

MODELAGEM DA INFORMAÇÃO E ESTRATÉGIAS METODOLÓGICAS PARA O APRENDIZADO DE MÁQUINA

Christian Jhulian Braga Quesada

Pesquisa de Mestrado “Inteligência Artificial (IA) e Arquitetura: caminhos possíveis de aplicação e melhoria no processo de projeto”

Orientadora: Profa. Dra. Anja Pratschke

04 de março de 2020

Resumo: Este artigo faz parte do material desenvolvido para o oitavo seminário de pesquisas em curso do Nomads.usp - Núcleo de Estudos sobre Habitares Interativos, cujo tema da edição é: método. Neste trabalho são abordadas algumas questões a respeito da caracterização e pré-processamento de dados para sua utilização em algoritmos de aprendizado de máquina, no intuito de estabelecer possíveis aplicações futuras para o campo de projeto em arquitetura, sobretudo no que se refere a modelagem da informação da construção (BIM). São brevemente apresentados dois paradigmas de aprendizado: supervisionado (preditivo) e não supervisionado (descritivo), além de abordar questões sobre o viés de representação e busca dos algoritmos. Também são abordadas questões de caracterização de dados (variáveis) segundo a classificação de tipos e escala.

Palavras-chave: Arquitetura, BIM, Inteligência Artificial, Aprendizado de Máquina, Métodos.

1. Introdução

A presente pesquisa de mestrado (em andamento) se trata de uma investigação sobre o estado da arte no campo da inteligência artificial (IA) e do Aprendizado de Máquina (AM), buscando relacionar possíveis aplicações para o contexto da arquitetura e urbanismo. Trata-se de uma investigação, como define David Sperling (2018), “de situações em que o pensar e o fazer arquitetura são potencializados por pares, arquitetos-aparatos tecnológicos” e não se trata, de forma alguma, da abstração de soluções projetuais sem a presença de arquitetos, resultantes apenas de ações realizadas por computadores autônomos.

Neste breve artigo, serão abordadas algumas questões a respeito da caracterização e pré-processamento de dados para utilização em algoritmos de aprendizado de máquina, procurando vislumbrar possíveis aplicações futuras para o campo de projeto em arquitetura, sobretudo, no que diz respeito à modelagem da informação da construção (BIM).

2. Sobre a Modelagem da Informação da Construção (BIM)

De acordo com Chuck Eastman et al. (2014), podemos definir a Modelagem da Informação da Construção (BIM - Building Information Modeling) “como uma tecnologia de modelagem e um conjunto associado de processos para produzir, comunicar e analisar modelos de construção” (p. 13), e que os Modelos de Construção são caracterizados por:

- **Componentes de construção**, que são representados com representações digitais inteligentes (objetos) que "sabem" o que eles são, e que podem ser associados com atributos (gráficos e de dados) computáveis e regras paramétricas.
- **Componentes que incluem dados que descrevem como eles se comportam**, conforme são necessários para análises e processos de trabalho, por exemplo, quantificação, especificação e análise energética.
- **Dados consistentes e não redundantes** de forma que as modificações nos dados dos componentes sejam representadas em todas as visualizações dos componentes.
- **Dados coordenados** de forma que todas as visualizações de um modelo sejam representadas de maneira coordenada.

Mais adiante, os autores do “Manual de BIM: um guia de modelagem da informação da construção para arquitetos, engenheiros, gerentes, construtores e incorporadores” (2014), comentam sobre como a tecnologia BIM poderá evoluir e quais serão os prováveis impactos no futuro da indústria da Arquitetura, Engenharia e Construção (AEC), e na sociedade como um todo. Segundo Eastman et al. (2014), as tendências tecnológicas incluem o desenvolvimento de verificações de conformidade e construtibilidade automatizadas, utilizando os modelos dos edifícios e o uso potencial de ferramentas de inteligência artificial:

O progresso tecnológico no poder computacional, as tecnologias de sensoriamento remoto, maquinário de produção controlado por computador, computação distribuída, tecnologias de troca de informação e outras vão abrir novas possibilidades que os desenvolvedores de softwares explorarão para suas próprias vantagens competitivas. Outra área técnica que pode introduzir desenvolvimentos adicionais que influenciam sistemas BIM é aquela a que se faz referência como inteligência artificial. Ferramentas BIM são plataformas convenientes para o renascimento do desenvolvimento de sistemas especialistas para uma gama de propósitos, como verificação de conformidade com códigos, revisões qualitativas, ferramentas inteligentes para comparação de versões, guias e tutoriais de projetos. (EASTMAN et al., 2014, p. 304-5)

