Chapter 3.

Andrzej CHMIELOWIEC[1]

IMPLEMENTATION OF THE MOUNTAIN CLUSTERING METHOD AND COMMENTS ON ITS PRACTICAL USE FOR DETERMINING CLUSTER CENTRES

Abstract

For certain applications a need arises to reduce a large set of measurement data and select a group of the most representative data. Such a situation occurs for example in the case of the fuzzy logic algorithms whose computational complexity makes them inapplicable to too large input sets. One of the methods to reduce a data set is to determine the centre of clusters, that is the elements being the optimum representation of the entire set. The purpose of this paper is to describe the operation of the potential method designed to locate the centres of clusters in the entire set of measurement data. We present the selection algorithm based on the assumption that in certain local environments data are normally distributed. This assumption proves to be correct in numerous practical applications; however, in some cases a different probability distribution may seem more appropriate. For these cases we will only hint at how one can try to modify the potential function to produce the most reliable effect. Along with the mathematical description of the method we also present the functionality of a dedicated software implemented for this purpose.

Keywords:
statistics, data analysis, measurement data processing, cluster estimation, software for cluster estimation

1. Introduction

Over the recent years, methods of acquiring large quantities of measurement data have developed significantly. The processing of ever larger data sets poses a major challenge for statistical analysis, inference and machine learning

[1*] Faculty of Mechanics and Technology, Rzeszow University of Technology, al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland, e-mail: achmie@prz.edu.pl

algorithms. One approach assumes that enormous data sets are partitioned into subsets called clusters represented by individual elements - cluster centres.

Section two presents the mountain clustering method designed to locate cluster centres. Section three discusses how this method is linked with statistics and probability, and provides guidelines on how to prepare measurement data properly. Finally, section four presents the results of the author's implementation of the subtractive mountain clustering method.

2. Mountain Clustering

Proposed by Yager and Filev [27, 28, 29], mountain clustering is one of the best methods to divide a set into a certain number of clusters/subsets. However, in order to create specific subsets, it is necessary to locate the so-called cluster centres. In general, any point of the measured space can be a cluster centre. Such an approach, however, results in computationally intensive algorithms of exponential complexity. The subtractive mountain clustering method [7], developed by Chiu and discussed in this article, takes a slightly different direction. It assumes that only an element of a set of data points can be a cluster centre. Thus, the method proposed by Chiu determines the starting points of clusters made up of a single point being the first element of a subset. Under the subtractive mountain clustering method, the searched space is limited to a divided set of points. However, due to its quadratic complexity with respect to the number of elements of a set, it is only used for medium-size sets [10].

The idea of subtractive mountain clustering is to determine function P for each point x_i representing its potential. It is therefore assumed that the potential in the i-th point of a set is given by

$$P(i) = \sum_{j=1}^{N} e^{-\alpha\|x_i - x_j\|^2}, \tag{1}$$

for $i = 1, \dots, N$, and $\alpha = 4 / (r_a)^2$ for constant $r_a > 0$. The form of the mountain function clearly shows that a data point with more neighbouring data points will have a higher potential value. This property makes subtractive mountain clustering much more resistant to disturbances caused by the emergence of random points compared with other clustering algorithms, such as C-means [17, 15] and FCM [4, 5, 6, 11].

After computing potential $P(i)$ of every data point, the data point with the highest potential is selected as the first cluster centre. Consequently, $c_1 = x_u$, where $u = \arg_i \max P(i)$, and $P(u)$ is given by P^* and is assumed to be the reference potential for the selection of new cluster centres. In addition, each time we select the next centre of the next cluster $c_k = x_u$ (for the relevant u), we revise the value of the mountain function assigned to particular points of a set in the following manner

$$P(i) = P(i) - P(u)e^{-\beta \|x_i - c_k\|^2}, \quad (2)$$

where $\beta = 4 / (r_b)^2$ for $r_b > 0$ is a constant defining the range of the mountain function. For the sake of practicality, we assume that $r_b > r_a$ and most often $r_b = 1.25\ r_a$. We continue to estimate new cluster centres until the potential of all points exceeds threshold $\varepsilon_d P^*$ for ε_d selected from range (0, 1).

The following algorithm presents how subtractive mountain clustering works [7, 8].

