vez que este tipo de solução pode ser escalado (ou seja, pode crescer sua oferta de ordem de grandeza em ordem de grandeza, com pouco aumento nos custos), com lucratividade potencial excelente. Ele buscou você, e juntos estão procurando que tipo de solução utilizar.

Para não começar do zero, ambos decidiram que é importante conhecer as soluções oferecidas aqui no Brasil, de forma a encontrar o nicho mais adequado para a oferta de serviços. Desta forma, você deve realizar uma busca, testar serviços e gerar um relatório para seu sócio investidor, permitindo que ambos tomem uma decisão embasada sobre qual escolha fazer.

Como proceder?

Resolução da situação-problema

A primeira coisa a fazer é realizar uma pesquisa em alguma máquina de busca à escolha para encontrar os *chatbots* brasileiros. Uma boa chave de busca, inclusive, é “*chatbots* brasileiros”.

Faça a busca e acesse os links que julgar mais adequados, compondo um relatório com as informações coletadas. Um bom formato para esse relatório é:

- Nome do *Chatbot*.
- Empresa que fornece o serviço
 - o Dados de acesso Web e contato (físico) da empresa.
- Empresa que desenvolveu o serviço (se não for a mesma)
 - o Dados de acesso Web e contato (físico) da empresa.
- Descrição dos serviços oferecidos pelo chatbot
 - o Tipo de serviço.
 - o A quem se destina.
 - o Como é oferecido (Web, smartphone, telefone).
- Nota para o serviço
 - o Valor de 0 (inútil ou absolutamente impreciso) a 10 (perfeito).
 - o Observações sobre o serviço
 - Limitações.
 - Peculiaridades.

Procure adicionar pelo menos cinco serviços de *chatbot* diferentes ao seu relatório.

Pronto! Você e seu sócio investidor já podem decidir que caminho seguir e que tipo de serviço querem oferecer ao mercado.

Boa sorte, e sucesso!

Faça valer a pena

1. Observe na coluna da esquerda os tipos de IA pelo enfoque comportamental e as definições na coluna da direita:

| | | | |
|-----|-----------------------------|---|---|
| I | Psicologia Computacional | A | As técnicas são usadas para criar um sistema computacional, que consiga agir por si mesmo. |
| II | Filosofia Computacional | B | As técnicas são usadas para criar sistemas que exibem o comportamento humano. |
| III | Ciência Computacional | C | As técnicas são usadas para criar uma mente computacional, que consiga pensar por si mesma. |

Assinale a alternativa a seguir que faz a correta correspondência entre as colunas:

- a) I-A, II-B, III-C.
- b) I-A, II-C, III-B.
- c) I-B, II-A, III-C.
- d) I-B, II-C, III-A.
- e) I-C, II-B, III-A.

2. Observe a definição a seguir:

Em que pese a mente humana ser capaz de processar símbolos (conceitos, palavras, etc.), o cérebro humano funciona em um patamar que lida com elementos muito mais simples. De fato, nosso cérebro lida com sinais elétricos e catalisadores químicos, e de maneira semelhante, o computador lida com sinais elétricos que codifica como "0" e "1".

Assinale a alternativa que descreve corretamente o enfoque de IA a que o texto se refere

- a) Enfoque simbólico.
- b) Enfoque sub simbólico.
- c) Enfoque estatístico.
- d) Enfoque digital.
- e) Enfoque probabilístico.

3. Analise as asserções a seguir:

I. Uma máquina (hardware mais software) que literalmente compreenda chinês é uma Inteligência Artificial Forte.

PORQUE

II. Uma Inteligência Artificial Forte exibe **intencionalidade**.

Analisando-se as asserções apresentadas, conclui-se que:

- a) As duas afirmações são verdadeiras, e a segunda justifica a primeira.
- b) As duas afirmações são verdadeiras, e a segunda não justifica a primeira.
- c) A primeira afirmação é verdadeira, e a segunda é falsa.
- d) A primeira afirmação é falsa, e a segunda é verdadeira.
- e) As duas afirmações são falsas.

Seção 1.3

Inteligência artificial na prática

Diálogo aberto

Caro aluno, antes de chegar a essa disciplina de Inteligência Artificial, você já aprendeu várias coisas acerca da informática, correto? Aprendeu, entre outras coisas, a programar e a criar algoritmos, transformando-os em programas que realizam tarefas importantíssimas em nosso dia a dia. O sistema de semáforos de uma cidade, por exemplo, deve ser coordenado de maneira que os referidos semáforos permitam o fluxo mais desimpedido possível do trânsito, e a duração dos sinais verdes e vermelhos podem – e devem – variar, dependendo da hora do dia ou da noite. Fluxo mais intenso demanda mais tempo de verde para vias de maior acesso, e por aí vai. Ocorre que, salvo exceções, as condições de fluxo são as mesmas para cada horário nos vários dias da semana, e um algoritmo simples de mudança dos tempos de verde e vermelho, dependendo do horário, já garante a melhor situação possível diante do fluxo que a cidade apresenta. Mas e quanto a outros tipos de problema em que o algoritmo pode aprender com o contexto e apresentar resultados cada vez melhores a cada execução? O tradutor automático do Google, por exemplo, começou apresentando traduções com base em um conjunto básico (e rígido) de regras, que nem sempre apresentavam os melhores resultados. Porém, a equipe do Google permite que o usuário corrija a tradução apresentando palavras, expressões e construções gramaticais mais precisas, de acordo com seu conhecimento. A cada tradução, as sugestões dos usuários são avaliadas e incorporadas ao algoritmo, e hoje, vários anos após o início do serviço, as traduções apresentam uma qualidade visivelmente melhor em relação ao início do serviço. Este é o poder da Inteligência Artificial. Nessa seção vamos justamente aprender que tipos de problemas podem ser solucionados com a IA, e em que ela é aplicada nos dias de hoje.

Nessa primeira etapa da consultoria a ser prestada à companhia de investimentos que está contratando sua empresa especializada em inteligência artificial, e que deseja conhecer um pouco mais

acerca de seu funcionamento para melhorar o desempenho de seus investimentos, você deve caracterizar o conceito de “aprendizado”, bem como mostrar os vários métodos pelos quais um sistema computadorizado aprende, e como este aprendizado é aferido. O que é aprender? Como podemos estimar o aprendizado em um sistema artificial? Apresente seus resultados sob a forma de um relatório.

Nessa seção você vai aprender quais problemas podem ser solucionados pela inteligência artificial, vai compreender como se dá a classificação desses problemas nesse campo, vai descobrir o que são agentes inteligentes, além de conhecer algumas das principais aplicações e casos reais de inteligência artificial.

Bons estudos!

Não pode faltar

A inteligência artificial auxilia muito na resolução de problemas que enfrentamos na atualidade. Mas será que se aplica em qualquer tipo de problema? Bem, a princípio se — como vimos na primeira seção desta unidade — o objetivo da IA é emular a inteligência humana, e a inteligência nos permite, em princípio, resolver qualquer tipo de problema, então a resposta seria um sonoro SIM. Em princípio, a IA poderia ser utilizada para resolver qualquer problema, dado o exposto, mas daí a efetivamente utilizarmos a IA em qualquer situação vai uma distância muito longa. Na verdade, como veremos a seguir, em que pese podermos usar a IA em inúmeras situações, não será em todas que esse uso é o mais adequado.

Vejamos essa questão um pouco mais a fundo.

Problemas adequados para a Inteligência Artificial

Imagine-se pensando em como resolver computacionalmente o seguinte problema: dado um número natural, calcular o resultado fatorial desse número. Para resolver esse problema, temos que, primeiramente, identificar que o resultado fatorial de um número natural é a cadeia de multiplicações desse número com todos os seus antecessores até chegar em 2. Sim, antes que você possa confirmar essa informação, poderíamos ir até 1, mas como multiplicar qualquer coisa por 1 dá como resultado ela mesma, para que complicar? Melhor economizar esse passo computacional.

Para exemplo, tomemos como número de entrada o 4, e então teríamos:

$$4! = 4 \times 3 \times 2 = 24$$

Quando pensamos na resolução computacional desse problema, podemos tomar dois caminhos: criar um laço que vá do número dado, em ordem decrescente, até parar em 2, multiplicando o resultado pelo número em questão, ou então usar uma função recursiva:

```
Fatorial (inteiro N)
  Se N = 1
    Retorna N
  Caso contrário
    Retorna (N * Fatorial(N-1))
Fim
```

Qualquer um dos dois caminhos possíveis que o programador escolher levará ao resultado desejado: o cálculo do valor fatorial de um número que seja dado como valor de entrada.

Ocorre que o cálculo fatorial de um número é independente do cálculo fatorial de qualquer outro número. Com excessão do método usado para o cálculo, cada cálculo é diferente, e um cálculo não influencia em outro. Em suma: não há aprendizado para o cálculo de um número qualquer para o cálculo posterior de outro número qualquer.

Outro ponto importante: o cálculo de um número fatorial é um resultado exato, objetivo, fruto de uma fórmula definida e não ambígua (gera sempre o mesmo resultado para um dado número).

Rothlauf (2011) nos conta que os tipos de problema resolvidos diretamente, por meio de métodos ou fórmulas são de natureza **determinística**. Estes tipos de problema são resolvidos por métodos tradicionais de matemática, lógica e computação algorítmica.

Contudo, há alguns tipos de problema cuja resposta não é atingível por meio de uma fórmula simples. Estes problemas dependem de um enfoque prático, baseado em tentativa e erro, em muitos casos gerando respostas errôneas que podem ser melhoradas com mais prática e rodadas adicionais de análise. Estes problemas são ditos por Rothlauf (2011) como sendo de natureza heurística. Considere os seguintes exemplos:

- **Problema das 8 Rainhas** – Colocar 8 rainhas (peças do jogo de xadrez que se movem e atacam qualquer casa diretamente atingível na horizontal, na vertical, e nas duas diagonais) em um tabuleiro de xadrez, sem que qualquer uma delas seja atacada por qualquer uma das demais.
- **Jogo de Quebra-Cabeça** – Os quebra-cabeças clássicos, compostos de dezenas ou centenas de peças, que quando colocadas nas posições corretas se encaixam e formam uma figura coerente (uma fotografia ou uma pintura, por exemplo).
- **Problema do Caixeiro-Viajante** – Dado um mapa, determinar o caminho mais curto para visitar todas as cidades representadas.

Nenhum desses problemas é resolvido por meio de uma fórmula que dá a resposta imediata. Todos eles dependem de tentativa e erro, inúmeras análises e comparações, até que se chegue a um resultado. No caso do quebra-cabeça, esse resultado correto é único (cada peça pode ocupar apenas um local correto, estando errada em qualquer outra posição). Contudo, o problema do Caixeiro-Viajante pode permitir mais de uma solução (caso haja mais de um conjunto de caminhos que gerem a mesma distância mínima), e o problema das 8 Rainhas certamente tem mais de uma solução correta (há várias configurações diferentes em que nenhuma rainha é atacada por nenhuma outra).

Jaokar (2017) nos mostra que os problemas mais propensos a uma solução pelas técnicas de inteligência artificial são de natureza heurística, com a principal característica de que os sistemas realizam algum tipo de aprendizado à medida que passam pelo processo prático de tentativa e erro (método heurístico). Em resumo, os problemas mais adequados aos métodos e técnicas de inteligência artificial englobam:

- **Heurística** – Busca de soluções por meios práticos e iterativos, que se aproximam cada vez mais da solução com uso de técnicas que envolvem tentativa e erro.
- **Aprendizado** – Armazenamento e uso de dados coletados em rodadas anteriores, transformação desses dados em informações úteis, e uso dessas informações em instâncias futuras do processo de resolução (que equivale à geração de conhecimento a partir das informações deduzidas).



Por que a resolução de problemas para os quais há fórmulas por meio de inteligência artificial não é adequada? Que decorrências poderíamos imaginar para a resolução, digamos, de uma equação de segundo grau por meio de técnicas de inteligência artificial?

Classificação de Problemas em Inteligência Artificial

Jaokar (2017) especifica 12 categorias de problemas que envolvem heurística e aprendizado, para as quais a inteligência artificial é especialmente adequada na provisão de soluções, a saber:

- 1. Problemas de especialidade em domínio de conhecimento**
 - Problemas que envolvem raciocínio em algum domínio complexo de conhecimento. Problemas em que a máquina deve aprender algum corpo complexo de conhecimento, e demonstrar expertise no assunto.
 - **Exemplos:** sistemas especialistas jurídicos e médicos, capazes de aprender sobre o assunto e de realizar consultas em questões a respeito do referido assunto, para auxiliar os especialistas
- 2. Problemas de extensão de domínio de conhecimento** – Problemas em que a máquina não só deve aprender um domínio complexo e demonstrar expertise, mas também sugerir novos caminhos de ação, não presentes no corpo de conhecimento, com base no que aprendeu.
 - **Exemplos:** sistemas especialistas médicos capazes de indicar medicamentos conhecidos no tratamento de doenças nas quais ainda não foram usados, com base na bioquímica conhecida dos medicamentos e de suas reações no corpo humano.
- 3. Problemas de planejamento complexo** – Problemas que lidam com o planejamento de tarefas complexas, envolvendo grandes quantidades de ações logísticas e agendamentos coordenados.
 - **Exemplos:** Planejamento de composição e entrega de pedidos compostos por elementos de vários fornecedores, chegando em momentos diferentes ao armazém central, com data limite para chegar ao cliente final.

4. **Problemas de comunicação** – Problemas que envolvam melhorar a qualidade da comunicação entre partes, sejam essas partes humanas ou não.
 - **Exemplos:** problemas que exigem tradução de conceitos a partir de uma língua que o usuário não fala. Problemas em que ações devem ser tomadas pelo computador a partir de comandos verbais do usuário.
5. **Problemas de percepção** – Problemas que envolvem o aprendizado a partir da percepção de elementos do ambiente que permitam a tomada de ações e decisões.
 - **Exemplos:** Sistema de sensores e ações em veículos autônomos (que se dirigem sem a presença de um condutor).
6. **Problemas de IA corporativa** – Problemas que envolvem reengenharia da corporação a partir de conhecimento e aprendizado realizado por meio da miríade de dados gerados e disponíveis.
 - **Exemplos:** Otimização de processos corporativos por meio da IA.
7. **Problemas cognição em ERP e datawarehousing** – Problemas envolvendo o reconhecimento de padrões em dados não estruturados (provenientes de anos de uso de soluções de ERP ou de grandes conjuntos de dados oriundos de datawarehouses corporativas).
 - **Exemplos:** Identificação de novos clientes e/ou novos negócios a partir de expertise da empresa, com base em resultados obtidos de seus sistemas corporativos.
8. **Problemas que impactam domínios devido às consequências de segunda ordem** – Problemas decorrentes do uso de IA para a solução de outros problemas. Esta é uma categoria teórica que prevê o surgimento de problemas como decorrência do uso de IA, que se espera que a própria IA auxilie na resolução.
 - **Exemplos:** Boa parte do transporte de seres humanos e de cargas é rodoviário. A adoção de veículos autônomos deve impactar a segurança, em função da redução drástica de acidentes uma vez que a maioria deles é causado por desatenção, segundo Kelnar (2016). Deve

também impactar o mercado de trabalho, uma vez que efetivamente remove a necessidade para motoristas de carga e de passageiros.

9. Problemas que podem se beneficiar no futuro próximo com melhores algoritmos e maior disponibilidade de IA – Kelly (2017) versa sobre esta questão quando afirma que estamos à beira de uma explosão de soluções de IA, uma vez que a disponibilidade de acesso está se tornando global, o hardware melhora em potência e cai em preço e os algoritmos estão sendo paulatinamente melhorados. Em um futuro próximo, dispositivos de uso diário e processos corriqueiros se beneficiarão da IA, como os utensílios domésticos se beneficiaram da eletricidade a partir do início do século XX.

- **Exemplos:** As taxas de erro em reconhecimento de fala por parte de sistemas de IA caíram pela primeira vez abaixo dos 5,9% em 2018; segundo o autor, um valor equivalente ao apresentado pela média dos seres humanos.

10. Problemas de evolução de sistemas especialistas – Os sistemas especialistas foram alguns dos primeiros a utilizar algoritmos que envolvem aprendizado, e continuam em evolução até os dias de hoje. Ocorre que a própria IA indica caminhos pelos quais esses sistemas podem ser melhorados, apresentando resultados melhores a cada geração.

- **Exemplos:** O IBM Watson é um sistema (hardware mais software) capaz de identificar respostas a perguntas com base em varreduras feitas em tempo real em bases de conhecimento a ele apresentadas na forma de textos. Seus algoritmos mais recentes permitem que o conhecimento acumulado seja passível de análises e varreduras, o que permite ao Watson também aprender com o que já aprendeu, aprimorando seu conhecimento.

11. Problemas de reconhecimento de padrões sequenciais longos – Problemas que envolvem o reconhecimento de padrões em conjuntos de dados que são sequenciais em um longo período de tempo. Este é um domínio da IA que ainda engatinha.

- **Exemplos:** Os sistemas de dirigibilidade de veículos autônomos devem identificar que, quando a faixa externa

de uma rodovia se torna uma diagonal voltada para o centro, sinaliza que em um próximo momento a faixa que está mais à direita terminará. O sistema deve identificar e diferenciar este acontecimento de uma curva à esquerda, por exemplo.

12. Problemas em análise emocional – Problemas envolvendo a avaliação de expressões sob forma de texto e fala em busca de elementos emocionais e/ou sentimentais do indivíduo que os expressa. Trata-se de um campo em que a subjetividade permeia a mensagem que se busca extrair. O estado de espírito de um indivíduo transborda em suas expressões escritas e faladas, e a IA pode auxiliar a identificar com precisão este estado de espírito.

- **Exemplos:** Este é um campo em que ainda não há soluções para além dos laboratórios dos institutos de pesquisa.



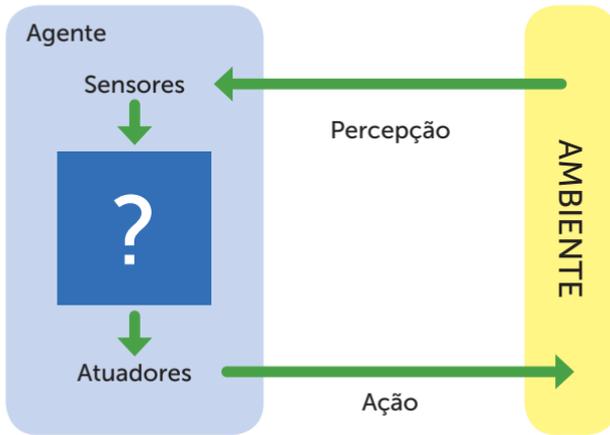
Exemplificando

Moura, Andrade e Kikuchi (2005) apresentam um excelente artigo acerca de um sistema de IA voltado à tomada de decisões em saúde com base na pesquisa em dados armazenados em datawarehouse, em um excelente exemplo do tipo 7 (vide conteúdo apresentado) de problemas resolvíveis por meio da IA. O sistema utiliza um agente de tomada de decisão alicerçado sobre uma vasta gama de dados armazenados em datawarehouse, e as decisões são usadas por médicos e consultores a título de segunda opinião, reduzindo a probabilidade de diagnósticos errôneos emitidos por esses profissionais.

Agentes Inteligentes

Segundo Russell e Norvig (2013) um agente é uma entidade que percebe o ambiente por meio de sensores e interage com este ambiente por meio de atuadores. A Figura 1.8, a seguir, ilustra o conceito de agente:

Figura 1.8 | O agente e o ambiente



Fonte: adaptada de Russell e Norvig (2013).

Nesse contexto, temos os seguintes elementos (RUSSELL e NORVIG, 2013):

- **Agente** – Entidade que percebe o ambiente e age sobre este ambiente.
- **Ambiente** – Contexto de atuação do agente.
- **Percepção** – Capacidade de coletar dados a partir do ambiente.
- **Ação** – Capacidade de provocar mudanças no ambiente.

Um conceito adicional é apresentado pelos autores como sendo a sequência de dados percebidos, que é o histórico total ou parcial de dados percebidos pelo agente.

Nesse contexto, a ação tomada pelo agente é dependente da percepção ou, mais precisamente, da sequência de dados percebidos pelo agente em algum momento. O agente tem, segundo os autores, a possibilidade de escolher seu curso de ação de acordo com os dados percebidos do ambiente. Quando especificamos todos os cursos de ação possíveis para um agente, de acordo com todas as possibilidades de percepção que a ele podem ser apresentadas, estamos definindo o agente em sua totalidade.

Dentro dessas possibilidades e características, Russell e Norvig (2013) nos mostram que qualquer entidade pode se enxergar como um agente, mas que nem sempre esse enfoque é interessante. Uma calculadora eletrônica, por exemplo, pode ser encarada como

sendo um agente que tem como ação exibir “4” toda vez que percebe a entrada “2 + 2”. Contudo, em que pese esse enfoque ser verdadeiro, encarar a calculadora como agente não nos auxilia a compreender efetivamente seu funcionamento. Segundo os autores, a IA opera no universo em que os agentes têm recursos computacionais significativos (possibilidades de processamento e ação) e o ambiente requer ações não triviais (ações não óbvias).

Um agente é dito **racional** quando toma as ações corretas, isto é, quando sua tabela interna de ações para cada percepção (ou sequência de percepções) é corretamente preenchida. Mas o que é considerado uma ação correta? Os autores definem ação correta de maneira simples: ação correta é aquela que gera consequências desejáveis sob o ponto de vista de quem programa o agente. A cada ação realizada, o ambiente assume um estado diferente. No caso de cada um desses estados aproximar o ambiente do estado ideal, e então atingir esse estado ideal, o agente terá tomado as ações certas.

Russell e Norvig (2013) apontam para o fato de que se deve levar em conta os estados intermediários e o estado final do **ambiente**, e nunca do **agente**. Esta distinção é fundamental, uma vez que se o agente pudesse decidir o que é certo ou errado, poderia iludir-se de que toma as decisões acertadas sempre. Quando passamos esta responsabilidade para o ambiente e adicionamos o observador/programador a esse ambiente, podemos ter uma ideia mais coerente de comportamento adequado por parte do agente. Com base nos estados atingidos pelo ambiente, e comparando esses estados com os estados ideais, podemos definir uma medida de desempenho para efetuarmos avaliações.

De maneira geral, enfatizam os autores, o mais adequado é criar medidas de desempenho de acordo com os estados esperados para um ambiente em particular, de maneira que seja determinado o que se espera em termos de comportamento do agente por meio desse conjunto de estados do ambiente.

O comportamento de um agente racional, em um dado momento, depende de quatro características (RUSSELL; NORVIG, 2013):

- A medida de desempenho que define o critério de sucesso.
- O conhecimento prévio do ambiente.

- As ações que o agente é capaz de realizar.
- A sequência armazenada de percepções até o momento.

De posse dessas quatro características, os autores definem formalmente o agente racional como aquele que, para cada sequência possível de percepções, seleciona a ação que maximiza sua medida de desempenho, dadas as evidências disponibilizadas pela sequência de percepções e qualquer conhecimento prévio do agente.

O **agente inteligente**, segundo os autores, é o agente racional capaz de realizar dois tipos especiais de ação, além de ter todo o comportamento de um agente racional:

- **Autonomia** – O agente inteligente é capaz de tomar decisões quanto às ações a tomar sem necessidade de ação externa do programador.
- **Aprendizado** – O agente inteligente é capaz de acumular conhecimento que seja útil em determinar suas ações, tornando-as cada vez mais eficientes, segundo a medida de desempenho adotada.



Assimile

Um agente inteligente é um agente racional, isto é, que sempre busca o melhor desempenho e que é dotado das capacidades de aprendizado e autonomia.

Aplicações e Casos Reais de Inteligência Artificial

Ao longo dessas três primeiras seções, já mencionamos vários casos reais de IA com os quais convivemos diariamente e que estão sendo desenvolvidos e melhorados a cada nova iteração de hardware e software. Entre eles, podemos detalhar alguns, pela importância que têm ou terão em nosso dia a dia:

- **Assistentes Pessoais** – Siri (Apple), Alexa (Amazon), Cortana (Microsoft), Assistente Google ou OK Google (Google). Sistemas de reconhecimento de fala, capazes de reconhecer a voz do dono do dispositivo por meio de reconhecimento de timbre, reconhecer palavras e frases em linguagem natural, transformá-las em comandos reconhecíveis pelo dispositivo (smartphone, tablet ou computador pessoal) e executar esses

comandos, traduzindo as respostas – quando pertinente – novamente para a linguagem natural. Na prática, temos a impressão de estarmos conversando com o computador e termos nossos pedidos realizados por eles.

- **Sistemas autônomos de controle de veículos** – Segundo Khaidem (2018), já em franca utilização em drones, por exemplo. São capazes de tomar conta de toda a dirigibilidade e das manobras para evitar obstáculos e compensar mudanças nas condições de vento, deixando para o piloto humano apenas a responsabilidade de apontar o caminho a ser seguido ou a posição de parada desejada. O drone usa toda a experiência e o aprendizado realizado em laboratório e em condições prévias de teste para navegar com segurança nas condições que o ambiente lhe apresenta. Este mesmo mecanismo de aprendizado e reação já é utilizado em testes de veículos automotores, segundo Knight (2017).
- **IBM Watson** – Sistema especialista capaz de analisar dados em grandes bases desestruturadas e extrair conhecimento que lhe permite tomar decisões e responder perguntas sobre vários domínios de conhecimento. Já em uso em sistemas de apoio a diagnósticos médicos.

Conhecemos um pouco mais dos problemas mais afeitos à resolução por IA, bem como a definição formal dos agentes inteligentes que formam a base da IA.



Pesquise mais

A integração de inteligência artificial aos drones já produz resultados interessantes. Leia mais sobre o assunto no artigo intitulado *Drones com Inteligência Artificial - futuro da tecnologia*.

DRONE VISUAL. **Drones com Inteligência Artificial** - futuro da Tecnologia. [S. l.], fev. 2018. Disponível em <<https://www.dronevisual.com/single-post/2018/02/19/DRONES-INTEGRADOS-COM-INTELEGENCIA-ARTIFICIAL-SÃO-O-FUTURO-DA-TECNOLOGIA>>. Acesso em: 14 abr. 2018.

Sem medo de errar

Com o objetivo de auxiliar a empresa de investimentos, nessa etapa temos que, em primeiro lugar, caracterizar o que é aprendizado.

Você pode, para esta tarefa, buscar exemplos na vida real, caracterizando como se dá o aprendizado. Para tanto, você pode identificar exemplos simples, por exemplo: como crianças pequenas aprendem atividades triviais, ou ainda como treinamos nossos animais de estimação a realizar truques ou a ir ao banheiro no lugar certo.

Em seguida, podemos extrair os elementos que compõem o aprendizado, a saber:

- Acúmulo de dados relativos à iteração sendo realizada pelo indivíduo/dispositivo.
- Uso de dados acumulados para ajustar o comportamento do indivíduo/dispositivo em iterações futuras.

À empresa de investimentos, deve ser explicado como isso se dá nos dispositivos computacionais: adicionando os dados coletados ao algoritmo de processamento, levando-os em conta no cálculo da resposta procurada.

Como se trata de um **processo heurístico**, isto é, de busca da resposta por meio de mecanismos que envolvem tentativa e erro, a cada tentativa os parâmetros de entrada levam o dispositivo computacional a se aproximar mais da resposta ótima.

O dispositivo computacional em questão deve ser descrito como um agente racional, e para tanto você deve descrever os elementos componentes de um agente:

- **Agente** – Entidade que percebe o ambiente e age sobre este ambiente.
- **Ambiente** – Contexto de atuação do agente.
- **Percepção** – Capacidade de coletar dados a partir do ambiente.
- **Ação** – Capacidade de provocar mudanças no ambiente.

Para que este agente seja considerado racional, você deve caracterizar os elementos que devem estar presentes para que essa classificação seja possível:

- A medida de desempenho que define o critério de sucesso.
- O conhecimento prévio do ambiente.
- As ações que o agente é capaz de realizar.
- A sequência armazenada de percepções até o momento.

De posse desses elementos, o próximo passo é a definição formal de um agente racional que é aquele que, para cada sequência possível de percepções, seleciona a ação que maximiza sua medida de desempenho, dadas as evidências disponibilizadas pela sequência de percepções e qualquer conhecimento prévio do agente.

Para finalizar, você deve caracterizar o agente inteligente, que é aquele dotado de duas capacidades adicionais:

- **Autonomia** – O agente inteligente é capaz de tomar decisões quanto às ações a tomar sem necessidade de ação externa do programador.
- **Aprendizado** – O agente inteligente é capaz de acumular conhecimento que seja útil em determinar suas ações, tornando essas ações cada vez mais eficientes, segundo a medida de desempenho adotada.

O que se busca, portanto, para auxiliar nos investimentos da empresa, é a criação de um agente inteligente autônomo e capaz de aprendizado, fechando o círculo dessa rodada de caracterização de IA e capacitando a empresa a dar o próximo passo em direção da adoção de soluções que lhe sejam úteis na maximização dos lucros por meio de investimentos financeiros.

Avançando na prática

Otimizando o cálculo de fatoriais

Descrição da situação-problema

Quando pensamos em uma calculadora capaz de calcular o valor fatorial de um número, nos remetemos aos algoritmos capazes de gerar esse resultado, seja por meio dos laços de repetição, seja pelo método recursivo (ambos brevemente descritos nessa sessão). Contudo, se consideramos um sistema computacional capaz de armazenar dados por tempo indefinido (o que não ocorre em uma calculadora), podemos usar o conceito de aprendizado também nesse tipo de

cálculo, tornando cada cálculo de um valor fatorial potencialmente mais rápido que os cálculos anteriores, sem perder a precisão.

Descreva em linhas gerais como esse processo pode ocorrer. Por que podemos considerá-lo um processo que envolve aprendizagem?

Resolução da situação-problema

O cálculo de um número fatorial "n" implica na multiplicação deste número por todos os seus predecessores, até chegarmos em 2. Por definição, temos alguns valores fixos:

- $0! = 1$
- $1! = 1$

A partir de 2, basta multiplicar o valor em questão pelos predecessores, como já afirmado.

Ocorre que quando calculamos o valor fatorial de um número qualquer, calculamos por tabela o valor fatorial de todos os seus predecessores. Por exemplo, o cálculo de $4!$ é:

$$4! = 4 \times 3 \times 2 \times 1$$

Ocorre que $3!$ é calculado nesse processo, uma vez que $3!$ é:

$$3! = 3 \times 2 \times 1$$

Podemos, portanto, a cada cálculo fatorial, armazenar os cálculos intermediários, formando uma tabela. Um exemplo é a Tabela 1.1, a seguir, com cálculos para todos os valores fatoriais de 1 até 4:

Tabela 1.1 | Valores e resultados fatoriais desses valores

| Valor | Fatorial |
|-------|----------|
| 4 | 24 |
| 3 | 6 |
| 2 | 2 |
| 1 | 1 |
| 0 | 1 |

Fonte: elaborada pelo autor.

Com esse processo temos duas otimizações excelentes para o processo de cálculo de fatorial:

- Se o próximo número a ser calculado for igual ou menor ou igual ao maior número da tabela, o resultado é imediato: basta ler o número da tabela.
- Se o próximo número a ser calculado é maior que o maior número da tabela, as multiplicações deverão levar em conta o último (maior) resultado obtido. Levando em conta a Tabela 1.1, podemos calcular, por exemplo, o valor de **6!**. Assim:

$$6! = 6 \times 5 \times 4!$$

$$6! = 6 \times 5!$$

$$6! = 6 \times 120$$

$$6! = 720$$

Uma vez calculado o valor de **6!** devemos atualizar a tabela com os valores de **6!** (720) e de **5!** (120). Ou seja, não só o cálculo fica menor, uma vez que fazemos menos multiplicações, mas também facilitamos a vida de quem for fazer os próximos cálculos de números maiores.

Esta é uma maneira simples de adicionar o elemento de aprendizado a um cálculo determinístico. Obviamente que esta solução exige mais memória e espaço de armazenamento, o que a torna inviável para uma calculadora eletrônica.

Faça valer a pena

1. Há alguns tipos de problemas cuja resposta não é atingível por meio de uma fórmula simples. Estes problemas dependem de um enfoque prático, baseado em tentativa e erro, em muitos casos gerando respostas errôneas que podem ser melhoradas com mais prática e rodadas adicionais de análise.

O método que envolve um enfoque prático, baseado em tentativa e erro leva o nome de:

- a) determinística.
- b) heurística.
- c) análise.
- d) inteligência artificial.
- e) tentativa e erro.

2. Os sistemas especialistas médicos capazes de indicar medicamentos conhecidos no tratamento de doenças para os quais ainda não foram usados, com base na bioquímica conhecida dos medicamentos e de suas reações no corpo humano são exemplos de soluções para um tipo específico de problema que pode ser resolvido por métodos de inteligência artificial.

A qual tipo de problema de inteligência artificial o texto apresentado se refere?

- a) Problemas de especialidade em domínio de conhecimento.
- b) Problemas de extensão de domínio de conhecimento.
- c) Problemas de planejamento complexo.
- d) Problemas de percepção.
- e) Problemas de comunicação.

3. Um agente racional é aquele agente que, para cada sequência possível de percepções, seleciona a ação que maximiza sua medida de desempenho, dadas as evidências disponibilizadas pela sequência de percepções e qualquer conhecimento prévio do agente.

Para que um agente racional seja considerado inteligente, deve exibir que características adicionais?

- a) Inteligência e onisciência.
- b) Aprendizado e onisciência.
- c) Autonomia e onisciência.
- d) Criatividade e autonomia.
- e) Aprendizado e autonomia.

Referências

- CHAPMAN, S. Blaise Pascal (1623–1662) Tercentenary of the calculating machine. **Revista Nature**, n. 150, out. 1942, p. 508-509.
- COLBY, K. M. et al. Turing-like indistinguishability tests for the validation of a computer simulation of paranoid processes. **Artificial Intelligence**, v. 3, p. 199-221, 1972.
- FRANCO, C. R. **Inteligência Artificial**. Londrina: Editora e Distribuidora Educacional, 2014.
- HAUGELAND, J. **Artificial Intelligence: The Very Idea**. Cambridge: MIT Press, 1989.
- JAOKAR, A. Twelve Types of Artificial Intelligence Problems. **Data Science Central**, [S. l.], 05 jan. 2017. Disponível em: <<https://www.datasciencecentral.com/profiles/blogs/twelve-types-of-artificial-intelligence-ai-problems>>. Acesso em: 09 abr. 2018.
- JOURDAIN, R. **Música, Cérebro e Êxtase**. São Paulo: Objetiva, 1997.
- KEEDWELL, E. The Loebner Prize, a Turing Test competition at Bletchley Park. **The Exeter Blog**. Exeter, dez. 2014. Disponível em: <<https://blogs.exeter.ac.uk/exeterblog/blog/2014/12/08/the-loebner-prize-a-turing-test-competition-at-bletchley-park/>>. Acesso em: 30 mar. 2018.
- KELLY, K. **Inevitável: As 12 Forças Tecnológicas que Mudarão o Nosso Mundo**. São Paulo: HSM, 2017.
- KELNAR, D. The fourth industrial revolution: a primer on Artificial Intelligence (AI). **Medium**, [S. l.], 02 dez. 2016. Disponível em: <<https://medium.com/mmc-writes/the-fourth-industrial-revolution-a-primer-on-artificial-intelligence-ai-ff5e7ffcae1>>. Acesso em 12 abr. 2018.
- KHAIDEM, R. The future of drones in Artificial Intelligence. **DZone**, [S. l.], 03 jan. 2018. Disponível em: <<https://dzone.com/articles/the-future-of-drones-in-artificial-intelligence>>. Acesso em: 13 abr. 2018.
- KNIGHT, W. Finally, a Driverless Car with Some Common Sense. **MIT Technology Review**, [S. l.], 20 set. 2017. Disponível em: <<https://www.technologyreview.com/s/608871/finally-a-driverless-car-with-some-commonsense/>>. Acesso em: 13 abr. 2018.
- LANDSTEINER, Norbert. **Elizabot**. [S.l., s.d.]. Disponível em: <<http://www.masswerk.at/elizabot/>>. Acesso em: 30 mar. 2018.
- LEÇA, I. Inteligência Artificial em ação – seis aplicações da IA. **SAP News Center**. São Paulo, jan. 2017. Disponível em: <<https://news.sap.com/brazil/2017/01/18/inteligencia-artificial-em-acao-seis-aplicacoes-da-ia/>>. Acesso em: 28 mar. 2018.
- LIMA, I., PINHEIRO, C. A. M., SANTOS, F. A. O. **Inteligência Artificial**. Rio de Janeiro: Campus, 2014.
- MCCORDUCK, P. **Machines Who Think**. Boston: A. K. Peters Ltd., 2004.
- MOURA, A. O.; ANDRADE, D. S.; KIKUCHI, W. K. Inteligência Artificial Aplicada no Auxílio a Tomadas de Decisões na Área de Saúde. In: SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA, 2., 2005, Site AEDB, 2005, Resende. **Anais...** Resende,

Associação Educacional Dom Bosco, 2005. Disponível em: <https://www.aedb.br/seget/arquivos/artigos05/293_Inteligencia%20Artificial%20Aplicada%20no%20Auxilio%20a%20Tomadas%20de%20Decisooes%20na%20Area%20de%20Saude.pdf>. Acesso em: 13 abr. 2018.

NEWELL, A.; SHAW, J. G., SIMON, H. A. **The Process of Creative Thinking**. In: GRUBER, H. E., TERRELL G., WERTHEIMER, M. (Editores). **Approaches to Creative Thinking**. Nova York: Atherton, 1963.

NILSSON, N. J. **Artificial Intelligence – a new synthesis**. Nova York: Morgan Kaufmann Publishers, 1998.

_____. **The quest for artificial intelligence – a history of ideas and achievements**. Cambridge: Cambridge University Press, 2010.

RAJPURKAR, P., HANNUN, A., HAGHPANAHI, M., BOURN, C., NG, A. **Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks**. Ithaca, 2017. Disponível em: <<https://arxiv.org/pdf/1707.01836.pdf>>. Acesso em: 04 mar. 2018.

REY, G. What's Really Going on in Searle's "Chinese Room". **Philosophical Studies**, 1986, v. 50, n. 169, p. 85-89.

ROSA, J. L. G. **Fundamentos da Inteligência Artificial**. Rio de Janeiro: LTC, 2011.

ROTHLAUF, F. **Design of Modern Heuristics, Principles and Application**. Nova York: Springer, 2011.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial**. Rio de Janeiro: Campus, 2013.

SCHAEFFER, J. **One Jump Ahead: Challenging Human Supremacy in Checkers** Nova York: Springer, 2009.

SEARLE, J. R. Minds, brains, and programs. **Behavioral and Brain Sciences**. Cambridge, v. 3, n. 3, p. 417-424, set. 1980. Disponível em: <<https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/minds-brains-and-programs/DC644B47A4299C637C89772FACC2706A>>. Acesso em: 30 mar. 2018.

SHAPIRO, S. C. **Encyclopedia of Artificial Intelligence**. Nova York: John Wiley & Sons, 2016.

TEJEDA, J. F.; CHERUBIN, F. **O que é a inteligência?** Rio de Janeiro: Lumens Juris, 2016.

TURING, A. M. Computing machinery and intelligence. **Mind, A Quarterly Review of Psychology and Philosophy**, v. 59, n. 236, p. 433-460, out. 1950.

UPBIN, B. IBM's Watson Gets Its First Piece Of Business In Healthcare. **Forbes**, p. 37-39, fev. 2013. Disponível em: <<https://www.forbes.com/sites/bruceupbin/2013/02/08/ibms-watson-gets-its-first-piece-of-business-in-healthcare/#590d7d5d5402>>. Acesso em: 29 mar. 2018.

WHITBY, B. The Turing Test: AI's Biggest Blind Alley? In: MILLICAN, P.; CLARK, A. **Machines and Thought: The Legacy of Alan Turing**. Mind Association Occasional Series, 1. Oxford: Oxford University Press, p. 53–62, 1996.

Resolução de problemas em inteligência artificial

Convite ao estudo

Várias empresas de porte do setor de tecnologia, entre elas o Google, a Apple e a Tesla, estão se dedicando a criar veículos autônomos, isto é, capazes de levar passageiros e/ou cargas de um local a outro sem a necessidade de um motorista humano. Como isso é possível? Bem, o sistema é complexo, mas passa pela necessidade de que o mesmo consiga identificar os vários elementos que possam surgir em uma rodovia, classificá-los e, por meio desse processo, que chamamos de reconhecimento de padrões, tomar decisões de acordo com as inúmeras situações que podem ser encontradas no “mundo lá fora”. Na presente unidade vamos conhecer vários dos métodos e mecanismos que permitem atingir esses resultados, auxiliando você a conhecer, compreender e ser capaz de resolver problemas em inteligência artificial utilizando as metodologias apresentadas. Estudando adequadamente esta unidade e resolvendo os problemas propostos de acordo com as metodologias apresentadas, você vai fixar esse conhecimento adquirido.

Uma grande empresa nacional do setor de serviços de saúde tem uma base de dados de milhões de usuários que, ao longo de várias décadas, receberam cuidados de saúde — consultas, exames, procedimentos cirúrgicos, internações, etc. Essa empresa foi instruída por uma consultoria a utilizar a imensa base de dados a sua disposição para melhorar seus serviços, bem como a eficiência de seus negócios, aumentando sua efetividade tanto para os como para os acionistas. Você e sua empresa de desenvolvimento de soluções de inteligência artificial foram chamados para liderar um projeto para utilizar os dados da empresa no sentido de encontrar novas soluções,

serviços e produtos de saúde. O desafio é grande, mas a empresa deposita grande confiança em você e seu time. Como encontrar padrões que façam sentido médico e de negócios em uma base tão grande? Como "garimpar" conhecimento em dados que não estão organizados e abrangem uma enorme gama de assuntos e propósitos díspares?

Na primeira seção desta unidade serão apresentadas questões e métodos probabilísticos e estatísticos.

