International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 2 Issue 4 ǁ April 2014 ǁ PP.21-25

Robust Fuzzy Data Clustering In An Ordinal Scale Based On A Similarity Measure

Yevgeniy Bodyanskiy[1], Oleksii Tyshchenko[1], Viktoriia Samitova[1]
[1]*(Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, Ukraine)*

ABSTRACT: *This paper is devoted to processing data given in an ordinal scale. A new objective function of a special type is introduced. A group of robust fuzzy clustering algorithms based on the similarity measure is introduced.*
Keywords - *ordinal value, feature vector, ordinal scale, similarity measure, clustering*

I. INTRODUCTION

The clustering task of multidimensional observations is an essential part of data mining, wherein it's assumed in its traditional formulation that each sample feature vector can belong to only one cluster. It's a more common case when a processed feature vector can belong to some classes at the same time with different membership levels. This situation is considered in fuzzy cluster analysis [1-3] which is widely used nowadays in many real-world applications which have to do with medicine, biology, economy, sociology, education, video processing etc.

The initial information to solve a standard fuzzy clustering task is a data sample which consists of N $n-$dimensional feature vectors $X=\{x(1),x(2),\dots x(k),\dots,x(N)\}$, $x(k)\in R^n$, k is an observation number at the "object-property" table. The clustering result is the partition of X into m overlapping classes with some membership levels $w_q(k)$ of the $k-$th feature vector $x(k)$ to the $q-$th cluster. It's recommended to transform all the feature components of the initial data while processed in such a way that they belong to some hypercube, for example $x(k)\in[0,1]^n$. In this way, the initial data takes the form of $x(k)=(x_1(k),\dots,x_i(k),\dots,x_n(k))^T$, $0\le x_i(k)\le 1,\ 1<m<N,\ 1\le q\le m,\ 1\le i\le n,\ 1\le k\le N.$

The situation is much more complicated when the initial data are set in an ordinal scale, i.e. a sequence of ordered linguistic variables like

$$x_i^1,x_i^2,\dots,x_i^j,\dots,x_i^{m_i},\ 1<\dots<j-1<j<j+1<\dots<m_i,$$

where x_i^j is a linguistic variable which corresponds to the $i-$th feature, j stands for a corresponding rank, m_i is a number of such ranks for the $i-$th feature. Wherein a rank number for each feature might be different, for example, there are two gradations in the case of $i=1$: «bad – good», there are three gradations in the case of $i=2$: «bad – satisfactorily – good», there are five gradations in the case of $i=3$: «very bad – bad – satisfactorily – good – excellent» etc. It's clear that usually $m_i<N.$ It should be noticed that people are much more likely to use an ordinal scale, rather than a numeric one.

The simplest way is to substitute ordinal values with their ranks though the loss of essential information is unavoidable because distances between the ranks are permanent and when dealing with linguistic variables it's rather hard to talk about a distance in general.

In this case, it seems more appropriate to preliminary map the initial linguistic variables into a numerical scale with the help of some transformations (fuzzy, frequency etc.) [4-6] with a minimal loss of information and to solve the fuzzy clustering task of numerical data.

It should be noticed that a feature of ordinal values is that values at the edges of the scale can occur less frequently (especially when m_i is rather big) than values which correspond to «average ranks». As a result the «edge» observations can be considered more likely as outliers than normal values. It's appropriate to use robust fuzzy clustering algorithms in this case [7-12] which are based on objective functions (metrics) of a special type which can suppress those outliers. Though it has been mentioned earlier that using a term "distance" for linguistic variables is rather unconvincing, therefore it's much more convenient to use a notion of "similarity" which satisfies "softer" requirements than a metric does.

In this way, a purpose of the current paper is robust fuzzy clustering methods' synthesis to process multidimensional data which are described with feature vectors and their components are ordered linguistic variables based on similarity functions of a special type which can suppress undesirable outliers. The initial information to solve the task is a sample $X = \{x(1), x(2), \dots x(k), \dots, x(N)\}$, where $x(k) = \{x_i^j(k)\}$, $i = 1, 2, \dots, n;\ \ j = 1, 2, \dots, m_i$ is a rank of a specific value of an ordinal variable in the $i-$th coordinate for the $k-$th sample object. The solution is the original sample partition of X into m intersecting classes $(1 < m < N)$ and membership levels' $w_q(k)$ calculation of the $k-$th feature vector to the $q-$th cluster, $1 \le q \le m$.