1. Select r_a,, r_b,, ε_u and ε_d.
2. Compute potential $P(i)$ of every point from set ($i = 1, \ldots, N$).
3. Select point x_u with the highest potential $P_u = P^*$ as the first cluster centre.
4. Assume that $k = 2$.
5. Then, keep repeating the following steps:
   a) Select point x_u with the highest potential P_u.
   b) If $P_u > \varepsilon_u P^*$, then x_u becomes the centre of the k-th cluster. If $\varepsilon_u P^* > P_u > \varepsilon_d P^*$, then x_u becomes centre c_k of the k-th cluster if it meets additional criteria (depending on how the algorithm is implemented).
   c) Assume that $k = k+1$.
   d) If $P_u > \varepsilon_d P^*$, end the clustering process - there are no more cluster centres.

Subtractive mountain clustering can be improved by incorporating a search by different values of α and β parameters. This way we obtain a least biased method [3]. We can even try to come up with a result similar to that produced by methods designed to estimate clusters with the lowest possible entropy [21, 20]. In addition, it is possible to replace the Gaussian potential function with the first-order Cauchy function [1]. A modified mountain function may also be used to estimate other types of clusters, for example circular shells [18]. If we add to that the option to replace the Euclidean distance with a kernel-induced distance [14], it turns out that subtractive mounting clustering is highly useful for estimating clusters in a given set.

The cluster centres determined using subtractive mountain clustering can be used to establish fuzzy inference rules for the purpose of various artificial intelligence algorithms [1, 19, 22, 23, 24]. In particular, they can be used to develop models predicting the behavior of various types of complex systems over time [16, 12, 9, 13], or in other words, to create machine learning algorithms.

3. Links of mountain clustering with statistics and probability

Let us now determine the relationship between equation (1) and the classical theory of probability. First, note that the mountain function is very similar to the probability density function for a normal distribution, which for $\mu = 0$ is given by

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right). \quad (3)$$

Taking into account the values of mountain function *P,* we can see that the following approximation is true

$$\exp(-\alpha x^2) = \exp\left(-\frac{4x^2}{r_a^2}\right) = \exp\left(-\frac{\frac{8}{9}x^2}{2(r_a/3)^2}\right) \cong \exp\left(-\frac{x^2}{2(r_a/3)^2}\right). \quad (4)$$

This means that r_a corresponds to approximately three standard deviations σ of a normal distribution presented in Figure 3.1. This interpretation of the mountain function means that clusters are in fact subsets of points concentrated around the centres according to a normal distribution with a given standard deviation. This correspondence would be accurate, if parameter α was defined as $4.5 / (r_a)^2$.

