

A Rough Set-Based Multiple Criteria Linear Programming Approach for Classification

Zhiwang Zhang¹, Yong Shi², Peng Zhang³, and Guangxia Gao⁴

¹ School of Information of Graduate University of Chinese Academy of Sciences, China;
Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences,
Beijing 100080, China
zzwmis@163.com

² Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences,
Beijing 100080, China; College of Information Science and Technology,
University of Nebraska at Omaha, Omaha NE 68182, USA
yshi@gucas.ac.cn

³ School of Information of Graduate University of Chinese Academy of Sciences, China;
Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences,
Beijing 100080, China
zhangpeng04@gmail.com

⁴ Foreign Language Department, Shandong Institute of Business and Technology, Yantai,
Shandong 264005, China
Gaoguangxia2006@126.com

Abstract. It is well known that data mining is a process of discovering unknown, hidden information from a large amount of data, extracting valuable information, and using the information to make important business decisions. And data mining has been developed into a new information technology, including regression, decision tree, neural network, fuzzy set, rough set, support vector machine and so on. This paper puts forward a rough set-based multiple criteria linear programming (RS-MCLP) approach for solving classification problems in data mining. Firstly, we describe the basic theory and models of rough set and multiple criteria linear programming (MCLP) and analyse their characteristics and advantages in practical applications. Secondly, detailed analysis about their deficiencies are provided respectively. However, because of the existing mutual complementarities between them, we put forward and build the RS-MCLP methods and models which sufficiently integrate their virtues and overcome the adverse factors simultaneously. In addition, we also develop and implement these algorithm and models in SAS and Windows platform. Finally, many experiments show that RS-MCLP approach is prior to single MCLP model and other traditional classification methods in data mining.

Keywords: Data mining, Rough Set, MCLP, Classification.

1 Introduction

Data mining has been used by many organizations to extract information or knowledge from large volumes of data and then use the valuable information to make critical business decisions. Consequently, analysis of the collected history data in data

warehouse or data mart can gain better insight into your customers and evaluation of the place in industry, improve the quality of decision-making and effectively increase competitiveness in market.

From the aspect of methodology, data mining can be performed through association, classification, clustering, prediction, sequential patterns, and similar time sequences [Han and Kamber, 2001]. For classification, data mining algorithms use the existing data to learn decision functions that map each case of the selected data into a set of predefined classes. Among various mathematical tools including statistics, decision trees, fuzzy set, rough set and neural networks, linear programming has been initiated in classification more than twenty years [Freed and Glover, 1981]. Given a set of classes and a set of attribute variables, one can use a linear programming model to define a related boundary value separating the classes. Each class is then represented by a group of constraints with respect to a boundary in the linear program. The objective function minimizes the overlapping rate of the classes or maximizes the distance between the classes. The linear programming approach results in an optimal classification. It is also flexible to construct an effective model to solve multi-class problems.

However, the MCLP model is not good at dimensional reduction and removing information redundancy, especially facing many attributes with a large number of data. To our joy, rough set can find the minimal attribute set and efficiently remove redundant information [Z. Pawlak, 1982]. Consequently, the developing approach of RS-MCLP to data mining is promising to overcome these disadvantages.

In this paper, we will give a full description of rough set-based MCLP method and model for classification in data mining. First a detailed introduction of MCLP model and rough set in the related work section is given, including the algorithms of MCLP model, rough set for feature selection and their virtues of classification. Then we put forth the methodology of the Rough set-based MCLP model after the analysis of their deficiencies respectively and implement the combinational model in SAS and Windows platform. And then we describe the advantages of the RS-MCLP model. Finally we present a comprehensive example in different data set and experimental conclusions.

2 Related Work

2.1 MCLP Approach for Classification

A general problem of data classification by using multiple criteria linear programming can be described as the following: Given a set of n variables or attributes in database $A = (a_1, a_2, \dots, a_n)$, let $A_i = (a_{i1}, a_{i2}, \dots, a_{in}) \in R^n$ be the sample observations of data for the variables, where $i = 1, 2, \dots, l$ and l is the sample size. If a given problem can be predefined as s different classes, C_1, C_2, \dots, C_s , then the boundary between the j th and $(j+1)$ th classes can be b_j , $j = 1, 2, \dots, s-1$. Then we determine the coefficients for an appropriate subset of the variables which can be represent the whole of decision space, denoted by $X = (x_1, x_2, \dots, x_m) \in R^m$ ($m \leq n$), and scalars b_j such that the separation of these classes can be described as follows:

(1) $A_i X \leq b_1, \forall A_i \in C_1$ and $b_{k-1} \leq A_i X \leq b_k, \forall A_i \in C_k, k = 2, \dots, s-1$, and $A_i X \geq b_{s-1}, \forall A_i \in C_s$. where $\forall A_i \in C_j, j = 1, 2, \dots, s$, means that the data case A_i belongs to the class C_j .

