Hybrid Fuzzy Clustering Using L_P Norms

Tomasz Przybyła[1], Janusz Jeżewski[2], Krzysztof Horoba[2], and Dawid Roj[2]

[1] Silesian University of Technology,
Institute of Electronics,
Akademicka 16, 44-101 Gliwice, Poland
Tomasz.Przybyla@polsl.pl
[2] Institute of Medical Technology and Equipment,
Biomedical Signal Processing Department,
Roosvelta 118, 41-800 Zabrze, Poland

Abstract. The fuzzy clustering methods are useful in the data mining applications. This paper describes a new fuzzy clustering method in which each cluster prototype is calculated as a value that minimizes introducted generalized cost function. The generalized cost function utilizes the L_p norm. The fuzzy meridian is a special case of cluster prototype for $p = 2$ as well as the fuzzy meridian for $p = 1$. A method for the norm selection is proposed. An example illustrating the performance of the proposed method is given.

1 Introduction

The goal of clustering is to find existing subsets in a set of objects $\mathbf{O} = \{o_1, \ldots, o_N\}$. The object set consists of unlabeled data, i.e. labels are not assigned to objects. 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 the objects of the $\mathbf{O}$ set are called *clusters* [1], [2].

In most cases, each o_i object from the $\mathbf{O}$ object set is represented by an $\mathbf{x}$ vector in the s-dimensional space i.e. $\mathbf{x} \in \mathbb{R}^s$. The set $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ is called the object data representation of $\mathbf{O}$. In such a case, the l-th component of the k-th feature vector $\mathbf{x}_k$ gives a measure of l-th feature (e.g. length of flower petal, age of car, 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 fuzzy means [2]. However, one of the most important inconvenience of the FCM method is its sensitivity to outliers i.e. there are feature vectors which 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 [3]. Another approach has been proposed by Krishnapuram and Keller [4], [5]. This clustering approach is based on possibilistic theory instead of the fuzzy sets theory. One of the most interesting modification has been proposed by Kersten. In this method the L_2 norm is replaced by the L_1 norm, and the cluster prototypes are computed as fuzzy medians [6].

N.T. Nguyen, C.-G. Kim, and A. Janiak (Eds.): ACIIDS 2011, LNAI 6591, pp. 187–196, 2011.

On the other hand, clustering method should be robust for data corrupted by outliers or (and) heavy–tailed distributed noise. The heavy tailed distribution is more suitable to model the impulsive noise than the Gaussian distribution [7], [8], [9], [10], [11]. One of the havy tailed distribution is the Cauchy distribution, where the location parameter is called the (sample) myriad [12]. The fuzzy myriads have been used as the cluster prototypes in the fuzzy c–myriads (FCMyr) clustering method [13]. Another example of the heavy tailed distribution is the Meridian distribution proposed by Aysal and Berner [14]. The location parameter for the Meridian distribution is called the (sample) meridian. In the adaptative 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. It is important that the Meridian distribution has heavier tails than the Cauchy distribution. Therefore, the Meridian distribution better describes the impulsive noise. The form of the cost function for a 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 generalized cost function is used to determine the cluster prototypes in the proposed clustering algorithm.

The paper is organized as follows. Section II gives the generalized cost function. The proposed clustering algorithm is introduced in Section III. Section IV illustrates numerical examples. The last section contains some conclusions and ideas for future research.

2 Generalized Cost Function

2.1 Weighted Myriad Cost Function

For the Cauchy distribution, the displacement parameter is called sample myriad. For the given set of N independent and identically distributed (i.i.d.) samples each obeying the Cauchy distribution with common scale parameter, the sample myriad is a value that minimizes the cost function Ψ_K defined as follows [17]:

$$\begin{aligned}\hat{\Theta}_K &= \arg\min_{\Theta\in\mathbb{R}} \Psi_K(\mathbf{x};\Theta) \\ &= \arg\min_{\Theta\in\mathbb{R}} \sum_{k=1}^{N} \log\left[K^2 + (x_k - \Theta)^2\right]\end{aligned}, \tag{1}$$

where: Θ is the location parameter, and K is the scale parameter. 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 $\{u_k\}_{k=1}^N$, the weighted myriad can be computed from the following expression

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

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

$$\lim_{K \to \infty} \hat{\Theta}_K = \frac{\sum_{k=1}^{N} u_k x_k}{\sum_{k=1}^{N} u_k} \ , \tag{3}$$

where $\hat{\Theta}_K = myriad\left\{u_k * x_k|_{k=1}^{N}; K\right\}_{k=1}^{N}$. This property is called myriad linear property [12], [17].