II. ORDINAL VARIABLES' MAPPING INTO A NUMERICAL SCALE

A linguistic variable $x_i^j(k)$ mapping into a numerical scale can be implemented in the simplest case with the help of the relative frequencies occurrence analysis of the $j-$th rank at the $i-$th feature [13]. If a sample contains N observations and the $j-$th rank is met N_i^j times then relative frequencies can be calculated easily

$$f_i^j = \frac{N_i^j}{N}$$

and then cumulative frequencies are used as linguistic values' estimates

$$F_i^1 = {N_i^1}/{N}, F_i^j = \sum_{l=1}^{j} f_i^l, \ j = 1, 2, \dots, m_i,$$

which means that numerical analogues can be introduced instead of the linguistic values $x_i^j(k)$

$$x_i^j(k) \to \tilde{x}_i(k) = F_i^j.$$

It is clear that the condition $0 \le \tilde{x}_i(k) \le 1$ is naturally fulfilled.

Basically, the relative frequencies can be used as such estimates, i.e. $x_i^j(k) \to \tilde{x}_i(k) = f_i^j$, wherein rarely occurring values $x_i^j(k)$ can be considered as outliers.

III. ROBUST FUZZY DATA CLUSTERING BASED ON A SIMILARITY MEASURE

It has been mentioned before that to solve a fuzzy data clustering task which contains outliers one can use objective functions of a special type. And the task solution has to do with these functions' minimization. It would be much more convenient to use the so-called "similarity measures" (SM) [14] instead of objective functions from a content point of view. The conditions which are used to these similarity measures are softer than those for the metrics:

$$\begin{cases} S(\tilde{x}(k), \tilde{x}(p)) \ge 0, \\ S(\tilde{x}(k), \tilde{x}(p)) = S(\tilde{x}(p), \tilde{x}(k)), \\ S(\tilde{x}(k), \tilde{x}(k)) = 1 \ge S(\tilde{x}(k), \tilde{x}(p)) \end{cases}$$

(the triangle inequality is absent), and the clustering task can be considered as the maximization of these measures.

Fig.1 illustrates the usage of a traditional Gaussian function as a similarity measure with different width parameters $\sigma^2 < 1$.

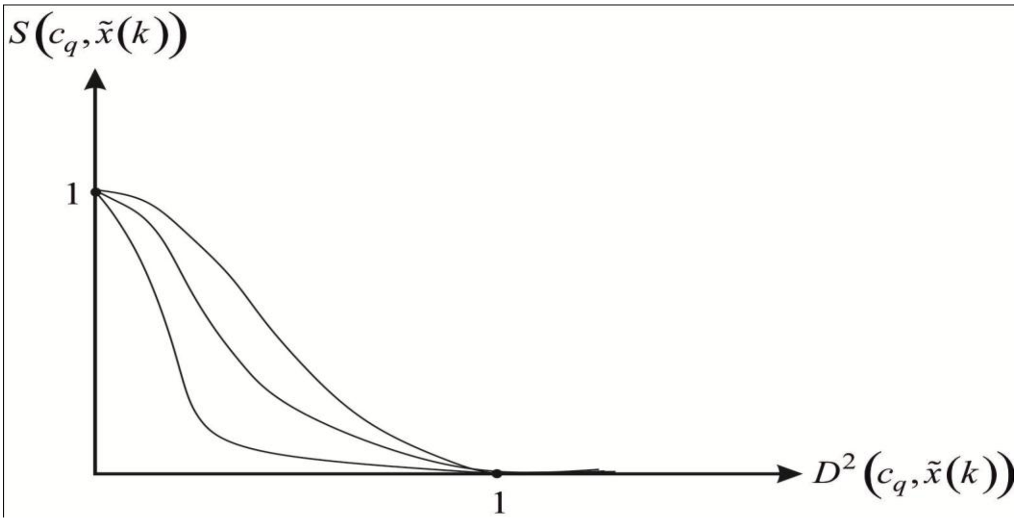

Fig.1. A traditional Gaussian function as a similarity measure

Selecting the width parameter σ^2 of the function

$$S\left(c_q, \tilde{x}(k)\right) = e^{-\frac{\left\|\tilde{x}(k)-c_q\right\|^2}{2\sigma^2}} = e^{-\frac{D^2\left(c_q, \tilde{x}(k)\right)}{2\sigma^2}} \quad (1)$$

