

Rough Set Generating Prediction Rules for Stock Price Movement

Hameed Al-Qaheri, Shariffah Zamoon, and Aboul Ella Hassanien
Department Of Quantitative Methods and Information Systems,
College of Business Administration, Kuwait University, Kuwait
Email: {Abo,alqaheri,szamoon}@cba.edu.kw

Ajith Abraham
Center for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology
Email: ajith.abraham@ieee.org, abraham.ajith@acm.org

Abstract

This paper presents rough sets generating prediction rules scheme for stock price movement. The scheme 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 applied to find all the reducts of the data. 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. A comparison between the obtained results using rough sets with decision tree and neural networks algorithms have been made. Rough sets show a higher overall accuracy rates reaching over 97% and generate more compact rules.

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. During the last two decades data mining [10] and computational intelligence techniques such as neural networks, fuzzy set, evolutionary algorithms, rough set theory, machine learning, rough sets, 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[23] and have paved the road for the increased usage of these techniques in various areas of economics and finance[7, 6, 4]. Examples of the utilization of these

techniques are the applications of genetic algorithms and genetic programming [20] for portfolio optimization [1], neural network in stocks selection [2] and predicting the S&P 100 index using rough sets [3] and various types of intelligent systems for making trading decisions [17, 19]. And other real world applications in the field of finance such as credit cards assessment, country risk evaluation, credit risk assessment, corporate acquisitions[23], business failure prediction, [25, 23], prediction of the financial health of the dot.com firms.[21]and bankruptcy prediction[22], customer segmentation [24] 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 techniques 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 [5, 8].

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

2 Rough sets: Foundations

Rough sets theory provides a novel approach to knowledge description and to approximation of sets. Rough theory was introduced by Pawlak during the early 1980s [26] and elaborated in [?]. It is based on the concept of approxi-

mation spaces and models of the sets and concepts. In rough sets theory, the data is collected in a table, called a decision table. Rows of a decision table correspond to objects, and columns correspond to features. In the data set, we assume that the 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 sets 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 mainstay 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 (also called elementary sets [?]). 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 .

Here we provide a brief explanation of the basic framework of rough set theory, along with some of the key definitions. A review of this basic material can be found in sources such as [26, ?].

3 Rough Sets 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: (1) 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, (2) Analysis and Rule Generating Phase(Activities in

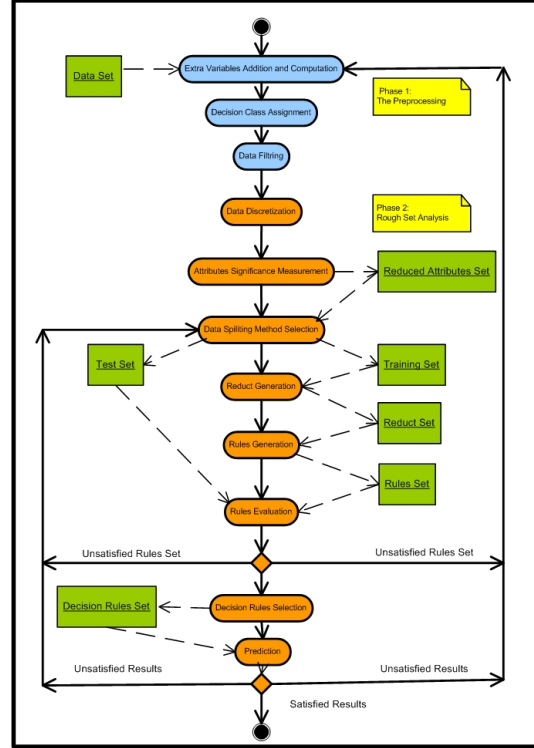


Fig. 1 Rough Sets Prediction Model

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., and (3) Classification and Prediction phase(Activities in Lighter Gray). This phase utilize the rules generated from the previous phase to predict the stock price movement.

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.

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 ob-

jects 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 [12]. 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 [27]. 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 paper, rough sets with boolean reasoning (RSBR) algorithm proposed by Zhong et al. [12] The main advantage of RSBR is that it combines discretization of real valued attributes and classification. The main steps of the RSBR discretization algorithm can be found on [8].