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 most frequent values in the input data set.

2.2 Weighted Meridian Cost Function

The random variable formed as the ratio of two independent zero–mean Laplacian distributed random variables is referred to as the Meridian distribution [14]. For the given set of N i.i.d. samples $\{x_k\}_{k=1}^{N}$ each obeying the Meridian distribution with the common scale parameter δ, the sample meridian $\hat{\beta}_\delta$ is given by [14]:

$$\begin{aligned}\hat{\beta}_\delta &= \arg \min_{\beta \in \mathbb{R}} \phi_\delta(\mathbf{x}; \beta) \\ &= \arg \min_{\beta \in \mathbb{R}} \sum_{k=1}^{N} \log\left[\delta + |x_k - \beta|\right]\end{aligned} \ , \tag{4}$$

where Φ_δ is the sample meridian cost function.

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

$$\begin{aligned}\hat{\beta}_\delta &= \arg \min_{\beta \in \mathbb{R}} \phi_\delta(\mathbf{x}, \mathbf{u}; \beta) \\ &= \arg \min_{\beta \in \mathbb{R}} \sum_{k=1}^{N} \log\left[\delta + u_k |x_k - \beta|\right] \cdot \\ &= meridian\left\{u_k * x_k|_{k=1}^{N}; \delta\right\}\end{aligned} \tag{5}$$

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 tends to infinity (i.e. $\delta \to \infty$), the weighted meridian is equivalent to the weighted median [14]. For the given data set of N i.i.d. samples $x_1, \cdots, x_N$ and assigned weights $u_1, \cdots, u_N$, the following equation holds true

$$\lim_{\delta \to \infty} \hat{\beta}_\delta = \lim_{\delta \to \infty} meridian\left\{u_k * x_k|_{k=1}^N; \delta\right\} = median\left\{u_k * x_k|_{k=1}^N\right\} \quad . \tag{6}$$

This property is called the median property. The second interesting case, called the modal property, occurs when the medianity parameter δ tends to zero. In this case, the weighted meridian $\hat{\beta}_\delta$ is equal to one of the most repeated values in the input data 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 cost function when the K parameter and the δ parameter tend to zero. Then, for the same data set $\mathbf{X}$, the value of weighted myriad is equal to the value of the weighed 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

$$||\mathbf{z}||_p = \left(\sum_{l=1}^{s} |z_l|^p\right)^{\frac{1}{p}} \quad , \tag{7}$$

where $\mathbf{z}$ is an s–dimensional real vector (i.e. $\mathbf{z} \in \mathbb{R}^s$). Applying the L_p norm to the weighted myriad cost function (2) or weighted meridian cost function (5), the generalized cost function can be expressed in the following form

$$\chi_\gamma^{(p)}(\nu) = \sum_{k=1}^{N} \log\left[\gamma + u_k||x_k - \nu||_p\right] \quad , \tag{8}$$

where $||\cdot||_p$ 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 (i.e. $\gamma = \sqrt{K}$).

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 (8), i.e.

$$\begin{aligned} \hat{\nu}_\gamma &= \arg\min_{\nu \in \mathbb{R}} \chi_\gamma^{(p)}(\nu) \\ &= \arg\min_{\nu \in \mathbb{R}} \sum_{k=1}^{N} \log\left[\gamma + u_k||x_k - \nu||_p\right] \end{aligned} \quad . \tag{9}$$

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

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

Properties of the $\hat{\nu}_\gamma$ value 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 special case of $\hat{\nu}_\gamma$, and for $p=2$ the weighted myriad is a special case of $\hat{\nu}_\gamma$.

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, respectively. In the rest of this 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 $\hat{\nu}_\gamma$ value will be interpreted as a fuzzy value.

2.4 Selection the L_p Norm

One of the most popular method for the empirical probability function estimation is the Parzen method [15]. For the given data set of N i.i.d. samples $x_1, \cdots, x_N$, the empirical probability density function (PDF) can be computed as follows

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{h} K\left(\frac{x - x_i}{h}\right), \tag{10}$$

where N is the number of samples, h is the smooth parameter, and $K(\cdot)$ is the kernel function.