3. Sobre o Aprendizado de Máquina (AM)

De acordo com a professora e pesquisadora Katti Faceli et al. (2019), nos últimos anos, com a crescente complexidade dos problemas a serem tratados computacionalmente e do enorme volume de dados gerados por diferentes setores, tornou-se evidente a necessidade de ferramentas computacionais mais sofisticadas, que fossem mais autônomas, que reduzissem a necessidade de intervenção humana e a dependência de especialistas. Dessa forma, os

computadores vêm sendo programados para aprender com a experiência passada, e para tal, empregam princípios de inferência como a indução, no qual se obtêm conclusões genéricas a partir de um conjunto particular de exemplos (conjunto de treinamento). Assim, os algoritmos de aprendizado de máquina aprendem a induzir uma função ou hipótese capaz de resolver um problema a partir dos dados que representam instâncias do problema a ser resolvido. A essa propriedade de uma hipótese continuar a ser válida para novos objetos dá-se o nome de capacidade de generalização de hipótese. Ou seja, para ser útil quando aplicada a novos dados, uma hipótese precisa apresentar boa capacidade de generalização. A esse processo de indução de uma hipótese (ou aproximação de função) a partir da experiência passada dá-se o nome de Aprendizado de Máquina (AM). (FACELI et al., 2019, p. 2-3)

Ainda segundo Faceli et al. (2019), um requisito importante para algoritmos de AM é que eles sejam capazes de lidar com dados imperfeitos, pois muitos conjuntos de dados apresentam algum tipo de problema, como presença de ruídos, dados inconsistentes, dados ausentes e dados redundantes, e os algoritmos de AM devem, idealmente, ser robustos aos problemas presentes nesses dados, minimizando sua influência no processo de indução de hipóteses. Entretanto, dependendo de sua extensão, esses problemas podem prejudicar o processo indutivo. Dessa forma, técnicas de pré-processamento são utilizadas com frequência para identificar e minimizar a ocorrência desses problemas. (FACELI et al., 2019, p. 4)

Técnicas de pré-processamento de dados são frequentemente utilizadas para melhorar a qualidade dos dados por meio da eliminação ou minimização dos problemas citados. Essa melhora pode facilitar o uso de técnicas de AM [Aprendizado de Máquina], levar à construção de modelos mais fiéis à distribuição real dos dados, reduzindo sua complexidade computacional, tornar mais fáceis e rápidos o ajuste de parâmetros do modelo e seu posterior uso. Isso pode, adicionalmente, facilitar a interpretação dos padrões extraídos pelo modelo. [...] Técnicas de pré-processamento de dados são úteis não apenas porque podem minimizar ou eliminar problemas existentes em um conjunto de dados, mas também porque podem tornar os dados mais adequados para sua utilização por um determinado algoritmo de AM. Por exemplo, alguns algoritmos de AM trabalham apenas com valores numéricos. (FACELI et al., 2019, p. 29)

4. Sobre os métodos de aprendizagem de máquina

Segundo Faceli et al. (2019), quando um algoritmo de aprendizado de máquina está aprendendo a partir de um conjunto de dados de treinamento, ele está procurando uma hipótese, no espaço de possíveis hipóteses, capaz de descrever as relações entre os objetos que melhor se ajuste aos dados de treinamento. Dessa forma, cada algoritmo de AM utiliza uma forma ou uma representação para descrever a hipótese induzida. Por exemplo, redes neurais artificiais representam uma hipótese por um conjunto de valores reais, associado aos pesos das conexões da rede, enquanto as árvores de decisão utilizam uma estrutura de árvore em que

cada nó interno é representado por uma pergunta referente ao valor de um atributo e cada nó externo está associado a uma classe. Dessa forma, a representação utilizada pode definir a preferência ou viés de representação do algoritmo, podendo restringir o conjunto de hipóteses que podem ser induzidas. Entretanto, além do viés de representação, os algoritmos de AM possuem também um viés de busca. O viés de busca de um algoritmo é a forma como o algoritmo busca a hipótese que melhor se ajusta aos dados de treinamento, ou seja, ele define como as hipóteses são pesquisadas no espaço de hipóteses. Por exemplo, alguns algoritmos que se utilizam da representação por árvores de decisão, podem ter como viés de busca a preferência por árvores de decisão com poucos nós. (FACELI et al., p. 5)

[...] cada algoritmo de AM possui dois vieses, um viés de representação e um viés de busca. O viés é necessário para restringir as hipóteses a serem visitadas no espaço de busca. Sem viés não haveria aprendizado/generalização. Os modelos seriam especializados para os exemplos individuais. Embora, à primeira vista, o viés pareça ser uma limitação dos algoritmos de AM, segundo Mitchell (1997), sem viés um algoritmo de AM não consegue generalizar o conhecimento adquirido durante seu treinamento para aplicá-lo com sucesso a novos dados. (FACELI et al., 2019, p. 5)

