clustering

Tomasz PRZYBYŁA*

HYBRID FUZZY CLUSTERING METHOD

A new hybrid clustering method based on a fuzzy myriad is presented. The proposed method could be considered as a generalisation of the well known fuzzy c-means method (FCM) proposed by Bezdek. Existing modifications of the FCM method, such as conditional clustering or partial supervised clustering can be applied to determine the objective function of the proposed method.

1. INTRODUCTION

The clustering aims at assigning a set of objects to clusters in such a way that objects within the same cluster have a high degree of similarity, while objects belonging to different clusters are dissimilar. The clustering methods can be divided into hierarchical and nonhierarchical (partitioning) methods. In this paper, clustering by minimisation of a criterion function will be considered. The most traditional clustering methods are "hard" partitioning i.e. every object belongs to one group. In such a partition boundaries among clusters are sharp. However, in practice, the boundaries are not strict but ambiguous. Thus, soft partitioning is more suitable in this case. However, the fuzzy set theory proposed by Zadeh [1] performs soft partitioning. The most popular method of fuzzy clustering is the fuzzy c-means (FCM) method proposed by Bezdek [2]. Unfortunately the FCM method is sensitive to presence of outliers and noise in clustered data. In real applications, the data are corrupted by noise and assumed models such a Gaussian distribution are never exact. This method is a prototype-based method, where the prototypes are weighted (fuzzy) means. The performance of a linear estimation of prototypes is optimal for the Gaussian model of data distribution. The Gaussian model is inadequate in an impulsive environment. Impulsive signals are more accurately modelled by distributions which density functions have heavier tails than the Gaussian distribution [3, 4].

This paper is divided into four sections. In the section 2, the weighted myriad (trated as fuzzy myriad) and its properties are described. Next, in the subsection 2.2 is introduced an objective function and its optimal arguments: cluster prototypes and partition matrix. The subsection 2.3 discusses the method of estimation of myriad linearity parameter value. The section 3 presents experimental results. Finally, in section 4 conclusions are presented.

* Silesian Technical University

2. HYBRID FUZZY CLUSTERING METHOD (HFCMYR)

2.1. WEIGHTED MYRIAD

Let us consider a set of N independent and identically distributed observations (iid), $\boldsymbol{X} = \{x_1, x_2, \cdots, x_N\}$, and a set of assigned weights $\boldsymbol{U} = \{u_1, u_2, \cdots, u_N\}$. A weighted myriad is the value that minimizes the weighted myriad objective function defined as follows:

$$\hat{\Theta} = \arg\min_{\Theta \in \Re} \sum_{k=1}^{N} \ln\left[K^2 + u_k (x_k - \Theta)^2\right]. \qquad (1)$$

The value of weighted myriad depends on the data set $\boldsymbol{X}$, the assigned weights $\boldsymbol{U}$ and the parameter K, called a linearity parameter. Two interesting cases may occur. First, when the K value tends to infinity (i.e. $K \to \propto$), then the value of weighted myriad converges with the weighted mean, that is

$$\lim_{K \to \infty} \hat{\Theta}_K = \frac{\sum_{k=1}^{N} u_k x_k}{\sum_{k=1}^{N} u_k}, \qquad (2)$$

where $\hat{\Theta}_K = myriad\{x_k \lozenge u_k; K\}_{k=1}^{N}$. This property is called myriad linear property [4].

Second case, called modal property, occurs when the value of K parameter tends to zero (i.e. $K \to 0$). In this case the value of the weighted myriad is always equal to one of the most frequent values in the input data set, i.e.:

$$\hat{\Theta}_0 = \arg\min_{x_j \in \aleph} \prod_{k=1,\, x_k \neq x_j}^{N} \left| x_k - x_j \right|, \qquad (3)$$

where:

$$\hat{\Theta}_K = \lim_{K \to \infty} \hat{\Theta}_K ,$$

and $\aleph$ is a set that contains the most frequent data in the input data set $\boldsymbol{X}$. The value $\hat{\Theta}_K$ is defined in the same way as in the linear property.

