Rough *c*-Regression Based on Optimization of Objective Function

Yasunori Endo[1], Akira Sugawara[2], and Naohiko Kinoshita[2]

[1] Faculty of Engineering, Information and Systems
University of Tsukuba
Tennodai 1-1-1, Tsukuba, Ibaraki, 305-8573, Japan
endo@risk.tsukuba.ac.jp
[2] Graduate School of Systems and Information Engineering
University of Tsukuba
Tennodai 1-1-1, Tsukuba, Ibaraki, 305-8573, Japan
{s1320626,s1220594}@u.tsukuba.ac.jp

Abstract. Clustering which is one of the pattern recognition methods is a technique automatically classifying data into some clusters. Various types of clustering are divided broadly into hierarchical and non-hierarchical clustering and crisp and fuzzy set theories have been applied to non-hierarchical clustering. Recently, clustering based on rough set theory has been attracted. Rough clustering represents a cluster by using two layers, i.e., upper and lower approximations. This paper proposes a *c*-regression method based on rough set representation which does regression analysis and clustering at the same time. Moreover, its effectiveness is shown through numerical examples.

1 Introduction

Computer system data has become large-scale and complicated in recent years due to progress in hardware technology, and the importance of data analysis techniques has been increasing accordingly. Clustering, which means a data classification method without any external criterion, has attracted many researchers as a significant data analysis technique.

Hathaway et al. proposed Hard and Fuzzy *c*-Regression (HCR and FCR) [1], which are clustering methods based on conventional regression model. With HCR and FCR, linear regression models are derived and belongingness or the membership grade of each object to each regression model is calculated. That is, those algorithms execute regression and clustering at same time.

FCR is a fuzzified HCR and fuzzy set representation plays very important role in FCR. Fuzzy set representation allows that an object belongs to two or more clusters. The belongingness is represented as a real value in a unit interval $[0, 1]$. Therefore, fuzzy set representation can be regarded as more flexible than HCR.

On the other hand, it is pointed out that "the fuzzy degree of membership may be too descriptive for interpreting clustering results." [2] In such cases, rough set representation is a more useful and powerful tool [3,4]. The basic concept of the rough set representation is based on two definitions of lower and upper approximations of a set.

V. Torra et al. (Eds.): MDAI 2013, LNAI 8234, pp. 272–283, 2013.

The lower approximation means that "an object surely belongs to the set" and the upper one means that "an object possibly belongs to the set". Clustering based on rough set representation could provide a solution that is less restrictive than conventional clustering and more descriptive than fuzzy clustering [5,2], and therefore the rough set based clustering has attracted increasing interest of researchers [6,7,8,9,10,11,2].

This paper proposes new clustering algorithms based on regression analysis and rough set representation and evaluate the algorithms through numerical examples.

2 Rough Sets

Let U be the universe and $R \subseteq U \times U$ be an equivalence relation on U. R is also called indiscernibility relation. The pair $X = (U, R)$ is called approximation space. If $x,\ y \in U$ and $(x, y) \in R$, we say that x and y are indistinguishable in X.

Equivalence class of the relation R is called elementary set in X. The family of all elementary sets is denoted by U/R. The empty set is also elementary in every X.

Each finite union of elementary sets in X is called composed set in X. The family of all composed sets is denoted by $\mathrm{Com}(X)$.

Since it is impossible to distinguish each element in an equivalence class, we may not be able to get a precise representation for an arbitrary subset $A \subset U$. Instead, any A can be represented by its lower and upper bounds. The upper bound $\overline{A}$ is the least composed set in X containing A, called the best upper approximation or, in short, upper approximation. The lower bound $\underline{A}$ is the greatest composed set in X containing A, called the best lower approximation or, briefly, lower approximation. The set $\mathrm{Bnd}(A) = \overline{A} - \underline{A}$ is called the boundary of A in X.

The pair $(\underline{A}, \overline{A})$ is the representation of an ordinary set A in the approximation space X, or simply a rough set of A. The elements in the lower approximation of A definitely belong to A, while elements in the upper bound of A may or may not belong to A.

