

# Proposta de Dinâmica de Caminhada no Algoritmo de Competição e Cooperação entre Partículas

Lucas Guerreiro\*, Fabricio Breve  
Universidade Estadual Paulista (UNESP)  
Rio Claro, Brasil  
guerreiroLuc@gmail.com, fabricio@rc.unesp.br

**Resumo**—O Algoritmo de Competição entre Partículas original apresenta caminhadas em um grafo, mantendo informações de dominância nos nós da rede. Neste trabalho é proposta uma nova dinâmica de caminhada, com dominâncias mantidas nas arestas do grafo, gerando uma nova abordagem de caminhada na rede. Apresenta-se neste trabalho detalhes da proposta, bem como os resultados atingidos em conjuntos de dados de *data sets* de aprendizado semi-supervisionado.

*Área: Inteligência Computacional*

## I. INTRODUÇÃO

Aprendizado semi-supervisionado, com destaque o aprendizado baseado em grafos, vem sendo uma das áreas mais estudadas e trabalhadas em Aprendizado de Máquina [1][2]. Estas abordagens propiciam um alto nível de acurácia com um custo relativamente baixo quando comparado com algoritmos supervisionados, por terem uma pequena parte de dados rotulados inicialmente [3].

O Algoritmo de Competição e Cooperação entre Partículas (PCC) [4] é um algoritmo baseado em grafos que se destaca por ter atingido uma alta taxa de classificação com menor custo computacional quando comparado aos algoritmos semi-supervisionados mais utilizados [5]. Esta foi a principal motivação na escolha do algoritmo para ser analisado e modificado, em busca de melhores taxas de classificação. Propostas anteriores de alterações no algoritmo, com a utilização de outras métricas de distância na construção do grafo, atingiram bons resultados [6][7]. Com base nisso, neste trabalho é proposta uma nova abordagem de caminhada de partículas, considerando que o modelo original sempre se utilizou dos nós do grafo para análise de dominância, podendo-se explorar novos métodos de caminhada de partículas no algoritmo.

## II. CONCEITOS E TÉCNICAS

A proposta deste trabalho é, portanto, armazenar as dominâncias relativas a cada classe componente da base nas arestas, conforme uma partícula caminha sobre tal aresta, atualiza-se as dominâncias de forma análoga à proposta original.

Com esta mudança é necessário um novo método de classificação dos nós não rotulados. Cada nó será dependente de todas as arestas às quais está conectado. Para isso, toma-se qual classe mais domina cada aresta conectada ao nó; a classe

que tiver maior dominância sobre as arestas será a classe atribuída ao nó.

## III. METODOLOGIA DE DESENVOLVIMENTO

Para analisar a eficácia do método proposto, o algoritmo é submetido aos *data sets* propostos por Chapelle [3], que são utilizados na análise de algoritmos semi-supervisionados. Compara-se então os resultados do algoritmo com os demais algoritmo no estado da arte. Uma breve descrição dos *data sets* pode ser vista na Tabela 1

TABELA 1 – BASES DE DADOS UTILIZADAS.

Data set	Classes	Parâmetros	Instâncias
g241c	2	241	1500
g241d	2	241	1500
Digit1	2	241	1500
USPS	2	241	1500
COIL	6	241	1500
BCI	2	117	400

Uma característica destes *data sets* é a subdivisão dos dados rotulados em grupos de 10 e de 100 exemplos rotulados. Tais grupos estão divididos em 12 subgrupos diferentes para cada *data set*, de forma a propor uma análise da média das execuções para todas as divisões propostas. Portanto, nas análises propostas neste trabalho, o algoritmo é executado 100 vezes para cada subconjunto, ao final é calculada a média dessas execuções para cada conjunto e comparada com os demais algoritmos.

Nestas execuções foram utilizadas as sete métricas de distância propostas em estudos anteriores [6]. A otimização de parâmetros foi feita por meio de algoritmos genéticos [8][9], identificando os melhores parâmetros para cada subdivisão das bases e executando-se o algoritmo com os valores de parâmetros otimizados.

## IV. RESULTADOS

Resultados de erros médios do experimento para cada um dos *data sets* da Tabela 1 podem ser vistos na Tabela 2 para o caso de 10 itens de dados rotulados e na Tabela 3 para 100

VII Workshop do Programa de Pós-Graduação em Ciência da Computação: "Interação entre Academia e Empresa", Unesp, Rio Claro, 14 e 15 de setembro de 2017.

itens de dados rotulados, comparados com o algoritmo PCC original.