2.2. OBJECTIVE FUNCTION

Let's consider a clustering category in which partitions of data set are built on the basis of some performance index, known also as an objective function [7]. The minimization of a certain objective function can be considered as an optimisation approach leading to some suboptimal configuration of the clusters. The main design challenge lies in formulating an objective function that is capable of reflecting the nature of the problem so that its minimization reveals a meaningful structure in the data set.

The proposed method is an objective functional method based on fuzzy c-partitions of a finite data set [2]. The proposed objective function can be an extension of classical within-groups sum of squared error objective function.

Assuming that a different value of K has been assigned to each cluster, the objective function of the proposed method can be described in the following way:

$$J_m(\mathbf{U},\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{N}\sum_{l=1}^{p} \ln\left[K_i^2 + u_{ik}^m (x_k(l) - v_i(l))^2\right], \tag{4}$$

where: K_i^2 is the myriad linear parameter assigned to the ***i***-th cluster and $1 \le i \le c$, $u_{ik} \in \boldsymbol{U}$ is the fuzzy partition matrix, $x_k(l)$ is the l-th feature of the k-th input data set, $v_i(l)$ is the l-th feature of the i-th cluster prototype and *m* is a fuzzy exponent.

The optimisation of the objective function J_m is completed with respect to the partition matrix and prototypes of the clusters. The first step is a constraint-based minimisation, which involves Lagrange multipliers to accommodate the constraints of the membership grades [2, 7].

The columns of partition matrix ***U*** are independent, so the minimisation of objective function (4) can be described as:

$$\begin{aligned} J_m(\boldsymbol{U},\boldsymbol{V}) &= \sum_{k=1}^{N}\sum_{i=1}^{c}\sum_{l=1}^{p} \ln\left[K^2 + u_{ik}^m (x_k(l) - v_i(l))^2\right] \\ &= \sum_{k=1}^{N} J_k \end{aligned} \tag{5}$$

The minimisation problem of (4) can be reduced to minimisation of N indenting components J_k. When a linear transformation $\ell(\cdot)$ is applied to the expression $K^2 + u_{ik}^m (x_k(l) - v_i(l))^2$ its variability range is changed to $(0,2]$, i.e.

$$0 < \ell\left(K^2 + u_{ik}^m (x_k(l) - v_i(l))^2\right) \le 2\,. \tag{6}$$

Based on (6) and representing the logarithm function by its power series, the minimized of the objective function can be reduced to the following expression:

$$J_k' = \sum_{i=1}^{c}\sum_{l=1}^{p}\left[L^2 + u_{ik}^m d_{ik}^2(l) - 1\right], \tag{7}$$

where

$L^2 + u_{ik}^m d_{ik}^2(l) = \ell\left(K^2 + u_{ik}^m (x_k(l) - v_i(l))^2\right)$,

and

$d_{ik}^2(l) = (x_k(l) - v_i(l))^2$.

When the Lagrange multipliers optimization method is applied to the (7) expression we obtain:

$$J'_k(\lambda, \boldsymbol{u}_k) = \sum_{i=1}^{c}\sum_{l=1}^{p}\left[L^2 + u_{ik}^m d_{ik}^2(l) - 1\right] - \lambda\left(\sum_{i=1}^{c} u_{ik} - 1\right), \tag{8}$$

where: λ is the Lagrange multiplier, $\boldsymbol{u}_k$ is the k^{th} column of partition matrix and the term $\lambda\left(\sum_{i=1}^{c} u_{ik} - 1\right)$ arises from definition of the partition matrix $\boldsymbol{U}$ [2,7]. When the gradient of (8) is equal to zero then, for the sets defined as:

$$\begin{aligned} &\underset{1\le k\le N}{\forall}\ \Im_k = \left\{i \mid 1 \le i \le c; \left\|\boldsymbol{x}_k - \boldsymbol{v}_i\right\|^2 = 0\right\} \\ &\tilde{\Im} = \{1,2,\ldots,c\} - \Im_k \end{aligned},$$