From the above description of rough sets, we can define the following conditions for clustering:

(C1) An object x can be part of at most one lower approximation.
(C2) If $x \in \underline{A}$, $x \in \overline{A}$.
(C3) An object x is not part of any lower approximation if and only if x belongs to two or more boundaries.

3 c-Regression

In this section, we explain regression analysis and its error evaluation. Next, we show some representative methods of c-regression.

3.1 Regression

Regression is a way to obtain a regression model which presents the best relation between given variables x and y.

$x = (x^1, \ldots, x^p) \in R^p$, $y \in R$, and $\beta \in R^{p+1}$ mean independent variable, dependent variable, and regression coefficient, and it is assumed that objects $(x_1, y_1), \ldots, (x_n, y_n)$ are given. By using a regression model $f(x;\beta)$, the object (x_k, y_k) is denoted by

$$y_k = f(x_k;\beta) + \varepsilon_k.$$

ε_k means an error between the regression model and each dependent variable y_k. In regression analysis, a regression model which minimizes the error ε_k is derived. We can consider various functions as $f(x;\beta)$. In this paper, we consider the linear regression model.

In c-regression, i regression models $f(x,\beta_i)$ $(i = 1, \ldots, c)$ are considered and each cluster C_i is represented by the i-th regression model. The regression model is defined as follows:

$$f(x;\beta) = \sum_{j=1}^{p} \beta_i^j x^j + \beta_i^{p+1}.$$

By putting $z = (x^1, \ldots, x^p, 1)^T$, we can rewrite the above equation as follows:

$$f(x;\beta_i) = z^T \beta_i.$$

$(\bullet)^T$ means transposition.

The clustering problem is which object x_k belongs to which cluster C_i.

3.2 Error Evaluation

There are some approaches to evaluate errors between each pair (x_k, y_k) and regression models. Two of most popular approaches are least square deviation (LS) and least absolute deviation (LAD). LS minimizes $\sum_{k=1}^{n} (d_{ki})^2$ and LAD minimizes $\sum_{k=1}^{n} |d_{ki}|$. Here,

$$d_{ki} = y_k - f(x_k;\beta_i).$$

From here, LS-(name of the method) and LAD-(name of the method) mean the method with LS and LAD as error evaluation, respectively.

3.3 RKR

Peters proposed RKR (Rough k-Regression) [10] which is not based on optimization of objective function. RKR is inspired by RKM (Rough k-Means) by Lingras [2].

Algorithm 1 (RKR)

RKR1 *Give initial lower and upper approximations.*
RKR2 *Calculate the optimal regression coefficients.*
RKR3 *Update lower and upper approximations.*
RKR4 *If the stop criterion satisfies, finish. Otherwise back to* **RKR2**.

LS-RKR

β_i is calculated as follows:

$$\beta_i = \left(\sum_{k=1}^{n} u_{ki} z_k z_k^T \right)^{-1} \sum_{k=1}^{n} u_{ki} y_k z_k,$$

where

$$u_{ki} = \begin{cases} \underline{w}, & (x_i \in \underline{C}_i) \\ \overline{w}. & (x_i \in \mathrm{Bnd}(C_i)) \end{cases}$$

$\underline{w}$ and $\overline{w}$ are given constants and $\underline{w} + \overline{w} = 1$.

Here, we introduce the notations ν_{ki} and u_{ki}. ν_{ki} and u_{ki} means the belongingness of x_k to $\underline{C}_i$ and $\mathrm{Bnd}(C_i)$, respectively. Those are calculated as follows:

$$\nu_{ki} = \begin{cases} 1, & ((T_k = \phi) \wedge (i = p)) \\ 0, & (\text{otherwise}) \end{cases}$$

$$u_{ki} = \begin{cases} 1, & ((T_k \neq \phi) \vee ((i = p) \wedge (i \in T_k))) \\ 0. & (\text{otherwise}) \end{cases}$$

Here,

$$p = \arg\min_i |d_{ki}|,$$

$$T_k = \left\{ i \mid \frac{|d_{ki}|}{|d_{kp}|} \leq \text{threshold}.\ i \neq p,\ i = 1, \ldots, c \right\}$$

threshold is an arbitrary constant.