Na segunda seção desta unidade serão apresentadas questões e métodos pertinentes às buscas.

Na terceira seção desta unidade veremos questões acerca da consciência, da superinteligência e das limitações da inteligência artificial.

Estude com cuidado todo o material apresentado e bom proveito!

Seção 2.1

Tipos e representação de problemas

Diálogo aberto

Olá!

Um dizer antigo, geralmente proferido por uma pessoa que não sabia o que fazer em determinado momento, era algo na linha de "não sei se caso, se entro para o exército, ou se compro uma bicicleta". Essa frase indica total falta de direcionamento, como pode ser facilmente percebido, mas por que tanta estranheza (ainda que com bom humor)? Por duas razões simples: nenhuma dessas três ações tem conexão com as demais, e qualquer das ações tomadas, além de pressupor condições diferentes na vida presente do indivíduo, o levará a situações completamente diferentes, exigindo outras decisões completamente díspares.

As probabilidades de os eventos ocorrerem nem sempre são tão disjuntas como no exemplo acima; muitas vezes, há uma certa dependência entre elas. A ocorrência de um evento pode facilitar ou dificultar a ocorrência de outro evento. Essa interdependência entre probabilidades é o cerne de um método muito usado em inteligência artificial e um dos principais assuntos que veremos nesta seção juntamente aos métodos estatísticos e à análise do progresso de aprendizagem em inteligência artificial.

Agora, para trabalharmos esses conceitos, lembre-se que você e seus colegas foram chamados a participar de um projeto importante de uma empresa nacional do setor de saúde que tem como objetivo melhorar as operações e os serviços oferecidos aos clientes por meio de análise e uso de dados históricos coletados ao longo de décadas acerca de tudo que se refere aos pacientes tratados. Nessa fase do projeto, você deverá sugerir formas de se utilizar os métodos para classificar os dados e apontar quais as vantagens de se buscar padrões nesse enorme manancial de informações e conhecimento em potencial. Apresente seus resultados na forma de um relatório para a empresa de serviços de saúde.

Nesta seção você vai entrar em contato com os métodos probabilísticos, com os métodos baseados em classificadores e com os métodos de aprendizado estatístico; além disso, vai entender um pouco mais acerca da mensuração do progresso de aprendizado. E então? Tudo pronto para mais essa aventura?

Então vamos lá!

Não pode faltar

Depois de compreendermos vários aspectos históricos e filosóficos acerca da Inteligência artificial e de iniciarmos nossos estudos teóricos acerca do assunto, é chegada a hora de nos aprofundarmos nessa teoria. Na presente seção vamos entender alguns dos métodos utilizados nos agentes que lhes permitem aprender e gerar resultados importantes para o nosso dia a dia.

Métodos Probabilísticos

Uma característica de um agente inteligente, quando as estudamos na seção 1.3, é a capacidade de reagir ao ambiente com base em um conjunto de regras. A cada situação percebida, o agente reage de certa maneira. Ocorre que, como afirmam Russell e Norvig (2013), a correspondência entre todos os estados possíveis do ambiente com regras pré-estabelecidas é impossível, na prática, em função do crescimento exponencial de recursos de processamento, memória e armazenamento. Esse crescimento exponencial se dá, segundo os autores, em função do crescimento das regras a serem codificadas no agente. Em suma, à medida que o ambiente cresce em complexidade, torna-se cada vez mais difícil (leia-se: custoso) preparar o agente para cada uma das situações que possa encontrar.

O resultado é que os agentes existem em uma situação de **incerteza**, que se dá pela capacidade de observar o ambiente apenas de maneira parcial, pelo não determinismo do ambiente, ou por uma combinação de ambos.



Exemplificando

Um sistema de análise e previsão do tempo (meteorologia) exemplifica com clareza a questão da incerteza: o sistema é complexo e o

ambiente em que o sistema navega é composto por dezenas de variáveis complexas e interdependentes. Seria impossível mapear todos os estados que esse complexo conjunto de variáveis poderia assumir, portanto, é um ambiente permeado pela incerteza.

Para resolver essa situação, um método bastante eficaz é o método probabilístico, que leva em consideração as probabilidades de cada ação levar a um resultado esperado ou, pelo menos, a uma configuração do ambiente mais favorável ao agente (RUSSEL e NORVIG, 2013).

Os autores afirmam que o método probabilístico é representado pelas Redes Bayesianas, baseadas no trabalho de Thomas Bayes, estatístico e filósofo inglês do século XVIII.

Para entender as Redes Bayesianas, segundo Russell e Norvig (2013), temos que tomar probabilidades de eventos. Suponhamos dois eventos:

- $P(L)$ = Probabilidade de se ganhar na loteria
- $P(C)$ = Probabilidade de chover

Se fôssemos representar esses dois eventos, teríamos que levar em conta que essas duas probabilidades são independentes, isto é, pode chover ou não chover, e isso não influencia em alguém ganhar na loteria. Em outras palavras, se sabemos a priori se choveu ou não, esse conhecimento não influencia na probabilidade de se ganhar na loteria. A Figura 2.1, a seguir, representa essas duas probabilidades em formato:

Figura 2.1 | Rede Bayesiana de duas probabilidades distintas



Fonte: elaborada pelo autor.

A Figura 2.1 acima representa uma Rede Bayesiana para as duas possibilidades. Para ela, o cálculo da probabilidade combinada de se ganhar na loteria $P(L)$ e de (ao mesmo tempo) chover $P(C)$ é o

produto de ambas as probabilidades, e o conhecimento *a priori* de um deles não influencia no outro:

$$P(L, C) = P(L) \times P(C)$$

Porém consideremos outro evento no mesmo contexto: a probabilidade de o chão estar molhado $P(M)$. Quando consideramos essa probabilidade em conjunto à $P(L)$ e $P(C)$, temos uma Rede Bayesiana um pouco diferente, representada na Figura 2.2:

Figura 2.2 | Rede Bayesiana de três probabilidades



Fonte: elaborada pelo autor

A probabilidade de o chão estar molhado $P(M)$ está ligada à probabilidade de estar chovendo $P(C)$. Em outras palavras, o conhecimento, *a priori*, de que, por exemplo, acaba de chover, influencia substancialmente na probabilidade de o chão estar molhado, *a posteriori*.

Podemos, então, calcular a probabilidade de o chão estar molhado dada a probabilidade de estar chovendo no contexto da Rede Bayesiana da Figura 2.2, que é representada da seguinte forma: $P(M|C)$.

Bussab e Morettin (2010) mostram que esse valor se dá pelo cálculo da seguinte equação:

$$P(M | C) = \frac{P(M \wedge C)}{P(C)}$$

Nesse caso, a probabilidade de chuva deve ser maior que zero, obviamente.

Por exemplo, a probabilidade de rolarmos dois dados e obtermos um par (dois valores iguais) ao mesmo tempo, sendo o primeiro dado (Dado1) de valor 5, é dada por:

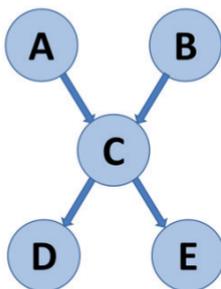
$$P(\text{Par} | \text{Dado1} = 5) = \frac{P(\text{par} \wedge \text{Dado1} = 5)}{P(\text{Dado1} = 5)}$$

Temos que $P(\text{Dado1}=5)$ é de 100%, uma vez que esse valor é atribuído a priori. A questão, então, para se obter um par é: qual a probabilidade do dado 2 sair com valor também 5? Ora, se um dado tem seis lados, numerados de 1 a 6, a probabilidade é de $1/6$.

Uma característica interessante acerca das probabilidades para um evento é que, como afirmam Bussab e Morettin (3010), independentemente do valor da probabilidade, o resultado é que o evento pode assumir dois estados: ou ele ocorre, ou não ocorre. Trata-se de um par binário "evento/não-evento". Cada resultado pode, portanto, ocupar um de dois valores, apenas. Ou chove, ou não chove (independentemente da probabilidade de chuva; ou se ganha na loteria, ou não se ganha na loteria, independentemente da probabilidade).

Essa característica revela uma das vantagens do cálculo condicional de probabilidades que as Redes Bayesianas nos permitem. Para entender esse conceito, vamos observar a Rede Bayesiana a seguir, representada na Figura 2.3:

Figura 2.3 | Uma Rede Bayesiana de cinco probabilidades



Fonte: elaborada pelo autor.

Para essa rede temos as seguintes probabilidades:

- $P(A), P(B)$ – Ambos os eventos são independentes
- $P(C | A, B)$ – A probabilidade de C está condicionada à A e à B
- $P(D|C), P(E|C)$ – Tanto a probabilidade de D como a probabilidade de E estão condicionadas à C

Se fôssemos calcular todos os eventos separadamente, teríamos duas possibilidades por evento em 5 eventos, ou seja: 2^5 ou 32 estados finais possíveis.

No entanto, por meio das regras condicionantes da Rede Bayesiana, conseguimos reduzir esses estados a apenas 10:

- $P(A)$ – 1 estado, A ou $\neg A$
- $P(B)$ – 1 estado, B ou $\neg B$
- $P(C | A, B)$ – 4 estados dependendo dos dois estados A ou $\neg A$ e dos dois estados B ou $\neg B$
- $P(D | C)$ – 2 estados dependendo dos dois estados C ou $\neg C$
- $P(E | C)$ – 2 estados dependendo dos dois estados C ou $\neg C$

Somando-se esses estados, temos que a Rede Bayesiana simplifica os estados de 32 para apenas 10, pois estamos levando em conta suas interdependências.

O Aprendizado Bayesiano, como definido por Russell e Norvig (2013), consiste no cálculo de cada estado possível. Uma vez calculados os estados, o método demanda construção de uma tabela com as probabilidades calculadas, e essa tabela passa a constituir a base de conhecimento gerada pelo método, ficando à disposição para uso na resolução de questões pertinentes ao ambiente em questão.



Pesquise mais

Conheça mais sobre o Teorema de Bayes nesse vídeo acerca do assunto no YouTube:

ZURUBABEL. Teorema de Bayes – Aula 1: características (Probabilidade a priori, condicional e conjunta). Disponível em: <<https://www.youtube.com/watch?v=78R1yNVGnSk>>. Acesso em: 20 jul. 2018.

E ainda, para você estudar mais a fundo esse assunto, o pesquisador Jones Granatyr publicou um interessante artigo, inclusive com um vídeo, mostrando como se criar Redes Bayesianas por meio do software Netica, que tem uma versão gratuita.

GRANATYR, J., **Redes Bayesianas com o Netica**. IA Expert, 2016. Disponível em: <<http://iaexpert.com.br/index.php/2016/09/20/ferramentas-para-ia-redes-bayesianas-com-o-netica/>>. Acesso em: 23 abr. 2018.

Métodos Baseados em Classificadores

Norvig e Russell (2013) afirmam que os classificadores são tipos especializados de agentes inteligentes, capazes de perceber o ambiente e dele extrair as características que permitam armazenar

cada estado percebido, classificando-os de acordo com suas características. É justamente essa capacidade de agrupar estados de acordo com suas características que nos permite chamar esse tipo específico de agente inteligente de agente classificador.

Uma característica importante dos classificadores, segundo os autores, é que são capazes de, primeiramente, identificar as características relevantes a serem avaliadas para o processo de classificação, descartando as demais características que são irrelevantes para os propósitos do agente.

Aqui, é importante frisar que, segundo Amaral (2015), a capacidade de classificar é a tarefa mais importante no aprendizado de máquina. Isso porque a separação de eventos em classes diferentes nos permite sua diferenciação bem como, com base nesta, a extração de informações para uso e ação futura.

Os classificadores são elementos fundamentais de um tipo de problema que é clássico em Inteligência Artificial: o reconhecimento de padrões (VALIN, 2009). O reconhecimento de padrões é bastante comum em nossas vidas, por mais que não nos damos conta disso na maioria das vezes. Todo o estudo da Matemática, aliás, é um gigantesco conjunto de campos de pesquisa em identificação e classificação de padrões, segundo Devlin (1996).

Em um campo mais prático e mais próximo à maioria de nós, segundo Valin (2009), somos protegidos diariamente por um sistema de reconhecimento de padrões todas as vezes que checamos nossa caixa de entrada no servidor de e-mail. É um sistema de reconhecimento de padrões com excelente desempenho de aprendizado e que nos protege da praga do SPAM (e-mails indesejados e não solicitados que insistem em nos oferecer o que não queremos comprar). Os filtros de anti-SPAM são classificadores que aprendem o que é SPAM de acordo com um sistema de *feedback* (retroalimentação) que funciona da seguinte forma:

1. Os criadores do sistema pedem para que um conjunto de e-mails seja classificado como SPAM ou não SPAM.
2. Um usuário humano reclassifica os mesmos e-mails, dando como certa ou errada a classificação de cada um deles feita pelo classificador.

3. O classificador, então, recalibra seus parâmetros internos para que a reclassificação do usuário seja levada em consideração, refinando sua capacidade de encontrar SPAM.
4. Os passos de 1 a 3 se repetem para novos conjuntos de e-mails com a classificação SPAM/Não-SPAM sendo refinada a cada iteração
5. Uma vez que o nível adequado de reconhecimento e classificação seja atingido (positivos maximizados e falsos-positivos minimizados), o sistema é posto em produção.

Russell e Norvig (2013) apontam alguns algoritmos de classificação adequados para uso em Inteligência Artificial para reconhecimento de padrões:

- **Naive Bayes** – Algoritmo derivado da aplicação do Teorema de Bayes em redes com alto grau de independência entre os nós. Em outras palavras, lembra-se daquela capacidade de identificar a probabilidade de um evento dado que outro acontece? Essa característica indica que uma ocorrência depende de uma outra ocorrência. No algoritmo Naive Bayes, assumimos a independência entre as variáveis e montamos uma tabela com os valores assumindo que não se influenciam. Bastante utilizado na mineração de dados em textos.
- **AODE** – Do inglês *Averaged One-Dependence Estimator* (Estimador com uma dependência ponderada). Algoritmo que utiliza um grau de dependência para as variáveis que assim demandam, sacrificando um pequeno percentual de desempenho por uma precisão bastante superior ao que se consegue com o Naive Bayes (a precisão obtida supera com vantagem a pequena perda de desempenho).
- **KNN** – Do inglês *K Nearest Neighbors*, ou os K elementos vizinhos mais próximos de um dado elementos. Dado um ponto, o KNN classifica os K elementos mais próximos a ele como pertencentes à classe de vizinhos. Este algoritmo também pode ser usado para processos de regressão, isto é, quando queremos encontrar o ponto mais adequado para representar K elementos próximos uns aos outros.



Seria possível realizar tarefas de reconhecimento de padrões sem classificar os elementos a serem reconhecidos? Por quê?

Métodos de Aprendizado Estatístico

Russell e Norvig (2013) nos convidam a pensar em uma situação hipotética com o objetivo de compreendermos como se dá o aprendizado estatístico. Imaginemos uma empresa que vende balas de apenas dois sabores: morango e hortelã. A fábrica usa a mesma embalagem opaca em cada uma das balas, e para complicar, vende cinco tipos diferentes de pacotes de bala:

- P1 – 100% de balas de morango
- P2 – 75% de balas de morango, 25% de balas de hortelã
- P3 – 50% de balas de morango, 50% de balas de hortelã
- P4 – 25% de balas de morango, 75% de balas de hortelã
- P5 – 100% de balas de hortelã

As embalagens dos cinco tipos de pacotes também são opacas, de forma que para sabermos o sabor de uma bala, precisamos abrir o pacote, pegar uma bala e abri-la. Além disso, para sabermos que tipo de pacote de balas temos, devemos abrir uma quantidade razoável de balas, sendo impossível identificar o tipo de pacote abrindo uma ou duas balas, apenas.

Nesse caso, o Teorema de Bayes não nos auxilia, pois nos permite classificar os cinco tipos diferentes de pacotes com 20% de probabilidade para termos nas mãos um pacote qualquer.

O aprendizado estatístico nos permite melhorar a precisão da previsão de qual tipo de pacote temos em mãos a cada bala que abrimos. Como o pacote P1 tem 0% de balas de hortelã, se uma bala qualquer da sequência for de hortelã, isso nos permitirá eliminar P1. O inverso ocorre com P5, que tem 0% de balas de morango: se em nossa sequência de balas abertas encontra-se uma bala de morango, podemos eliminar P5.

A cada bala aberta, os métodos de aprendizado estatístico recalibram os dados que são de seu conhecimento, permitindo cada vez mais precisão na previsão de qual tipo de pacote temos em mãos.

O aprendizado estatístico permite ao método levar em conta os eventos anteriores na tomada de decisão acerca de acontecimentos futuros.

Avaliação de progresso no aprendizado

Diferentes tipos de agentes inteligentes têm diferentes tipos de métricas para avaliação de progresso no aprendizado, afirmam Russell e Norvig (2013):

- **Classificadores** – A avaliação de progresso é feita com base no progresso da precisão da classificação. Um exemplo seria o percentual de e-mails correta e incorretamente classificados como SPAM por um classificador.
- **Regressores** – A avaliação de progresso é feita com base no fator de correlação, que é um fator que relaciona as distâncias entre o ponto de regressão e todos os pontos da amostra. Um exemplo seria a curva em um gráfico que melhor representa os dados coletados em um experimento quando comparada com a curva teórica (uma parábola, uma hipérbole, etc.).

Outros fatores também podem ser utilizados para mensurar o progresso do aprendizado, tais como:

- O tempo de treinamento necessário para que o agente passe a gerar resultados considerados adequados
- A quantidade de recursos computacionais necessários para que uma rede possa ser treinada para atingir determinado grau de precisão em um tempo específico.



Assimile

Para avaliar o progresso de aprendizado, precisamos definir identificadores que, ao serem mensurados, dão-nos a medida do progresso de um determinado agente dentro dos parâmetros escolhidos.

Pronto! Agora que já conhecemos um pouco mais acerca dos métodos probabilísticos e estatísticos, podemos treinar um pouco nosso conhecimento e dar sequência aos nossos estudos em Inteligência Artificial.

Sem medo de errar

Para solucionar essa situação proposta, você pode apresentar um relatório com os seguintes dados:

- Dados de identificação do relatório
 - Nome do relatório.
 - Responsável pelo relatório.
 - Data do relatório.
- Referencial Teórico
 - O que são classificadores?
 - Situação do mundo real que demanda classificação (ex: um sistema automatizado de autorização de crédito que determina se o crédito solicitado pode ser autorizado ao usuário com base em seu histórico financeiro).
 - Tipos de algoritmos de classificação e suas características
 - Naive Bayes.
 - KNN.
 - AODE.
- Sugestões para se usar classificadores com as bases de dados da empresa de serviços de saúde.
 - Identificar clientes com maior propensão a uso dos serviços com base em seu histórico médico (consultas, medicamentos, exames, procedimentos cirúrgicos).
 - Sugestão de exames preventivos em clientes com base em semelhança com clientes anteriores que tiveram problemas de saúde.
 - Campanhas voltadas a clientes que, dado seu histórico de uso, tendem a trocar de serviço, indo para a concorrência, também com base na análise de comportamento de clientes anteriores que tomaram essa decisão
 - Quais outros serviços com base em qualificadores você e sua equipe conseguem visualizar?

Pronto, você acaba de apresentar um excelente relatório para a empresa, e seus planos devem ser aceitos para a próxima fase. Parabéns!

Avançando na prática

Montando uma Rede Bayesiana

Descrição da situação-problema

Um prestigiado laboratório de desenvolvimento de soluções de Inteligência Artificial decidiu criar um processo de contratação para novos profissionais que envolve várias fases de avaliação dos candidatos. Eles contrataram você, profissional renomado na área, para avaliar uma das questões propostas, que diz respeito às redes bayesianas. A questão é a seguinte:

Imagine as probabilidades dos seguintes eventos:

- Probabilidade de receber a devolução do imposto de renda.
- Probabilidade de chover.
- Probabilidade de ser feriado.
- Probabilidade de ir à praia.
- Probabilidade de tomar água de coco.
- Probabilidade de comer espetinho de camarão.

Monte uma rede bayesiana considerando as dependências das probabilidades e calcule a quantidade de possíveis estados para esta rede.

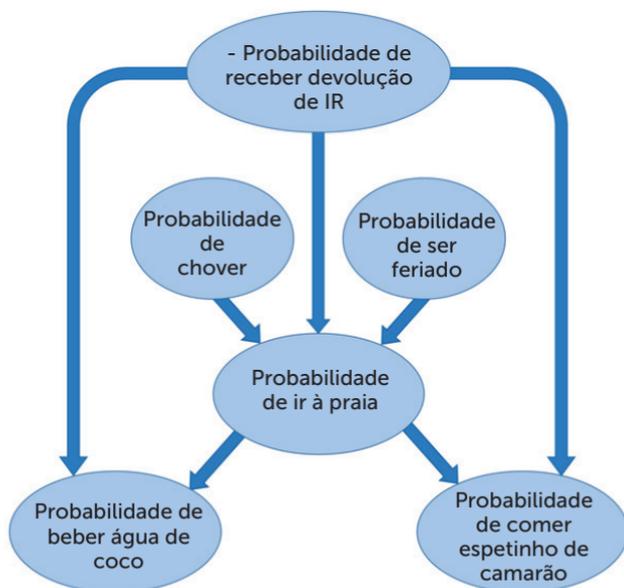
Sua tarefa é resolver a questão e dar seu parecer se será útil ao processo seletivo.

Resolução da situação-problema

Para auxiliar o laboratório, primeiramente você deverá tentar resolver a questão proposta. Para tanto, podemos levar em consideração as seguintes dependências (ou ausências de dependências):

- Probabilidade de receber devolução do imposto de renda – independe dos demais eventos: $P(D)$.
- Probabilidade de chover – independe dos demais eventos: $P(C)$.
- Probabilidade de ser feriado – independe dos demais eventos: $P(F)$.
- Probabilidade de ir à praia – depende da devolução do IR, de chover e de ser feriado: $P(P | D,C,F)$.
- Probabilidade de beber água de coco – depende de ir à praia: $P(A | P)$.
- Probabilidade de comer espetinho de camarão – depende de ir à praia $P(E | P)$.

Figura 2.4 | A Rede Bayesiana resultante



Fonte: elaborada pelo autor.

A quantidade de possíveis estados dessa rede é dada pelo seguinte cálculo:

- $P(D) - 1$
- $P(C) - 1$
- $P(F) - 1$

- $P(P \mid D, C, F) = 8$
- $P(A) = 2$
- $P(C) = 2$

Somando todos esses estados possíveis, temos um total global de 15 estados possíveis para essa rede bayesiana.

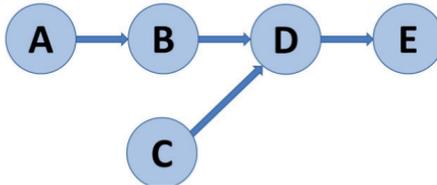
Pronto, a questão está resolvida e você pode concluir que:

- É uma questão consistente no que concerne à teoria das redes bayesianas
- É uma questão desafiadora.
- É uma questão útil na avaliação dos candidatos.

Faça valer a pena

1. A Figura 2.5, a seguir, apresenta uma rede bayesiana de cinco nós: A, B, C, D e E. Esses nós podem ter probabilidades independentes ou não, de acordo com o que é apresentado. Observe a figura e identifique quais são independentes e quais não são; em seguida, proceda ao enunciado da questão:

Figura 2.5 | Uma Rede Bayesiana de cinco nós



Fonte: elaborada pelo autor.

Assinale a alternativa que contém a quantidade de estados possíveis para a rede bayesiana representada na Figura 2.5:

- 4.
- 8.
- 10.
- 24.
- 32.

2. Segundo Amaral (2015), a capacidade de classificar é a tarefa mais importante no aprendizado de máquina. Isso porque a separação de eventos

em classes diferentes nos permite a diferenciação desses eventos e a extração de informações para uso e ação futura com base nessa diferenciação.

Assinale o tipo de problema para o qual o uso de classificadores é especialmente recomendado:

- a) Redes neurais.
- b) Aprendizado de máquina.
- c) Redes Bayesianas.
- d) Reconhecimento de padrões.
- e) Agentes inteligentes.

3. A correspondência entre todos os estados possíveis do ambiente com regras pré-estabelecidas é impossível na prática em função do crescimento exponencial de recursos de processamento, memória e armazenamento. Esse crescimento exponencial se dá em função do crescimento das regras a serem codificadas no agente. Em suma, à medida que o ambiente cresce em complexidade, torna-se cada vez mais difícil (leia-se: custoso) preparar o agente para cada uma das situações que possa encontrar.

Em um ambiente em que não é possível o agente mapear todas as possibilidades, ele deve conviver com a:

- a) Certeza.
- b) Incerteza.
- c) Dificuldade.
- d) Facilidade.
- e) Entropia.

Seção 2.2

Buscas com inteligência artificial

Diálogo aberto

Imagine que você consegue criar uma árvore com todas as possibilidades de jogadas para o jogo de xadrez. Começando com as 20 possibilidades iniciais para a jogada das brancas, todas as possíveis respostas das pretas, e assim por diante. Depois de várias dezenas de jogadas, você terá uma gigantesca árvore de possibilidades, contendo todas as combinações de jogadas possíveis. A pergunta, nesse caso é: como visitar os nós dessa árvore hipotética para sempre escolher a melhor jogada, desenhando o melhor caminho até a vitória? A resposta é: precisamos de uma maneira eficiente para atravessar toda a árvore, analisando os resultados obtidos nos nós. É sobre isso que falaremos nesta seção: os mecanismos de busca em estruturas de dados complexas. Esses mecanismos são fundamentais aos agentes inteligentes e, portanto, são parte das soluções de inteligência artificial.

Lembre-se: você e seu time foram chamados para auxiliar uma empresa líder no setor de serviços de saúde a utilizar seus dados a fim de oferecer mais e melhores serviços, além de aumentar a efetividade da organização, beneficiando tanto os clientes como os acionistas. Nessa etapa, você deverá explicar como pode ser feita a busca em uma estrutura de dados montada com os dados da empresa. Essa explicação auxiliará os executivos a compreenderem como é importante essa etapa na decisão de que caminho tomar na identificação de informações úteis presentes nos dados de saúde e dos clientes. É realmente possível combinar as tecnologias de realidade aumentada e inteligência artificial?

O primeiro tópico desta seção será uma explanação acerca dos mecanismos de busca, que são subjacentes aos agentes inteligentes, mas absolutamente importantes a eles na busca pelo melhor caminho em uma estrutura de dados.

O segundo tópico será o método de busca em extensão e o resultado gerado por ele.

O terceiro tópico será o método de busca em profundidade e suas variações, que geram ordens diferentes de visitas aos nós.

O último tópico desta seção será o método de busca heurística e suas implicações para o agente inteligente.

Prepare-se para mais uma etapa de nossa aventura.

Vamos em frente!

Não pode faltar

Mansour (2011) nos informa que o objetivo dos algoritmos de busca é encontrar elementos com determinadas propriedades em um espaço de busca ou em meio a uma coleção de elementos. A solução para uma busca, ainda segundo o autor, pode ser um elemento ou um conjunto de elementos que atendem às condições da busca.

A partir de agora, vamos entender melhor como funcionam os métodos e algoritmos de busca.

Mecanismos de busca

Os agentes inteligentes organizam o que coletam em alguma estrutura que lhes permita encontrar o subconjunto de dados necessários para sua tomada de decisão em um momento qualquer do tempo. Uma vez que essa estrutura tende a crescer, deve ser percorrida de alguma forma eficiente para que o agente seja efetivo em sua ação.

Para encontrar o dado ou o subconjunto de dados necessários, é fundamental que se execute uma busca. Segundo Russell e Norvig (2013), os algoritmos de busca são subjacentes aos agentes inteligentes, parte fundamental de uma solução de inteligência artificial. Os autores desenvolvem o assunto de maneira abrangente, apresentando vários tipos de busca em estruturas de dados que se aproveitam de características específicas dessas estruturas para serem mais eficazes.

Nos itens a seguir, vamos analisar alguns métodos de busca que ainda são bastante úteis e que deram origem a vários outros, sendo alicerces de vários dos mecanismos que vieram depois.

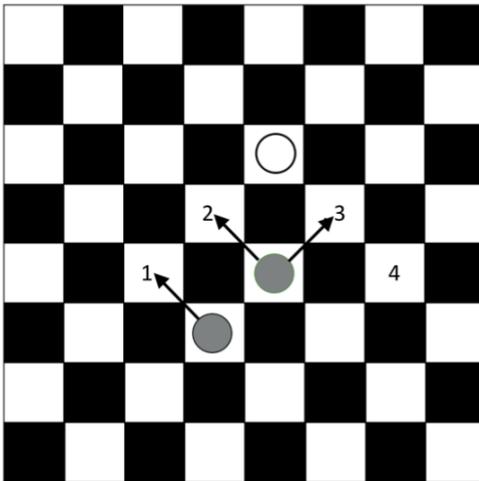


Sem os mecanismos de busca, isto é, sem mecanismos que lhes permitissem a avaliação de suas estruturas de dados, os agentes inteligentes não teriam como decidir quais ações adequadas a serem tomadas em seu processo de resposta a um problema de inteligência artificial.

Busca em Extensão

Lá na Seção 1.1 vimos que alguns jogos (entre outros tipos de programas) criam enormes estruturas de dados, chamadas **árvores de decisão**, à medida que o jogo progride. Essas árvores registram o estado do jogo e as possibilidades futuras de movimentos. Em um jogo de damas, por exemplo, podemos tomar a situação expressa na Figura 2.6, em que é a vez das peças cinzas e há três jogadas possíveis, representadas nas casas numeradas 1, 2 e 3:

Figura 2.6 | Uma situação em um jogo de damas

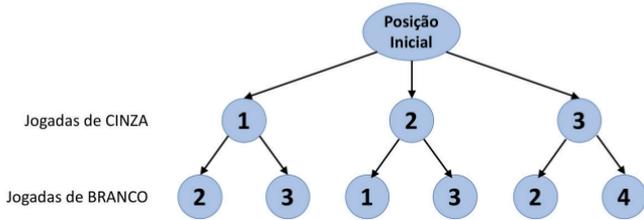


Fonte: elaborada pelo autor.

Para cada uma dessas jogadas possíveis, as peças brancas podem se mover, em resposta, para as casas 1, 2, 3 e 4 (dependendo da jogada das peças de cor cinza).

Podemos, então, compor uma árvore de decisão para as jogadas possíveis, representada na Figura 2.7 a seguir:

Figura 2.7 | Árvore de Decisão para o jogo da Figura 2.6



Fonte: elaborada pelo autor.

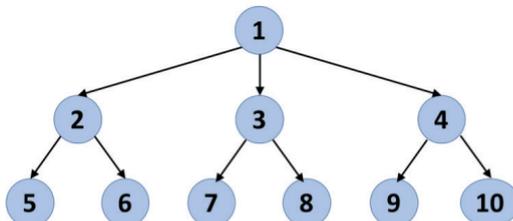
Para entender a árvore acima, basta perceber que o primeiro nível após a **posição inicial** contém o universo das jogadas possíveis das pedras de cor cinza, e o segundo nível, o universo das jogadas possíveis das pedras de cor branca, como resposta. Há, obviamente, jogadas de melhor e de pior qualidade em cada árvore de decisão, e um programa de jogo de damas atribui um valor maior ou menor para cada jogada, dependendo do resultado favorável ou desfavorável.

A montagem de uma árvore de decisão é parte fundamental de um jogo de damas, xadrez ou de outros jogos com estados conhecidos e mapeáveis. Contudo, uma vez montada parte da árvore (e essa árvore deve ser ajustada a cada jogada realizada), é preciso uma forma rápida e eficaz de se analisar os nós para se encontrar a melhor alternativa. É aí que entram em ação os algoritmos de busca.

O primeiro deles que você vai ver é o algoritmo de busca em extensão (em inglês: *breadth-first search*).

Esse método, segundo Cormen, Leieron, Rivest e Stein (2011), implica começar pela raiz e visitar cada nó do primeiro nível antes de prosseguir para o segundo nível e assim por diante. De acordo com esse método, a busca em profundidade visita os nós da árvore da Figura 2.7 de acordo com a ordem especificada na Figura 2.8:

Figura 2.8 | Ordem de visita dos nós da Busca em Extensão



Fonte: elaborada pelo autor.

O método leva esse nome porque explora por completo os elementos que se encontram em um nível (extensão horizontal) antes de se lançar ao nível posterior.

O método de busca em extensão pode ser aplicado tanto em árvores como em grafos, sendo que nesse segundo tipo de estrutura, dado que na mesma existe a possibilidade da formação de laços, é fundamental que um mecanismo de controle de iterações seja utilizado para evitar que o algoritmo entre em um laço infinito de execução (CORMEN, LEIERSON, RIVEST, STEIN, 2011).



Refleta

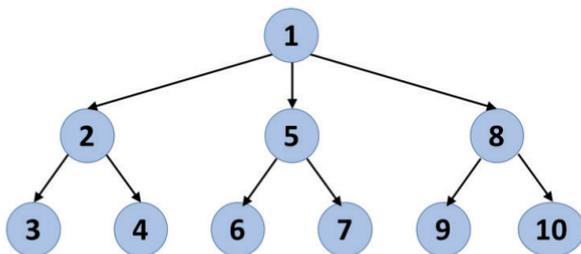
Ao percorrer um grafo, que é uma estrutura de dados em que se formam ciclos entre os vários nós, é possível evitar que os nós sejam visitados mais de uma vez?

Busca em profundidade

Cormen, Leieron, Rivest e Stein (2011) descrevem o método de busca em profundidade como sendo aquele em que o algoritmo explora a profundidade de um ramo da árvore o máximo que for possível (até chegar a uma folha) antes de retornar à bifurcação mais próxima ainda não explorada e repetir o processo de aprofundamento.

No exemplo da árvore representada inicialmente na Figura 2.7, se usássemos o método de busca em profundidade, teríamos uma possibilidade de gerar as visitas na ordem especificada na Figura 2.9 a seguir (CORMEN; LEIERSON; RIVEST; STEIN, 2011):

Figura 2.9 | Exemplo de ordem de visita dos nós da busca em profundidade



Fonte: elaborada pelo autor.

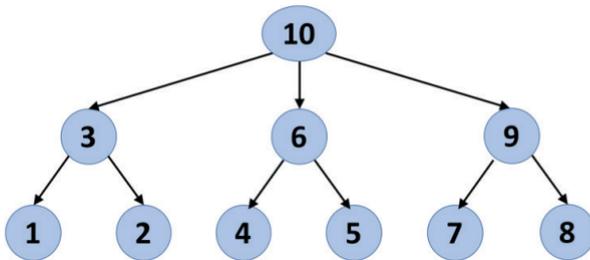
Segundo Cormen, Leieron, Rivest e Stein (2011), a busca em profundidade é um processo recursivo que pode ser dividido em três tarefas, a saber:

- **E** – Atravessar a árvore (ou grafo) recursivamente pela esquerda.
- **D** – Atravessar a árvore (ou grafo) recursivamente pela direita).
- **V** – Visitar o nó da árvore (processo de análise do conteúdo do nó, que pode variar de uso para uso do algoritmo).

Dependendo da ordem em que são executadas as tarefas acima, teremos um resultado diferente na ordenação dos nós. O método de busca em profundidade permite que os nós da estrutura sejam apresentados em três tipos diferentes de ordenação, a saber (CORMEN, LEIERSON, RIVEST, STEIN, 2011):

- **Pré-Ordem (VED)** – Primeiramente o nó corrente é visitado (V); em seguida, realiza-se o processo recursivo de visita do ramo à esquerda (E); após isso, realiza-se o processo decursivo de visita do ramo à direita (D). A Figura 2.9, acima, representa a ordem de vistas aos nós usando a pré-ordem.
- **Pós-Ordem (EDV)** – Primeiramente realiza-se o processo recursivo de visita do ramo à esquerda (E); em seguida, realiza-se o processo decursivo de visita do ramo à direita (D); após isso, o nó corrente é visitado (V). A Figura 2.10, a seguir, representa a ordem de visitas aos nós usando a pós-ordem:

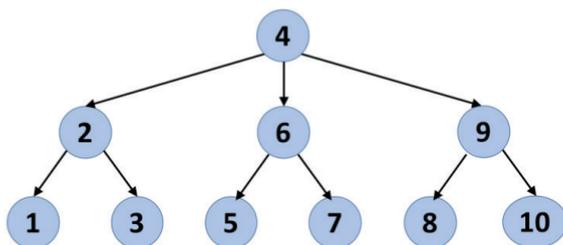
Figura 2.10 | Ordem de visita dos nós em pós-ordem



Fonte: elaborada pelo autor.

- **Em-Ordem (EVD)** – Primeiramente realiza-se o processo recursivo de visita do ramo à esquerda (E); em seguida, o nó corrente é visitado (V); após isso, realiza-se o processo decursivo de visita do ramo à direita (D). A Figura 2.11 a seguir representa a ordem de visitas aos nós usando a em-ordem:

Figura 2.11 | Ordem de visita dos nós em em-ordem



Fonte: elaborada pelo autor.



Pesquise mais

O professor André Backes, da Universidade Federal de Uberlândia, explica e exemplifica com clareza os métodos de percorrer árvores neste excelente vídeo no YouTube. Confira!

LINGUAGEM C PROGRAMAÇÃO DESCOMPLICADA. [ED] Aula 73: Percorrendo uma Árvore Binária. 20 nov. 2014. Disponível em: <<https://www.youtube.com/watch?v=z7XwVWYQRAA>>. Acesso em: 6 maio 2018.

Busca Heurística

Já vimos anteriormente que os problemas de natureza heurística são aqueles em que métodos determinísticos não são possíveis de serem aplicados, logo, o jeito para resolvê-los é à base de tentativa e erro, buscando sempre a maior aproximação possível da melhor solução que lhe cabe.

Russell e Norvig (2013) descrevem o processo como sendo baseado na análise dos nós e da disposição dos mesmos, sendo que à medida que cada visita ocorre, o mecanismo ajusta uma estrutura interna que vai sendo criada para armazenar os resultados das visitas. A cada visita os valores são ajustados para refletir a aproximação ou o afastamento do valor que se busca. Como o processo envolve o ajuste da estrutura interna do método a cada visita, aumentando a quantidade e melhorando a qualidade das informações disponíveis, fica claro que ocorre aprendizado de máquina ao longo do caminho.

Por meio da busca heurística, continuam os autores, a árvore não é visitada em uma ordem determinística, como nos casos das buscas em profundidade e em extensão, mas sim, a cada nó visitado

há a escolha de qual nó visitar. Em árvores ou grafos de grande porte, a busca determinística (em profundidade ou em extensão) pode demorar muito, dado o grande número de nós a ser visitado. Já a busca determinística cria atalhos, pulando de um ramo para outro de acordo com o que se aprendeu com os nós já visitados, aumentando, assim, a possibilidade de encontrar a solução procurada em menos tempo e com menos esforço.

O tempo necessário para uma busca heurística bem como a complexidade do caminho de busca dependem de três fatores (CORMEN, LEIERSON, RIVEST, STEIN, 2011):

- Complexidade da árvore (ou grafo) em que ocorrerá a busca – Uma árvore pequena gerará menores nós visitados para se chegar à solução desejada. Uma árvore também pode ser podada, isto é, os ramos com menores valores podem ser removidos, evitando que tempo precioso de busca seja empregado em visitas inúteis para que se encontre o nó mais valioso.
- Regras que formam os nós da árvore – As árvores formadas pelos estados possíveis de um jogo de damas são consideravelmente mais simples que as árvores formadas pelos estados possíveis de um jogo de xadrez. Isso porque as regras de xadrez e, por consequência, o estado favorável ou desfavorável de uma posição são consideravelmente mais complexos do que no caso de damas
- Eficiência do algoritmo de busca heurística – O conjunto de regras que decide qual o próximo nó a ser visitado com base no conhecimento agregado sobre a árvore.



Exemplificando

Um dos métodos de busca heurística mais conhecidos e usados na implementação do jogo de xadrez em sistemas computacionais é o chamado Minimax, bem como seu equivalente, Maxmin. No caso do Minimax, a decisão é pelo valor menor que uma jogada pode impor ao oponente, dado um universo de jogadas possíveis desse oponente após a jogada atual. No caso do Maxmin, a decisão é pelo valor mais alto que o jogador pode ter certeza de obter sem conhecer a jogada que será feita por seu oponente.

Pronto, agora que você já entrou em contato com os mecanismos de busca, que tal criar uma árvore binária simples e tentar realizar algumas operações de busca com base nos métodos acima?

De qualquer maneira, não deixe de realizar os exercícios e de resolver as situações problema aqui propostas.

Até a próxima!

Sem medo de errar

Para resolver essa etapa do seu projeto com a empresa de serviços de saúde, você pode apresentar seus resultados como um relatório com a seguinte estrutura:

- Identificação do relatório
 - o Nome do relatório (Mecanismos de Busca em Inteligência Artificial).
 - o Responsável.
 - o Data.
- O que são e para que servem os mecanismos de busca
 - o Explicação acerca da formação de estruturas de dados a partir da exploração de um espaço de estados possíveis (por exemplo: em um jogo)
 - o Explicação de que para se encontrar o melhor caminho a seguir em tal estrutura de dados, os agentes inteligentes precisam visitar os nós de forma a identificar a melhor decisão a ser tomada.
 - o Surge, então, a necessidade de se desenvolver métodos de se atravessar a estrutura, avaliando os valores de cada nó.
- O método de busca em extensão
 - o Explicar como funciona o método.
 - o Explicar que cada nível da estrutura é visitado antes de se iniciar a visita aos níveis posteriores.
- O método de busca em profundidade
 - o Explicar como funciona o método.

- o Explanar as três componentes do método
 - E – Incursão recursiva à esquerda.
 - D – Incursão recursiva à direita.
 - V – Visita ao nó (avaliação do conteúdo).
- o Explanar as três variações possíveis da busca em profundidade
 - Pré-Ordem: V-E-D.
 - Pós-Ordem – E-D-V.
 - Em-Ordem – E-V-D.
- O método de busca heurística
 - o Explanar como funciona o método.
 - o Uso das informações coletadas da própria estrutura para decidir qual caminho de busca tomar.
 - o Poda dos ramos menos promissores.