the values of partition matrix are described by:

$$\underset{1\le i\le c}{\forall}\ \underset{1\le k\le N}{\forall}\ u_{ik} = \begin{cases} \left[\sum_{j=1}^{c}\left(\dfrac{\left\|\boldsymbol{x}_k - \boldsymbol{v}_i\right\|}{\left\|\boldsymbol{x}_k - \boldsymbol{v}_j\right\|}\right)^{2/(m-1)}\right]^{-1} & if \quad \Im_k = \varnothing \\ 0 & if \quad \underset{i\in\tilde{\Im}_k}{\forall} \\ 1 & if \quad \Im_k \ne \varnothing \end{cases}, \tag{9}$$

where: $\|\bullet\|$ is an Euclidean norm, and $\mathbf{v}_i$ are prototypes $1 \le i \le c$.

For the fixed number of clusters c, parameters m and K and the fixed partition matrix $\boldsymbol{U}$, the prototype values minimizing (4) are fuzzy myriads described as follows:

$$v_i(l) = \arg\min_{\Theta\in\Re} \sum_{k=1}^{N} \ln\left[K^2 + u_{ik}^m\left(x_k(l) - \Theta\right)^2\right], \tag{10}$$

where: i is the cluster number ($1 \le i \le c$), and l is the component (feature) number ($1 \le l \le p$).

2.3. K VALUE ESTIMATION

The α-stable distribution is a generalization of the Gaussian distribution (α=2) or the Cauchy distribution (α=1). So, methods for evaluating parameters of α-stable distribution can be applied for the mentioned distributions.

Assuming, that $\boldsymbol{x}$ is α-stable random variable and $\mathbf{y} = \ln|\mathbf{x}|$, the following dependence can be proofed [5]

$$Var(\mathbf{y}) = \frac{\pi^2}{6}\left(\frac{1}{\alpha^2} + \frac{1}{2}\right), \tag{11}$$

where: $0 \le \alpha \le 2$.

For the Gaussian distribution (α=2), the *K* value should approach infinity. For *K*>50, differences between the fuzzy myriad and the fuzzy mean can be omitted, therefore at the Gaussian point is set at K=50. For the α<1, the fuzzy myriad estimator should be as selective as possible (*K* should approach 0). The following relation between α and *K* values has been proposed [6]

$$K = \sqrt{\frac{\alpha}{2-\alpha}}. \tag{12}$$

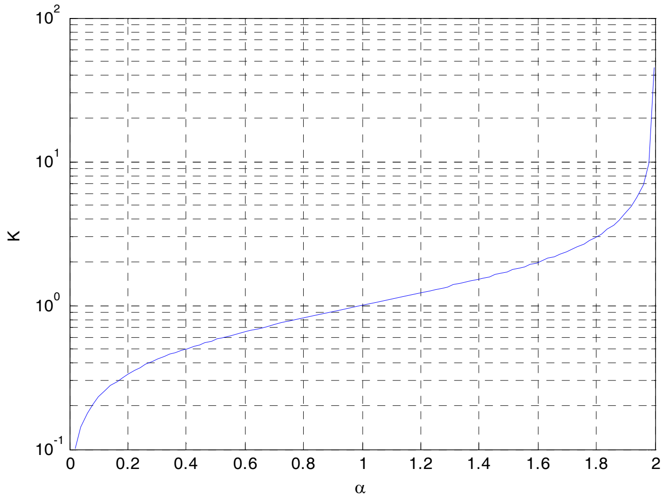

Fig. 1 Proposed $\alpha - K$ curve.

For the data set ***X***, the estimation for each cluster can be done in the following way

1. For the ***i***-th cluster, for each element $1 \le l \le p$ of feature vectors belonging to the ***i***-th cluster, compute the K_l parameter.
2. Finally, the value of K_i is computed as $K_i = \min_{1 \le l \le p} K_l$.

