***Possibilistic clustering with seeds***

Antoine et al. (2018) [[1](#_ENREF_1)] introduced an innovative semi-supervised clustering method called SPFCM. This algorithm integrates label constraints while maintaining a possibilistic partition. Adopting a possibilistic framework facilitates the articulation of various forms of uncertainty. The benefits of such a framework within SPFCM are twofold: First, constraints are articulated under the guise of the probability of cluster membership. This representation allows an expert to integrate partial information into the clustering algorithms, exceeding the capabilities of all other semi-supervised adaptations of k-means and FCM. Second, creating a possibilistic partition by SPFCM enables comprehensive insights into the data set and simplifies, for example, the identification of outliers. In practical applications, prior information is accessible with different levels of certainty. This investigation proposes using soft label knowledge, which refers to objects assigned a degree of membership in a particular cluster. Such information can be obtained through expert input or automatically through background knowledge.

Let be the a priori possibility that belongs to the cluster and takes values ​​in the interval [0, 1]. This value is equal to 0 if it is certain that the object does not belong to the cluster , while a value of is a high probability that belongs to the cluster , even if different degrees of assumption regarding membership of other clusters are possible.  In the context of the SFPCM algorithm, a fundamental requirement is to ensure that the possibilistic value closely matches . In certain scenarios, restrictions are relaxed to prevent abrupt disruptions in the structure that can lead to inconsistent results. Therefore, the penalty term is integrated into the PFCM algorithm to create a new algorithm called SPFCM. It allows handling soft label restrictions. The objective function of this method is presented as follows:

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

Such that:

|  |  |
| --- | --- |
|  | **(2)** |
|  | **(3)** |

In Eq. (1), the coefficients and are positive. Subject to the provisions in condition (2), and when m is an even integer greater than 1, the variable 0 serves as a trade-off coefficient that balances the unsupervised retrieved intrinsic structure with the considerations imposed by the constraints. The variable allows the selection of constrained values exclusively within the penalty term:

|  |  |
| --- | --- |
|  |  |

The improvement of on is achieved by treating and as fixed constants. Since the penalty term included for SPFCM does not have probabilistic partition values, the adjustment of the membership degrees is carried out using Eq. (4).

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

To minimize objective function with respect to , the variables and are kept constant. The independence of the columns and rows of allows the possibility of updating each individual value separately. By setting to optimize the optimization process, the problem is converted into a quadratic form. The derivative is then calculated:

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

Provided that and , classified as positive semidefinite with respect to . Consequently, the following update for is obtained when the derivative is set to 0.

|  |  |
| --- | --- |
|  |  |

The primary limitation of the SPFCM algorithm pertains to its susceptibility to the quality and quantity of labeled data, which can substantially affect the results of clustering. In instances where the labeled data is inadequate or fails to accurately represent the overall distribution, it may result in suboptimal clustering efficacy.

[1] V. Antoine, J. A. Guerrero, T. Boone, and G. Romero, "Possibilistic clustering with seeds," in *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 8-13 July 2018 2018, pp. 1-7, doi: 10.1109/FUZZ-IEEE.2018.8491655.