

## Geração de ensembles de classificadores com base em análise da complexidade do problema de classificação

### Doutorado

Marcos Monteiro Junior, Alceu de Souza Britto Junior

**Contexto:** Com a grande quantidade de dados e conseqüentemente problemas mais complexos de classificação, foi necessário melhorar os métodos de classificação de dados. Nesse contexto surgiram diferentes maneiras para minimizar esse problema, tal como a combinação de vários classificadores. É possível conseguir resultados superiores utilizando mais de um classificador, levando em conta a eficiência e a precisão. A combinação de classificadores ditos fracos muitas vezes pode apresentar um resultado de classificação melhor que um único classificador forte e com menos esforço computacional. Essa técnica denomina-se sistema de múltiplos classificadores (SMC). O processo para criação de um SMC pode ser generalizado em três etapas principais: geração, seleção e integração. Na geração é onde os classificadores são gerados, na seleção selecionados e na integração a decisão é tomada de acordo com os resultados obtidos. Existem diferentes maneiras de criar um *ensemble*, (grupo de classificadores) de classificadores, deve-se definir o se o *ensemble* a ser gerado é heterogêneo ou homogêneo. São exemplos de classificadores homogêneos *Bagging*, *Boosting*, *Random Subspaces*. Relacionado aos classificadores, a complexidade do problema é uma medida que interfere diretamente na qualidade, na geração de classificadores e na acurácia. Além disso, quanto mais complexo o problema mais difícil é gerar um *ensemble* de classificadores capazes de se distinguirem no espaço de complexidade. **Objetivo Geral:** Mediante o contexto exposto, esse trabalho tem como objetivo propor uma nova técnica de geração de ensembles de classificadores baseada em análise de complexidade do problema de classificação. Para isso, busca-se implementar uma técnica capaz de explorar melhor o espaço de complexidade do problema gerando *ensemble* de classificadores capaz de cobrir melhor o espaço de características. **Questão de pesquisa:** O uso de análise da complexidade do problema possibilita melhorar a dispersão dos *bags* no espaço de características para melhorar a geração do *ensemble* de classificadores? **Método de pesquisa:** 1) Levantamento bibliográfico sobre técnicas de geração de pool; 2) estudo das medidas de complexidade de dados existentes na literatura; 3) seleção de um método de otimização a ser utilizado (AG, PSO, ...); 4) implementação de uma nova abordagem de geração de pool com base na análise de complexidade de dados; 5) Avaliação da abordagem criada usando protocolo experimental pré-definido. **Resultados Preliminares:** Não há.

**Palavras-chaves:** Geração de classificadores, Complexidade dos dados, Ensembles