The proposed strategy of the K_i value estimation can be considered as a maximum selectivity strategy.

In some cases, an estimation of the α value based on (11) results incorrect values (i.e. the α is the complex value). In such cases, the value of *K* parameter is fixed at 2.0.

2.4. CLUSTERING DATA WITH HFCMYR CLUSTERING METHOD

The hybrid clustering method can be described as follows:

1. for the given data set $\mathbf{X}=\{\mathbf{x}_1,\ldots\mathbf{x}_N\}$ where $\mathbf{x} \in \Re^p$, fix the number of clusters $c \in \{2,\ldots,N-1\}$, the fuzzyfing exponent $m \in [1,\infty)$ and the tolerance limit ε. Initialise

randomly the partition matrix $\boldsymbol{U}$ and fix initial values of the K parameter for each cluster, fix l=0,

2. calculate the prototype values $\boldsymbol{V}$, as weighted myriads. A weighted myriad has to be calculated for each feature of $\boldsymbol{v}_i$ based on (10),
3. update the partition matrix $\boldsymbol{U}$ using (9),
4. update $K_i, 1 \le i \le c$, based on (11) and (12),
5. if $\left\| \mathbf{U}^{(l+1)} - \mathbf{U}^{(l)} \right\| < \varepsilon$ then stop the clustering algorithm, otherwise $l=l+1$ and go to (2°).

3. NUMERICAL EXPERIMENTS

To illustrate the performance of the proposed method, HFCMyr method is run on two data sets. The first data set is an artificial two-group data in 2D space. The well known IRIS data set proposed by Fisher [8] has been chosen as the second data set.

In all evaluated simulations the fuzzyfier m=2 was used. The Fuzzy C-Means method, proposed by Bezdek [2], has been chosen as the reference method. The Figure 2 shows the first data set. This data set is fully synthetic with well-defined centre of groups in 2D, located at $[0.3 \quad 0.3]^T$ and $[0.8 \quad 0.3]^T$ respectively. The true value group centres are marked by triangles. Outliers have been added at the point $[100.0 \quad 100.0]^T$. The number of outliers varies from 0 (no outliers) up to 32 (number of points when the proposed method has crashed down). The test data set with outliers has not been presented because of its illegibility.

For the synthetic data set, the following values have been fixed: the number of clusters c=2, and the tolerance limit $\varepsilon=10^{-6}$. The results for the proposed method are presented in the table 1, and the results of reference method in the table 2.