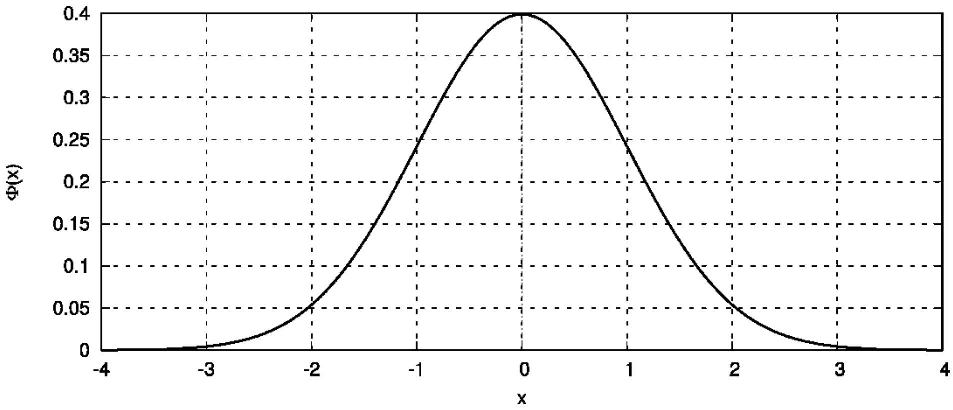


Fig. 3.1. Normal distribution curve for $\mu = 0$, $\sigma = 1$

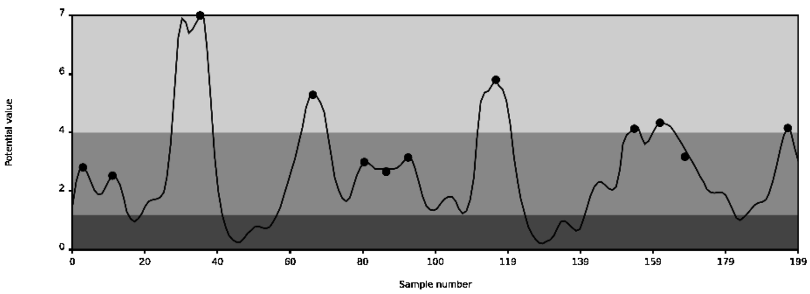
Now, it is easy to imagine that the Gaussian mountain function is replaced with another continuous probability distribution. This way we can better match the division into clusters in terms of points having a different probability distribution on individual coordinates. We can even try to define the mountain function for each coordinate of a point separately.

Another major conclusion from the above observations is that we can attempt to replace values $\|x_i - x_j\|$ with any other metric - not necessarily the Euclidean distance. Speaking of metrics, we must emphasize how important the process of normalizing coordinates of a point is to subtractive mountain clustering. If points x_i have only one coordinate, then normalization does not affect the results obtained. However, if points x_i have two or more coordinates, a failure to normalize individual coordinates may result in there being only one coordinate significantly affecting the final result. For example, let us imagine that the values of the first coordinate are from range [0, 1] and the values of the second coordinate are from range [99, 100]. It is not difficult to find such examples in the measuring practice. Note that, in principle, only the second coordinate having high absolute values has any significant effect on the value of the mountain function. That is why, before employing subtractive mountain clustering, it is important to scale down the values of each coordinate and move them, for example, to variability range [0, 1]. Then, the weight of each coordinate of a point will be identical and will have the same effect on the value of the mountain function.

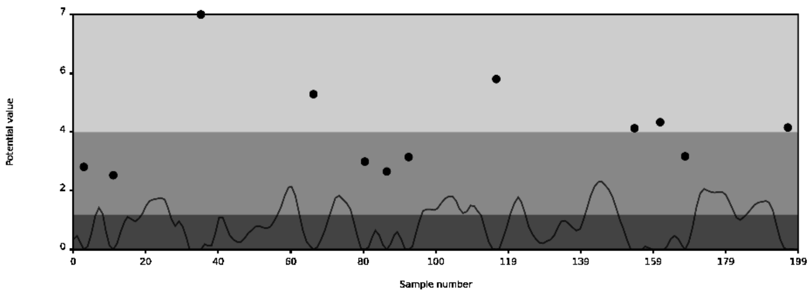
4. Application for Estimating Cluster Centres

The subtractive mountain clustering algorithm has been implemented in the form of a browser application with graphical presentation of results. The implementation is available for download from the author's website: https://achmie.v.prz.edu.pl/materialy-do-pobrania/materialy-ogolnodostepne/mountain-clustering-3.html

Figure 3.2 shows the result produced in the application for a sample set of 200 data points with three coordinates. The individual stages of determining the subsequent cluster centres are presented in Figure 3.5 at the end of the article.



a)



b)

Fig. 3.2. Values of the mountain function for a sample of 200 three-dimensional points: a) values of the mountain function when the algorithm starts to run, b) values of the mountain function after all cluster centres have been located

In order to illustrate how the implemented method works, a version of the application dedicated exclusively to two-dimensional data has been developed. Its main purpose is to show the readers what cluster centres should really be.

