LHC-S3FCM

In 2019, Gan et al. [1] proposed a Local Homogeneous Consistent Safe Semi-Supervised Clustering (LHC-S3FCM) algorithm based on FCM and SSFCM\_s [2], where the prior knowledge is provided in the form of class labels. Within the framework of the LHC-S3FCM algorithm, an initial local graph is systematically constructed to effectively model the interrelationships between labeled sample and its nearest homogeneous unlabeled samples, utilizing insights garnered from unsupervised clustering methodologies. Subsequently, a graph-based regularization expression is formulated to facilitate the alignment of predictions from labeled samples with those of their local homogeneous neighbors. This algorithm effectively mitigates the potential risks associated with labeled samples. Furthermore, LHC-S3FCM positively exploits the labeled samples by restricting the corresponding outputs to be the given class labels when the labeled ones may be helpful.

The advantages of this algorithm in comparison to the existing established semi-supervised clustering methodologies, encompass: (1) The potentially hazardous labeled samples are safely not freely utilized through local homogeneous consistency within this framework. (2) The introduced graph-based regularization term can be seamlessly integrated into other objective function-based semi-supervised clustering approaches. A limitation of this algorithm is the introduction of a novel regularization parameter, denoted as λ2.

This algorithm initially identifies the *p* nearest neighbors of the labeled samples based on the Euclidean distance metric and subsequently determines the homogeneous unlabeled samples within the *p* nearest neighbors utilizing the clustering outcomes derived from FCM. The objective function of LHC-S3FCM algorithm is defined as follow:

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

Subject to:

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

In Eq. (1), refers to the number of data samples, is the number of labeled samples and is the number of clusters. and represent the regularization parameters that govern the balance between FCM and SSFCM. In particular, the latter two components serve to constrain the clustering predictions to align with the specified class labels and the results derived from the homogeneous neighboring instances. also, is the relationship between the labeled and unlabeled samples, which is defined as follows:

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

where signifies the collection of *p* closest neighbors corresponding to , and and , respectively, denote the samples that are labeled and those that remain unlabeled. is the predicted labels obtained by FCM. By resolving the Lagrange equation, the updating equations (Eq. (4) and Eq. (6)) are derived for , and , correspondingly:

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

where and .

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

where and .

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

[1] H. Gan, Y. Fan, Z. Luo, and Q. Zhang, "Local homogeneous consistent safe semi-supervised clustering," *Expert Systems with Applications,* vol. 97, pp. 384-393, 2018/05/01/ 2018, doi: <https://doi.org/10.1016/j.eswa.2017.12.046>.

[2] W. Pedrycz and J. Waletzky, "Fuzzy clustering with partial supervision," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),* vol. 27, no. 5, pp. 787-795, 1997.