

Universidade de Pernambuco
Programa de Pós-Graduação em Engenharia da
Computação (PPGEC)

Proposta de Tese de Doutorado

Área: Computação Inteligente

Título: Uso de aprendizagem semi-supervisionada para tarefas com escassez de dados rotulados

Orientador – Carmelo José Albanez Bastos Filho (carmelo.filho@upe.br)

Coorientador- Felipe Farias (felipefariax@gmail.com)

Devido à escassez de dados rotulados em muitas tarefas, bem como o alto custo de rotulação desses dados em termos de tempo e recursos, paradigmas de aprendizado que utilizam dados não rotulados têm um alto valor na pesquisa em aprendizado de máquina. Um desses paradigmas é o aprendizado semi-supervisionado, que busca aproveitar tanto dados rotulados quanto não rotulados [1].

Apesar de sua aparente utilidade, existem vários problemas em aberto nessa área que dificultam a adoção desse paradigma [2]. Alguns desses desafios incluem:

- Generalização de modelos: quando há poucos dados rotulados, garantir que os modelos generalizem bem para novos exemplos é complexo;
- Especificação de modelos: é difícil especificar modelos capazes de aprender a distribuição de dados não rotulados;
- Adaptação a diferenças de distribuição: lidar com diferenças na distribuição entre dados rotulados e não rotulados é um desafio constante.

Neste contexto, a presente pesquisa propõe uma revisão sistemática da literatura sobre aprendizado semi-supervisionado. Serão investigados métodos indutivos [3] e transdutivos [4] tradicionais e do estado-da-arte; será realizada a comparação de tais métodos para uma mesma tarefa e, finalmente, a concepção e desenvolvimento de um método novo dentro do paradigma de aprendizado semi-supervisionado. Esperamos que os resultados dessa pesquisa enriqueçam áreas com escassez de dados rotulados.

Referências Bibliográficas

- [1] Zhu, X. (2008). *Semi-supervised learning literature survey*. Technical Report. 1530, University of Wisconsin-Madison.
- [2] Cholaquidis, A., Fraiman, R., & Sued, M. (2018). On semi-supervised learning. *ArXiv*. /abs/1805.09180.
- [3] Xu, L., & Schuurmans, D. (2005) Unsupervised and semi-supervised multi-class support vector machines. In *Proceedings of the 20th National Conference on Artificial Intelligence* (Vol. 5, p. 13).
- [4] Zhu, X. (2005). *Semi-supervised learning with graphs*. Ph.D. thesis, Carnegie Mellon University.