

O que é aprendizado de máquina?

André Carlos Ponce de Leon Ferreira de Carvalho

Resumo

Aprendizado de Máquina (AM) investiga como computadores podem aprender a resolver problemas utilizando experiências passadas, representadas por conjuntos de dados. Para isso, fazem uso do conhecimento de diferentes áreas, dentre as quais se destacam a computação, a matemática e a probabilidade e estatística. O aprendizado ocorre por meio de algoritmos de AM, que são projetados para diferentes tarefas e baseados em diferentes paradigmas. Como algoritmos de AM têm sido empregados com sucesso em um grande número de aplicações em setores variados, em particular no setor privado, cresce a demanda por especialistas nessas áreas. Apesar do aumento de oportunidades de formação, estudos realizados em vários países apontam que, no mundo inteiro, a demanda por profissionais cresce mais rápido do que a quantidade de profissionais formados, e a diferença aumenta a cada ano. Uma alternativa para lidar com esse problema é automatizar o uso da AM, o AM automatizado.

Abstract

Machine Learning (ML) investigates how computers can learn to solve problems using past experiences, represented by data sets. For such, it uses knowledge from different areas, among which computing, mathematics and probability and statistics stand out. Learning takes place through ML algorithms, which are designed for different tasks and based

on different paradigms. As ML algorithms have been successfully employed in a large number of applications in different sectors, particularly in the private sector, the demand for specialists in these areas grows. Despite the increase in training opportunities, studies carried out in several countries show that, worldwide, the demand for professionals grows faster than the number of trained professionals, and the difference increases every year. An alternative to deal with this problem is to automate the use of ML, the automated ML.

1. Introdução

A pesquisa científica, a melhor forma de adquirir e validar conhecimento que permita o avanço da sociedade, evoluiu e é desenvolvida por meio de paradigmas que regem a produção de conhecimento. Conforme HEY (2009), milhares de anos atrás, a ciência era empírica. Neste paradigma, o conhecimento buscava descrever fenômenos naturais, sendo adquirido por meio de experimento e observação dos resultados de relações causa-efeito. A ciência era basicamente experimental. Centenas de anos atrás, ganhou força o paradigma teórico, no qual a descoberta de conhecimento ocorria de forma analítica, por meio de uso de modelos e deduções. No meio do século passado tomou corpo o paradigma computacional, que, por meio da utilização de sistemas digitais, buscava explicar e simular fenômenos complexos. O uso de computadores levou a importantes descobertas em praticamente todas as áreas de conhecimento.

Desde que os computadores foram inventados, é perguntado se eles podem aprender. A palavra aprendizado faz parte do nosso cotidiano. Estamos acostumados a ela, ou suas palavras derivadas, desde crianças e somos continuamente lembrados de sua importância durante toda a nossa vida, por estar associada ao crescimento, desenvolvimento, amadurecimento e envelhecimento. E não as vinculamos apenas aos seres humanos, mas também a animais, principalmente os de estimação, e, agora, aos computadores, máquinas.

Assim como no aprendizado “tradicional”, em que são investigadas metodologias e técnicas que levem a um melhor aprendizado, mensurado pela aquisição de conhecimento ou habilidade, no Aprendizado de Máquina (AM) estudamos como computadores, ou máquinas, no seu sentido mais amplo, podem aprender a fazer algo.

O termo Aprendizado de Máquina foi popularizado pelo cientista da computação Artur Samuel, no seu artigo *Some Studies in Machine Learning Using the Game of Checkers* (SAMUEL, 1959), quando definiu que AM *é o campo de estudo que dá aos computadores a habilidade de aprender sem ser explicitamente programado*. Esse artigo descreve um programa de AM, desenvolvido pelo cientista, para jogar damas.

Outras definições de AM apresentadas ao longo do tempo seguem uma linha semelhante, como (ARKIN, 1998, p. 386):

- Uma máquina de aprendizado, definida de maneira ampla, é qualquer dispositivo cujas ações são influenciadas por experiências anteriores (NILSSON, 1965);
- Qualquer mudança em um sistema que o permita ter um melhor desempenho na segunda vez em que ele repita uma mesma tarefa, ou outra retirada da mesma população (SIMON, 1983);
- Modificação de uma tendência comportamental por meio de experiência (MISH, 1984);
- Uma melhoria na capacidade de processar informação a partir da atividade de processar informação (TANIMOTO, 1990);
- Capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência (MITCHELL, 1997).