Analysis and Rule Generating Phase

As we mentioned before, 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. 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 relevant attributes [9, 15].

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 [9, 11, 14, 15]. A decision table may have more than one

reduct. And 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 paper, 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. [9, 13].

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 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.

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. 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. Global strength measure of strengths of rules defined in [9] 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 Experimentation

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.

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. And in our statistical analysis we used the mean and the median to detect the outliers in our data set. 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:

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

Table (1) 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

Table 1. A Partial Set of the generated rules

Rule form
Rule1 - 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) $\implies d = 0$
Rule2 - 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) $\implies d = 1.0$
Rule3 - 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) $\implies d = -1.0$

Table 2. Number of generated rules

Method	Generated rule numbers
Neural networks	630
Rough sets	371

the number of times a rule appears in all reducts, number of generated reducts, and the support the strength of a rule.

Table (2) shows the number of generated rules using rough sets and for the sake of comparison we have also generated rules using neural network. Table (2) indicates that the number of rules generated using neural networks is much larger than the rough sets.

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 [18] and to check the performance of our method, we calculated the confusion matrix between the predicted classes and the actual classes as shown in Table (3). The confusion matrix is a table sum-

Table 3. 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

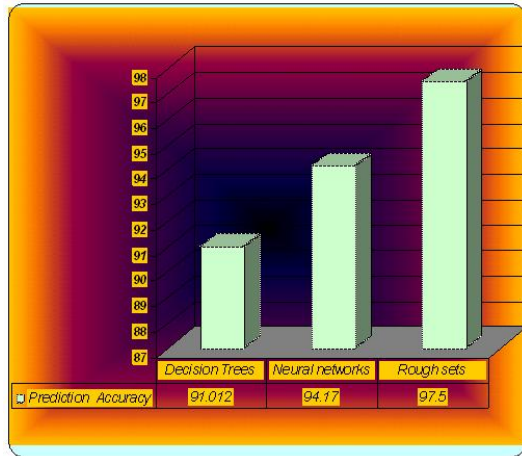


Fig. 2: Comparative analysis in terms of the prediction accuracy

marizing the number of true positives, true negatives, false positives, and false negatives when using classifiers to classify the different test objects. Figure (2) shows the overall prediction accuracy of well known two approaches compared with our rough set approach. It shows that the rough sets 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

The paper presented prediction rules for stock price movement 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 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.

References

[1] R.J. Bauer Jr., (1994) Genetic Algorithms and Investment Strategies, Wiley, New York, 1994.
 [2] G. Mani, K.K. Quah, S. Mahfoud, D. Barr (1995) An analysis of neural-network forecasts from a

large-scale, real-world stock selection system. Proceedings of the IEEE/IAFE 1995 Conference on Computational Intelligence for Financial Engineering (CIFER95), IEEE, New Jersey, 1995, pp. 72-78.
 [3] Skalko C., (1996) Rough sets help time the OEX. Journal of Computational Intelligence in Finance 4 (6) (1996) 2027.
 [4] Andreas Graflund (2002) Bayesian Inference Approach to Testing Mean Reversion in the Swedish Stock Market. Available at SSRN: <http://ssrn.com/abstract=250089>
 [5] Francis E. H. Tay, and Lixiang Shen (2002) Economic and financial prediction using rough sets model. European Journal of Operational Research Volume 141, Issue 3, 16 September 2002, pp. 641-659
 [6] Jovina Roman and Akhtar Jameel (1996) Backpropagation and Recurrent Neural Networks in Financial Analysis of Multiple Stock Market Returns. Proceedings of the 29th Hawaii International Conference on System Sciences (HICSS) Volume 2: Decision Support and Knowledge-Based Systems, pp. 454-
 [7] Shuxiang Xu and Ming Zhang (2005) Data mining - an adaptive neural network model for financial analysis. Third International Conference on Information Technology and Applications, ICITA2005, vol. 1(4-7), pp. 336-340.
 [8] Hameed Al-Qaheri, Aboul Ella Hassanien and Ajith Abraham.; Discovering Stock Price Prediction Rules Using Rough Sets. Neural Network World Journal, 2008 (to appear)
 [9] 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.), Rough Set Methods and Applications. Physica Verlag (2000) pp. 49–88.
 [10] Cios K., and Pedrycz W., and Swiniarski R., (1998) Data Mining Methods for Knowledge Discovery. Kluwer Academic Publishers, 1998.
 [11] Gediga G. Duentsch I. (2000) Statistical Techniques for Rough Set Data Analysis. In: Rough Sets: New Developments. Physica Verlag, pp. 545-565. 907-911.
 [12] Qizhong, Zhang (2007) An Approach to Rough Set Decomposition of Incomplete Information Systems. 2nd IEEE Conference on Industrial Electronics and Applications, ICIEA2007, 23-25 May 2007, pp. 2455-2460.