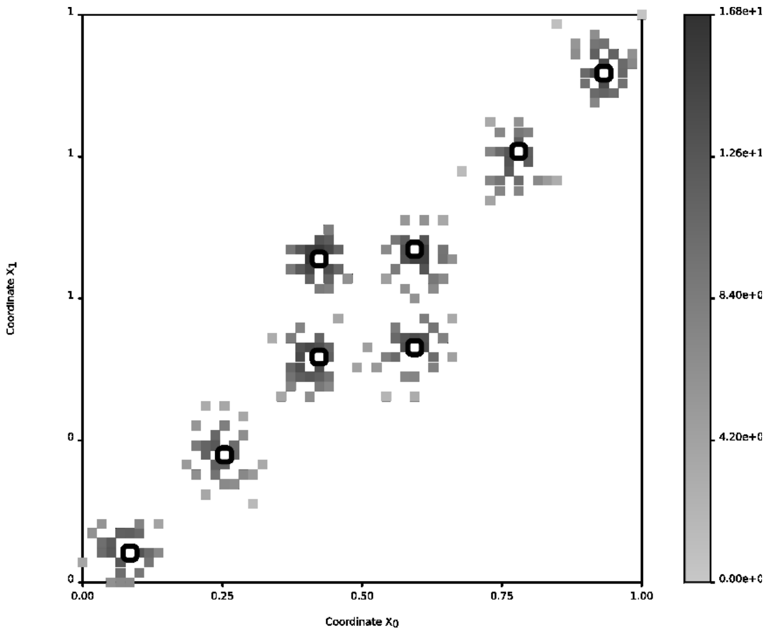
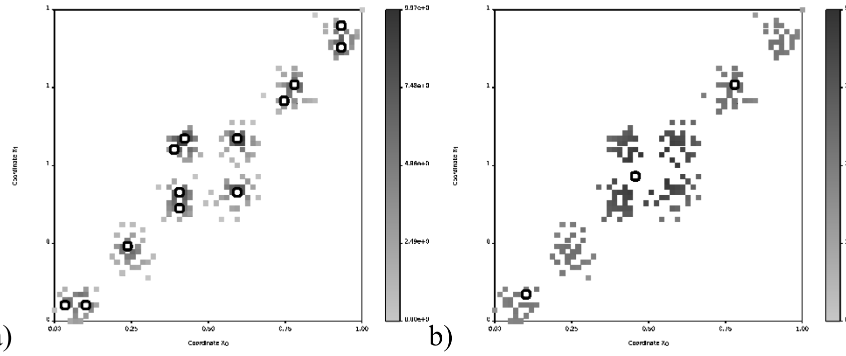


Fig. 3.3. The cluster centres located in the centres of individual clusters of points - the effect of a proper selection of radiuses $r_a = 0.10$ and $r_b = 0.14$

Figure 3.3 is a perfect illustration of the above. It shows that for well-chosen parameters r_a and r_b cluster centres should occur more or less in the centres of clusters of points. For poorly chosen parameters there may be too many or too few cluster centres as shown in Figures 3.4 a) and 4 b) respectively. Proper selection of these parameters is impossible without a thorough knowledge of the measurement data. These parameters can be, for example, in close correlation with certain settings and properties of the measuring device or may be related to the statistical distribution of a given coordinate. The two-dimensional version of the application features a button making it possible to disable normalization of the coordinates used in individual measurements. This allows us to check how much the results of this method depend on the absolute values of coordinates (as mentioned in the previous section).



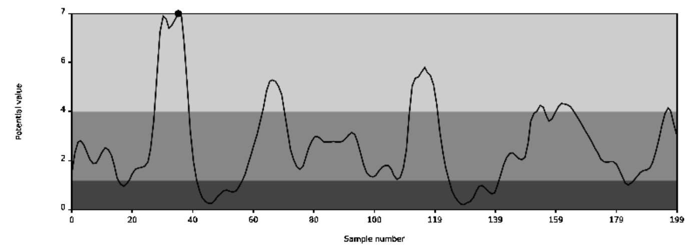
a) b)

Fig. 3.4. Cluster centres located too densely and too sparsely due to poor choice of radiuses: a) too densely for $r_a = 0.06$ and $r_b = 0.08$, b) too sparsely for $r_a = 0.28$ and $r_b = 0.36$

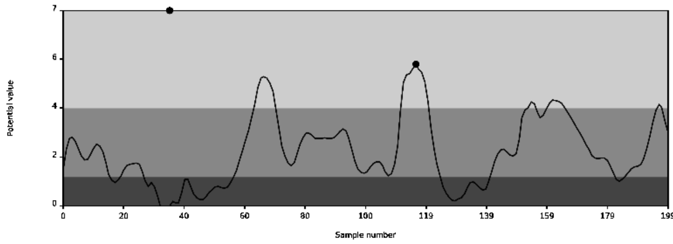
5. Conclusions

The article proposes a practical approach to subtractive mountain clustering and presents an IT tool designed for its implementation. Sections three and four discuss practical issues related to the proper use of the method in question. In particular, special attention is paid to the need to normalize all coordinates before starting the clustering process. The second important conclusion is drawn from the practical examples showing how the final results are influenced by poor choice of input parameters such as r_a and r_b.