AM é uma das várias subáreas de conhecimento da Inteligência Artificial (IA), que inclui entre elas, em ordem alfabética, algoritmos de busca, lógica matemática, lógica nebulosa, processamento de linguagem natural, ontologias, processamento de linguagem natural, raciocínio baseado em casos, raciocínio probabilístico, robótica, sistemas multiagentes, vida artificial e visão computacional.

Por ser fortemente dirigida a dados, AM é também uma subárea de conhecimento da Ciência de Dados (CD) (FLACH, 2012). Na CD, algoritmos de AM exercem um papel importante na construção de modelos capazes de extrair conhecimento de conjuntos de dados.

Este artigo está organizado da seguinte forma. A próxima seção apresenta alguns conceitos e aspectos relevantes de CD, área em que AM tem sido mais empregado. A Seção 3, após descrever formalmente como ocorre o processo de aprendizado, discorre sobre a experiência, representada por dados, necessária para a resolução de problemas, que podem ser descritos por tarefas, e como o sucesso na resolução pode ser avaliado pelo desempenho da solução proposta, o modelo induzido.

2. Ciência de Dados

Como fruto do terceiro paradigma, foram criados grupos de pesquisa na área espalhados por todo o mundo, e, em 1953, foi criado o primeiro curso de graduação em Computação, pela Universidade de Cambridge, no Reino Unido. Dezesesseis anos depois, em 1969, foi criado o primeiro curso de graduação em Computação no Brasil. Desde então, a comunidade de Computação cresceu e abriu várias frentes, entre elas a IA.

A CD é resultado do crescente uso da computação para apoiar o uso de técnicas estatísticas para a análise de dados. O termo foi usado pela primeira vez em 1974, pelo cientista da computação dinamarquês Peter Naur, ganhador do ACM A. M. Turing Award de 2005, equivalente ao prêmio Nobel para a área da Computação, com seu livro *Concise Survey of Computer Methods* (NAUR, 1974). Desde então, a CD se estabeleceu como uma importante área de conhecimento, com seus avanços reportados em conferências e periódicos dedicados ao tema.

O amadurecimento da área de CD impulsionou várias subáreas relacionadas à análise de dados, que incluem amostragem, tratamento, transformação, pré-processamento e modelagem. A subárea de modelagem investiga a indução de modelos, principalmente por meio de algoritmos de AM.

Avanços tecnológicos nos dispositivos utilizados para coleta de dados, processamento de dados, armazenamento de dados e transmissão de dados estão por trás do fenômeno denominado *Big Data*, inicialmente definido por três Vs: Volume de dados, Variedade das fontes e dos formatos dos dados e Velocidade de geração de dados. Isso promoveu o crescimento da área de CD. Um avanço tão forte que a crescente demanda por profissionais da área de Ciência de Dados motivou o surgimento, a partir de 2013, de centenas de cursos de graduação em CD no exterior, vários oferecidos por renomadas universidades. No Brasil, os primeiros cursos foram criados em 2020, menos da metade do tempo entre a criação do primeiro curso de Computação no exterior e no Brasil.

Assim como o paradigma computacional proveu ferramentas e apoiou o desenvolvimento de várias outras áreas do conhecimento, a CD, com sua busca de conhecimento novo e relevante em conjuntos de dados, tornou-se uma colaboradora das pesquisas e aplicações em quase todas as áreas, promovendo a abertura de novas frentes de pesquisa, o desenvolvimento de novos e sofisticados algoritmos, executados em máquinas computacionais de alto desempenho, e permitindo trabalhar com grandes volumes de dados, fomentando a eScience.

A forte inserção dos sistemas computacionais que fazem intenso uso de dados, como é o caso da CD, nas várias áreas do conhecimento foi considerada já em 2007 pelo cientista da computação americano James Nicholas Gray (Jim Gray), que ganhou o ACM A. M. Turing Award em 1998, o surgimento de um quarto paradigma da ciência (HEY, 2009).

Segundo ele, o quarto paradigma “tira benefício da grande quantidade de dados disponível para pesquisas e estende nossa capacidade de interrogar, entender e manipular os mundos físico e virtual”.

