fuzzy clustering, fuzzy meridian, fuzzy myriad, generalized cost function, unsupervised clustering

Tomasz PRZYBYŁA[1], Dawid ROJ[2], Janusz JEŻEWSKI[2], Adam MATONIA[2]

GENERALIZED FUZZY CLUSTERING METHOD

This paper presents a new hybrid fuzzy clustering method. In the proposed method, cluster prototypes are values that minimize the introduced generalized cost function. The proposed method can be considered as a generalization of fuzzy c-means (FCM) method as well as the fuzzy c-median (FCMed) clustering method. The generalization of the cluster cost function is made by applying the L_p norm. The values that minimize the proposed cost function have been chosen as the group prototypes. The weighted myriad is the special case of the group prototype, when the L_p norm is the L_2 (Euclidean) norm. The cluster prototypes are the weighted meridians for the L_1 norm. Artificial data set is used to demonstrate the performance of proposed method.

1. INTRODUCTION

Let us consider a set of objects $\boldsymbol{O}=\{o_1, \dots ,o_N\}$. The object set consists of unlabeled data, i.e. labels are not assigned to objects. The goal of clustering is to find existing subsets in the $\boldsymbol{O}$ set. Objects from one group have a high degree of similarity, while they have a high degree of dissimilarity with objects from other groups. Subsets that are found among objects of the $\boldsymbol{O}$ set are called *clusters* [4,10]. In most cases, each o_i object from the $\boldsymbol{O}$ object set is represented by an x vector in the s-dimensional space, i.e. $x \in \Re^s$. The set $\boldsymbol{X}=\{\boldsymbol{x}_1, \dots ,\boldsymbol{x}_N\}$ is called the object data representation of $\boldsymbol{O}$. In such case, the l^{th} component of the k^{th} feature vector $\boldsymbol{x}_k$ gives a measure of the l^{th} feature (e.g. length of flower petal, age, weight) of the k^{th} object o_k.

One of the most popular clustering method is the fuzzy c-means (FCM) method. In this method, cluster prototypes are computed as the fuzzy means [10]. However, one of the most important inconvenience of the FCM method is its sensitivity to outliers i.e. there are feature vectors which component (or components) have quite different value compared to other feature vectors. There are many modifications for the limitation of the outliers influence. In the first modification, the L_2 norm is replaced by the L_1 norm and by the generalized L_p norm [8]. Another approach has been proposed by Krishnapuram and Keller [12,13]. This clustering approach is based on possibilistic theory instead on fuzzy theory. Another modification has been proposed by Kersten. In this method the L_2 norm is replaced by L_1 norm, and the cluster prototypes are computed as fuzzy medians [11].

The use of heavy tailed distribution to model the impulsive noise gives better results than the use of Gaussian distribution [5,6,7,9]. One of the heavy tailed distribution is the Cauchy distribution, where the location parameter is called (sample) myriad [1]. The fuzzy myriads have been used as the cluster prototypes in the fuzzy c-myriads (FCMyr) clustering method [15]. Another example of the heavy tailed distribution is the Meridian distribution proposed by Aysal and Berner [3]. The location parameter for the Meridian distribution is called (sample) meridian. In the adaptive fuzzy c-meridians (AFCMer) clustering method, the cluster prototypes were computed as fuzzy meridians [16]. The myriad is the maximum likelihood estimator of the location parameter for the Cauchy distribution, so is the meridian for the Meridian distribution. The form of cost function for sample myriad is very similar to the sample meridian cost function. The L_2 norm is used for the myriad cost function, where for the meridian cost function, the L_1 norm is used.

In this paper, the generalized cost function is presented. In the proposed cost function, the L_p norm is used. Assuming p=2 the generalized cost function becomes the myriad cost function, while for p=1 the proposed cost function becomes the meridian cost function. Such a generalization is used to determine the cluster prototypes in the proposed clustering algorithm.

[1] Silesian University of Technology, Institute of Electroncis, Akademicka 16, 44-100 Gliwice, Poland,
[2] Institute of Medical Technology and Equipment, Biomedical Signal Processing Department, Roosevelta 118, 41-800 Zabrze, Poland.

