



**TÍTULO:**

**OTIMIZAÇÃO DE PARÂMETROS NO MODELO DE COMPETIÇÃO E COOPERAÇÃO DE PARTÍCULAS**

**AUTORES:**

Thiago Costa Leme Rodrigues, Fabricio Aparecido Breve, Departamento de Estatística, Matemática Aplicada e Computação, Universidade Estadual Paulista (UNESP), Rio Claro (SP), thiago.clrodrigues@gmail.com, bolsa PIBIC.

**INTRODUÇÃO:**

Aprendizado Semi-Supervisionado é a categoria de aprendizado de máquina na qual se utiliza tanto dados rotulados quanto não rotulados. Na abordagem de caminhada de partículas, cada item é representado por um vértice e as arestas são criadas com base em uma medida de similaridade, no caso a distância euclidiana (Breve et al., 2012). São criadas partículas para cada item rotulado, e elas caminham pelo grafo em um processo de cooperação (entre partículas da mesma classe) e competição (entre partículas de classes diferentes). Cada grupo de partículas tenta dominar a maior quantidade possível de vértices utilizando a regra aleatório-gulosa.

**OBJETIVOS:**

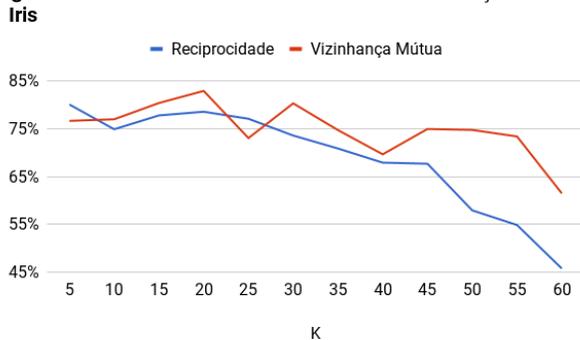
O principal objetivo deste projeto é explorar novas dinâmicas no modelo de caminhada de partículas, incluindo mudanças na estrutura dos grafos gerados e diferentes modelos de construção dos grafos, para tentar encontrar parâmetros que apresentem melhor acurácia de classificação e/ou desempenho.

**MATERIAL E MÉTODOS:**

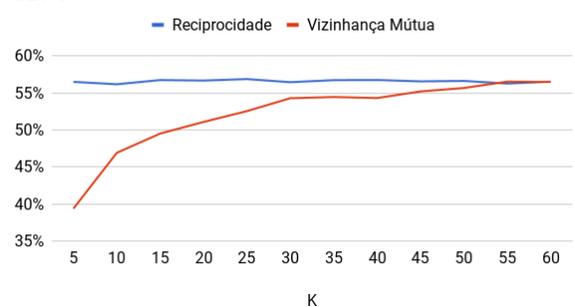
No primeiro modelo avaliado foi mantida a condição de reciprocidade presente inicialmente, já no segundo dois nós somente são vizinhos se um nó A for vizinho de um nó B e B também for previamente vizinho de A (se essa condição não é satisfeita não há ligação entre esses nós), modelo chamado de vizinhança mútua. Modelos avaliados em função do parâmetro  $k$ , número mínimo de vizinhos de cada nó presente no grafo.

**RESULTADOS:**

**Figura 1.** Percentuais de acerto da Base Iris em função dos dois métodos.



**Figura 2.** Percentuais de acerto do SET-5.



**DISCUSSÃO:**

Analisando os dois métodos propostos, observa-se que geralmente com o aumento do valor de  $k$  o percentual tende a diminuir quando se utiliza o método da reciprocidade. Por outro lado, com a utilização do método da vizinhança mútua, o percentual tende a aumentar. O uso de valores de  $k$  maiores no método da vizinhança mútua possibilita resultados melhores do que valores menores, principalmente porque esse método trabalha com a proposta de formar ligações entre nós apenas quando os dois já são previamente vizinhos um do outro, fato que acaba deixando nós sozinhos (sem vizinhos), algo que impossibilita a classificação deles e, também, acaba prejudicando a classificação da base como um todo. Ao se usar valores de  $k$  maiores esse problema é minimizado. É visível que para cada base utilizada obteve-se um valor de  $k$  ótimo diferente, fato que pode ser explicado pelas diferenças existentes entre as bases. Dentre as quais a mais significativa é o fato de que a distribuição dos elementos entre as classes não é uniforme na maioria das bases e, inclusive, em algumas a diferença de elementos entre as classes é gritante, deixando algumas classes com um número de elementos muito baixo, dificultando consideravelmente a classificação deles. Além disso também deve-se levar em consideração a diferença de elementos presentes e a quantidade de atributos fornecidos, já que esses fatores interferem na montagem da rede e, portanto, podem vir a criar redes com formatos diversos como, por exemplo, uma rede completamente dispersa, completamente conexa, ou ainda com uma distinção bem evidente de comunidades.

**REFERÊNCIAS:**

Breve, F. & Zhao, L. (2012). Particle competition and cooperation in networks for semi-supervised learning with concept drift. In Neural Networks (IJCNN), The 2012 International Joint Conference on (pp. 1–6).