3. Tarefas de Aprendizado de Máquina

Uma definição formal de como um algoritmo aprende é dada por Tom Mitchell. De acordo com Mitchell, “um algoritmo, ou programa de computador, é dito **aprender** a partir de uma experiência *E*

com respeito a alguma classe de tarefas T e medida de desempenho P , se seu desempenho em tarefas de T , medido por P , melhora com a experiência E^n (MITCHELL, 1997).

Em AM, uma experiência é representada por um conjunto de dados, em que cada dado é um exemplo de experiência passada, sendo por isso chamado de exemplo, mas também conhecido por instância ou objeto. Dependendo da forma como cada objeto é representado, o conjunto de dados é estruturado ou não estruturado.

Conjuntos de dados estruturados são aqueles em que os objetos são organizados em tabelas atributo-valor, em que cada coluna representa um atributo e tem, idealmente, um valor para cada objeto, como a ilustrada pela Tabela 1, que mostra um conjunto de dados formado por 6 objetos, as 6 linhas da tabela, cada um deles uma fruta, e 5 atributos.

As 4 primeiras colunas da tabela, que descrevem diferentes aspectos de cada fruta, são chamadas de atributos preditivos (covariáveis, valores de entrada). A última coluna, que rotula qual é a fruta, é chamada de atributo alvo (valor de saída). Como cada objeto nesse conjunto de dados tem um rótulo associado – a identificação da fruta –, a tabela representa um conjunto de dados rotulado. Aqui, para destacar que a coluna fruta apresenta o atributo alvo, seus valores estão em negrito. Caso os dados não tivessem rótulo e, conseqüentemente, a tabela que os descreve não tivesse a coluna para o atributo alvo, ela representaria um conjunto de dados não rotulado.

TABELA 1. Conjunto de dados estruturados

Peso	Largura	Altura	Score de cor	Fruta
192	8,4	7,3	0,55	Maçã
144	6,8	7,4	0,75	Laranja
166	6,9	7,3	0,93	Maçã
162	7,5	7,1	0,83	Maçã
160	7,0	7,4	0,81	Laranja
160	6,9	7,3	0,79	Laranja

Um conjunto de dados não estruturado é aquele que não representa os exemplos por tabelas atributo-valor. Nesse caso, os exemplos podem ser representados por imagens, sons, vídeos, sequências biológicas, entre os mais comuns. Assim, um conjunto de dados não estruturado relacionado ao conjunto de dados estruturado da Tabela 1 poderia ser a imagem de cada fruta, e o rótulo, o nome da fruta.

Os problemas a serem resolvidos por AM podem ser vistos como tarefas a serem realizadas. As principais tarefas nos quais os problemas se enquadram são descritivas, preditivas ou prescritivas.

Nas tarefas descritivas, algoritmos de AM são aplicados a um conjunto de dados para descrever aspectos presentes nos dados. A tarefa descritiva mais comum é a de agrupamento de dados, que busca organizar os objetos presentes em um conjunto de dados em grupos de objetos semelhantes. Outro exemplo de tarefa descritiva é descobrir itens frequentes, que são itens que aparecem juntos muitas vezes em uma transação. Um exemplo é que itens são comprados em conjunto por clientes de um supermercado, em que cada compra é uma transação.

Os problemas associados a tarefas preditivas são resolvidos induzindo, e gerando, modelos preditivos a partir de um conjunto de dados rotulados. As principais tarefas preditivas são de classificação, quando o modelo preditivo deve ser capaz de prever corretamente a categoria ou classe de um objeto, por exemplo, para verificar se um paciente está saudável ou doente, e problemas de regressão, quando o modelo preditivo deve prever um valor numérico, em geral um valor real, como qual o valor justo para a venda de um imóvel.

Nas tarefas prescritivas, o problema a ser resolvido segue o sentido inverso das tarefas preditivas, tendo o seguinte formato: tendo em vista o que eu quero que meu modelo dê como resposta, valor do atributo alvo ou um dado valor desejado, que valores eu devo usar para os atributos preditivos. Um exemplo de tarefa prescritiva é a produção de novos materiais que tenham um ou mais propriedades. Por exemplo, que elementos químicos devem ser combinados para produzir um plástico

difícil de quebrar ao cair no chão e de derreter ao entrar em contato com o fogo.

