A Generic Scheme for Generating Prediction Rules Using Rough Sets

Hameed Al-Qaheri[1], Aboul Ella Hassanien[1,2], and Ajith Abraham[3]

[1] Department of Quantitative Methods and Information Systems
College of Business Administration, Kuwait University, Kuwait
[2] Information Technology Department, FCI, Cairo University
5 Ahamed Zewal Street, Orman, Giza, Egypt
alqaheri@cba.edu.kw, a.hassanien@fci-cu.edu.eg, abo@cba.edu.kw
[3] Center for Quantifiable Quality of Service
Norwegian University of Science and Technology
O.S. Bragstads plass 2E, NO-7491 Trondheim, Norway
ajith.abraham@ieee.org, abraham.ajith@acm.org

Abstract. This chapter presents a generic scheme for generating prediction rules based on rough set approach for stock market prediction. To increase the efficiency of the prediction process, rough sets with Boolean reasoning discretization algorithm is used to discretize the data. Rough set reduction technique is applied to find all the reducts of the data, which contains the minimal subset of attributes that are associated with a class label for prediction. Finally, rough sets dependency rules are generated directly from all generated reducts. Rough confusion matrix is used to evaluate the performance of the predicted reducts and classes. For comparison, the results obtained using rough set approach were compared to that of artificial neural networks and decision trees. Empirical results illustrate that rough set approach achieves a higher overall prediction accuracy reaching over 97% and generates more compact and fewer rules than neural networks and decision tree algorithm.

1 Introduction

Over the last few decades statistical techniques such as regression and Bayesian models and econometric techniques have dominated the research activities in prediction. Data mining [10] and computational intelligence techniques such as neural networks, fuzzy set, evolutionary algorithms, rough set theory, machine learning, multi-criteria decision aid (MCDA), etc., emerged as alternative techniques to the conventional statistical and econometric models and techniques that have dominated this field since the 1930s [56] and have paved the road for the increased usage of these techniques in various areas of economics and finance[45, 26, 21]. Examples of the utilization of these techniques are the applications of genetic algorithms and genetic programming [22] for portfolio optimization [5], neural network in stocks selection [33] and predicting the S&P 100 index using rough sets [46] and various types of intelligent systems for making trading decisions [1, 2, 3, 8, 16, 27, 28, 29, 34, 50, 51]. Other real world

A. Abraham, R. Falcón, and R. Bello (Eds.): Rough Set Theory, SCI 174, pp. 163–186.
springerlink.com

applications in the field of finance such as credit cards assessment, country risk evaluation, credit risk assessment, corporate acquisitions[56], business failure prediction, [32, 56, 11], prediction of the financial health of the dot.com firms.[7]and bankruptcy prediction[35], customer segmentation [9] are but few examples showing the diversity of the coverage of these new techniques.

In recent years, and since its inception, rough set theory has gained momentum and has been widely used as a viable intelligent data mining and knowledge discovery technique in many applications including economic, financial and investment areas. Applications of rough sets in economic and financial prediction can be divided into three main areas: database marketing, business failure prediction and financial investment [17, 4].

Database marketing is a method of analyzing customer data to look for patterns among existing preferences and to use these patterns for a more targeted selection of the customers [17, 15]. It is based on the principle that through collecting and organizing information about a business, one can reduce the cost of the businessŠs marketing efforts and increase profit. Database marketing is characterized by enormous amounts of data at the level of the individual consumer. However, these data have to be turned into information in order to be useful. To this end, several different problem specifications can be investigated. These include market segmentation, cross-sell prediction, response modelling, customer valuation and market basket analysis. Building successful solutions for these tasks requires applying advanced data mining and machine learning techniques to find relationships and patterns in historical data and using this knowledge to predict each prospect's reaction to future situations. The rough set model has been applied in this domain (see [41, 25]).

Business failure prediction [32, 44, 56, 11], of the financial health of the dot.com firms [7] and bankruptcy prediction[35], are examples of an important and challenging issue that has served as the impetus for many academic studies over the past three decades[32]. Recently, there has been a significant increase in interest in business failure prediction, from both industry and academia. Financial organizations, such as banks, credit institutes, clients, etc. need these predictions for evaluating firms in which they have an interest[17]. Accurate business failure prediction models would be extremely valuable to many industry sectors, particularly in financial investment and lending institutes. Despite the fact that Discriminant analysis has been the most popular approach, there are also a large number of alternative techniques available such as rough sets [12, 52].

Many financial analysis applications [45] such as *financial investment* employ predictive modeling techniques, for example, statistical regression, Bayesian approach and neural networks [45, 26, 21], to create and optimize portfolios and to build trading systems. Building trading systems using the rough set model was studied by several researchers. Ziarko et al. [54], Golan and Edwards [20] applied the rough set model to discover strong trading rules from the historical database of the Toronto stock exchange. Reader may refer to [17] for a detailed review of applications of rough sets in financial domain.

Despite the many prediction attempts using rough set models, prediction still remains a challenging and difficult task to perform specially within complicated, dynamic and often stochastic areas such as economic and finance. In response to this challenge, this chapter presents a generic scheme for generating prediction rules using rough set. The scheme, which could be applied in various areas of economic and finance such as stock price movement prediction, etc., is expected to extract knowledge in the form rules to guide the decision maker in making the right decision, say buy, hold or sell in the area of stock trading and portfolio management. To increase the efficiency of the prediction process, rough sets with Boolean reasoning discretization algorithm is used to discretize the data. Rough set reduction technique is, then, applied to find all reducts of the data which contains the minimal subset of attributes that are associated with a class used label for prediction. Finally, rough set dependency rules are generated directly from all generated reducts. Rough confusion matrix is used to evaluate the performance of the predicted reducts and classes.

This chapter is organized as follows. Sect. 2 gives a brief introduction to rough sets. Sect. 3 discusses the proposed rough set prediction model in detail. Experimentation is covered in Sect. 4 including data preparation and its characteristic, analysis, results and discussion of the results and finally, conclusions are provided in Sect. 5.

2 Rough Sets: Foundations

Rough set theory , a new intelligent mathematical tool proposed by Pawlak [37, 38, 39], is based on the concept of approximation spaces and models of sets and concepts. The data in rough set theory is collected in a table called a decision table. Rows of the decision table correspond to objects, and columns correspond to features. In the data set, we also assume that a set of examples with a class label to indicate the class to which each example belongs are given. We call the class label a decision feature, the rest of the features are conditional. Let $\mathcal{O}, \mathcal{F}$ denote a set of sample objects and a set of functions representing object features, respectively. Assume that $B \subseteq \mathcal{F}, x \in \mathcal{O}$. Further, let $[x]_B$ denote:

$$[x]_B = \{y : \ x \sim_B \ y\}\,.$$

Rough set theory defines three regions based on the equivalent classes induced by the feature values: lower approximation $\underline{B}X$, upper approximation $\overline{B}X$ and boundary $BND_B(X)$. A lower approximation of a set X contains all equivalence classes $[x]_B$ that are subsets of X, and upper approximation $\overline{B}X$ contains all equivalence classes $[x]_B$ that have objects in common with X, while the boundary $BND_B(X)$ is the set $\overline{B}X \setminus \underline{B}X$, *i.e.*, the set of all objects in $\overline{B}X$ that are not contained in $\underline{B}X$. So, we can define a rough set as any set with a non-empty boundary.

