

Fichamento de Artigo

Leandro Baêta Lustosa Pontes
Profº. Dr. Francisco de Assis Boldt
Pesquisa em Computação Aplicada

IFES, Serra, 18/02/2021

1 Fichamento de arquivo

1. O que o artigo propõe? Qual é a novidade que ele apresenta?

Propõe um procedimento sistemático para comparar de forma justa os escores de desempenho experimental para métodos de aprendizado de máquina para o diagnóstico de falhas baseado em vibração, que se concentra na análise de sinais de vibração para identificar possíveis falhas em um equipamento.

É proposta uma diretriz algorítmica na forma de um framework de validação, analisando criticamente a divisão usual treinamento-validação-teste. Tendo em vista, que na grande maioria das publicações científicas relacionadas, a precisão estimada e os critérios de desempenho semelhantes são os únicos parâmetros de qualidade apresentados. Além do enviesamento dos resultados ao reutilizar a base de teste, já utilizada no treinamento do agente inteligente. No entanto, o algoritmo que deu origem a esses resultados é em grande parte enviesado, baseado em métodos de validação simples e na reciclagem de padrões idênticos encontrados em conjuntos de dados de teste, que anteriormente foram usados para treinamento. E o super ajuste de hiperparâmetros, que introduzem resultados superotimistas aos experimentos.

Uma das principais hipóteses e novidade apresentada por este trabalho é mostrar que uma validação cruzada aninhada, embora não esteja sujeita a pontuações de desempenho superotimistas, é sempre melhor do que a validação cruzada convencional.

2. Como os experimentos do artigo são modelados para corroborar a hipótese proposta?

A partir de um modelo de acurácia aninhado, onde o dataset é dividido em treino e teste, e depois cada dataset de treino é dividido internamente em treino' e validação. A cada iteração externa são realizadas n iterações internas. Para cada ciclo interno os fold são alterados, o que era conjunto de validação, passa a ser treino' e vice-versa, até que todas as possibilidades de fold sejam executadas e as performances obtidas são gravadas em uma matriz. Ao fim de todos os ciclos internos é calculada a média de todas as execuções e o melhor conjunto de hiperparâmetros é guardado pelo loop externo. A cada ciclo externo é avaliado se o conjunto de hiperparâmetros atual deve ser mantido, ou ajustado por um melhor.

3. Quais os pontos fortes do artigo?

- Review de artigos que fazem análise de sinais de vibração e apresentar os principais problemas com esses estudos.
- Propor um framework que permite a construção de um modelo algorítmico, que evite o viés de similaridade, verificar a significância estatística das diferenças dos resultados e que garanta a reprodutibilidade do experimento.

- A demonstração com 3 datasets gaussianos de 8 variáveis e duas classes A e B, demonstrando que em geral os modelos enviesados tendem a ser superotimistas e os modelos de acurácia aninhados corrigem um pouco a performance do modelo e eles tendem a se igualar conforme o número de amostras crescem, datasets com poucas amostras são prejudiciais para qualquer modelo.
- As técnicas para aumentar a reprodutibilidade, para evitar o viés de similaridade e para realizar a análise estatística dos resultados para a comparação com outros estudos.

4. Quais os pontos fracos dele?

- Existem modelos mais sofisticados para calcular a performance final para os modelos de acurácia aninhados, mas isso não é explorado pelos autores.
- O framework proposto carrega um problema intrínseco, pois quando um sinal de aquisição é dividido em muitas amostras, uma amostra carrega informações da própria aquisição, não apenas informações de falhas. Portanto, a validação cruzada aninhada pode produzir resultados otimistas demais.
- Tendo em vista que a exigência $J = K * R.30$ não ficou satisfeita, não foram realizados testes estatísticos. Logo, uma investigação mais detalhada deve ser empregada para detectar diferenças significativas de desempenho.
- As diferenças de acurácias não foram suficientes para rejeitar a hipótese de desempenhos semelhantes entre os métodos classificadores.
- Existe uma dificuldade em distribuir os padrões nas dobras, portanto, uma escolha meticulosa é necessária ao projetar os experimentos.
- Cenários de produção são mais difíceis de se generalizar, pois sensores podem falhar, novos pontos de ajustes são introduzidos e em um contexto de aprendizado supervisionado, pode-se haver o erro humano ao rotular a fonte de dados corretamente.

5. O artigo está relacionado de alguma forma com o seu projeto de dissertação? Por quê?

- Sim, algo que percebe-se é que muitos trabalhos científicos relacionados ao meu projeto de dissertação, não apresentam claramente nos seus artigos, ou as vezes no modelo criado, como os datasets foram tratados, tornando difícil a tarefa de reproduzir os experimentos realizados.
- Outro problema é em relação aos hiperparâmetros utilizados nos experimentos, geralmente os autores deixam os toolbox configurados de forma default e com isso não informam como o modelo gerou o resultado. E no final não conseguimos verificar e esse dado foi omitido, ou de fato foi utilizada a configuração default.
- A solução que é dada para a comparação entre diferentes métodos de acurácia, com o intuito de controlar o erro familiar, aplicando a análise estatística poderá ser utilizado no meu projeto de dissertação.