O desempenho dos modelos induzidos em geral está relacionado a quão bem um algoritmo de AM que os induz desempenha sua tarefa. O quão relevantes são os grupos encontrados por um algoritmo de agrupamento, qual a proporção em que um modelo preditivo de classificação acerta a classe correta para um conjunto de valores dos atributos preditivos e quão bons foram os valores de atributos preditivos escolhidos para que o atributo alvo tenha o valor desejado.

No entanto, em várias aplicações, também é importante medir o desempenho de outras formas, como:

- **Custo computacional (tempo)**: quanto tempo é necessário para induzir um modelo e/ou para aplicar um modelo previamente induzido a novos dados. Em várias aplicações, o tempo é uma medida crítica. Exemplo dessas situações são o tempo necessário para que um algoritmo aprenda a defender uma rede de computadores de um ataque cibernético e o tempo para que o modelo recomende uma manobra em uma aeronave. Nesses casos, é inclusive admitido uma piora no desempenho preditivo na realização da tarefa para não ultrapassar o limite de tempo permitido.
- **Custo computacional (memória)**: qual o espaço de memória necessário para armazenar um modelo gerado por um algoritmo de AM. Muitas vezes o modelo precisa caber em um espaço restrito, como na memória de um celular.
- **Interpretabilidade**: avalia a facilidade de entender os mecanismos por trás do funcionamento de modelos e como eles tomam decisões. Com a Lei Geral de Proteção aos Dados (LGPD), em que as pessoas têm direito de entender como uma decisão que a afeta foi tomada, esta forma de avaliação tem ganhado espaço. Um exemplo clássico é o direito que uma pessoa que se candidatou para um crédito financeiro, ao ter seu pedido negado, tem de saber a razão para tal.

A avaliação por apenas uma das formas muitas vezes não é suficiente. Por isso, com frequência, são avaliados o desempenho e a qualidade de soluções de AM utilizando duas ou mais dessas formas.

Um exemplo de aprendizado por AM para resolver um problema associado a uma tarefa seria a busca por uma solução de um problema de diagnóstico médico. A busca da solução é uma tarefa preditiva por ter como meta prever o diagnóstico de novos pacientes a partir dos resultados de seus exames clínicos. Mais ainda, é uma tarefa de classificação, que pode ter duas classes: paciente saudável ou doente.

A experiência utilizada para o aprendizado pode ser representada por um conjunto de dados médicos de um grupo de pacientes, em que cada objeto corresponde a um paciente. O estado de saúde de cada paciente pode ser representado por resultados de um conjunto de exames clínicos, cada exame clínico sendo, portanto, um atributo preditivo. Assumindo que o conjunto de dados possui diagnóstico (rótulo), atributo alvo, para cada paciente, o conjunto de dados é rotulado.

A aplicação de um algoritmo de AM a esses dados deve gerar um modelo preditivo capaz de associar cada conjunto de resultados dos exames a um diagnóstico. Esse modelo busca aprender o conhecimento utilizado pelo médico para prever o diagnóstico de novos pacientes depois. A qualidade do modelo gerado pode ser avaliada tanto pelo desempenho preditivo do modelo, ou seja, quantos diagnósticos para novos pacientes ele acerta, quanto pela sua interpretabilidade, que permite ao médico validar como o modelo chegou a seu diagnóstico.

4. Algoritmos de Aprendizado de Máquina

O aprendizado ocorre quando um algoritmo de AM é aplicado a um conjunto de dados. A princípio, algoritmos de AM não diferem dos algoritmos escritos em uma linguagem de programação para resolver dada tarefa por meio de uma sequência de passos. Mas existe uma clara diferença entre eles: em um algoritmo convencional, esta sequência de passos tem por meta resolver de forma clara um problema específico,

como calcular a raiz quadrada de um valor numérico, sem que haja um processo de aprendizado. A sequência de passos em um algoritmo de AM deve aprender um modelo matemático a partir de um conjunto de dados. Espera-se que o modelo represente o conhecimento desses dados e que possa ser utilizado para realizar a tarefa almejada quando aplicado a novos dados.