The indiscernibility relation $\sim_B$ (or by Ind_B) is a fundamental principle of rough set theory. Informally, $\sim_B$ is a set of all objects that have matching descriptions. Based on the selection of B, $\sim_B$ is an equivalence relation partitions a set of objects $\mathcal{O}$ into equivalence classes. The set of all classes in a partition is denoted by $\mathcal{O}/\sim_B$ (also by $\mathcal{O}/Ind_B$). The set $\mathcal{O}/Ind_B$ is called the quotient set. Affinities between objects of interest in the set $X \subseteq \mathcal{O}$ and classes in a partition can be discovered by identifying those classes that have objects in common with X. Approximation of the set X begins by determining which elementary sets $[x]_B \in \mathcal{O}/\sim_B$ are subsets of X.

In the following subsections, we provide a brief explanation of the basic framework of rough set theory, along with some of the key definitions. For a detailed review of the basic material, reader may consult sources such as [37, 38, 39].

2.1 Information System and Approximation

Definition 1. *(Information System) Information system is a tuple (U, A), where U consists of objects and A consists of features. Every $a \in A$ corresponds to the function $a : U \to V_a$ where V_a is a's value set. In applications, we often distinguish between conditional features C and decision features D, where $C \cap D = \emptyset$. In such cases, we define decision systems (U, C, D).*

Definition 2. *(Indiscernibility Relation) Every subset of features $B \subseteq A$ induces indiscernibility relation*

$$Ind_B = \{(x, y) \in U \times U : \forall_{a\in B}\, a(x) = a(y)\}$$

For every $x \in U$, there is an equivalence class $[x]_B$ in the partition of U defined by Ind_B.

Due to the imprecision, which exists in real world data, there are sometimes conflicting classification of objects contained in a decision table. The conflicting classification occurs whenever two objects have matching descriptions, but are deemed to belong to different decision classes. In such cases, the decision table is said to contain inconsistencies.

Definition 3. *(Lower and Upper Approximation)*
In rough set theory, approximations of sets are introduced to deal with inconsistency. A rough set approximates traditional sets using a pair of sets named the lower and upper approximation of the set. Given a set $B \subseteq A$, the lower and upper approximations of a set $Y \subseteq U$, are defined by equations (1) and (2), respectively.

$$\underline{B}Y = \bigcup_{x:[x]_B \subseteq X} [x]_B. \tag{1}$$

$$\overline{B}Y = \bigcup_{x:[x]_B \cap X \neq \emptyset} [x]_B. \tag{2}$$

Definition 4. *(Lower Approximation and positive region) The positive region $POS_C(D)$ is defined by*

$$POS_C(D) = \bigcup_{X:X \in U/Ind_D} \underline{C}X.$$

$POS_C(D)$ is called the positive region of the partition U/Ind_D with respect to $C \subseteq A$, i.e.*, the set of all objects in U that can be uniquely classified by elementary sets in the partition U/Ind_D by means of C [40].*

Definition 5. *(Upper Approximation and Negative Region) The negative region $NEG_C(D)$ is defined by*

$$NEG_C(D) = U - \bigcup_{X:X \in U/Ind_D} \overline{C}X,$$

i.e.*, the set of all all objects that can be definitely ruled out as members of X.*

Definition 6. *(Boundary region) The boundary region is the difference between upper and lower approximation of a set X that consists of equivalence classes having one or more elements in common with X. It is given as follows:*

$$BND_B(X) = \underline{B}X - \bar{B}X \tag{3}$$

2.2 Reduct and Core

Often we wonder whether there are features in the information system, which are more important to the knowledge represented in the equivalence class structure than other features and whether there is a subset of features which by itself can fully characterize the knowledge in the database. Such a feature set is called a reduct. Calculation of reducts of an information system is a key issue in RS theory [38, 39, 42] and we use reducts of an information system in order to extract rule-like knowledge from an information system.

Definition 7. *(Reduct) Given a classification task related to the mapping $C \rightarrow D$, a reduct is a subset $R \subseteq C$ such that*

$$\gamma(C, D) = \gamma(R, D)$$

and none of proper subsets of R satisfies analogous equality.

Definition 8. *(Reduct Set) Given a classification task mapping a set of variables C to a set of labeling D, a reduct set is defined with respect to the power set $P(C)$ as the set $R \subseteq P(C)$ such that $Red = \{A \in P(C) : \gamma(A, D) = \gamma(C, D)\}$. That is, the reduct set is the set of all possible reducts of the equivalence relation denoted by C and D.*

Definition 9. *(Minimal Reduct) A minimal reduct $R_{minimal}$ is the reduct such that $\|R\| \leq \|A\|, \forall A \in R$. That is, the minimal reduct is the reduct of least cardinality for the equivalence relation denoted by C and D.*

Definition 10. *(Core) Attribute $c \in C$ is a core feature with respect to D, if and only if it belongs to all the reducts. We denote the set of all core features by $Core(C)$. If we denote by $R(C)$ the set of all reducts, we can put:*

$$Core(C) = \bigcap_{R \in R(C)} R \tag{4}$$

The computation of the reducts and the core of the condition features from a decision table is a way of selecting relevant features. It is a global method in the sense that the resultant reduct represents the minimal set of features which are necessary to maintain the same classification power given by the original and complete set of features. A straight forward method for selecting relevant features is to assign a measure of relevance to each feature and then select the features with higher values. And based on the generated reduct system, we generate a list of rules that will be used for building the classifier model which will be able to identify new objects and assign them the correct class label corresponding decision class in the reduced decision table (i.e. the reduct system). Needless to say, the calculation of all the reducts is fairly complex (see [47, 23, 48]).

2.3 Significance of the Attribute

The significance of features enables us to evaluate features by assigning a real number from the closed interval [0,1], expressing the important a feature in an information table. Significance of a feature a in a decision table DT can be evaluated by measuring the effect of removing of the feature a in C from feature set C on a positive region defined by the table DT. As shown in definition 2.3, the number $\gamma(C, D)$ express the degree of dependency between feature C and D or accuracy of approximation of U/D by C.. The formal definition of the significant is given as follows:

Definition 11. *(Significance) For any feature $a \in C$, we define its significance ζ with respect to D as follows:*

$$\zeta(a, C, D) = \frac{|POS_{C \setminus \{a\}}(D)|}{|POS_C(D)|} \tag{5}$$

Definitions 7-11 are used to express the importance of particular features in building the classification model. For a comprehensive study, reader may consult [49]. An important measure is to use frequency of occurrence of features in reducts. One can also consider various modifications of Definition 7, for example approximate reducts, which preserve information about decisions only to some degree [47]. Further more, positive region in Definition 4 can be modified by allowing for the approximate satisfaction of inclusion $[x]_C \subseteq [x]_D$, as proposed, e.g., in VPRS model [53]. Finally, in Definition 2, the meaning of $IND(B)$ and $[x]_B$ can be changed by replacing equivalence relation with similarity relation, especially useful when considering numeric features. For further reading, see [38, 42].

2.4 Decision Rules

In the context of supervised learning, an important task is the discovery of classification rules from the data provided in the decision tables. These decision rules not only capture patterns hidden in the data but also can be used to classify new unseen objects. Rules represent dependencies in the dataset, and represent extracted knowledge, which can be used when classifying new objects not present in the original information system. Once reducts were found, the job of creating definite rules for the value of the decision feature of the information system is practically done. To transform a reduct into a rule, one has to bind the condition feature values of the object class from which the reduct originated to the corresponding features of the reduct. To complete the rule, a decision part comprising the resulting part of the rule is added. This is done in the same way as for the condition features. To classify objects, which has never been seen before, rules generated from a training set are used. These rules represent the actual classifier. This classifier is used to predict classes to which new objects are attached. The nearest matching rule is determined as the one whose condition part differs from the feature vector of re-object by the minimum number of features. When there is more than one matching rule, a voting mechanism is used to choose the decision value. Every matched rule contributes votes to its decision value, which are equal to the number of times objects are matched by the rule. The votes are added and the decision with the largest number of votes is chosen as the correct class. Quality measures associated with decision rules can be used to eliminate some of the decision rules.