Introducting non–negative cost function Ψ as the measure of fit the empirical PDF $\hat{f}$ to the generalized cost function χ, the L_p norm can be determined from

$$\begin{aligned} p &= \arg \min_{p \in \{1,2\}} \min_{x \in X_i} \Psi_p(x) \\ &= \arg \min_{p \in \{1,2\}} \sum_{k=1}^{N} \|\hat{f}(x) - f_p(x;\gamma)\|_2 \end{aligned}, \tag{11}$$

where

$$f_p(x;\gamma) = \begin{cases} \left(\frac{\gamma}{2}\right) \frac{1}{(\gamma + |x|)^2} & \text{if } p=1, \\ \left(\frac{\gamma}{\pi}\right) \frac{1}{\gamma^2 + x^2} & \text{if } p=2. \end{cases}$$

For $p=1$, function $f_p(x;\gamma)$ describes the Meridian distribution and for $p=2$ describes the Cauchy distribution .

The method of the L_p norm determination can be described as follows:

1. For the input data samples $x_1, x_2, \ldots, x_N$, fix the the kernel function $K(\cdot)$, the smooth parameter h, and the γ parameter,
2. For $p = 2$ compute the myriad based on (1) and compute the value of function $\Psi_2(x)$,
3. For $p = 1$ compute the meridian based on (4) and compute the value of function $\Psi_1(x)$,
4. The $L_p = L_2$ norm if $\Psi_2(x) < \Psi_1(x)$; otherwise the $L_p = L_1$ norm.

3 Hybrid 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 [2], [18]. 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 [2],[18]. 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

$$J_m^{(p)}(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{c} \sum_{k=1}^{N} \sum_{l=1}^{s} \log\left[\gamma + u_{ik}^m ||x_k(l) - v_i(l)||_p\right] \quad , \tag{12}$$

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–sample to the i–th cluster, the $\mathbf{U}$ 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 $J_m^{(p)}$ is completed with respect to the partition matrix $\mathbf{U}$ and the prototypes of the clusters $\mathbf{V}$. By minimizing (12) using Lagrangian multipliers, the following new membership u_{ik} update equation

$$u_{ik} = \left(\sum_{j=1}^{c} \left(\frac{||\mathbf{x}_k - \mathbf{v}_i||_p}{||\mathbf{x}_k - \mathbf{v}_j||_p} \right)^{1/(m-1)} \right)^{-1} \quad , \tag{13}$$

can be derived. For the case, where $||\mathbf{x}_k - \mathbf{v}_i||_p = 0$, then $u_{ik} = 1$ and $u_{jk} = 0$ for $j \in \{1 \cdots c\} - \{i\}$.

For the fixed number of clusters c and the partition matrix $\mathbf{U}$ as well as for the exponent m, the prototype values minimizing (12) are the values described as follows

$$v_i(l) = \arg \min_{\nu \in \mathbb{R}} \sum_{k=1}^{N} \log \left[\gamma + u_{ik}^m ||x_k(l) - \nu||_p \right] \quad , \tag{14}$$

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

3.1 Clustering Data with the Hybrid Clustering Method

The proposed hybrid clustering method can be described as follows:

1. For the given data set $\mathcal{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$ where $\mathbf{x}_i \in \mathbb{R}^s$, fix the number of clusters $c \in \{2, \cdots, N\}$, the fuzzyfing exponent $m \in [1, \infty)$ and assume the tolerance limit ε. Initialize randomly the partition matrix $\mathbf{U}$ and fix the value of parameter γ, fix $l = 0$,
2. Select the appropriate L_p norm for each cluster and each feature based on (11),
3. for the obtained L_p norms calculate the prototype values $\mathbf{V}$ for each feature of $\mathbf{v}_i$ based on (14),
4. update the partition matrix $\mathbf{U}$ using (13),
5. if $\left\| \mathbf{U}^{(l+1)} - \mathbf{U}^{(l)} \right\| < \varepsilon$ then STOP the clustering algorithm, otherwise $l = l + 1$ and go to (3).

