

Detecção de Comunidades Sobrepostas utilizando Modelos Inspirados pela Natureza

Ivan Scaranello Cartolano, Fabricio Aparecido Breve, Campus Rio Claro, Instituto de Geociências e Ciências Exatas, Ciências da Computação, ivan_6365@hotmail.com, Bolsa de Iniciação Científica concedida pela Fapesp.

Palavras Chave: *Aprendizado de Máquina, Aprendizado Semi-supervisionado, Detecção de Comunidades.*

Introdução

O aprendizado de máquina é uma técnica que permite que o programa de computador aprenda de uma experiência 'E', com relação a uma tarefa 'T' e sua respectiva performance 'P'. Neste caso, a performance 'P' medida em relação à tarefa 'T' aumenta após a experiência 'E'.

Material e Métodos

Muitos algoritmos começam transformando a sua base de dados em um grafo, esse conceito é usado neste trabalho.

Cada indivíduo da base será um nó do grafo, e como critério para as arestas, são definidos os K vizinhos mais próximos de cada nó, e essa proximidade é dada pela distância euclidiana entre eles: sejam $[a_1, a_2, \dots, a_n]$ e $[b_1, b_2, \dots, b_n]$ os dados dos indivíduos A e B respectivamente, a distância euclidiana 'D' entre eles é dada por:

$$D = (a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2$$

Definidas as distâncias, o grafo pode ser representado por uma matriz de adjacência. Se o indivíduo j estiver entre os K vizinhos do indivíduo i, ou vice-versa, então $mat[i,j] = 1$, caso contrário, $mat[i,j] = 0$;

Resultados e Discussão

Para melhor visualização, a gravura abaixo representa parte do grafo criado a partir da base wine do site

<http://archive.ics.uci.edu/ml/datasets/Wine>

Com essa estrutura, podemos estudar a relação entre os nós, classificando-os. Supondo, por exemplo, que a base seja formada por x classes diferentes de indivíduos, rotulando uma parte deles e usando o grafo, este estudo pretende identificar a classe do restante dos indivíduos da base de dados.

Para isso, é criada uma partícula para cada indivíduo rotulado, partículas que vem de indivíduos da mesma classe formam um time, e os times percorrem o grafo, marcando território e competindo pelos nós.

Ao término das interações, o que se espera é que o time dominante em cada nó corresponda a classe deste nó.

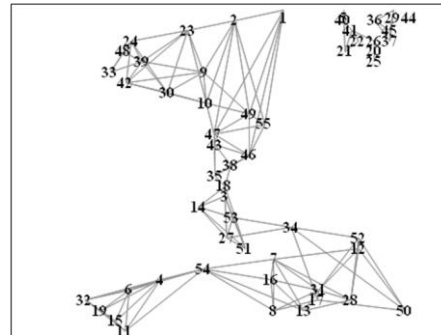


Figura 1. Representação do grafo realizado a partir da base de dado Wine

Conclusões

Grafos ajudam a representar estruturas topológicas não triviais, e assim, são de fundamental importância para esse estudo de comunidades sobrepostas que se pretende fazer com aprendizado de máquina.

Agradecimentos

Agradecimentos pelo apoio da Agência fomentadora de pesquisa Fapesp.