3 Rough Set Prediction Model (RSPM)

Figure 1 illustrates the overall steps in the proposed Rough Set Prediction Model(RSPM) using a UML Activity Diagram where a square or rectangular represents a data object, a rounded rectangular represents an activity, solid and dashed directed lines indicate control flow and data object flow respectively. Functionally, RSPM can be partitioned into three distinct phases:

- *Pre-processing phase(Activities in Dark Gray).* This phase includes tasks such as extra variables addition and computation, decision classes assignments, data cleansing, completeness, correctness, attribute creation, attribute selection and discretization.
- *Analysis and Rule Generating Phase(Activities in Light Gray).* This phase includes the generation of preliminary knowledge, such as computation of object reducts from data, derivation of rules from reducts, rule evaluation and prediction processes.
- *Classification and Prediction phase (Activities in Lighter Gray).* This phase utilize the rules generated from the previous phase to predict the stock price movement

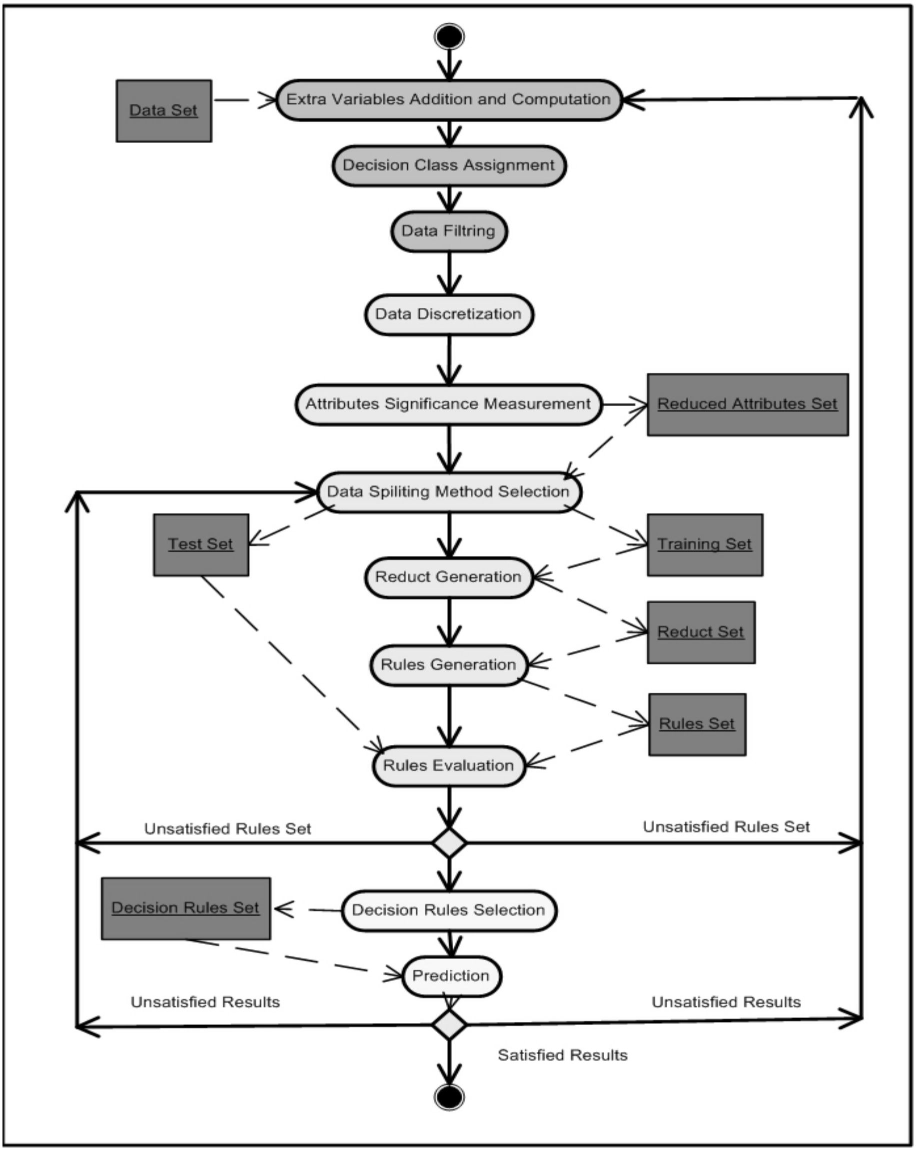

Fig. 1. General overview of rough set prediction model

3.1 Pre-processing Phase

In this phase, the decision table required for rough set analysis is created. In doing so, a number of data preparation tasks such as data conversion, data cleansing, data completion checks, conditional attribute creation, decision attribute generation, discretization of attributes are performed. Data splitting is also performed which created two randomly generated subsets, one subset for analysis containing 75% of the objects in the data set and one validation containing the remainder 25% of the objects. It must be emphasized that data conversion performed on the initial data must generate a form in which specific rough set tools can be applied.

Data Completion and Discretization Processes

Data Completion

Often, real world data contain missing values. Since rough set classification involves mining for rules from the data, objects with missing values in the data set may have undesirable effects on the rules that are constructed. The aim of the data completion procedure is to remove all objects that have one or more missing values. Incomplete data or information systems exist broadly in practical data analysis, and approaches to complete the incomplete information system through various completion methods in the preprocessing stage are normal in data mining and knowledge discovery. However, these methods may result in distorting the original data and knowledge, and can even render the original data to be un-minable. To overcome these shortcomings inherent in the traditional methods, we used the decomposition approach for incomplete information system (i.e. decision table)proposed in [43].

Data Discretization

When dealing with attributes in concept classification and prediction, it is obvious that they may have varying importance in the problem being considered. Their importance can be pre-assumed using auxiliary knowledge about the problem and expressed by properly chosen weights. However, in the case of using the rough set approach to concept classification and prediction, it avoids any additional information aside from what is included in the information table itself. Basically, the rough set approach tries to determine from the data available in the information table whether all the attributes are of the same strength and, if not, how they differ in respect of the classifier power.

Therefore, some strategies for discretization of real valued features must be used when we need to apply learning strategies for data classification (*e.g.*, equal width and equal frequency intervals). It has been shown that the quality of learning algorithm is dependent on this strategy, which has been used for real-valued data discretization [14]. It uses data transformation procedure which involves finding cuts in the data sets that divide the data into intervals. Values lying within an interval are then mapped to the same value. Performing this process leads to reducing the size of the attributes value set and ensures that the rules that are mined are not too specific. For the discretization of continuous-valued attributes, we adopt, in this chapter, rough sets with boolean reasoning (RSBR) algorithm proposed by Zhong et al. [43] The main advantage of RSBR is that it combines discretization of real-valued attributes and classification. For the main steps of the RSBR discretization algorithm, reader may consult [4].

3.2 Analysis and Rule Generating Phase

Analysis and Rule Generating Phase includes generating preliminary knowledge, such as computation of object reducts from data, derivation of rules from reducts, and prediction processes. These stages lead towards the final goal of generating rules from information system or decision table.

