***A Novel Semi-Supervised Fuzzy C-Means Clustering Method***

As a center-based clustering algorithm, FCM can be sensitive to the choice of initial centers, often leading to suboptimal solutions by getting stuck in local minima. To address this issue, Kunlun et al. (2009) [[1](#_ENREF_1)] modified the classic FCM by using a semi-supervised method. They use a small amount of labeled data as supervised information to artificially improve the initial center. Based on Basu's concept [[2](#_ENREF_2)], this study introduced a seed set containing some labeled data. First, the author partitioned the seed set and used the center of each partition as the cluster center, optimizing the FCM objective function through the EM algorithm. It is assumed that the seed set represents all K clusters, and there is typically at least one seed point in each cluster.

The Constrained FCM algorithm can be explained in detail as follows:

Let denotes the dataset, and assume represent a subset of seed data (labeled data), where each seed point has a known cluster assignment. The goal is to partition the dataset into K clusters by applying a semi-supervised method. The algorithm follows these steps:

1. Initialize the cluster centers by computing the average of the seed points assigned to each cluster.

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

where *​* is the subset of seed points assigned to cluster .

1. Compute the distance between data points in X and the cluster centers in . Subsequently, calculate the initial objective function value and update the membership matrix .
2. Repeat the process until convergence.

4-a) Update the cluster centers as follows:

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

4-b) Calculate the objective function using:

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

In Eq. (3), is the fuzzifier degree (), and represents the Euclidean distance between th sample and the th cluster center.

4-c) Update membership as follows:

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

[1] L. Kunlun, C. Zheng, C. Liping, and Z. Rui, "A novel semi-supervised fuzzy c-means clustering method," in *2009 Chinese Control and Decision Conference*, 17-19 June 2009 2009, pp. 3761-3765, doi: 10.1109/CCDC.2009.5191706.

[2] S. Basu, A. Banerjee, and R. J. Mooney, "Semi-supervised clustering by seeding," in *Proceedings of the nineteenth international conference on machine learning*, 2002, pp. 27-34.