LAD-RKR

The optimal solution to β_i can be calculated from solving the following linear programming problem for each i:

$$\sum_{k=1}^{n} u_{ki} r_{ki} \to \min, \tag{1}$$

$$\text{s.t.} \quad y_k - f(x_k; \beta_i) \leq r_{ki},$$

$$y_k - f(x_k; \beta_i) \geq -r_{ki},$$

$$r_{ki} \geq 0, \quad (k = 1, \ldots, n)$$

where

$$u_{ki} = \begin{cases} \underline{w}, & (x_i \in \underline{C}_i) \\ \overline{w}. & (x_i \in \mathrm{Bnd}(C_i)) \end{cases}$$

$\underline{w}$ and $\overline{w}$ are given constants and $\underline{w} + \overline{w} = 1$.

How to calculate v_{ki} and u_{ki} is same as LS-RKR.

The two parameters $\underline{w}$ and threshold give The both methods LS-RKR and LAD-RKR flexibility. That is an advantage of the both methods that we can obtain various outputs.

On the other hand, we do not know how to determine the parameters because there is no standard of those parameters. Moreover, there is no evaluation criteria for their outputs in RKR. Outputs of many clustering algorithms based on optimization of objective function strongly depend on initial values, therefore we have to evaluate the outputs by certain criteria. Those matters are disadvantages of RKR. So, we will propose new clustering algorithms based on rough set representation and regression model, which overcome the above disadvantages of RKR.

4 Proposed Method: RCR

In this section, we propose RCR ((Rough c-Regression) based on optimization of objective function. RCR is based on optimization of objective function. In our algorithm, the number of parameters is one and the objective function can be used as evaluation criteria for the outputs.

The flow of algorithms of LS-RCR (LS-Rough c-Regression) and LAD-RCR (LAD-Rough c-Regression) are same, so we call those algorithms as RCR in a lump. Here, $N = \{v_{ki} \mid k = 1, \ldots, n,\ i = 1, \ldots, n\}$ and $U = \{u_{ki} \mid k = 1, \ldots, n,\ i = 1, \ldots, n\}$.

Algorithm 2 (RCR)

RCR1. *Give initial lower and upper approximations and calculate the optimal regression coefficients.*
RCR2. *Calculate N and U which minimize the objective function with fixing β.*
RCR3. *Calculate β which minimizes the objective function with fixing N and U.*
RCR4. *If the stop criterion satisfies, finish. Otherwise back to* **RCR2**.

LS-RCR

The objective function of LS-RCR is defined as follows:

$$J(N, U, \beta) = \sum_{i=1}^{c} \sum_{k=1}^{n} (v_{ki}\underline{w} + u_{ki}\overline{w}) d_{ki}^2$$

The constraints are as follows:

$$\begin{aligned} & v_{ki} \in \{0, 1\}, \quad u_{ki} \in \{0, 1\}, \\ & \underline{w} + \overline{w} = 1, \\ & \sum_{i=1}^{c} v_{ki} = 0 \Rightarrow \sum_{i=1}^{c} u_{ki} \geq 2, \\ & \sum_{i=1}^{c} u_{ki} = 0 \Rightarrow \sum_{i=1}^{c} v_{ki} = 1. \end{aligned} \tag{2}$$

Those constraints obviously satisfy the above conditions **C1**, **C2** and **C3**. Actually, (2) is rewritten as

$$\sum_{i=1}^{c} \nu_{ki} = 0 \Rightarrow \sum_{i=1}^{c} u_{ki} = 2.$$

When an object belongs to some boundaries, its belonging to two boundaries makes the objective function minimize in comparison with three or more boundaries.