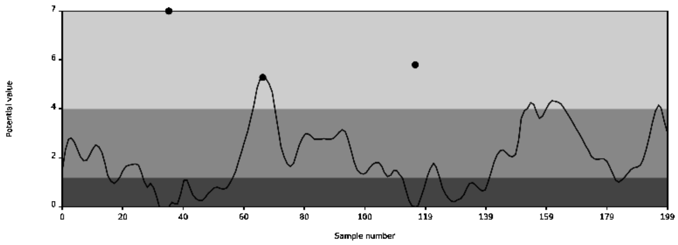
The study was carried out with the use of apparatus purchased from funds of the project 'Establishment of the Scientific and Research Inter-University Laboratory in Stalowa Wola', realized as part of the Operational Programme Eastern Poland 2007–2013, Priority axis I 'Modern Economy', Measure 1.3 'Supporting Innovativeness' as per contract No. POPW.01.03.00-18-016/12-00.



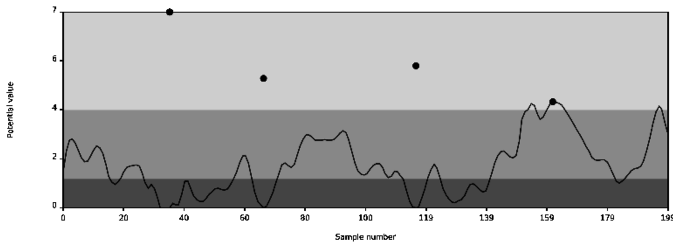
a)


Potential value
Sample number
0
2
4
6
7
0 20 40 60 80 100 119 139 159 179 199

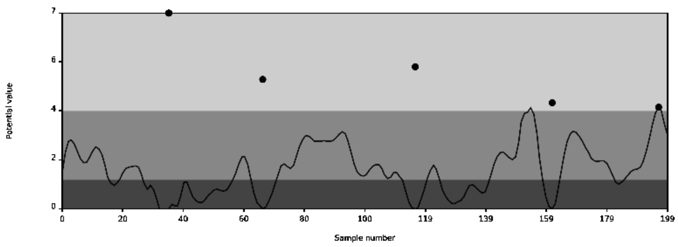
b)


Potential value
Sample number
0
2
4
6
7
0 20 40 60 80 100 119 139 159 179 199

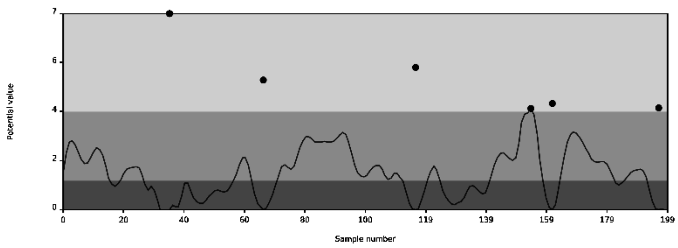
c)


Potential value
Sample number
0
2
4
6
7
0 20 40 60 80 100 119 139 159 179 199

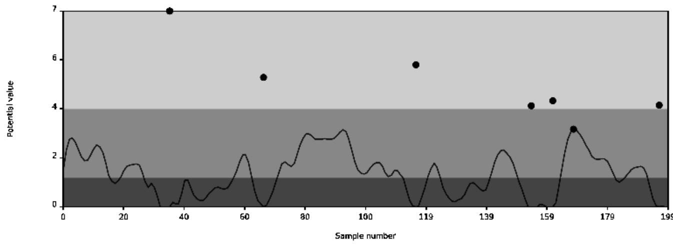
d)


Potential value
Sample number
0
2
4
6
7
0 20 40 60 80 100 119 139 159 179 199

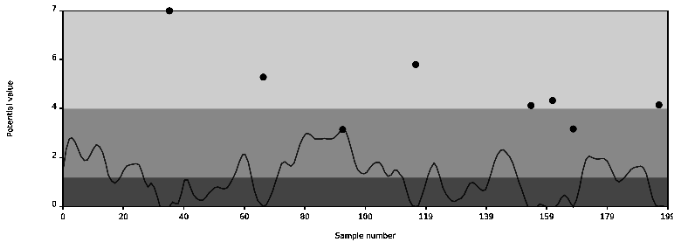
e)