For a binary classification we need to choose a boundary b to separate two classes: G (Goods) and B (Bads); For the purpose of simplification we present only descriptions about binary classification, and which can be extended easily into multiple classification circumstances. That is:

(1') $A_i X \leq b, A_i \in G$ and $A_i X \geq b, A_i \in B$. Where A_i are the vector value of the subset of the variables.

For better separation of Goods and Bads, someone considered the two measurements of the overlapping degree with respect to A_i and the distance where A_i departed from its adjusted boundary b respectively [Freed and Glover, 1981]. Subsequently, Glover introduced the two factors above in models [Glover, 1990]. Consequently, we have the following conclusions:

Let α_i be the overlapping degree as above describing, and we want to minimize the sum of α_i , then the primal linear programming can be written as:

$$(2) \text{ Minimize } \sum_i \alpha_i, \text{ subject to: } A_i X \leq b + \alpha_i, A_i \in G \text{ and } A_i X \geq b - \alpha_i, A_i \in B.$$

Let β_i be the distance as above defined too, and we want to maximize the sum of β_i , then the primal linear programming can be expressed as:

$$(3) \text{ Maximize } \sum_i \beta_i, \text{ subject to: } A_i X \geq b - \beta_i, A_i \in G \text{ and } A_i X \leq b + \beta_i, A_i \in B.$$

If considering the two measurements in classification simultaneously, we will get hybrid multiple criteria linear programming model as follows:

(4) *Minimize* $\sum_i \alpha_i$ and *Maximize* $\sum_i \beta_i$, subject to: $A_i X \leq b + \alpha_i - \beta_i, A_i \in G$ and $A_i X \geq b - \alpha_i + \beta_i, A_i \in B$. Where A_i are given, X and b are unrestricted, and α_i and $\beta_i \geq 0$.

Furthermore, the compromise solution approach has been used to improve the above model (4) in business practices [Shi and Yu, 1989]. It is assumed that the ideal value of $-\sum_i \alpha_i$ be α^* ($\alpha^* > 0$), at the same time, the ideal value of $\sum_i \beta_i$ be β^* ($\beta^* > 0$). Then, if $-\sum_i \alpha_i > \alpha^*$, the regret measure is defined as $-d_\alpha^+ = \alpha^* + \sum_i \alpha_i$ ($d_\alpha^+ \geq 0$); otherwise, it is 0. If $-\sum_i \alpha_i < \alpha^*$, the regret measure is also written as $d_\alpha^- = \alpha^* - \sum_i \alpha_i$ ($d_\alpha^- \geq 0$); otherwise it is 0.

Thus, we have $\alpha^* + \sum_i \alpha_i = d_\alpha^- - d_\alpha^+$ and $|\alpha^* + \sum_i \alpha_i| = d_\alpha^- + d_\alpha^+$. Similarly, we have $\beta^* - \sum_i \beta_i = d_\beta^- - d_\beta^+$ and $|\beta^* - \sum_i \beta_i| = d_\beta^- + d_\beta^+, d_\beta^+ \geq 0, d_\beta^- \geq 0$. To sum up, the improved MCLP model which we use for modeling in this paper may be expressed as:

$$(5) \text{ Minimize : } d_\alpha^- + d_\alpha^+ + d_\beta^- + d_\beta^+,$$

$$\text{Subject to: } \alpha^* + \sum_i \alpha_i = d_\alpha^- - d_\alpha^+ \text{ and } \beta^* + \sum_i \beta_i = d_\beta^- - d_\beta^+, \\ A_i X = b + \alpha_i - \beta_i, A_i \in G \text{ and } A_i X = b - \alpha_i + \beta_i, A_i \in B.$$

Where A_i , α^* and β^* are given, X and b are unrestricted, and $\alpha_i, \beta_i, d_\alpha^-, d_\alpha^+, d_\beta^-, d_\beta^+ \geq 0$.