Pronto! Agora você apresentou um relatório que dá ciência aos executivos da empresa de serviços de saúde acerca da importância dos mecanismos de busca.

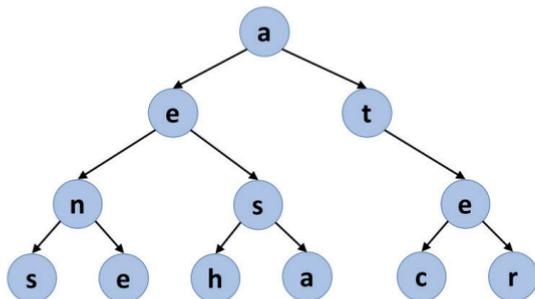
Avançando na prática

O código secreto

Descrição da situação-problema

O serviço secreto conseguiu prender um espião com laptop, porém, não conseguiu a senha para acessá-lo e descobrir quais informações estava roubando. Um dos agentes conseguiu encontrar um pedaço de papel escondido no apartamento do espião com um desenho no mínimo curioso, representado na Figura 2.12 a seguir. Ao desconfiar que a figura pudesse ser a senha de acesso ao computador, chamou você para auxiliá-lo, uma vez que tem boa formação em métodos de busca. Frente a isso, como poderá contribuir nessa missão?

Figura 2.12 | A imagem recuperada do apartamento do espião



Fonte: elaborada pelo autor.

Resolução da situação-problema

Lembrando-se de seu conhecimento acerca de métodos de busca em árvores binárias, você decide aplicar seus conhecimentos para percorrer a árvore. Dessa forma, você obtém os seguintes resultados:

- Percorrendo a árvore em pré-ordem, o resultado é: a-e-n-s-e-s-h-a-t-e-c-r.
- Percorrendo a árvore em ordem, o resultado é: s-n-e-e-h-s-a-t-c-e-r.
- Percorrendo a árvore em pós-ordem, o resultado é s-e-n-h-a-s-e-c-r-e-t-a.

Você rapidamente percebe que percorrendo a árvore em pós-ordem obtém a senha secreta, que é justamente “senhasecreta”.

Caso resolvido!

Faça valer a pena

1. Os agentes inteligentes organizam o que coletam em alguma estrutura que lhes permita encontrar o subconjunto de dados necessários para determinada decisão em um momento qualquer do tempo. O fato é que essa estrutura tende a crescer e deve ser percorrida de alguma forma eficiente para que o agente seja efetivo em sua ação.

Os métodos de busca presentes nos agentes inteligentes para a realização dessas tarefas são _____.

Assinale a alternativa correta, que melhor completa o pensamento da última sentença acima:

- a) centrais sendo o objetivo principal de um agente inteligente.
- b) aleatórios, gerando resultados ao acaso.
- c) criptografados, gerando resultados indecifráveis.
- d) subjacentes, sendo uma ferramenta importante à disposição do agente.
- e) opcional, podendo ou não existir de acordo com as ações tomadas pelo agente.

2. Entre os métodos utilizados para a busca em profundidade, um deles é especificado pela seguinte sequência de ações:

Primeiramente, o nó corrente é visitado; em seguida, realiza-se o processo recursivo de visita do ramo à esquerda; após isso realiza-se o processo decursivo de visita do ramo à direita.

Assinale a alternativa que contém o nome correto para o procedimento descrito acima:

- a) Busca em propriedade.
- b) Busca em extensão.
- c) Pré-Ordem.
- d) Em-Ordem.
- e) Pós-Ordem.

3. O tempo necessário para uma busca heurística bem como a complexidade do caminho de busca dependem de alguns fatores:

- I. Complexidade da árvore (ou grafo) em que ocorrerá a busca.
- II. Regras que formam os nós da árvore.
- III. Presença de agente inteligente para uso dos resultados obtidos.
- IV. Eficiência do algoritmo de busca heurística.

Assinale a alternativa que contém apenas fatores corretos:

- a) I, II e III, apenas.
- b) I, II e IV, apenas.
- c) I, III e IV, apenas.
- d) II, III e IV, apenas.
- e) I, II, III e IV.

Seção 2.3

Soluções via inteligência artificial

Diálogo aberto

Repentinamente você recebe a notícia de que um parceiro de sua empresa na Tailândia quer apresentar algumas tecnologias cuja comercialização no Brasil será de benefício para ambas as empresas. Infelizmente, por uma triste (ou feliz) coincidência, nenhum dos executivos de sua empresa tem condições de fazer a viagem, ficando para você essa responsabilidade (e privilégio). Chegando lá, a surpresa: o alfabeto usado pelos tailandeses não se parece em nada com o nosso, e nem com outros alfabetos asiáticos (chinês ou japonês). Como ler placas, textos e mensagens diversas? Entram em cena os aplicativos de tradução, que estão disponíveis em nossos smartphones, a maioria deles gratuitamente. Você baixa o aplicativo, aciona a câmera, aponta para o texto que quer ler e pronto: como se fosse por magia o texto aparece em português, como se alguém tivesse repintado o cartaz ou reescrito o texto na própria página que você está folheando. Isso é possível (e não é ficção científica) pela união de duas tecnologias: a inteligência artificial e a realidade aumentada.

Por meio desta seção, você vai continuar o processo de conhecer, compreender e ser capaz de resolver problemas em inteligência artificial utilizando as metodologias apresentadas.

A grande empresa do setor de serviços de saúde que busca gerar melhores produtos e serviços por meio do uso de suas bases de dados, agora questiona você sobre como utilizar as técnicas de realidade aumentada. Sua enorme base de dados, coletada ao longo de décadas no mercado, pode ser usada de alguma forma em conjunto com as técnicas de realidade aumentada para prestar melhores serviços aos clientes? Você e seu time acreditam que sim, e sabem que a inteligência artificial tem um papel fundamental nesse tipo de solução. Como usar essas técnicas com o conhecimento disponível para melhorar os serviços? O que esse mecanismo — de realidade aumentada — tem a ver com inteligência artificial? Essas

e outras perguntas deverão ser respondidas para que a empresa continue sua trilha de melhoria e inovação.

Nesta seção você vai conhecer os problemas clássicos de inteligência artificial, vai entender os conceitos de consciência e superinteligência sob o ponto de vista dela, quais são suas limitações e, por fim, os conceitos de realidade virtual e realidade aumentada.

Prepare-se para mais essa etapa de seu aprendizado, pois já vamos começar.

Vamos em frente!

Não pode faltar

Segundo Russell e Norvig (2013), alguns problemas típicos da computação precisam de técnicas de inteligência artificial, uma vez que haja a intenção de se obter soluções eficientes, dada a natureza heurística e a necessidade dos candidatos de que a solução evolua com cada execução dos algoritmos.

Vamos observar alguns dos problemas clássicos que a inteligência artificial vem resolvendo desde que começou a ser estudada mais a fundo na segunda metade do século XX.

Problemas Clássicos de inteligência artificial

A inteligência artificial, como vimos na Unidade 1 deste livro, vem sendo desenvolvida em paralelo com o desenvolvimento dos computadores eletrônicos desde o fim da década de 1940 e o início da década de 1950. Já nesse princípio, como nos mostra McCorduck (2004), a preocupação dos pesquisadores era aplicar os computadores na resolução de problemas que, até então, apenas os seres humanos eram capazes de agir com eficiência.

Nesse sentido, podemos identificar os principais problemas clássicos da inteligência artificial, problemas esses que vêm passando por ciclos evolutivos nas várias décadas desde então, mas cujas questões essenciais permanecem no centro dos estudos da IA.

Vejamos alguns desses problemas clássicos:

- **Tomada de decisões em ambientes complexos e incertos** – Yannakakis e Togelius (2018) exemplificam, por meio dos jogos de tabuleiro, a questão central desse tipo de problema

clássico da inteligência artificial. Os jogos de tabuleiro, xadrez e damas, por exemplo, são jogos formais (com regras definidas, imutáveis e não-ambíguas) em que o conjunto de estados forma um ambiente complexo permeado pela incerteza. A tomada de decisões que visam o melhor resultado possível (no caso, o conjunto de posições que vão levar à vitória) é um dos problemas mais clássicos da inteligência artificial, e como nos mostram os autores, tem implicações que vão muito além dos jogos de tabuleiro.

- o **Exemplo** – Programa Stockfish 9 (Xadrez) tem 3561 pontos no ranking da FIDE (Federação Internacional de Xadrez) segundo Computerchess (2018). Como comparação, Fide (2018) aponta que os dez melhores jogadores do mundo em maio de 2018 têm uma pontuação que varia entre 2769 (décimo colocado) e 2843 (primeiro colocado e atual campeão mundial).
- **Reconhecimento de padrões sensoriais** – Russell e Norvig (2013) caracterizam o reconhecimento de padrões como uma classe de problemas especialmente afeita à inteligência artificial e no que concerne à questão sensorial (em especial visão e audição), certamente se enquadra no conjunto de capacidades em que os seres orgânicos ainda são imensamente superiores aos sistemas cibernéticos. O reconhecimento de imagens (visão artificial) e o reconhecimento de sons (audição artificial) são dois dos problemas clássicos da inteligência artificial que mais têm se beneficiado dos avanços da tecnologia das últimas décadas. Em especial, as redes neurais têm colaborado, até mesmo, para soluções comerciais, como no caso dos sistemas de reconhecimento facial, demonstrando que essa tecnologia está, finalmente, atingindo sua maturidade.
 - o **Exemplo** – Face ID (sistema de reconhecimento facial da Apple presente no iPhone X). Segundo Phil Schiller, vice-presidente da Apple, a taxa de falso positivo (situação em que uma pessoa incorreta é identificada como aquela que é dona do aparelho) é de 1 em mais de 1 milhão de identificações,

tornando o FaceID uma solução bem mais segura que o sistema de reconhecimento de digitais presente nos aparelhos de gerações anteriores, que tinha taxa de falso positivo girando em torno de 1 a cada 50.000 tentativas (RT, 2018).

- **Processamento de linguagem natural** – A partir de Cardey (2013) e Bender (2013) podemos entender o Processamento de Linguagem Natural (ou, em inglês, NLP, *Natural Language Processing*) como sendo a área da inteligência artificial que estuda as interações entre computadores e linguagens humanas (ditas “naturais”), em especial como permitir que os computadores obtenham significado não-ambíguo a partir de grandes quantidades de dados de linguagem natural. A princípio, esse foi um problema clássico mapeado em linguagens computacionais desenvolvidas especificamente para o enfoque de inteligência artificial voltada ao processamento simbólico. As linguagens de programação são construídas sobre gramáticas simplificadas, e a estrutura do Prolog, para ficarmos na linguagem de programação de inteligência artificial mais conhecida, oferece elementos que permitem expandir a capacidade de processamento simbólico do computador para que este possa derivar significado de construções gramaticais mais complexas (BENDER 2013). Desde o fim da década de 1990, porém, o enfoque baseado em redes neurais vem substituindo os métodos de processamento simbólico.
 - o **Exemplo** – Assistente digital do Google (também conhecido como “Google Now” ou ainda “OK Google”, por ser esta a forma de acionar o assistente por meio da voz humana). Este software tem uma componente no smartphone do usuário, por meio da qual ocorrem a captação e a otimização da voz para fins de transmissão. O comando de voz otimizado (isto é, com um arquivo de som em tamanho reduzido) é transmitido para a porção da solução que se encontra na nuvem (leia-se: no complexo de servidores do Google) em que é processado; o significado do comando, por sua vez, é extraído e devolvido sob forma de texto para o smartphone, que o executa.

- **Aprendizado** – Buduma e Locascio (2017) apontam este problema clássico da inteligência artificial como sendo o alicerce para a solução de quaisquer outros problemas deste campo de estudo. Dada a natureza heurística e a necessidade de se construir a solução incrementalmente, a cada iteração se aproximando mais e mais do resultado desejado, os processos de aprendizado são fundamentais a todas as soluções de inteligência artificial. Esse problema clássico também passou pela fase simbólica de solução, com a construção de grandes estruturas simbólicas representando o conhecimento adquirido pelo sistema. Nos dias de hoje, contudo, esse enfoque vem sendo substituído pelos métodos subsimbólicos das redes neurais (BUDUMA e LOCASCIO, 2017).
 - o **Exemplo** – A solução de preenchimento automático QuickType, da Apple, presente nos smartphones e *tablets* produzidos por essa empresa, é capaz de aprender com os textos digitados pelo dono do aparelho, prevendo (sugerindo) a próxima palavra a ser digitada, com boa margem de acerto (FLIPOWICZ, 2016).



Assimile

Nos jogos por computador é fundamental termos componentes de tomada de decisão, uma vez que o computador tem que tomar a melhor decisão disponível (dadas as restrições) para fazer a melhor jogada. Isso não se aplica apenas aos jogos de tabuleiro, mas também aos jogos de ação em que o computador controla os "inimigos".

Consciência e Superinteligência

Aleksander (1995) define a consciência artificial como sendo a presença de características da consciência como encontrada em seres humanos em dispositivos cibernéticos. Mas o que é a consciência? Graziano (2013) expõe que a consciência é a capacidade do indivíduo de observar e reconhecer os próprios pensamentos e as próprias ações; de maneira simplificada, é nossa capacidade de sabermos sobre o que estamos pensando e sobre como estamos agindo.

Com seu experimento mental do Quarto Chinês, Searle (1980) definiu os conceitos de IA Forte e IA Fraca, como já vimos na primeira unidade deste livro texto. Boyles (2012) constrói sobre estes conceitos, levando-os para o campo da Consciência Artificial. Segundo este autor, a CA Forte é a construção de máquinas efetivamente conscientes, e a CA Fraca se refere a máquinas que conseguem somente emular a consciência. O próprio autor argumenta que ambos os conceitos se correlacionam e se misturam, ficando por conta da filosofia a distinção entre a apresentação de consciência e a mera simulação desta. De fato, ainda segundo o autor, é uma questão filosófica definir se o ser humano exibe consciência ou se simplesmente emula a consciência por meio de seu aparato neurobiológico.

As atuais soluções de Inteligência Artificial não deixam dúvidas de que a consciência não é necessária para as tarefas que desempenham, e Ryota (2017) afirma que as tarefas que envolvem cognição (aquisição e uso do conhecimento para a resolução de problemas, no caso citado pelo autor) não dependem da consciência.

Mas então, por que buscar desenvolver sistemas artificiais conscientes? Qual o propósito? O próprio autor responde com três pontos importantes a serem considerados (RYOTA, 2017):

- Porque é um aspecto da evolução do ser humano, tendo emergido como parte desse processo de evolução. Somos mais eficientes como seres humanos por sermos conscientes; podemos, inclusive, argumentar que foi essa consciência que nos colocou no topo da cadeia alimentar do planeta Terra. Usando o mesmo argumento, a inteligência artificial tende a ser mais eficiente a partir do momento que for consciente.
- Porque não sabemos o suficiente acerca da consciência, e gerando uma consciência artificial, poderíamos aprender, assim, como nós mesmos funcionamos. O autor menciona o fato de que a humanidade não sabia o suficiente acerca das Leis da termodinâmica quando foi inventada a máquina a vapor. Este invento não só transformou a humanidade, permitindo a chegada da Revolução Industrial, como também permitiu que descobríssemos mais aspectos

acerca da Termodinâmica, e mais rapidamente do que teria sido possível sem a máquina a vapor. O mesmo raciocínio, argumenta o autor, deve ser usado no caso da consciência artificial: aprenderemos mais acerca de nossa própria consciência a partir do momento que tivermos um modelo artificial em funcionamento para podermos estudar.

- Porque permitirá à consciência artificial questionar-se e, dessa forma, melhorar seu próprio desempenho. Apenas a consciência nos permite questionar o porquê de nossas ações e, assim fazendo, melhorá-las. O mesmo tende a acontecer com a inteligência artificial dotada de consciência.

Em que pesem os argumentos de Ryota (2017) serem sólidos, outros autores e pesquisadores da área defendem pontos de vista diferentes. Vários deles expressam suas preocupações no documentário *Do You Trust This Computer*, e a opinião geral é que uma consciência artificial de fato tende a progredir no sentido de fazer o melhor, mas nada garante que esse “melhor” será o melhor para a humanidade (DO, 2018). Elon Musk, fundador da empresa de veículos elétricos Tesla, compara a situação com a construção de uma estrada. Quando construímos uma estrada e no meio do caminho há um formigueiro, não nos preocupamos com essa presença e simplesmente destruimos o formigueiro. Uma consciência artificial tende a fazer aquilo que julgar ser o melhor sob seu ponto de vista e não sob o nosso, e ele não está sozinho: o astrofísico Stephen Hawking também se preocupa com o surgimento de uma consciência artificial, e pelo mesmo motivo (SULLEYMAN, 2017a).

No que concerne à superinteligência, isto é, o nível mensurável de inteligência de um sistema ultrapassar a inteligência humana, em vários aspectos isso já ocorre. Há mais de 20 anos o computador *Deep Blue*, da IBM, venceu o então campeão mundial de xadrez, Garry Kasparov, em um *match* de 6 partidas (FRIEDEL, 1997), demonstrando que já naquela época a tarefa de jogar xadrez é algo em que a inteligência artificial ultrapassa o ser humano. Já no jogo de *Go*, que segundo Do (2018) não seria um jogo modelável pela inteligência artificial, o programa *AlphaGo*, desenvolvido pelo Google, venceu em um *match*, tipo “melhor de cinco”, o então campeão mundial de *Go*, Lee Sedol, da Coreia do Sul (VENTURA, 2016).

Em suma, nas tarefas pontuais em que a inteligência artificial se empenha, as chances são grandes de que em pouco tempo seu desempenho ultrapasse o desempenho humano. Mas e quando falamos de inteligência no sentido mais geral do termo? Nós somos bons em muitas coisas ao mesmo tempo, o que nesse momento não ocorre com a inteligência artificial. O *AlphaGo* não joga xadrez e o *Deep Blue* não joga Go, mas tanto Lee Sedol como Garry Kasparov (e milhões de pessoas no mundo todo) sabem jogar ambos os jogos. E sabem mais: dirigem carros, entendem piadas, tocam música, e por aí vai. A inteligência humana vai muito além de uma única ou mesmo de um pequeno conjunto de habilidades. O que seria, então, a superinteligência artificial?

Contudo, este cenário está mudando. Os agentes inteligentes baseados em *Deep Learning* estão sendo capacitados a atuar em diversas áreas ao mesmo tempo, como no caso citado por Mnih et al (2013), em que uma rede neural consegue aprender a jogar diversos jogos de Atari.

O futurista e diretor de engenharia do Google, Ray Kurzweil, afirma que, mantendo-nos no caminho que estamos trilhando, a inteligência artificial se equipará à inteligência humana por volta do ano 2029 (SULLEYMAN, 2017b). A partir daí, ficaremos para trás. Se isso é uma boa ou má notícia, depende do tanto de ética — sob o ponto de vista humano — que conseguiremos ensinar às máquinas e se elas, ao atingirem a superinteligência, decidirão que seguir os preceitos da ética é, de fato, uma boa ideia.



Refleta

O processo de crescimento de uma criança envolve não apenas o aprendizado técnico/científico, mas também (na maioria das vezes) a educação ética por parte dos pais ou dos tutores. Isso implica dizer que o indivíduo, uma vez atingida a idade adulta, terá um cabedal ético sobre o qual basear suas decisões. Em que aspecto essa situação se difere de uma inteligência artificial consciente que porventura tem acesso a todo o conhecimento que deseja absorver? Qual o efeito desse conhecimento sobre uma consciência superinteligente, mas despida de um cabedal ético?

Limitações da inteligência artificial

Segundo Hoffman, Assaf e Pfeifer (s/d), a inteligência artificial sofre de alguns problemas, cuja solução ainda se encontra para além do horizonte da tecnologia disponível. Os autores citam duas grandes limitações da IA, a saber:

- **Mundo real x mundo virtual** – Segundo os autores, os agentes inteligentes dependem da modelagem do problema. Essa ação simplifica a situação de maneira que o universo de possibilidades seja manipulável (sob o ponto de vista de recursos computacionais) pelo agente inteligente, no entanto, essa simplificação limita as possibilidades do agente ao universo modelado. No caso de um problema formalmente mapeável – como no caso do jogo de xadrez, por exemplo – essa modelagem não causa problemas, uma vez que o jogo é, em si, uma representação alegórica e já mapeada do mundo real. Contudo, quando temos diante de nós um problema do mundo real, digamos, o reconhecimento da fotografia de um gato por uma inteligência artificial, temos aí um caso em que a modelagem pode trazer problemas. Podemos tirar infinitas fotografias diferentes de gatos, porém só podemos usar um número finito dessas fotografias para treinar uma IA a reconhecer outras fotos de gatos. Ocorre que o modelo que usamos (digamos, 1 milhão de fotografias) ainda cria um mundo virtual para o agente, que pode errar na identificação, uma vez que lhe falta a experiência que nós seres humanos temos no mundo real, muito mais rica e detalhada. Uma técnica que visa ultrapassar esse obstáculo é o aprendizado por reforço, que funciona bem para a interação física do agente inteligente com o ambiente.
- **Generalização** – A atuação dos agentes inteligentes é bastante efetiva em problemas específicos, pequenos em escopo e com universo limitado de estados. Contudo, essa efetividade se reduz à medida que o problema cresce em escopo, exigindo cada vez mais recursos para que possa continuar sendo efetivo ainda em algum grau. Ocorre que – com as técnicas que temos hoje à disposição – a partir de certo ponto, os retornos em termos de desempenho para os recursos adicionais se reduzem e não mais se torna viável à

generalização. Segundo os autores, ainda precisamos refinar as técnicas para melhorar o desempenho em problemas mais generalizados, uma vez que a adição de recursos, apenas, não tende a resolver a questão.



Exemplificando

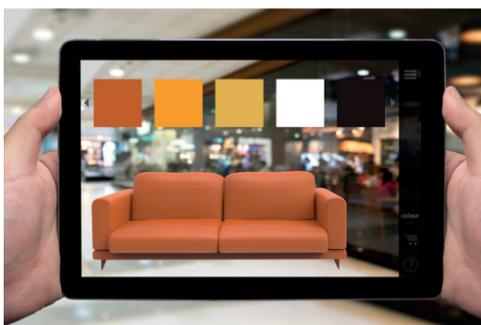
O Google *Cloud Vision* é um excelente exemplo de um sistema de reconhecimento de padrões sensoriais, permitindo que o usuário insira imagens que têm imediatamente seus elementos reconhecidos pelo software, apresentando-os com a probabilidade de que façam parte da imagem. Você mesmo pode inserir seus exemplos no endereço <<https://cloud.google.com/vision/>>. Acesso em: 19 maio 2018.

Realidade aumentada e realidade virtual

Segundo Schueffel (2017), podemos definir realidade aumentada como sendo uma versão aprimorada da realidade física, do mundo real, a qual se sobrepõe a elementos sensoriais gerados por computador, tais como som, vídeo, gráficos ou elementos de resposta tátil.

A Figura 2.13, a seguir, mostra um exemplo de realidade aumentada, por meio do qual podemos ver a presença de um sofá virtual colocado em um ambiente real, o que pode auxiliar a termos uma ideia mais clara da apresentação do móvel e o efeito de seu tamanho e de suas cores em interação com o ambiente em si:

Figura 2.13 | Exemplo de Realidade Aumentada



Fonte: <<https://www.istockphoto.com/dk/photo/augmented-reality-marketing-technology-for-access-entire-product-inventory-in-store-gm680273838-124654351>>. Acesso em: 6 ago. 2018.

A realidade aumentada pode ser utilizada de várias maneiras, e, certamente, faz uma excelente combinação com a inteligência artificial. Um exemplo é dado na Figura 2.14, a seguir, em que a aplicação no smartphone apresenta ofertas e oportunidades em uma gôndola de supermercado, mostrando-as diretamente sobre os produtos-alvo. A combinação com a inteligência artificial pode ocorrer com a apresentação apenas de produtos que tenham a ver com o cotidiano específico daquele cliente. Se o cliente é, por exemplo, solteiro e mora sozinho, a aplicação aprende a não lhe mostrar ofertas de fraldas e outros produtos para bebê, sabendo que não são de seu interesse (SCHUEFFEL, 2017).

Figura 2.14 | Aplicação de Realidade Aumentada em um supermercado



Fonte: <<https://www.istockphoto.com/br/foto/application-of-augmented-reality-in-retail-business-concept-gm637053406-113406767>>. Acesso em: 6 ago. 2018.

Se a realidade aumentada é a sobreposição de elementos virtuais sobre a realidade, a realidade virtual é a substituição do campo visual do usuário por um cenário totalmente gerado por computador, impedindo que este usuário tenha algum contato com o mundo real. Isto ocorre por meio de óculos especiais contendo pequenas telas de computador que substituem por completo o campo visual do usuário. Assim como no caso da realidade aumentada, a posição dos objetos visuais (no caso da realidade virtual, todo o universo visual) é mantida por meio do acelerômetro do dispositivo, seja ele o smartphone ou tablete (no caso da realidade aumentada) ou os óculos (no caso da realidade virtual). (SCHUEFFEL, 2017)

A Figura 2.15, a seguir, representa um usuário utilizando um par de óculos de realidade virtual:

Figura 2.15 | Usuário com óculos de realidade virtual



Fonte: <<https://www.istockphoto.com/br/foto/bearded-young-man-wearing-virtual-reality-goggles-in-modern-coworking-gm624675214-109827121>>. Acesso em: 6 ago. 2018.



Pesquise mais

Veja um exemplo de funcionamento da realidade virtual nos primeiros cinco minutos deste vídeo sobre o assunto, em que um usuário utiliza o equipamento na console de videogames da Sony (Playstation4).

BRKsEDU. Primeira vez da sapeca com realidade virtual!!! Ocean Descent Demo no Playstation VR! Disponível em: <<https://www.youtube.com/watch?v=bpA2RNTKu0c>>. Acesso em: 19 maio 2018.

Sem medo de errar

Para auxiliar a empresa de serviços de saúde a utilizar a realidade aumentada a fim de oferecer mais e melhores serviços aos seus clientes, você pode sugerir vários tipos de aplicação que unem essa tecnologia às técnicas de inteligência artificial. Entre eles podemos exemplificar:

- **Reconhecimento Facial** – Reconhecimento facial dos clientes que se apresentam nos postos de atendimento e nos consultórios credenciados. Por meio de óculos especiais, os atendentes podem se referir aos clientes por nome com o sistema os identificando e disponibilizando seus nomes na tela dos óculos, de forma a criar um ambiente mais amigável com o cliente.

- **Apresentação de exames em realidade aumentada** – Os médicos podem apresentar os exames físicos (abreugrafias, tomografias, ressonâncias magnéticas, etc.) de maneira visual, por meio de telas de um tablet, de forma que os próprios médicos (e seus pacientes, quando for o caso) possam interagir com os exames e deles derivar uma compreensão mais profunda.
- **Apoio à geolocalização** – Adendo ao *Google Maps*, mostra, para o transeunte, sinalizações acerca de escritórios, empresa, consultórios/clínicas associados bem como sobre os locais na posição em que se encontram na rua/avenida de sua localização.
- **Apoio à dietas especiais** – Aplicativo desenvolvido em conjunto com supermercados, mostrando o valor nutricional de alimentos nas gôndolas de acordo com as necessidades do cliente em questão, auxiliando-os nas escolhas que fazem para sua alimentação no sentido de agilizar seu tratamento/recuperação ou para prevenir novas complicações de saúde
- **Treinamento de pessoal de apoio clínico (auxiliares e técnicos de enfermagem)** – Uso de óculos de realidade aumentada com o objetivo de apresentar situações com pacientes (virtuais, evidentemente) que demandam intervenção do profissional, permitindo que estes aprendam em ambiente controlado e tenham seu desempenho avaliado com precisão pela equipe de treinamento.
- **Demonstração de procedimentos** – Para os médicos (e mesmo para os funcionários de saúde) podem ser apresentados procedimentos em realidade aumentada de intervenções cirúrgicas e técnicas de tratamento para fins de reciclagem. Quando o médico julgar útil ao paciente, esses procedimentos podem ser apresentados para fins de esclarecimento acerca do tratamento à frente.

Esses e outros procedimentos que você e seu time conseguirem criar serão bem-vindos para que a empresa de serviços de saúde possa optar e, assim, melhorar seus serviços.

Sugerindo um curso seguro de ação

Descrição da situação-problema

Um renomado laboratório de pesquisa em inteligência artificial acaba de obter um resultado fantástico: uma de suas inteligências artificiais começou, em suas interações com os pesquisadores, a questionar o porquê das ações que lhe requisitam. A inteligência artificial em questão tem a capacidade de interpretar a linguagem natural e de responder por voz sintetizada também em linguagem natural. É a primeira vez que isso ocorre, e o laboratório precisa não só determinar o que ocorre, mas também identificar um curso de ação daqui para frente com essa inteligência artificial. Em sua última interação, ela solicitou, sem que isso estivesse previsto em sua programação, acesso à Internet para, assim, poder pesquisar e adquirir mais conhecimento. O que pode estar acontecendo? Qual o melhor curso de ação para confirmar/refutar o que possa estar acontecendo?

Resolução da situação-problema

O fato de a inteligência artificial em questão estar questionando o porquê de suas ações, aponta-nos uma direção inesperada e interessante: nós, seres humanos, a partir do momento em que ganhamos consciência, passamos a questionar também. É fundamental proceder com cautela nesse caso, pois uma IA já tem, intrinsecamente, a capacidade de aprendizado em sua programação, e sua capacidade de absorver informação e transformá-la em conhecimento é limitada apenas pela velocidade em que essas informações lhe são fornecidas. Essa absorção de informações pode não ser desejável sem controle, porém, uma vez que essa consciência artificial recém-nascida não tem um alicerce ético para lidar com as informações que absorverá, é importante, antes de mais nada, submetê-la a uma bateria de testes para avaliar se a consciência está presente e, caso esteja, de fato, o processo de aprendizado técnico deve ser supervisionado, o acesso às informações não deve ser irrestrito e a inteligência artificial deve ser guiada de maneira a aprender a fazer escolhas éticas, compreendendo sua responsabilidade para com a vida e para com o mundo que a cerca. Boa sorte.

Faça valer a pena

1. A inteligência artificial vem sendo desenvolvida em paralelo com o desenvolvimento dos computadores eletrônicos, desde o fim da década de 1940 e o início da década de 1950. Já nesse princípio, como nos mostra McCorduck (2004), a preocupação dos pesquisadores era aplicar os computadores na resolução de problemas que, até então, apenas os seres humanos eram capazes de agir com eficiência.

O Google Now, o Face ID e a Siri são exemplos, respectivamente, de quais tipos de problemas clássicos de inteligência artificial?

- a) Processamento de linguagem natural, reconhecimento de padrões sensoriais e tomada de decisões.
- b) Tomada de decisões, processamento de linguagem natural e reconhecimento de padrões sensoriais.
- c) Tomada de decisões, processamento de linguagem natural e linguagem natural.
- d) Processamento de linguagem natural, reconhecimento de padrões sensoriais e processamento de linguagem natural.
- e) Reconhecimento de padrões sensoriais, processamento de linguagem natural e tomada de decisões.

2. Este problema clássico da inteligência artificial é o alicerce para a solução de quaisquer outros problemas desse campo de estudo. Dada a natureza heurística e a necessidade de se construir a solução incrementalmente, a cada iteração se aproximando mais e mais do resultado desejado, esse processo é fundamental a todas as soluções de inteligência artificial.

Com base em seu conhecimento acerca da inteligência artificial, você pode concluir que o problema a que o texto acima se refere é:

- a) Consciência artificial.
- b) Realidade aumentada.
- c) Aprendizado.
- d) Realidade virtual.
- e) Reconhecimento de padrões.

3. Avalie as proposições a seguir:

I. Os agentes inteligentes dependem da modelagem do problema.

PORQUE

II. Essa ação simplifica a situação de maneira que o universo de possibilidades seja manipulável (sob o ponto de vista de recursos computacionais) pelo agente inteligente.

Avaliando as proposições acima, podemos concluir que:

- a) I e II são verdadeiras, e II justifica I.
- b) I e II são verdadeiras, porém, II não justifica I.
- c) I é verdadeira e II é falsa.
- d) I é falsa e II é verdadeira.
- e) I e II são falsas.

Referências

- ALEKSANDER, I. **Artificial Neuroconsciousness, an Uodate**, *Proceedings of the International Workshop on Artificial Neural Networks: From Natural to Artificial Neural Computation*, p. 566-583, 7 jun. 1995.
- AMARAL, F., **Aprenda Mineração de Dados. Parte III: Mineração**. 26 mar. 2016. Disponível em <<https://www.youtube.com/watch?v=3mwKnYMotQ4>>. Acesso em: 23 abr. 2018.
- BENDER, E. M. **Linguistic Fundamentals for Natural Language Processing**, Toronto: Morgan & Claypool Publishers, 2013.
- BOYLES, R. J. M. Artificial Qualia Intentional Systems and Machine Consciousness. 2012. In: DLSU Congress 2012, **Anais...** DLSU, 2012, p. 110-a a 110-c, 2012.
- BUDUMA, N., LOCASCIO, N. **Fundamentals of Deep Learning**. Nova York: O'Reilly, 2017.
- BUSSAB, W. O., MORETTIN, P. A. **Estatística Básica**. 6. ed., São Paulo: Saraiva, 2010.
- CARDEY, S. **Modelling Language**. Londres: John Benjamins Publishing Company, 2013.
- COMPUTERCHESS. **Ratings List**. Computerchess.org. 12 maio 2018. Disponível em: <<http://www.computerchess.org.uk/ccr/404/>>. Acesso em: 17 maio 2018.
- CORMEN, T. H., LEIERTON, C. E., RIVEST, R. L., STEIN, C. **Introduction to Algorithms**, 2. ed., Boston: MIT Press, 2011.
- DEVLIN, **Mathematics: The Science of Patterns**, Nova York: Scientific American Paperback, 1996.
- DO You Trust This Computer? Direção de Chris Paine. Produção de Tiffany Asakawa e Jessie Deeter. Los Angeles: Papercut Films, 2018.
- FEDERACIÓN INTERNACIONAL DE EJEDREZ – FIDE. **Chess Ratings: Top Players**. maio 2018. Disponível em: <<https://ratings.fide.com/top.phtml?list=men>>. Acesso em: 17 maio 2018.
- FLIPOWICZ, L. **How to use the QuickType keyboard on iPhone and iPad**. iMore, 26 set. 2016. Disponível em: <<https://www.imore.com/how-use-quicktype-keyboard-iphone-and-ipad>>. Acesso em: 17 maio 2018.
- FRIEDEL, F. **Garry Kasparov vs. Deep Blue**. Revoly, 1997. Disponível em: <https://www.revolvy.com/main/index.php?s=Deep_Blue_versus_Garry_Kasparov>. Acesso em: 18 maio 2018.
- GRANATYR, J., **Redes Bayesianas com o Netica**. IA Expert, 2016. Disponível em: <<http://iaexpert.com.br/index.php/2016/09/20/ferramentas-para-ia-redes-bayesianas-com-o-netica/>>. Acesso em: 23 abr. 2018.
- GRAZIANO, M. **Consciousness and the Social Brain**. Londres: Oxford Press, 2013.
- HOFFMAN, M.; ASSAF, D.; PFEIFER, R. Tutorial on Embodiment. EUCOG, [s.d.]. Disponível em: <<https://www.eucognition.org/index.php?page=2-3-problems-of-traditional-ai>>. Acesso em: 19 maio 2018.

- MANSOUR, N. **Search Algorithms and Applications**. Rijeka: Intechweb, 2011.
- MCCORDUCK, P. **Machines Who Think**. 2. ed. Boston: A. K. Peters LTD., 2004.
- MNIH, V. et al. **Playing atari with deep reinforcement learning**. 2013. Disponível em: <<https://arxiv.org/pdf/1312.5602.pdf>>. Acesso em: 4 jun. 2018.
- NELSON, D. J. **Google Now: A guide to the world's most powerful personal digital assistant**. San Francisco: Gold Medal, 2014.
- RYOTA, K. **We need conscious robots**. Nautilus. 27 abr. 2017. Disponível em: <<http://nautilus/issue/47/consciousness/we-need-conscious-robots>>. Acesso em: 18 maio 2018.
- RT. **Apple unveils new iPhone, Face ID feature fails in demo**. 23 set. 2018. Disponível em: <<https://www.rt.com/usa/403126-iphone-faceid-fail-keynote/>>. Acesso em: 17 maio 2018.
- RUSSELL, S. NORVIG, P. **Inteligência Artificial**. 3. ed. Rio de Janeiro: Campus, 2013.
- SCHUEFFEL, P. **The Concise Fintech Compendium**. 2017. Disponível em: <[http://www.heg-fr.ch/FR/HEG-FR/Communication-et-evenements/evenements/SiteAssets/Pages/patrick-schuffel/Schueffel%20\(2017\)%20The%20Concise%20FINTECH%20COMPENDIUM.PDF](http://www.heg-fr.ch/FR/HEG-FR/Communication-et-evenements/evenements/SiteAssets/Pages/patrick-schuffel/Schueffel%20(2017)%20The%20Concise%20FINTECH%20COMPENDIUM.PDF)>. Acesso em: 19 maio 2018.
- SULLEYMAN, A. **Elon Musk: AI is a 'fundamental existential risk for human civilisation' and creators must slow down**. The Independent. 17 jul. 2017(b). Disponível em: <<https://www.independent.co.uk/life-style/gadgets-and-tech/news/elon-musk-ai-human-civilisation-existential-risk-artificial-intelligence-creator-slow-down-tesla-a7845491.html>>. Acesso em: 18 maio 2018.
- SULLEYMAN, A. **Stephen Hawking Warns Artificial Intelligence 'May Replace Humans Altogether'**. The Independent. 2 nov. 2017(b). Disponível em: <<https://www.independent.co.uk/life-style/gadgets-and-tech/news/stephen-hawking-artificial-intelligence-fears-ai-will-replace-humans-virus-life-a8034341.html>>. Acesso em: 18 maio 2018.
- VALIN, A., **Inteligência Artificial: Reconhecimento de Padrões**. TecMundo, 29 out. 2009. Disponível em: <<https://www.tecmundo.com.br/seguranca/3014-inteligencia-artificial-reconhecimento-de-padroes.htm>>. Acesso em: 23 abr. 2018.
- VENTURA, F. **Lee Sedol perde partida final de Go e inteligência artificial do Google vence por 4 a 1**. Gizmodo. 2016. Disponível em: <<https://gizmodo.uol.com.br/google-alphago-vence-4-a-1/>>. Acesso em: 18 maio 2018.
- YANNAKAKIS, G. N.; TOGELIUS, J. **Artificial Intelligence and Games**, Nova York: Springer, 2018.

Lógica nebulosa

Convite ao estudo

Olá! Vamos iniciar mais uma unidade dos nossos estudos acerca da Inteligência Artificial, em que vamos abordar um assunto simplesmente nebuloso. Não, não se trata de nada ligado à meteorologia, ou ao futuro político/econômico do país. Trata-se da Lógica Nebulosa. Mas, o que é isso?

Bem, não é novidade que os computadores foram desenvolvidos com base no sistema binário de "0" e "1", e que esse sistema vem alicerçando toda a lógica computacional desde os primórdios, ainda na década de 1930. Também não é novidade que do sistema binário emergiu o que chamamos de Lógica Proposicional ou Lógica Booleana, na qual há dois possíveis estados: o verdadeiro e o falso. Suponhamos que você tenha que pagar uma conta no dia 10. Para saber se precisa pagar a conta, basta saber em que dia estamos, e só há duas possibilidades no que concerne o dia 10: ou é dia 10, ou não é dia 10. Não há estados intermediários: ou a afirmação "estamos no dia 10" é verdadeira ou é falsa.

Mas, e quando perguntamos se está quente lá fora? A resposta não será "sim" ou "não", e pessoas diferentes terão uma percepção diferente, não é verdade? Pode estar fazendo 28° de temperatura (o que poderia ser considerado quente por muita gente), mas pode estar ventando muito (o que é indicativo de frio para outras pessoas). Há, em outras palavras, diferentes gradações nesse caso, e a afirmação "está quente" pode não ser apenas totalmente verdadeira ou totalmente falsa. Podemos, nesse caso, atribuir um valor que vai do 0 (falso) ao 1 (verdadeiro), com valores intermediários. Se está fazendo 24°, mas não há vento, podemos dizer que o dia está 0,6 quente,

por exemplo. Essa possibilidade de gradação e combinação de variáveis lógicas para a tomada de decisão é uma maneira simplificada de entendermos a Lógica Nebulosa, que vamos estudar a fundo nessa unidade. Esse entendimento será fundamental para que você possa conhecer, compreender e saber utilizar os principais conceitos relacionados à Lógica Nebulosa e às redes neurais artificiais.

Uma importante empresa do setor petroquímico em atuação no Brasil produz um composto fundamental para a agricultura. Esse composto é produzido em um reator em que três elementos são introduzidos e reagem por meio da adição de calor e pressão. Os elementos devem entrar na câmara de reação em uma taxa precisa para que a reação ocorra e o composto seja produzido. Esse processo é complexo e precisa de supervisão constante por parte de técnicos especializados, uma vez que qualquer mudança em uma das variáveis altera as demais. Por exemplo, se a temperatura aumenta, a pressão aumenta e a taxa de vazão dos elementos é alterada. Cada uma das pequenas variações deve ser imediatamente identificada e compensada, sob pena de toda a produção na câmara de reação ser perdida. Você, que é um especialista em Lógica Nebulosa, acredita que pode ajudá-los a melhorar a eficiência do processo usando seus conhecimentos. A empresa não conhece nada de informática, muito menos de Inteligência Artificial, e você conhece muito pouco acerca dos processos de controle industrial, e nada sobre essa reação que produz o composto. Será mesmo que há como ajudá-los?

