Undersampling Representativo de Classe Dominante por Fator de Qualidade Baseado em Multiplicadores de Lagrange

Paulo Cirino

October 4, 2017

Abstract

1 Introdução

Esse trabalho é fundamento em um algorítimo de *fuzzy clustering*, à ser publicado, que foi criado para acelerar o método *Fuzzy C Means*. Ele atinge seu objetivo por meio da remoção de pontos da bordas com o auxilio de um fator de qualidade, que diz respeito a pertinência de cada amostra à todos os *clusters*.

Os algorítimo de *fuzzy clustering* permitem que uma amostra de um *data set* pertença, ao mesmo tempo, à múltiplos agrupamentos. O nível que uma amostra pertence a cada *cluster* é tradicionalmente chamado de **pertinência** $\mu_i(x_j)$, que é a pertinência da amostra x_j para o *cluster* i.

A função de custo J, associada à problemas de *fuzzy clustering*, pode ser definida em 1.

$$\begin{aligned} &\min && J \\ &\text{sujeito a} && \sum_{k=1}^{c} u_{ik} = 1, \quad k = 1, 2, ..., N \end{aligned} \tag{1}$$

Onde J é definido em 2.

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^2 d_{ik}^2 \tag{2}$$

Nessa situação μ_{ik}, é a pertinência da amostra k em relação ao centro i. Adotando a solução de Multiplicadores de Lagrange, a nova função de custo assume a forma descrita em 3, com derivadas parciais 4 e 5.

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} \left[u_{ik}^2 d_{ik}^2 - \lambda (\sum_{m=1}^{c} u_{mk} - 1) \right] \tag{3}$$

$$\frac{\partial J}{\partial \lambda} = \sum_{m=1}^{c} u_{mk} - 1 : \frac{\partial J}{\partial \lambda} = 0 \implies \sum_{i=1}^{c} u_{ik} = 1 \tag{4}$$

$$\frac{\partial J}{\partial u_{st}} = 2u_{st}d_{st}^2 - \lambda : \frac{\partial J}{\partial u_{st}} = 0 \implies u_{st} = \frac{\lambda}{2d_{st}^2} \tag{5}$$

Assim, a equação 6, representa cada um dos multiplicadores de Lagrange do *data set*.

$$\lambda_k = \frac{2}{\sum_{j=1}^{c} \frac{1}{d_{jk}^2}}, \quad k = 1, 2..., N \tag{6}$$

Assim é possível definir uma medida de qualidade para cada amostra, descrita na equação, 7. A medida q_k de qualidade, é obtida para cada amostra $\mathbf{x}_k$ de $\mathbf{X} = \{x_i \in \mathbb{R} | i = 1...N\}$, e representa uma medida de incerteza da pertinência μ_{ik}.

$$q_k = 1 - c^c \prod_{i=1}^{c} \frac{1}{\mu_{ik}} \tag{7}$$

Substituindo a equação 5 em 7, podemos representar q_k em 9

$$q_k = 1 - \frac{2}{\lambda_k} c^c \prod_{i=1}^{c} d_{ik}^2 \tag{8}$$

2 Método

3 Resultados

4 Conclusão