Relevant Attribute Extraction and Reduction

One of the important aspects in the analysis of decision tables is the extraction and elimination of redundant attributes and also the identification of the most important attributes from the data set. Redundant attributes are attributes that could be eliminated without affecting the degree of dependency between the remaining attributes and the decision. The degree of dependency is a measure used to convey the ability to discern objects from each other. The minimum subset of attributes preserving the dependency degree is called reduct. The computation of the core and reducts from a decision table is, in a way, selecting the relevant attributes [6, 48].

In decision tables, there often exist conditional attributes that do not provide (almost) any additional information about the objects. These attributes need to be removed in order to reduce the complexity and cost of decision process [6, 18, 42, 48]. A decision table may have more than one reduct. Any of these reducts could be used to replace the original table. However, finding all the reducts from a decision table is NP-complete but fortunately, in applications, it is usually not necessary to find all of them – one or a few of them are sufficient. Selecting the best reduct is important. The selection depends on the optimality criterion associated with the attributes. If a cost function could be assigned to attributes, then the selection can be based on the combined minimum cost criteria. But in the absence of such cost function, the only source of information to select the reduct from is the contents of the table. In this chapter, we adopt the criteria that the best reducts are the those with minimal number of attributes and – if there are more such reducts – with the least number of combinations of values of its attributes cf. [6, 36].

In general, rough set theory provides useful techniques to reduce irrelevant and redundant attributes from a large database with a lot of attributes. The dependency degree (or approximation quality, classification quality) and the information entropy are two most common attribute reduction measures in rough set theory. In this chapter, we use the dependency degree measure to compute the significant features and measuring the effect of removing a feature from the feature sets. [24].

Computation of the Reducts

A reduced table can be seen as a rule set where each rule corresponds to one object of the table. The rule set can be generalized further by applying rough set value reduction method. The main idea behind this method is to drop those redundant condition values of rules and to unite those rules in the same class. Unlike most value reduction methods, which neglect the difference among the classification capabilities of condition attributes, we first remove values of those attributes that have less discrimination factors. Thus more redundant values can be reduced from decision table and more concise rules can be generated. The main steps of the Rule Generation and classification algorithm are outlined in Algorithm-1:

Algorithm 1. Reduct Generation algorithm

Input: information table (ST) with discretized real valued attribute.
Output: reduct sets $R_{final} = \{r_1 \cup r_2 \cup .... \cup r_n\}$

```
1: for each condition attribute c ∈ C do
2:     Compute the correlation factor between c and the decisions attributes D
3:     if the correlation factor > 0 then
4:         Set c as relevant attributes.
5:     end if
6: end for
7: Divide the set of relevant attribute into different variable sets.
8: for each variable sets do
9:     Compute the dependency degree and compute the classification quality
10:    Let the set with high classification accuracy and high dependency as an initial
       reduct set.
11: end for
12: for each attribute in the reduct set do
13:    Calculate the degree of dependencies between the decisions attribute and that
       attribute.
14:    Merge the attributes produced in previous step with the rest of conditional
       attributes
15:    Calculate the discrimination factors for each combination to find the highest
       discrimination factors
16:    Add the highest discrimination factors combination to the final reduct set.
17: end for
18: repeat
19:    statements 12
20: until all attributes in initial reduct set are processed
```

Rule Generation from a Reduced Table

The generated reducts are used to generate decision rules. The decision rule, at its left side, is a combination of values of attributes such that the set of (almost) all objects matching this combination have the decision value given at the rule's right side. The rule derived from reducts can be used to classify the data. The set of rules is referred to as a classifier and can be used to classify new and unseen data. The main steps of the Rule Generation and classification algorithm are outlined as Algorithm-2):

Algorithm 2. Rule Generation

Input: reduct sets $R_{final} = \{r_1 \cup r_2 \cup .... \cup r_n\}$
Output: Set of rules

```
for each reduct r do
    for each corresponding object x do
        Contract the decision rule (c_1 = v_1 ∧ c_2 = v_2 ∧ .... ∧ c_n = v_n) ⟶ d = u
        Scan the reduct r over an object x
        Construct (c_i, 1 ≤ i ≤ n)
        for every c ∈ C do
            Assign the value v to the corresponding attribute a
        end for
        Construct a decision attribute d
        Assign the value u to the corresponding decision attribute d
    end for
end for
```

The quality of rules is related to the corresponding reduct(s). We are especially interested in generating rules which cover largest parts of the universe U. Covering U with more general rules implies smaller size rule set.

3.3 Classification and Prediction Phase

Classification and prediction is the last phase of our proposed approach. We present a classification and prediction scheme based on the methods and techniques described in the previous sections. Figure 2 illustrates the classification scheme for a construction of particular classification and prediction algorithm. To transform a reduct into a rule, one only has to bind the condition feature values of the object class from which the reduct originated to the corresponding features of the reduct. Then, to complete the rule, a decision part comprising the resulting part of the rule is added. This is done in the same way as for the condition features. To classify objects, which has never been seen before, rules generated from a training set will be used. These rules represent the actual classifier. This classifier is used to predict to which classes new objects are attached. The nearest matching rule is determined as the one whose condition part differs from the feature vector of re-object by the minimum number of features. When there is more than one matching rule, we use a voting mechanism to choose the decision value. Every matched rule contributes votes to its decision value, which are equal to the t times number of objects matched by the rule. The votes are added and the decision with the largest number of votes is chosen as the correct class. Quality measures associated with decision rules can be used to eliminate some of the decision rules.

The global strength defined in [6] for rule negotiation is a rational number in [0, 1] representing the importance of the sets of decision rules relative to the considered tested object. Let us assume that $T = (U, A \bigcup (d))$ is a given decision table, u_t is a test object, $Rul(X_j)$ is the set of all calculated basic decision rules