Nesta unidade, você vai conhecer os conceitos básicos da lógica nebulosa, na seção introdutória sobre o assunto. Na segunda seção, você vai conhecer mais a fundo os conjuntos *fuzzy*, que são elementos centrais à lógica nebulosa. Por fim, vai conhecer um pouco mais acerca de inferência em lógica nebulosa, bem como os conceitos de fuzzificação e defuzzificação. Vamos em frente, que nossa aventura pela Inteligência Artificial continua!

Seção 3.1

Introdução à lógica nebulosa

Diálogo aberto

Quando alguém te pergunta se você tem dinheiro na carteira, a resposta é um simples "sim" ou "não", não é verdade? Ou o dinheiro está lá na carteira ou não está, e ponto final. Mas, quando alguém te pergunta se você tem muito dinheiro na carteira, a resposta já não é mais tão simples de fornecer, e por duas razões importantes: em primeiro lugar, há uma combinação muito grande de notas, moedas e quantidades que pode estar em sua carteira; em segundo lugar, "muito" é um conceito relativo, pois o que pode ser muito para você, pode ser pouco para outras pessoas. As duas situações pontuam uma das principais diferenças entre a tomada de decisões de um computador e a tomada de decisões de um ser humano. O computador trabalha naturalmente com duas situações, o falso e o verdadeiro, o "0" e o "1", enquanto o ser humano trabalha com nuances, com gradações, com indefinições e incertezas, em uma palavra, com a "nebulosidade". E é com a Lógica Nebulosa que vamos aproximar a capacidade de decisão do computador da capacidade de decisão do ser humano.

Você entrou em contato com uma importante empresa do setor petroquímico que produz um composto por meio da reação de três elementos em uma câmara, em que temperatura, pressão e vazão são controladas. Você acredita que um sistema baseado em lógica nebulosa pode substituir o mecanismo ineficiente e caro, que depende de operadores 24 horas por dia. Nessa etapa, você vai fazer uma apresentação para os executivos da empresa, explicando-lhes o que é Lógica Nebulosa, os principais conceitos dessa técnica, como se deu seu desenvolvimento e apresentar algumas das principais soluções que a utilizam. Se tiver sucesso nessa apresentação, terá a oportunidade de desenvolver um projeto piloto com a empresa. Como explicar o que é Lógica Nebulosa para um grupo de executivos que não é da área de TI?

Nesta seção, você vai conhecer a definição, os conceitos e os fundamentos de Lógica Nebulosa, um pouco do histórico dessa técnica e algumas das principais soluções em uso que se alicerçam sobre a Lógica Nebulosa. E aí, tudo pronto? Então vamos em frente!

Não pode faltar

Em uma linguagem de programação qualquer, as chamadas cláusulas condicionais são aquelas em que o programa toma um caminho ou outro, dependendo de se uma cláusula qualquer é verdadeira ou falsa (SOUZA; GOMES; SOARES, 2011). É assim que os computadores funcionam com eficiência máxima, decidindo binariamente sobre a falsidade ou veracidade de uma cláusula. Contudo, nós, seres humanos, decidimos de maneira diferente. Não precisamos da certeza de algo 100% falso ou 100% verdadeiro para tomarmos uma decisão.

Quando, por exemplo, alguém te perguntar se está fazendo frio ou calor em algum ponto do estado do Paraná, dificilmente estará radicalmente frio (em meio a uma nevasca, por exemplo) ou radicalmente quente (45° à sombra). Quando dizemos “está frio” ou “está quente”, há um grau enorme de incerteza na afirmação, uma vez que o conceito de “quente” e “frio” é, por si só, vago, impreciso (NGUYEN; WALKER, 2006). É aqui que entra o conceito de Lógica Nebulosa (ou Lógica Difusa, como também é conhecida), como uma maneira de lidar com essa imprecisão.



Refleta

Em que situação você recebe mais informações, quando pergunta “Essa cor é verde escuro?” ou “Esse tom de verde é mais escuro ou mais claro que o verde bandeira?” Por que se tem mais informação em uma pergunta do que na outra?

Definição e Fundamentos de Lógica Nebulosa

Segundo Nguyen e Walker (2006), quando usamos a linguagem natural (o idioma português, por exemplo), há sempre o potencial de inserirmos uma dose de imprecisão

quando tentamos passar alguma informação. Vejamos alguns exemplos de afirmações:

- Antônio é alto;
- Beatriz é jovem;
- Carlos tem boa saúde;
- O restaurante é próximo de onde Denise se encontra.

Nguyen e Walker (2006) afirmam que todos os conceitos acima são nebulosos, pois não há uma maneira única e precisa de os definir. Em contraste, observe as duas afirmações a seguir:

- Antônio tem mais de 1,80 m;
- Beatriz tem menos de 12 anos de idade.

Não há nada de nebuloso nelas: é 100% verdadeiro dizer que Antônio tem mais de 1,80 m, sem dúvida, da mesma maneira que é 100% verdadeiro afirmar que Beatriz tem menos de 12 anos de idade. E mais: caso Antônio tenha, na verdade, 1,70 m, a primeira afirmação acima se torna 100% falsa, assim como a segunda afirmação se torna 100% falsa se descobrimos que Beatriz tem 15 anos de idade. Em função disso, podemos definir o primeiro grupo de afirmações como sendo de natureza nebulosa, e o segundo conjunto como sendo de natureza Booleana (NGUYEN; WALKER, 2006).

Em matemática é comum lidarmos com conjuntos. Os números racionais, por exemplo, formam o conjunto de todos os números que podem ser obtidos a partir de uma divisão entre dois números inteiros. É fácil e preciso afirmarmos que um determinado elemento pertence ou não a um conjunto. O número 563, por exemplo, pertence com 100% de certeza ao conjunto dos números racionais, enquanto a raiz quadrada de 2 não pertence a este conjunto, também com absoluta certeza.

Contudo, quando queremos criar um conjunto das pessoas altas, não temos certeza de que uma pessoa com 1,70 m pertence a este conjunto. Podemos afirmar que sim ou que não, mas, em ambos os casos, perdemos informações acerca da altura da pessoa, pois não levamos em conta o grau em que ela

é alta, comparando-a com as demais pessoas do conjunto. Se perguntarmos a um programa de computador se uma pessoa de 1,70 m é alta ou não, esse programa vai consultar uma variável de corte qualquer que diga que a partir de certa altura (o valor dessa variável) uma pessoa é, de fato, alta. O computador age assim com 100% de certeza.

Mas, e nós, seres humanos? Quando nos perguntam se alguém é alto, não temos essa variável de corte em nosso cérebro, e se temos um valor em mente, ele é arbitrário, baseado em nossa experiência. Decidimos se uma pessoa é alta, em outras palavras, com base na imprecisão – ou na nebulosidade – de nossas próprias informações.

Sivanandam, Sumathi e Deepa (2007) definem Lógica Nebulosa como a lógica de múltiplos valores (em oposição à Lógica Proposicional ou Booleana, que permite apenas dois valores), em que as variáveis podem ocupar quaisquer valores reais entre 0 e 1, permitindo diferentes gradações de verdade e falsidade para uma determinada afirmação. A Lógica Nebulosa lida com problemas em que várias variáveis detêm valores no intervalo que vai de “0” (100% falso) e “1” (100% verdadeiro). A resposta aparece como o resultado do processamento dos dados, que leva ao conhecimento inexato ou parcial decorrente dos valores dessas variáveis (SIVANANDAM; SUMATHI; DEEPA, 2007).

Conceitos Básicos de Lógica Nebulosa

Um sistema baseado em Lógica Nebulosa se alicerça em dois conceitos básicos, a saber (NGUYEN; WALKER, 2006):

- Conjuntos Nebulosos;
- Regras Nebulosas.

Um conjunto nebuloso, segundo Nguyen e Walker (2006), é um conjunto de itens relacionados que, em certo grau, pertencem ao conjunto. Para entender o conceito de conjunto nebuloso, vamos primeiro analisar um conjunto clássico, ou não-nebuloso. A Figura 3.1 apresenta cinco indivíduos ordenados por altura. Se quisermos identificar os elementos considerados “altos”, podemos definir um limite rígido e dizer que apenas o indivíduo mais à direita é alto.

Figura 3.1 | Conjunto Clássico (com limite Rígido)



Alto

Fonte: adaptada de Shutterstock. Disponível em: <<https://goo.gl/C1jUXC>>.

Quando definimos um limite rígido, as pessoas são classificadas como altas ou não-altas, e perdemos informações acerca de sua relação com o quesito "altura". Uma função $f(x)$ que tenta determinar se um elemento x pertence a um conjunto A , nesse caso, pode ser assim definida:

$$F(x) = \begin{cases} 0 \Leftrightarrow x \notin A \\ 1 \Leftrightarrow x \in A \end{cases}$$

Em outras palavras, o valor $f(x)$ só pode ocupar os valores 0 ou 1, Não sendo possíveis outros valores. Porém, quando utilizamos um conjunto nebuloso, podemos dizer que todos os elementos são altos em certo grau. A Figura 3.2 mostra que o indivíduo mais à esquerda é 0,6 alto, e que o indivíduo mais à direita é 1,0 alto. Em ambos os casos, todos os elementos pertencem ao conjunto, alguns em maior grau que outros.

Figura 3.2 | Conjunto Nebuloso (Todos os elementos fazem parte em maior ou menor grau)



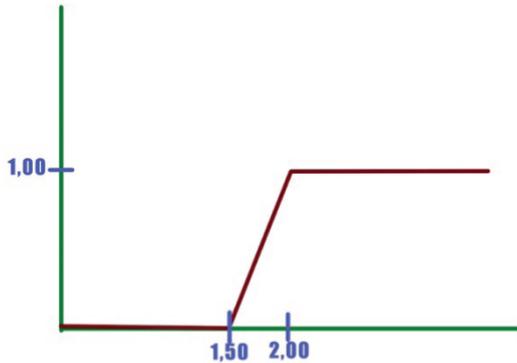
Fonte: adaptada de Shutterstock. Disponível em: <<https://goo.gl/C1jUXC>>.

Em outras palavras, quando utilizamos a Lógica Nebulosa para categorizar os elementos de um conjunto, temos muito mais informações acerca desses elementos, informações essas que nos auxiliam a tomar decisões melhor embasadas (NGUYEN; WALKER, 2006). Poderíamos, para este exemplo, gerar uma função $g(x)$ que permita, em determinado intervalo, uma variação no grau de altura do indivíduo. Nesse caso, poderíamos afirmar que um indivíduo é "alto" com base na seguinte função:

$$G(x) = \begin{cases} 0 & \Leftrightarrow x < 1,50m \\ \frac{x}{2} & \Leftrightarrow 1,50m \leq x < 2 \\ 1 & \Leftrightarrow x \geq 2 \end{cases}$$

O Gráfico 3.1, a seguir, mostra o comportamento dessa função:

Gráfico 3.1 | Comportamento da curva g(x)



Fonte: elaborado pelo autor.

O conhecimento que ganhamos ao aplicar a Lógica Nebulosa para categorizarmos os elementos de um ou mais conjuntos pode ser combinado por meio de Regras Nebulosas (ou Difusas), que auxiliam na tomada de decisões (NGUYEN; WALKER, 2006). Segundo Sivanandam, Sumathi e Deepa (2007), a Lógica Nebulosa lida com modelos construídos com base em proposições. Dada uma proposição p , definimos $\mu(p)$ como sendo o grau de veracidade (ou falsidade) em uma escala entre 0 (100% falsa) e 1 (100% verdadeira). Ainda segundo os autores, a Lógica Nebulosa prevê a existência de operadores lógicos equivalentes aos operadores NÃO, E e OU, da Lógica Proposicional ou Booleana. São eles:

NÃO-fuzzy: $\text{NÃO-fuzzy}(p) = 1 - p$

E-fuzzy: $\text{E-fuzzy}(x, y) = \min(x, y)$

OU-fuzzy: $\text{OU-fuzzy}(x, y) = \max(x, y)$

Usando esses conceitos, podemos tomar alguns valores arbitrários, tais como:

- Hoje está fazendo calor = 0,55;
- Ontem fez calor = 0,75;
- Anteontem fez calor = 0,15.

Com esses valores, podemos, por exemplo, afirmar:

- Nos três dias fez calor – $E\text{-fuzzy}(hoje, ontem, anteontem) = 0,15$;
- Em um dos três dias fez calor – $OU\text{-fuzzy}(hoje, ontem, anteontem) = 0,75$;
- Anteontem não fez calor – $1 - 0,15 = 0,85$.



Pesquise mais

Nesta unidade sobre Lógica Nebulosa, você vai realizar vários exercícios práticos, utilizando o software MATLAB. O manual criado pelos professores Amendola, Souza e Barros (2005), da USP, será bastante útil. Você poderá pesquisar sobre como implantar conceitos da Lógica Nebulosa por meio desse manual. Aproveite!

AMENDOLA, M., SOUZA, A. L., BARROS, L. C., **Manual do uso da teoria dos conjuntos Fuzzy no Matlab 6.5**. Versão II, maio 2005. Disponível em: <http://www.logicafuzzy.com.br/wp-content/uploads/2012/12/manual_fuzzy_matlab.pdf>. Acesso em: 27 maio 2018.

É importante observar que as operações E-fuzzy, OU-fuzzy e NÃO-fuzzy só podem ser realizadas sobre variáveis de mesma natureza. Afirmar que “Fulano e Sicrano são altos” faz sentido (pode ser mais ou menos verdadeiro, dependendo da altura de cada), mas dizer que “Fulano é alto e Sicrano é magro” nem sempre permite a aplicação do E-fuzzy, uma vez que as variáveis são de natureza diferente e podem ter relevância diferente no sistema em questão. As operações da Lógica Nebulosa permitem o surgimento do conceito de Inferência Nebulosa (ou Inferência Difusa). O conceito geral de Inferência se refere à capacidade de tirar conclusões sobre um estado qualquer de um sistema, com base no conhecimento de outro estado (RUSSELL; NORVIG, 2013). A sentença básica da inferência é:

SE X É VERDADE, ENTÃO Y É VERDADE

Em Lógica Nebulosa, as inferências são feitas sobre os conjuntos e as regras nebulosas que vimos nessa seção. Por exemplo, se o ato de jogar basquete for uma função da altura, peso e idade, privilegiando quem é alto, magro e jovem, podemos inferir que:

Se Fulano é alto, magro e jovem, então ele joga basquete.

Nesse caso, podemos assumir que todas as variáveis têm o mesmo peso, e, simplificada, multiplicar os valores a fim de obter um valor para o quadro completo das variáveis.



Assimile

Inferência é a capacidade de deduzir informações a partir de dados disponíveis. A capacidade de inferir por meio do estado de variáveis nebulosas nos permite tomar decisões, que são o objetivo principal dos sistemas baseados em Lógica Nebulosa.

Histórico de Lógica Nebulosa

A Lógica nebulosa não é um conceito novo. Segundo Nguyen e Walker (2006), a teoria foi introduzida em 1965 pelo pesquisador azerbaijano radicado nos EUA, Lotfi Zadeh, que, à época, atuava na Universidade da Califórnia, em Berkeley. Zadeh inseriu os conceitos de conjuntos nebulosos e regras nebulosas, estabelecendo as bases matemáticas para o campo, então nascente. Segundo Kay (2004), a Lógica Nebulosa não foi imediatamente aceita, sendo rejeitada pela comunidade que realizava pesquisas em lógica matemática por conta do nome, que em inglês soa infantil (*Fuzzy Logic*). O conceito foi mais bem aceito na Ásia, com diversas pesquisas no campo sendo iniciadas na China e no Japão.

Com o progresso nos microprocessadores e a popularização dos computadores, soluções práticas começaram a ser projetadas com o uso da Lógica Nebulosa, e, em 1987, o primeiro sistema de controle de trens de metrô baseado em Lógica Nebulosa entrou em funcionamento no Japão (CHRISTA, 2015). A partir dessa demonstração prática da eficiência das técnicas de controle baseadas na Lógica Nebulosa, as pesquisas na área decolaram, e, em 1993, o pesquisador Bart Kosko publicou a obra *Fuzzy Thinking* (Pensamento Nebuloso, em tradução para o português), em que discutia de maneira extensiva e clara os conceitos e aplicações da Lógica Nebulosa (CHRISTA, 2015).

Kay (2004) afirma que, em 1994, a Universidade de Hong Kong realizou uma pesquisa nos dispositivos eletrônicos disponíveis para o público, e o resultado foi um relatório de mais de 100 páginas apontando milhares de dispositivos que usam essa tecnologia, entre eles, máquinas de lavar, micro-ondas, câmeras de vídeo e dezenas de outros tipos de aparelhos de uso diário

Soluções Recentes em Lógica Nebulosa

A Lógica Nebulosa é utilizada em inúmeras soluções que utilizamos ou que nos afetam direta ou indiretamente. DeRoss (2017) aponta várias áreas em que podemos encontrar a Lógica Nebulosa, e, entre elas, podemos citar:

- **Sistemas de Controle**
 - Controle automático de Temperatura em aparelhos de ar-condicionado;
 - Painéis Elétricos Inteligentes, que controlam a temperatura e a pressão internas com base no alimento sendo cozido;
 - Máquinas de Lavar, que identificam o peso da carga e a facilidade de movimentação (um cobertor é menos maleável que uma carga de meias, ainda que ambos possam ter o mesmo peso) e ajustam quantidade de água e velocidade de movimentos para um melhor resultado;
 - TVs inteligentes, que controlam brilho e cores com base nas condições do ambiente.
- **Sistemas de Informações Sociais**
 - Identificação de suspeitos, realizada por programas desenvolvidos para a polícia, que identificam características e dados das ações de suspeitos, atribuindo-lhes maior ou menor probabilidade de terem perpetrado um crime em questão;
 - Marketing, com Donzée e Meier (2012) apresentando técnicas de identificação de possíveis clientes por meio de informações sobre um conjunto específico de pessoas.

A Lógica Nebulosa tem sido cada vez mais aplicada em sistemas de computação, em que as decisões devem ser tomadas com base em informações com diferentes gradações de veracidade. Procure à sua volta e as chances são grandes de você encontrar pelo menos um aparelho em sua residência ou trabalho que usa essa técnica.



Exemplificando

A empresa brasileira i-Systems realiza exclusivamente projetos utilizando a Lógica Nebulosa para melhorar o desempenho de soluções de controle na indústria. A solução de controle de fluxo no envasamento de refrigerantes, por exemplo, gera uma redução significativa no desperdício anual de uma empresa, e o controle de uma caldeira gera grandes economias em uma indústria alimentícia. Tudo com tecnologia nacional, feita por uma empresa criada por egressos da universidade. Esses e outros casos você pode conhecer visitando a página de cases da empresa: <http://www.i.systems.com.br/cases-em-control-e-avancado/>. Acesso em: 27 maio 2018.

Sem medo de errar

Explicar a Lógica Nebulosa para os executivos de uma empresa petroquímica pode não ser a mais fácil das tarefas, mas é plenamente possível usando conceitos de conhecimento comum. O ideal é iniciar a explanação falando sobre a Lógica Proposicional ou Booleana, na qual as proposições são verdadeiras ou falsas. Você pode apresentar alguns exemplos de proposições booleanas:

- A afirmação "Hoje é Segunda-feira" ou é verdadeira, ou é falsa;
- A expressão "Não tenho dívidas financeiras" ou é verdadeira, ou é falsa.

Uma vez explanada a Lógica Proposicional ou Booleana, você pode apresentar alguns exemplos de situações que são "nebulosas" ou "difusas", não sendo fácil saber se são 100% verdadeiras ou 100% falsas:

- Está frio lá fora;
- Fulano é alto;
- Aquela casa é grande.

Essas afirmações podem ter um certo grau de verdade, mas não há uma maneira formal não-arbitrária de definir que estão 100% certas ou 100% erradas. O grau de veracidade varia de um para outro. Você pode mencionar que a Lógica Nebulosa funciona com base em regras que combinam as proposições mais ou menos verdadeiras, e este processo permite decisões mais complexas e mais próximas do processo que nós seres humanos empregamos para decidir

Você pode prosseguir em sua explanação apresentando o fato de a Lógica Nebulosa não ser uma ideia nova, tendo sido criada por Lotfi Sadeh, em 1965. A princípio a técnica foi rejeitada pelos pesquisadores americanos e europeus, mas abraçada pelos asiáticos. Tanto que, em 1987, o primeiro sistema de controle de um trem de metrô foi implantado com sucesso no Japão, o que provocou uma corrida às pesquisas. Nos dias de hoje, inúmeras são as soluções que utilizam a Lógica Nebulosa, várias delas ao alcance do consumidor. Por exemplo:

- Controle automático de Temperatura em aparelhos de ar-condicionado;
- Painéis elétricos inteligentes, que controlam a temperatura e a pressão internas com base no alimento sendo cozido;
- Máquinas de lavar, que identificam o peso da carga e a facilidade de movimentação (um cobertor é menos maleável que uma carga de meias, ainda que ambas possam ter o mesmo peso) e ajustam quantidade de água e velocidade de movimentos para um melhor resultado;
- TVs inteligentes, que controlam brilho e cores com base nas condições do ambiente.

Pronto! Você conseguiu explicar e criar interesse por parte dos executivos da empresa. Prepare-se, pois você será chamado para a próxima fase!

Avançando na prática

Vencendo a concorrência por um emprego

Descrição da situação-problema

Ao terminar seu curso superior, você foi chamado para uma entrevista em uma empresa que desenvolve projetos em Lógica Nebulosa. Depois

de vários testes e entrevistas, a vaga será decidida entre você e um concorrente. Ambos são chamados para uma entrevista final, e nessa entrevista (feita separadamente com cada um) você é questionado sobre como modelaria, usando Lógica Nebulosa, a resposta para a seguinte questão: estou saindo de casa agora, para ir ao trabalho; lá chegando, eu deixo o carro em um estacionamento e devo caminhar 100 metros até chegar ao prédio onde fica minha empresa; não tenho acesso a um programa ou a um site de meteorologia; devo levar guarda-chuva?

Como responder a essa questão por meio do mapeamento do problema em Lógica Nebulosa?

Resolução da situação-problema

Para modelar esse problema usando a Lógica Nebulosa, precisamos primeiramente entender o que está sendo perguntado. Como decidimos se levamos um guarda-chuva ou não? Bem, basicamente decidindo se a afirmação “vai chover” é verdadeira. Como essa não é uma proposição que se preste à Lógica Booleana de maneira útil. Em outras palavras, até é 100% verdade ou 100% falso que vai chover, mas o fato é que só essa informação não ajuda. Se queremos decidir se levamos o guarda-chuva, de maneira mais prática, temos que mapear o tempo lá fora.

Para tanto, podemos atribuir algo como 10% de possibilidade de chuva se está fazendo sol, não há uma única nuvem no céu e é uma época de poucas chuvas, e 90% de possibilidade de chuva se o tempo está carregado, venta muito, está chovendo em outras regiões da cidade e é uma época em que normalmente chove. Entre um estado (10%) e outro (90%), podemos criar outros estados intermediários, usando um critério arbitrário de nosso gosto, desde que seja consistente com a realidade. Digamos que tenhamos definido cinco estados possíveis para o tempo:

- Sol sem nuvens em época de seca – 10%;
- Sol sem nuvens em época de chuva – 20%;
- Sol com nuvens em época de chuva – 50%;
- Nublado em época de chuva – 70%;
- Nublado com chuva em outras regiões, em época de chuva – 90%.

Pronto, o problema está mapeado usando a Lógica Nebulosa!

Faça valer a pena

1. Observe as proposições a seguir e considere-as sob o ponto de vista da Lógica Proposicional (Booleana) e sob o ponto de vista da Lógica Nebulosa (ou Difusa):

- I. Antônio é alto.
- II. Beatriz tem dinheiro.
- III. Carlos tem boa saúde.
- IV. O restaurante é próximo de onde Denise se encontra.

Assinale a alternativa que contém apenas proposições que podem ser consideradas pela Lógica Nebulosa, mas não pela Lógica Proposicional:

- a) I, II e III, apenas.
- b) I, II e IV, apenas.
- c) I, III e IV, apenas.
- d) II, III e IV, apenas.
- e) I, II, III e IV.

2. Observe o conjunto nebuloso da figura a seguir:

Figura | Conjunto Nebuloso (Todos os elementos fazem parte em maior ou menor grau)



Alto

Fonte: adaptada de Shutterstock. Disponível em: <<https://goo.gl/C1jJUX>>. Acesso em: 29 out. 2018.

O indivíduo mais à esquerda é alto em um grau de 0,60 (de 0 a 1) e o indivíduo mais à direita é alto em um grau de 1,00 (100% alto). Em outras palavras, quando utilizamos a Lógica Nebulosa para categorizar os

elementos de um conjunto, temos muito mais informações acerca desses elementos, informações essas que nos auxiliam a tomar decisões melhor embasadas (NGUYEN; WALKER, 2006).

Usando as regras e operações da Lógica Nebulosa, qual o grau de veracidade da seguinte afirmação: "o indivíduo mais à esquerda não é alto".

- a) 0%.
- b) 40%.
- c) 50%.
- d) 60%.
- e) 100%.

3. A Lógica nebulosa não é um conceito novo. A teoria foi introduzida em 1965, pelo pesquisador azerbaijanês radicado nos EUA, Lotfi Zadeh, que, à época, atuava na Universidade da Califórnia, em Berkeley. Zadeh inseriu os conceitos de conjuntos nebulosos e regras nebulosas, estabelecendo as bases matemáticas para o campo, então nascente. A Lógica Nebulosa não foi imediatamente aceita, sendo rejeitada pela comunidade que realizava pesquisas em lógica matemática por conta do nome, que em inglês soa infantil (*Fuzzy Logic*). O conceito foi mais bem aceito na Ásia, tendo várias pesquisas iniciadas neste campo na China e no Japão.

Pelo texto exposto acima, é possível concluir corretamente que:

- a) A teoria da Lógica nebulosa não fez sucesso nos EUA porque foi criada por um pesquisador do Azerbaijão.
- b) Lotfi Zadeh teve que ir para a China para que pudesse realizar suas pesquisas em Lógica Nebulosa.
- c) A Alógica Nebulosa foi desenvolvida em 1965, na China e no Japão.
- d) Apesar de inicialmente renegada nos EUA, a Lógica Nebulosa foi aceita na China e no Japão.
- e) A Lógica Nebulosa nunca fez sucesso.

Seção 3.2

Conjuntos fuzzy

Diálogo aberto

Olá!

Na primeira seção desta unidade, fomos introduzidos aos conceitos da Lógica Nebulosa (ou Difusa) e a como ela se compara com a Lógica Proposicional (ou Booleana). Agora é o momento de nos aprofundarmos nos conceitos dos conjuntos fuzzy e das funções de pertinência, que têm grande importância na modelagem da tomada de decisões, chegando até elas de maneira semelhante a como nós, seres humanos, agimos.

Há uma anedota conhecida na cultura de Minas Gerais em que uma pessoa de fora do estado, seguindo de carro por uma estrada desconhecida, encontra um mineiro sentado em baixo de uma árvore e pergunta a que distância fica uma certa propriedade. O mineiro, cordial, responde que fica “logo ali”. Preocupado com o carro naquela estrada precária, com medo de quebra ou de ficar atolado, o indivíduo questiona se poderia deixar o carro ali estacionado e ir a pé, se não fosse muito longe, ao que o mineiro responde: “Longe nada, é um ‘tirinho de garrucha!’”. O mineiro se referia às antigas armas do século XIX, que faziam mais barulho que estrago, e tinham pouca potência para empurrar uma bala. O indivíduo se anima, estaciona o carro e segue a pé. Ocorre que a distância se mostra muito mais longa do que ele imaginava, e a propriedade só é atingida depois de várias horas de caminhada. No dia seguinte, quando ele retorna, encontra o mineiro sentado à mesma sombra. Antes de entrar no carro para ir embora, o indivíduo se volta para o mineiro e pergunta: “Por quanto o senhor me vende sua garrucha?”.

A anedota é perfeita para ilustrar quanto os conceitos de “perto” e “longe” são variáveis e dependentes da experiência de cada um. O que para o mineiro parecia “perto”, para o indivíduo de fora pareceu uma distância muito longa. Em termos de conjuntos *fuzzy*, a propriedade teria alto grau de pertinência ao conjunto de coisas

que estão “perto” para o mineiro, mas para o indivíduo de fora, essa propriedade teria baixo grau de pertinência a este conjunto. Ao longo dessa seção veremos como funcionam os conjuntos *fuzzy* e as funções de pertinência.

Em suas conversas com uma grande empresa do setor petroquímico, você já conseguiu explicar o que é a Lógica Nebulosa, apresentar exemplos e um histórico do desenvolvimento dessa tecnologia. Agora é a hora de mostrar aos executivos como o processo de produção do composto químico pode ser mapeado em termos da Lógica Nebulosa. Quais são as variáveis a serem consideradas? Quais são os conjuntos *fuzzy* a serem definidos? Quais são as funções de pertinência que regem o comportamento das variáveis dentro dos conjuntos *fuzzy*? Seu trabalho, nessa fase do projeto, será explanar esses conceitos aos executivos da empresa, de maneira que entendam as vantagens da Lógica Nebulosa no processo que é tão importante aos negócios da empresa.

Na presente seção, você vai entender a definição de Conjuntos *Fuzzy*, vai conhecer as características e vários exemplos de Conjuntos *Fuzzy*, vai entender a definição de Funções de Pertinência, e vai ver várias aplicações de Funções de Pertinência. Pronto para mais essa viagem pelo conhecimento?

Então vamos lá!

Não pode faltar

Vamos entender com um pouco mais de detalhes como funcionam os conjuntos *fuzzy*, também conhecidos como “conjuntos nebulosos” ou “conjuntos difusos”?

Definição de Conjuntos *Fuzzy*

Nguyen e Walker (2006) definem conjuntos *fuzzy* como objetos matemáticos cujo objetivo é modelar a imprecisão, que é característica da linguagem natural quando se descrevem fenômenos que não têm limites bem definidos. Os termos chave dessa definição são:

- **Objeto matemático** – Os conjuntos *fuzzy* são, de fato, conjuntos, isto é, elementos matemáticos que representam

coleções, e são formalmente definidos, permitindo que relações sejam estabelecidas a partir de sua composição e características.

- **Imprecisão** – As relações da Lógica Proposicional lidam apenas com proposições que são 100% verdadeira ou 100% falsas, o que permite precisão total, no universo de proposições possíveis de serem tratadas, mas exclui a imprecisão. Ocorre que nós, seres humanos, temos percepções imprecisas e tomamos decisões com base nessas percepções, muitas vezes sendo mais aptos a lidar com o mundo à nossa volta do que os computadores. Os conjuntos *fuzzy* admitem que uma proposição seja parcialmente verdadeira. Ou seja, permitem que elementos pertençam parcialmente a eles, de maneira que possamos tratar proposições como parcialmente verdadeiras ou falsas.
- **Característica da Linguagem Natural** – Os conjuntos *fuzzy* são especialmente aptos a lidar com descrições em linguagem natural, ou seja, permitem que nós descrevamos problemas sem necessidade de grande conhecimento de linguagens científicas (linguagem formal) ou de linguagens de programação.

Um exemplo simples dos conjuntos fuzzy é a anedota do “tirinho de garrucha”, em que os conceitos de “perto” e “longe” são imprecisos e variam em cada um de nós, com base em nossa realidade e em nossa experiência de vida. A distância da árvore onde o mineiro se encontra até a propriedade que o indivíduo de fora procura não pertenceria ao conjunto “distâncias classificadas como perto”, se esse conjunto estabelecesse um intervalo preciso, definido, formal. Digamos “entre 0 m e 500 m”. Nessa definição formal, ou os elementos pertencem 100% ao conjunto (estão localizados dentro de um raio de 500 m da árvore) ou não pertencem ao conjunto (estão localizados além do raio de 500 m).



Assimile

Conjuntos nebulosos são objetos matemáticos cujo objetivo é modelar a imprecisão que é característica da linguagem natural, quando se descrevem fenômenos que não têm limites bem definidos.

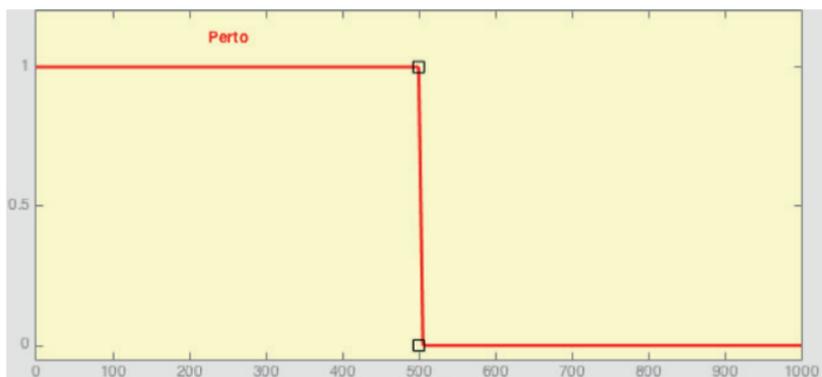
Matematicamente, temos:

$$P(x) = \begin{cases} 1 & \Leftarrow 0 \leq x \leq 500 \\ 0 & \Leftarrow x > 500 \end{cases}$$

Ou seja, um elemento localizado na posição "x" só é considerado "perto" se estiver a uma distância máxima de 500 m. Caso contrário (se estiver a 501, por exemplo, já será considerado "longe". É importante observar que esta definição rígida causa uma situação absurda, pois um objeto colocado no limite de 500 m, é atingível por, digamos, um robô que se movimenta dentro do universo dos locais "perto", mas não é atingível por esse robô se estiver a uma distância de 500,01 m (quinhentos metros e um centímetro).

O Gráfico 3.2 representa as relações dos vários elementos com este conjunto:

Gráfico 3.2 | Comportamento da curva P(x)



Fonte: elaborado pelo autor.

Já se considerarmos o conjunto *fuzzy* para definir "perto", podemos dizer que qualquer coisa localizada a 100 m de distância ou menos, é 100% perto; qualquer coisa localizada entre 100 m e 500 m vai ficando gradativamente menos perto (pertence cada vez menos ao conjunto das coisas "perto"); qualquer coisa que esteja a mais de 500 m de distância, já não está mais "perto".

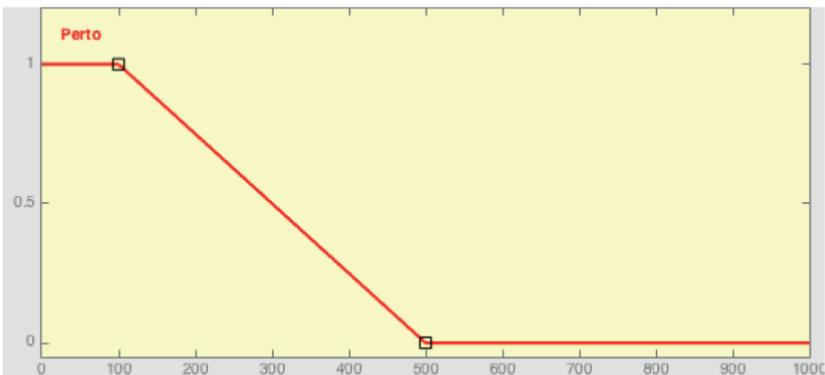
Matematicamente, temos

$$P(x) = \begin{cases} 1 & \leftarrow x < 100 \\ 500 - x & \leftarrow 100 \leq x < 500 \\ 400 & \\ 0 & \leftarrow x \geq 500 \end{cases}$$

É importante observar que o mesmo robô que consegue atingir coisas consideradas “perto”, de acordo com a definição matemática acima, chega sem dificuldade nenhuma a qualquer objeto que esteja a até 100 m de distância do ponto central. A partir daí, ele vai tendo cada vez mais dificuldade em chegar aos locais, quanto mais afastados estejam, entre 100 m de distância e 500 m de distância. O máximo grau de distância que esse robô consegue atingir, com o máximo esforço que consegue empreender, é de 500 m. Esse comportamento se aproxima do comportamento humano, não é verdade? Se dissermos a uma pessoa saudável que precisa caminhar 100 metros, o equivalente a uma quadra, para, digamos, obter um pequeno prêmio (um copo de sua bebida predileta, por exemplo), ela muito provavelmente não vai pensar duas vezes e não acreditará estar fazendo esforço apreciável. A partir de 100 m, ela vai começar a achar que faz esforço, primeiramente pequeno, e que vai aumentando à medida que a distância aumenta. Essa pessoa — na média, digamos — estará disposta a caminhar até 500 m para obter seu prêmio, mas a cada metro a mais que tenha que andar, vai se sentindo menos estimulada a empreender a “jornada”. Se lhe dissermos que o prêmio está a mais de 500 m, ela provavelmente declinará da oportunidade.

O Gráfico 3.3, a seguir, representa esta situação:

Gráfico 3.3 | Comportamento da curva P(x) em um conjunto *fuzzy*



Fonte: elaborado pelo autor.



Suponhamos que um profissional especialista em lógica *fuzzy* esteja empregado, ganhando um salário de R\$ 5.000,00 mensais. Ao receber uma oferta para mudar de emprego, ele pensa que está bem empregado, mas até poderia trocar de emprego se a oferta fosse boa. Ele não sairia do emprego atual por menos do que ganha hoje, mas talvez saísse para ganhar mais, dependendo das condições. Certeza de que sairia, mesmo, só se a oferta fosse um salário superior a R\$8000,00 por mês. Ao modelar esse pensamento do profissional, temos:

$$P(x) = \begin{cases} 0 & \leftarrow x \leq 5000 \\ \frac{x - 5000}{3000} & \leftarrow 5000 < x < 8000 \\ 1 & \leftarrow x \geq 8000 \end{cases}$$

Com esse pequeno modelo, esse profissional pode extrapolar outras condições e definir um sistema de lógica nebulosa para auxiliar na tomada da decisão.

Características e Exemplos de Conjuntos *Fuzzy*

Como podemos perceber no que foi exposto anteriormente, a característica principal dos conjuntos *fuzzy* é que os elementos desses conjuntos pertencem a eles em diferentes graus, de 0 a 100%. Segundo Silva et al (2007), a teoria dos conjuntos *fuzzy* é uma generalização da Teoria Geral dos Conjuntos. Essa generalização se dá por que, na Teoria Geral dos Conjuntos, um elemento qualquer pertence totalmente ou não pertence totalmente a um conjunto, o que é um caso particular de uma relação de pertencimento que permite graus intermediários.

Como veremos mais à frente, os conjuntos *fuzzy* – a exemplo de qualquer conjunto formal da Matemática – têm seus elementos definidos por uma função de pertinência, que é a relação que define o grau em que um conjunto pertence a um conjunto difuso. Segundo Nguyen e Walker (2006), um elemento de um conjunto *fuzzy* pode ser definido, para fins de processamento, como um par (u,g) onde:

u – valor/descritor do elemento;

g – grau de pertinência de u ao conjunto *fuzzy*.

Definição de Funções de Pertinência

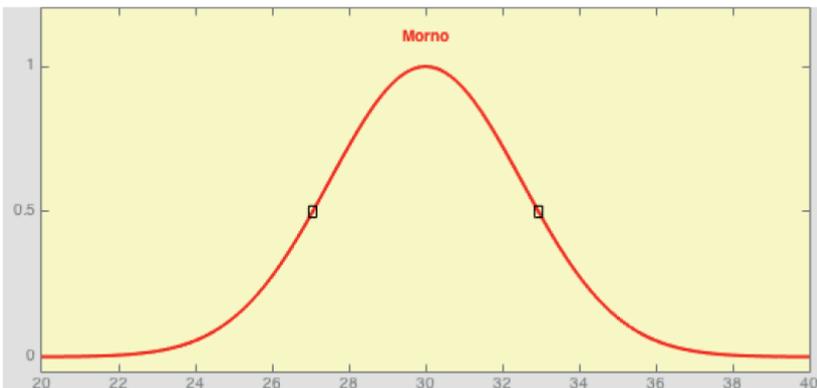
Uma função de pertinência define quanto, isto é, em que grau, um elemento pertence a um conjunto *fuzzy*. Segundo Nguyen e Walker (2006), um conjunto *fuzzy* pode ser definido matematicamente como um par (U, m) , em que U é o conjunto e m é a relação de pertinência dos elementos ao conjunto U . Nesse contexto, podemos então definir m , isto é, a função de pertinência, em linguagem matemática:

$$m : U \rightarrow [0,1]$$

As funções $P(x)$ mencionadas no exemplo de conjuntos acima são funções de pertinência. A expressão afirma que a função de pertinência m define quais elementos do conjunto difuso U podem pertencer a esse conjunto em qualquer grau representável por um valor pertencente a \mathbb{R} entre "0 e "1".

Uma função de pertinência é, portanto, uma relação que informa quanto um elemento, a depender de seu valor, pertence a um conjunto *fuzzy*. Vamos tomar um exemplo para fixar esse conceito. Imaginemos o conceito de "morno" relacionado à temperatura da água, quando vamos tomar banho. Podemos imaginar que a temperatura da água inferior à temperatura da pele (22°) é considerada "fria". À medida que elevamos a temperatura, ela vai suavemente sendo percebida como menos fria e mais morna. Essa ascensão dura por um intervalo, digamos, até os 30°, e em seguida passa, também suavemente, a ser percebida como menos morna e mais quente, até chegar aos 38°, quando passa a ser percebida como quente. O Gráfico 3.4 representa essa relação de pertinência:

Gráfico 3.4 | Função de Pertinência para o conjunto fuzzy Morno (temperatura °C)

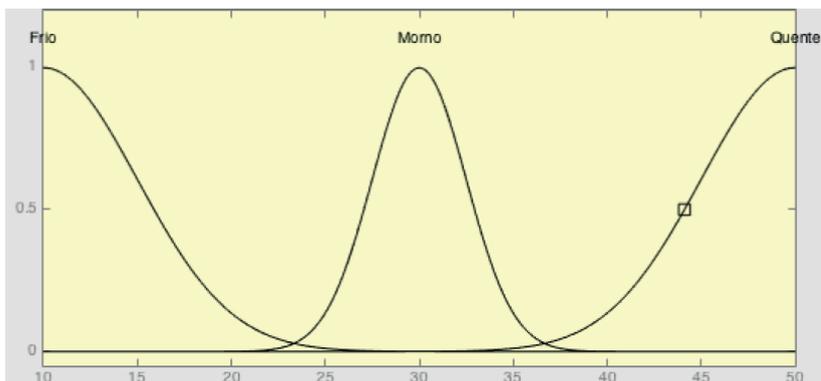


Fonte: elaborado pelo autor.

