***Neighborhood information-based semi-supervised fuzzy C-means employing feature-weight and cluster-weight learning***

Jasim et al. (2024) [[1](#_ENREF_1)] introduced a novel semi-supervised fuzzy C-means clustering method to address the limitations identified in the SSFCM algorithm. They designed a new objective function based on the Feature-weighting and Cluster weighting scheme. Integrating class label information into the objective function enhances the algorithm's grasp of the initial cluster count, improves the approximation of initial center vectors, and correctly assigns labeled samples to corresponding clusters. The local weighting of the features gives the most important features more weight, while cluster weighting leads to balanced clusters during the clustering process. Additionally, considering weights for clusters results in reduced sensitivity of initial centers. This approach also applies a spatial penalty term to the membership function, enabling a sample's clustering to be influenced by its neighbors, facilitating the formation of better clusters, and enhancing efficiency. Furthermore, the method employs a kernel metric and avoids the performance impacts of noisy and outlier samples by utilizing a non-Euclidean distance measure.

Although this algorithm performs well in clustering, the increase in computational complexity due to double-weighting mechanisms and sensitivity to parameter settings can be considered among its disadvantages. In addition, the algorithm may struggle with scalability and efficiency on large datasets, potentially limiting its practical applications.

The objective function of the neighborhood information-based semi-supervised fuzzy C-means algorithm is expressed as follows:

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

Subject to

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

In the first term of the objective function, , M, and K represent the number of data samples, the dimensionality of the feature space, and the number of clusters, respectively. The matrix is the membership matrix, where each entry corresponds to the membership degree of the th sample to the *k*th cluster center. The matrix indicates the set of centers associated with the clusters indexed by . In additionally, the matrix assigns weights to features, where denotes the weight of the th feature in the th cluster. Finally, in the vector , each entry represents the weight of the th cluster.

The parameter is constrained within the range and , while the parameter is limited to the range . The specific values of the parameters and are determined by the user. In the second term of the objective function, acts as a regularization parameter that incorporates local neighborhood information, where denotes the set of neighbors for the *n*th sample, and represents the number of these neighbors. In the third term of the objective function is a regularization parameter that integrates class information. Here, is a label indicator vector of length , where if is labeled and 0 otherwise. Additionally, represents the fuzzy membership degrees of the labeled samples, where if sample belongs to class , and otherwise. The expression represents a non-Euclidean distance metric, which is defined as follows:

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

where indicates inverse of the variance of *m*th feature of dataset.

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

By applying the closed-form solution to Eq. (1), the following update equations are derived for , , , and , respectively:

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

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

|  |  |
| --- | --- |
|  | (7) |

|  |  |
| --- | --- |
|  | (8) |

[1] A. K. Jasim, J. Tanha, and M. A. Balafar, "Neighborhood information based semi-supervised fuzzy C-means employing feature-weight and cluster-weight learning," *Chaos, Solitons & Fractals,* vol. 181, p. 114670, 2024/04/01/ 2024, doi: <https://doi.org/10.1016/j.chaos.2024.114670>.