4 Numerical Experiments

In the numerical experiments the fuzzfing exponent has been fixed to $m = 2$, and the tolerance limit $\varepsilon = 10^{-5}$, and as tke kernel function the Gaussian kernel was chosen. For a computed set of prototype vectors $\mathbf{V}$ the clustering accuracy has been measured as the Frobenius norm distance between the true centers μ and the prototype vectors. The matrix $\mathbf{A}$ is created as $||\mu - \mathbf{V}||_F$, where $||\mathbf{A}||_F$:

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

4.1 Selection the L_p Norm

The purpose of this experiment is to investigate the proposed method of the L_p selection. Two artifical data set have been generated. The first data set includes noise with outliers. Figure 1 shows the data set and the shapes of empirical PDF and $f_p(x; \gamma)$ function. Figure 2 shows the second data set. This data set includes noise without outliers.

For the both data set, the values of the fit function Ψ_p for $\gamma = 3$ are presented in Table 2. It can be seen, that for the data set with outliers, the Meridian distribution better describes the data set than the Cauchy distribution. Generally, this means, that the sample meridian is a more robust estimator than the myriad estimator. This confirms the fact, that the Mridian distribution is better suited for impulsive noise. In the other hand, for data set without outliers, the Cauchy distribution gives better fit to the empirical PDF than the Maridian distribution.

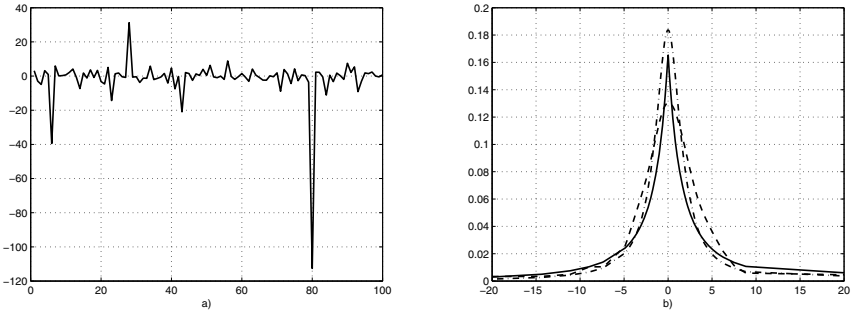

Fig. 1. a) Data set with noise and outliers, b) the Meridian distribution (solid line), the Cauchy distribution (dotted line) and the empirical PDF (dashed line)

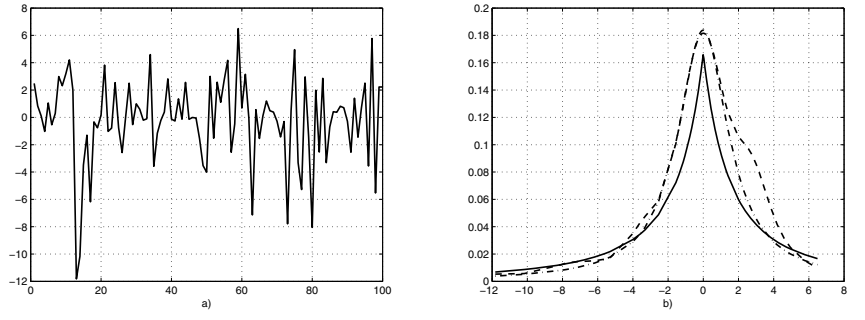

Fig. 2. a) Data set with noise without outliers, b) the Meridian distribution (solid line), the Cauchy distribution (dotted line) and the empirical PDF (dashed line)

Table 2. Values of the $\Psi_p(x)$ function for $\gamma = 3$

data set	$\Psi_1(x)$	$\Psi_2(x)$
with outliers	0.0447	0.0703
without outliers	0.1347	0.0289

4.2 Data Clustering

The example involves three heavy–tailed and overlapped groups of data. The whole data have been generated by a pseudo–random generator. The data set includes 3 groups described by different distributions. The true group centers are: $\mu_1 = [-5, 5]^T$, $\mu_2 = [0, 0]^T$, and $\mu_3 = [10, 0]^T$. The first data set includes overlaping groups generated by Gaussian distributed random generator. In the second data set, the first group data were generated by the Cauchy distributed random generator, while the two others were generated by the Gaussian random generator. In the third data set, the first group data were generated by the Gaussian distributed random generator, while the two other groups were generated

Table 3. The difference among computed cluster centers $\mathbf{V}$ and the true centers μ

	Data set		
γ	I	II	III
1.0	0.6070	2.4653	0.7283
2.0	0.5198	1.0037	0.5852
5.0	0.4792	1.0566	0.5318
10.0	0.4722	1.1900	0.5199
20.0	0.4706	1.4435	0.4953

by the Cauchy distributed random generator. The obtained results for different values of the γ parameter are presented in Table 3.

