

Determinação das funções de pertença em *fuzzy clustering* através da Matriz-U de um “Self-Organizing Map”

Miguel Loureiro¹, Victor Lobo^{1,2}, Fernando Bação¹
mloureiro@isegi.unl.pt, vlobo@isegi.unl.pt, bacao@isegi.unl.pt

¹Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa

²Escola Naval, Marinha Portuguesa

Palavras chave: Self-Organizing Map, Fuzzy Clustering, FCM, Data Mining.

O *Fuzzy C-Means* (FCM) é um algoritmo de *clustering* frequentemente utilizado. Neste algoritmo, as funções de pertença de um indivíduo aos clusters são essencialmente função das distâncias do indivíduo aos centros dos clusters. Esta definição de função de pertença é probabilística, visto que a soma da pertença de um indivíduo a todos os clusters totaliza a unidade. Por outro lado, como esta função de pertença é apenas função da distância, não tem em consideração as variações de densidade que podem ser importantes na definição de clusters.

Este paper tem como objectivo introduzir uma nova forma de determinar as funções de pertença, através de um Self-Organizing Map (SOM), mais concretamente utilizando a respectiva matriz-U.

O SOM é uma rede neuronal não supervisionada, que pode ser considerada uma projecção não linear do espaço n -dimensional dos dados, num espaço geralmente bi-dimensional de output. A partir de um SOM é possível obter uma “matriz-U” onde estão representadas as distâncias entre neurónios vizinhos.

A nossa proposta consiste em treinar um SOM com o conjunto dos dados em estudo, e obter a respectiva “matriz-U”. A partir da matriz-U serão calculadas as novas funções de pertença. A função de pertença de um ponto x a um centróide c será obtida do seguinte modo:

- 1- Mapear os dois pontos para a matriz-U, obtendo as respectivas coordenadas (u_x e u_c)
- 2- Calcular o trajecto T de menor custo Q entre os dois pontos u_x e u_c . O custo é definido como sendo o somatório do valor absoluto das diferenças na matriz U ao longo do trajecto.
- 3- A função de pertença de x a c será $1/(1+Q)$.

Embora computacionalmente dispendioso, este método permite que as funções de pertença deixem de ser função umas das outras, e permite detectar clusters baseados em variações de densidade dos dados. Por exemplo neste método um indivíduo poderá ter uma pertença de 100% a dois clusters, caso não haja diferenças significativas entre a densidade da distribuição entre os três pontos em consideração.

- Bezdek, J. C., Keller, J., Krisnapuram, R., & Pal, M. (1999). *Fuzzy models and algorithms for pattern recognition and image processing*: Kluwer Academic Publishers.
- Kohonen, T. (1995). *Self-organizing maps* (1st ed.). Berlin-Heidelberg: Springer.
- Vesanto, J. (1999). Som-based data visualization. *Intelligent Data Analysis*, 3, 111-126.
- Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3), 586-600.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8, 338-353.