The paper is organized as follows. Section 2 gives the generalized cost function. The proposed clustering algorithm is introduced in Section 3, and Section 4 contains numerical examples. The last section contains some conclusions and ideas for future research.

2. GENERALIZED COST FUNCTION

2.1. FUZZY MYRIAD

The probability density function (PDF) of the Cauchy distribution is described in the following way:

$$f(x;\Theta)=\left(\frac{K}{\pi}\right)\frac{1}{K^2+(x-\Theta)^2}, \tag{1}$$

where: Θ is the location parameter, and K is the scaling factor (*K>0*).

For the given set of N independent and identically distributed (i.i.d.) samples each obeying the Cauchy distribution with the common scale parameter K, the sample myriad $\hat{\Theta}_K$ is a value that minimizes the cost function Ψ_K defined as follows [2]:

$$\begin{aligned}\hat{\Theta}_K &= \arg\min_{\Theta\in\Re}\Psi_K(\mathbf{x};\Theta)\\ &= myriad\left(x_k\,|_{k=1}^{N};\Theta\right)\end{aligned}, \tag{2}$$

where $\Psi_K=\sum_{k=1}^{N}\log\left[K^2+(x_k-\Theta)^2\right]$. By assigning non-negative weights to the input samples, the weighted myriad $\hat{\Theta}_K$ is derived as a generalization of the sample myriad. For the N i.i.d. observations $\{x_k\}_{k=1}^{N}$ and the weights $\{u_k\}_{k=1}^{N}$, the weighted myriad can be computed from the following expression:

$$\begin{aligned}\hat{\Theta}_K &= \arg\min_{\Theta\in\Re}\sum_{k=1}^{N}\log\left[K^2+u_k(x_k-\Theta)^2\right]\\ &= myriad\left(u_k * x_k\,|_{k=1}^{N};\Theta\right)\end{aligned}. \tag{3}$$

The value of weighted myriad depends on the data set ***x***, the assigned weights ***u*** and the parameter K, called a linearity parameter. Two interesting cases may occur. First, when the K value tends to infinity (i.e. $K\to\infty$), the value of weighted myriad converges with the weighted mean, that is:

$$\lim_{K\to\infty}\hat{\Theta}_K=\frac{\sum_{k=1}^{N}u_k x_k}{\sum_{k=1}^{N}u_k}, \tag{4}$$

where: $\hat{\Theta}_K = myriad\left(u_k x_k\,|_{k=1}^{N};K\right)$.

This property is called myriad linear property [1,2]. Second interesting 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 weighted myriad is always equal to one of most frequent values in the input data set, i.e.:

$$\hat{\Theta}_0=\arg\min_{x_j\in\Xi}\prod_{k=1,\,x_k\neq x_j}^{N}\left|x_k-x_j\right|, \tag{5}$$

where: $\hat{\Theta}_0=\lim_{K\to 0}\hat{\Theta}_K$

and Ξ 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. WEIGHTED MERIDIAN

The random variable formed as the ratio of two independent zero-mean Laplacian distributed random variables is referred to as the Meridian distribution [3]. The PDF form of the Meridian distribution is given by:

$$f(x;\delta)=\left(\frac{\delta}{2}\right)\frac{1}{(\delta+|x|)^2}\ , \tag{6}$$

where: δ is the scaling factor (δ > 0).

For the given set of i.i.d. samples $\{x_k\}_{k=1}^N$ each obeying the Meridian distribution with the common scale parameter δ, the sample meridian is given by [3]:

$$\begin{aligned}\hat{\beta}_\delta &= \arg\min_{\beta\in\Re}\sum_{k=1}^{N}\log\left[\delta+|x_k-\beta|\right]\\ &= \arg\min_{\beta\in\Re}\phi_\delta(\mathbf{x};\beta)\\ &= meridian\left(x_k\ |_{k=1}^N;\delta\right)\end{aligned}\ , \tag{7}$$

where ϕ_δ is the sample meridian cost function. Parameter δ is called the medianity parameter.