- [13] Ning S. Ziarko W. Hamilton J. and Cercone N. (1995) Using Rough Sets as Tools for Knowledge Discovery. In: First International Conference on Knowledge Discovery and Data Mining KDD'95, U.M. Fayyad, R. Uthurusamy (eds.), Montreal, Que., Canada, AAAI, pp.263-268.
- [14] Polkowski L. (2003) Rough Sets: Mathematical Foundations. Physica-Verlag.
- [15] Starzyk J.A. Dale N. and Sturtz K. (2000) A mathematical foundation for improved reduct generation in information systems. In Knowledge and Information Systems Journal, Springer, pp.131-147.
- [16] Lixiang Shen and Han Tong Loh (2004) Applying rough sets to market timing decisions. Decision support Systems journal, vol. 37, pp. 583-597.
- [17] A. Abraham, B. Nath and P.K. Mahanti In: V.N. Alexandrov et al., Editors, Hybrid Intelligent Systems for Stock Market Analysis, Computational Science, Springer-Verlag, Germany, USA (2001), pp. 337345
- [18] Li J. and Cercone N. (2005) A Rough Set Based Model to Rank the Importance of Association Rules. In: RSFDGrC'2005 (2), LNAI, Springer, pp.109-118.
- [19] J. Van den Berg, U. Kaymak and W.-M. Van den Bergh, Financial markets analysis by using a probabilistic fuzzy modelling approach, Int. J. Approximate Reasoning 35 (2004) (3), pp. 291305
- [20] Grosan,C. and A. Abraham, Stock Market Modeling Using Genetic Programming Ensembles, Genetic Systems Programming, Nadia Nedjah et al. (Eds.), Studies in Computational Intelligence, Springer Verlag Germany, ISBN: 3-540-29849-5, pp. 131-146, 2006.
- [21] Bose, Indranil (2006). Deciding the financial health of dot-coms using rough sets. Information and Management. Vol. 43, Iss7, pp. 835 - 846.
- [22] Thomas E McKee(2003) Rough Sets Bankruptcy Prediction Models versus Auditor Signalling Rates. Journal of Forecasting. Vol 22, Iss 8, pp 569-586
- [23] Zopounidis, C., Doumpos, M.(2002) Multi-group discrimination using multi-criteria analysis: Illustrations from the field of finance. European Journal of Operational Research. Vol. 139, Iss. 2, pp. 371-389.
- [24] Cheng, Ching-Hsue, Chen, You-Shyang. Classifying the segmentation of customer value via RFM model and RS theory. Expert Systems with Applications xxx (2008) xxx-xxx (article in progress)
- [25] Lin, Ching-Hsue, Wang, Yao-Tien, Wu, Chih-Hung, Chuang, Chun-Ling (2008) Developing a business failure prediction model via RST, GRA and CBR. Expert Systems with Applications xxx (2008) xxx-xxx (article in press)
- [26] Z. Pawlak, *Rough Sets. Theoretical Aspects of Reasoning About Data*, Kluwer, The Netherlands, 1991.
- [27] J. Dougherty, R. Kohavi, M. Sahami, "Supervised and Unsupervised Discretization of Continuous Features", in *Proc. of the XII International conference on Machine Learning*, pp. 294-301, 1995.