



Interações entre Big data e cloud computing



Algoritmos de Aprendizado de Máquina para minerar os dados

O que são algoritmos?

Bloco 1

Deivid Sardinha



➤ Algoritmos

- Instruções.
- Sequência lógica.
- Entrada, processamento e saída.

Figura 1 – Representação de solução



Fonte: rawpixel.com/Freepik.com.



► Os algoritmos de análise de dados

- C4.5.
- JRip.
- K-Médias.
- SVM (máquina de vetores de suporte).
- KNN (vizinhos K-próximos).



➤ C4.5

- O algoritmo da árvore de escolha C4.5 produz árvores que utilizam características persistentes ou ostensivas.
- O algoritmo escolhe, para cada *hub* interno, uma propriedade que melhor segmenta o subconjunto atual das informações de entrada.
- O procedimento de separação e conquista é recebido na geração do modelo, com uma determinação dos atributos subjacentes realizados de acordo com as regras do pacote. Então, a árvore salva apenas os atributos importantes para a divisão.



➤ JRip

- Algoritmo *The Repeated Incremental Pruning to Produce Error Reduction* (RIPPER) refere-se a uma variante aprimorada do algoritmo de poda incremental de erro reduzida (IREP), recebendo uma forma de poda para lidar com a diminuição de erros e produzir regras adequadas.
- Este algoritmo depende de um método de redução de informações que compreende isolar as informações de preparação em um subconjunto em desenvolvimento e um subconjunto de poda.
- No ponto em que o subconjunto em desenvolvimento faz várias diretrizes, a poda é usada para diminuir o erro de caracterização.



Algoritmos de Aprendizado de Máquina para minerar os dados

Categoria de Aprendizado de Máquina

Bloco 2

Deivid Sardinha



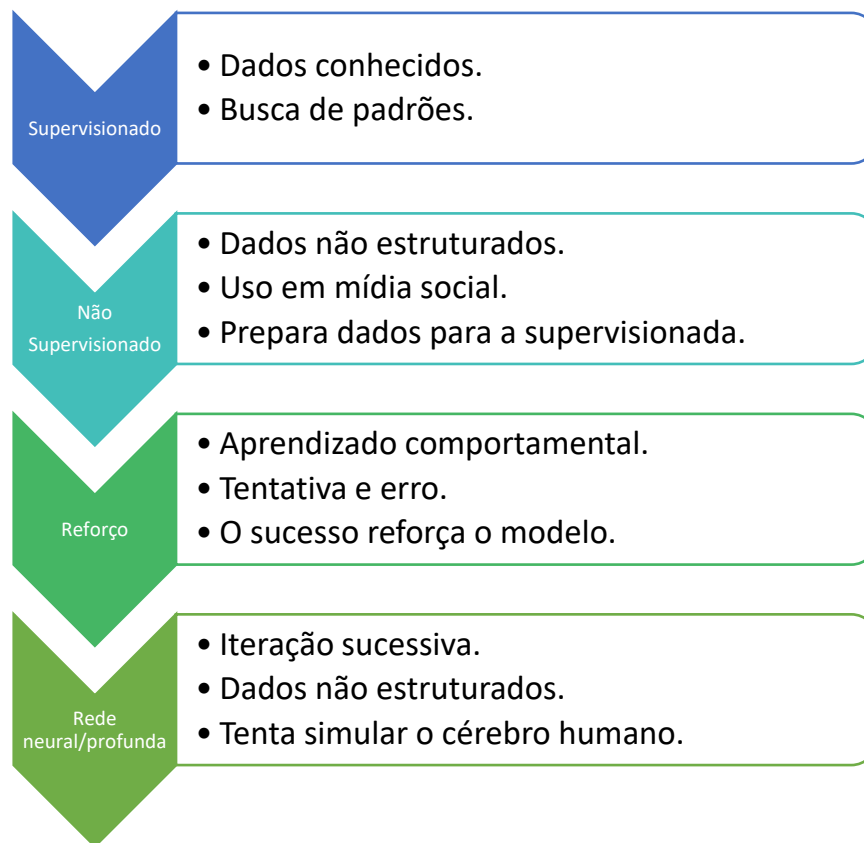
➤ Categoria de Aprendizado de Máquina

- O Aprendizado de Máquina recebe abordagens diferentes com o intuito de melhorar o desempenho dos modelos preditivos.
- Elas variam de acordo com o problema de negócios envolvido.