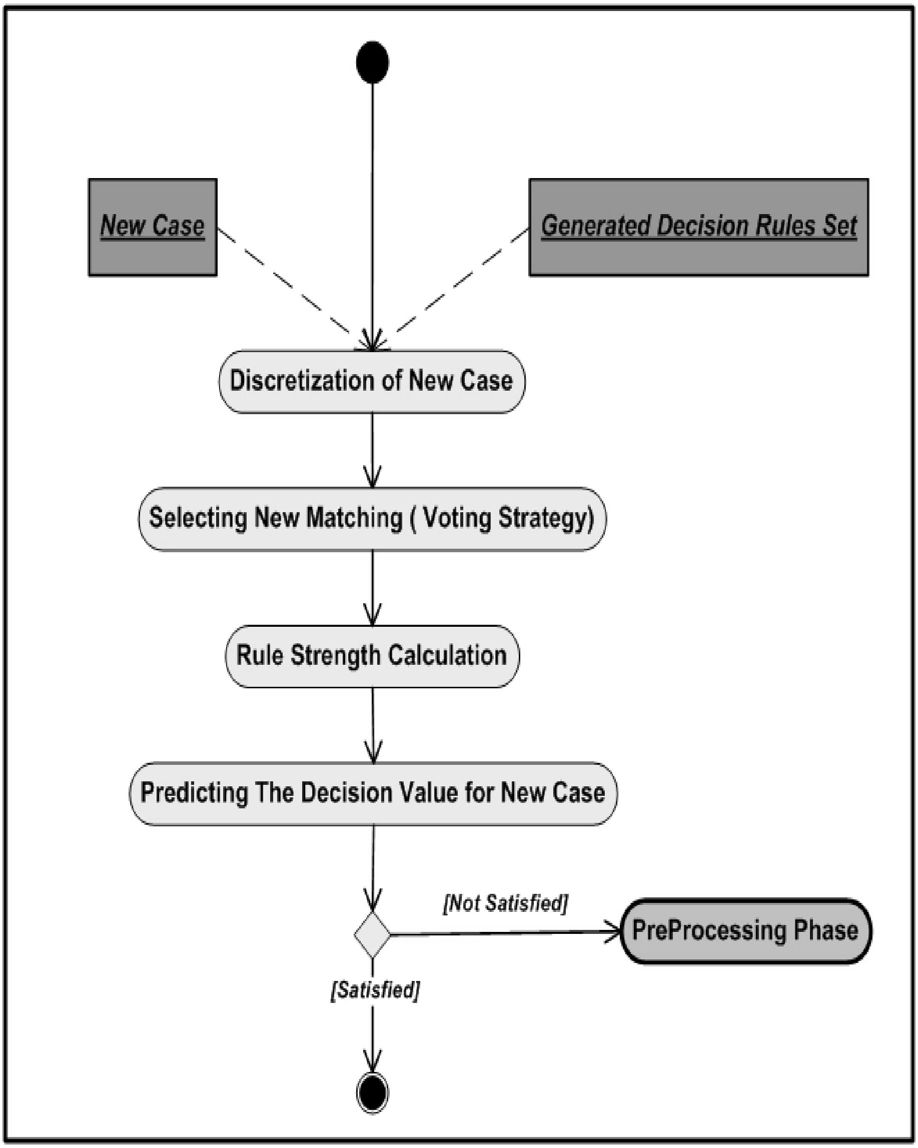

Fig. 2. Rough set classification and prediction scheme

for T, classifying objects to the decision class $X_j(v_d^j = v_d)$, $MRul(X_j, u_t) \subseteq Rul(X_j)$ is the set of all decision rules from $Rul(X_j)$ matching tested object u_t. The global strength of decision rule set $MRul(X_j, u_t)$ is defined by the following form [6]:

$$MRul(X_j, u_t) = \frac{\left|\bigcup_{r \subset MRul(X_j, u_t)} |Pred(r)|_A \cap |d = v_d^j|_A\right|}{\left||d = v_d^j|A\right|}.$$

Measure of strengths of rules defined above is applied in constructing classification algorithm. To classify a new case, rules are first selected matching the new object. The strength of the selected rule sets is calculated for any decision class, and then the decision class with maximal strength is selected, with the new object being classified to this class.

4 Experimental Results

4.1 Data Set and Its Characteristics

To test and verify the prediction capability of the proposed RSPM, the daily stock movement of a banking stock traded in Kuwait Stock Exchange and spanning over a period of 7 years (2000-2006), were captured. Figure 3 depicts a sample of the stock's daily movements.

Sector	stk	Ticker	Date	Last	High	Low	Vol	Trade	Value
Banking	102	GBK	01/02/2000	410	410	400	150000	9	60800
Banking	102	GBK	01/03/2000	405	405	405	140000	4	56700
Banking	102	GBK	01/04/2000	405	405	405	1010000	31	409050
Banking	102	GBK	01/05/2000	405	405	405	370000	7	149850
Banking	102	GBK	01/11/2000	400	400	400	130000	5	52000
Banking	102	GBK	01/12/2000	400	400	400	10410000	14	4164000
Banking	102	GBK	15/01/2000	400	400	400	990000	16	396000
Banking	102	GBK	16/01/2000	400	400	400	1450000	27	580000
Banking	102	GBK	17/01/2000	400	400	400	1740000	30	696000
Banking	102	GBK	18/01/2000	400	400	400	1550000	22	620000

Fig. 3. A sample of the stock daily movement

Table 1 shows the attributes used in the creation of the rough set decision table, where MA: Moving average of price,UP: Upward price change, Dw:Downward price change; P_i: closing price. The first five attributes in the Table, i.e. Last(or Closing Price), High, Low, Trade, and Value) were extracted from the stock daily movement. The other important attributes in the table were compiled from the literature [31] along with the formula for their computation. The decision attributed, D, in Table 1, which indicates the future direction of the the data set, is constructed using the following formula:

$$Dec_{att} = \frac{\sum_{i=1}^{i=n}((n+1) - i).sign[close(i) - close(0)]}{\sum_{i=1}^{n} i} \tag{6}$$

where close (0) is today's closing price and close (i) is the ith closing price in the future. Equation (1) specifies a range -1 to +1 for Dec_{att} A value of +1 indicate

Table 1. Stock price movement decision table

Attribute	Attribute description
Last	closing price
High	High price
Low	Low price
Trade	
Value	
Lag_i, $i = 1..6$	An event occurring at time $t + k$ $(k > 0)$
	is said to lag behind event occurring at time t,
$Aver_5$	moving average of 5 days for close price
Momentum	$P_i - P_{i-4}$
Disparity in 5 days	$\frac{P_i}{MA_5} * 100$
Price Osculiator	$OSCP = 100 - \frac{100}{1+\frac{\sum_{i=0}^{n-1} UP_{i-1}/n}{\sum_{i=0}^{n-1} DW_{i-1}/n}}$
RSI (relative strength index)	$= 100 - \frac{100}{\sum_{i=0}{n-1} UP_i/n}$
ROC	rate of change $\frac{P_i - P_{i-n}}{P_i} * 100$
D	Decision attribute

that every day up to n days in the future, the market closed higher than today. Similarly, -1 indicates that every day up to n days in the future, the market closed lower than today.

Figure 4 presents a snapshot of the 21 index for the period covering from Jan. 1st 2000 to Jan. 31th 2000, and the fluctuation of the $Dec_a tt$. Figure 5 illustrates part of the calculated daily stock movement time series data set according the attributes described in Table 1.

4.2 Analysis, Results and Discussion

For many data mining tasks, it is useful to learn about the general characteristics of the given data set and to identify the outliers - samples that are not consistent with the general behavior of the data model. Outlier detection is important because it may affect the classifier accuracy. As such we performed several descriptive statistical analysis, such as measures of central tendency and data dispersion. In our statistical analysis, we used the mean and the median to detect the outliers in our data set. Table 2 represents the statistical analysis and essential distribution of attributes, respectively.