(here $c_q - (n \times 1)$ is a coordinates' vector of the $q-$th cluster prototype), one could suppress the influence of widely separated observations from the prototype, that's basically impossible to fulfill with the help of the traditional Euclidean metrics

$$D^2\left(c_q, \tilde{x}(k)\right) = \left\|\tilde{x}(k) - c_q\right\|^2.$$

Taking into consideration the objective function based on the similarity measure (1)

$$E_s\left(w_q(k), c_q\right) = \sum_{k=1}^{N}\sum_{q=1}^{m} w_q^{\beta}(k) S\left(c_q, \tilde{x}(k)\right) = \sum_{k=1}^{N}\sum_{q=1}^{m} w_q^{\beta}(k) e^{-\frac{\left\|\tilde{x}(k)-c_q\right\|^2}{2\sigma^2}}$$

(here $\beta > 0$ is a fuzzifier which is used in the fuzzy clustering theory [2, 3]), standard probabilistic constraints

$$\sum_{q=1}^{m} w_q(k) = 1,$$

the Lagrange function

$$L_S\left(w_q(k), c_q, \lambda(k)\right) = \sum_{k=1}^{N}\sum_{q=1}^{m} w_q^{\beta}(k) e^{-\frac{\left\|\tilde{x}(k)-c_q\right\|^2}{2\sigma^2}} + \sum_{k=1}^{N} \lambda(k)\left(\sum_{q=1}^{m} w_q(k) - 1\right) \quad (2)$$

(here $\lambda(k)$ is an undetermined Lagrange multiplier) and solving the Karush-Kuhn-Tucker system of equations, we get the solution

$$\begin{cases} w_q(k) = \dfrac{S\left(c_q, \tilde{x}(k)\right)^{\frac{1}{\beta-1}}}{\sum_{l=1}^{m} S\left(c_l, \tilde{x}(k)\right)^{\frac{1}{\beta-1}}}, \\ \lambda(k) = -\left(\sum_{l=1}^{m} \beta S\left(c_l, \tilde{x}(k)\right)^{\frac{1}{\beta-1}}\right)^{\beta-1}, \\ \nabla_{c_q} L_S\left(w_q(k), c_q, \lambda(k)\right) = \sum_{k=1}^{N} w_q^{\beta}(k) e^{-\frac{\left\|\tilde{x}(k)-c_q\right\|^2}{2\sigma^2}} \cdot \dfrac{\tilde{x}(k) - c_q}{\sigma^2} = \vec{0}. \end{cases} \quad (3)$$

The last equation of (3) doesn't have any analytical solution, that's why one should use the Arrow-Hurwicz-Uzava procedure to find a saddle point of the Lagrangian function (2). We obtain the algorithm after using this procedure

$$\begin{cases} w_q(k+1) = \dfrac{S\left(c_q(k), \tilde{x}(k+1)\right)^{\frac{1}{\beta-1}}}{\sum_{l=1}^{m} S\left(c_l(k), \tilde{x}(k+1)\right)^{\frac{1}{\beta-1}}}, \\ c_q(k+1) = c_q(k) + \eta(k+1)\nabla_{c_q} L_S\left(w_q(k+1), c_q(k)\right) = \\ = c_q(k) + \eta(k+1) w_q^{\beta}(k+1) e^{-\frac{\left\|\tilde{x}(k+1)-c_q(k)\right\|^2}{2\sigma^2}} \cdot \dfrac{\tilde{x}(k+1) - c_q(k)}{\sigma^2} \end{cases}$$

where $\eta(k+1)$ is a learning rate parameter.