Owing to the following characteristics, MCLP models are more popular correspondingly than traditional nonlinear models, a) simplicity, from algorithm to model results MCLP are very easy to understand and explain. b) flexibility, user may freely input different parameters to adjust model performance and get better effects. c) generalization, because of systematic consideration to the best trade-off between minimizing the overlapping degree and maximizing the distance departed from boundary, the model will gain better classification correct rate and generalization of training set and test set.

2.2 Rough Sets-Based Feature Selection Method

On account of the deficiency which MCLP model failed to make sure and remove the redundancy in variables or attributes set. That is to say the model is not good at giving judgment on attributes which are useful and important or unnecessary and unimportant relatively. However, rough set methods have an advantage in this aspect.

Rough set theory which was developed by Z. Pawlak is a new mathematical analysis method for dealing with fuzzy and uncertain information and discovering knowledge and rules hid in data or information [Z. Pawlak, 1982]. Besides, knowledge or attribute reduction is one of the kernel parts of rough sets, and it can efficiently reduce the redundancy in knowledge base or attribute set.

For supervised learning a decision system or decision table may often be the form $A = (U, A \cup \{d\})$, where U is a nonempty finite set of objects called the universe, A is a nonempty finite set of attributes, $d \notin A$ is the decision attribute. The elements of A are called conditional attributes or simple conditions.

And a binary relation $R \subseteq X \times X$ which is reflexive (i.e. an object is in relation with itself xRx), symmetric (i.e. if xRy then yRx) and transitive (if xRy and yRz then xRz) is called an equivalence relation. The equivalence class of an element $x \in X$ consists of all objects $y \in X$ such that xRy .

Let $A = (U, A)$ be an information system, then with any $B \subseteq A$ there is associated an equivalent relation $IND_A(B) : IND_A(B) = \{(x, x') \in U^2 \mid \forall a \in B, a(x) = a(x')\}$, here $IND_A(B)$ is called B -indiscernibility relation. If $(x, x') \in IND_A(B)$, then objects x and x' are indiscernible from each other by attributes from B . Then the equivalence classes of the B -indiscernibility relation are denoted $[x]_B$.

An equivalence relation induces a partitioning of the universe U . These partitions can be used to build new subsets of the universe. Subsets that are most often of interest have the same value of the outcome attribute [J. Komorowski, L. Polkowski, A. Skowron, 1998].

In a word, rough set is a powerful tool of data analysis with the virtue as follows: a) not using the prior knowledge, traditional analysis methods (i.e. fuzzy set, probability and

statistics method) are also used to process uncertain information, but it is necessary for them to provide additive information or prior knowledge. However, rough set only make use of information in data set. b) expressing and processing uncertain information effectively, on the basis of equivalent relation and indiscernibility relation it can reduce redundant information and gain the minimal reduction of knowledge or attribute and discover simplifying knowledge and rules. c) missing value, rough set can be avoided of the effects because of missing value in data set. d) high performance, it can rapidly and efficiently process the large number of data with many variables or attributes.

3 A Rough Set-Based MCLP Approach for Classification

3.1 The Limitation of Rough Sets and MCLP Methods

Although rough set has many advantages just like the above mentioned, it is short of the fault tolerance and generalization in new data case. Besides, it only deals with discrete data. However, MCLP model good at those aspects.

Similarly, MCLP model can gain the better compromise solution on condition that it got the better control of the trade-off of between minimizing the overlapping degree and maximizing the distance departed from boundary. That is to say it do not attempt to get optimal solution but to gain the better generalization by means of using regret measurement and to seek for non inferior solution. Nevertheless, MCLP model only can deal with continuous and discrete data with numeric type. In addition, it also failed to treat with missing value and get the reduced conditional attribute set by the aid of its function.

3.2 Rough Set-Based MCLP Approach for Classification

According to the above analysis, we can find the difference and the existing mutual complementarities between them and combine rough set and MCLP model in data mining.

In general, MCLP model can not reduce the dimensions of input information space. Moreover, it result in too long training time when the dimensions of input information space is too big, and to some extent the model will not gain solution of primal problems. However, rough set theory can discover the hided relation among data, remove redundant information and gain a better dimensionality reduction.