The sample meridian can be generalized to the weighted meridian by assigning non-negative weights to the input samples. So, the weighted meridian is given by:

$$\begin{aligned}\hat{\beta}_\delta &= \arg\min_{\beta\in\Re}\sum_{k=1}^{N}\log\left[\delta+u_k|x_k-\beta|\right]\\ &= \arg\min_{\beta\in\Re}\phi_\delta(\mathbf{x},\mathbf{u};\beta)\\ &= meridian\left(u_k * x_k\ |_{k=1}^N;\delta\right)\end{aligned}\ . \tag{8}$$

The behavior of the weighted meridian significantly depends on the value of its medianity parameter δ. Two interesting cases may occur. The first case occurs when the value of the medianity parameter tend to infinity (i.e. $\delta\to\infty$), the weighted meridian is equivalent to the weighted median [3]. For the given data set of N i.i.d. samples $x_1, \dots, x_N$ and assigned weights $u_1, \dots, u_N$, the following equation holds true:

$$\begin{aligned}\lim_{\delta\to\infty}\hat{\beta}_\delta &= \lim_{\delta\to\infty} meridian\left(u_k * x_k\ |_{k=1}^N;\delta\right)\\ &= median\left(u_k * x_k\ |_{k=1}^N\right)\end{aligned}\ . \tag{9}$$

This property is called the median property. The second interesting case, called the modal property, occurs when the medianity parameter δ tends to zero (i.e. $\delta\to 0$). In this case, the weighted meridian $\hat{\beta}_\delta$ is equal to the most repeated values in the input data set. Furthermore:

$$\lim_{\delta\to 0}\hat{\beta}_\delta = \arg\min_{x_j\in X_i}\left[\frac{1}{u_j^r}\prod_{k-1,\, x_k\neq x_j}^{N}|x_k-x_j|\right]\ , \tag{10}$$

where: X_i is the set of the frequently repeated values, and r is the number of occurrences of a member of X_i in the sample set.

2.3. GENERALIZED COST FUNCTION

Comparing the properties of the weighted myriad cost function and weighted meridian cost function common features can be found. One of them is the behavior of the both function when the K parameter and the δ parameter tend to zero. Then, for the same data set $\boldsymbol{X}$ and the same weights $\boldsymbol{U}$, the value of weighted myriad is equal to the value of the weighted meridian. Another common feature of both functions is their similar form, but the weighted myriad cost function uses the L_2 norm, while the weighted meridian cost function uses the L_1 norm.
Let the L_p norm be defined as follows:

$$\|\mathbf{z}\|_p = \left(\sum_{l=1}^{s} |z_l|^p \right)^{1/p} , \tag{11}$$

where $\boldsymbol{z}$ is an s-dimensional real vector (i.e. $\mathbf{z} \in \Re^s$).

Applying the L_p norm to the weighted myriad cost function (3) or weighted meridian cost function (8), the generalized cost function can be expressed in the following form:

$$\chi_\gamma^{(p)}(\nu) = \sum_{k=1}^{N} \log\left[\gamma + u_k \|x_k - \nu\|_p\right] \tag{12}$$

where: $\|\bullet\|_p$ is the L_p norm to the p power, and parameter γ corresponds to medianity parameter δ for p=1, and corresponds to linearity parameter K for p=2. It should be mentioned, that for p=1 the γ parameter is equal to medianity parameter δ, but for p=2 parameter γ is equal to the square root of the linearity parameter K.

For the given data set $\{x_k\}_{k=1}^N$ and the assigned weights $\{u_k\}_{k=1}^N$, let the $\hat{\nu}_\gamma$ be the value minimizing the cost function (12), i.e.:

$$\begin{aligned} \hat{\nu}_\gamma &= \arg\min_{\nu\in\Re} \chi_\gamma^{(p)}(\nu) \\ &= \arg\min_{\nu\in\Re} \sum_{k=1}^{N} \log\left[\gamma + u_k \|x_k - \nu\|_p\right] . \end{aligned} \tag{13}$$

Properties of the $\hat{\nu}_\gamma$ are presented in Table 1. The function $\chi_\gamma^{(p)}(\nu)$ can be regarded as a generalized cost function. For p=1 a weighted meridian is a s special case of $\hat{\nu}_\gamma$, and for p=2 the weighted myriad is a special case of $\hat{\nu}_\gamma$.