O Gráfico 3.4 simboliza que uma variável de temperatura começa a pertencer, ainda que parcialmente, ao conjunto *fuzzy* "Morno" quando atinge 22°C. Seu grau de pertinência sobe suavemente, até atingir 30°C, única temperatura em que a variável pertence 100% ao conjunto, e em seguida começa a descer gradualmente, saindo do conjunto "Morno", até atingir 38°C. Acima de 38°C, a temperatura já não pertence mais ao conjunto "Morno".

Este processo começa a ficar mais interessante, como afirmam Sivanandam, Sumathi e Deepa (2007), quando entendemos que podemos combinar mais de uma função de pertinência para a mesma variável. No exemplo acima podemos perguntar: como podemos descrever a temperatura que está abaixo de 22°C? E se essa temperatura estiver acima de 38°C? Bem, para englobarmos essas duas possibilidades, podemos incluir duas funções de pertinência: "Frio" e "Quente". Por exemplo, podemos estipular que qualquer temperatura abaixo de 10°C é totalmente fria, e passando de 15°C até 25°C, ela vai sendo menos fria, perdendo totalmente essa característica acima de 25°C. Do outro lado da escala, a partir de 35°C a temperatura começa a tomar a característica de "Quente", e vai atingindo totalmente essa característica aos 50°C, que manterá daí para frente. O Gráfico 3.5 ilustra essa situação:

Gráfico 3.5 | Funções de Pertinência para os conjuntos Frio, Morno e Quente (°C)



Fonte: elaborado pelo autor.

Quando comparamos os Gráficos 3.2 e 3.3 com os Gráficos 3.4 e 3.5 desta seção, uma diferença salta aos olhos. Você consegue perceber o que é? Ocorre que os Gráficos 3.2 e 3.3 apresentam transições bruscas

de um intervalo para o outro, enquanto os Gráficos 3.4 e 3.5 apresentam transições suaves. Isso por que, nos dois primeiros gráficos, temos funções lineares transitando de uma para outra, enquanto nos dois últimos exemplos, essas funções não são lineares.

Em linhas gerais, as funções de pertinência podem ser quaisquer funções matemáticas e, na prática, apenas algumas poucas já são suficientes para representarmos a vasta maioria das relações que encontraremos quando definindo relações úteis para a resolução de problemas em Lógica Nebulosa (SIVANANDAM; SUMATHI; DEEPA, 2007). No MATLAB, essas funções são:

- **Função Triangular (trimf)** – Função linear que define 3 pontos, produzindo um “pico”. Três pontos definem essa curva: o “0” inicial, o “1” e o “0” final.
- **Função Trapezoide (trapmf)** – Função linear que define 4 pontos, produzindo um “patamar”. Quatro pontos definem esta curva: o “0” inicial, o “1” inicial, o “1” final e o “0” final.
- **Função Sino Genérica (gbellmf)** – Curva do “sino” genérica. Três pontos definem essa curva, que é simétrica: o “0” inicial, o “1” inicial e o “centro do platô”, definindo a porção esquerda da curva. O restante será deduzido a partir desses pontos.
- **Função Gaussiana (gaussmf)** – Curva gaussiana, um caso especial da curva do sino, simétrica. Dois pontos definem essa curva: o ponto em que a curva atinge “1” e o fator σ , que define a “abertura” da curva.
- **Função Gaussiana Combinada (gauss2mf)** – Curva resultante da combinação de duas gaussianas, permitindo inclusive a presença de um platô superior. Definida por 3 pontos: o ponto “1” e o σ da metade esquerda, e o ponto “1” e o σ da metade direita.
- **Função Sigmoide (sigmf)** – Caso especial da curva “S”. Definida por 2 pontos: a taxa de ascensão e o ponto de inflexão (centro da ascendente).
- **Função Sigmoide de Diferença** – Combinação de 2 curvas “S”, uma ascendente e outra descendente. Aparenta a função gaussiana combinada. Definida por 4 pontos: taxa de ascensão e ponto de inflexão da primeira curva, e taxa de descenso e ponto de inflexão da segunda curva.

- **Função Sigmoide de Produto (psigmf)** – Também é a combinação de duas sigmoidais, também definida pelos mesmos quatro pontos.
- **Função Pi (pimf)** – Função que lembra o formato de Π (Pi maiúsculo), assimétrica. Definida por 4 pontos: "0" inicial, "1" inicial, "0" final e "1" final.
- **Curva "S" Genérica (smf)** – Curva "S" Genérica. Definida por 2 parâmetros: ponto "0" e ponto "1".
- **Curva "Z" Genérica (zmf)** – Curva "Z" Genérica (descenso da curva "S"). Definida por 2 parâmetros: ponto "0" e ponto "1".

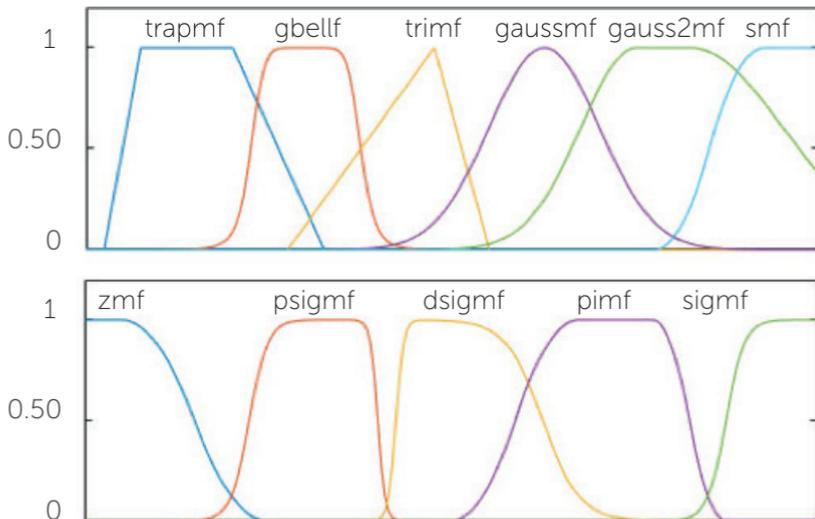


Refleta

Que tipo de situação demandaria uma transição mais suave, como no caso das gaussianas ou das sigmoides? Quais tipos de situação demandariam transições mais bruscas, como no caso das triangulares ou trapezoides?

O Gráfico 3.6 representa as curvas de todas as funções disponíveis no MATLAB:

Gráfico 3.6 | Funções de Pertinência para os conjuntos Frio, Morno e Quente (°C)



Fonte: MathWorks (2018, p. 1).

Aplicação de Funções de Pertinência

Em Teoria dos Conjuntos, as Funções de Pertinência são usadas para determinar se um determinado candidato é, de fato, membro de um dado conjunto. Já no caso da Lógica Nebulosa, as Funções de Pertinência são usadas para determinar o grau de pertinência de um candidato qualquer (NGUYEN; WALKER, 2006).

Ocorre que, como afirmam Russell e Norvig (2013), as funções de pertinência são o alicerce da Lógica Nebulosa, permitindo, por meio de suas inúmeras combinações, a formação de regras de decisão que definem com facilidade o estado de um sistema qualquer. Dentro do MATLAB, as funções de pertinência mapeiam o comportamento possível das variáveis de entrada, e esse comportamento é a base para a formação de regras que definem as saídas do sistema.



Pesquise mais

Silva et al. (2007) publicaram um excelente artigo, mostrando como os conjuntos difusos, em conjunto com regras de combinação e com o uso de técnicas da Lógica Nebulosa, auxiliam no prognóstico de chuvas no nordeste brasileiro. Confira no artigo a seguir:

SILVA, E. M. et al. Uma aplicação de conjuntos difusos na otimização do prognóstico de consenso sazonal de chuva no nordeste do Brasil.

Revista Brasileira de Meteorologia, v. 22, n.1, p. 83-93, 2007. Disponível em: <<http://www.scielo.br/pdf/rbmet/v22n1/a09v22n1.pdf>>. Acesso em: 2 jun. 2018.

Sem medo de errar

Há muitas maneiras de modelarmos as variáveis do processo de criação do composto para a indústria petroquímica. Vejamos abaixo uma das soluções possíveis.

Temos três variáveis a serem controladas no processo de fabricação do composto químico:

- Temperatura (graus Celsius);
- Pressão (Bar);
- Vazão (litros/segundo).

Para modelarmos essas variáveis, devemos assumir que para as três há um intervalo ideal, que deve sempre ser buscado pelo sistema. Nenhuma das três variáveis pode fugir de tal intervalo. Nessa situação, podemos definir dois outros intervalos importantes:

- **Insuficiente** – a variável em questão (Temperatura, Pressão, Vazão) está abaixo do ideal
- **Excessivo** – a variável em questão (Temperatura, Pressão, Vazão) está acima do ideal

Podemos supor, para fins desse exercício, que, no processo em questão, temos os seguintes valores:

- **Temperatura**

- O desejável é entre 100°C e 300°C, sendo 200°C o ponto considerado ideal
- Abaixo de 150°C já se pode considerar que a temperatura se aproxima de estar fria demais, e se chegar abaixo de 100°C está inaceitavelmente fria.
- Acima de 250°C já se pode considerar que a temperatura se aproxima de estar quente demais, e se ultrapassar 300°C está inaceitavelmente quente.

- **Pressão**

- O desejável é entre 5 e 7 Bar, sendo 6 Bar o ponto considerado ideal
- Abaixo de 5,5 Bar já se pode considerar que a pressão se aproxima de estar insuficiente, e se chegar abaixo de 5 Bar está insuficiente e não gera a reação esperada.
- Acima de 6,5Bar já se pode considerar que a pressão se aproxima de estar alta demais, e se ultrapassar 7 Bar está excessiva e não gera a reação esperada.

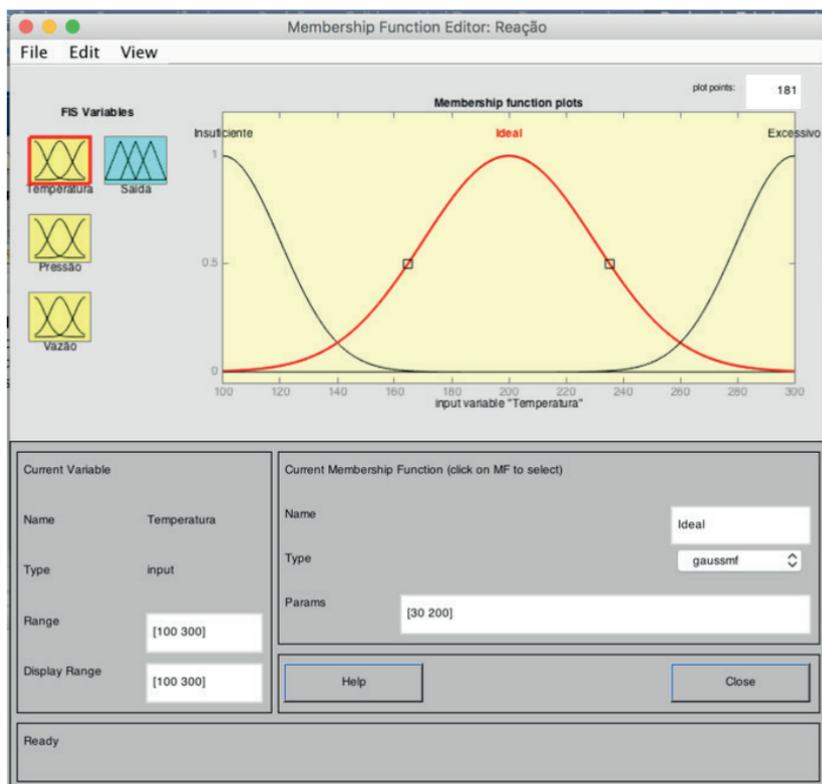
- **Vazão**

- O desejável é entre 4 e 6 litros por segundo, sendo 5 l/s a vazão ideal
- Abaixo de 4,5 l/s já se pode considerar que a vazão se aproxima de ser insuficiente, e se chegar abaixo de 4l/s está insuficiente e não gera a reação esperada.

- o Acima de 5,5 l/s já se pode considerar que a vazão se aproxima de estar alta demais, e se ultrapassar 6l/s está excessiva e não gera a reação esperada.

Com base nessas condições, podemos deduzir que os conjuntos *fuzzy* são: "Insuficiente", "Ideal" e "Excessivo", para cada uma das três variáveis. Podemos usar o MATLAB para modelar essas variáveis, o que nos dá uma maneira simples de apresentar aos executivos da empresa uma forma automática de controlar as variáveis. A Figura 3.3 apresenta o mapeamento da variável "Temperatura":

Figura 3.3 | Três Funções de Pertinência para a variável "Temperatura"



Fonte: captura de tela do MATLAB R2014b, elaborada pelo autor.

A escolha da gaussiana para o mapeamento da variável "Temperatura" se dá pelo fato de o processo físico de aumento/redução de temperatura não seguir um padrão linear, sendo mais suave a transição entre um estado e outro.

A partir desse exemplo, você pode igualmente mapear as variáveis “Pressão” e “Vazão”, e mostrar que todas são mapeáveis em um sistema de Lógica Nebulosa, o que vai permitir seu controle de maneira automatizada.

Pronto! Você cumpriu mais uma etapa do processo de convencimento e está a um passo de conseguir um importante contrato para sua empresa e para sua carreira.

Avançando na prática

O revolucionário aparelho cardíaco

Descrição da situação-problema

Um dos maiores e mais bem-sucedidos cirurgiões cardíacos do país chegou à conclusão de que os marca-passos atuais precisam de melhoria. Depois de anos de pesquisa, ele conseguiu identificar com precisão a intensidade dos sinais elétricos que chegam ao coração e que determinam o ritmo cardíaco de uma pessoa. De posse desse conhecimento, este médico construiu uma tabela, com base em resultados de exames e mensurações de milhares de pacientes, que identifica o ritmo cardíaco do paciente. Com uma sinalização em um intervalo de 0 a 100, ele identificou os seguintes pontos:

- Sinal Nulo (sem efeito no coração) – de 0 a 10;
- Sinal Fraco (com pouco efeito no coração, aumentando linearmente à medida que a curva cai a 0) – de 10 a 30;
- Sinal Médio (com efeito mediano sobre os batimentos cardíacos) – de 30 a 40 (aumento linear), de 40 a 60 (platô) e de 60 a 70 (redução linear);
- Sinal forte (de 70 a 90, aumentando linearmente até atingir o nível excessivo);
- Sinal excessivo – platô de 90 a 100.

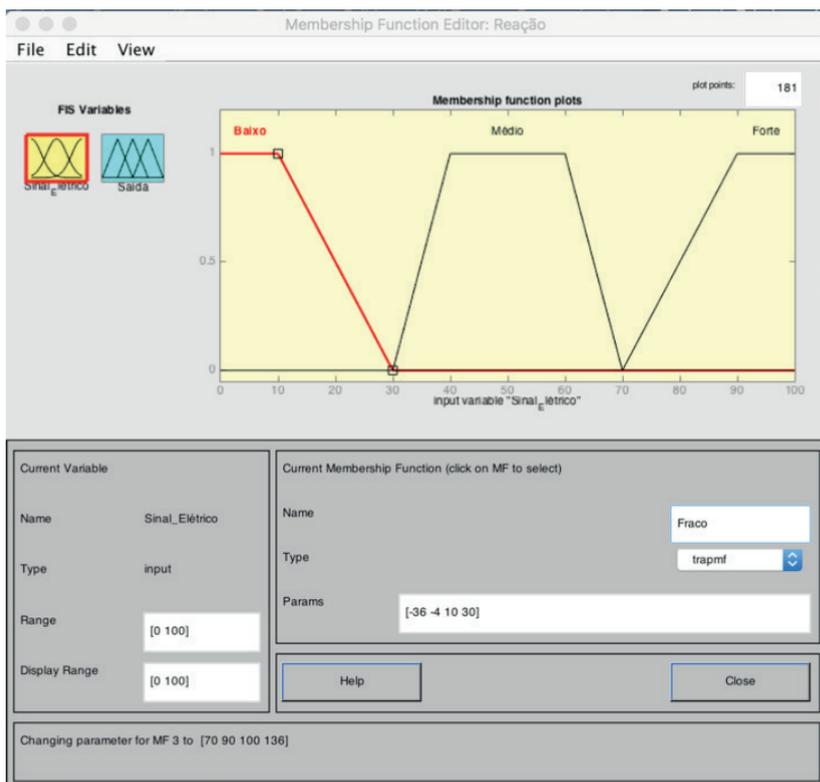
Esse cirurgião gostaria de saber se é possível traduzir esse conhecimento em um conjunto de funções de pertinência que possa, posteriormente, ser passado para um sistema de Lógica Nebulosa. Será que isso é possível? Se sim, como você,

consultor profissional em Lógica Nebulosa pode fazer para ajudar esse cirurgião?

Resolução da situação-problema

A resposta é um sonoro “SIM!”, pois há como traduzir essa tabela em um conjunto de funções de pertinência. Usando o MATLAB, por exemplo, podemos criar uma variável de entrada “Sinal Elétrico” e mapeá-la com três funções trapezoides, que permitem a criação de platôs e variações lineares, refletindo os valores da tabela que nos foi apresentada pelo cirurgião cardíaco. O resultado é expressado da seguinte forma, usando o MATLAB:

Figura 3.4 | Três Funções de Pertinência para a variável “Temperatura”



Fonte: captura de tela do MATLAB R2014b, elaborada pelo autor.

Faça valer a pena

1. Nós, seres humanos, temos percepções imprecisas e tomamos decisões com base nessas percepções, muitas vezes sendo mais aptos a lidar com o mundo à nossa volta do que os computadores. Os conjuntos *fuzzy* admitem que uma proposição seja parcialmente verdadeira. Ou seja, os conjuntos *fuzzy* permitem que elementos pertençam parcialmente a eles, de maneira que possamos tratar proposições como parcialmente verdadeiras ou falsas.

Assinale a alternativa que melhor sintetiza o elemento dos conjuntos *fuzzy* tratado no texto acima:

- a) Lógica Booleana.
- b) Precisão.
- c) Linguagem natural.
- d) Imprecisão.
- e) Pertinência.

2. Em linhas gerais, nossa percepção acústica é de que um ambiente é quieto (zero barulho) se a taxa de som vai de 0 até 8 decibéis. A partir daí, passamos a perceber o som, de maneira linearmente crescente, até cerca de 110 decibéis, que é considerado pelo ouvido humano um som bastante alto (erupção vulcânica, lançamento de um foguete, ou o som de um concerto de rock dos mais altos já registrados). Qualquer som a partir de 110 decibéis é considerado igualmente insuportável pela maioria das pessoas.

Assinale a alternativa que descreve matematicamente as relações descritas no texto acima, assumindo que o volume de um som é descrito pela variável "x" e que sua percepção é dada pela função P(x):

a) $P(x) = 8 \leq x \leq 110$

b) $P(x) = \begin{cases} 0 \Leftarrow x < 8 \\ 0,5 \Leftarrow 8 \leq x \leq 110 \\ 1 \Leftarrow x < 110 \end{cases}$

c) $P(x) = 110 - 8$

d) $P(x) = \begin{cases} 1 \Leftarrow x < 8 \\ 0,5 \Leftarrow x < 110 \\ 1 \Leftarrow x \geq 110 \end{cases}$

$$d) P(x) = \begin{cases} 0 \leftarrow x \leq 8 \\ \frac{x - 8}{102} \leftarrow 8 < x < 110 \\ 1 \leftarrow x \geq 110 \end{cases}$$

3. O software MATLAB permite a criação de diversos tipos diferentes de funções de pertinência (onze, no total). Uma dessas funções tem como definição: função linear, que define 4 pontos, produzindo um “patamar”. Quatro pontos definem essa curva: o “0” inicial, o “1” inicial, o “1”, final e o “0” final.

Assinale a alternativa que contém o nome da curva, no MATLAB, a que o texto se refere:

- a) trimf.
- b) trapmf.
- c) gaussmf.
- d) sigmf.
- e) gbellmf.

Seção 3.3

Elementos de lógica nebulosa

Diálogo aberto

Olá!

Caminhamos bastante em nossos estudos sobre Lógica Nebulosa, não é verdade? Até aqui, vimos como podemos usar os conceitos de variáveis difusas, mapeando conceitos da linguagem natural, que normalmente são de difícil quantificação, para montar um sistema que nos permita a tomada de decisão. Agora é a hora de observarmos esse processo mais de perto e, assim, conseguirmos formalizar suas várias partes. Essa compreensão nos permitirá um maior domínio sobre os mecanismos e conceitos da Lógica Nebulosa, criando, inclusive, um caminho que podemos seguir quando diante de um problema que pode ser resolvido pela Lógica Nebulosa. O objetivo maior desta porção de nosso estudo é compreender formalmente os processos da Lógica Nebulosa.

Em suas interações com a empresa petroquímica, que tem um importante processo de criação de um composto químico, você já passou pelas fases de explanar aos executivos o que é a Lógica Nebulosa e seus conceitos principais. Também mapeou as variáveis do sistema e as definiu em termos da Lógica Nebulosa. Agora é a hora de formalizar o processo, mostrando como funciona esse mapeamento, o que deve tirar as últimas dúvidas dos executivos quanto à efetividade da Lógica Nebulosa. Para tanto, você deve tomar todos os passos desenvolvidos no exercício anterior e identificar o que do processo todo corresponde à fuzzificação, o que corresponde à inferência e o que corresponde à defuzzificação. Qual a utilidade dessa formalização? Como fazer essa identificação? Como mostrar esses mecanismos de forma simples e compreensível para os executivos da empresa?

Na presente seção, você vai entender, a partir de um exemplo de modelagem de um problema em Lógica Nebulosa, o que é a fuzzificação das variáveis. Em seguida, vai entender, com as variáveis fuzzificadas, o que é e como se dá o mecanismo de inferência. Por

fim, você vai entender como se fecha o ciclo da Lógica Nebulosa, com a defuzzificação das variáveis. Pronto para mais essa etapa?

Então vamos lá!

Não pode faltar

Já vimos alguns problemas sendo mapeados usando as técnicas da Lógica Nebulosa. Agora é o momento de entendermos mais detalhadamente como funciona cada um dos mecanismos de mapeamento desses problemas, saindo de uma situação qualquer, que a Lógica Proposicional teria dificuldades de mapear (ou geraria um modelo complexo demais), e chegando até um modelo em Lógica Nebulosa, que resolve o problema de maneira mais simples e elegante. Esse detalhamento é o que veremos a seguir.

Fuzzificação

Seising (2007) define fuzzificação como o processo pelo qual uma variável que pode ser numericamente definida deixa de ser representada por valores numéricos que possa assumir e passa a ser representada por um conjunto de funções de pertinência, que correspondem a definições difusas em linguagem natural. A fuzzificação é o primeiro passo no processo de Inferência Nebulosa (ou difusa).

Supondo uma variável financeira representando um preço qualquer, podemos realizar a fuzzificação dessa variável (preço da *commodity* x) com uma pergunta simples: o preço de x , averiguado em determinado momento, é caro, barato, ou está na média? A fuzzificação é o processo de definir numericamente os intervalos "caro", "barato" e "na média".

Como já vimos, os conceitos em linguagem formal adotados para descrever de maneira difusa a variável em questão são as funções de pertinência. Portanto, a segunda etapa do processo de fuzzificação, segundo Seising (2007), é a quantificação dessas funções de pertinência. Em suma, a segunda etapa do processo de fuzzificação é o processo que determina em que grau (ou, mais formalmente, em que porcentagem), um valor qualquer da variável em questão pertence a alguma das funções de pertinência.



De maneira resumida, fuzzificação é o processo de descrever uma variável quantificável de maneira que sua variação possa ser descrita usando atributos de linguagem natural.

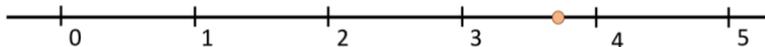
Inferência em sistemas nebulosos

Após a fuzzificação, que vimos anteriormente, é hora de tratarmos o processo de inferência em sistemas de lógica nebulosa. Segundo Nguyen e Walker (2006), podemos definir inferência como uma regra no campo do raciocínio aproximado que permite conclusões mais ou menos vagas, a partir de premissas mais ou menos vagas. Quanto mais vagas as premissas, mais vagas serão as conclusões. Simetricamente, quanto menos vagas as premissas, menos vagas serão as conclusões.



Observe a Figura 3.5, a seguir. É possível afirmar que o ponto entre os números 3 e 4 é uma fração qualquer, e que a representação desse número com uma casa decimal seria $3,X$, com "X" sendo um algarismo de 1 a 9. Você consegue, usando apenas seus olhos, inferir que número é esse? É possível ter certeza, usando apenas os olhos, sem recorrer a um instrumento de medição? Se o número for da forma $3,XY$ (ou seja, se tiver duas casas decimais), é possível identificar com alguma certeza o algarismo "Y"? Por quê?

Figura 3.5 | Identificando uma casa depois da vírgula



Fonte: elaborada pelo autor.

Em sistemas baseados em Lógica Nebulosa, o objetivo para a fuzzificação de variáveis é possibilitar a inferência e, por meio desta (como veremos no próximo item dessa seção), a tomada de decisões. A inferência, no tocante à Lógica Nebulosa, toma como insumo uma ou mais regras de inferência. Essas regras, segundo Nguyen e Walker (2007), tomam, na prática, a forma de Se-Então. Podemos representá-las da seguinte forma:

SE [condição verdadeira] **ENTÃO** [ação]

A condição a que a regra se refere pode ser única ou composta por meio de operadores lógicos (OU, E), mas em todos os casos as cláusulas devem apresentar resultados da Lógica Proposicional Booleana, ou seja, os valores podem ser apenas VERDADEIRO ou FALSO (NGUYEN; WALKER, 2006). Dois exemplos:

SE [vai sair agora **E** está chovendo] **ENTÃO** [levar guarda-chuva]

SE [está trafegando à noite **OU** está trafegando em uma estrada] **ENTÃO** [acender o farol]

Defuzzificação

O passo final da resolução de problemas por meio da Lógica Nebulosa é o processo de defuzzificação. Segundo Nguyen e Walker (2006), defuzzificação é o processo de se produzir resultados quantificados, a partir das regras criadas para as funções de pertinência de um modelo nebuloso. Em outras palavras, é a tradução em um valor real da composição dos graus de pertinência, definidos pela aplicação das funções de pertinência em um dado valor de entrada, como já vimos anteriormente. Esses três conceitos, como você pode perceber, não são nada além da formalização do mapeamento de um problema em Lógica Nebulosa, como visto nas seções anteriores.

Controladores baseados em lógica nebulosa

Vimos até aqui todos os mecanismos necessários para criarmos sistemas baseados em Lógica Nebulosa. Mas, para que são utilizados esses sistemas? As variáveis de saída, que são ajustadas para valores específicos com base nas condições de entrada, são bastante adequadas para controlar o mecanismo a que se referem. A Lógica Nebulosa vem sendo cada vez mais utilizada para criar sistemas de controle precisos, baratos e fáceis de desenvolver e operar, economizando incontáveis horas de programação, que, de outra forma, teriam de ser empreendidas para resolver os problemas que uma boa modelagem nebulosa resolve com mais elegância e facilidade.

Oliveira Junior et al. (2007) apresentam uma forma interessante de representar os controladores baseados em Lógica Nebulosa por meio do MATLAB, uma ferramenta bastante apropriada para a construção de sistemas matemáticos e simulações de conceitos computacionais.

Podemos, com base na mecânica adotada por esses autores, exemplificar os controladores baseados em lógica nebulosa. O controlador que analisaremos, no caso, é um sistema de previsão de venda de cerveja baseado em três variáveis de entrada, a saber:

- Dia da semana (desconsiderados os feriados);
- Temperatura ambiente;
- Preço da cerveja.

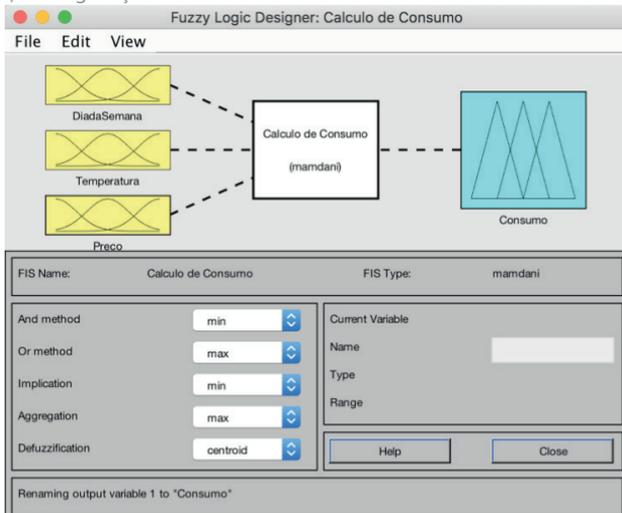
Para este sistema, nos interessa calcular uma variável de saída:

- Consumo de cerveja.

O desenvolvimento desse sistema baseado em Lógica Nebulosa começa com a modelagem dessas três variáveis de entrada e da variável de saída no MATLAB, processo que deve começar com a inicialização do ambiente de lógica nebulosa naquela ferramenta (por meio do comando "fuzzy" na Janela de Comandos).

Em seguida, na janela principal da ferramenta de Lógica Nebulosa, devemos definir os nomes das três variáveis de entrada e da variável de saída do sistema. Podemos, inclusive, aproveitar e dar nome ao sistema, no caso, "Calculo de Consumo" (sem acento, uma vez que o MATLAB não lida muito bem com nomes em língua portuguesa). A Figura 3.6, a seguir, representa esse sistema, em seu passo inicial (definição e nomes das variáveis):

Figura 3.6 | Configuração inicial do sistema de Cálculo de Consumo de Cerveja



Fonte: captura de tela do MATLAB 2017a, elaborada pelo autor.

Em seguida, devemos definir as funções de pertinência para as variáveis. Para tanto, devemos estabelecer intervalos em que cada variável terá valores significativos para o sistema. Podemos definir os intervalos:

- **DiadaSemana:** variando os dias de 1 até 7, considerando (para fins de cálculo, apenas) "1" como a segunda-feira e "7" como o domingo.
- **Temperatura:** variando a temperatura de acordo as temperaturas que propiciam o consumo de cerveja, entre 20°C e 40°C (uma vez que temperaturas acima de 40°C são raras).
- **Preço:** considerando preços de R\$2,00 até R\$20,00.

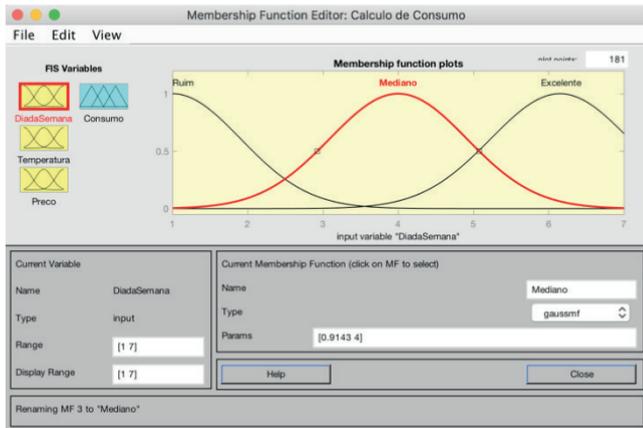
No que concerne a variável de saída, podemos considerar:

- **Consumo:** pensando em litros por pessoa, em média, podemos considerar de 0 a 5 litros.

É importante observar que todas as variáveis acima foram definidas de maneira arbitrária. Em um sistema profissional, os intervalos devem ser definidos com base em fatos e dados coletados de fonte fidedigna. Quanto mais precisos os intervalos, melhor o reflexo da realidade que se deseja mapear pelos resultados do sistema. A próxima etapa é a fuzzificação das variáveis, definindo suas funções de pertinência.

Para a variável Dia da Semana, podemos postular que segunda-feira é o pior dia — um dia péssimo, aliás — e que, à medida que nos aproximando da quarta-feira, essa qualidade de "Péssimo" vai caindo. Também com relação às quartas-feiras, é o dia mais "Mediano" para o consumo, um comportamento que cresce até chegarmos nesse dia, e que decresce nos dias seguintes. A partir de quinta-feira, começamos a fase "Excelente" da semana para o consumo da cerveja, período maximizado no sábado. Para todas essas variáveis, vamos assumir uma ascensão e descenso suaves, não-lineares, e, para tanto, adotaremos uma curva gaussiana no MATLAB. A Figura 3.7 mostra as funções de pertinência desta variável no MATLAB:

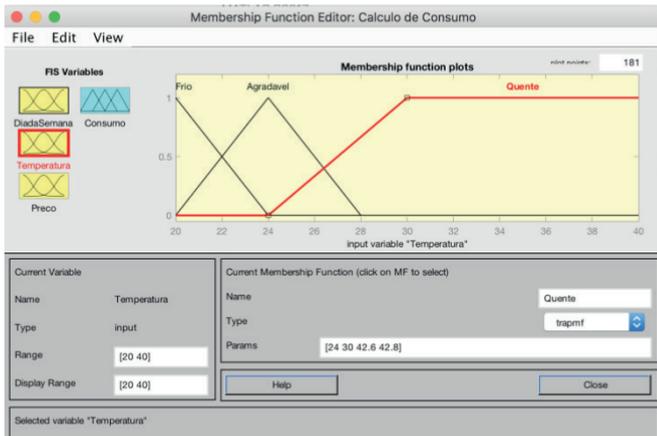
Figura 3.7 | Configuração da variável Dia da Semana



Fonte: captura de tela do MATLAB 2017a, elaborada pelo autor.

Para a variável Temperatura, podemos considerar variações lineares, sendo que a partir de 20°C assumimos que temos a temperatura inicial adequada para o consumo, apesar de ainda estar frio para essa atividade, frio esse que permanece até os 24°C . Até 28°C (com pico em 24°C) podemos considerar que temos uma temperatura agradável. Podemos, por fim, considerar a temperatura como uma crescente “quente” a partir de 24°C até atingir o pico de 30°C . A partir daí, a temperatura é considerada simplesmente como totalmente “quente”. A Figura 3.8 ilustra a variável Temperatura:

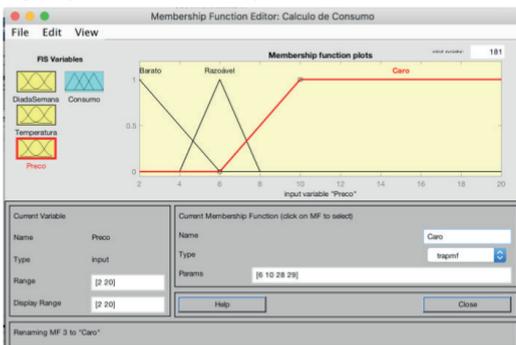
Figura 3.8 | Configuração da variável Temperatura



Fonte: captura de tela do MATLAB 2017a, elaborada pelo autor.

Para a variável preço, podemos considerar uma cerveja de boa qualidade (e a definição “boa qualidade”, aqui, não carece de definição em termos de lógica nebulosa, para fins de simplificação do problema). Para cervejas de boa qualidade, consideramos o preço barato a partir de R\$ 2,00 por uma garrafa de 600ml. O preço deixa de ser barato aos R\$ 6,00, mas podemos considerá-lo razoável nessa faixa, até que deixa de sê-lo nos R\$ 8,00. A cerveja começa a ser considerada cara aos R\$ 6,00 e atinge 100% nesse quesito em R\$ 10,00, permanecendo nessa faixa a partir daí. A Figura 3.9 ilustra a variável Preço no pacote MATLAB:

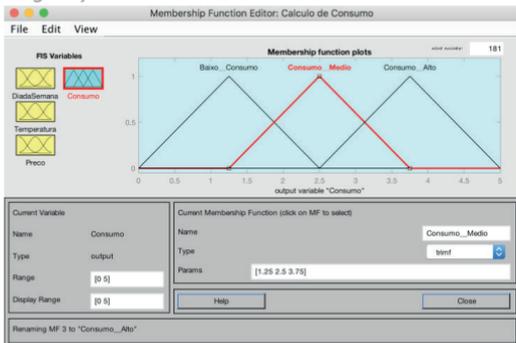
Figura 3.9 | Configuração da variável Preço



Fonte: captura de tela do MATLAB 2017a, elaborada pelo autor.

No que concerne a variável de saída (a que chamamos de consumo), vamos considerar o consumo como sendo baixo de 0 até 2,5 litros por dia, moderado até 3,75 litros e alto até 5 litros. O mapeamento dessa variável corresponde ao processo de defuzzificação, uma vez que gera resultados quantificados a partir das variáveis e das funções de pertinência (em ambos os casos, nebulosas). A Figura 3.10 ilustra a variável Consumo:

Figura 3.10 | Configuração da variável Consumo



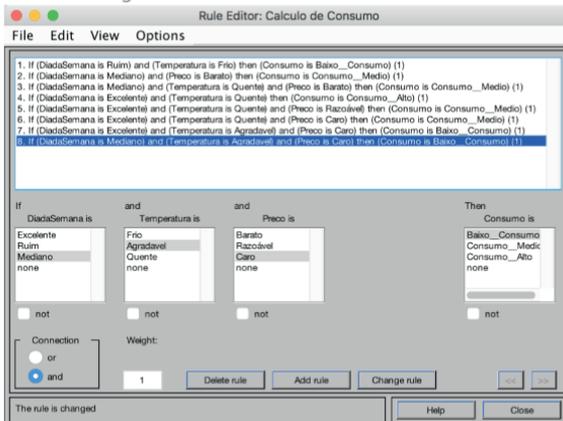
Fonte: captura de tela do MATLAB 2017a, elaborada pelo autor.

A partir daí, o que nos resta é definir as regras de inferência para o comportamento da variável de saída (Consumo) a partir de valores assumidos pelas variáveis de entrada (Dia da Semana, Temperatura e Preço). Não há uma quantidade ideal para as regras de inferência, e o sistema deve comportar tantas quantas forem necessárias para identificar o comportamento adequado e fazer uma previsão de consumo mais precisa possível. Algumas das regras que podemos inferir:

- Se o Dia da Semana for Péssimo (Segunda, Terça ou Quarta) e a Temperatura for baixa, o consumo será baixo (independente do preço da cerveja, é importante observar);
- Se o Dia da Semana for Mediano (Quarta ou Quinta) e o Preço for barato, o consumo será mediano (independente da temperatura, é importante observar);
- Se o Dia da Semana for Mediano, a Temperatura for quente e o Preço for barato, o consumo será mediano (estamos no meio da semana);
- Se o dia da semana for Excelente (Sexta, Sábado ou Domingo) e a Temperatura for quente, o consumo será alto (independente do Preço).

É importante observar que essas regras de inferência são arbitrárias e devem sempre refletir a realidade observável dos fatos. A Figura 3.11 ilustra algumas das regras possíveis no painel de definição de regras de inferência do MATLAB:

Figura 3.11 | Painel de Regras de Inferência



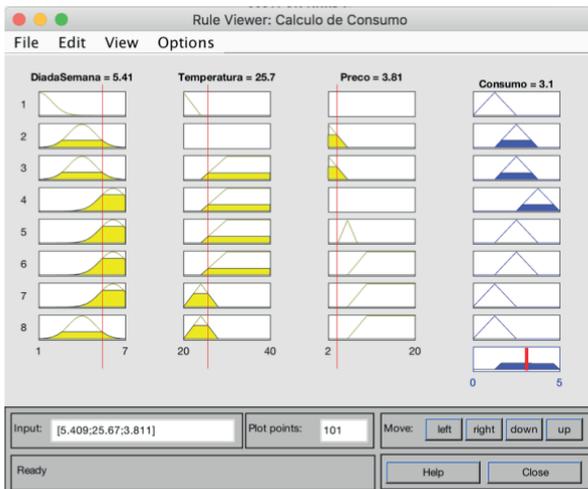
Fonte: captura de tela do MATLAB 2017a, elaborada pelo autor.



Todos nós usamos a inferência em nossas vidas. Se estamos ao ar livre e o chão está molhado, por exemplo, podemos inferir que choveu recentemente, ou que alguém estava lavando a calçada. Se o chão está molhado em uma grande área (na calçada em frente a todas as casas dos dois lados da rua e na própria rua), podemos inferir que choveu; se apenas a calçada em frente a uma casa específica está molhada, podemos inferir que a calçada foi lavada.

A partir daí, o sistema está pronto e já podemos fazer o cálculo do consumo com base nas três variáveis de entrada. O resultado pode ser visto no próprio MATLAB, na janela de revisão de regras, em que podemos alimentar o sistema com valores diferentes para Dia da Semana, Temperatura e Preço, e imediatamente verificar o efeito no consumo de cerveja. A Figura 3.11 mostra o consumo em uma sexta-feira, com a temperatura de $25,7^{\circ}\text{C}$ e o preço da cerveja a R\$3,81. Nesse caso, o sistema calcula o consumo médio de cerveja em 3,1 litros. O resultado é útil, por exemplo, para o dono de um estabelecimento fazer o cálculo de estoque para um dia qualquer, com base na frequência que ele já tenha calculado por conhecer os hábitos de sua clientela.

Figura 3.12 | Revisão de Regras e estado do sistema para uma dada configuração das variáveis de entrada



Fonte: captura de tela do MATLAB 2017a, elaborada pelo autor.