➤ Abordagens

Figura 2 – Abordagem de Aprendizado de Máquina



Fonte: elaborada pelo autor.



Análise na internet

- A análise de dados na web tem um componente em crescimento constituído pelas plataformas e pelos serviços de computação em nuvem (*cloud computing*).
- A mineração de links e detecção de comunidades (*cluster*) são exemplos de análise de rede. Nelas, procura-se prever ou detectar nós de uma rede.



Teoria em Prática

Bloco 3

Deivid Sardinha



➤ Reflita sobre a seguinte situação

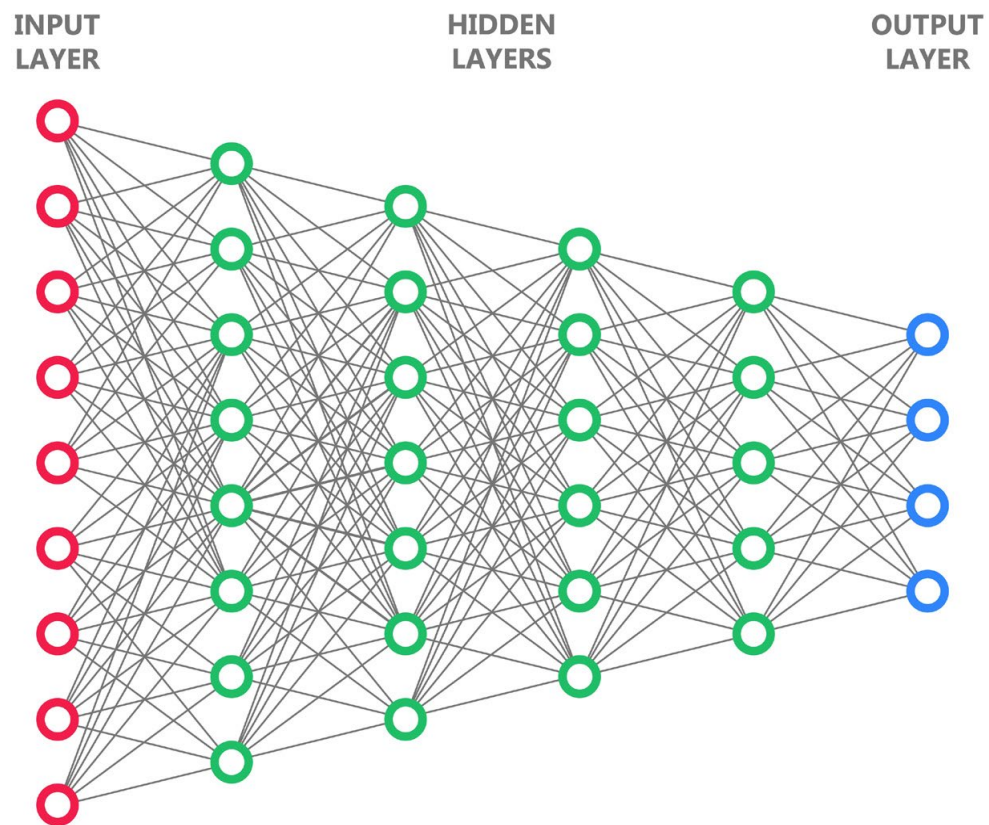
Vejamos o algoritmo Convolutional Neural Networks, conhecido como ConvNets. Esses algoritmos levaram o resultado de análise com imagem a um nível muito alto. Eles trabalham com a verificação e identificação facial sem restrições. Essa técnica está sendo estudada extensivamente nos últimos anos, pois suas aplicações práticas são enormes. Os algoritmos mais recentes e de melhor desempenho de verificação de faces representam faces com recursos extremamente completos. Modelos de aprendizado profundo, como o ConvNets, são eficazes para extrair recursos visuais de alto nível e são muito usados nas aplicações de verificação de face. Os ConvNets são ensinados para classificar as faces disponíveis pelas suas identidades. Cada ConvNet recebe um caminho facial como entrada e extrai recursos profundos das camadas inferiores. No processo, os números de recurso vão reduzindo ao longo dos níveis mais baixos de extração de recursos, enquanto, ao mesmo tempo, os recursos mais globais e de alto nível são formados nos níveis superiores. No final do último nível, as informações da identidade são ricas e diretamente preditas com inúmeras classes de identidade, podendo chegar a milhares e, assim, identificar uma face. Para mais detalhes, consulte Sun *et al.* (2014).