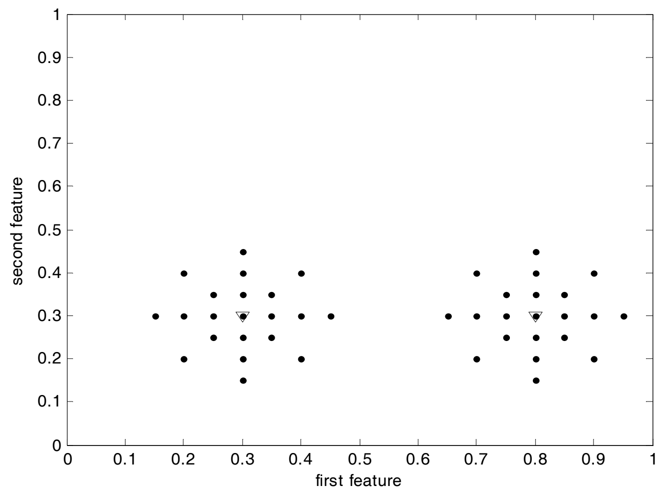

Fig.2 The synthetic data set

Table 1. The prototype values for synthetic data set and the proposed method

HFCMyr				
Number of outliers	$\boldsymbol{V_1}$	K_1	$\boldsymbol{V_2}$	K_2
0	$[0.2987\ 0.3000]^T$	2.0	$[0.8013\ 0.3000]^T$	2.0
5	$[0.3103\ 0.3113]^T$	2.0	$[0.8028\ 0.3011]^T$	0.6454
10	$[0.3219\ 0.3227]^T$	2.0	$[0.8036\ 0.3016]^T$	0.5672
15	$[0.3336\ 0.3342]^T$	2.0	$[0.8044\ 0.3022]^T$	0.5469
20	$[0.3455\ 0.3459]^T$	2.0	$[0.8050\ 0.3029]^T$	0.5423
25	$[0.3578\ 0.3582]^T$	2.0	$[0.8056\ 0.3036]^T$	0.5443
30	$[0.3708\ 0.3711]^T$	2.0	$[0.8058\ 0.3043]^T$	0.5495
32	$[0.5500\ 0.3000]^T$	2.0	$[100.0\ 100.0]^T$	2.0

Table 2. The prototype values for synthetic data set and the reference method

FCM		
Number of outliers	$\boldsymbol{V_1}$	$\boldsymbol{V_2}$
0	$[0.3000\ 0.3000]^T$	$[0.8013\ 0.2987]^T$
5	$[0.5500\ 0.3000]^T$	$[100.0\ 100.0]^T$

For the IRIS data set, the following values have been fixed: the number of clusters $c=3$, and the tolerance limit $\varepsilon=10^{-6}$.

Results are shown in the tables 3 and 4 for the proposed and the reference methods, respectively.

Table 3. Results for IRIS data set and the proposed method

	I	II	III	**K**
$\boldsymbol{X_1}$	50			2.0
$\boldsymbol{X_2}$		45	5	2.0
$\boldsymbol{X_3}$		7	43	2.0

Table 4. Results for IRIS data set and the reference method

	I	II	III
X_1	50		
X_2		47	3
X_3		13	37

The results obtained from the proposed method and the reference method is very similar for the data without outliers. An increase of the outliers improves selectivity of cluster estimation in the proposed method, while in the reference method incorrect (different than expected) results are obtained. The critical number of outliers that causes breaking results from the proposed method is 31, whereas the reference methods break down for data with 5 outliers.

4. CONCLUSIONS

This paper has dealt with clustering of data corrupted by noise and outliers. The well-known existing methods, for example Bezdek's FCM are sensitive on outliers, hence the obtained groups can be different than primary expected. Therefore, methods which results are intuitively correct (the same or very similar to expected) are worth searching for.

Results of the proposed method are more accurate than the reference method outputs. A nonlinear estimation of group prototypes has increased robustness of the clustering method. The cost of the increased robustness and flexibility is the longer computational time.

ACKNOWLEDGEMENTS

The author would like to thank the anonymous referees for their valuable suggestions.

BIBLIOGRAPHY

[1] Zadeh L., A., *Fuzzy sets*, Information and Control 8 (1965), 338-353.
[2] Bezdek, J. C., *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York, (1981)
[3] Arce G. R. Kalluri S., *Fast algorithm for weighted myriad computation by fixed{point search*, IEEE Transactions on Signal Processing 48 (2000), no. 1, 159-171.
[4] Arce, G. R. Kalluri S., *Robust frequency-selective filtering using weighted myriad filters admitting real-valued weights*, IEEE Transactions on Signal Processing 49 (2001), no. 11, 2721-2733.
[5] Georgiu, P. G., Tsakalides, P., Kyriakakis C., *Alpha-Stable Modeling of Noise and Robust Time-Delay Estimation in the Presence of Impulsive Noise*, IEEE Transactions on Multimedia 1 (1999), no. 3, 291-301.
[6] Gonzales, J. G., Arce, G. R., *Optimality of the Myriad Filter in Practical Impulsive-Noise Environments.* IEEE Transactions on Signal Processing 49 (2001), no. 2, 438-441.
[7] Pedrycz W., *Knowledge-Based Clustering*. Wiley-Interscience, (2005).
[8] Fisher R. A., Iris Data Set, http://www.ics.uci.edu/~mlearn/MLRepository.html.