

Introdução e Motivação

O estado atual das redes neurais e do aprendizado profundo difundiu o emprego de métodos de aprendizado de máquina supervisionado na resolução de problemas diversos. Entretanto, este tipo de solução não pode ser amplamente utilizado devido ao seu funcionamento opaco, o que impede a compreensão exata e não permite garantir a confiabilidade dos resultados.

Este trabalho dedica-se a clarificar o funcionamento dos métodos de aprendizado de máquina supervisionado, seguindo uma abordagem prática e concreta por meio do desenvolvimento de um arcabouço de testes, generalizável e aplicável para qualquer problema, modelo e dados. A efetividade do arcabouço foi validada aplicando-o a um modelo de GNN para o problema de previsão de tráfego.

Fundamentação Teórica

De modo genérico e simplificado, todo algoritmo de aprendizado de máquina supervisionado segue o esquema abaixo [1] [4].



A partir deste esquema, identificou-se **4 elementos fundamentais** a todo problema solucionado com aprendizado de máquina supervisionado e **2 elementos derivados** do processo de aprendizado. Cada elemento fundamental possui restrições associadas, as quais estão implícitas nos conjuntos \mathcal{V} , \mathcal{X}' , \mathcal{X} e \mathcal{Y} .

Elementos Fundamentais	Elementos Derivados
Dados brutos ¹	A função aprendida ²
Padrões definidos ³	Os valores previstos ⁴
Dados de treinamento	
Um modelo	

Tabela 1. Elementos fundamentais e derivados. (1) Dados na forma em que foram mensurados ou obtidos. (2) Representada pelos parâmetros aprendidos. (3) Transformações aplicadas sobre os dados brutos para gerar x_i . (4) Representados pelos resíduos do modelo [3] (válido para problemas de regressão).

Arcabouço de Testes

Concepção dos Testes

Ideação: Os testes têm como objetivo explicitar comportamentos de um modelo de interesse. Este processo é realizado verificando-se um atributo de algum dos elementos fundamentais e derivados, através do treinamento do modelo com um conjunto de dados, e os resultados são interpretados com base nos valores do erro e das métricas. A fim de estabelecer uma base robusta para interpretação dos resultados, selecionam-se modelos de comparação. Ademais, deve-se definir um contexto de execução dos testes, fixando hiperparâmetros e transformações sobre os dados, com o objetivo de que estes fatores não influenciem nos resultados.

Efetividade dos Testes: Apesar de cada teste atuar somente sobre um atributo de um dos elementos fundamentais ou derivados, eles têm capacidade de revelar informações sobre outros elementos além do elemento efetivamente testado. Os tipos de informações explicitadas estão na Figura 1.

Modelos de Comparação

Idealmente, os modelos de referência devem ser selecionados seguindo algum parâmetro de escolha compatível com o tipo de comportamento ou resultado que se deseja verificar do modelo testado. Propomos os seguintes parâmetros de escolha:

- Escolha baseada em complexidade: Modelos são selecionados com base na quantidade de parâmetros, camadas, operadores, entre outros fatores relacionados a complexidade do modelo.
- Escolha baseada em desempenho: Selecionam-se modelos que obtiveram bons resultados para o problema abordado no teste.
- Escolha baseada em variedade: Selecionam-se modelos com tipos de arquiteturas distintas.

O Arcabouço

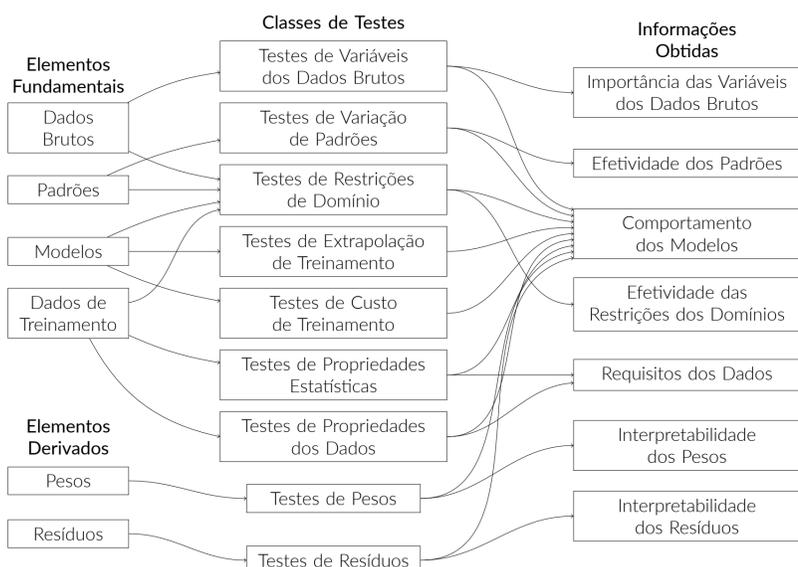


Figura 1. Representação do arcabouço de testes. Cada elemento fundamental ou derivado induz a geração de uma classe cujos testes exploram algum tipo de atributo deste elemento. As arestas entre as classes e os elementos indicam qual elemento é testado pela classe. As arestas entre as classes de testes e os tipos de informação relacionam as principais informações obtidas pela classe. Ressaltasse que as classes podem revelar mais informações além das indicadas pelas arestas.

Dados Sintéticos

O arcabouço se beneficia da utilização de dados sintéticos para representar cenários menos prováveis de se encontrar em dados reais, e para gerar dados com determinadas propriedades de interesse.

Geração dos Dados: Por simplicidade, estes dados podem ser gerados diretamente de funções base (como distribuições de probabilidade e exponenciais complexas), embutindo alguma propriedade de interesse nos valores gerados.