Em seguida, quanto ao paradigma de aprendizado a ser adotado para lidar com uma tarefa, podemos classificar em dois tipos de modelos: o supervisionado (preditivo); e o não supervisionado (descritivo). As tarefas supervisionadas se distinguem pelo tipo de rótulos dos dados: discreto, no caso de classificação; e contínuo, no caso de regressão. As tarefas descritivas são genericamente divididas em: agrupamento, em que os dados são agrupados de acordo com sua similaridade; sumarização, cujo objetivo é encontrar uma descrição simples e compacta para um conjunto de dados; e associação, que consiste em encontrar padrões frequentes de associações entre os atributos de um conjunto de dados. No entanto, é importante observarmos que, apesar dessa divisão básica de modelos em preditivos e descritivos, um modelo preditivo também provê uma descrição compacta de um conjunto de dados e um modelo descritivo pode prover também previsões após ser validado. (FACELI et al., 2019, p. 7)

Em tarefas de previsão, a meta é encontrar uma função (também chamada de modelo ou hipótese) a partir dos dados de treinamento que possa ser utilizada para prever um rótulo ou valor que caracterize um novo exemplo, com base nos valores dos seus atributos de entrada. Para isso, cada objeto de conjunto de treinamento deve possuir atributos de entrada e saída. [...] Os algoritmos ou métodos de AM utilizados nessa tarefa induzem modelos preditivos. Esses algoritmos seguem o paradigma de aprendizado supervisionado. O termo supervisionado vem da simulação da presença de um “supervisor externo”, que conhece a saída (rótulo) desejada para cada exemplo (conjunto de valores para os atributos de entrada). Com isso, o supervisor externo pode avaliar a capacidade da hipótese induzida de prever o valor de saída para novos exemplos. [...] Em tarefas de descrição, a meta é explorar ou descrever um conjunto de dados. Os algoritmos de AM utilizados nessas tarefas não fazem uso do atributo de saída. Por isso, seguem o paradigma de aprendizado não supervisionado. Uma tarefa descritiva de agrupamento de dados, por exemplo, tem por meta encontrar grupos de objetos semelhantes no conjunto de dados. Outra tarefa descritiva é encontrar regras

de associação que relacionam um grupo de atributos a outro grupo de atributos. (FACELI et al., 2019, p. 6)

Por fim, cabe ressaltar que existem outras metodologias de aprendizado de máquina além das citadas neste trabalho, como por exemplo: o aprendizado semissupervisionado, o aprendizado ativo e aprendizado por reforço.

5. Sobre os métodos de caracterização de dados

Segundo Faceli et al. (2019), apesar do crescente número de bases de dados disponíveis, na maioria das vezes não é possível utilizar algoritmos de AM diretamente sobre esses dados, sendo que técnicas de pré-processamento são frequentemente utilizadas para tornar os conjuntos de dados mais adequados para o uso de algoritmos de AM. Vários algoritmos têm dificuldades em utilizar os dados no seu formato original, e para tratar desse problema, são realizadas transformações nos dados originais antes que eles sejam utilizados pelo algoritmo. Um exemplo dessa transformação é a transformação de valores simbólicos em valores numéricos. (FACELI et al., 2019, p. 10-11)

A análise das características presentes em um conjunto de dados permite a descoberta de padrões e tendências que podem fornecer informações valiosas que ajudem a compreender o processo que gerou os dados. Muitas dessas características podem ser obtidas por meio da aplicação de fórmulas estatísticas simples. Outras podem ser observadas por meio do uso de técnicas de visualização. (FACELI et al., 2019, p. 12)

De acordo com o pesquisador Júlio Cesar Pereira (2004), existe uma distinção básica na natureza das variáveis (conjunto de dados), sendo que algumas delas são: qualitativas (de alocação numa categoria) e outras quantitativas (de posição numa escala), o que remete à diferença entre medidas derivadas e fundamentais, permitindo assim, que as variáveis sejam classificadas em quatro tipos básicos: **quantitativo discreto** (números inteiros, sem frações, como em contagens, constituindo um conjunto finito de dados); **quantitativo contínuo** (números que podem assumir valores fracionários, com intervalos de valores conhecidos, mas com um conjunto infinito de valores possíveis); **qualitativo categórico nominal** (categorias, sendo que cada categoria é independente, sem relação com as outras); e **qualitativo categórico ordinal** (categorias, sendo que cada categoria mantém uma relação de ordem com as outras).

