

Otimização de Parâmetros no Modelo de Competição e Cooperação de Partículas

Thiago Costa Leme Rodrigues, Fabricio Aparecido Breve

Universidade Estadual Paulista Campus de Rio Claro, Instituto de Geociências e Ciências Exatas (IGCE), Departamento de Estatística, Matemática Aplicada e Computação, Ciências da Computação

thiago.clrodrigues@gmail.com, bolsa PIBIC – Bolsa Reitoria (RT)

Palavras Chave: *Aprendizagem de Máquina, Caminhada de Partículas, Grafos.*

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 na regra aleatório-gulosa de seleção de vértices vizinhos a serem visitados, para tentar encontrar parâmetros que apresentem melhor acurácia de classificação e/ou desempenho.

Material e Métodos

Para que fossem atingidos os objetivos, inicialmente foi implementado o modelo original de caminhada de partículas, o qual posteriormente sofreu alterações na construção do grafo e na regra aleatório-gulosa. Para simulações computacionais foram usadas bases de dados extraídas do repositório UCI (Bache & Lichman, 2013).

Os parâmetros avaliados nesse estudo foram: k , número mínimo de vizinhos no grafo; *exponencial* (e), que é aplicado sobre a distância em saltos entre o vértice de origem da partícula e os vértices candidatos à visita.

Resultados e Discussão

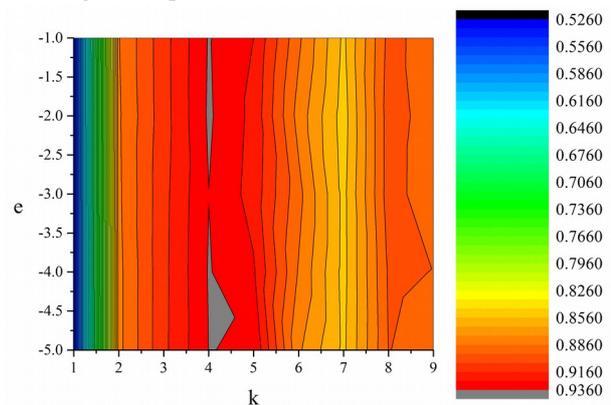
Na Tabela 1 e Figura 1 são apresentados resultados parciais de percentuais de acerto em função dos parâmetros k e e referentes a base de dados Iris, que apresenta 150 elementos com 4 atributos cada, sendo eles as larguras e comprimentos das sépalas e pétalas da planta.

XXVIII Congresso de Iniciação Científica

Tabela 1. Percentuais de acerto em função dos parâmetros k e e .

$k \backslash e$	-2	-3	-4	-5	-6
1	51,333	51,333	51,333	51,333	51,333
2	83,320	83,413	83,620	83,633	83,667
3	89,300	89,573	89,380	89,440	89,473
4	93,207	93,320	93,260	93,267	93,327
5	82,100	82,193	82,360	82,300	82,240
6	81,240	81,313	81,107	81,347	81,293
7	81,847	81,787	81,720	81,653	81,900
8	77,847	77,820	77,940	77,960	78,007
9	77,540	77,613	77,360	77,527	77,500
10	74,673	74,693	74,680	74,653	75,080

Figura 1. Curvas de contorno para percentuais de acerto em função dos parâmetros k e e .



Com base na Tabela 1 e Figura 1 é possível observar que os valores $k = 4$, $e = -3$, por exemplo, possibilitam obtenção das maiores percentuais de acerto.

Conclusões

Conclui-se que as modificações resultaram em melhorias, sendo possível obter maiores percentuais de acerto. Como por exemplo, na base Iris, os resultados obtidos com $k = 4$ foram melhores em relação ao valor original $k = 3$.

Agradecimentos

À FAPESP (#2016/05669-4), ao CNPq (#475717/2013-9) e à UNESP pelo apoio financeiro.

Bache, K. & Lichman, M. (2013). UCI machine learning repository.
Breve, F., Zhao, L., Quiles, M., Pedrycz, W., & Liu, J. (2012). *Knowledge and Data Engineering, IEEE Transactions on*, 24(9), 1686–1698.