Table 1. Properties of $\hat{\nu}_\gamma$ estimator.

γ	p=1	p=2
$\gamma \to 0$	most repeated value in the input data set	
$0 < \gamma < \infty$	$\hat{\nu}_\gamma = meridian\left(u_k * x_k \,\|_{k=1}^N ; \gamma\right)$	$\hat{\nu}_\gamma = myriad\left(u_k * x_k \,\|_{k=1}^N ; \sqrt{\gamma}\right)$
$\gamma \to \infty$	$\hat{\nu}_\gamma = median\left(u_k * x_k \,\|_{k=1}^N\right)$	$\hat{\nu}_\gamma = mean\left(u_k * x_k \,\|_{k=1}^N\right)$

Assuming without loss of generality that the weights are in the unit interval (i.e. $u_k \in [0,1]$ where $1 \le k \le N$), the weights can be interpreted as membership degrees. Then, a weighted myriad $\hat{\Theta}_K$ or a weighted meridian $\hat{\beta}_\delta$ can be interpreted as a fuzzy myriad or fuzzy meridian. In the rest paper, the weights will be treated as a membership degrees and the weighted myriad and weighted meridian will be interpreted as fuzzy myriad and fuzzy meridian. Also, the $\hat{v}_\gamma$ value will be interpreted as a fuzzy value.

3. GENERALIZED CLUSTERING METHOD

Let us 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 [1,2,16]. The minimization of a certain objective function can be considered as an optimization approach leading to suboptimal configuration of the clusters. The main design challenge is 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 based on fuzzy c-partitions of the finite data set [4,14]. The suggested objective function can be an extension of the classical functional of within-group sum of an absolute error. The objective function of the chosen method can be described in the following way:

$$J_m^{(p)}(\mathbf{U},\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{N}\sum_{l=1}^{s}\log\left[\gamma + u_{ik}^m \left\| x_k(l) - v_i(l) \right\|_p\right], \tag{14}$$

where: c is the number of clusters, N is the number of the data samples, s is the number of features describing the clustered objects. The γ parameter controls the behavior of cluster prototypes, $u_{ik} \in \mathbf{U}$ is the membership degree of the k^{th} sample to the i^{th} cluster, the $\boldsymbol{U}$ is the fuzzy partition matrix, $x_k(l)$ represents the l^{th} feature of the k^{th} input data from the data set, and m is the fuzzyfing exponent called the fuzzyfier.

The optimization objective function $J^{(p)}{}_m$ is completed with respect to the partition matrix $\boldsymbol{U}$ and the prototypes of the clusters $\boldsymbol{V}$. By minimizing (14) using Lagrangian multipliers, the following new membership u_{ik} update equation can be derived:

$$u_{ik} = \left(\sum_{j=1}^{c} \left(\frac{\left\| \mathbf{x}_k - \mathbf{v}_i \right\|_p}{\left\| \mathbf{x}_k - v_j \right\|_p} \right)^{1/(m-1)} \right)^{-1}, \tag{15}$$

For the case, where $\left\| \mathbf{x}_k - \mathbf{v}_i \right\|_p = 0$, then u_{ik}=1 and u_{jk}=0 for $j \in \{1 \ldots c\}-\{i\}$. For the fixed number of clusters c and the partition matrix $\boldsymbol{U}$ as well as for the exponent m, the prototype values minimizing (14) are the values described as follows:

$$v_i(l) = \arg\min_{v \in \Re} \sum_{k=1}^{N} \log\left[\gamma + u_{ik}^m \left\| x_k(l) - v_i(l) \right\|_p\right], \tag{16}$$

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

3.1. CLUSTERING DATA WITH THE GENERALIZED CLUSTERING METHOD

The proposed clustering algorithm can be described as follows:

1. For the given data set $\boldsymbol{X}=\{\boldsymbol{x}_1\ldots\boldsymbol{x}_N\}$ where $\boldsymbol{x}_k \in \Re^s$, fix the number of clusters $c\in\{2,\ldots, N\}$, the fuzzyfing exponent $m\in[1, \infty)$ and assume the tolerance limit ε. Initialize randomly the parition matrix $\boldsymbol{U}$, fix the value of γ, and set l=0.
2. Calculate the prototype values $\boldsymbol{V}$ for each feature of $\boldsymbol{v}_i$ based on (16),
3. Update the partition matrix $\boldsymbol{U}$ using (15),
4. if $\left\|\mathbf{U}^{(l+1)}-\mathbf{U}^{(l)}\right\|<\varepsilon$ then STOP the clustering, otherwise $l=l+1$ and go to (2).

4. NUMERICAL EXPERIMENTS

In the numerical experiment the fuzzy exponent has been fixed to m=2, and the tolerance limit $\varepsilon=10^{-5}$. The values of the γ parameter have been taken from the set $\gamma\in\{100,10,1,0.1\}$. For a computed set of prototype vectors $\boldsymbol{V}$ the clustering accuracy has been measured as the Frobenius norm distance between the true centers $\boldsymbol{\mu}$ and the prototype vectors. The matrix $\boldsymbol{A}$ is created as $\|\mu-\mathbf{V}\|_F$, where $\|\mathbf{A}\|_F$:

$$\|\mathbf{A}\|_F=\left(\sum_{i,k}A_{ik}^2\right)^{1/2}.$$

Table 2. The Frobenius norm between true centers and prototype vectors obtained for different number of disturbing samples and different norms.

p=1					p=2			
N_i	γ=100	γ=10	γ=1	γ=0.1	γ=100	γ=10	γ=1	γ=0.1
0	0	0	0	0	0.03	0.03	0.03	0.03
10	0	0	0	0	0.07	0.07	0.06	0.06
20	0	0	0	0	0.1	0.1	0.1	0.09
30	0	0	0	0	0.16	0.16	0.15	0.11
40	0.05	0.05	0.05	0.05	0.18	0.18	0.17	0.14
50	0.14	0.14	0.15	0.12	0.32	0.32	0.31	0.24
75	0.08	0.1	0.07	0.07	0.22	0.22	0.22	0.17
100	0.05	0.05	0.05	0.05	0.23	0.23	0.22	0.16
200	0.35	0.17	0.2	0.15	0.34	0.34	0.34	0.3

The purpose of this experiment is to investigate the ability to detect groups in data set. For this purpose, various number of disturbing samples were added to data set with two well-separated groups in 2D space. The center of first group is located at $[0.3\ 0.3]^T$ and the center of the second group is located at $[0.8\ 0.3]^T$. The disturbing samples are realizations of evenly distributed random variable on $[0\ 1]\times[0\ 1]$ intervals. The Figure 1 shows an example of corrupted data set, where the "two cluster" data are plotted as (×) and the corruption samples are plotted as (•).The obtained results are presented in the Table 2, where the left column includes the number of disturbing samples described by N_i.

It can be noticed, that the significant deterioration of clustering results occurs for N_i=200 disturbing samples. For the number of disturbing samples lower than 40 and the L_1 norm , the noise samples do not have an influence on the clustering results. For the L_2 norm, the number of disturbing samples lower than 20 do not have an influence on the obtained results. When the number of disturbing samples is increased, clustering results are wore. Still, they are acceptable, especially for the L_1 norm.

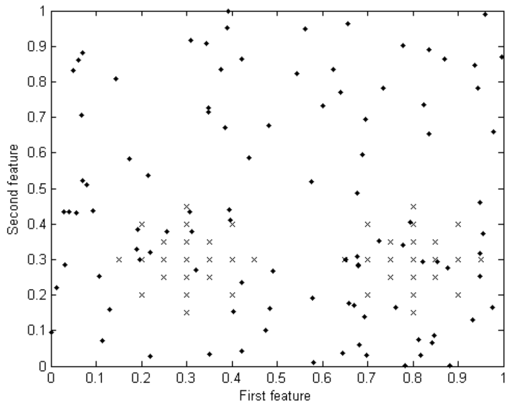

Fig.1. *Hidden groups* of data with 100 corruption samples.