Potential value
Sample number
0
2
4
6
7
0 20 40 60 80 100 119 139 159 179 199

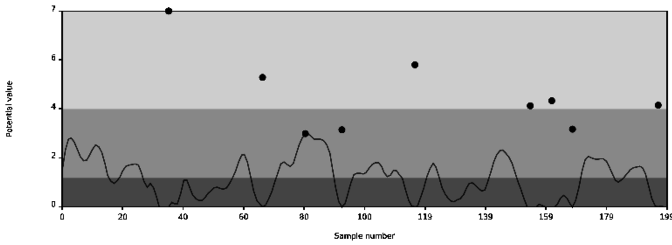
f)


Potential value
Sample number
0
2
4
6
7
0
20
40
60
80
100
119
139
159
179
199

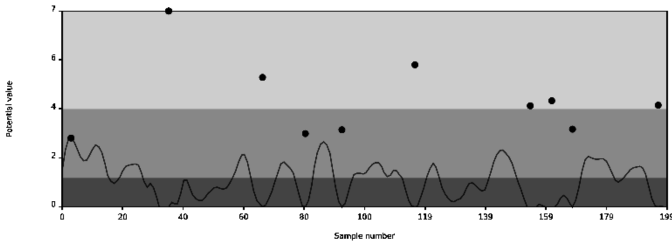
g)


Potential value
Sample number

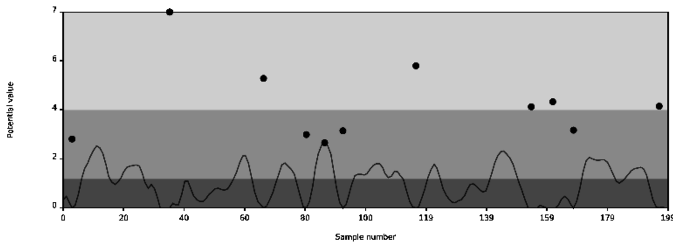
h)


Potential value
Sample number

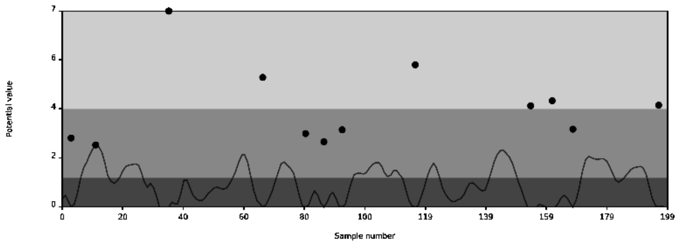
i)


Potential value
Sample number

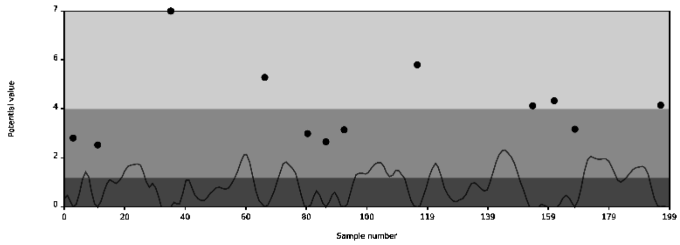
j)


Potential value
Sample number

k)



l)



m)

Fig. 3.5. Subsequent steps of the subtractive mountain clustering method on the example of a sample set of 200 three-dimensional data points; letters from a) to m) mark figures showing the modified mountain function curves and selected cluster centres

References

[1] Angelov P.P., Filev D.P. (2004). An approach to online identification of Takagi-Sugeno fuzzy models. IEEE Trans. Syst. Man. Cybern. Vol. 34(1). pp. 484-498.

[2] Astrom K.J., Wittenmark B. (1984). Computer Controlled Systems: Theory and Design, Prentice Hall Professional Technical Reference.

[3] Beni G., Liu X. (1994). A least biased fuzzy clustering method. IEEE Trans. Pattern. Anal. Mach. Intell. Vol. 16(9). pp. 954-960.

[4] Bezdek J. (1974). Cluster validity with fuzzy sets. J. Cybernetics. Vol. 3(3). pp. 58-71.

[5] Bezdek J. (1981). Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York.