Small values of parameter γ affect the selectivity of the determination of the cluster prototypes. The selected norm does not influence the results because the for small values of γ the $\hat{\nu}$ value tends to most frequent sample in the data set. For the used data sets, the best results were obtained for $1 < \gamma < 10$.

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 hybrid clustering method is presented. The word *hybrid* stands for different cluster estimation which is dependent on two parameters. The proposed method can be treated as a generalization of two clustering methods: the fuzzy c–means method and the fuzzy c–medians method.

The presented generalization of the cost function allows the application of the L_p norm, where $1 < p < 2$ or $p < 1$. In such cases, it is difficult to interpret and identify the value $\hat{\nu}$.

The current work solves the local minima problem and the performance of the cluster centers estimation for large data sets.

Acknowledgment

This work was partially supported by the Ministry of Science and Higher Education resources in 2010–2012 under Research Project NN518 411138.

References

1. Kaufman, L., Rousseeuw, P.: Finding Groups in Data. Wiley–Interscience, Chichester (1990)
2. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York (1981)
3. Hathaway, R.J., Bezdek, J.C., Hu, Y.: Genalized Fuzzy c-Means Clustering Strategies Using L_p Norm Distances. IEEE Trans. on Fuzzy Sys. 8, 576–582 (2000)
4. Krishnapuram, R., Keller, J.M.: A Possibilistic Approach to Clustering. IEEE Trans. on Fuzzy Sys. 1, 98–110 (1993)

5. Krishnapuram, R., Keller, J.M.: The Possibilistic C–Means Algorithm: Insights and Recomendations. IEEE Trans on Fuzzy Sys. 4, 385–396 (1996)
6. Kersten, P.R.: Fuzzy Order Statistics and Their Application to Fuzzy Clustering. IEEE Trans. on Fuzzy Sys. 7, 708–712 (1999)
7. Huber, P.: Robust statistics. Wiley, New York (1981)
8. Dave, R.N., Krishnapuram, R.: Robust Clustering Methods: A Unified View. IEEE Trans. on Fuzzy System 5, 270–293 (1997)
9. Chatzis, S., Varvarigou, T.: Robust Fuzzy Clustering Using Mixtures of Student's–t Distributions. Pattern Recognition Letters 29, 1901–1905 (2008)
10. Frigui, H., Krishnapuram, R.: A Robust Competitive Clustering Algorithm With Applications in Computer Vision. IEEE Trans. Pattern Analysis and Machine Intelligence 21, 450–465 (1999)
11. Sun, J., Kaban, A., Garibaldi, J.M.: Robust mixture clustering using Pearson type VII distribution. Pattern Recognition Letters 31, 2447–2454
12. Arce, G.R., Kalluri, S.: Fast Algorithm For Weighted Myriad Computation by Fixed Point Search. IEEE Trans. on Signal Proc. 48, 159–171 (2000)
13. Przybyła, T.: Fuzzy c–Myriad Clustering Method. System Modeling Control, 249–254 (2005)
14. Aysal, T.C., Barner, K.E.: Meridian Filtering for Robust Signal Processing. IEEE Trans. on Signal Proc. 55, 3949–3962 (2007)
15. Parzen, E.: On Estimation Of A Probability Density Function And Mode. Ann. Math. Stat. 33, 1065–1076 (1962)
16. Przybyła, T., Jeżewski, J., Horoba, K.: The Adaptive Fuzzy Meridian and Its Application to Fuzzy Clustering. In: Advances in Intelligent and Soft Computing, vol. 57, pp. 247–256. Springer, Heidelberg (2009)
17. Arce, G.R., Kalluri, S.: Robust Frequency–Selective Filtering Using Weighted Myriad Filters admitting Real–Valued Weights. IEEE Trans. on Signal Proc. 49, 2721–2733 (2001)
18. Pedrycz, W.: Konwledge–Based Clustering. Wiley–Interscience, Chichester (2005)