5. CONCLUSIONS

In many cases, the real data are corrupted by noise and outliers. Hence, the clustering methods should be robust for noise and outliers. In this paper the generalized clustering method has been presented. The word *generalized* stands for different cluster estimation which is dependent on two parameters. The proposed method can be treated as a generalization of two clustering method: the fuzzy c-means method and fuzzy c-medians method. The presented generalization of the cost function allows the application of the L_p norm, where $1 \le p \le 2$ or $p \le 1$. In such cases, it is difficult to interpret and identify the value $\hat{v}$. The current work solves the local minima problem and the performance of the cluster centers estimation for large data set.

ACKNOWLEDGEMENT

This work was supported in part by the Ministry of Sciences and Higher Education resources in 2010-2012 under Research Project N N518 4111838.

BIBLIOGRAPHY

[1] ARCE G.R., KALLURI S., Fast algorithm for weighted myriad computation by fixed point search, IEEE Transactions on Signal Processing, Vol. 48, 2000, pp. 159–171.
[2] ARCE G.R., KALLURI S., Robust frequency-selective filtering using weighted myriad filters admitting real-valued weights, IEEE Transactions on Signal Processing, Vol. 49, 2001, pp. 2721–2733.
[3] AYSAL T.C., BARNER K.E., Meridian filtering for robust signal processing, IEEE Transactions on Signal Processing, Vol. 55, 2007, pp. 3949–3962.
[4] BEZDEK, J.C., Pattern recognition with fuzzy objective function algorithms, Published by Plenum Press in New York, 1981.
[5] CHATZIS S., VARVARIGOU Th., Robust fuzzy clustering using mixtures of Student's-t distributions, Pattern Recognition Letters, Vol. 29, 2008, pp.1901–1905.
[6] DAVE R.N., KRISHNAPURAM R., Robust clustering methods: a unified view, IEEE Transactions on Fuzzy Systems, Vol. 5, 1997, pp. 270–293.
[7] FRIGUI H., KRISHNAPURAM R., A robust competitive clustering algorithm with application in computer vision, IEEE Transaction On Pattern Analysis and Machine Intelligence, Vol. 21, 1999, pp. 450–465.
[8] HATHAWAY R.J., BEZDEK J.C., HU Y., Generalized fuzzy c-means clustering strategies using L_p norm distances, IEEE Transactions on Fuzzy Systems, Vol. 8, 2000, pp. 576–582.
[9] HUBER P., Robust statistics, Published by Wiley in New York, 1981.
[10] KAUFMAN L., ROUSSEEUW P., Finding groups in data, Published by Wiley-Interscience, 1990.

[11] KERSTEN P.R., Fuzzy order statistics and their application to fuzzy clustering, IEEE Transactions on Fuzzy Systems, Vol. 7, 1999, pp. 708–712.
[12] KRISHNAPURAM R., KELLER J.M., A possibilistic approach to clustering, IEEE Transactions on Fuzzy Systems, Vol. 1, 1993, pp. 98–110.
[13] KRISHNAPURAM R., KELLER J.M., The possibilistic c-means algorithm: insights and recommendations, IEEE Transactions on Fuzzy Systems, Vol. 4, 1996, pp. 385–396.
[14] PEDRYCZ W., Knowledge-based clustering, Published by Wiley-Interscience, 2005.
[15] PRZYBYŁA T., Fuzzy c-myriad clustering method, System Modeling Control, 2005, pp. 249–254.
[16] PRZYBYŁA T., JEŻEWSKI J., HOROBA K., The adaptive fuzzy meridian and its application to fuzzy clustering, Advances in Intelligent and Soft Computing, Vol. 51, 2009, pp. 247–256.