Now, we describe how to calculate the optimal solutions to ν_{ki} and u_{ki}. For each object x_k, we first assume that x_k belongs to the lower approximation of a cluster which corresponds to the closest regression model $f(x, \beta_p)$, that is,

$$p = \arg\min_i d_{ki}^2.$$

In this case, the objective function is calculated as follows:

$$J_\nu = \nu_{kp} \underline{w} d_{kp}^2.$$

Next, We assume that x_k belongs to two boundaries. In this case, we can find the closest regression model $f(x, \beta_p)$ and the second closest one $f(x, \beta_q)$, that is,

$$q = \arg\min_{i, i \neq p} d_{ki}^2.$$

The objective function is calculated as follows:

$$J_u = u_{kp} \overline{w} d_{kp}^2 + u_{kq} \overline{w} d_{kq}^2.$$

Finally, ν_{ki} and u_{ki} is calculated as follows:

$$\nu_{ki} = \begin{cases} 1, & ((J_\nu < J_u) \wedge (i = p)) \\ 0, & (\text{otherwise}) \end{cases}$$

$$u_{ki} = \begin{cases} 1, & ((J_\nu > J_u) \wedge ((i = p) \vee (i = q))) \\ 0. & (\text{otherwise}) \end{cases}$$

Next, we consider the optimal solution to β_i. From $\frac{\partial J}{\partial \beta_i^j} = 0$, we obtain

$$\beta_i = \left(\sum_{k=1}^{n} (\nu_{ki} \underline{w} + u_{ki} \overline{w}) z_k z_k^T \right)^{-1} \cdot \sum_{k=1}^{n} (\nu_{ki} \underline{w} + u_{ki} \overline{w}) y_k z_k.$$

LAD-RCR

The objective function of LAD-RCR is defined as follows:

$$J(N, U, \beta) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\nu_{ki} \underline{w} + u_{ki} \overline{w}) |d_{ki}|.$$

The constraints are the same as LS-FCR.

The optimal solutions to v_{ki} and u_{ki} can be obtained by replacing d_{ki}^2 in LS-FCR to $|d_{ki}|$.

The optimal solution to β_i is calculated from solving the following linear programming problem:

$$\begin{aligned}
&\sum_{k=1}^{n}(v_{ki}\underline{w} + u_{ki}\overline{w})r_{ki} \to \min,\\
&\text{s.t.} \quad y_k - f(x_k;\beta_i) \le r_{ki},\\
&\qquad y_k - f(x_k;\beta_i) \ge -r_{ki},\\
&\qquad\qquad r_{ki} \ge 0. \quad (k = 1,\ldots,n)
\end{aligned}$$

5 Numerical Examples

5.1 Preparation

Evaluation of Outputs of RKR

RKR has no evaluation criteria, so we can not evaluate the outputs of RKR. Therefore, we consider the objective function of RCR and the following function based on the objective function of HCR as the evaluation criterion as follows:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{n}\left(v_{ki}d_{ki}^2 + \frac{u_{ki}}{\sum_{j=1}^{c}u_{kj}}d_{ki}^2\right) \tag{3}$$

(Name of the method)-R and (name of the method)-H in the results means that the objective function of RCR and (3) are used as the evaluation criterion, respectively.

Datasets

We prepare two artificial datasets and one real dataset to compare the proposed methods with the conventional ones. All datasets are in two-dimensional Euclidean space. Artificial dataset 1 consists of two line shaped clusters and 20 messed noise objects, and each cluster has 100 objects (Fig. 1). Artificial dataset 2 consists of two line shaped clusters which are very close to each other and 15 noise objects at random, and each cluster has 15 objects (Fig. 2). The real dataset represents GDP and energy consumption of Asian countries from 1973 to 1992 (Fig. 3). Horizontal and vertical axes mean real GDP (10^9\$) and primary energy consumption (10^6ton), respectively.