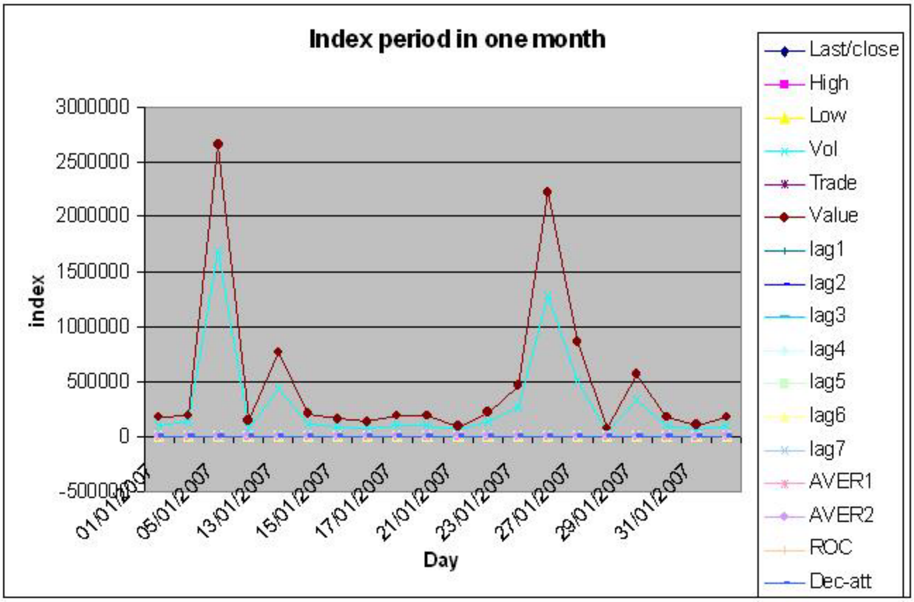

Fig. 4. Snapshot of 21 index for the period covering January 2000

Date	Last/close	High	Low	Vol	Trade	Value	lag1	lag2	lag3	lag4	lag5
01/01/2007	1720	1720	1700	105000	10	178700	1720	1720	1720	1720	1720
04/01/2007	1500	1500	1480	130000	12	194000	1480	1480	1520	1520	1520
05/01/2007	1580	1620	1560	1685000	76	2657300	1580	1620	1580	1600	1600
07/01/2007	1740	1780	1740	82500	8	143950	1780	1740	1740	1740	1700
13/01/2007	1800	1820	1760	435000	24	773000	1800	1840	1860	1840	1780
14/01/2007	1800	1800	1780	117500	9	209200	1800	1800	1840	1860	1840
15/01/2007	1780	1780	1760	95000	25	168050	1800	1800	1800	1840	1860
16/01/2007	1780	1800	1760	72500	20	128750	1780	1800	1800	1800	1840
17/01/2007	1760	1760	1740	110000	11	191600	1780	1780	1800	1800	1800
20/01/2007	1760	1760	1740	110000	11	191600	1760	1780	1780	1800	1800
21/01/2007	1760	1760	1700	55000	6	94900	1760	1760	1780	1780	1800
22/01/2007	1740	1740	1700	132500	24	226450	1760	1760	1760	1780	1780
23/01/2007	1740	1740	1700	270000	24	464050	1740	1760	1760	1760	1780
24/01/2007	1760	1780	1720	1280000	91	2216600	1740	1740	1760	1760	1760
27/01/2007	1740	1760	1720	502500	42	869100	1760	1740	1740	1760	1760
28/01/2007	1740	1740	1720	40000	3	69200	1740	1760	1740	1740	1760

Fig. 5. Samples of the banking sector data - after post processing

We reach the minimal number of reducts that contains a combination of attributes which has the same discrimination factor. The final generated reduct sets, which are used to generate the list of rules for the classification are:

$$\{\textbf{high, low, last, momentum, disparity in 5 days, Roc}\}$$

A natural use of a set of rules is to measure how well the ensemble of rules is able to classify new and unseen objects. To measure the performance of the

Table 2. Statistical results of the attributes

Attribute	Mean	Std. Dv	Median	Correlation with decision class
Last-Close	497.8	145.17	490.0	0.255
High	498.9	145.6	490	0.2500
Low	493.7	143.5	485.0	0.24
Vol	626189.3	1314775.6	240000	0.097
Trade	13.3	15.12	8.0	0.185
Value	322489.3	674862.3	118900.0	0.1065
Lag1	522.25	94.5	490.0	-0.0422
Lag2	493.8	0.4828	490.0	0.0055
Lag3	496.4	148.5	490.0	0.092
Aver5	501.5	103.6	488.0	0.075
Momentum	2.44	163.1	0.0	0.266
Disparity in 5 days	99.0	25.2	100.3	0.28
Price Osculator	.0002	0.095	0.006	0.156
RSI	49.8	1.4.36	49.8	-0.035
ROC	-4.7	21.5	0.0	-0.365

rules is to assess how well the rules perform in classifying new cases. So we apply the rules produced from the training set data to the test set data.

The following present the rules in a more readable format:

R1: IF Closing Price(Last) = (403 **OR** 408) **AND**
High = (403 **OR** 408) **AND**
Low = (3 **OR** 8) **AND**
momentum = (403 **OR** 408) **AND**
disparityin5dayes = (100.48700 **OR** 100.60700) **AND**
ROC = (−0.50505 **OR** 0.51021)
THEN Decision Class is 0.0

Table 3 shows a partial set of the generated rules. These obtained rules are used to build the prediction system.

Several runs were conducted using different setting with strength rule threshold. Rule importance and rule strength measures are used to obtain a sense of the quality of the extracted rules. These measures are chosen according to the number of times a rule appears in all reducts, number of generated reducts, and

Table 3. A partial set of the generated rules

Rule number	Rule form
R1	Last/close=(403 or 408) AND High=(403 RO 408) AND Low=(403 or 408) AND momentum=(3 OR 8) AND disparityin5dayes=(100.48700 or 100.60700) AND ROC=(-0.50505 or 0.51021) $\Longrightarrow d = 0$
R2	Last/close=(398 or 403) AND High=(398 or 403) AND Low=(393 or 398) AND momentum=(-2 or 3) AND disparityin5dayes=(125.19600 or 125.43000) AND ROC=(-0.50505 or 0.51021) $\Longrightarrow d = 0$
R3	Last/close=(403 or 408)) AND High(403 or 408) AND Low=(398 or 403) AND momentum(3 or 8) AND disparityin5dayes=(100.93900 or 101.01500) AND ROC=(0.51021) $\Longrightarrow d = 1.0$
R4	Last/close=(378 or 385) AND High(378 or 385) AND Low=(378 or 385)) AND momentum=(-25 or -17) AND disparityin5dayes=(97.70110) AND ROC=(-0.50505) $\Longrightarrow d = -1.0$
R5	Last/close=(183 or 370) AND High=(368, 373) AND Low=(183, 368) AND momentum=(-37, -32) AND disparityin5dayes=(113.76700 or 120.81700) AND ROC=(-0.50505) $\Longrightarrow d = 1.0$
R6	Last/close=(403, 408) AND High=(403 or 408) AND Low=([398 or 403)) AND momentum=(-2 or 3) AND disparityin5dayes=(100.24500 or 100.27300) AND ROC=(0.51021) $\Longrightarrow d = 1.0$

Table 4. Number of generated rules

Method	Generated rule number
Neural networks	630
Rough sets	371

Table 5. Model prediction performance (confusion matrix)

Actual	Predict Class1	Predict Class2	Predict Class3	Accuracy
Class1 (-1)	39	1	0	0.975 %
Class2 (0)	0	76	0	1.0 %
Class3 (+1)	0	2	34	0.94%
	1.0	.962	1.0	0.9802 %

the support the strength of a rule. The rule importance and Rule Strength are given by the following forms:

Rule Importance. Rule Importance measures ($Importance_{rule}$) is used to assess the quality of the generated rule and it is defined as follows:

$$Importance_{rule} = \frac{\tau_r}{\rho_r}, \tag{7}$$

where τ_r is the number of times a rule appears in all reducts and ρ_r is the number of reduct sets.

Rule Strength. The strength of a rule, $Strength_{rule}$, states how well the rule covers or represent the data set and can be calculated as follows:

$$Strength_{rule} = \frac{Support_{rule}}{|U|}, \tag{8}$$

where $|U|$ denotes the number of all objects in the training data or objects in the universe in general. The strength of a rule states how well the rule covers or represents the data set.

Table 4 shows the number of generated rules using rough sets and for the sake of comparison we have also generated rules using neural network. Table 4 indicates that the number of rules generated using neural networks is much larger than that of the rough set approach.

