Fişa suspiciunii de plagiat / Sheet of plagiarism's suspicion	Indexat la: 38/01

Opera suspicionată (OS) Suspicious work	Opera autentică (OA) Authentic work

OS	Pruteanu, E., Ababei, S., Culea, G., Puiu, G., Angheluţ, M., ”Objective function based fuzzy clustering”, *Modelling and Optimization in the Machines Building Field MOCM*, vol.1, no.14, 2008, pg. 229-232.
OA	Turčan, A., Ocelíková, E., ”Fuzzy Clustering Applied on Mobile Agent Behaviour Selection”, 2005, Disponibil la: http :// bmf . hu / conferences / SAMI2005 / ocelikova.pdf.

Incidenţa minimă a suspiciunii / Minimum incidence of suspicion	
p.229:28 – p.231:8	p.2:13 – p.3:24
Fişa întocmită pentru includerea suspiciunii în Indexul Operelor Plagiate în România de la www.plagiate.ro	

Precizare:

Prin notaţia p.229:28 - p.231:8 se înţelege că fragmentul de text preluat fără indicarea provenienţei în opera suspicionată este cuprins integral în opera autentică între rândul 13 al pag.2 şi rândul 24 al pag.3.

OBJECTIVE FUNCTION BASED FUZZY CLUSTERING

PRUTEANU E., ABABEI Şt, CULEA G., PUIU GABRIEL, ANGHELUT M

University of Bacău

Abstract: Clustering algorithms, in general, aim to divide unlabeled data set into groups of data, which are similar in some way. While hard clustering algorithms outcome in not overlapping clusters, fuzzy approach to clustering allows each datum to be member of one or more clusters. Goal of this paper is to give overview on most typical fuzzy clustering algorithms.

Keywords: fuzzy, data, clustering, algorithms

1. CLUSTERING

Them main purpose of clustering is to divide a set of objects into reasonable groups (called clusters). The division of objects is based on measuring of similarity or dissimilarity between the pair of objects. Thus, result of clustering is a set of clusters, where object within one cluster are more similar to each other, than to object in another cluster.

Clustering methods are used in many branches ranging from data-mining through market research to remote sensing. Many different approaches were employed in clustering algorithms. These algorithms are divided into several subsets according to any typical property of algorithm.

One of the properties is fuzziness of the algorithm. Using this property, three different types of algorithms are known.

Hard clustering algorithms partition data set into clusters, where one datum belongs only to one cluster. *Fuzzy clustering* algorithms allow each datum to belong into several clusters with different degree of membership. *Possibility clustering* algorithms, like fuzzy clustering algorithms, allow one datum to belong to several clusters with different degree of membership, but it is not guaranteed that datum belongs to any cluster.

2. FUZZY CLUSTERING ALGORITHMS

As said before, fuzzy clustering algorithms allows each datum to belong into one or more clusters with certain membership degree. Larger membership values indicate higher confidence in the assignment of the datum to the cluster.

2.1. General Fuzzy Clustering Algorithm

(1) Select an initial fuzzy partition of the N objects into K clusters by selecting the $N \times K$ membership matrix U. An element u_{ij} of this matrix represents the grade of membership of object $\mathbf{x}_i$ in cluster $\mathbf{c}_j$. Typically, $u_{ij} \in [0,1]$.

(2) Using U, find the value of a fuzzy criterion function, e.g., a weighted squared error criterion function, associated with the corresponding partition. One possible fuzzy criterion function is

$$J(\mathbf{Z},\mathbf{U},\mathbf{V})=\sum_{i=1}^{C}\sum_{k=1}^{N}\mu_{ik}^{m}\left\|z_k-v_i\right\|_{\mathbf{A}_i}^{2}$$

Reassign patterns to clusters to reduce this criterion function value and recompute U.

(3) Repeat step (2) until entries in U do not change significantly.

The most popular fuzzy clustering algorithm is the fuzzy *c*-means (FCM) algorithm. Even though it is better than the hard *k*-means algorithm at avoiding local minima, FCM can still converge to local minima of the squared error criterion. The design of membership functions is the most important problem in fuzzy clustering; different choices include those based on similarity decomposition and centroids of clusters.

2.2. Gustafson-Kessel Algorithm

The Gustafson-Kessel (GK) algorithm is an extension of the FCM algorithm. The GK algorithm employs an adaptive distance norm in order to detect the clusters with the different shapes.

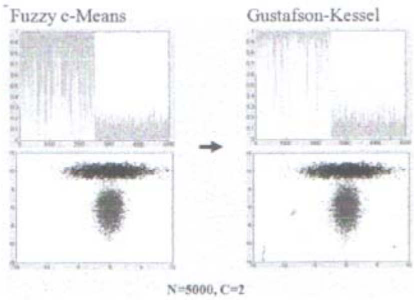

Fig. 2. Gustafson-Kessel Algorithms

Contrary to the FCM algorithm the GK algorithm gives the clusters with the different sizes in the different sizes in the different dimensions (clusters are ellipsoids). In this case every cluster has its own norm matrix A_i. The matrices A_i are an optimization variable in following objective function:

$$J(\mathbf{Z},\mathbf{U},\mathbf{V},A)=\sum_{i=1}^{C}\sum_{k=1}^{N}\mu_{ik}^{m}(z_k-v_i)A_i(z_k-v_i)^{T}$$

But the objective function cannot be minimized directly because J depends on Ai linearly. It means that J can be made as small as the considered positive definite matrices Ai. Therefore Ai is constrained by:

$$\det|A_i|=\rho_i$$

The following result is obtained:

$$A_i=\left[\rho_i\det(F_i)\right]^{1/n}F_i^{-1}$$

where n is the dimension of the data (patterns) and Fi is the fuzzy covariance matrix of the i-th cluster and the fuzzy covariance matrix is defined as:

$$F_i = \frac{\sum_{k=1}^{N} \mu_{ik}^m (z_k - v_i)(z_k - v_i)^T}{\sum_{k=1}^{N} \mu_{ik}^m}$$

Such these cluster shape provides an easier and more accurate coverage of the pattern space and in the following section is shown that this method is better for a generating of the membership functions and the rules.
The GK algorithm is also data scale independent, i.e. if the data in any dimension are multiplied by a constant than the relative coordinates of the cluster centers and the matrix of membership degrees are identical. Moreover the convergence of this algorithm is ensured, but locally only.