Dezenas de milhares de algoritmos têm sido propostos, cada um utilizando heurísticas, regras e suposições específicas que definem como ajustar os parâmetros de seus modelos para resolver um conjunto de problemas. Esses algoritmos podem ser organizados de acordo com a abordagem ou paradigma adotado. Os quatro principais paradigmas são:

- **Supervisionado:** busca um modelo capaz de associar a cada grupo de valores dos atributos preditivos (valores de entrada) de um objeto em um conjunto de dados o valor correto para o correspondente atributo alvo (valor de saída). O termo supervisionado se deve à simulação de um supervisor, ou professor, que, durante o processo de aprendizado, apresenta a saída desejada. É importante observar que em algumas aplicações tem-se mais de um atributo alvo, mas isso é tema para um próximo artigo. O aprendizado supervisionado é quase sempre usado para tarefas preditivas.
- **Não supervisionado:** não utiliza o valor do atributo alvo, não contando assim com um supervisor ou professor. Em geral, busca organizar os dados utilizando apenas os valores dos atributos preditivos. O aprendizado não supervisionado é quase sempre usado para tarefas descritivas.
- **Semi-supervisionado:** utilizado quando um conjunto de dados é parcialmente rotulado, ou seja, apenas um subconjunto dos objetos possui um valor para o atributo alvo. Isso geralmente ocorre quando o custo de rotular um objeto é elevado.
- **Por reforço:** os algoritmos aprendem observando o resultado de ações tomadas, reforçando os resultados desejados por um meca-

nismo de recompensa e penalizando os resultados indesejados por um mecanismo de punição. Com isso, o aprendizado ocorre por tentativa e erro.

Todo algoritmo de AM tem um viés indutivo, uma ideia preconcebida da melhor maneira de resolver um problema. Embora possa parecer uma característica negativa, o viés é necessário para que ocorra o aprendizado. Ele restringe o espaço de busca do algoritmo por uma boa solução. Do contrário, essa busca poderia ter uma duração muito longa.

Os principais vieses dos algoritmos de AM são de busca, que definem que tipos de modelos serão inicialmente testados pelo algoritmo e como novos modelos serão explorados, e de representação, que definem o formato que os modelos devem assumir.

Como cada algoritmo de AM tem seu viés indutivo, cada um deles vai se adequar melhor a um determinado tipo de problema ou distribuição de dados. Assim, não existe um único algoritmo que seja sempre melhor que os demais (Wolpert, 1996; Pardalos, 2021), pois o desempenho dos algoritmos depende de quão bem eles combinam com as propriedades dos dados utilizados.

Algoritmos de AM possuem hiperparâmetros, cujos valores são definidos por seu usuário, que afetam o seu funcionamento e restringem ainda mais o espaço de busca dos modelos, tornando a tarefa computacionalmente viável. Quando aplicados a um conjunto de dados, com dada combinação de valores para seus hiperparâmetros, os algoritmos, por sua vez, ajustam valores dos parâmetros de um modelo para encontrar aqueles que levem a uma melhor estimativa de desempenho futuro, quando o modelo for aplicado a novos dados não vistos durante o aprendizado.

O resultado obtido na resolução de um problema utilizando AM depende ainda de várias decisões que extrapolam a aplicação de um algoritmo de AM. É necessário selecionar e aplicar técnicas específicas para identificar e tratar problemas relacionados à qualidade dos dados. Para citar os passos mais comuns, é preciso transformar os dados para tipos

e formatos adequados para o algoritmo a ser utilizado, pré-processar os dados de forma a resolver problemas como a maldição da dimensionalidade, lidar com dados desbalanceados e validar os modelos gerados. A sequência de passos necessários para uma boa solução de AM deu origem ao termo AM de ponta-a-ponta.

Apesar do AM levar à automatização de várias atividades nas mais diversas áreas de conhecimento, a boa utilização do AM, eficaz e eficiente, ainda está restrita a especialistas, que – importante lembrar – não são formados na mesma velocidade em que cresce a demanda. Isso restringe o desenvolvimento de soluções baseadas em AM a quem tem recursos para contratar serviços desses profissionais. Um bom profissional utiliza as técnicas e os algoritmos mais adequados, e de forma correta. Um modo de minimizar esse problema, democratizando o acesso ao AM, é automatizar o uso eficiente e eficaz de AM, o que pode ser feito usando o próprio AM. Essa alternativa é a mola propulsora de uma subárea de pesquisa denominada Aprendizado de Máquina Automatizado (AutoML).

