

# SELEÇÃO DINÂMICA COM MEDIDAS DE COMPLEXIDADE À NÍVEL DE INSTÂNCIA EM RANDOM FORESTS

## Mestrado

Gustavo Yamamoto Bonacina, Alceu de Souza Britto Júnior

**Contexto:** Os problemas atuais apresentam uma quantidade de dados e um nível de complexidade que desafiam os métodos existentes de reconhecimento de padrões. Criar um classificador monolítico para cobrir a maioria destes problemas é impraticável. Pesquisas recentes tem focado na geração de Sistemas de Classificadores Múltiplos (MCSs) e evidenciado que a seleção de classificadores específicos para cada base de teste (seleção dinâmica) tem gerado resultados melhores. Random forest é um MCS composto por árvores de decisão, onde cada árvore varia conforme os valores de um vetor aleatório amostrado de forma independente, com a mesma distribuição para todas as árvores. Ele é considerado um dos métodos de conjunto de classificadores mais bem sucedido atualmente por ser rápido, robusto à ruídos e não sofrer sobreajuste, entre outras vantagens. Embora seu desempenho seja satisfatório, ainda pode haver um ganho considerável nos resultados, conforme mostram os oráculos calculados. **Objetivo Geral:** Este projeto visa melhorar as taxas de precisão com a implementação da seleção dinâmica utilizando medidas de complexidade à nível de instância<sup>1</sup> no algoritmo random forest. **Método de Pesquisa:** Diferentes medidas de complexidade serão implantadas no processo de seleção dos classificadores gerados. Os testes serão realizados inicialmente em 16 bases de dados, sendo 4 delas sintéticas. As outras 12 foram utilizadas em outro estudo<sup>2</sup> que também visa melhorar os resultados do método. Assim, serão observados os resultados e alguns de seus coeficientes, como a taxa de erro, taxa de acerto e erro relativo absoluto, serão armazenados para posterior comparação com outras variações do algoritmo. **Resultados Preliminares:** O possível ganho na precisão foi evidenciado com o cálculo do oráculo das 12 bases de dados e posterior comparação com a taxa de precisão obtida pelo random forest. O ganho na precisão ultrapassa os 26% e apresentou uma média de 13%.

---

<sup>1</sup> Smith, M. R., Martinez, T., & Giraud-Carrier, C. (2014). An instance level analysis of data complexity. *Machine learning*, 95(2), 225-256.

<sup>2</sup> Tsymbal, A., Pechenizkiy, M., & Cunningham, P. (2006). Dynamic integration with random forests. In *Machine Learning: ECML 2006* (pp. 801-808). Springer Berlin Heidelberg.

**Palavras-chave:** reconhecimento de padrões, classificação, seleção dinâmica, random forests, medidas de complexidade