In practical applications, because rough set is sensible to the noise of data, the performance of model will be the bad when we apply the results learning from data set without noise to data set including noise. That is to say, rough set have the poor generalization. Nevertheless, MCLP model is provided with the good noise suppression and generalization.

Therefore, according to their characteristics of the complementarities, the integration of rough set and MCLP model will produce a new hybrid model or system. At the same time, rough set is used as a prefixion part which is responsible for data preprocessing and MCLP will be regarded as the classification and prediction system where use the reduced information by rough set in the model or system. Consequently, the system structure of the rough set-based MCLP model for classification and prediction may be presented as follows in Figure 1:

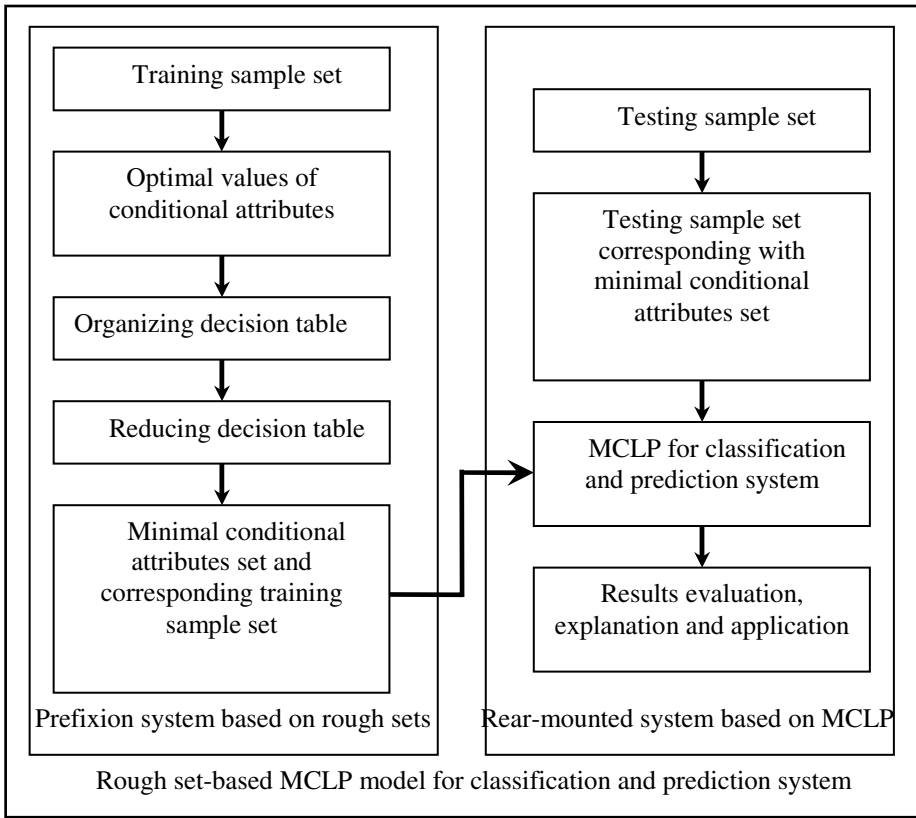


Fig. 1. Rough set-based MCLP model for classification

Firstly, we derive the attributes or variables from the source data sets collected on the basis of classification requirements and quantify these attributes. Then a data set which is composed of quantified attributes or variables may be represented as a table, where each row represents a case, an event, an observation, or simply an object. Every column represents an attribute or a variable that can be measured for each object, therefore, this table is called a decision table (or a decision system, information system). In addition, attributes may be categorized into conditional attributes and decision attributes. Meanwhile, reduction in decision table includes both reducing conditional attributes and decision rules. As far as the conditional attribute reduction is concerned, we need to check the consistency of decision table after removing an attribute. If the decision table is consistent, we will remove such attribute and finally gain the minimal attribute set. However, as for decision rule reduction we must examine the rest of training set to find which attribute is redundant after deleting the repetitions information in sample set, and then the minimal decision table or system will be produced by means of removing redundant and repetitious information. Of course, we may complete decision rule reduction ahead of the conditional attribute reduction and finally get the minimal decision table. In succession, we need create new training set

based on the minimal decision attribute and the corresponding primal data over again, where the set preserve the important attributes have an affects on performance of classification model. Then we use the data set to learn and train the MCLP model. Similarly, we need to create new testing set based on the minimal decision attribute and the corresponding primal data. Finally, we use the data set to test classifier learning from the above data set, get the results of prediction and give evaluation and explanation for these results.