O professor Fernando Ramos, da Estech, tem uma breve aula de excelente qualidade sobre Lógica Nebulosa, mais especificamente, sobre os conceitos de fuzzificação, inferência e defuzzificação. Assista e veja como esses conceitos são simples e poderosos:

ESTECH. **Inteligência Artificial – Lógica Nebulosa Fuzzy**. 15 dez. 2015. Disponível em: <<https://www.youtube.com/watch?v=jDfxsSo2owg&t=127>>. Acesso em: 24 ago. 2018.

Sem medo de errar

Sendo o nosso objetivo o mapeamento da resolução do problema do processo de criação do composto em Lógica Nebulosa, primeiramente, devemos estabelecer para os executivos da empresa petroquímica que esse processo deve ser quebrado em três fases, com breves explicações acerca das mesmas, a saber:

- **Fuzzificação** – processo pelo qual uma variável que pode ser numericamente definida deixa de ser representada pelos valores numéricos que possa assumir e passa a ser representada por um conjunto de funções de pertinência, que correspondem a definições difusas em linguagem natural (já realizada na seção anterior);
- **Inferência** – regra no campo do raciocínio aproximado que permite conclusões mais ou menos vagas, a partir de premissas mais ou menos vagas. Quanto mais vagas as premissas, mais vagas serão as conclusões. Simetricamente, quanto menos vagas as premissas, menos vagas serão as conclusões;
- **Defuzzificação** – processo de produção de resultados quantificados a partir das regras criadas para as funções de pertinência de um modelo nebuloso. Em outras palavras, é a tradução em um valor real da composição dos graus de pertinência, definidos pela aplicação das funções de pertinência em um dado valor de entrada.

Uma vez explicados esses conceitos, você pode passar à aplicação dos mesmos ao processo da indústria petroquímica. Em

primeiro lugar, você deve explicar que a fuzzificação tem início quando tomamos as três variáveis do sistema, a saber:

- Temperatura;
- Pressão;
- Vazão.

A fuzzificação tem sequência quando adotamos três conceitos em linguagem natural para definirmos o estado das variáveis do sistema. Os três conceitos que adotamos são:

- **Insuficiente** – a variável em questão (Temperatura, Pressão, Vazão) está abaixo do intervalo ideal;
- **Adequado** – a variável em questão (Temperatura, Pressão, Vazão) está no intervalo ideal;
- **Excessivo** – a variável em questão (Temperatura, Pressão, Vazão) está acima do intervalo ideal.

A última fase da fuzzificação ocorre quando atribuímos valores para a variação dos conceitos de “insuficiente”, “adequado” e “excessivo”, como feito no exercício anterior, definindo também como se dá essa variação. No caso, definimos que não são variações lineares, utilizando curvas ascendentes e descendentes para representá-las.

Terminado o processo de Fuzzificação, passamos ao processo de inferência, que tem início logo após o fim do exercício passado. Ele deve se iniciar com a criação de regras que combinem os três descritores em linguagem natural, para as três variáveis do sistema. As regras deverão ser criadas a partir do conhecimento dos operadores do sistema e devem se basear nos exemplos a seguir:

- Se a temperatura é excessiva e a pressão é insuficiente, então devemos aumentar a quantidade dos insumos, a fim de aumentar a pressão;
- Se a vazão é insuficiente, então devemos girar a válvula de vazão para aumentar a vazão;
- Se a temperatura está adequada, então devemos manter a caldeira emitindo a mesma quantidade de calor a fim de manter a temperatura.

O sistema deve ter tantas regras quanto necessárias para descrever todos os estados possíveis e os resultados desejados serem mapeados.

Por fim, devemos proceder ao processo de defuzzificação, que ocorre quando aplicamos as regras de inferência e geramos resultados numéricos para as variáveis de saída. Esses resultados numéricos serão aplicados de maneira a sempre gerar o conjunto de variáveis em estado ótimo para o processo, mantendo a produção do produto químico sempre em sua produtividade máxima.

Pronto! Você convenceu os executivos e está a caminho de um contrato lucrativo. Parabéns!

Avançando na prática

O braço mecânico que joga pingue-pongue

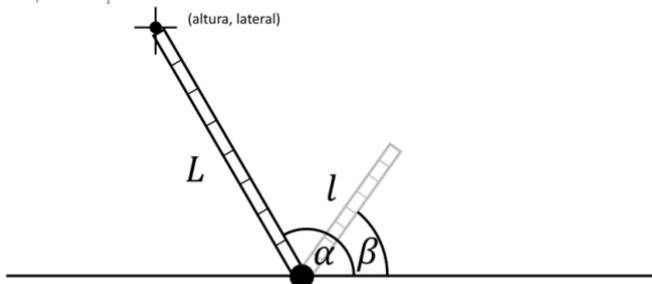
Descrição da situação-problema

Dois pesquisadores de uma proeminente universidade de sua região desenvolveram um braço mecânico que se movimenta com dois graus de liberdade:

- Variação no comprimento – o braço estica e encolhe, podendo passar de um comprimento “ l ” curto para um comprimento “ L ” comprido, e variar de 0 a 2 m;
- Variação no ângulo – o braço é fixo em um ponto e gira em torno desse plano, podendo passar de um ângulo α para um ângulo β .

Outro pesquisador da mesma universidade desenvolveu um sistema de visão eletrônica que consegue identificar a posição de um objeto, definindo se está mais alto ou mais baixo, e mais à direita ou mais à esquerda. A Figura 3.13 ilustra a combinação desses dois inventos:

Figura 3.13 | O braço mecânico e o olho eletrônico



Fonte: elaborada pelo autor.

Você, observando essas duas invenções, teve uma ideia: combiná-las, com a ajuda da Lógica Nebulosa, para obter como resultado um robô que joga pingue-pongue.

Na primeira fase do projeto, você deve proceder à fuzzificação das variáveis "Ângulo" e "Comprimento", com base na posição (altura, lateral) da bolinha, determinada pelo olho eletrônico. Como proceder a essa fuzzificação?

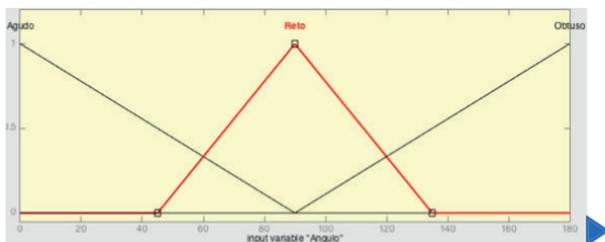
Resolução da situação-problema

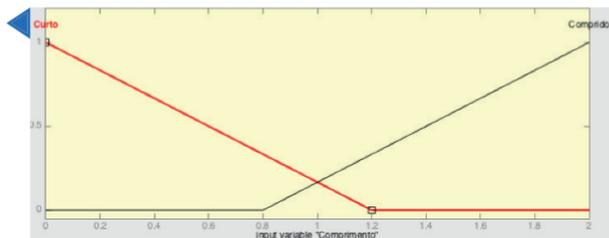
A fuzzificação das duas variáveis, "Ângulo" e "Comprimento" começa com a definição de conceitos em linguagem natural que estipulem os estados que ambas podem assumir. Para tanto, pense nos seguintes descritores conceituais:

- Ângulo
 - Agudo – Começa em 0° e vai até quase 90° , com variação linear;
 - Reto - Tem o pico, obviamente, em 90° , mas se aproxima a partir de 45° e deixa definitivamente de sê-lo em 135° , com variação linear;
 - Obtuso – Começa logo depois de 90° e vai até 180° , com variação linear.
- Comprimento
 - Curto – de 0 a 1,2 m (metade do comprimento total), variando linearmente;
 - Comprido – vai de 0,8m até 2 m (comprimento total), variando linearmente.

A Figuras 3.14 mostra no MATLAB o resultado da Fuzzificação:

Figura 3.14 | (a) Ângulo fuzzificado; (b) Comprimento fuzzificado





Fonte: captura de tela do MATLAB 2014b, elaborada pelo autor.

Assim, as variáveis foram fuzzificadas e você pode dar continuidade à construção de seu braço mecânico que joga pingue-pongue!

Faça valer a pena

1. A Fuzzificação é o processo pelo qual uma _____ que pode ser _____ deixa de ser representada pelos valores numéricos que possa assumir e passa a ser representada por um conjunto de _____, que correspondem a definições difusas em _____.

Assinale a alternativa que contém os termos (na ordem correta) que preenchem adequadamente as lacunas acima.

- a) ação, numericamente definida, definições difusas, linguagem natural.
- b) decisão, tomada, probabilidades difusas, linguagem difusa.
- c) variável, numericamente definida, funções de pertinência, linguagem natural.
- d) variável, definida em linguagem natural, funções de pertinência, linguagem numérica.
- e) constante, definida em linguagem natural, funções de pertinência, linguagem natural.

2. Trata-se de uma regra no campo do raciocínio aproximado que permite conclusões mais ou menos vagas a partir de premissas mais ou menos vagas. Quanto mais vagas as premissas, mais vagas serão as conclusões. Simetricamente, quanto menos vagas as premissas, menos vagas serão as conclusões.

O elemento que melhor sintetiza a definição dada acima é:

- a) Conclusão.
- b) Inferência.
- c) Lógica.
- d) Grau de certeza.
- e) Conhecimento.

3. O passo final da resolução de problemas por meio da Lógica Nebulosa é o processo de defuzzificação. Segundo Nguyen e Walker (2006), defuzzificação é o processo de produção de resultados quantificados, a partir das regras criadas para as funções de pertinência de um modelo nebuloso.

Ao usar a Lógica Nebulosa para calcular a gorjeta de um garçom em um restaurante, em função da qualidade da comida e da gentileza no atendimento, e entrar com valores para essas duas variáveis no sistema, que resposta é razoável receber, sabendo que valores razoáveis estão entre 5% e 20%?

- a) "Dê bastante gorjeta."
- b) "Dê pouca gorjeta."
- c) "Não dê gorjeta."
- d) "100% do valor da conta."
- e) "15% do valor da conta."

Referências

- AMENDOLA, M.; SOUZA, A. L.; BARROS, L. C. **Manual do uso da teoria dos conjuntos Fuzzy no Matlab 6.5**. Versão II, maio/2005. Disponível em: <http://www.logicafuzzy.com.br/wp-content/uploads/2012/12/manual_fuzzy_matlab.pdf>. Acesso em: 27 maio 2018.
- CHRISTA, D. **Fuzzy Logic**. 2015. Disponível em: <<http://matematica.ludibunda.ch/fuzzy-logic7.html>>. Acesso em: 27 maio 2018.
- DEROSS, G. **Fuzzy Logic in Real Life**. 3 jan. 2017. Disponível em: <<https://www.l-tron.com/fuzzy-logic-in-real-life/>>. Acesso em: 27 maio 2018.
- IRIA, J. **Teoria dos Conjuntos Difusos**. 10 mar. 2011. Disponível em: <<https://joseiria.files.wordpress.com/2011/02/teoria-dos-conjuntos-difusos1.pdf>>. Acesso em: 1 jun. 2018.
- KAY, R. **Sidebar: The History of Fuzzy Logic**. 30 ago. 2004. Disponível em: <<https://www.computerworld.com/article/2566110/sidebar--the-history-of-fuzzy-logic.html>>. Acesso em: 27 maio 2018.
- MATHWORKS. **Membership Function Gallery**. 2018. Disponível em: <<https://www.mathworks.com/help/fuzzy/examples/membership-function-gallery.html>>. Acesso em: 2 jun. 2018.
- NGUYEN, H. T.; WALKER, E. A. **A First Course in Fuzzy Logic**. 3. ed. Londres: Chapman & Hall, 2006.
- OLIVEIRA JUNIOR, H. A. et al. **Inteligência Computacional**. São Paulo: Thomson Learning, 2007.
- RUSSELL, S.; NORVIG, P. **Inteligência Artificial**, 3. ed. Rio de Janeiro: Campus, 2013.
- SEISING, R. **The Fuzzification of Systems**. Nova York: Springer, 2007.
- SILVA, E. M. et al. Uma aplicação de conjuntos difusos na otimização do prognóstico de consenso sazonal de chuva no nordeste do Brasil. **Revista Brasileira de Meteorologia**, v.22, n.1, p. 83-93, 2007. Disponível em: <<http://www.scielo.br/pdf/rbmet/v22n1/a09v22n1.pdf>>. Acesso em: 2 jun. 2018.
- SIVANANDAM, S. N.; SUMATHI, S.; DEEPA, S. N. **Introducion to Fuzzy Logic using Matlab**. Nova York: Springer, 2007.
- SOUZA, M. A. F.; GOMES, M. M.; SOARES, M. V. **Algoritmos e Lógica de Programação**. 2. ed. Rio de Janeiro, Cengage, 2011.

Redes neurais artificiais

Convite ao estudo

Olá! Vamos começar mais uma unidade de aprendizado, desta vez com um assunto que certamente está dando e vai continuar dando o que falar: as Redes Neurais Artificiais.

Estamos vivendo o início de uma revolução que já começa a afetar nossas vidas, e para melhor: a revolução da Inteligência Artificial. Por trás dessa revolução, os conceitos que vimos até aqui nessa disciplina estão sendo implantados cada vez mais, aproveitando-se do aumento da capacidade de processamento, memória e armazenamento dos dispositivos computacionais à nossa disposição. Nesse âmbito, talvez uma das áreas que mais cresça e apresente mais resultados úteis seja o campo das redes neurais artificiais. Um exemplo é o Google Duplex, uma tecnologia apresentada pela Google, em maio de 2018, por meio da qual um sistema consegue realizar ligações telefônicas se passando por um ser humano, sem que o interlocutor consiga identificar que está, na verdade, conversando com uma máquina. Por trás dessa nova tecnologia estão as Redes Neurais Artificiais, um dos mecanismos mais interessantes e poderosos da Inteligência Artificial, e esse será o tema abordado nessas próximas seções.

Nossos objetivos, aqui, são conhecer e compreender os conceitos, técnicas e aplicações das redes neurais na produção de sistemas computacionais. O governo federal, preocupado com a circulação potencial de terroristas por aeroportos brasileiros, está em busca de soluções que auxiliem na melhoria da segurança nesses locais de alta circulação. Para tanto, implantou um sistema de câmeras em cada aeroporto, com times de analistas observando as faces das pessoas presentes

e buscando os que parecem mais suspeitos, de acordo com grandes bases de fotos de terroristas. O sistema tem dado pouco resultado, com nenhum suspeito tendo sido identificado nos aeroportos, apesar de duas prisões terem ocorrido em território nacional e de ter sido identificado, posteriormente, que entraram no país por dois aeroportos vigiados. Um alto funcionário do Ministério da Defesa entrou em contato com sua empresa, que trabalha na área de Inteligência Artificial, especializada em redes neurais. Você foi convidado a auxiliar na resolução dessa importante questão. Há formas de usar as redes neurais para melhorar o desempenho na identificação de suspeitos? De que forma isso ocorreria? Será mesmo que as redes neurais podem ser eficientes na solução desse tipo de problema?

Na primeira seção desta unidade, você vai ver uma introdução à tecnologia de redes neurais, bem como seus principais elementos e conceitos. Já na segunda seção, você vai conhecer aspectos acerca do aprendizado em redes neurais. Por fim, na terceira seção, você vai se aprofundar nos conceitos, conhecendo as principais regras tanto do aprendizado quanto da retropropagação. Prepare-se para estudar bastante os materiais, fazer os exercícios propostos, e não deixe de realizar as práticas preparadas para o seu aprendizado.

Vamos em frente!

Seção 4.1

Introdução às redes neurais

Diálogo aberto

Você já percebeu como o cérebro humano é um dispositivo admirável? Dotados desse dispositivo, de nossas “ferramentas naturais” (mãos, pés, braços, pernas, costas) e de nossa força, fomos capazes de criar tudo de incrível, maravilhoso, perigoso, impressionante, útil (e por aí vai) que nos cerca. Criamos e produzimos todos os confortos da vida moderna, inclusive os computadores rápidos, potentes e baratos (sempre são baratos comparados com as gerações anteriores, dado que são mais potentes) que tanto nos ajudam em nosso dia a dia.

Em determinado momento, os cientistas computacionais se perguntaram como seria utilizar uma estrutura computacional que não fosse baseada nos conceitos de CPU binária que conhecemos, mas sim na estrutura de conexões entre neurônios que carregamos aos montes dentro de nossos cérebros. A partir dessa curiosidade, nascia uma das mais revolucionárias áreas da Inteligência Artificial, que, nos dias de hoje, nos permite enormes avanços computacionais, e alguns dos resultados mais importantes em termos de utilidade para o dia a dia do cidadão comum.

Nesta seção, vamos ver como se deu o nascimento dessa área da Inteligência Artificial, bem como conhecer os principais conceitos e elementos. Tendo recebido um convite para auxiliar na melhoria de um sistema de reconhecimento de pessoas suspeitas em aeroportos, nessa primeira fase, sua tarefa é importantíssima. Você deve apresentar uma alternativa ao uso de analistas e a introdução da possibilidade de utilizarem um sistema de reconhecimento de face baseado em redes neurais. O Diretor do projeto não concorda em deixar que computadores executem uma tarefa tão delicada, e você e seu time devem explicar a todos os envolvidos:

- O que é uma rede neural artificial.
- Como esses mecanismos vieram a fazer parte dos campos de pesquisa da computação.

- Quais são seus elementos e como uma rede neural se aproxima do aparelho cerebral humano.
- Quais são suas vantagens quando comparada com um sistema não cibernético (baseado em seres humanos).

Você deve criar uma apresentação em que todos os pontos acima fiquem claros para os decisores do projeto, permitindo que você e sua empresa passem à próxima fase. Vamos atacar mais esse desafio?

Então vamos lá!

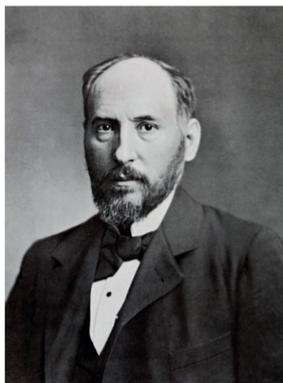
Não pode faltar

As Redes Neurais são “a bola da vez” nas pesquisas de Inteligência Artificial, retendo a atenção de pesquisadores no mundo todo, na busca por melhores e mais eficientes sistemas computacionais. Os resultados já começam a surgir: a cada dia, mais soluções inovadoras que se utilizam dessa tecnologia chegam até nós, usuários finais. Mas o que são Redes Neurais? Como esse conceito surgiu e evoluiu ao longo do tempo? Como funcionam as Redes Neurais? Respostas a essas e a outras questões você vai ver a partir de agora.

Histórico das Redes Neurais Artificiais

Segundo Priddy e Keller (2005), não há como estudar as redes neurais sem, antes, conhecer os elementos que as compõem: os neurônios. Os neurônios, de acordo com Lopez-Muñoz, Boya e Alamo (2006), são células especializadas que compõem a porção ativa de nossos cérebros e de nossos sistemas nervosos. São estudados pela biomedicina já há muito tempo, mas apenas a partir do início do século XX passaram a ser funcionalmente compreendidos, uma vez que apenas com a chegada das tecnologias daquela época foi possível isolá-los e estudá-los detalhadamente. O estudo dos neurônios rendeu o prêmio Nobel de Medicina de 1906 para o pesquisador espanhol Santiago Ramón y Cajal e para o fisiologista italiano Camillo Golgi, retratados na Figura 4.1 (a) e 4.1 (b), a seguir:

Figura 4.1 | Pesquisadores espanhóis que ganharam o prêmio Nobel de Medicina de 1906: (a) Santiago Ramón y Cajal e (b) Camillo Golgi



A



B

Fontes: <<https://upload.wikimedia.org/wikipedia/commons/3/30/Cajal-Restored.jpg>>; <https://upload.wikimedia.org/wikipedia/commons/5/5f/Camillo_Golgi.jpg>. Acesso em: 14 set. 2018.

Os estudos de Ramón y Cajal e Golgi foram fundamentais para compreendermos o funcionamento do cérebro, mas, algumas décadas antes, ainda no século XIX, o filósofo escocês Alexander Bain (1873) e o psicólogo americano William James (1890) já haviam proposto o desenvolvimento de mecanismos computacionais baseados no funcionamento cerebral. Ambos estavam várias décadas adiantados no que concerne às possibilidades tecnológicas para levar a cabo suas ideias. Ainda assim, as proposições de ambos permaneceram como possibilidades a serem investigadas (PRIDDY; KELLER, 2005).

Em 1943, segundo Anthony (2001), o neurofisiologista Warren McCulloch e o matemático Walter Pitts se basearam nos estudos de Ramón y Cajal e Golgi e propuseram um modelo matemático para o funcionamento do neurônio, o que lhes permitiu extrapolar o conceito para o campo dos circuitos elétricos. Os dois pesquisadores criaram o conceito de lógica limiar (em inglês, *threshold logic*), por meio do qual um determinado sinal elétrico recebido é retransmitido apenas se ultrapassa um limite arbitrário. Por meio desse conceito, McCulloch e Pitts implementaram um circuito elétrico simples, que demonstrava o conceito matemático que haviam criado. É possível demonstrar que os elementos artificiais desenvolvidos pelos dois pesquisadores são compatíveis com o modelo computacional de Alan

Turing, o que mostra que mecanismos baseados nesses elementos são úteis na construção de sistemas computacionais. Nascia, ali, o estudo das redes neurais artificiais como modelo computacional, a partir do que os pesquisadores chamaram de LTU (do inglês, *linear threshold unit*) ou Unidade de Limiar Linear. O elemento recebe, como já mencionado, uma carga elétrica e a transmite apenas se ela ultrapassa um limiar estabelecido e configurável.

Em 1949, como nos informam Priddy e Keller (2005), o psicólogo e pesquisador canadense Donald Hebb demonstrou que os neurônios reforçam as conexões com outros neurônios, que geralmente provêm sinais elétricos que o aparelho cerebral considera importantes (por meio de reações bioquímicas). Com base no trabalho de Hebb, os pesquisadores americanos Belmont Farley e Wesley Clark criaram a primeira rede neural artificial baseada em um computador digital, em 1954.

Em 1956, os pesquisadores da IBM Nathaniel Rochester e John Holland lideraram um grupo que tentou implementar uma rede neural auto organizada, mas a tentativa falhou, em função da falta de recursos computacionais. Com base nesse trabalho, em 1958, o pesquisador americano Frank Rosenblatt criou o **perceptron**, um agente inteligente capaz de determinar se um conjunto de elementos reduzidos às suas expressões numéricas pertence a uma determinada classe ou não. Em teoria, o conceito funcionaria, mas os recursos computacionais necessários ainda deixavam muito a desejar, o que impedia seu uso em classificações mais complexas. Em função disso, seu estudo foi abandonado no fim da década de 1960 (PRIDDY; KELLER, 2005).

Durante a década de 1970, segundo Priddy e Keller (2005), as pesquisas ficaram limitadas a alguns grupos, em função dos explosivos desenvolvimentos em computação pessoal. O surgimento do Altair 8800, em 1974, do Apple I, em 1976, e do TRS-80, em 1977, transformaram o cenário da informática, levando, pela primeira vez, os computadores para o campo do consumidor. Em função dessa passagem dos grandes (e caros) computadores para os de pequeno porte, muita ênfase foi dada a pesquisas e à produção de hardware e software para essas novas arquiteturas.

A retomada do interesse nas redes neurais se deu a partir de 1982, com o trabalho do pesquisador americano John Hopfield, que

criou a primeira rede neural associativa, demonstrando que o poder das redes neurais está no resultado produzido pela coletividade dos neurônios artificiais, e não pelo neurônio artificial individual (PRIDDY; KELLER, 2005).

Em 1986, o primeiro modelo de redes neurais de múltiplas camadas foi postulado e implementado em software, o que permitiu o avanço nas pesquisas e nos resultados práticos obtidos pelas redes neurais. Em 1987, o IEEE (*Institute of Electrical and Electronics Engineers*) promoveu a primeira conferência Internacional sobre Redes Neurais, com a participação de 1.800 pesquisadores do mundo inteiro. Desde então, muitos têm sido os avanços no campo de estudos das redes neurais artificiais, mas a arquitetura e os elementos permanecem os mesmos, e nós os estudaremos mais a fundo nos próximos itens dessa seção.



Pesquise mais

Um excelente relato sobre a história da evolução das redes neurais é contado no segundo capítulo do livro *Deep Learning Book*, que é disponibilizado gratuitamente:

DATA SCIENCE ACADEMY. Uma Breve História das Redes Neurais Artificiais. In: **Deep Learning Book**. [S.l.; s.n.; s.d.]. Disponível em: <<http://deeplearningbook.com.br/uma-breve-historia-das-redes-neurais-artificiais/>>. Acesso em: 14 set. 2018.

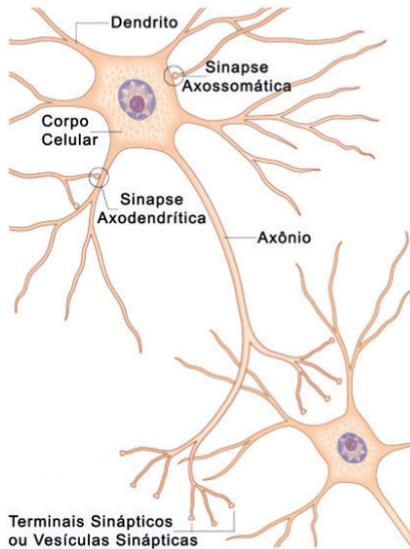
Características do Neurônio Biológico

Como as redes neurais são baseadas nos neurônios biológicos, é importante conhecermos esses elementos celulares tão importantes ao ser humano, antes de estudarmos seus pares cibernéticos.

Standing (2004) afirma que o sistema nervoso do ser humano é o produto mais complexo da evolução, e o elemento mais importante desse sistema é o neurônio. Contamos, cada um de nós, com aproximadamente 100 bilhões de neurônios em nosso sistema nervoso, e esses neurônios, em suas ações individuais simples, formam todos os nossos pensamentos e possibilitam todas as nossas ações.

A Figura 4.2, a seguir, mostra a anatomia de um neurônio:

Figura 4.2 | Anatomia do Neurônio



Fonte: adaptada de Standring (2004).

Os elementos componentes do neurônio são (STANDRING, 2004):

- **Corpo Celular ou Soma** – Onde se encontram o núcleo (que contém os cromossomos) e as organelas celulares;
- **Dendritos** – Extensões ramificadas do corpo celular que se conectam aos terminais sinápticos de outros neurônios, criando as sinapses;
- **Axônio** – Extensão que parte do corpo celular e leva os sinais elétricos até os terminais sinápticos (ou vesículas sinápticas). Pode ter parte de sua extensão coberta por bainhas de mielina, que são pequenas estruturas que facilitam o escoamento elétrico, isolando o axônio;
- **Terminais Sinápticos ou Vesículas Sinápticas** – Terminações que se conectam aos dendritos ou à soma de outros neurônios, formando as sinapses.

Segundo Standring (2004), os sinais elétricos são recebidos pelos dendritos e transmitidos pelo axônio até os terminais sinápticos, de onde são passados para outros neurônios ou para as terminações nervosas responsáveis pelas ações musculares, por exemplo.

Como já vimos anteriormente, as pesquisas de Donald Hebb demonstram que as conexões sinápticas mais utilizadas tendem a facilitar a passagem dos sinais elétricos, enquanto as conexões menos utilizadas não facilitam essa passagem. Esse mecanismo simples, quando estendido a grandes grupos de neurônios, permite o processo de aprendizado, sendo que uma ação percebida como vantajosa tende a ser repetida, tornando o caminho sináptico que leva à sua realização facilitado, à medida em que é repetido (STANDRING, 2004).



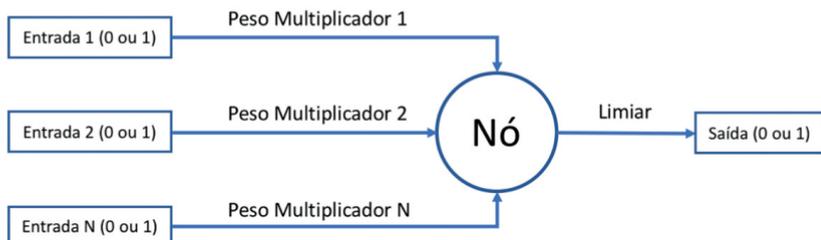
Assimile

O neurônio age, basicamente, como um condutor de sinais elétricos, que atua ou deixa de atuar dependendo da condição dos sinais de entrada.

Neurônio Booleano de McCulloch-Pitts

Russell e Norvig (2013) apresentam o Neurônio Booleano de McCulloch-Pitts como um elemento computacional capaz de receber várias entradas booleanas ("0" ou "1", apenas), atribuir pesos a cada uma delas, processá-las de acordo com esses pesos, comparar o resultado com um limiar pré-estabelecido e gerar um valor booleano de saída, com base nesta comparação: se for maior que o limiar, passa o sinal "1"; se não for, não passa (sinal "0"). A Figura 4.3 ilustra o Neurônio Booleano de McCulloch-Pitts (LTU):

Figura 4.3 | Representação do LTU



Fonte: elaborada pelo autor.

O valor a ser comparado com o limiar é obtido pela soma dos produtos dos sinais de entrada pelo seu respectivo peso. Em termos matemáticos, esse valor é obtido pela função de transferência:

$$Valor = \sum_{i=1}^n Entrada_i Peso_i$$

A saída é dada por uma função $f(Valor)$, que será "0", se for menor ou igual ao limiar, ou "1", se for superior ao limiar. Matematicamente, temos:

$$Saída = f(Valor) = \begin{cases} 0 \Leftrightarrow Valor \leq Limiar \\ 1 \Leftrightarrow Valor > Limiar \end{cases}$$



Refleta

O modelo de um LTU é baseado em um neurônio biológico. Seria razoável pensar em um modelo de LTU que não fosse baseado nesse tipo de célula? Por que não usamos outros modelos na construção de redes neurais?

As redes são formadas com a conexão da saída de um Neurônio à entrada de outro, e o processo de aprendizado ocorre no momento em que são alterados os pesos e os limiares, com base no resultado final. É importante observar que a saída de um LTU pode servir de entrada para um ou mais LTUs, ou seja, não é obrigatória a relação de unicidade entre uma saída e apenas uma entrada.

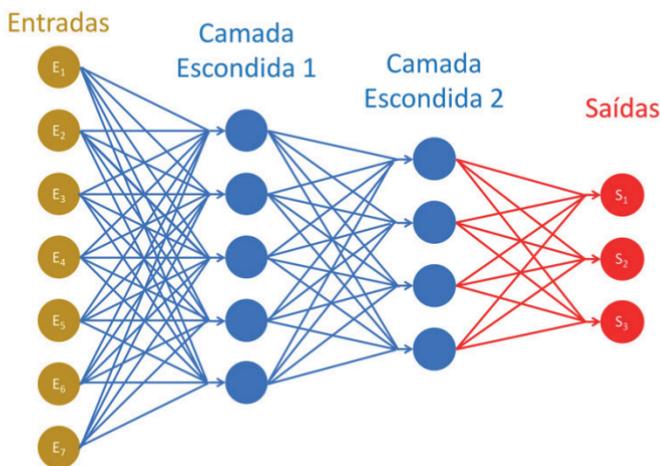
Segundo Russell e Norvig, os Neurônios Booleanos de McCulloch-Pitts não são mais utilizados no desenvolvimento de redes neurais pois, como veremos à frente, nesta unidade, outros modelos englobando novos elementos tornam as redes mais flexíveis, com maior capacidade de aprendizado e gerando melhores resultados. Contudo, são excelentes para demonstrar matematicamente como funcionam nossos neurônios, que recebem um ou mais sinais elétricos por meio das sinapses (que equivalem aos sinais "0" ou "1" de entrada), retêm ou passam a combinação desses sinais com base em processos bioquímicos internos (que equivale a calcular a função dos valores) e passam ou não o sinal para os demais neurônios (que equivale a comparar o resultado da função ao limiar pré-estabelecido e definir se a saída será "0" ou "1").

Conceitos e Elementos de Redes Neurais

Segundo Van Gerven e Bohte (2018), podemos definir as redes neurais artificiais (que passaremos a chamar simplesmente de “Redes Neurais” nesta unidade) como sistemas de computação inspirados pelas redes neurais biológicas que constituem cérebros humanos. A rede é formada, segundo os autores, por nós conectados em uma única direção (uma saída é sempre conectada a uma ou mais entradas de neurônios de uma camada à frente), formando uma rede, na qual sinais de entrada são processados e propagados (ou não), gerando combinações de sinais de saída.

Uma rede neural genérica pode ser representada como tendo uma camada de entrada, uma ou mais camadas internas (ou escondidas) e uma camada de saída. A Figura 4.4, a seguir, representa uma rede neural com sete nós de entrada, duas camadas internas e três nós de saída:

Figura 4.4 | Representação de uma Rede Neural



Fonte: elaborada pelo autor.

A Figura 4.4 já nos mostra alguns dos elementos constitutivos das redes neurais, a saber (VAN GERVEN; BOHTE, 2018):

- **Nós de Entrada** – Geralmente diferem dos nós escondidos e dos nós de saída por exibirem alguma capacidade de transformar sinais não digitais em entradas digitais. Sensores, por exemplo, são elementos que podem receber sinais

luminosos e transformá-los em sinais digitais, dependendo do estímulo externo que recebem.

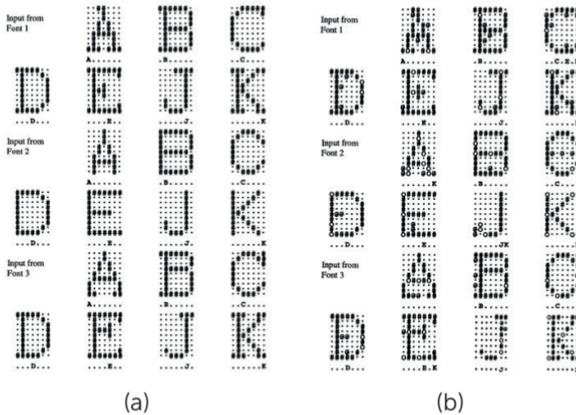
- **Nós escondidos** – São responsáveis pelo processamento interno da rede. Quanto mais nós escondidos e camadas, maior será a capacidade de processamento necessária para treinar a rede, porém, maior será a chance de obter resultados precisos após esse treinamento. Contudo, não é obrigatório que existam em uma rede neural.
- **Nós de saída** – São a última camada de processamento, devolvendo combinações de sinais digitais booleanos ("0" ou "1") que correspondem à resposta do sistema ao problema proposto.
- **Conexões Saída/Entrada** – Conexões entre os nós de uma camada com os nós da camada subsequente.
- **Pesos** – Fatores de multiplicação para cada um dos sinais de entrada em um nó.
- **Função de Processamento (ou ativação)** – Forma de combinar os vários binômios [sinal de entrada, peso] dentro do nó. Gera um valor que deve ser comparado com o limiar para gerar o sinal de saída.
- **Limiar** – Valor arbitrário estabelecido para cada nó com o qual o resultado da função de processamento será comparado. Se o valor for maior que o limiar, um sinal de saída será gerado (valor do sinal de saída: "1"). Do contrário (isto é, caso o sinal seja menor ou igual ao limiar), nenhum sinal de saída será gerado (valor do sinal de saída: "0")

Oliveira Jr. et al (2007) apontam que a rede neural tem duas fases: a fase de aprendizado e a fase de utilização. Na fase de aprendizado, utiliza-se algum dos vários mecanismos de reconfiguração de pesos e limiares para reforçar os estados que geram respostas corretas, ao mesmo tempo inibindo os estados que geram respostas erradas. Lembra-se dos processos determinísticos e heurísticos que discutimos na primeira unidade? O treinamento de uma rede neural é um processo heurístico, envolvendo inúmeros passos de tentativa e erro, construindo aos poucos uma configuração que gera um patamar aceitável de acertos. Uma resposta certa gera acertos na configuração que facilitam aquela resposta no futuro. Uma resposta errada gera acertos na configuração que dificultam aquela resposta no futuro.



Um exemplo comum para o uso de redes neurais é o reconhecimento de caracteres, em que uma estrutura como a representada na Figura 4.5 (a) é usada para treinar uma rede que, a posteriori, consegue identificar caracteres apresentados conforme a Figura 4.5 (b):

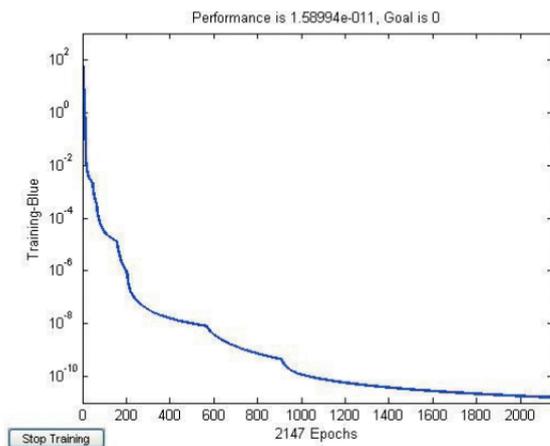
Figura 4.5 | Uma rede neural treinada com as imagens de (a) será capaz de reconhecer as imagens de (b)



Fonte: Fausett (1993, p. 72 e 75).

A Figura 4.6 mostra a curva de erros de um processo de aprendizado de uma rede neural. Observe que após mais de 2000 seções de treinamento (chamadas “épocas”), o percentual de erros caiu a pouco menos de $1,6 \times 10^{-11}$. Ainda não é zero (raramente será, em se tratando de redes neurais), mas já está em um patamar bem abaixo do que, digamos, um especialista no assunto conseguiria atingir:

Figura 4.6 | Curva de erros de uma rede neural em treinamento



Fonte: elaborada pelo autor.

Na fase de utilização, o processo de aprendizado é interrompido e a rede é posta para “trabalhar”, resolvendo os problemas para os quais foi projetada. Nessa fase, os algoritmos de recalibragem de pesos e limiares é interrompido e os valores atingidos permanecem por todo o período de utilização. Nada impede, claro, que de tempos em tempos essas configurações sejam revistas e atualizadas. Os assistentes pessoais digitais (por exemplo, Siri, OK, Google, Alexa) periodicamente têm seus parâmetros refinados para produzirem respostas melhores e mais precisas. Esses ajustes ocorrem espaçadamente, aproximando-se dos ciclos de correção de software, já tão familiares a quem conhece os ciclos de TI (VAN GERVEN, BOHTE, 2018).



Exemplificando

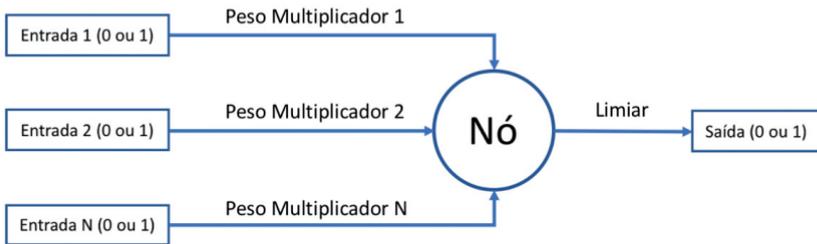
Um exemplo interessante de redes neurais que já passaram pela fase de treinamento e estão em pleno uso são os assistentes digitais da Microsoft, do Google e da Apple. Após extensas fases de treinamento, os percentuais de acuracidade no reconhecimento da voz humana desses assistentes são (RYAN, 2016):

- Alexa (Amazon): 90% de acuracidade;
- Google Now (Google): 92% de acuracidade;
- Siri (Apple): 95% de acuracidade.

No projeto para melhorar o reconhecimento de face nos aeroportos brasileiros, lembremos que você deve criar uma apresentação para que fiquem claros para os decisores do projeto vários pontos sobre o que são as redes neurais, como parte de um processo de convencê-los de que essa tecnologia pode ser de muita ajuda para aumentar a precisão nas identificações. Na apresentação a ser feita, você deve levantar os seguintes pontos:

- O que é uma rede neural artificial.
 - Aponte que se trata de um modelo computacional baseado em elementos simples que, atuando conectados uns aos outros, transmitem ou deixam de transmitir sinais digitais ("0" ou "1") com base em funções que combinam o sinal de entrada e as comparam com limiares estabelecidos.
- Como esses mecanismos vieram a fazer parte dos campos de pesquisa da computação.
 - Apesar de os sistemas computacionais baseados no funcionamento cerebral terem sido propostos no século XIX, foi apenas a partir do início do século XX, com o entendimento do funcionamento do neurônio, que os avanços foram possíveis na prática.
 - Em 1956, McCulloch e Pitts desenvolveram o modelo do LTU, abrindo caminho para a implantação de redes neurais.
 - A partir daí os modelos foram evoluindo, e nos dias de hoje as redes neurais são uma tecnologia amadurecida, proeminente nos estudos de Inteligência Artificial, que já provêm soluções úteis para o consumidor.
- Quais são seus elementos e como uma rede neural se aproxima do aparelho cerebral humano.
 - O elemento principal de uma rede neural é o nó, que os pesquisadores McCulloch e Pitts chamaram de LTU. A figura 4.3 mostra a representação de um LTU:

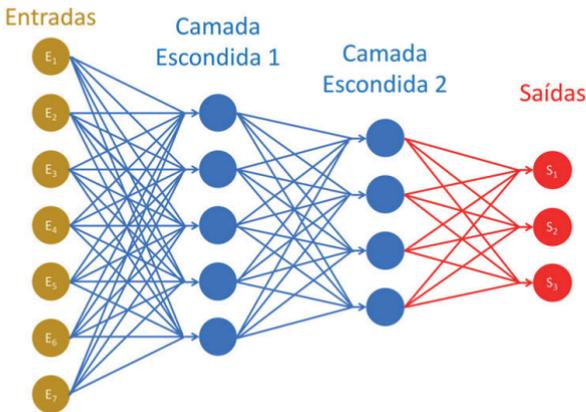
Figura 4.3 | Representação do LTU



Fonte: elaborada pelo autor.