Evaluation Method

As quantitative evaluation, we show the total sum of errors L, i.e.,

$$L = \sum_{k=1}^{n}\sum_{i=1}^{c}u'_{ki}\,(d_{ki})^2\,.$$

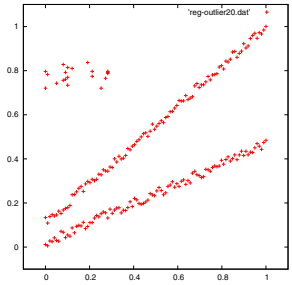

Fig. 1. Artificial dataset 1

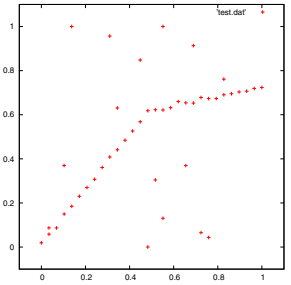

Fig. 2. Artificial dataset 2

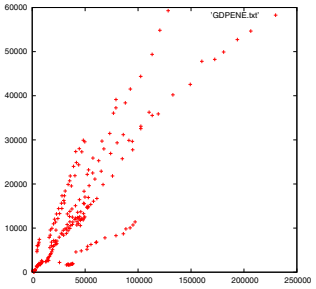

Fig. 3. GDP dataset

Here u'_{ki} means belongingness of x_k to C_i after assigning all x_k to their final clusters. Therefore, $u'_{ki} \in \{0, 1\}$. The noise objects are not included into the calculation. In case of RKR and RCR, we calculate L by two ways. One is that L is calculated without no noise objects which belong to certain boundaries, and another is that L is calculated with the object.

Parameters

In each algorithms, we prepare 1000 initial values and set parameters as $m = 3$, $c = 2$ for the artificial datasets, $c = 3$ for GDP dataset, $\underline{w} = 0.7$ and threshold $= 1.1$ in RKR, and $\underline{w} = 0.7$ or $\underline{w} = 0.8$ in RCR.

5.2 The First Artificial Dataset

Table 1 shows the total sum of errors by each methods. The values in parentheses in the table means total sum of errors between no noise objects which belong to certain boundaries and the closest regression model.

From the results, it is obvious that least absolute deviation is more robust than least square deviation. In particular, the proposed algorithms output better results. The reason is that the belongingness of noise objects to each clusters is obviously larger than $\overline{w}$ in case of RKR and RCR. For instance, the belongingness of noise objects to clusters in

LAD-FCR is about 0.5 and the belongingness in LAD-HCR is 1, while $\overline{w}$ in LAD-RCR is 0.3 or 0.2. In other words, the parameter $\overline{w}$ ($\underline{w}$) is very important role in RCR.

Table 1. Total sum of errors for Artificial dataset 1

Method	Total sum of errors
LS-HCR	1.934552
LS-FCR	2.120955
LS-RKR-H	1.747893 (1.955494)
LS-RKR-R	1.706390 (1.964700)
LS-RCR ($\underline{w} = 0.7$)	**1.651619 (1.956077)**
LS-RCR ($\underline{w} = 0.8$)	**1.139004 (2.013713)**
LAD-HCR	0.040824
LAD-FCR	0.043938
LAD-RKR-H	0.040824 (0.040824)
LAD-RKR-R	0.040824 (0.040824)
LAD-RCR ($\underline{w} = 0.7$)	**0.037886 (0.037886)**
LAD-RCR ($\underline{w} = 0.8$)	**0.033636 (0,033636)**

5.3 The Second Artificial Dataset

Table 2 shows the total sum of errors by each methods. The values in parentheses in the table means total sum of errors between no noise objects which belong to certain boundaries and the closest regression model. The results for this dataset also show that least absolute deviation is better than least square deviation.