Dados de Referência: Dados gerados diretamente de alguma função base, ou que são um representante extremo de alguma propriedade. São importantes pois estabelecem uma base de comparação para outros dados gerados da mesma função ou que possuem a mesma propriedade.

Classes de Dados: Dados sintéticos e reais podem ser agrupados em classes, com base em critérios de união. Os critérios de união devem se basear em similaridade (por exemplo, dados de distribuição normal) ou em propriedades (como periodicidade ou linearidade).

Processos de Teste

O arcabouço possui dois processos de testes. O primeiro processo é adequado quando o atributo que se deseja testar é facilmente controlável através dos dados, enquanto o segundo é mais adequado para o caso oposto.

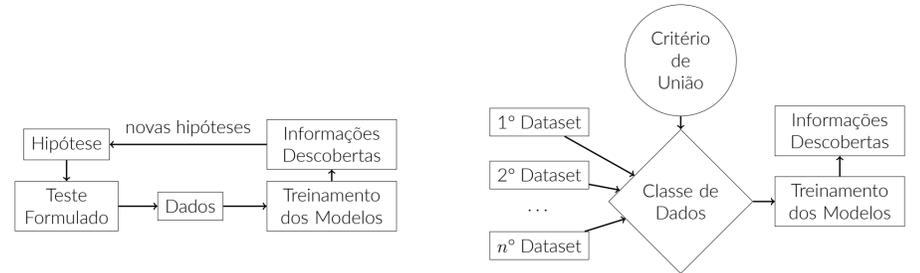


Figura 2. 1º Processo de Teste - Uma hipótese guia a formulação do teste, o qual guia a geração dos dados, e, com base nos dados, treinam-se os modelos e obtêm-se informações que conduzem a mais hipóteses.

Figura 3. 2º Processo de Teste - Os modelos são treinados com conjuntos de uma mesma classe, e em caso de similaridade de resultados, o critério de união afeta o modelo.

Aplicação do Arcabouço

Metodologia

Modelos de Interesse e de Comparação: Testou-se o modelo de GNN descrito no artigo [5], e, como modelos de comparação, selecionou-se a regressão linear e um modelo de FNN.

Informações Necessárias: Os modelos necessitam de uma série temporal S com dados de fluxo de veículos capturados por sensores. Além disso, a GNN necessita de um dígrafo fortemente conexo, aperiódico e ponderado representando uma cadeia de Markov das conexões dos sensores. Os pesos do dígrafo são calculados com base na velocidade dos veículos.

Dados Utilizados: Selecionou-se dados de janeiro de 2023 do conjunto de dados PEMSd3 [2], capturados por 331 sensores. Este conjunto possui intervalo de amostragem de 5 minutos.

Padrões Definidos: Aplicou-se uma interpolação linear a S e uma normalização aos dados brutos. Em seguida, cada tupla (x_t, y_t) dos dados de treinamento é gerada da seguinte forma: y_t corresponde a amostra de tempo t de S e x_t é gerado concatenando dados de curto, médio e longo prazo anteriores a y_t . Os dados de curto prazo correspondem as duas amostras anteriores a y_t , e os dados de médio e longo prazo correspondem respectivamente a conjuntos de amostras de S com período total de 1 e 7 dias, e intervalo de amostragem de 30 e 60 minutos.

Resultados Obtidos

A aplicação do arcabouço neste contexto revelou as seguintes informações:

Melhor Modelo: Na maioria dos testes realizados, a regressão linear apresentou desempenho igual ou superior à FNN e à GNN, demonstrando também um comportamento mais estável e robusto a diferentes propriedades dos dados, como dados não normalizados, presença de valores negativos, e valores inteiros ou arredondados.

Influência de Dependências Espaciais e Temporais: O modelo de GNN é pouco sensível às dependências espaciais, aprendendo majoritariamente as características temporais. Diminuir o intervalo de amostragem dos dados melhora a performance da GNN. Além disso, os 3 modelos aprendem o padrão temporal principalmente por meio dos dados de curto prazo.

Propriedades dos Dados: Propriedades estatísticas dos dados não afetam os modelos. A presença de ruído não afeta os resultados quando a magnitude destes é até 1% do maior valor dos dados, e tem pouco impacto nos resultados quando a magnitude é até 10% do maior valor. Além disso, os modelos são robustos a dados com alta dimensionalidade e obtêm bons resultados com dados com periodicidade (mesmo a regressão linear), porém não conseguem prever com qualidade dados com relações não lineares.

Extrapolação de Treinamento: Os dados utilizados possuem pouca variação no padrão periódico, o que permite que os 3 modelos prevejam com qualidade dados de períodos muito distantes do conjunto de treinamento. No entanto, caso haja padrões não vistos, a performance dos modelos é fortemente prejudicada, necessitando de muitos dados para a previsão ruim ser absorvida nas métricas.

Quantidade de Dados de Treinamento: Os 3 modelos geram uma boa previsão com apenas algumas dezenas de dados de treinamento.

Referências

- Yaser S. Abu-Mostafa, Malik Magdon-Ismael e Hsuan-Tien Lin. *Learning From Data*. AMLBook, 2012.
- Caltrans. *Performance Measurement System (PeMS)*. 2024. URL: <http://pems.dot.ca.gov>.
- Singer Morettin. *Estatística e Ciência de Dados*. Grupo Editorial Nacional, 2021.
- Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, 1998.
- Yang Zhang, Tao Cheng e Yibin Ren. "A graph deep learning method for short-term traffic forecasting on large road networks". Em: *Computer-Aided Civil and Infrastructure Engineering* 34.10 (2019), pp. 877–896. DOI: <https://doi.org/10.1111/mice.12450>.