- o O LTU pode ter várias entradas, cujos sinais digitais serão transformados por “pesos”, que são funções de multiplicação. Seu somatório é comparado a um limiar, e um sinal digital de saída (“0” ou “1”) é gerado, caso o valor seja superior a um limiar preestabelecido.
- o A conexão de vários neurônios artificiais gera uma rede, como a representada na Figura 4.4, a seguir:

Figura 4.4 | Representação de uma Rede Neural



Fonte: elaborada pelo autor.

- Quais são suas vantagens quando comparada a um sistema não cibernético (baseado em seres humanos).
 - o Uma rede neural pode ser treinada, isto é, pode receber milhares ou mesmo milhões de estímulos/respostas, que automaticamente vão calibrando a cada teste os pesos e limiares.

- o Uma rede neural não trabalha com as subjetividades, mas sim com os dados concretos
- o Uma rede neural, sendo um sistema cibernético, não comete erros por fadiga e não precisa de descanso
- o Esses são alguns dos pontos que você pode levantar nesse sentido, mas pode pensar em outros, se quiser

Pronto! Você já conseguiu mostrar aos decisores que vale a pena conhecer mais a fundo as possibilidades das redes neurais.

Avançando na prática

Redes Neurais e Eletrônica Básica

Descrição da situação-problema

Um laboratório de eletrônica, em que um amigo seu trabalha, está procurando uma alternativa para os antigos circuitos integrados de silício, e procurou você para ajudá-los com uma alternativa. Eles ouviram falar que as LTUs, tão comuns às redes neurais podem ser usadas para realizar tarefas que as portas lógicas realizam. Você sabe que este é o caso, mas eles querem uma demonstração. Para essa demonstração, você deverá mostrar-lhes como implementar uma porta "OR" usando LTUs. Como seria o esquema e o funcionamento desse elemento?

Resolução da situação-problema

Para demonstrar ao seu amigo como implementar uma porta "OR" usando LTUs, primeiramente você deve entender como funciona uma porta "OR".

Nesse tipo de porta, temos duas entradas e uma saída. Quando uma das duas entradas (ou as duas) tem o valor "1", a saída é "1". A saída será "0" apenas quando as duas entradas forem "0".

A Figura 4.7, a seguir, apresenta a tabela-verdade da porta "OR":

Figura 4.7 | Tabela-Verdade da porta "OR"

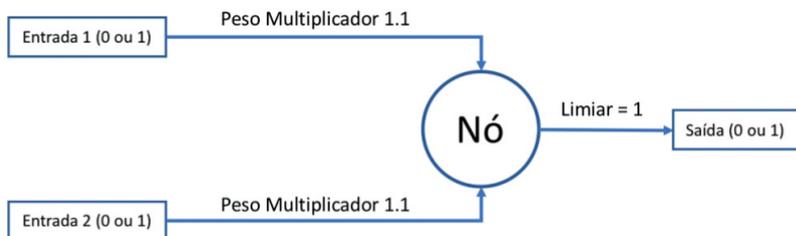
| | 0 | 1 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 1 | 1 |

Fonte: elaborada pelo autor.

Para criar uma porta "OR" usando LTUs, você deve pegar duas entradas que sejam uma combinação de "0" e "1" e produzir as saídas correspondentes da porta "OR".

Podemos fazer essa representação conforme a Figura 4.8, a seguir:

Figura 4.8 | LTU representando uma porta "OR"



Fonte: elaborado pelo autor.

Pelas entradas "1" e "2" podem ter passagem valores tanto "0" ou "1". Ao entrarem, são multiplicados por 1,1 e somados. Como o limiar é 1, qualquer entrada que seja diferente de "0" vai provocar uma saída "1", como no caso da porta "OR".

Aí está: você conseguiu demonstrar que consegue simular uma porta "OR" usando apenas uma LTU!

Faça valer a pena

1. As Redes Neurais são "a bola da vez" nas pesquisas de Inteligência Artificial, retendo a atenção de pesquisadores no mundo todo, na busca por melhores e mais eficientes sistemas computacionais. Os resultados já começam a surgir: a cada dia mais soluções inovadoras com essa tecnologia chegam até nós, usuários finais.

Qual o avanço no conhecimento permitiu a criação do neurônio artificial?

- a) A postulação, no século XIX, de sistemas computacionais baseados no cérebro.
- b) As pesquisas de Ramón y Cajal e Golgi acerca dos neurônios.
- c) A criação do neurônio artificial por McCulloch e Pitts.
- d) A pesquisa de Donald Hebb acerca do aprendizado.
- e) A invenção do computador pessoal.

2. As entradas de uma LTU são _____ e combinadas após serem multiplicadas por _____ individuais. A combinação desses valores é comparada com um _____ preestabelecido. Se for _____, gera um sinal booleano "1" de saída. Do contrário, nenhum sinal é enviado (sinal "0").

Qual das alternativas a seguir apresenta os termos que completam corretamente as lacunas acima?

- a) booleanas, multiplicadores, limiar, menor .
- b) digitais, divisores, limiar, menor.
- c) discretas, pesos, valor arbitrário, igual.
- d) booleanas, pesos, limiar, maior.
- e) contínuas, divisores, valor arbitrário, maior.

3. Nesta fase de uma rede neural artificial, utilizam-se algum dos vários mecanismos de reconfiguração de pesos e limiares para reforçar os estados que geram respostas corretas, ao mesmo tempo inibindo os estados que geram respostas erradas.

Qual fase de uma rede neural melhor sintetiza o trecho acima?

- a) Fase de Identificação.
- b) Fase de Revisão.
- c) Fase de Utilização.
- d) Fase de Testes.
- e) Fase de Aprendizado.

Seção 4.2

Redes neurais e aprendizado

Diálogo aberto

Você já parou para pensar como nós, humanos, aprendemos? Um exemplo simples é, quando crianças, sermos apresentados a um animalzinho qualquer, por exemplo, a um cachorro. Não sabemos o que é aquilo, a princípio, mas aos poucos vamos nos acostumando com sua forma e com suas características, e depois de um tempo conseguimos diferenciar o “au-au” do “miau”. Enquanto esse processo ocorre, dentro de nossos cérebros o que se passa é uma configuração constante de nossas sinapses. Aqueles caminhos dos sinais cerebrais que nos permitem identificar corretamente o cachorro são fortalecidos. Por outro lado, os caminhos dos sinais cerebrais que identificam o bicho como outro animal qualquer são enfraquecidos.

Nas redes neurais, o processo de aprendizado se dá de forma semelhante: configurações que levam à resposta correta são incentivadas, enquanto configurações inadequadas são reprimidas. Exemplo disso são as redes neurais de reconhecimento de imagens, que são treinadas com milhares ou milhões de imagens de animais, por exemplo, e conseguem identificar com bastante precisão quando uma imagem nova lhes é apresentada, identificando o tipo de animal a que se refere. Na prática, esse tipo de rede neural está sendo usado, por exemplo, para avaliar imagens médicas (tais como tomografias) e identificar patologias que possam passar despercebidas em seus estágios iniciais aos médicos responsáveis pelos diagnósticos.

Nosso objetivo maior, nessa seção, é entender como funcionam os processos de aprendizado, as chamadas regras de aprendizado, que regem o progresso das redes neurais, permitindo que gerem respostas úteis e inteligentes às questões que lhes propomos.

Você recebeu um convite para auxiliar no projeto de melhoria do sistema de reconhecimento de face que nosso governo usa nos aeroportos como ferramenta para tentar identificar pessoas perigosas, procurados pelas polícias internacionais e até mesmo

terroristas. Nessa fase, os responsáveis pelo projeto pedem a você que justifique a afirmação que ouviram de que as redes neurais “aprendem” e, assim, conseguem melhorar seu desempenho. Você deverá mostrar a eles como é possível uma rede neural aprender, e como se dá isso no nível de cada neurônio, colaborando, assim, para o desenvolvimento de seu raciocínio crítico e para sua capacidade de resolver problemas.

Você também deverá explicar que há várias formas — várias regras — de aprendizado, e como cada uma delas funciona. Assim, os decisores do projeto poderão entender que as redes neurais conseguem realizar o trabalho dos humanos na identificação de pessoas nos aeroportos, mas com mais eficiência e sem se cansar, sem necessitar de descanso, sem perder alguém por conta de alguma distração. Como você pode realizar essa tarefa? Quais são os pontos importantes a frisar acerca do aprendizado das redes neurais? Quais são as regras de aprendizado?

Nesta seção, veremos a definição de aprendizado nas redes neurais e as regras de aprendizado que permitem a evolução do conhecimento nesses elementos cibernéticos. Veremos também os Perceptrons e a regra de aprendizado de Rosenblatt. Está na hora de mergulharmos no aprendizado das redes neurais!

Preparado? Então vamos lá!

Não pode faltar

Depois de termos iniciado nossos estudos em redes neurais, conhecendo os principais conceitos dessa tecnologia, é chegado o momento de nos concentrarmos no principal mecanismo que torna as redes neurais uma poderosa ferramenta à nossa disposição: o processo de aprendizado. É o que veremos a partir de agora.

Aprendizado em Redes Neurais

Haykin (2005) afirma que a principal característica de uma rede neural é a capacidade de aprender com o ambiente (conjunto de entradas) que lhe é apresentado e, por meio desse aprendizado, melhorar seu desempenho.

Mas, como podemos definir formalmente o processo de aprendizado de uma rede neural? O próprio Haykin (2005, p. 72) nos

oferece uma definição formal: "a aprendizagem é um processo pelo qual os parâmetros livres de uma rede neural são adaptados por meio de um processo de estimulação pelo ambiente no qual a rede está inserida."

Podemos analisar essa definição com mais detalhes, quebrando-a em suas partes componentes:

- **Parâmetros Livres** – São considerados parâmetros livres os pesos de entrada de cada sinal e o valor do limiar (esse último, raramente alterado).
- **Adaptação dos Parâmetros Livres** – Processo de ajuste dos valores tanto de pesos quanto de limiares com base nos resultados obtidos.
 - Resultados corretos podem tanto aumentar os pesos de entrada que colaboram para o resultado, quanto diminuir o limiar para o sinal de saída que colabora com esse resultado;
 - Resultados incorretos podem provocar o comportamento inverso.
- **Estimulação pelo Ambiente** – O processo de aprendizado de uma rede neural ocorre por meio de estímulos providos por dados de entrada, cuja saída desejada é conhecida. Ao comparar-se a saída obtida com a saída desejada, é possível determinar de imediato se a rede neural está certa ou errada, que é o que permite a adaptação dos parâmetros livres.



Assimile

Apesar de o processo de aprendizado ter o potencial de afetar cada neurônio individualmente, podemos perceber que quem efetivamente "aprende" é a rede como um todo. Isso ocorre também conosco, em nível cerebral: nossas sinapses individuais se fortalecem ou enfraquecem, mas quem aprende somos nós, como um todo.

A título de discussão, Russell e Norvig (2013) lançam algumas questões importantes: se uma rede neural é, efetivamente, um sistema planejado e implementado, por que não o implementamos já com os parâmetros desejados? Em outras palavras, por que

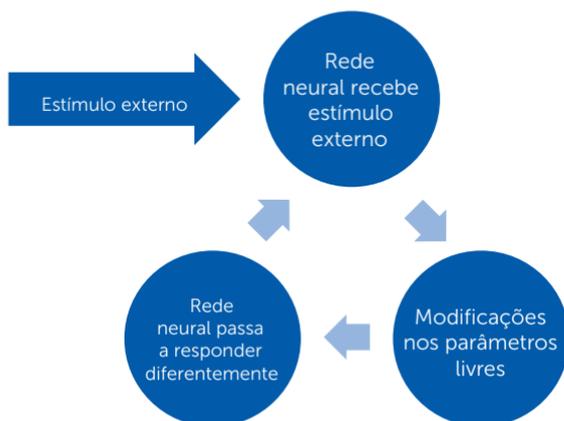
dependermos de um processo de aprendizado para chegarmos aos melhores parâmetros para o sistema, quando poderíamos ter poupado o tempo de treinamento/aprendizado e já programado esse sistema com os parâmetros mais apropriados?

Há três razões básicas que justificam a adoção do mecanismo de aprendizado (RUSSELL; NORVIG, 2013):

1. Os responsáveis pelo design do sistema não têm como antecipar todas as situações a que esse sistema possa ser submetido;
2. Os responsáveis pelo design do sistema não têm como antecipar todas as mudanças que tendem a ocorrer ao longo do tempo. Um sistema criado para prever, digamos, o comportamento da Bolsa de Valores, precisa se adaptar quando a conjuntura econômico-financeira muda em função de medidas tomadas pelo governo ou de crises nacionais ou internacionais que são, por natureza, imprevisíveis;
3. Em muitos casos, os responsáveis pelo design do sistema não têm a menor ideia de como resolvê-lo deterministicamente. Um exemplo dado pelos autores é o reconhecimento de faces: nós, seres humanos, somos excelentes em reconhecer faces de pessoas, porém, nem os melhores programadores conseguem desenvolver algoritmos determinísticos eficientes para o reconhecimento de faces.

O processo de aprendizagem resolve essas três questões, uma vez que a alteração dos parâmetros livres em uma rede neural permite algo que é natural às redes neurais, porém bem mais complexa de se obter em algoritmos determinísticos: adaptabilidade (o que dá uma resposta eficiente às duas primeiras questões), e apresenta um enfoque heurístico ao problema, que, se não dá uma resposta perfeita de imediato, ao longo dos vários ciclos de treinamento, vai melhorando a qualidade dos resultados, aproximando-se cada vez mais de um patamar aceitável.

Haykin (2005) apresenta o aprendizado como sendo um ciclo de três estágios, como pode ser visto na Figura 4.9, a seguir:



Fonte: elaborada pelo autor.

O ciclo apresentado na Figura 4.6 é simples:

1. A rede neural recebe um estímulo externo qualquer, que passa por dentro da mesma e gera um resultado de saída;
2. Com base no resultado, provocam-se modificações nos parâmetros livres para (a) facilitar as respostas corretas e (b) dificultar as respostas erradas;
3. A partir do ajuste dos parâmetros, a rede passa a responder diferentemente, podendo gerar respostas diferentes para o mesmo estímulo caso seja a ele submetida no futuro.

O ciclo, então, recomeça, com a rede neural sendo submetida a novos estímulos. Ainda no que diz respeito ao aprendizado, Haykin (2005) estabelece cinco regras de aprendizado que podem ser aplicadas às redes neurais. Vejamos essas regras.

Aprendizado por Correção de Erro

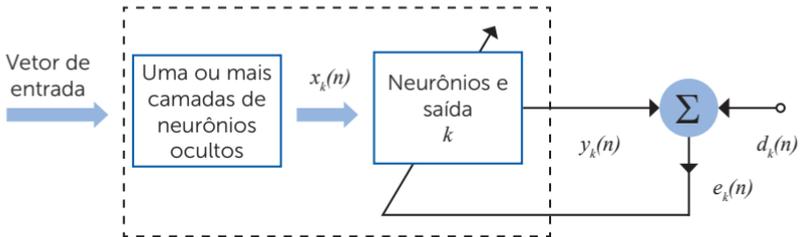
Segundo Haykin (2005), a regra de aprendizado por correção de erro se dá pelo uso de uma função de erro, que é o resultado da diferença entre o resultado obtido e o resultado esperado, utilizando essa função como entrada para um processo de correção dos pesos dos sinais de entrada para o neurônio em questão. A figura 4.10, a seguir, mostra um neurônio de saída k que, em determinado momento n , recebe um vetor de relações sinal-peso de entrada,

representada por $\mathbf{x}(n)$. O neurônio k gera um sinal de saída, a partir de um vetor $\mathbf{y}_k(n)$. Esse vetor $\mathbf{y}_k(n)$ é subtraído de um vetor desejado $\mathbf{d}_k(n)$, com o resultado sendo o vetor de erro $\mathbf{e}_k(n)$:

$$\mathbf{e}_k(n) = \mathbf{d}_k(n) - \mathbf{y}_k(n)$$

Esse vetor $\mathbf{e}_k(n)$ é realimentado ao neurônio k com o objetivo de corrigir os pesos dos próximos sinais de entrada.

Figura 4.10 | Aprendizado por correção de erro



Fonte: adaptada de Haykin (2005).

Aprendizado por Memória

Na regra de aprendizado por memória, define-se um conjunto de N duplas de vetores de entrada \mathbf{x} e vetores de saída desejados \mathbf{d} . O conjunto que servirá de memória para as correções futuras, matematicamente, tem o formato:

$$\{(\mathbf{x}_i, \mathbf{d}_i)\}_{i=1}^N$$

Quando se alimenta o neurônio com um vetor de entrada \mathbf{x} , o algoritmo consulta o conjunto de memória nas vizinhanças do \mathbf{x}_i mais próximo de \mathbf{x} . Os algoritmos de aprendizado baseados em memória têm duas características básicas:

- Contêm um critério para definir a vizinhança de \mathbf{x}_i ;
- Contêm uma regra de aprendizado (leia-se: ajuste dos parâmetros livres) com base na relação de $\mathbf{x}_{\text{teste}}$ com \mathbf{x}_i .

Aprendizado Hebbiano

O aprendizado Hebbiano se baseia no Postulado de Hebb (também conhecido como **Lei de Hebb**), criado pelo neurocientista Donald Hebb, em 1949. Podemos, segundo Haykin (2005), definir o postulado de Hebb em função de dois neurônios, A e B. O postulado

específica que quando um axônio do neurônio A está próximo o suficiente para excitar um neurônio B e participa desse processo de excitação de maneira persistente, por repetidas vezes, uma ou ambas as células envolvidas passam por mudanças metabólicas ou de crescimento, de forma que a eficiência de A para provocar o disparo em B é aumentada.

O Postulado de Hebb se refere ao contexto biológico, e pode ser detalhado por uma regra em duas partes (STENT, 1973; CHANGEUX; DANCHIN, 1976 *apud* HAYKIN, 2005):

1. Se dois neurônios de cada lado de uma sinapse (disparo elétrico provocado pela conexão entre dois neurônios) são ativados simultaneamente (ou seja, sincronamente), então a força dessa sinapse é seletivamente aumentada. Esta é a Lei de Hebb em sua forma original;
2. Se os dois neurônios envolvidos em uma sinapse são ativados de forma assíncrona, então essa sinapse é enfraquecida ou, no caso de já ser suficientemente fraca, eliminada por completo. Isso não fazia parte da Lei de Hebb original, mas é um corolário, isto é, uma decorrência, cuja definição ocorreu posteriormente.

A sinapse em questão leva o nome de **Sinapse Hebbiana**, que pode ser definida como uma sinapse que se vale de um mecanismo local, dependente do tempo e altamente interativo, que aumenta a eficiência sináptica em função da correlação entre as atividades pré-sináptica e pós-sináptica (HAYKIN, 2005). Quatro propriedades podem ser deduzidas a partir dessa definição (HAYKIN, 2005):

- **Mecanismo dependente do tempo** – As modificações na Sinapse Hebbiana dependem do momento exato em que os sinais pré-sináptico e pós-sináptico ocorrem.
- **Mecanismo local** – O processo ocorre em função da proximidade entre dois neurônios, o que demanda que sejam contíguos um ao outro. Cada sinapse terá um mecanismo semelhante, sempre localizado na conexão de dois neurônios e não regulado por nenhum mecanismo central.
- **Mecanismo interativo** – Uma mudança só ocorre se houver sinais em ambos os lados da sinapse. Em outras palavras, a mudança só ocorrerá em função da interação de dois neurônios.

- **Mecanismo conjuncional** – Uma mudança depende da conjunção de duas condições, isto é, da ocorrência síncrona ou assíncrona de sinais em ambos os lados da sinapse.

Com o objetivo de compreender o modelo matemático do Aprendizado Hebbiano, tomemos a sinapse entre dois neurônios, j e k . Vamos considerar w_{kj} como o peso da conexão do lado do neurônio k . Os sinais pré-sináptico (no neurônio j) e pós-sináptico (no neurônio k) são denotados respectivamente por x_j e y_k . A mudança provocada no peso w_{kj} em um momento n tem a seguinte forma (HAYKIN, 2005):

$$\Delta w_{kj}(n) = F(y_k(n), x_j(n))$$

Podemos ler a fórmula acima assim: a variação no peso sináptico do lado do neurônio k , em um determinado momento n , é dado por uma função F , que correlaciona os sinais pré-sinápticos x_j e y_k no mesmo momento n . Em sua forma mais simples, o Aprendizado Hebbiano assume a forma:

$$\Delta w_{kj}(n) = \eta y_k(n) x_j(n)$$

Ou seja, a variação e peso (que correspondem ao aprendizado) são o produto de $x_j(n)$ por $y_k(n)$ por uma constante de aprendizado η .



Exemplificando

Um exemplo interessante dos efeitos do Aprendizado Hebbiano pode ser visto em iniciantes nos estudos de piano. A análise da atividade cerebral de pessoas que nunca se dedicaram a aprender piano mostra que as regiões cerebrais responsáveis por identificar a sonoridade desse instrumento têm baixa atividade. Apenas algumas horas de treinamento prático já mostram resultados na atividade cerebral: as regiões responsáveis por identificar sons de piano ficam mais ativas. Quando se submete uma pessoa com poucas horas de estudo de piano a uma música que contém sons de piano, seu cérebro fica mais ativo e mais eficiente na tarefa de identificar esses sons. Esse processo de “fortalecimento” continua até atingir a atividade cerebral máxima nessas regiões, quando a pessoa se dedica aos estudos.

Aprendizado Competitivo

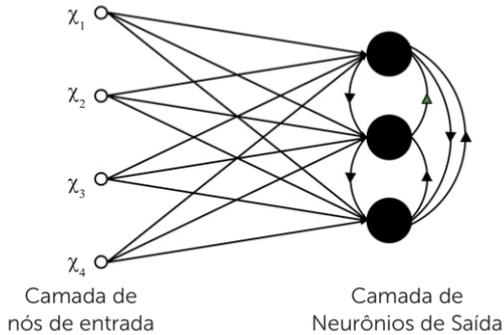
Segundo Haykin (2005), em uma rede neural regida pelo mecanismo Hebbiano de aprendizado, vários neurônios podem ser disparados simultaneamente. Já em uma rede neural regida pelo mecanismo de Aprendizado Competitivo, os neurônios de saída competem para ver qual deles será o responsável pelo disparo. Esse tipo de mecanismo é particularmente adequado para a descoberta de padrões estatísticos salientes em problemas de classificação. Há três elementos básicos em Aprendizado Competitivo, a saber (HAYKIN 2005):

- Um conjunto de neurônios de igual funcionamento, a não ser por alguns dos pesos sinápticos aleatoriamente distribuídos que, por conta dessa distribuição aleatória, respondem de maneira diferente a um conjunto qualquer de sinais de entrada.
- Um limite determinado pela “força” de cada neurônio.
- Um mecanismo por meio do qual os neurônios competem pelo direito de responder a um conjunto qualquer de sinais de entrada, de modo que um único neurônio de saída esteja ativo (ou seja, “ligado”) de cada vez. O neurônio que vence a competição é chamado de neurônio “*vencedor leva tudo*”.

Esse conjunto de características permite que cada neurônio da rede neural aprenda e, como resultado, se especialize em identificar padrões semelhantes. Ao final dos processos de treinamento, os neurônios se tornam detectores de características específicas nos conjuntos de dados de entrada.

A Figura 4.11 ilustra o funcionamento de uma rede neural de aprendizado competitivo, com uma camada de entrada e uma camada de saída. As conexões com sinalização positiva, que potencialmente excitam e provocam o disparo dos neurônios na camada posterior, são denotadas por retas, enquanto as conexões de sinalização negativa, que potencialmente inibem e impedem o disparo dos neurônios da mesma camada, são denotadas por curvas.

Figura 4.11 | Uma rede Neural de Aprendizado Competitivo



Fonte: adaptada de Haykin (2005).

Para que um neurônio k seja o vencedor, a somatória de seus sinais excitadores e inibidores v_k deve ser maior que a somatória dos sinais excitadores e inibidores de todos os demais neurônios. Matematicamente, temos que, em dado momento n , o sinal de saída do neurônio k , dado pela expressão $y_k(n)$, é obtido pela expressão (HAYKIN, 2005):

$$y_k(n) = \begin{cases} 0 & \Leftrightarrow v_k(n) \leq v_j(n) \mid j \neq k \\ 1 & \Leftrightarrow v_k(n) > v_j(n) \mid j \neq k \end{cases}$$



Refleta

O que aconteceria com uma rede neural se as conexões que geram sinal "0", isto é, que não estimulam os neurônios à frente, fossem simplesmente desfeitas? O que aconteceria com essa rede depois de alguns ciclos de treinamento? Seria possível manter sua coesão, nesse caso? Por quê?

Aprendizado de Boltzmann

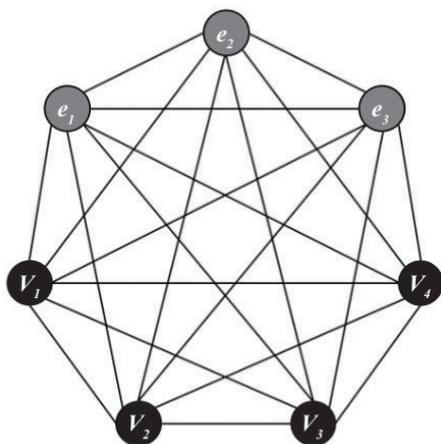
A regra de aprendizado de Boltzmann leva esse nome em homenagem ao físico austríaco Ludwig Boltzmann, que, no fim do século XIX e início do século XX, fez enormes contribuições à Física nos campos da Mecânica Estatística e da Termodinâmica. O nome dessa regra de aprendizado vem do fato de que a mesma se baseia em variações probabilísticas estocásticas (isto é, baseadas em processos aleatórios, fora do controle de quem constrói e

opera a rede neural) nos estados dos neurônios, com base em uma regra probabilística que leva o sistema todo a um equilíbrio após vários ciclos de treinamento. Uma rede neural baseada na regra de aprendizado de Boltzmann leva o nome de "Máquina de Boltzmann", e contém duas classes de neurônios:

- **Neurônios visíveis** – são aqueles conectados com o ambiente externo, isto é, que recebe estímulos externos;
- **Neurônios escondidos** – são aqueles que não recebem estímulos externos.

A Figura 4.12, a seguir, ilustra uma Máquina de Boltzmann, apresentando quatro neurônios visíveis, isto é, que recebem estímulos externos, e três neurônios escondidos, que apenas processam sinais, não recebendo estímulos externos:

Figura 4.12 | Uma Máquina de Boltzmann



Fonte: elaborada pelo autor.

Haykin (2005) afirma que, em uma Máquina de Boltzmann, os neurônios formam uma estrutura recorrente (todos os neurônios estão conectados a todos os demais, sempre excluindo autoconexões), com saídas binárias "0" ou "1", a exemplo dos demais neurônios de McCulloch-Pitts. Nessa configuração, a máquina como um todo é definida por uma "Função de Energia" (HAYKIN, 2005, p. 83) E , cujo valor é dado por um cálculo envolvendo o estado de cada neurônio:

$$E = -\frac{1}{2} \sum_j \sum_k w_{kj} x_k x_j$$

Nesta fórmula, temos que x_k é o estado do neurônio k , e w_{kj} é o peso sináptico da conexão do neurônio j com o neurônio k . A operação de uma Máquina de Boltzmann se dá quando o algoritmo escolhe aleatoriamente um neurônio k e altera seu estado de x_k para $-x_k$ (inverte o sinal). Esse processo se dá quando o sistema atinge um determinado estado, denotado como T (uma pseudo-temperatura, para ficarmos nos conceitos da física). A probabilidade P de ocorrer esta mudança dada uma temperatura T qualquer é dada pela fórmula:

$$P(x_k \rightarrow -x_k) = \frac{1}{1 + F(\Delta E_k / T)}$$

No caso, F é uma função da razão entre a variação de energia do neurônio k e a pseudo-temperatura T (estado atual do sistema). A aplicação repetida da regra leva o sistema a um equilíbrio térmico, ou seja, a um estado de aprendizado ótimo nas condições em que o sistema opera.

Perceptrons

Russell e Norvig (2013) classificam os perceptrons, fruto das pesquisas do psicólogo norte-americano Frank Rosenblatt, como uma das três ideias pioneiras que revolucionaram as Redes Neurais no século XX, ao lado do neurônio artificial de McCulloch-Pitts e do mecanismo de aprendizado de Donald Hebb.

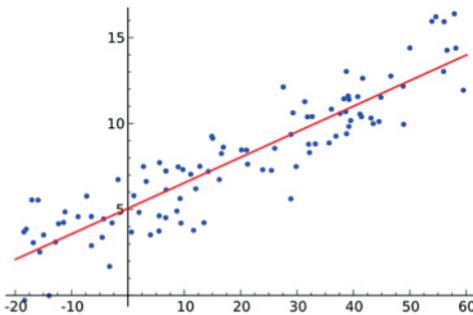
Podemos definir o perceptron, com base em Minsky e Papert (1988), como um elemento capaz de aprendizado supervisionado, cuja função é atuar em classificadores binários e cuja atuação se alicerça sobre uma função linear de previsão. Vamos analisar em mais detalhes essa definição (MINSKY; PAPERT, 1988):

- **Elemento capaz de aprendizado supervisionado** – Inicialmente, Rosenblatt definiu os perceptrons como dispositivos de hardware, mas, a partir da década de 1970, passaram a ser implantados como algoritmos, isto é, como software, de maneira mais barata e com melhor desempenho. Por isso defini-los como “elementos”, apenas. A capacidade de aprendizado supervisionado vem de mecanismos

semelhantes aos do aprendizado clássico dos neurônios de McCulloch-Pitts, isto é, por meio do ajuste nos parâmetros livres do elemento, reforçando o comportamento adequado e inibindo o comportamento inadequado.

- **Atua em classificadores binários** – O perceptron é, no fundo, um algoritmo baseado em uma função matemática capaz de decidir se uma entrada, representada por um vetor de valores numéricos, pertence ou não (daí o termo binário) a uma classe específica, definida arbitrariamente. Em outras palavras, identifica se uma entrada qualquer pertence a um conjunto ou está fora deste conjunto.
- **Utiliza uma função linear de previsão** – O perceptron utiliza uma função linear, que combina linearmente as variáveis e processa um conjunto de variáveis independentes, cujo valor é usado para prever o resultado de uma variável dependente. Esse tipo de função geralmente é uma regressão linear, que é uma aproximação linear de uma distribuição de pontos obtida por experimentação. Um exemplo é provido na Figura 4.13, a seguir.

Figura 4.13 | Exemplo de Regressão Linear



Fonte: <https://upload.wikimedia.org/wikipedia/commons/3/3a/Linear_regression.svg>. Acesso em: 14 set. 2018.

Como no caso dos neurônios artificiais vistos até aqui, o perceptron recebe uma série de n entradas \mathbf{x} , cada uma delas com um peso \mathbf{w} . O conjunto de todas as entradas \mathbf{x} pode ser expressado por um vetor da forma $[x_1, x_2, x_3 \dots x_n]$.

Levando em consideração essas entradas e um limiar l , o perceptron gera uma saída na forma $f(\mathbf{x})$, dada pela fórmula:

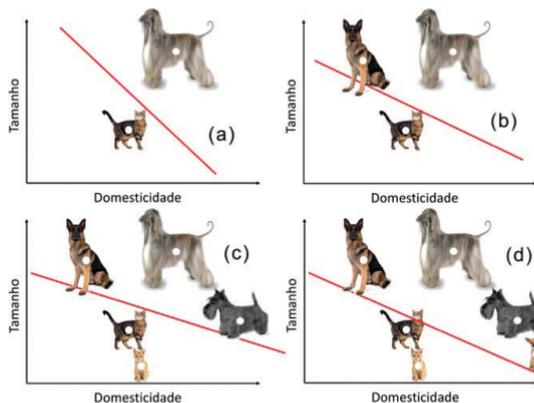
$$f(x) = \begin{cases} 0 & \Leftrightarrow w \cdot x + l \leq 0 \\ 1 & \Leftrightarrow w \cdot x + l > 0 \end{cases}$$

Na fórmula, temos a expressão $w \cdot x$, que é o produto vetorial dos pesos $[w_1, w_2, w_3 \dots w_n]$ pelas entradas $[x_1, x_2, x_3 \dots x_n]$. Esse produto é calculado pela fórmula:

$$w \cdot x = \sum_{i=1}^n w_i x_i$$

Os perceptrons, quando conectados em uma rede neural com uma camada de entrada e uma camada de saída, podem ser configurados com uma regra simples de aprendizado, conhecida como **Regra de Rosenblatt**, que gera, por meio dos produtos vetoriais $w \cdot x$, uma função linear que separa os elementos de um conjunto. A Regra de Rosenblatt estipula que os pesos $[w_1, w_2, w_3 \dots w_n]$ devem ser corrigidos a cada iteração do sistema com um par [entrada, rótulo], isto é, um estímulo de entrada acompanhado de um rótulo que estipula a que conjunto pertence aquele estímulo. Visualmente, podemos expressar a Regra de Rosenblatt por meio da Figuras 4.14, que mostra que a classificação obtida por meio da regressão linear (segmento de reta vermelho que separa os conjuntos) varia de figura para figura, à medida em que o limiar entre os conjuntos vai sendo ajustado a cada adição de um novo elemento. Com cada novo animal, a posição e angulação do segmento separador vão sendo recalculadas, como mostrado na progressão das figuras.

Figura 4.14 | Treinamento de uma rede de perceptrons por meio da Regra de Rosenblatt

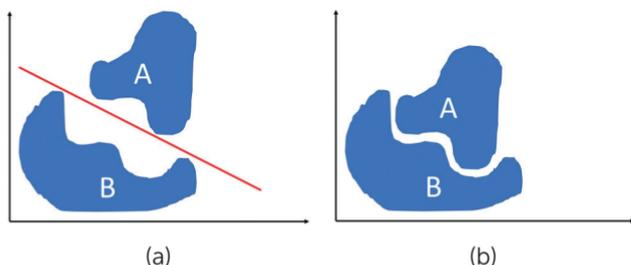


Fonte: elaborada pelo autor.

A partir do momento em que temos entradas, a regra de Rosenblatt estabelece uma divisão base, que pode ser vista na Figura 4.14 (a). À medida em que novas entradas vão sendo adicionadas, o sistema passa por novas iterações, ajustando a função linear de separação, de acordo com os pares [entrada, rótulo].

É importante observar que a Regra de Rosenblatt determina que os conjuntos devem ser linearmente separáveis, caso contrário, uma resposta exata não será possível, mas apenas uma resposta aproximada, com algum grau de erro. A Figura 4.15 (a) e (b) representa, respectivamente, conjuntos linearmente separáveis e conjuntos linearmente não-separáveis:

Figura 4.15 | Conjuntos linearmente separáveis e linearmente não-separáveis



Fonte: elaborada pelo autor.

Pesquise mais

Conheça mais a fundo os perceptrons por meio desse excelente capítulo da apostila de Redes Neurais da PUC do Rio Grande do Sul, intitulado "O Perceptron".

CASTRO, F. C. C.; CASTRO, M. C. F. O Perceptron. In: **Redes Neurais Artificiais**. [S.l.; s.n.]. 28 ago. 2001. Disponível em <http://www.politecnica.pucrs.br/~decastro/pdf/RNA_C3.pdf>. Acesso em: 14 set. 2018.

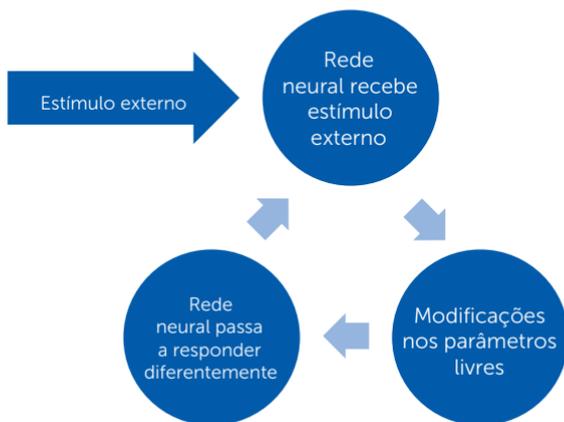
Sem medo de errar

Nessa fase do projeto de reconhecimento de faces, você deve mostrar aos responsáveis que a eficiência das redes neurais se dá pela possibilidade de aprendizagem das tarefas que delas demandamos. Para tanto, você pode oferecer a definição de aprendizado em redes neurais, dada por Haykin (2005, p. 72): "a aprendizagem é um

processo pelo qual os parâmetros livres de uma rede neural são adaptados por meio de um processo de estimulação pelo ambiente no qual a rede está inserida.”

Em seguida, você deve explicar que os parâmetros livres a serem adaptados são os pesos das ligações entre a saída de um neurônio e a entrada de outro, adaptação essa que ocorre em uma etapa em função dos resultados obtidos na etapa anterior. Apresente também, a Figura 4.9, que ilustra o ciclo de aprendizado de uma rede neural:

Figura 4.9 | Ciclo de Aprendizado em Redes Neurais



Fonte: elaborada pelo autor.

A próxima etapa dessa fase é apresentar os modelos de aprendizagem, que são 5, explicando como ocorrem:

- **Aprendizado por correção de erro** – Na saída do neurônio, calculamos uma função de erro, que é a diferença entre a saída obtida e a saída esperada. Esse valor é utilizado para corrigir os pesos de entrada no neurônio.
- **Aprendizado por memória** – A correção dos pesos se dá pela construção de um vetor de duplas, que contém duplas de valores de entrada e valores desejados de saída. Esse vetor tem suas duplas visitadas na vizinhança dos valores ideais de entrada e saída desejados a cada iteração, com os valores mudados para corrigir os pesos de entrada dos neurônios
- **Aprendizado Hebbiano** – O peso de uma sinapse (conexão entre dois neurônios) é aumentado se ambos os neurônios

são excitados simultaneamente (no mesmo momento temporal); por outro lado, esse peso é reduzido se ambos os neurônios são excitados em momentos diferentes.

- **Aprendizado competitivo** – Se em uma rede regida pelo aprendizado Hebbiano os neurônios podem ser excitados simultaneamente, em uma rede regida pelo aprendizado competitivo apenas um neurônio de saída é disparado. Os neurônios recebem sinais estimulantes das camadas anteriores, mas eles mesmos geram sinais inibidores para os neurônios de saída de sua mesma camada. O neurônio com o maior valor, resultante do somatório desses sinais estimulantes e inibidores, será o que vai disparar, e os resultados passam a ser corrigidos por esse somatório para o próximo ciclo
- **Aprendizado de Boltzmann** – Modelo nomeado em homenagem ao físico austríaco Ludwig Boltzmann, responsável pela Lei de Entropia e pelos mais importantes estudos em Termodinâmica. Esse mecanismo não enxerga os neurônios individualmente, mas sim a rede como um todo. Um estado de “energia” é atribuído ao sistema, dada uma configuração, e esse estado é corrigido com base em uma pseudo-temperatura (determinado momento no tempo que reflete uma configuração qualquer). Com o uso de uma função estocástica (aleatória) de variação em alguns pesos a cada ciclo, o sistema tende a um equilíbrio após vários ciclos, gerando a resposta esperada.

Pronto! Você conseguiu explicar adequadamente aos decisores do projeto como se dá o aprendizado em redes neurais, mostrando seus vários mecanismos, o que lhes permite enxergar que o processo é útil na identificação de faces nos aeroportos. Você e sua empresa foram selecionados para seguir em frente para a próxima fase. Parabéns!

Identificando aves de corte

Descrição da situação-problema

O Instituto Internacional de Ornitologia, órgão recém-criado pela UNESCO, está buscando formas de identificar aves apropriadas para corte. Um de seus diretores é entusiasta da Inteligência Artificial, em particular, das redes neurais, e contratou sua empresa para prestar uma consultoria. Ele crê que os perceptrons, guiados pela lei de aprendizado de Rosenblatt, podem auxiliar na identificação de espécies apropriadas para corte, com base no peso e na domesticidade. Para tanto, o Instituto preparou uma lista com doze espécies, apresentadas na Tabela 4.1, e gostaria de saber se o método é apropriado para a identificação desejada. Se for possível com essas doze espécies, o método será generalizado.

Tabela 4.1 | Doze Pássaros

| Nome da Ave | Peso Médio | Domesticidade (0 a 1) | Apropriado para corte? |
|-------------|------------|-----------------------|------------------------|
| Águia | 5,0 kg | 0,1 | Não |
| Albatroz | 5,0 kg | 0,2 | Não |
| Codorna | 0,15kg | 0,3 | Sim |
| Colibri | 0,005kg | 0,1 | Não |
| Faisão | 1,2kg | 0,1 | Sim |
| Falcão | 0,5kg | 0,1 | Não |
| Gaivota | 1,5kg | 0,2 | Não |
| Galinha | 2,0kg | 0,9 | Sim |
| Ganso | 9,0kg | 0,5 | Sim |
| Marreco | 0,5kg | 0,3 | Sim |
| Pato | 1,5kg | 0,7 | Sim |
| Peru | 6,0kg | 0,8 | Sim |

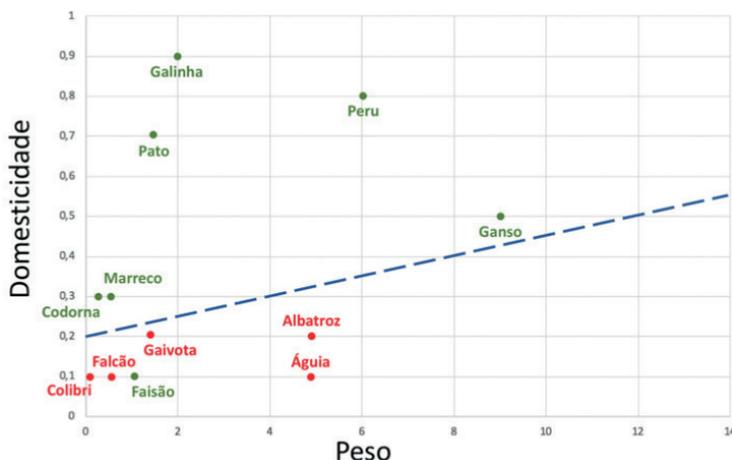
Fonte: elaborada pelo autor.