Moreover, the results show the outputs by LS-RCR are similar to LS-HCR or S-FCR. The reason is that RCR has essentially three kinds of belongingness, ν_{ki}, u_{ki}, and $\underline{w}$. The larger $\underline{w}$ is, the more easily objects belong to boundaries, because RCR is based on optimization of objective function. Objects easily belong to lower approximations when $\underline{w} = 0.7$, and consequently, the output by RCR is similar to HCR. On the other hand, objects easily belong to boundaries when $\underline{w} = 0.8$, and the output by RCR is similar to FCR.

For the dataset, the best output is by LAD-RCR with $\underline{w} = 0.8$. The reason is that there is a few objects which must be considered when solving linear programming problem for each cluster and consequently the optimal solutions can be easily calculated. In addition, it is important that $\underline{w}$ has a big influence on the outputs by RCR.

5.4 GDP Dataset

We normalize original data into $[0, 1] \times [0, 1]$ and apply each algorithms. Table 3 shows the total sum of errors by each methods. The values in parentheses in the table means total sum of errors between no noise objects which belong to certain boundaries and the closest regression model. From the results, we can not see large difference between least square deviation and least absolute one. It can be considered to be better to apply least square deviation with considering calculation cost.

Table 2. Total sum of errors for Artificial dataset 2

Method	Total sum of errors
LS-HCR	0.373295
LS-FCR	0.156386
LS-RKR-H	0.373295 (0.373295)
LS-RKR-R	0.327269 (0.392041)
LS-RCR ($\underline{w} = 0.7$)	**0.327269 (0.392041)**
LS-RCR ($\underline{w} = 0.8$)	**0.125412 (0.157550)**
LAD-HCR	0.183568
LAD-FCR	0.049798
LAD-RKR-H	0.183568 (0.183568)
LAD-RKR-R	0.183568 (0.183568)
LAD-RCR ($\underline{w} = 0.7$)	**0.143662 (0.183568)**
LAD-RCR ($\underline{w} = 0.8$)	**0.001095 (0.001095)**

Table 3. Total sum of errors for GDP dataset

Method	Total sum of errors
LS-HCR	0.229599
LS-FCR	0.232759
LS-RKR-H	0.219475 (0.229734)
LS-RKR-R	0.219475 (0.219475)
LS-RCR ($\underline{w} = 0.7$)	**0.217145 (0.229859)**
LS-RCR ($\underline{w} = 0.8$)	**0.163177 (0.231176)**
LAD-HCR	0.238877
LAD-FCR	2.083101
LAD-RKR-H	0.238877 (0.238877)
LAD-RKR-R	0.230992 (0.239120)
LAD-RCR ($\underline{w} = 0.7$)	**0.223766 (0.239134)**
LAD-RCR ($\underline{w} = 0.8$)	**0.034578 (0.938140)**

Moreover, it is obvious that the outputs by LAD-FCR and LAD-RCR with $\underline{w} = 0.8$ are incorrect. We can consider three reasons. The first is that each object belongs to all clusters by LAD-FCR and consequently, the regression models gather. The second is that objects easily belong to boundaries by LAD-RCR and then, the outputs by LAD-RCR with $\underline{w} = 0.8$ is similar to LAD-FCR. The third is that objects are dense nearby origin, so it is difficult to classify those object clearly.

The cluster which is represented by a line with the largest slope can be regarded as a group of energy-consuming advanced countries, and the cluster which is represented by a line with the smallest slope can be regarded as a group of energy-conservation countries. The belongingness of each country depends on error evaluation. For instance, Nepal belongs to a group of middle countries with least square deviation, while it belongs to a group of energy-conservation countries with least absolute deviation. However, there is little difference between total sum of errors by least square deviation an least absolute one. Therefore, we can interpret it as the possibility that Nepal could be either.

6 Conclusion

This paper proposed new c-regression based on optimization of objective function and rough set representation, and evaluate the proposed algorithms through numerical examples.

We believe that LAD-RCR is useful in comparison with other algorithms. LAD-RCR reduces the influence of noise objects by making those objects belong to boundaries. Consequently, LAD-RCR derives better regression model. If boundaries is considered as noise clusters, RCR can be regarded as a kind of noise clustering.