3.3 The Characteristics of the Rough Set-Based MCLP Approach

As for the rough set-based on MCLP approach, data preprocessing which is used rough set method leads to the minimal attributes set and the absent redundant information. Accordingly, the quantity of data which are used for MCLP model will be reduced and the speed of the system is increased remarkably.

On the other hand, in this approach we regard MCLP model as rear-mounted system, which will possess better fault tolerance and interference suppression capabilities, because we apply rough set method to remove the correlative attributes and the overlapping information. That is to say, the dimensional reduction leads to creating a good comprehensive evaluation function and outputting assessment results for us.

4 Algorithm of Rough Set-Based MCLP Approach

In this part, we give an algorithm which implements the Rough Set-based MCLP model for classification (RS-MCLP), describing it in the following:

Input: $A = (a_1, a_2, \dots, a_n)$, // Sample observations of data for the variables

$Y = \{0,1\} \in R$, // Target variable

b , // Initial boundary value

α^* , // Negative sum of the overlapping degree

β^* // Sum of the distance where A_i departed from its adjusted boundary b

Output:

$X = (x_1, x_2, \dots, x_m) \in R^m$ ($m \leq n$), // Weights of the minimal attribute set

α_i ($1 \leq i \leq l$), // The overlapping degree of each observation or case

β_i ($1 \leq i \leq l$) // The distance departed from boundary of each observation

Processing flow: // Rough Set-based MCLP model for classification

Step one: Data preprocessing and using Rough Set theory to compute and get the minimal attribute set:

(1) Discretize all of the continuous variables and merge all of the unreasonable intervals of discrete variables using ChiMerge technologies or other related methods. [Liu, H. & H. Motoda, 1998].

(2) Compute the minimal attribute set $A' = (a_1, a_2, \dots, a_m)$ ($m \leq n$) using Rough Set theory and methods on basis of the discrete variable or attribute set.

Step two: Firstly partition data set, and then compute the weights of variables using MCLP model and gain the ordering values of each observation in classification:

(1) Divide the data set with the minimal attribute set into training set and testing set independently.

(2) Train and Learn MCLP model which is described in model (5) above on training set.

(3) Check and validate the performance of classification. If the result of classification is satisfied, the flow goes on; otherwise, change the parameters of models and turn the step to (2).

(4) In the end, compute the results of classification on testing set.

Step three: Transform the results above into probabilities and evaluate the classification accuracy rate:

(1) According to the results, we use logistic regression to compute the probabilities of each instance. That is: $\log(p/(1-p)) = A'X + \varepsilon$, where p is probability of target class, A' is the minimal attribute set obtained by Rough Set methods, X is the weights or coefficients of variables gained by MCLP approach and ε is an interception item.

(2) Compute and give the evaluation of classification, namely type I and II error rate, the total misclassification rate, KS index and Gini index as well.

5 Experimentation and Comparison of Results

In this experimentation, data sets source from the UCI Knowledge Discovery in Database Archive, an online repository of large data sets which encompasses a wide variety of data types. In the repository, we select five data sets related with medical diagnosis and treatments: 1) breast cancer data: the table have 286 instances, 9 attributes or variables (i.e. age, menopause, tumor size, node cap, malignant degree, left or right, etc.) and 1 class attribute which show breast cancer whether is recurrent or not; 2) heart disease data: the data set includes 270 instances, 13 attributes (i.e. age, sex, chest pain type, blood pressure, blood sugar, heart rate, if exercise induced angina, etc.) and 1 target attribute; 3) Lung Cancer data: the data set includes 32 instances, 56 variables and 1 class attribute which possesses 3 types. 4) Wisconsin breast cancer data: the data set includes 699 instances, 9 attributes (i.e. clump thickness, cell Size, cell shape, marginal adhesion, bare nuclei, bland chromatin, normal nucleoli, mitoses, etc.) and 1 target variable which shows the cancer is benign or malignant; 5) SPECTF Heart data: the dataset describes diagnosing cardiac Single Proton Emission Computed Tomography (SPECT) images. Each of the patients is classified into two categories: normal and abnormal. And the database of 267 SPECT image sets (patients) was processed to extract features that summarize the original SPECT images.