Segundo Faceli et al. (2019), o tipo de um atributo diz respeito ao grau de quantização nos dados, e a escala indica a significância relativa dos valores. Dessa forma, a escala define as operações que podem ser realizadas sobre os valores do atributo, e podem ser classificados como: nominais, ordinais, intervalares e racionais (ou proporcionais). Os dois primeiros são do tipo qualitativo e os dois últimos são do tipo quantitativo. (FACELI et al., 2019, p. 13-15)

Conhecer o tipo/escala dos atributos auxilia a identificar a forma adequada de preparar os dados e posteriormente modelá-los. [...] O tipo define se o atributo representa quantidades, sendo então denominado quantitativo ou numérico, ou qualidades, quando é chamado de qualitativo, simbólico ou categórico, pois os valores podem ser associados a categorias. (FACELI et al., 2019, p. 13)

É importante observar também que uma medida quantitativa possui, além do valor numérico, uma unidade, como por exemplo: metro, quilograma ou graus. No processo de extração de conhecimento, se um atributo (dado ou variável) tiver valor 100, o valor em si não diz nada, não sabemos se é uma medida em centímetros, metros ou jardas, por exemplo, sendo que essa informação é importante na avaliação do conhecimento adquirido. Sendo assim, os atributos quantitativos ou numéricos podem assumir valores binários, inteiros ou reais, e por outro lado, atributos qualitativos são, geralmente, representados por um número finito de símbolos ou nomes. (FACELI et al., 2019, p. 14)

6. Considerações Finais

De acordo com Júlio Cesar Pereira (2004), a escolha do tipo de variável a ser utilizado pelo pesquisador dependerá da natureza de seu objeto de estudo e dos recursos de que dispõe para mensuração. Por exemplo, um evento qualitativo que comporta medida quantitativa contínua é a cor: embora seja frequentemente medida por variável categórica nominal (branco, preto, azul etc.), ela comporta mensurações contínuas para matiz e saturação, que podem ser realizadas desde que o pesquisador disponha de equipamento necessário. (PEREIRA, 2019, p. 44-45). Ou seja, o tipo de variável utilizada pelo pesquisador pode condicionar as oportunidades de análise:

A mensuração de um objeto de estudo pode realizar-se com dados de uma ou mais dimensões, que sejam inerentes à sua própria natureza (medidas fundamentais) ou de projeção de sua imagem (medidas derivadas). Sua descrição, análise ou comparação com outro objeto poderão ser feitas por medidas originais ou por processamento de combinação sintética delas. [...] A assunção de premissas (postulados, proposições, axiomas) é inerente à estrutura do conhecimento científico e condiciona o entendimento de objetividade e verdade, ou seja, de representação do real. Repensar essas premissas deve auxiliar o pesquisador a conceber suas estratégias de investigação. A distinção entre ser e atributo, de que se ocupa a filosofia, constitui-se em elemento importante para que o pesquisador defina a delimitação de seu objeto de estudo. (PEREIRA, 2004, p. 40)

Dessa forma, antes de aplicarmos algoritmos de AM a um conjunto de dados (seja ele qual for) é importante que esses dados sejam analisados e pré-processados. Essa análise e pré-processamento, pode ser realizada por diversas técnicas, e permite uma melhor compreensão da distribuição dos dados e pode dar suporte à escolha de formas de abordagem do problema. (FACELI et al., 2019, p. 28)

REFERÊNCIAS BIBLIOGRÁFICAS

EASTMAN, Chuck; TELCHOLZ, Paul; SACK, Rafael; LISTON, Hathleen. **Manual de BIM: um guia de modelagem da informação da construção para arquitetos, engenheiros, gerentes, construtores e incorporadores.** [tradução: Cervantes Gonçalves Ayres Filho... et al.]; Revisão técnica: Eduardo Toledo Santos. - Dados eletrônicos. Porto Alegre: Bookman, 2014.

FACELI, Katti; LORENA, Ana Carolina; GAMA, João; CARVALHO, André Carlos Ponce de Leon Ferreira de. **Inteligência artificial: uma abordagem de aprendizado de máquina.** Reimpr. – Rio de Janeiro: LTC, 2019.

PEREIRA, Júlio Cesar Rodrigues. **Análise de dados qualitativos: estratégias metodológicas para as Ciências da Saúde, Humanas e Sociais.** 3 ed. 1. Reimpr. – São Paulo: Editora da Universidade de São Paulo (EDUSP), 2004.

SPERLING, David. Prefácio. In: CELANI, M. G. C.; SEDREZ, M. (ORG.). **Arquitetura contemporânea e automação: prática e reflexão.** São Paulo: ProBooks, 2018.