[6] Bezdek J., Hathaway R., Sabin M., Tucker W. (1987). Convergence theory for fuzzy c-means: Counterexamples and repairs. The Analysis of Fuzzy Information, Bezdek J. (ed), CRC Press, Vol. 3, Chap. 8.

[7] Chiu S.L. (1994). Fuzzy model identification based on cluster estimation. J. Intell. Fuzzy Syst. Vol. 2(3). pp. 267-278.

[8] Chiu S.L. (1994). A cluster estimation method with extension to fuzzy model identification. In: Proc IEEE Int. Conf. Fuzzy Syst., Orlando, FL, Vol. 2. pp. 1240-1245.

[9] Crowder R.S. (1990). Predicting the Mackey-Glass time series with cascade-correlation learning. In Proc. 1990 Connectionist Models Summer School, Carnegie Mellon University, pp. 117-123.

[10] Dave R.N., Krishnapuram R. (1997) Robust clustering methods: A unified view. IEEE Trans. Fuzzy Syst. Vol. 5(2). pp. 270-293.

[11] Dunn J. (1974). A fuzzy relative of the ISODATA process and its use in detecting compact, well separated cluster. J. Cybernetics. Vol. 3(3). pp. 32-57.
[12] Jang J.S.R. (1993). ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans. on Systems, Man. & Cybernetics. Vol. 23(3). pp. 665-685.
[13] Kikuchi S., Nanda R., Perincherry V. (1994). Estimation of trip generation using the fuzzy regression method. 1994 Annual Meeting of Transportation Research Board, Washington, D.C.
[14] Kim D.W., Lee K.Y., Lee D., Lee K.H. (2005). A kernel-based subtractive clustering method. Pattern Recogn. Lett. Vol. 26. pp. 879-891.
[15] Linde Y., Buzo A., Gray R.M. (1980). An algorithm for vector quantizer design. IEEE Trans. Commun. Vol. 28. pp. 84-95.
[16] Mackey M., Glass L. (1977). Oscillation and chaos in physiological control systems. Science. Vol. 197. pp. 287-289.
[17] MacQueen J.B. (1967). Some methods for classification and analysis of multivariate observations. In: Proc 5th Berkeley Symp on Math Statistics and Probability, University of California Press, Berkeley, pp. 281-297.
[18] Pal N.R., Chakraborty D. (2000). Mountain and subtractive clustering method: Improvements and generalizations. Int. J. Intell. Syst. Vol. 15. pp. 329-341.
[19] Powell M.J.D. (1987). Radial basis functions for multivariable interpolation: a review, Algorithms for approximation, Clarendon Press, New York, NY.
[20] Rose K. (1998). Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. Proc. IEEE. Vol. 86(11). pp. 2210-2239.
[21] Rose K., Gurewitz E., Fox G.C. (1990). A deterministic annealing approach to clustering. Pattern Recogn. Lett. Vol. 11(9). pp. 589-594.
[22] Strobach P. (1990). Linear Prediction Theory: A Mathematical Basis for Adaptive Systems, Springer-Verlag New York, Inc., Secaucus, NJ.
[23] Sugeno M., Tanaka K. (1991). Successive identification of a fuzzy model and its applications to prediction of a complex system, Fuzzy Sets and Systems, Vol. 42(3), pp. 315-334.
[24] Takagi T., Sugeno M. (1985). Fuzzy identification of systems and its application to modeling and control. IEEE Trans. on Systems, Man. & Cybernetics. Vol. 15. pp. 116-132.
[25] Wang L.X., Mendel J.M. (1992). Generating fuzzy rules by learning from example. IEEE Trans. on Systems, Man. & Cybernetics. Vol. 22(6).
[26] Wang L.X. (1993). Training of fuzzy logic systems using nearest neighborhood clustering. Proc. 2nd IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE). pp. 13-17.
[27] Yager R.R., Filev D.P. (1993). Learning of fuzzy rules by mountain clustering. Proc. SPIE Conf. on Applications of Fuzzy Logic Technology. pp. 246-254.
[28] Yager R.R., Filev D.P. (1994). Generation of fuzzy rules by mountain clustering. J. Intell. Fuzzy Syst. Vol. 2(3). pp. 209-219.
[29] Yager R.R., Filev D.P. (1994). Approximate clustering via the mountain method. IEEE Trans. Syst. Man. Cybern. Vol. 24(8). 1279-1284.