Measuring the performance of the rules generated from the training data set in terms of their ability to classify new and unseen objects is also important. Our measuring criteria were Rule Strength and Rule Importance [30] and to check

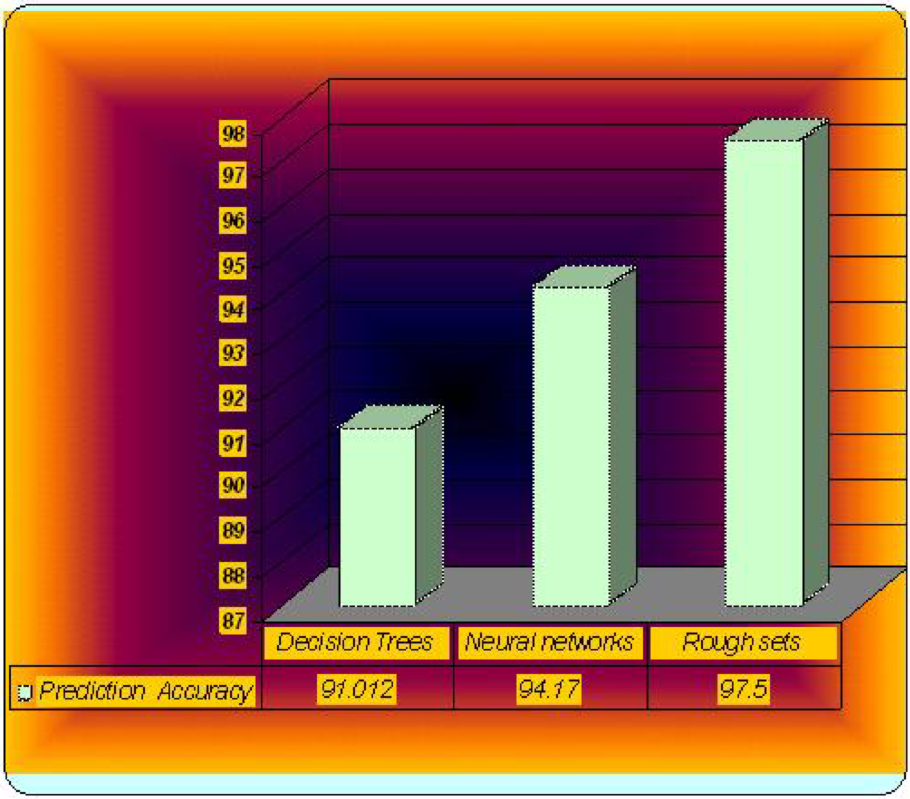

Fig. 6. Comparative analysis in terms of the prediction accuracy

the performance of our method, we calculated the confusion matrix between the predicted classes and the actual classes as shown in Table 5. The confusion matrix is a table summarizing the number of true positives, true negatives, false positives, and false negatives when using classifiers to classify the different test objects.

Figure 6 shows the overall prediction accuracy of well known two approaches compared with the proposed rough set approach. Empirical results reveal that the rough set approach is much better than neural networks and ID3 decision tree. Moreover, for the neural networks and the decision tree classifiers, more robust features are required to improve their performance.

5 Conclusions and Future Research

This chapter presented a generic stock price prediction model using rough set theory. The model was able to extract knowledge in the form of rules from daily stock movements. These rules then could be used to guide investors whether to buy, sell or hold a stock. To increase the efficiency of the prediction process, rough sets with Boolean reasoning discretization algorithm is used to discretize the data. Rough set reduction technique is, then, applied to find all reducts of the data which contains the minimal subset of attributes that are associated with a class used label for prediction. Finally, rough set dependency rules are generated directly from all generated reducts. Rough confusion matrix is used to evaluate the performance of the predicted reducts and classes.

Using a data set consisting of daily movements of a stock traded in Kuwait Stock Exchange, a preliminary assessment showed that performance of the rough set based stock price prediction model, given the limited scoped of the data set, was highly accurate and as such this investigation could lead to further research using a much larger data set consisting of the entire Kuwait Stock Exchange, which would in turn prove the model's generalizability that the model is accurate and sufficiently robust and reliable as a forecasting and prediction model. For comparison purposes, the results obtained using rough sets were compared to those generated by neural networks and decision tree algorithms. It was shown, using the same constrained data set, that rough set approach has a higher overall accuracy rates and generate more compact and fewer rules than neural networks. A future research, based on this finding, could be to implement a hybrid approach using rough sets as reducts generator and neural networks for knowledge discovery and rule generator utilizing the rough set reducts.

References

1. Abraham, A., Nath, B., Mahanti, P.K.: Hybrid intelligent systems for stock market analysis. In: Alexandrov, V.N., Dongarra, J., Juliano, B.A., Renner, R.S., Tan, C.J.K. (eds.) ICCS-ComputSci 2001. LNCS, vol. 2074, pp. 337–345. Springer, Heidelberg (2001)
2. Abraham, A., Philip, N.S., Nath, B., Saratchandran, P.: Performance analysis of connectionist paradigms for modeling chaotic behavior of stock indices. In: Second International Workshop on Intelligent Systems Design and Applications, Computational Intelligence and Applications, pp. 181–186. Dynamic Publishers Inc., USA (2002)
3. Abraham, A., Philip, N.S., Saratchandran, P.: Modeling chaotic behavior of stock indices using intelligent paradigms. Int. J. Neural Parallel Sci. Comput. 11(12), 143–160 (2003)
4. Al-Qaheri, H., Hassanien, A.E., Abraham, A.: Discovering Stock Price Prediction Rules Using Rough Sets. Neural Network World Journal (in press, 2008)
5. Bauer, R.J.: Genetic Algorithms and Investment Strategies. Wiley, New York (1994)
6. Bazan, J., Nguyen, H.S., Nguyen, S.H., Synak, P., Wróblewski, J.: Rough Set Algorithms in Classification Problem. In: Polkowski, L., Tsumoto, S., Lin, T.Y. (eds.) ICCV-WS 1999, pp. 49–88. Physica Verlag (2000)
7. Indranil, B.: Deciding the financial health of dot-coms using rough sets. Information and Management 43(7), 835–846 (2006)
8. Chan, W.S., Liu, W.N.: Diagnosing shocks in stock markets of southeast Asia, Australia, and New Zealand. Math. Comput. Simulation 59(13), 223–232 (2002)