5. Aprendizado de Máquina Automatizado

Para o desenvolvimento de uma boa solução de AM, é necessário fazer várias escolhas, que devem estar bem fundamentadas. Adicionalmente, uma parte considerável é manual, força bruta, cansativa, exige atenção e é repetitiva. As mesmas razões que motivam o uso de AM motivam a automatização de seu uso.

Várias abordagens para AutoML têm sido propostas nos últimos anos, com grandes empresas da área de tecnologia da informação criando grupos de pesquisa para desenvolver suas próprias ferramentas (HE, 2021). Elas podem ser divididas em: baseadas em otimização, baseadas em meta-aprendizado e a combinação das duas.

As abordagens baseadas em otimização são as mais comuns e utilizam técnicas de otimização para definir *pipelines*, otimizar modelos, selecionar atributos preditivos e ajustar hiperparâmetros (Hutter, 2019; Mantovani, 2019). Mais recentemente, a otimização tem sido utilizada

para construir novos algoritmos de AM, combinando partes de algoritmos existentes em menos de um segundo, em vez de meses ou anos.

As baseadas em meta-aprendizado usam algoritmos de AM para induzir modelos preditivos capazes de recomendar uma ou um subconjunto de alternativas dentre um conjunto de alternativas disponíveis (Vilalta, 2002, 2005; Brazdil, 2021). Meta-aprendizado investiga o desenvolvimento de sistemas de recomendação baseados em AM capazes de recomendar os algoritmos e/ou as técnicas mais adequadas para uma dada aplicação. Na maioria das vezes, estuda como os algoritmos de Aprendizado de Máquina podem aumentar sua eficiência por meio da experiência de aprendizado.

Meta-aprendizado difere do AM convencional, chamado aqui de aprendizado no nível base, no escopo de seu nível de aprendizado. Enquanto o aprendizado no nível base trabalha em um conjunto de dados por vez, o aprendizado no meta-nível é pautado no acúmulo de experiência do desempenho obtido em múltiplas aplicações de um ou mais algoritmos de AM. De maneira geral, é possível afirmar que o meta-aprendizado foca na relação entre domínios ou tarefas e estratégias de aprendizado. Meta-aprendizado tem sido aplicado com sucesso em vários tipos de tarefas, incluindo agrupamento de dados (Pimentel, 2019, 2020), estratégias de aprendizado ativo (Santos, 2019), técnicas de segmentação de imagens (Aguiar, 2019), algoritmos de otimização (Kanda, 2016), para citar algumas.

Considerações Finais

Cada vez mais produtos e serviços fazem uso de AM, aumentando sua presença nas nossas atividades cotidianas e subsidiando decisões que afetam nossa vida. Para evitar ou reduzir abusos, o AM deve ser utilizado de forma responsável, produzindo soluções justas, eticamente corretas e não preconceituosas. Não são poucos os casos reportados de uso indevido de AM. É importante observar que isso não ocorre apenas com AM, tem ocorrido, ocorre e ocorrerá com cada nova tecnologia.

Não é eliminando a tecnologia que os problemas desaparecerão, mas com o investimento em educação pública de qualidade, que nos tornará cada vez mais preparados para conviver bem com ela.

Referências

AGUIAR, G. J.; MANTOVANI, R. G.; MASTELINI, S. M.; DE CARVALHO, A. C. P. L. F.; BARBON JR., S. A meta-learning approach for selecting image segmentation algorithm. **Pattern Recognition Letters** 128: 480-487, 2019.

ARKIN, R. C. **Behavior-based Robotics**. 1ª edição. MIT Press: Cambridge, MA, USA. 1998.

BRAGA, A. DE P.; LUDERMIR, T. B.; DE CARVALHO, A. C. P. L. F. **Redes Neurais Artificiais: Teoria e Aplicações**. 2ª edição. LTC-GrupoGen: Rio de Janeiro. 2007.

BRAZDIL, P.; VAN RIJN, J. N.; SOARES, C.; VANSCHOREN, J. **Metalearning: Applications to Data Mining**, Series: Cognitive Technologies, Springer, 2ª edição. 2021.

DE CARVALHO, A. C. P. L. F. Inteligência Artificial: riscos, benefícios e uso responsável. **Estudos Avançados** 35(101), 21-36. 2021.

