**Semi-supervised Kernel-Based Fuzzy C-Means**

Zhang et al. (2004) proposed a semi-supervised kernel-based fuzzy c-means algorithm known as , which is based on the kernel-based fuzzy c-means clustering algorithm (KFCM). The is a semi-supervised algorithm, and this paper focuses on applying it for classification tasks, not clustering. The authors compared the proposed algorithm with trained classical classifiers with a small number of labeled data, such as k-nearest neighbor (k-NN) [[1](#_ENREF_1)] And support vector machines (SVM) [[1](#_ENREF_1)]. The algorithm offers several significant advantages. It uses kernel-based methods to handle non-linear data and increases classification accuracy. Moreover, as a semi-supervised algorithm, it effectively uses labeled and unlabeled data, especially when labeled data is scarce.

Suppose there is a lot of labeled and unlabeled data. In this research, every vector is considered in space. Therefore, all labeled and unlabeled data can be displayed as the total matrix.

|  |  |
| --- | --- |
|  | (1) |

In this context, and represent labeled or unlabeled data, respectively. Also, the variables and denote the number of labeled and unlabeled data, respectively. The total number of data points is represented by the relation . In the design of the classifier approach, such as the k-nearest neighbor classifier, only is used to train the classification function, which is then applied to label the variable .

Likewise, the fuzzy c-partition of in Eq. (2) can be represented in matrix form as follows:

|  |  |
| --- | --- |
|  | (2) |

In this case, the value of the component in is beforehand and is generally set to 1 if the data point is labeled with class , and 0 otherwise. From , an initial set of cluster centers can be calculated as follows:

|  |  |
| --- | --- |
|  | (3) |

As a result, the membership value in is updated as follows:

|  |  |
| --- | --- |
|  | (4) |

Finally, the cluster centers are updated as follows:

|  |  |
| --- | --- |
|  | (5) |

The Gaussian kernel is employed for algorithm. The parameter for this kernel method calculated as follows:

|  |  |
| --- | --- |
|  | (6) |

Here, represents the number of clusters, denotes the labeled or unlabeled data, is the total number of data points (both labeled and unlabeled), and is the centroid of the data points.

[1] R. Duda, P. Hart, and D. Stork, "Pattern Classification John Wiley & Sons Inc," *New York,* pp. 630-633, 2001.