Putting the fuzzifier value $\beta = 2$, we get to a robust fuzzy c-means modification (FCM [1]) based on the similarity measure:

$$\begin{cases} w_q(k+1) = \dfrac{S\left(c_q(k), \tilde{x}(k+1)\right)}{\sum\limits_{l=1}^{m} S\left(c_l(k), \tilde{x}(k+1)\right)}, \\ c_q(k+1) = c_q(k) + \eta(k+1) w_q^2(k+1) e^{-\frac{\left\|\tilde{x}(k+1)-c_q(k)\right\|^2}{2\sigma^2}} \cdot \dfrac{\tilde{x}(k+1)-c_q(k)}{\sigma^2}. \end{cases}$$

Using then the accelerated machine time concept, one could introduce a robust adaptive fuzzy clustering procedure like

$$\begin{cases} w_q^{(\tau+1)}(k+1) = \dfrac{S\left(c_q^{(\tau)}(k), \tilde{x}(k)\right)^{\frac{1}{\beta-1}}}{\sum\limits_{l=1}^{m} S\left(c_l^{(\tau)}(k), \tilde{x}(k)\right)^{\frac{1}{\beta-1}}}, \\ c_q^{(Q)}(k) = c_q^{(0)}(k+1), \\ c_q^{(\tau+1)}(k+1) = c_q^{(\tau)}(k+1) + \eta(k+1)\left(w_q^{(Q)}(k)\right)^{\beta} e^{-\frac{\left\|\tilde{x}(k+1)-c_q^{(\tau)}(k+1)\right\|^2}{2\sigma^2}} \cdot \dfrac{\tilde{x}(k+1)-c_q^{(\tau)}(k+1)}{\sigma^2} \end{cases}$$

where $\tau = 0,1,\ldots,Q$ is the accelerated machine time. This time is so that Q machine time iterations are accomplished when two neighboring observations $\tilde{x}(k)$ and $\tilde{x}(k+1)$ are fed.

A decision on membership of each $\tilde{x}(k)$ to a specific cluster c_q is made according to the similarity measure maximum value.

A robust probabilistic [15] fuzzy clustering algorithm can be performed in the same manner based on the criterion

$$E_s\left(w_q(k), c_q, \mu_q\right) = \sum_{k=1}^{N}\sum_{q=1}^{m} w_q^{\beta}(k) S\left(c_q, \tilde{x}(k)\right) + \sum_{q=1}^{m} \mu_q \left(1 - w_q(k)\right)^{\beta} \tag{4}$$

where a parameter $\mu_q \geq 0$ determines the distance where a membership level takes a value of 0,5, which means that

$$\left\|\tilde{x}(k) - c_q\right\|^2 = \mu_q,$$

then

$$w_q(k) = 0,5.$$

Solving the maximization task (4), we can get

$$\begin{cases} w_q(k+1) = \left(1 + \left(\dfrac{S^{-1}\left(c_q(k), \tilde{x}(k+1)\right)}{\mu_q(k)}\right)\right)^{-1}, \\ c_q(k+1) = c_q(k) + \eta(k+1) w_q^{\beta}(k+1) e^{-\frac{\left\|\tilde{x}(k+1)-c_q(k)\right\|^2}{2\sigma^2}} \cdot \dfrac{\tilde{x}(k+1)-c_q(k)}{\sigma^2}, \\ \mu_q(k+1) = \dfrac{\sum\limits_{p=1}^{k+1} w_q^{\beta}(p) S^{-1}\left(c_q(k+1), \tilde{x}(p)\right)}{\sum\limits_{p=1}^{k+1} w_q^{\beta}(p)}, \end{cases}$$

when $\beta = 2$:

$$\begin{cases} w_q(k+1) = \dfrac{1}{1 + \dfrac{S^{-1}\left(c_q(k), \tilde{x}(k+1)\right)}{\mu_q(k)}}, \\ c_q(k+1) = c_q(k) + \eta(k+1) w_q^2(k+1) e^{-\frac{\left\|\tilde{x}(k+1) - c_q(k)\right\|^2}{2\sigma^2}} \cdot \dfrac{\tilde{x}(k+1) - c_q(k)}{\sigma^2}, \\ \mu_q(k+1) = \dfrac{\sum_{p=1}^{k+1} w_q^2(p) S^{-1}\left(c_q(k+1), \tilde{x}(p)\right)}{\sum_{p=1}^{k+1} w_q^2(p)}. \end{cases}$$