When least absolute deviation is used instead of least square deviation, the RCR algorithm has an advantage of robustness against noise, while calculation cost increases because linear programming problems must be solved to calculate the optimal solutions. Moreover, there is another disadvantage that it is harder to obtain good results than least square deviation. We think that this causes the linear programming problems.

The parameter $\underline{w}$ has a big influence on the results in the proposed methods. Therefore, when the proposed methods are used, it is necessary to choose an adequate value of $\underline{w}$.

In future papers, we have to consider two matters. The first is about boundaries of RCR. In the proposed algorithms, each object belongs to just two boundaries when it does not belong to any lower approximations. Then, we have to evaluate the validity. The second is about way to choose error evaluation. There are no indication for it. The indication is necessary when we use RCR, so we have to consider it.

Acknowledgment. We would like to gratefully thank Professor Sadaaki Miyamoto of the University of Tsukuba, and Associate Professor Yuchi Kanzawa of Shibaura Institute of Technology for their advice. We would also like to sincerely thank Professor Yoji Uchiyama of the University of Tsukuba for his providing GDP Dataset.

References

1. Hathaway, R.J., Bezdek, J.C.: Switching regression models and fuzzy clustering. IEEE Trans. on Fuzzy Systems 1(3), 195–204 (1993)
2. Lingras, P., Peters, G.: Rough clustering. WIREs Data Mining and Knowledge Discovery 1(1), 64–72 (2011)
3. Pawlak, Z.: Rough Sets. International Journal of Computer and Information Sciences 11(5), 341–356 (1982)
4. Inuiguchi, M.: Generalizations of Rough Sets: From Crisp to Fuzzy Cases. In: Proceedings of Rough Sets and Current Trends in Computing, pp. 26–37 (2004)
5. Pawlak, Z.: Rough Classification. International Journal of Man-Machine Studies 20, 469–483 (1984)
6. Hirano, S., Tsumoto, S.: An Indiscernibility-Based Clustering Method with Iterative Refinement of Equivalence Relations. Journal of Advanced Computational Intelligence and Intelligent Informatics 7(2), 169–177 (2003)
7. Lingras, P., West, C.: Interval Set Clustering of Web Users with Rough K-Means. Journal of Intelligent Information Systems 23(1), 5–16 (2004)
8. Mitra, S., Banka, H., Pedrycz, W.: Rough-Fuzzy Collaborative Clustering. IEEE Transactions on Systems Man, and Cybernetics, Part B, Cybernetics 36(5), 795–805 (2006)

9. Maji, P., Pal, S.K.: Rough Set Based Generalized Fuzzy *C*-Means Algorithm and Quantitative Indices. IEEE Transactions on System, Man and Cybernetics, Part B, Cybernetics 37(6), 1529–1540 (2007)
10. Peters, G.: Rough clustering and regression analysis. In: Yao, J., Lingras, P., Wu, W.-Z., Szczuka, M.S., Cercone, N.J., Ślęzak, D. (eds.) RSKT 2007. LNCS (LNAI), vol. 4481, pp. 292–299. Springer, Heidelberg (2007)
11. Mitra, S., Barman, B.: Rough-fuzzy clustering: An application to medical imagery. In: Wang, G., Li, T., Grzymala-Busse, J.W., Miao, D., Skowron, A., Yao, Y. (eds.) RSKT 2008. LNCS (LNAI), vol. 5009, pp. 300–307. Springer, Heidelberg (2008)
12. Nakayama, Y., Miyamoto, S.: Least Square Deviations and Least Absolute Deviations in *c*-Regression Models. In: Proc. of Fuzzy System Symposium, vol. 17, pp. 583–586 (2001) (in Japanese)
13. Miyamoto, S., Umayahara, K., Nemoto, T., Takata, O.: Algorithm of Fuzzy *c*-Regression Based on Least Absolute Deviations. Journal of Japan Society for Fuzzy Theory and Systems 12(4), 578–587 (2000) (in Japanese)