Diante da apresentação do ConvNet, pesquise outros exemplos de algoritmos e suas aplicações e tente criar possíveis usos. Escolha um exemplo e estude como foi possível chegar a ele.



➤ Redes neurais

Figura 3 – Redes neural de vários níveis



Fonte: all_is_magic/iStock.com.

Dicas do(a) Professor(a)

Bloco 4

Deivid Sardinha





Leitura Fundamental

Prezado aluno, as indicações a seguir podem estar disponíveis em algum dos parceiros da nossa Biblioteca Virtual (faça o login através do seu AVA). Algumas indicações também podem estar disponíveis em sites acadêmicos como o Scielo, repositórios de instituições públicas, órgãos públicos, anais de eventos científicos ou periódicos científicos, acessíveis pela internet.

Isso não significa que o protagonismo da sua jornada de autodesenvolvimento deva mudar de foco. Reconhecemos que você é a autoridade máxima da sua própria vida e deve, portanto, assumir uma postura autônoma nos estudos e na construção da sua carreira profissional.

Por isso, te convidamos a explorar todas as possibilidades da nossa Biblioteca Virtual e além! Sucesso!





Indicação de leitura 1

A análise de dados que agrupa elementos com características comuns é muito utilizada em *Big Data* para extrair conhecimento de conjuntos de dados. Essa análise é chamada de clusterização, sua aplicação é facilitada pela disponibilidade de armazenamento, processamento e alto fluxo de dados. Este artigo apresenta um método para ajudar a compreensão de *cluster* com a intenção de melhorar a compreensão e interpretação desses agrupamentos. Recomendo atenção no item 2, *Aprendizado de máquina não supervisionado*, pois algoritmos que tratam esses tipos de dados são muito utilizados.

Referência:

METZ, J.; MONARD, M. C. Clustering hierárquico: uma metodologia para auxiliar na interpretação dos clusters. In: XXV Congresso da Sociedade Brasileira de Computação. **Anais [...]**. São Leopoldo, Sociedade Brasileira de Computação, 22 a 29 de julho de 2005.





Indicação de leitura 2

O texto trata vários conceitos envolvidos com algoritmo de *Machine Learning*, fundamentos, estrutura e forma de atuação. Destina-se a pesquisar e compreender como funcionam. É sugerida especialmente a leitura que trata de algoritmo supervisionado e rede neural.

Referência:

BROWNLEE, J. A Tour of Machine Learning Algorithms. **Machine Learning Mastery**, San Juan, 12 de agosto de 2019.



► Dica do(a) Professor(a)

Os algoritmos são interessantes pois estão sempre se renovando e novos surgem a todo instante, contudo, há alguns que são conhecidos há milhares de anos. No entanto, poucos algoritmos são perfeitamente entendidos. Existem algoritmos complicados, difíceis e complicados, elegantes, simples e fáceis (SEEDGEWICK, 1983).





Referências

LAVALLE, S. *et al.* Big Data, analytics and the path from insights to value. **Analytics Magazine**, [s.l.], 2011. Disponível em: <http://analytics-magazine.org/big-data-analytics-and-the-path-frominsights-to-value/>. Acesso em: 5 abr. 2019.

MARR, B. How Jaguar Land Rover Is Getting Ready For The 4th Industrial Revolution: AI & Autonomous Cars. **Forbes**, Jersey City, 2018. Disponível em: <https://www.forbes.com/sites/bernardmarr/2018/10/26/how-jaguar-land-rover-is-getting-ready-for-the-4th-industrial-revolution-autonomous-vehicles/?sh=4fe3a2933a5e>. Acesso em: 9 abr. 2019.

NILSSON, N. J. **Introduction to machine learning**. An early draft of a proposed textbook. Stanford University Stanford, 1998. Disponível em: <https://ai.stanford.edu/~nilsson/MLBOOK.pdf>. Acesso em: 30 ago. 2019.

SEdgeWICK, R. **Algorithms**. Library of Congress Cataloging in Publication Data. Massachusetts: Addison-Wesley Publishing Company, 1983. Disponível em: <http://dsp-book.narod.ru/Algorithms.pdf>. Acesso em: 30 ago. 2019.



Bons estudos!