And finally introducing the accelerated time we get the procedure

$$w_q^{(\tau+1)}(k) = \frac{1}{1 + \left(\dfrac{S^{-1}\left(c_q^{(\tau)}(k), \tilde{x}(k)\right)}{\mu_q^{(\tau)}(k)}\right)^{\frac{1}{\beta-1}}},$$

$$\begin{cases} c_q^{(Q)}(k) = c_q^{(0)}(k+1), \\ c_q^{(\tau+1)}(k+1) = c_q^{(\tau)}(k+1) + \eta(k+1)\left(w_q^{(Q)}(k)\right)^{\beta} e^{-\frac{\left\|\tilde{x}(k+1) - c_q^{(\tau)}(k+1)\right\|^2}{2\sigma^2}} \cdot \dfrac{\tilde{x}(k+1) - c_q^{(\tau)}(k+1)}{\sigma^2}, \\ \mu_q^{(\tau+1)}(k) = \dfrac{\sum_{p=1}^{k}\left(w_q^{(\tau+1)}(p)\right)^{\beta} S^{-1}\left(c_q^{(\tau+1)}(k), \tilde{x}(p)\right)}{\sum_{p=1}^{k}\left(w_q^{(\tau+1)}(p)\right)^{\beta}}. \end{cases}$$

IV. Conclusion

The group of robust fuzzy clustering algorithms based on the similarity measure is introduced. These algorithms are designated for multivariate observations' processing. The observations are given in an ordinal scale. This approach is based on the linguistic variables' mapping into a numerical scale and the modification of the well-known fuzzy c-means method which suppresses outliers. The proposed algorithms can be implemented easily. In fact, these algorithms are gradient optimization procedures of a special type objective functions.

References

[1] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (N.Y.: Plenum, 1981).

[2] F. Hoeppner, F. Klawonn, and R. Kruse, *Fuzzy-Clusteranalysen* (Braunschweig: Vieweg, 1997).

[3] F. Hoeppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Clustering Analysis: Methods for Classification, Data Analysis and Image Recognition* (Chichester: John Wiley & Sons, 1999).

[4] R.K. Brouwer, Fuzzy set covering of a set or ordinal attributes without parameter sharing, *Fuzzy Sets and Systems*, *157*, 2006, 1775-1786.

[5] M. Lee, Mapping of ordinal feature values to numerical values through fuzzy clustering, *Proc. on Fuzzy Systems "FUZZ-IEEE 2008" (IEEE World Congress on Computational Intelligence)*, Hong Kong, 2008, 732–737.

[6] Ye.V. Bodyanskiy, V.O. Opanasenko, O.N. Slipchenko, Fuzzy clusterization of the data given in an ordinal scale, *Systemy obrobky informacii*, *4(62)*, 2007, 5-9.

[7] R.N. Dave, R. Krishnapuram, Robust clustering methods: A unified view, *IEEE Trans. on Fuzzy Systems*, *5*, 1997, 270-293.

[8] K. Tsuda, S. Senda, M. Minoh, and K. Ikeda, Sequential fuzzy cluster extraction and its robustness against noise, *Systems and Computers in Japan*, *28*, 1997, 10-17.

[9] Ye. Bodyanskiy, Computational intelligence techniques for data analysis, *"Lecture Notes in Informatics"*, *72*, 2005, 15-36.

[10] Ye. Bodyanskiy, Ye. Gorshkov, I. Kokshenev, and V. Kolodyazhniy, Robust recursive fuzzy clustering algorithms, *Proc. East West Fuzzy Colloquium 2005*, Zittau/Goerlitz, 2005, 301-308.

[11] Ye. Gorshkov, I. Kokshenev, Ye. Bodyanskiy, V. Kolodyazhniy, and O. Shilo, Robust recursive fuzzy clustering-based segmentation of biomedical time series, *Proc. 2006 Int. Symp. on Evolving Fuzzy Systems*, Lancaster, UK, 2006, 101-105.

[12] I. Kokshenev, Ye. Bodyanskiy, Ye. Gorshkov, and V. Kolodyazhniy, Outlier resistant recursive fuzzy clustering algorithm, *Computational Intelligence: Theory and Applications. Advances in Soft Computing*, *38,* 2006, 647-652.

[13] M. Lee, R.K. Brouwer, Likelihood based fuzzy clustering for data sets of mixed features, *Proc. 2007 IEEE Symp. on Foundation of Computational Intelligence*, 2007, 544-549.

[14] J.J. Sepkovski, Quantified coefficients of association and measurement of similarity, *Int. J. Assoc. Math*., *6(2)*, 1974, 135-152.

[15] R. Krishnapuram, J.M. Keller, The possibilistic C-means algorithm: Insights and recommendations, *IEEE Trans. on Fuzzy Systems*, *4*, 1996, 385-393.