TABELA 2 – ERROS MÉDIOS PARA 10 ITENS ROTULADOS (%).

	g241c	g241d	Digit1	USPS	COIL	BCI
PCC	37,57	43,94	9,94	17,44	58,65	47,66
Método Proposto	45,98	45,34	18,65	16,63	60,55	47,15

TABELA 3 – ERROS MÉDIOS PARA 100 ITENS ROTULADOS (%).

	g241c	g241d	Digit1	USPS	COIL	BCI
PCC	24,20	23,93	2,65	4,65	14,85	44,38
Método Proposto	37,49	37,06	5,15	9,03	25,96	49,28

Na Tabela 4 pode-se ver o posicionamento do algoritmo proposto no ranking, apresentado por Breve et al. [4], de algoritmos semi-supervisionados com 10 itens rotulados e na Tabela 5 o posicionamento para 100 itens de dados rotulados, assim como os resultados do algoritmo PCC original. O método proposto bem como seus resultados são adicionados aos rankings nas tabelas abaixo que compara dezoito métodos de aprendizado semi-supervisionado [3][4].

TABELA 4 – RANKING PARA 10 ITENS ROTULADOS.

	g241c	g241d	Digit1	USPS	COIL	BCI
PCC	4	6	6	7	4	6
Método Proposto	12	8	15	5	6	4

TABELA 5 – RANKING PARA 100 ITENS ROTULADOS.

	g241c	g241d	Digit1	USPS	COIL	BCI
PCC	8	5	3	2	8	9
Método Proposto	12	12	15	14	14	18

Conforme pode-se observar, o algoritmo proposto atingiu erros menores do que muitos dos algoritmos mais utilizados em vários casos, com classificação melhor do que o algoritmo PCC original nas bases UCI e USPS para 10 itens rotulados, tendo eficácia boa nas demais bases com 10 itens de dados rotulados. Ainda, apesar de estar um pouco abaixo do algoritmo PCC, o algoritmo proposto ainda obteve bons índices de classificação na base COIL, de 6 classes, para 10 itens de dados rotulados. Contudo, para 100 itens rotulados a acurácia do algoritmo proposto não esteve entre as melhores posições, sendo no

melhor dos casos a décima segunda melhor. Isto evidencia que o método proposto atua melhor em casos de poucos dados rotulados.

## V. CONSIDERAÇÕES FINAIS

Neste trabalho foi apresentada uma nova abordagem de caminhada de partículas para o modelo de Competição e Cooperação entre Partículas. Considera-se o método proposto como positivo por ter tido, principalmente, bons índices de classificação quando trabalhando com poucos itens de dados rotulados. A acurácia do algoritmo para 100 itens de dados rotulados não esteve entre as melhores, porém ainda considera-se a abordagem positiva pela apresentação de resultados de uma técnica diferente e, ainda, por superar alguns dos algoritmos semi-supervisionados mais famosos utilizados atualmente.

Trabalhos futuros na linha do projeto apresentado podem ser feitos a partir de maior refinamento do algoritmo genético utilizado na definição de parâmetros, já que o algoritmo se mostra muito sensível aos parâmetros utilizados. Outra proposta pode ser feita na mescla das características do modelo PCC original com o método proposto, extraindo quando a dinâmica de nós é a melhor com os casos em que o método proposto neste trabalho tem melhor índice de classificação.

## REFERÊNCIAS

- [1] ALPAYDIN, E.. Introduction to Machine Learning. MIT Press, 2004.
- [2] MITCHELL, T. Machine Learning. McGraw Hill, 1997.
- [3] CHAPELLE, O.; SCÖLKOPF, B.; ZIEN, A. Semi-Supervised Learning. Adaptive Computation and Machine Learning. MIT Press, 2006.
- [4] BREVE, F. A.; ZHAO, L.; QUILES, M. G.; PEDRYCZ, W.; LIU, J. Particle Competition and Cooperation in Networks for Semi-Supervised Learning. IEEE Transactions on Knowledge and Data Engineering, v. 24, p. 1686-1698, 2012.
- [5] BREVE, F. A.; QUILES, M. G.; ZHAO, L. Interactive Image Segmentation of Non-contiguous Classes Using Particle Competition and Cooperation. 15th International Conference on Computational Science and Its Applications (ICCSA 2015), v. 9155, p. 203-216, 2015.
- [6] GUERREIRO, L.; BREVE, F. A. Analysis of the Influence of Distance Metrics on the Semi-supervised Algorithm of Particle Competition and Cooperation. XI Workshop de Visão Computacional (WVC2015), p. 77-82, 2015.
- [7] GUERREIRO, L.; BREVE, F. A. Analyzing and Inferring Distance Metrics on the Particle Competition and Cooperation Algorithm. In: Gervasi O. et al. (eds) Computational Science and Its Applications – ICCSA 2017. ICCSA 2017. Lecture Notes in Computer Science, vol 10409. Springer, Cham, 2017.
- [8] HOLLAND, J. H. Genetic algorithms and the optimal allocation of trials. SIAM Journal of Computing, 2(2), 68-105, 1973.
- [9] MITCHELL, M. An Introduction to Genetic Algorithms. MIT Press, 1998.