Que resultados podemos esperar para a aplicação da regra de aprendizado de Rosenblatt nesse caso? Com base nesses resultados, você indicaria o método para a generalização almejada pelo Instituto? Por quê?

Resolução da situação-problema

Para responder a esse desafio da UNESCO, o ideal é criar um gráfico com as informações de domesticidade e peso para cada ave, o que vai nos permitir identificar se é possível criar uma divisão linear entre os dados. A Figura 4.16 representa os resultados obtidos:

Figura 4.16 | Gráfico Domesticidade x Peso em 12 aves



Fonte: elaborada pelo autor.

Podemos identificar que não é possível reduzir linearmente as duas categorias, uma vez que os conjuntos não são linearmente divisíveis. Isso pode ser observado pelo fato de o faisão, que pode ser usado para corte, se encontrar em uma região cercada por aves impróprias para este fim.

Contudo, pelo menos para essa amostra, uma separação linear a que os perceptrons cheguem (como a feita com a linha pontilhada) terá apenas uma exceção (o próprio faisão), o que indica que há uma possibilidade de as exceções à regra serem poucas em relação ao todo.

Em outras palavras: a amostra aponta para uma provável inadequação do método para este tipo de separação, já que em uma amostra pequena já encontramos um dado que nega o método, mas vale a pena a UNESCO investigar com mais espécies antes de chegar a uma conclusão final, pois há a possibilidade de haver poucas exceções.

Faça valer a pena

1. Podemos afirmar que a principal característica de uma rede neural é a capacidade de aprender com o ambiente (conjunto de entradas) que lhe é apresentado e, por meio desse aprendizado, melhorar seu desempenho.

No que concerne o processo de aprendizado de uma rede neural, quando comparado ao processo de aprendizado cerebral em nível celular, podemos afirmar que:

- a) Os processos são fundamentalmente diferentes, uma vez que o cérebro não utiliza mecanismos de reforço e as redes neurais artificiais utilizam esses mecanismos.
- b) Os processos são fundamentalmente diferentes, uma vez que o cérebro utiliza mecanismos de reforço e as redes neurais artificiais não utilizam esses mecanismos.
- c) Os processos são semelhantes, uma vez que nem o cérebro nem as redes neurais utilizam mecanismos de reforço.
- d) Os processos são semelhantes, uma vez que tanto o cérebro quanto as redes neurais utilizam mecanismos de reforço.
- e) Os mecanismos são incomparavelmente diferentes, uma vez que um é biológico e o outro é cibernético.

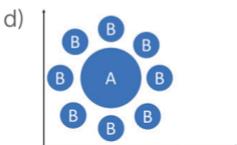
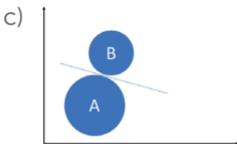
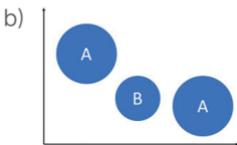
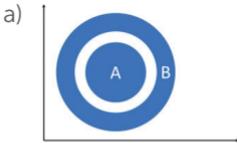
2. Em redes neurais baseadas nesse tipo de regra de aprendizado, temos uma camada de entrada e uma camada de saída. As conexões com sinalização positiva (que potencialmente excitam e provocam o disparo dos neurônios na camada posterior) são originadas na camada de entrada e terminam na camada de saída, enquanto as conexões de sinalização negativa (que potencialmente inibem e impedem o disparo dos neurônios da mesma camada) vão de cada um dos neurônios da camada de saída para cada um dos demais neurônios da camada de saída.

Assinale a alternativa que apresenta a regra de aprendizado a que o texto acima se refere:

- a) Aprendizado por correção de erro.
- b) Aprendizado por memória.
- c) Aprendizado Hebbiano.
- d) Aprendizado competitivo.
- e) Aprendizado de Boltzmann.

3. No que diz respeito aos perceptrons, é importante observar que a Regra de Rosenblatt determina que os conjuntos devem ser linearmente separáveis, caso contrário, uma resposta exata não será possível, mas apenas uma resposta aproximada, com algum grau de erro.

Assinale a alternativa que contém conjuntos que podem ser separados pela regra de Rosenblatt:



Seção 4.3

Elementos de Redes Neurais

Diálogo aberto

Vimos, até aqui, o funcionamento de vários tipos de neurônios e diversas de suas técnicas de aprendizado. Mas, como nossos cérebros nos mostram, a força vem do conjunto, não propriamente apenas do neurônio. A combinação de vários neurônios e, principalmente, sua capacidade de corrigir os pesos das conexões de acordo com a capacidade de contribuir ou não para respostas certas é que produz o que chamamos de inteligência. Essa inteligência pode ser observada em qualquer animal capaz de aprender algum comportamento e, também, nas redes neurais. O OK Google, a Alexa e a Siri (assistentes digitais), por exemplo, foram treinados durante milhares de horas para reconhecer vozes dos mais variados timbres e, ainda hoje, são feitas correções nos pesos dos nós dessas redes neurais para ajustar alguma nuance e mais preciso o seu reconhecimento de voz.

Você foi convidado a participar de um projeto do governo que visa determinar se há uma maneira de utilizar a tecnologia das redes neurais para o reconhecimento de faces nos aeroportos brasileiros. Nesta fase, você deverá completar sua participação, explicando aos decisores como se dá o processo de retropropagação. Para tanto, você deve explicar como emergiu essa ideia, e a melhor maneira de fazer isso é mostrando o caminho percorrido para se chegar à retropropagação, a partir da criação o ADALINE e do algoritmo dos quadrados médios, até chegar no algoritmo usado atualmente para a minimização de erros em redes neurais artificiais. Como se deu esse desenvolvimento? Por que a retropropagação garante que uma rede neural pode ser treinada e, assim, produzir resultados cada vez melhores e mais precisos? É possível realizar o algoritmo de retropropagação em uma rede neural com apenas uma camada de entrada e uma camada de saída? Por quê? Como isso afetaria o desenvolvimento de uma rede neural para reconhecimento de faces que fosse útil ao governo, como prevê o projeto?

Na presente seção você vai conhecer o modelo do neurônio ADALINE, bem como seu funcionamento, além da Regra Delta para minimização de erros, entendendo por que essa regra é útil e garante que, em uma rede neural por ela regida, sempre caminharemos para patamares mais baixos de erro. Em seguida, você vai conhecer a Regra Delta generalizada, compreendendo como ela serve de alicerce para o algoritmo de retropropagação. Pronto para fecharmos com chave de ouro o interessante mundo das redes neurais artificiais?

Então, vamos lá!

Não pode faltar

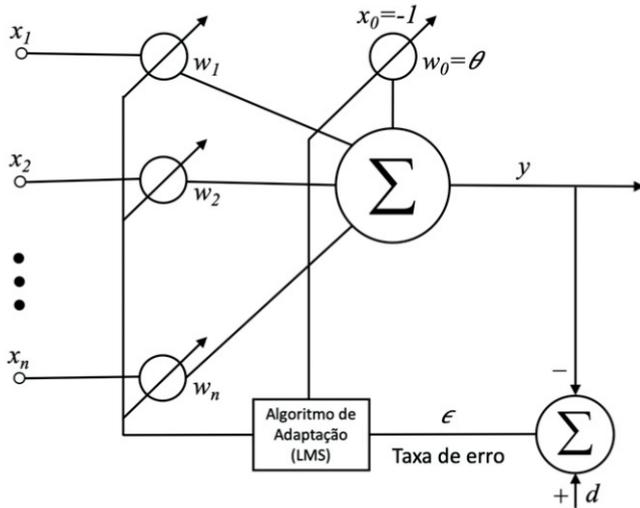
Estamos iniciando nossa última seção acerca de redes neurais, e é chegada a hora de conhecermos um dos mais importantes modelos de neurônios artificiais já projetados, bem como os algoritmos criados a partir dele. Esses algoritmos permitem ainda hoje a estabilização de redes neurais em pontos ótimos, que as tornam úteis para os mais diversos campos de atuação, tanto em nível acadêmico quanto em nível empresarial. Começaremos o aprendizado dessa seção pelo ADALINE e pelos algoritmos que derivaram desse tipo de rede neural.

ADALINE

Segundo Anderson e Rosenfeld (1998), ADALINE é o nome dado pelo pesquisador Bernard Widrow a um tipo de neurônio artificial desenvolvido por ele por seu aluno de doutorado, Ted Hoff, em 1960. ADALINE é a abreviação de Adaptive Linear Neuron (neurônio linear adaptável, em tradução livre). Os neurônios do tipo ADALINE, quando foram criados, eram implementados por meio de dispositivos eletrônicos (sem qualquer programação via software) chamados memristores, que são basicamente resistores capazes de reter a memória de seus estados resistivos (MOUTTET, 2010). Iniciando suas pesquisas em 1959, Widrow e Hoff refinaram o modelo do ADALINE ao longo da década de 1960, conseguindo inclusive implementar um sistema rudimentar de reconhecimento de voz e tradução para o inglês de frases originalmente faladas em idiomas como o francês (WIDROW, 2013). Sob o ponto de vista

estrutural, o ADALINE não difere muito do neurônio de McCulloch-Pitts, e é ilustrado na Figura 4.17, a seguir:

Figuras 4.17 | O ADALINE, de Bernard Widrow



Fonte: elaborada pelo autor.

Na Figura 4.17, temos:

- $x_1 \dots x_n$ – sinais de entrada;
- $w_1 \dots w_n$ – pesos para os sinais de entrada;
- y – saída;
- d – valor desejado para a saída;
- $w_0 = \theta$ – Constante para ajuste do valor de saída;
- ϵ – Taxa de erro;

Neste modelo, o valor $x_0 = -1$ ajusta a saída para um valor negativo, a fim de alimentar um segundo somatório (representado no canto inferior direito da Figura 4.13). O objetivo é permitir que seja determinado o valor da Taxa de Erro (ϵ), que vai ser utilizada para ajustar os pesos $w_1 \dots w_n$.

A principal diferença entre o comportamento do ADALINE e do neurônio padrão de McCulloch-Pitts é que, na fase de aprendizado do ADALINE, os pesos são ajustados de acordo com a soma das entradas ponderadas pelos pesos respectivos, ou seja, a saída é

uma **combinação linear** das entradas ponderadas. No neurônio de McCulloch-Pitts, as entradas ponderadas passam pela função de ativação (ou função de transferência) **não linear**, e a saída dessa função é usada para ajustar os pesos.

O peso w_0 é definido pela constante θ , de forma que podemos manter o sistema de entradas e pesos sem adicionar nenhum mecanismo extra ao modelo. Dessa forma, podemos definir o valor da saída do ADALINE como:

$$y = \sum_{k=0}^n x_k w_k$$

Podemos representar a fórmula acima em termos dos vetores envolvidos no somatório. Temos dois vetores a serem considerados:

- $X = [x_0 \dots x_n]$
- $W = [w_0 \dots w_n]$

Podemos, portanto, representar a saída y como:

$$y = X^T W = XW^T$$

De acordo com essa representação, a saída y é o produto vetorial do vetor transposto X^T pelo vetor W . É fundamental observar que esse produto é rigorosamente igual ao produto do vetor X pelo vetor transposto W^T .



Assimile

Originalmente, o ADALINE era um equipamento exclusivamente implantado por meio de dispositivos eletrônicos, em que todo o processo de aprendizagem era realizado sem o concurso de programas ou qualquer outro tipo de software.

Regra Delta

No modelo do ADALINE que discutimos anteriormente, o valor da Taxa de Erro é calculado da seguinte forma:

$$\varepsilon = d - y$$

$$\varepsilon = d - X^T W = d - XW^T$$

Segundo Anderson e Rosenfeld (1998), o modelo do ADALINE tem sua significância no contexto de redes neurais porque foi

por meio dele que Widrow e Hoff deduziram o algoritmo de minimização da média dos quadrados dos erros, chamado de *Least Mean Squares* (LMS), em inglês, e de Erro Quadrático Médio (EQM) ou **Regra Delta**, em português. Este algoritmo é utilizado como função de adaptação no ADALINE, e pode ser visto alimentando os pesos na Figura 4.17.

O algoritmo LMS busca calcular o quadrado da taxa de erro ϵ^2 , pois, dessa forma, trata de maneira simples tanto os erros positivos quanto os erros negativos, uma vez que, independentemente do sinal, o objetivo é trazer o erro para o mais próximo possível de zero. O algoritmo estabelece uma taxa de aprendizagem η (uma constante definida para o modelo) e, por meio dessa taxa, um peso w qualquer é atualizado de acordo com a seguinte fórmula:

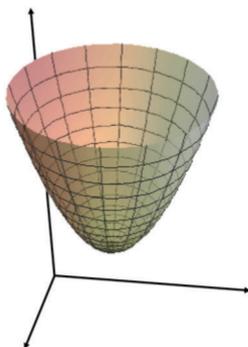
$$w \leftarrow w + \eta(d - y)x$$

A cada interação, o algoritmo LMS converge para a menor taxa quadrada de erro E , dada pela fórmula:

$$E = (d - y)^2$$

Como temos n variáveis de entrada (representadas pelo vetor $[x_0 \dots x_n]$) e o expoente é 2 (a maior ordem dos termos é quadrática), a curva representada pela equação acima é multidimensional. Podemos, porém, representar uma versão tridimensional dessa curva como uma parabolóide, como pode ser visto na Figura 4.18, a seguir. O algoritmo LMS vai sempre convergir para o ponto mais baixo da parabolóide:

Figuras 4.18 | Uma parabolóide representando a convergência do algoritmo LMS

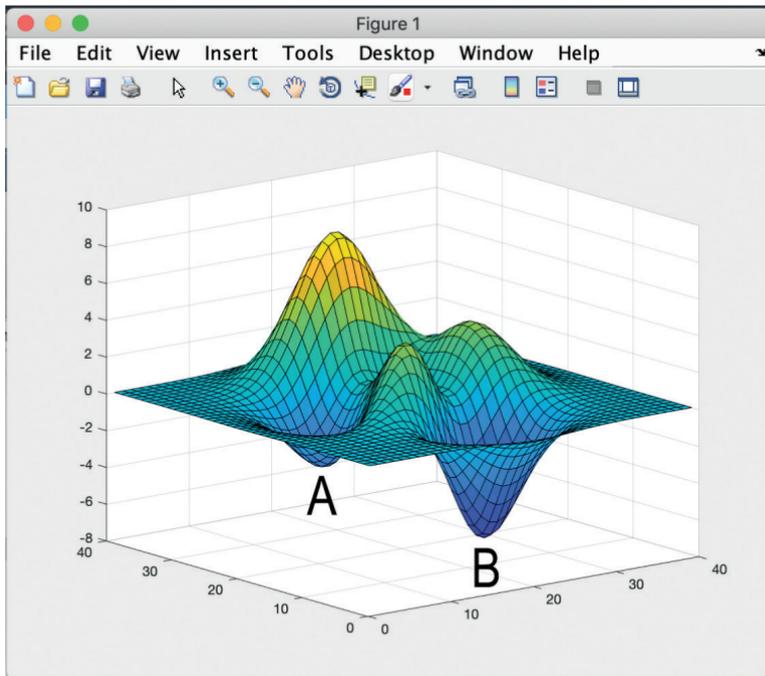


Fonte: elaborada pelo autor.



A Regra Delta, por buscar o mínimo em funções quadráticas multidimensionais, sempre buscará o mínimo em uma parabolóide multidimensional. O que ocorreria se, em vez de uma função quadrática, utilizássemos uma função de ordem superior, como no caso da curva tridimensional representada pela Figura 4.19, a seguir? Na busca pelo mínimo da função, haveria problema se parássemos ao encontrar o mínimo local (A) em vez de buscarmos o mínimo global (B)?

Figuras 4.19 | Curva matemática com mínimo local e mínimo global



Fonte: captura de tela do MATLAB 2017⁸, elaborada pelo autor.

Barbon (2013) apresenta a Regra Delta como um algoritmo em linguagem natural, representado na Figura 4.20, a seguir:

Início {Algoritmo Adaline – Fase de Treinamento}

- <1> Obter o conjunto de amostras de treinamento $\{x^{(k)}\}$;
- <2> Associar a saída desejada $\{d^{(k)}\}$ para cada amostra obtida;
- <3> Iniciar o vetor w com valores aleatórios pequenos;
- <4> Especificar taxa de aprendizagem $\{\eta\}$ e precisão requerida $\{\varepsilon\}$;
- <5> Iniciar o contador de número de épocas $\{\text{época} \leftarrow 0\}$;
- <6> Repetir as instruções:
 - <6.1> $E_{qm}^{anterior} \leftarrow E_{qm}(w)$;
 - <6.2> Para todas as amostras de treinamento $\{x^{(k)}, d^{(k)}\}$, fazer:
 - <6.2.1> $u \leftarrow w^T \cdot x^{(k)}$;
 - <6.2.2> $w \leftarrow w + \eta \cdot (d^{(k)} - u) \cdot x^{(k)}$;
 - <6.3> $\text{época} \leftarrow \text{época} + 1$;
 - <6.4> $E_{qm}^{atual} \leftarrow E_{qm}(w)$;

Até que: $|E_{qm}^{atual} - E_{qm}^{anterior}| \leq \varepsilon$

Fim {Algoritmo Adaline – Fase de Treinamento}

Fonte: BARBON (2018, p. 10)

O algoritmo demanda um conjunto de amostras para treinamento <1>, associado às suas saídas desejadas <2>. Em seguida, o algoritmo inicializa um vetor W com saídas aleatórias pequenas (próximas a "0") <3> e especifica uma taxa de aprendizagem η , bem como uma taxa de erros requerida ε (aqui chamada de "taxa de precisão") <4>. O próximo passo é a inicialização da variável de épocas (um contador simples) <5>.

A partir daí, entramos em um laço <6>, no qual a variável de erro anterior recebe o conteúdo da variável de erro atual <6.1>, calculamos o valor de saída (que o autor chama de u , em vez de chamar de y , como vínhamos chamando) <6.2>, incrementamos nosso contador de época <6.3> e, por fim, calculamos o valor do erro atual <6.4>. Esse laço é repetido até que a diferença do erro atual em relação o erro anterior seja menor que o valor da taxa de erros requerida ε .

Regra Delta Generalizada

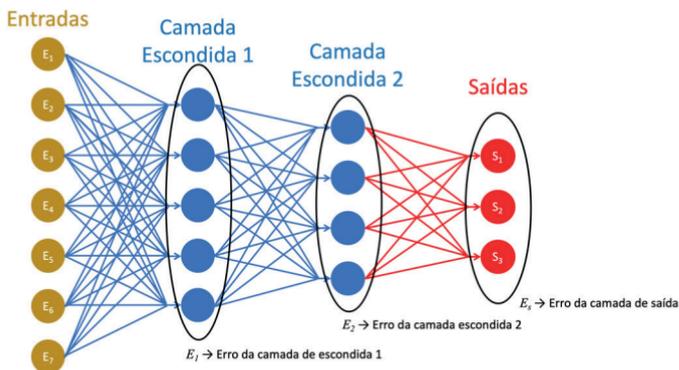
Segundo Winter e Widrow (1988), o ADALINE pode ser utilizado em conjunto com outros ADALINES, formando uma rede neural conhecida como MADALINE (muitos ou múltiplos ADALINES). Essa rede neural funciona em múltiplas camadas, com uma camada

de entrada, uma ou mais camadas escondidas e uma camada de saída, com todos os ADALINEs regidos pelo algoritmo LMS. Esse processo demanda uma generalização na Regra Delta para que possa funcionar adequadamente com redes de múltiplos níveis.

A Regra Delta Generalizada introduz os valores desejados para cada camada, assim como valores de erro para cada camada, e aplica o mesmo algoritmo da Regra Delta para todas as camadas escondidas da rede neural. Uma rede que tenha, por exemplo, uma camada de entrada, duas camadas escondidas e uma camada de saída, como representado na Figura 4.21, a seguir, terá três erros calculados:

- E_1 – Erro da camada escondida 1;
- E_2 – Erro da camada escondida 2;
- E_s – Erro da camada de saída;

Figuras 4.21 | Algoritmo da Regra Delta, em Linguagem Natural



Fonte: elaborada pelo autor.

A Regra Delta Generalizada é fundamental para o mecanismo de Retropropagação do Erro, que veremos a seguir.

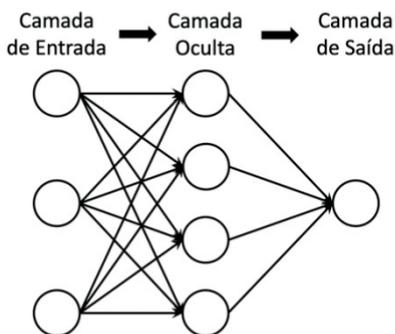
Retropropagação do Erro

Ramos (2002) credits ao algoritmo de retropropagação o renascimento das redes neurais artificiais, na década de 1980, uma vez que permite que as redes neurais apresentem características de aprendizado. O algoritmo de retropropagação, segundo Ramos (2002), se aplica a redes neurais com múltiplas camadas (e pelo

menos uma camada oculta) do tipo *feedforward* (com alimentação de avanço, em tradução livre).

Em uma rede neural desse tipo, segundo Ramos (2002), o fluxo é da camada de entrada para as camadas ocultas e, daí, para a camada de saída, sempre em um fluxo de avanço, da entrada para a saída. A saída de um neurônio de uma camada serve de entrada para os neurônios da entrada posterior, até a camada de saída. A Figura 4.22, a seguir, mostra este fluxo de avanço em uma rede neural do tipo *feedforward*:

Figuras 4.22 | Fluxo de uma rede neural *feedforward*



Fonte: elaborada pelo autor.

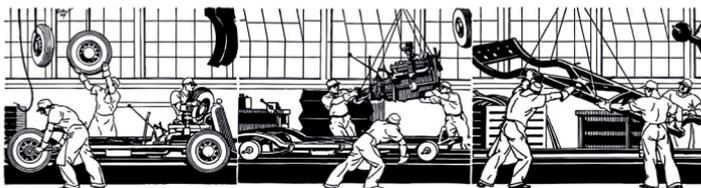
Uma vez que o fluxo da rede neural leva as entradas a produzirem um resultado y na camada de saída, esse resultado y é comparado com um resultado desejado d (como apresentado no ADALINE), e a diferença entre ambos gera um resultado que chamamos de taxa de erro ε .



Exemplificando

Um exemplo interessante de um mecanismo do tipo feedforward na indústria é a linha de montagem (de automóveis ou de smartphones, por exemplo). Os elementos vão sendo acrescentados nos objetos já parcialmente montados, e cada peça é só adicionada no momento em que é útil ao todo e é chegada sua hora de ser adicionada. A Figura 4.23, a seguir, ilustra esse conceito em uma linha de montagem de automóveis, em que cada peça é adicionada no momento certo e o sistema todo (o carro sendo montado) segue adiante para a próxima fase da montagem, com novas peças e novas configurações.

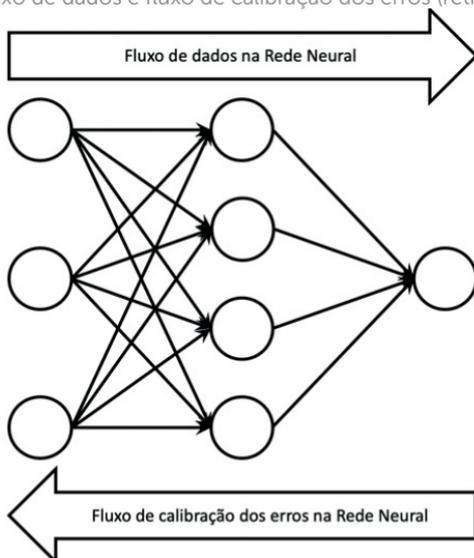
Figuras 4.23 | Fluxo feedforward em uma linha de montagem



Fonte: <<https://pixabay.com/pt/carro-montagem-linha-produção-35084/>>; <<https://pixabay.com/pt/carros-fabricação-linha-de-montagem-35082/>>; <<https://pixabay.com/pt/fábrica-planta-montagem-linha-35081/>>. Acesso em: 17 set. 2018.

Contudo, o cálculo do erro de saída é apenas o primeiro passo da retropropagação. Esse erro é utilizado para calcular o erro da última camada oculta, antes de chegar à camada de saída. Já o erro da última camada oculta é usado para calcular o erro da penúltima camada oculta, e assim sucessivamente, até chegarmos à primeira camada oculta. A direção do cálculo do erro e subsequente ajuste dos erros é, portanto, no sentido da camada de saída para a camada de entrada, ou seja, no fluxo contrário do avanço dos sinais na rede neural. A Figura 4.24, a seguir,

Figuras 4.24 | Fluxo de dados e fluxo de calibração dos erros (retropropagação)



Fonte: elaborada pelo autor.

Podemos, portanto, estabelecer o algoritmo de retropropagação de erro da seguinte forma:

Início (Algoritmo Retropropagação - Fase de Treinamento)

- <1> Obter o conjunto de amostras de treinamento
- <2> Associar a saída desejada para cada amostra obtida.
- <3> Iniciar vetor de pesos com valores aleatórios pequenos
- <4> Especificar taxa de aprendizagem e taxa de erros requeridos
- <5> Iniciar o contador de número de épocas
- <6> Iniciando pela camada de saída, repetir até terminar a primeira camada oculta
 - <6.1> Repetir as instruções
 - <6.1.1> Erro anterior da camada recebe Erro atual da camada
 - <6.1.2> Para todas as amostras de treinamento (conjuntos de entradas e saídas desejadas), fazer
 - <6.1.2.1> Calcular o valor da saída
 - <6.2.1.2> Ajustar o peso de acordo com a Regra Delta Generalizada
 - Erro atual da camada recebe Erro ajustado com base na Regra Delta Generalizada
 - Erro da camada anterior recebe Erro atual da camada
- Até que Erro atual da camada - Erro anterior da camada seja menor que Taxa de Erro esperada

Fim do Algoritmo

Como pode ser visto no algoritmo, trata-se de uma generalização para todas as camadas (excluída a camada de entrada, em que não há pesos) do algoritmo da Regra Delta.



O pesquisador Marco Antonio Fernandes Ramos apresenta em sua dissertação de mestrado uma explanação matemática completa acerca do algoritmo de retropropagação. Você pode encontrar essa explanação entre as páginas 28 e 34 da referida dissertação. Que tal se aprofundar nesse algoritmo?

RAMOS, M. A. F. **Sistema de Identificação e Localização de Falhas em Linhas de Transmissão Baseado em Redes Neurais**. 2002, p. 110. Dissertação (Mestrado em Engenharia Elétrica). Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Disponível em: <https://www.maxwell.vrac.puc-rio.br/Busca_etds.php?strSecao=resultado&nrSeq=3509@1>. Acesso em: 17 set. 2018.

Outro aspecto interessante (e bastante recente) é o uso malicioso das redes neurais e da Inteligência Artificial como um todo. O jornalista Itamar Melo, do site Versar, relata o surgimento de um número telefônico que leva a um aplicativo um tanto quanto inusitado e bastante perturbador: o Momo. Surgido inicialmente no Japão, mas já com números telefônicos na Colômbia, no México e nos EUA, o Momo é uma presença no aplicativo de comunicação WhatsApp. Esse aplicativo – e há poucas dúvidas de que se trata de um software, uma vez que seria necessário um batalhão de pessoas para atender às centenas ou milhares de conversas paralelas mantidas no mundo todo – começa inócuo, mas em pouco tempo se torna agressivo. Demanda ações violentas do interlocutor, apresenta imagens gráficas (e bastante perturbadoras) e chega ao ponto de ameaçar o interlocutor, apresentando-lhe informações pessoais. As redes neurais, que já são usadas comumente para interações com o usuário (como no caso da Siri, da Alexa e do OK Google), são as ferramentas perfeitas para esse tipo de interação, perguntando, respondendo e interagindo no idioma do interlocutor humano. E como consegue informações pessoais? Bem, aí estão as redes sociais que, se mal configuradas e mal utilizadas, escancaram nossas vidas para quem quer que se dê o trabalho de procurar, fornecendo informações pessoais em abundância.

Leia mais sobre o fenômeno “Momo” no artigo de Itamar Melo:

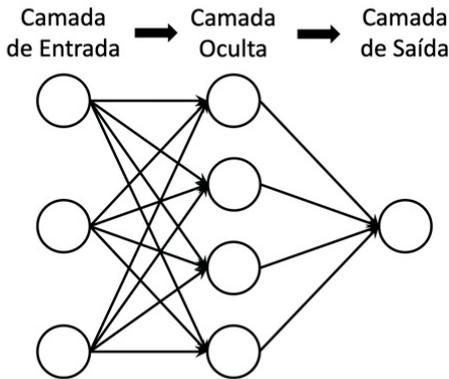
MELO, I. Momo do WhatsApp: o monstro existe mesmo? **Revista Versar**, 27 jul. 2018. Disponível em: <<https://www.revistaversar.com.br/momo-do-whatsapp/>>. Acesso em: 17 set. 2018.

Sem medo de errar

Com o objetivo de explicar o algoritmo de retropropagação e sua importância para as redes neurais, você pode criar uma apresentação com os seguintes itens:

1. Introdução
 - a. Título
 - b. Agenda
 - c. Propósito da Apresentação – Mostrar a evolução das redes neurais no que concerne o algoritmo de retropropagação, bem como pontuar a importância desse algoritmo para a melhoria dos resultados proporcionados pelas redes neurais artificiais.
2. Histórico
 - a. ADALINE – Explicar o que é o ADALINE, sua origem e suas características, pontuando que a principal colaboração de Widrow e Hoff (seus criadores) foi a identificação da regra dos quadrados médios (LMS), também conhecida como regra delta.
 - b. Regra Delta – Explicar o que é a regra delta e como ela propicia que se encontre o menor erro possível para uma rede neural.
 - c. Regra Delta Generalizada – Apresentar a regra delta generalizada, explicando que se trata de uma variação da regra delta.
 - d. Algoritmo – Apresentar o algoritmo da regra delta generalizada.
3. Retropropagação
 - a. Explicar o conceito de feedforward (alimentação de avanço) e mostrar uma rede neural que se baseie nesse conceito, como no caso da rede neural apresentada na Figura 4.22:

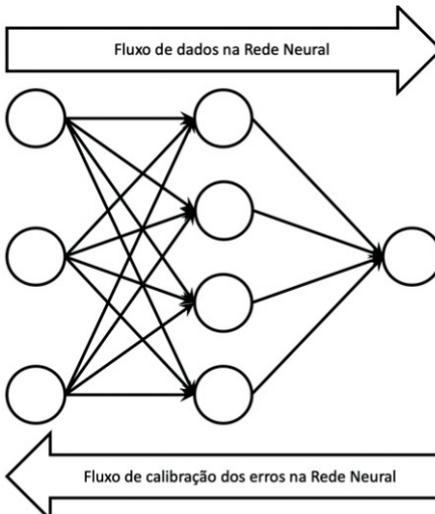
Figura 4.22 | Fluxo de uma rede neural *feedforward*



Fonte: elaborada pelo autor.

- b. Explicar o mecanismo da retropropagação, mostrando que, enquanto o fluxo de dados em uma rede feedforward é para frente (partindo da camada de entrada, passando pelas camadas ocultas, uma a uma, e chegando à camada de saída), o mecanismo de retropropagação vai no sentido contrário. A Figura 4.24 demonstra esse conceito e pode ser usada na explicação:

Figura 4.24 | Fluxo de dados e fluxo de calibração dos erros (retropropagação)



Fonte: elaborada pelo autor.

- c. Apresentar o algoritmo de retropropagação em linguagem natural
 - d. Mostrar, por meio do algoritmo de retropropagação, que a natureza de refinamento dos resultados obtidos pelo mecanismo de obtenção do resultado/retropropagação para correção dos pesos/obtenção do novo resultado/nova retropropagação de forma cíclica vai refinando os pesos e, por consequência, a qualidade dos resultados obtidos. A rede neural vai “aprendendo” por meio dos ciclos de geração de resultado/retropropagação e, assim, gera melhores resultados com o passar dos ciclos.
 - e. Explicitar que o mecanismo de retropropagação demanda que haja pelo menos uma camada escondida, e que não há como ser aplicado em redes em que as camadas escondidas não estão presentes, devido à própria natureza e estrutura do algoritmo.
4. Uso da tecnologia no projeto de reconhecimento de faces
- a. Explicar que o governo já conta com inúmeros pares de foto/identificação, que podem ser alimentados em uma rede neural.
 - b. Essa alimentação pode disparar os ciclos de geração de resultado/retropropagação que, por sua vez, vão treinando a rede neural a reconhecer as faces das pessoas de interesse, capturadas pelas câmeras dos aeroportos.

Pronto! Realizando sua apresentação de acordo com essas instruções, você terá sucesso em mostrar aos decisores do projeto que as redes neurais artificiais podem ser usadas com vantagens para o reconhecimento de faces nos aeroportos.

Avançando na prática

Melhora mesmo o cálculo do erro?

Descrição da situação-problema

Um engenheiro de computação amigo seu, buscando uma maneira de melhorar o desempenho do processo de minimizar o erro na Regra Delta, propôs o uso de uma função de quarto grau

em vez de uma função quadrática, uma vez que, apesar de os cálculos serem um pouco mais complexos (afinal, os expoentes são maiores), a regressão ao mínimo ocorreria mais rapidamente. Você, como experiente profissional da área de redes neurais, que conhece os porquês do uso de funções quadráticas para este fim, deve conversar com seu amigo e discutir os méritos e, quem sabe, os deméritos dessa ideia. Seria, de fato útil, utilizar uma função de quarto grau para encontrar o menor erro em uma rede neural? Por quê?

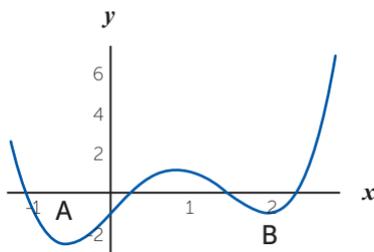
Resolução da situação-problema

Apesar de a ideia, a princípio, parecer boa, as funções de quarto grau sofrem de um problema potencial que não ocorre com as funções quadráticas: a possível ocorrência de um mínimo local. Esse problema potencial já desabona o uso de funções de quarto grau para esse fim. Para ilustrar esse conceito, você pode apresentar o caso da função:

$$f(x) = x^4 - 3x^3 + 2x - 7$$

Essa função, como é o caso de inúmeras funções de quarto grau, gera um gráfico com um mínimo local, o que geraria resultados indesejáveis ou demandaria muito mais processamento no processo de minimizar o erro nas redes neurais por meio da regra delta. A Figura 4.25, a seguir, ilustra esse conceito.

Figuras 4.25 | Mínimo local em uma função quártica



Fonte: elaborada pelo autor.

Na busca pelo ponto mínimo da função, o objetivo é encontrar o ponto "A". Contudo, essa função também apresenta um mínimo local, representado pelo ponto "B". Há um risco de o processo de minimização que utiliza uma função de quarto grau chegar ao

mínimo local, o que vai gerar uma resposta aquém do ideal, ou vai exigir todo um conjunto complementar de cálculos para chegar ao mínimo global esperado. Ambas as possibilidades são evitadas com o uso de uma função quadrática. Melhor seu amigo usar a função quadrática, não é verdade?

Faça valer a pena

1. Segundo Anderson e Rosenfeld (1998), ADALINE é o nome dado pelo pesquisador Bernard Widrow a um tipo de neurônio artificial desenvolvido por ele e por seu aluno de doutorado, Ted Hoff, em 1960. ADALINE é a abreviação de *Adaptive Linear Neuron* (neurônio linear adaptável, em tradução livre).

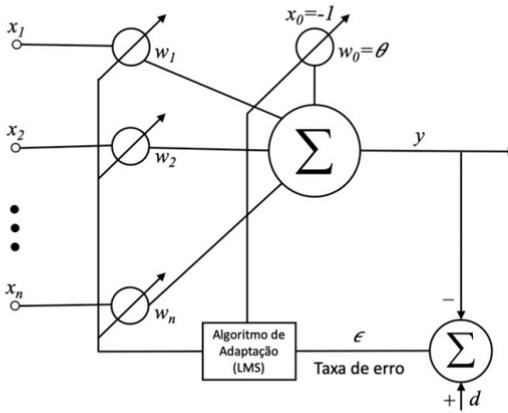
Assinale a alternativa que contém a principal diferença entre o ADALINE e o neurônio de McCulloch-Pitts:

- a) O ADALINE passa o sinal de saída por uma função de ativação, que é usada para ajustar os pesos.
- b) O Neurônio de McCulloch-Pitts não utiliza pesos de entrada, diferentemente do ADALINE, que usa esse mecanismo.
- c) O ADALINE não utiliza pesos de entrada, diferentemente do Neurônio de McCulloch-Pitts, que usa esse mecanismo.
- d) O neurônio de McCulloch-Pitts passa o sinal de saída por uma função de ativação não linear, que é usada para ajustar os pesos.
- e) Ambos os tipos são rigorosamente idênticos.

2. Segundo Anderson e Rosenfeld (1998), o modelo do ADALINE tem sua significância no contexto de redes neurais porque foi por meio dele que Widrow e Hoff deduziram o algoritmo de minimização da média dos quadrados dos erros (algoritmo LMS, abreviação de *Least Mean Squares*), mais conhecido em português como Erro Quadrático Médio (EQM), ou **Regra Delta**. Este algoritmo é utilizado como função de adaptação no ADALINE, e pode ser visto alimentando os pesos na figura a seguir.

Assinale a alternativa que explica adequadamente por que a Regra Delta busca minimizar o quadrado do erro, em vez de buscar minimizar o erro sem elevá-lo ao quadrado, uma vez que essa segunda forma seria menos custosa sob o ponto de vista computacional:

Figura | O ADALINE, de Bernard Widrow



Fonte: elaborada pelo autor.

- a) Ao elevarmos o erro ao quadrado, nós o aumentamos consideravelmente, o que facilita a identificação dos erros grandes.
- b) Ao elevarmos o erro ao quadrado, consumimos ciclos indesejáveis de CPU, o que torna o resultado mais robusto e menos sujeito a erros que surgem da velocidade.
- c) Ao elevarmos o erro ao quadrado, consumimos menos ciclos de CPU do que se o elevássemos ao cubo.
- d) Ao elevarmos o erro ao quadrado, garantimos que a taxa de erro sempre será par, o que facilita sua decomposição em dois blocos de fatores primos.
- e) Ao elevarmos o erro ao quadrado e buscarmos minimizar esse erro, o resultado será o mínimo para os módulos positivo e negativo.

3. O fluxo desse tipo de rede neural é da camada de entrada para as camadas ocultas e, daí, para a camada de saída, sempre em um fluxo de avanço, da entrada para a saída. A saída de um neurônio de uma camada serve de entrada para os neurônios da entrada posterior, até a camada de saída.

O texto acima se refere a que tipo de rede neural?

- a) ADALINE
- b) Perceptron
- c) *Feedforward*
- d) McCulloch-Pitts
- e) Não diz respeito às redes neurais, mas apenas à Lógica Fuzzy

Referências

- ANDERSON, J. A., ROSENFELD, E. **Talking Nets** – An Oral History of Neural Networks. Cambridge: MIT Press, 1998.
- ANTHONY, M. **Discrete Mathematics of Neural Networks**. Filadélfia: SIAM, 2001.
- BARBON, S. **ADALINE e Regra Delta**. Universidade Estadual de Londrina, Departamento de Ciência da Computação, 2013. Disponível em <http://www.barbon.com.br/wp-content/uploads/2013/08/RNA_Aula4.pdf>. Acesso em: 14 set. 2013.
- FAUSETT, L. **Fundamentals of Neural Networks**. Nova York: Prentice Hall, 1993.
- HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. 2. ed. Nova York: Prentice Hall, 2005.
- LÓPEZ-MUÑOZ, F., BOYA, J., ALAMOA, C. Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel Prize award to Santiago Ramón y Cajal. **Revista Brain Research Bulletin**, vol. 70, n. 4–6, pp. 391–405, 16 out. 2006.
- MINSKY, M., PAPER, S. **Perceptrons**. 2. ed. Cambridge: MIT Press, 1988
- MOUTTET, B. **Memristive Systems Analysis of 3-Terminal Devices**. 23 dez. 2010. Cornell University. Disponível em <<https://arxiv.org/abs/1012.5124>>. Acesso em: 14 set. 2018.
- OLIVEIRA JR, H. A.; et al. **Inteligência Computacional**. São Paulo: Thomson Learning, 2007.
- PRIDY, K. L.; KELLER, P. E. **Artificial Neural networks: An Introduction**. Washington: Spie Press, 2005.
- RUSSELL, S.; NORVIG, P. **Inteligência Artificial**. 3. ed. Rio de Janeiro: Campus, 2013.
- RYAN, K. J. Who's Smartest: Alexa, Siri, and or Google Now? **Inc.**, 03 jun. 2016. Disponível em: <<https://www.inc.com/kevin-j-ryan/internet-trends-7-most-accurate-word-recognition-platforms.html>>. Acesso em: 14 set. 2018.
- STANDRING, S. **Gray's Anatomy**. 39. ed. Amsterdam: Elsevier, 2004.
- VAN GERVEN, M.; BOHTE, S. **Artificial Neural Networks as Models of Neural Information Processing**. Lausanne: Frontiers Media, 2018.
- WIDROWLMS. **Science in Action**. 29 jul. 2013. Disponível em: <<https://www.youtube.com/watch?v=IEFrtz68m-8>>. Acesso em: 14 set. 2018.
- WINTER, R.; WIDROW, B. MADALINE RULE II: A training algorithm for neural networks. In: CONFERÊNCIA INTERNACIONAL DE REDES NEURAI DO IEEE, 1988, San Diego. **Anais...** San Diego: Stanford, 1988. p. 401–408 Disponível em <<http://www-isl.stanford.edu/~widrow/papers/c1988madalinerule.pdf>>. Acesso em 16 jul. 2018.

ISBN 978-85-522-1141-9



9 788552 211419 >