

# Aprendizado Semi-Supervisionado utilizando Competição e Cooperação entre Partículas em Dígrafos

**Autor:** Bruno Fouz Valente (brunofouzv@gmail.com), **Orientador:** Prof. Dr. Fabricio Aparecido Breve (fabricio@rc.unesp.br), UNESP Campus Rio Claro, Bacharelado em Ciências da Computação, Bolsa PIBIC/Reitoria (RT).

Palavras Chave: *Aprendizado de Máquina, Aprendizado Semi-Supervisionado*

## Introdução

Aprendizado de máquina é uma subárea da ciência da computação, evoluída dos estudos de reconhecimento de padrões em Inteligência Artificial, que tem como objetivo criar algoritmos que detectam e aprendem um comportamento comum numa base de dados a partir de observações, similar à forma como o ser humano desenvolve o seu aprendizado. Esta subárea tem se tornado muito importante nos últimos anos para o processamento automatizado da grande quantidade de dados que é produzido no dia-a-dia [1].

O modelo de competição e cooperação entre partículas [2] é uma abordagem de aprendizado de máquina, na categoria de aprendizado semi-supervisionado baseado em grafos, que é inspirado pela natureza.

## Objetivos

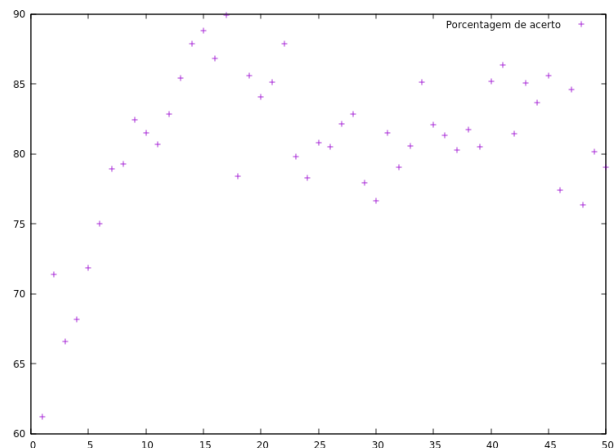
O objetivo deste projeto é de estender o modelo de competição e cooperação entre partículas para aplicação em dígrafos, formados a partir da distância euclidiana entre os itens da base de dados. A hipótese é de que, dependendo da base de dados, o modelo possa ter uma melhoria na acurácia dos resultados da classificação ou de desempenho na execução.

## Material e Métodos

O algoritmo do modelo proposto foi desenvolvido com base no modelo original de competição e cooperação entre partículas, para posteriormente executar testes e experimentos comparando os dois modelos. As bases de dados da UCI Machine Learning Repository [3] e de Chapelle [4] foram utilizadas com a finalidade de avaliar a acurácia de classificação.

## Resultados e Discussão

Os resultados preliminares dos testes realizados com as bases de dados citadas seguem abaixo.



**Figura 1.** Estudo dos K vizinhos mais próximos na base Wine (K = 17 é o valor mais ótimo).

Base de dados	Modelo Original	Modelo com Dígrafos
Wine	<b>94,27%</b>	85,5%
chappelle-g241c	90,78%	<b>93,27%</b>
chappelle-Digit1	88,51%	<b>91,79%</b>

**Tabela 1.** Resultados da acurácia nos dois modelos.

## Conclusões

Os resultados mostram que, dependendo da natureza da base de dados utilizada, há um aumento na acurácia da classificação.

## Agradecimentos

Os autores gostariam de agradecer à FAPESP (#2016/05669-4), ao CNPq (#475717/2013-9) e à PROPE pelo auxílio financeiro.

<sup>1</sup> Olivier Chapelle, Bernhard Schölkopf, e Alexander Zien. *Semi-Supervised Learning*, The MIT Press, **2006**.

<sup>2</sup> Fabricio Breve, Liang Zhao, Marcos Quiles, Witold Pedrycz e Jiming Liu. *Particle Competition and Cooperation in Networks for Semi-Supervised Learning*, IEEE Transactions on Knowledge and Data Engineering, Vol. 24, No. 9, **2012**.

<sup>3</sup> Bache and Lichman. *Iris data set*.

<https://archive.ics.uci.edu/ml/datasets/Iris>, UCI Machine Learning Repository.

<sup>4</sup> Olivier Chapelle. *The Benchmark Data Sets*.

<http://olivier.chapelle.cc/ssl-book/benchmarks.html>.