FLAME clustering

From Wikipedia, the free encyclopedia

> A **major contributor to this article appears to have a close connection** with its subject. It may require cleanup to comply with Wikipedia's content policies, particularly neutral point of view. Please discuss further on the talk page. *(August 2010)*

Fuzzy clustering by Local Approximation of MEmberships (FLAME) is a data clustering algorithm that defines clusters in the dense parts of a dataset and performs cluster assignment solely based on the neighborhood relationships among objects. The key feature of this algorithm is that the neighborhood relationships among neighboring objects in the feature space are used to constrain the memberships of neighboring objects in the fuzzy membership space.

Contents [hide]

Description of the FLAME algorithm   [edit]

The FLAME algorithm is mainly divided into three steps:

1. Extraction of the structure information from the dataset:
   1. Construct a neighborhood graph to connect each object to its K-Nearest Neighbors (KNN);
   2. Estimate a density for each object based on its proximities to its KNN;
   3. Objects are classified into 3 types:
      1. Cluster Supporting Object (CSO): object with density higher than all its neighbors;
      2. Cluster Outliers: object with density lower than all its neighbors, and lower than a predefined threshold;
      3. the rest.
2. Local/Neighborhood approximation of fuzzy memberships:
   1. Initialization of fuzzy membership:
      1. Each CSO is assigned with fixed and full membership to itself to represent one cluster;
      2. All outliers are assigned with fixed and full membership to the outlier group;
      3. The rest are assigned with equal memberships to all clusters and the outlier group;
   2. Then the fuzzy memberships of all type 3 objects are updated by a converging iterative procedure called *Local/Neighborhood Approximation of Fuzzy Memberships*, in which the fuzzy membership of each object is updated by a linear combination of the fuzzy memberships of its nearest neighbors.
3. Cluster construction from fuzzy memberships in two possible ways:
   1. One-to-one object-cluster assignment, to assign each object to the cluster in which it has the highest membership;
   2. One-to-multiple object-clusters assignment, to assign each object to the cluster in which it has a membership higher than a threshold.

The optimization problem in FLAME   [edit]

The Local/Neighborhood Approximation of Fuzzy Memberships is a procedure to minimize the Local/Neighborhood Approximation Error (LAE/NAE) defined as the following:

$$E(\{p\}) = \sum_{x \in X} \left\| p(x) - \sum_{y \in \mathcal{N}(x)} w_{xy} p(y) \right\|^2$$

where X is the set of all type 3 objects, $p(x)$ is the fuzzy membership vector of object x, $\mathcal{N}(x)$ is the set of nearest neighbors of x, and w_{xy} with $\sum_{y \in \mathcal{N}(x)} w_{xy} = 1$ are the coefficients reflecting the relative proximities of the nearest neighbors.

The NAE can be minimized by solving the following linear equations with unique solution which is the unique global minimum of NAE with value zero:

$$p_k(\boldsymbol{x}) - \sum_{\boldsymbol{y} \in \mathcal{N}(\boldsymbol{x})} w_{\boldsymbol{xy}} p_k(\boldsymbol{y}) = 0, \quad \forall \boldsymbol{x} \in \boldsymbol{X}, \quad k = 1, ..., M$$

where M is the number of CSOs plus one (for the outlier group). The following iterative procedure can be used to solve these linear equations:

$$\boldsymbol{p}^{t+1}(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \mathcal{N}(\boldsymbol{x})} w_{\boldsymbol{xy}} \boldsymbol{p}^t(\boldsymbol{y})$$

A simple illustration on a 2-Dimension testing dataset   [edit]
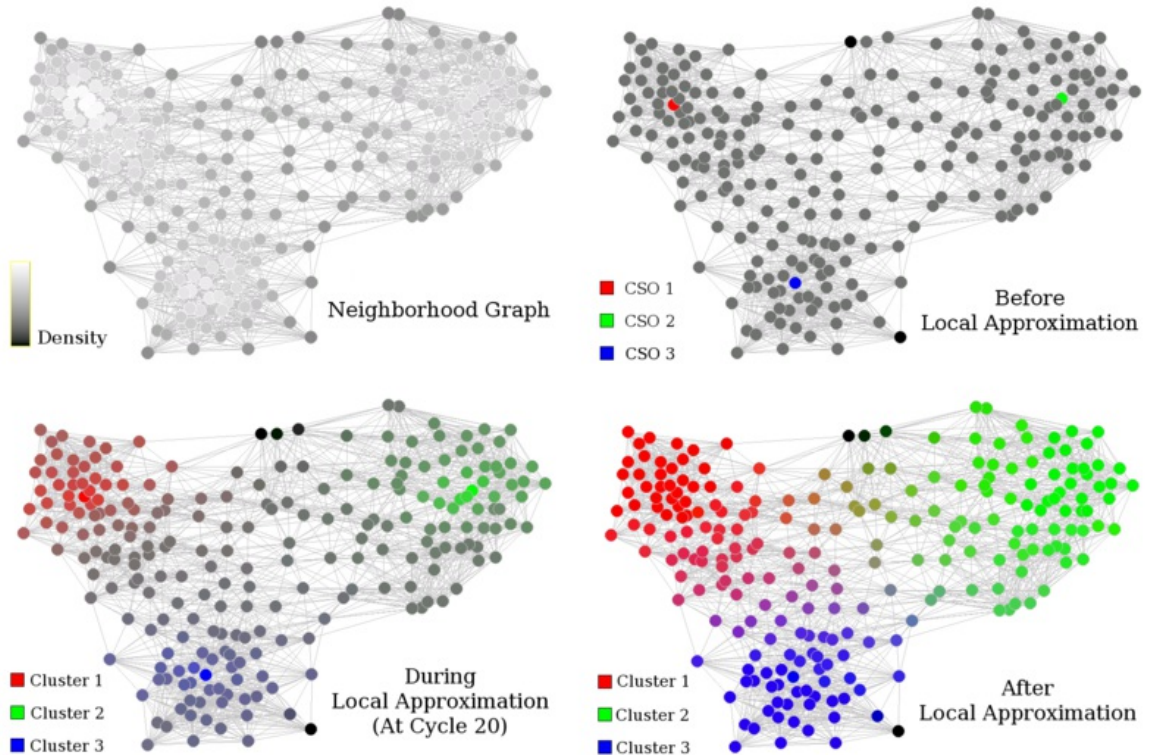
See also   [edit]

- Data clustering
- Fuzzy clustering

External links   [edit]

- BMC Bioinformatics (2007): FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data
- FLAME source codes in C released under FreeBSD-like license on GoogleCode

Categories: Data clustering algorithms