2.3. Fuzzy Maximum Likelihood Estimates Clustering

The fuzzy maximum likelihood estimates clustering (FMLE) uses a distance norm based on the fuzzy maximum likelihood estimates:

$$D_{ik} = \frac{[\det(F_i)]^{\frac{1}{2}}}{P_i} e^{\frac{(z_k - v_i)^T F_i^{-1} (z_k - v_i)}{2}}$$

Where F_i is covariance matrix and P_i is the prior probability of the i-th cluster defined as:

$$P_i = \frac{1}{N} \sum_{k=1}^{N} \mu_{ik}$$

The FMLE algorithm is able to detect clusters of the varying shapes, sizes and densities. But the FMLE is more sensitive to an initial condition because it converges to a near local optimum due to the exponential function.

This algorithm also allows us to modify the properties of the results in the same way as GK algorithm. The cluster shapes can be restricted and the volumes of clusters can be different. But the interpretation of the required restrictions of cluster shapes and volumes of clusters is more difficult than for the GK algorithm and it is not so clear. But the using of this modification allows to prevent from very thin or very small or very big clusters.

2.4. Clustering algorithms with linear proto-types

Contrary to the GK and the FMLE algorithms using linear prototypes defined by matrix A, these algorithms use r-dimensional linear of non-linear subspaces of data space as prototypes.

The basic algorithm is called fuzzy c-varieties (FCV). The main idea is to measure the distances of data from r-dimensional linear subspaces of R^n. The criterion is:

$$J(Z,U,V_r) = \sum_{k=1}^{N} \sum_{i=1}^{C} \mu_{ik}^m D_{rik}^2$$

And

$$D_{rik}^2 = \left\| z_k - v_i \right\|^2 - \sum_{j=1}^{r} \left\langle z_k - v_i, v_{ij} \right\rangle^2$$

where V_r is a set of C linear r-dimensional varieties

$V_r = \{V_{r1}, V_{r2}, \ldots, V_{rc}\}$ and D_{ik}

is the orthogonal distance z_k to the linear variety V_{ri}.

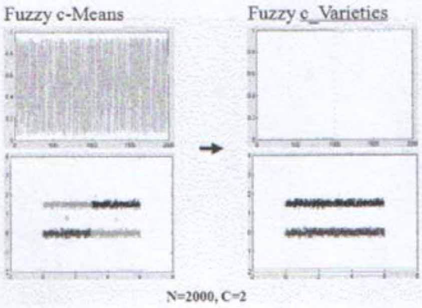

Fig. 2. Fuzzy C-Varieties Algorithm

The fuzzy c-varieties algorithm is the generalization of the FCM and the GK. The main disadvantage of the FCV algorithm is that the size of the linear variety cannot be constrained. Thus the algorithm tends to connect collinear clusters together in spite of that the cluster can be divided very well. The algorithm gives bad results when the size of the varieties is different from cluster to cluster.

3. CONCLUSION

Fuzzy clustering is a powerful tool for clustering data, giving good results. However, all of these algorithms still rely to some extent on rigid assumptions about the data which must be provided by user. These assumptions appear both explicitly as fixed numerical values and implicitly in the way, that user represents the input to the algorithm. While prior knowledge to application can help user choose appropriate assumptions, oftentimes user must alternate between running the clustering algorithm and updating assumptions based on the result.

REFERENCES

[1] J. C. Bezdek, Pattern Recognition With Fuzzy Objective Function. New York: Plenum, 1981.

[2] Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum. New York, 1981

[3] Gath, I., Geva, A.B.: Unsupervised optimal Fuzzy Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, p.773-781

[4] R. N. Dave, " Use of the adaptive fuzzy clustering algorithm to detect lines in digital images," Intell. Robots Comput. Vision VIII, vol. 1192, pt. 2, pp. 600-611, Nov. 1989.

[5] D. E. Gustafson and W.C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in Proc. IEEE Conf. Decision Contr., San Diego, CA, 1979.

[6] *Xuejian, Xiong; Kap Luk, Chan; Kian Lee, Tan* Similarity-driven cluster merging method for unsupervised fuzzy clustering . ACM International Conference Proceeding Series, Proceedings of the 20th conference on Uncertainty in artificial intelligence 2004

MODELLING AND OPTIMIZATION IN THE MACHINES BUILDING FIELD

Volume edited by **ROMANIAN TECHNICAL SCIENCES ACADEMY** and **UNIVERSITY OF BACĂU** in co-operation with **POLITEHNICA UNIVERSITY OF BUCHAREST** and **TECHNICAL UNIVERSITY OF IAŞI**

CHIEF EDITOR:

EDITORIAL BOARD:

SCIENTIFIC REVIEWERS:

Articles appearing in MODELLING AND OPTIMIZATION IN THE MACHINES BUILDING FIELD are abstracted by: PROQUEST (CSA – SUA), VINITI (RUSIA)

See also in MOCM:

- About MOCM
- Issues of MOCM [14]
- Peer Review for MOCM
- Instructions for Authors of MOCM articles
- Browse abstracts for MOCM
- Search in MOCM
- Contact MOCM editorial board