After we choose the above data sets, first each table is divided into two parts: training set and testing set. Then we turn to train MCLP model using the training set respectively, which is built on SAS development environment, and the system rapidly give the solutions of the models. Afterwards, when the results of models are applied to testing set, we will obtain the classification result and the corresponding evaluation indexes. As is represented in the Table 1, we provide the detailed results: the number of attributes used by model, type I error rate, type II error rate and the total misclassification rate as well.

Table 1. The results and comparison of MCLP model and Rough Set-base MCLP model

Model	Data Set	Number of Attributes	Type I Error Rate (%)	Type II Error Rate (%)	Mis-Classification Rate (%)
MCLP	Breast Cancer	9	29	36	35
	Heart Disease	13	66	17	39
	Lung Cancer	56	67	15	36
	WBreast Cancer	9	3	19	11
	SPECTF Heart	44	41	56	24
RS-MCLP	Breast Cancer	7	28	36	32
	Heart Disease	5	36	20	27
	Lung Cancer	3	22	30	26
	WBreast Cancer	5	2	20	11
	SPECTF Heart	3	33	29	23

For the sake of comparison and validation our new model, we also train Rough Set-based MCLP (RS-MCLP) model in similar way. However, the difference lies in: first we discretize the related attributes in data set where includes some continuous attributes, and then get the reduction decision tables using rough set theory and methods. Furthermore, we train MCLP model on the reduction data set again, and the classification results and evaluation indexes are listed in Table 1. At the same time, for the convenient comparison, the number of variable is provided too. In addition, we use these parameters in MCLP, and they are: the overlapping degree $\alpha^* = 0.01$, the distance departed from its adjusted boundary $\beta^* = 300000$ and the class boundary $b = 1$.

In a word, through the comparison with the classification results of the different models, we find the accuracy of classification is not decreasing. However, the RS-MCLP approach can significantly reduce the number of variable in model and the computational complexity.

6 Conclusions and Future Work

This paper provides a new data mining model and its applications in the different decision systems or tables, and experiments show that the rough set-based MCLP model for classification is prior to single MCLP model and Rough Set method. That is to say, after rough set attribute reduction and its removing redundant information, the speed and the performance of MCLP model will be considerably improved and increased. Besides, we plan to implement a rough set-based fuzzy MCLP model for classification and attempt to extend it to regression methods or unsupervised learning approaches in the future.

Acknowledgements. This research has been partially supported by a grant from National Natural Science Foundation of China (#70621001, #70531040, #70501030, #70472074, #10601064), 973 Project #2004CB720103, Ministry of Science and Technology, China, and BHP Billiton Co., Australia.

References

1. Komorowski, J., Polkowski, L., Skowron, A.: *Rough Sets: A Tutorial*. Springer, Heidelberg (1998)
2. Maimon, O., Rokach, L.: *Data Mining and Knowledge Discovery Handbook*, pp. 535–558. Springer, Heidelberg (2005)
3. Sawaragi, Y., Nakayama, H., Tanino, T.: *Theory of Multiobjective Optimization*, Mathematics Science and Engineering, vol. 176. Academic Press, London (1985)
4. Freed, N., Glover, F.: Simple but Powerful Goal Programming Models for Discriminant Problems. *European Journal of Operational Research* 7, 44–60 (1981)
5. Shi, Y., Wise, M., Luo, M., Lin, Y.: Data Mining in Credit Card Portfolio Management: A Multiple Criteria Decision Making Approach. In: *Advance in Multiple Criteria Decision Making in the New Millennium*, pp. 427–436. Springer, Berlin (2001)
6. Kou, G., Xiantao, L., Peng, Y., Shi, Y., Morganwise, Xu, W.: Multiple criteria linear programming approach to data mining. *Models, algorithm designs and software development, Optimization Methods and Software* 18(4), 453–473 (2003)
7. Swiniarski, R.W., Skowron, A.: Rough set methods in feature selection and recognition. *Pattern Recognition Letters* 24, 833–849 (2003)
8. Pawlak, Z.: Rough sets. *International Journal of Computation and Information Sciences* 11, 341–356 (1982)