9. Cheng, C., Chen, Y.: Classifying the segmentation of customer value via RFM model and RS theory. Expert Systems with Applications (article in progress) (2008)
10. Cios, K., Pedrycz, W., Swiniarski, R.: Data Mining Methods for Knowledge Discovery. Kluwer Academic Publishers, Dordrecht (1998)
11. Dimitras, A.I., Slowinski, R., Susmaga, R., Zopounidis, C.: Business failure prediction using rough sets. European Journal of Operational Research 114(2), 263–281 (1999)
12. Dimitras, A.I., Slowinski, R., Susmagab, R., Zopounidisa, C.: Business failure prediction using rough sets. European Journal of Operational Research 114(2), 263–280 (1999)
13. Dougherty, J., Kohavi, R., Sahami, M.: Supervised and Unsupervised Discritization of Continuous Features. In: Proc. of the XII International conference on Machine Learning, San Francisco, pp. 294–301 (1995)
14. Dougherty, J., Kohavi, R., Sahami, M.: Supervised and Unsupervised Discritization of Continuous Features. In: Proc. of the XII International conference on Machine Learning, pp. 294–301 (1995)
15. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery: An overview. In: Fayyad, U.M., et al. (eds.) Advances in Knowledge Discovery and Data Mining, p. 129. AAAI, New York (1996)
16. Francis, E.S., Cao, L.J.: Modified support vector machines in financial time series forecasting. Neurocomputing 48(14), 847–861 (2002)
17. Francis, E.H., Lixiang, S.: Economic and financial prediction using rough set model. European Journal of Operational Research 141(3), 641–659 (2002)
18. Gediga, G., Duentsch, I.: Statistical Techniques for Rough Set Data Analysis. In: Rough Sets: New Developments, pp. 545–565. Physica Verlag (2000)
19. Golan, R.: Stock market analysis utilizing rough set theory. Ph.D. Thesis, Department of Computer Science, University of Regina, Canada (1995)
20. Golan, R., Ziarko, W.: A methodology for stock market analysis utilizing rough set theory. In: Proceedings of the IEEE/IAFE Computational Intelligence for Financial Engineering, pp. 32–40 (1995)
21. Graflund, A.: Bayesian Inference Approach to Testing Mean Reversion in the Swedish Stock Market (2002) SSRN: `http://ssrn.com/abstract=250089`
22. Grosan, C., Abraham, A.: Stock Market Modeling Using Genetic Programming Ensembles, Genetic Systems Programming. In: Nedjah, N., et al. (eds.) Studies in Computational Intelligence, pp. 131–146. Springer, Germany (2006)
23. Guoyin, W.: Rough reduction in algebra view and information view. International Journal of Intelligent Systems 18(6), 679–688 (2003)
24. Hassanien, A.E., Abdelhafez, M.E., Own, H.S.: Rough Set Data Analysis in Knowledge Discovery: A Case of Kuwaiti Diabetic Children Patients, Advances in Fuzzy System (in press, 2008)
25. Jiajin, H., Chunnian, L., Chuangxin, O., Yao, Y.Y., Zhong, N.: Attribute reduction of rough sets in mining market value functions. In: IEEE/WIC International Conference Proceedings on Web Intelligence, WI 2003, pp. 470–473 (2003)
26. Jovina, R., Akhtar, J.: Backpropagation and Recurrent Neural Networks in Financial Analysis of Multiple Stock Market Returns. In: Proceedings of the 29th Hawaii International Conference on System Sciences (HICSS), Decision Support and Knowledge-Based Systems, vol. 2, p. 454 (1996)
27. Kim, K.J., Han, I.: Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. Expert Systems Applications 19(2), 125–132 (2000)

28. Leigh, W., Modani, N., Purvis, S., Roberts, T.: Stock market trading rule discovery using technical charting heuristics. Expert Systems and Applications 23(2), 155–159 (2002)
29. Leigh, W., Purvis, R., Ragusa, J.M.: Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network and genetic algorithm: a case study in romantic decision support. Decision Support Syst. 32(4), 361–377 (2002)
30. Li, J., Cercone, N.: A rough set based model to rank the importance of association rules. In: Ślęzak, D., Yao, J., Peters, J.F., Ziarko, W.P., Hu, X. (eds.) RSFDGrC 2005. LNCS, vol. 3642, pp. 109–118. Springer, Heidelberg (2005)
31. Lixiang, S., Loh, H.T.: Applying rough sets to market timing decisions. Decision support Systems journal 37, 583–597 (2004)
32. Lin, C., Wang, Y., Wu, C., Chuang, C.: Developing a business failure prediction model via RST, GRA and CBR. Expert Systems with Applications (2008) (article in press)
33. Mani, G., Quah, K.K., Mahfoud, S., Barr, D.: An analysis of neural-network forecasts from a large-scale, real-world stock selection system. In: Proceedings of the IEEE/IAFE 1995 Conference on Computational Intelligence for Financial Engineering (CIFER 1995), pp. 72–78. IEEE, Los Alamitos (1995)
34. Maqsood, M.R., Abraham, A.: Neural network ensemble method for weather forecasting. Neural Computing and Applications 13(2), 112–122 (2004)
35. McKee, T.E.: Rough Set Bankruptcy Prediction Models versus Auditor Signalling Rates. Journal of Forecasting 22(8), 569–586 (2003)
36. Ning, S., Ziarko, W., Hamilton, J., Cercone, N.: Using Rough Sets as Tools for Knowledge Discovery. In: Fayyad, U.M., Uthurusamy, R. (eds.) First International Conference on Knowledge Discovery and Data Mining KDD 1995, pp. 263–268. AAAI, Montreal (1995)
37. Pawlak, Z.: Rough Sets. Int. J. Computer and Information SCI. 11, 341–356 (1982)
38. Pawlak, Z.: Rough Sets- Theoretical aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
39. Pawlak, Z., Grzymala-Busse, J., Slowinski, R., Ziarko, W.: Rough sets. Communications of the ACM 38(11), 89–95 (1995)
40. Pawlak, Z., Skowron, A.: Rudiments of rough sets. Information Sciences 177, 3–27 (2007)
41. Poel, D.: Rough sets for database marketing. In: Polkowski, L., Skowron, A. (eds.) Rough Sets in Knowledge Discovery 2, pp. 324–335. Physica-Verlag (1998)
42. Polkowski, L.: Rough Sets: Mathematical Foundations. Physica-Verlag (2003)
43. Qizhong, Z.: An Approach to Rough Set Decomposition of Incomplete Information Systems. In: 2nd IEEE Conference on Industrial Electronics and Applications, ICIEA 2007, pp. 2455–2460 (2007)
44. Salcedo-Sanza, S., Fernandez-Villacanasa, J.L., Segovia-Vargasb, J., Bousono-Calzona, C.: Genetic programming for the prediction of insolvency in non-life insurance companies. Computers & Operations Research 32, 749–765 (2005)
45. Shuxiang, X., Ming, Z.: Data mining - an adaptive neural network model for financial analysis. In: Third International Conference on Information Technology and Applications, ICITA 2005, vol. 1(4-7), pp. 336–340 (2005)
46. Skalko, C.: Rough sets help time the OEX. Journal of Computational Intelligence in Finance 4(6), 20–27 (1996)
47. Ślęzak, D.: Various approaches to reasoning with frequency-based decision reducts: a survey. In: Polkowski, L., Tsumoto, S., Lin, T.Y. (eds.) Rough Sets in Soft Computing and Knowledge Discovery: New Developments, Physica Verlag (2000)

48. Starzyk, J.A., Dale, N., Sturtz, K.: A mathematical foundation for improved reduct generation in information systems. In: Knowledge and Information Systems Journal, pp. 131–147. Springer, Heidelberg (2000)
49. Swiniarski, R., Skowron, A.: Rough set methods in feature selection and recognition. Pattern Recognition Letters 24(6), 833–849 (2003)
50. Tsaih, R., Hsu, Y., Lai, C.C.: Forecasting S&P 500 stock index futures with a hybrid AI system. Decision Support Syst. 23(2), 161–174 (1998)
51. Van den Berg, J., Kaymak, U., Van den Bergh, W.M.: Financial markets analysis by using a probabilistic fuzzy modelling approach. Int. J. Approximate Reasoning 35(3), 291–305 (2004)
52. Wang, Z., Li, H., Deng, X.: Empirical Study of Business Failure Prediction Based on Rough Set. In: International Conference on Management Science and Engineering, ICMSE 2006, pp. 848–852 (2006)
53. Ziarko, W.: Variable Precision Rough Set Model. Journal of Computer and Systems. Sciences 46(1), 39–59 (1993)
54. Ziarko, W., Golan, R., Edwards, D.: An application of datalogic/R knowledge discovery tool to identify strong predictive rules in stock market data. In: Proceedings of AAAI Workshop on Knowledge Discovery in Databases, Washington, DC, pp. 89–101 (1993)
55. Zhong, N., Skowron, A.: Rough set based Knowledge Discovery Process. International Journal of Applied Maths and Computer Science 11(3), 603–619 (2001)
56. Zopounidis, C., Doumpos, M.: Multi-group discrimination using multi-criteria analysis: Illustrations from the field of finance. European Journal of Operational Research 139(2), 371–389 (2002)