FACELI, K.; LORENA, A. C.; GAMA J.; ALMEIDA, T. A.; DE CARVALHO, A. C. P. L. F. **Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina**. 2ª edição. LTC-GrupoGen: Rio de Janeiro. 2021.

FLACH, P. **Machine Learning: The Art and Science of Algorithms that Make Sense of Data**. Cambridge University Press. 2012.

Xin He, Kaiyong Zhao, Xiaowen Chu. AutoML: A survey of the state-of-the-art, **Knowledge-Based Systems**, Volume 212, 106622, 2021.

HEY, T.; STEWART T., TOLLE, K. **The Fourth Paradigm: Data-Intensive Scientific Discovery**. Microsoft Research: Redmond, Washington. 2009.

HUTTER, F., KOTTHOFF, L.; VANSCHOREN, J. (Editors). **Automated Machine Learning**. Springer International Publishing, 2019.

KANDA, J.; DE CARVALHO, A. C. P. L. F.; HRUSCHKA, E. R.; SOARES, C.; BRAZDIL, P. Meta-learning to select the best meta-heuristic for the Traveling Salesman Problem: A comparison of meta-features. **Neurocomputing** 205: 393-406, 2016.

MANTOVANI, R. G.; ROSSI, A. L. D.; ALCOBAÇA, E.; VANSCHOREN, J.; DE CARVALHO, A. C. P. L. F. A meta-learning recommender system for hyperparameter tuning: Predicting when tuning improves SVM classifiers. **Inf. Sci.** 501: 193-221, 2019.

MISH, F. C. WEBSTER'S NINTH NEW COLLEGIATE DICTIONARY. Merriam-Webster series. Springfield, Mass. 1984.

MITCHELL, T. M. **Machine Learning**. McGrawHill: New York, 1997.

NILSSON, N. J. **Learning machines**. McGrawHill: New York. 1965.

PIMENTEL, B. A.; DE CARVALHO, A. C. P. L. F. A new data characterization for selecting clustering algorithms using meta-learning. **Inf. Sci.** 477: 203-219, 2019.

PIMENTEL, B. A.; DE CARVALHO, A. C. P. L. F. A Meta-learning approach for recommending the number of clusters for clustering algorithms. **Knowl. Based Syst.** 195: 105682, 2020.

SAMUEL, A. L. Some Studies in Machine Learning Using the Game of Checkers. **IBM Journal of Research and Development**, vol. 3, no. 3, pp. 210-229, July 1959.

SANTOS, D. P.; PRUDÊNCIO, R. B. C.; DE CARVALHO, A. C. P. L. F. Empirical investigation of active learning strategies. **Neurocomputing** 326-327: 15-27, 2019.

SIMON, H. A. "Why Should Machines Learn?" *In* Michalski, R.S.; Carbonell, J.G.; Mitchell, T. M. (Editors). **Machine Learning, Symbolic Computation**. Springer: Berlin, Heidelberg. 1983.

TANIMOTO, S. L. **The elements of artificial intelligence using Common Lisp**. Computer Science Press, Inc.: USA. 1990.

VILALTA R.; DRISSI, Y. (2002). A perspective view and survey of metalearning. **Artificial Intelligence Review** 18(2) 77-95.2002.

VILALTA R.; GIRAUD-CARRIER C.; BRAZDIL P. "Meta-Learning: Concepts and Techniques." *In* Maimon, O.; Rokach, L. (Editors). **Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers**. Springer Publishers. 2005.

WOLPERT, D. H. The lack of a priori distinctions between learning algorithms. **Neural Computation**, 8(7) 1341-1390. 1996.

André Carlos Ponce de Leon de Carvalho · Professor Titular do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (ICMC-USP), campus São Carlos, Bolsista de Produtividade em Pesquisa 1A do CNPq e Vice-Presidente da Sociedade Brasileira de Computação (SBC). É membro do Comitê de Governança da Estratégia Brasileira de Inteligência Artificial, do Comitê Diretivo para América Latina e Caribe da International Network for Government Science Advice (INGSA) e do Strategy and Partnerships Board of the UKRI Centre for Doctoral Training in Accountable, Responsible and Transparent AI (ART-AI), University of Bath, Reino Unido, do Artificial Intelligence Technical Committee of the International Federation for Information Processing e do Te Ipu o te Mahara Artificial Intelligence Research Group, the University of Waikato, Nova Zelândia.