The extraction of trading rules from stock market data using rough sets

Kyoung-jae Kim and Ingoo Han

Graduate School of Management, Korea Advanced Institute of Science and Technology, 207-43 Cheongryangri-Dong, Dongdaemun-Gu, Seoul 130-012, Korea
E-mail: kkjkgsm.kaist.ac.kr

Abstract: We propose the rough set approach to the extraction of trading rules for discriminating between bullish and bearish patterns in the stock market. Rough set theory is quite valuable for extracting trading rules because it can be used to discover dependences in data while reducing the effect of superfluous factors in noisy data. In addition, it does not generate a signal to trade when the pattern of the market is uncertain because the selection of reducts and the extraction of rules are controlled by the strength of each reduct and rule. The experimental results are encouraging and show the usefulness of the rough set approach for stock market analysis with respect to profitability.

Keywords: rough sets, trading rules, stock market timing

1. Introduction

Stock market prediction is the long-cherished desire of investors, speculators and industries. Although many studies have investigated the prediction of price movements in the stock market, financial time series are too complex and noisy to forecast. Many researchers have predicted price movements in the stock market using artificial intelligence (AI) techniques during past decades.

The earliest studies are mainly focused on applications of artificial neural networks (ANNs) to stock market prediction (Ahmadi, 1990; Kamijo & Tanigawa, 1990; Kimoto et al., 1990; Yoon & Swales, 1991; Trippi & DeSieno, 1992; Choi et al., 1995). Recent research tends to hybridize several AI techniques (Hiemstra, 1995; Tsaih et al., 1998).

Some researchers tend to include novel factors in the learning process. Kohara et al. (1997) incorporated prior knowledge to improve the performance of stock market prediction. Lee and Jo (1999) developed a candlestick chart analysis expert system for predicting the best stock market timing. They reported that the average hit ratio of applied rules was 72%.

Although a vast number of papers exist addressing the predictability of stock market returns, most of the proposed models rely on accurate forecasting of the level of the underlying stock index or its return. However, it is more important to detect market timing, when to buy and sell stocks, than to predict daily price movement because trade driven by a certain forecast with a small forecast error may not be as profitable as trade guided by accurate market timing. Forecasting methods based on minimizing forecasting error may not be adequate for meeting investors’ objectives (Leung et al., 2000). In addition, investors in the stock market generally do not trade every day. If they trade their equities every day, they are charged a tremendously high fee for trade.

This study proposes a rule-based technique, the so-called rough set approach, to detect stock market timing. Rough set theory is quite valuable for extracting trading rules. First, it can generate profitable rules of market timing because it not only handles noise well but also eliminates irrelevant factors (Ruggiero, 1997). Second, the rough set approach is appropriate for detecting market timing because it does not generate a signal for trade when the pattern of the market is uncertain, using the control function of rough sets. In addition, it does not make any assumption about the distribution of data.

The rest of the paper is organized into five sections. The next section reviews the basic concept of rough set theory and the applications in business and finance. Section 3 describes the method of trading rule extraction and reviews the concept of market timing in the stock market. In the fourth section, we describe the research data and experiments. In the fifth section, empirical results are summarized and discussed. Conclusions and future research issues are presented in the final section.

2. Rough sets and their applications in business

The following presents some basic concepts of rough set theory as described by prior research. A detailed explanation may be found in the references in this paper.
2.1. Basic concepts

The concept of rough sets proposed by Pawlak (1982) assumes that there is some information which can be associated with every object of the universe. The central concept behind rough sets is collections of rows that have the same values for one or more attributes. These sets, called elementary sets, are said to be indiscernible (Ruggiero, 1997). The so-called indiscernible relation from this view is the mathematical basis of rough sets.

The elementary set forms a basic granule of knowledge about the universe. Any subset of the universe can be expressed either precisely or roughly (Dimitras et al., 1999). In the latter case, a certain subset can be characterized by two ordinary sets which are called the lower and the upper approximation. For each concept $Y$, the greatest definable set contained in $Y$ and the least definable set containing $Y$ can be computed. The former set is called the lower approximation of $Y$ (Pawlak et al., 1995). The lower approximation consists of all of the objects that certainly belong to the concept while the upper approximation consists of any objects that possibly belong to the concept (Ruggiero, 1997). The concept $Y$ is called rough if the upper approximation is not equal to the lower approximation (Jagielska et al., 1999).

In general, knowledge about objects can be represented in the shape of an information table. The rows of the table are objects, the columns are attributes and the intersections of rows and columns are filled with attribute values. The important point in the rough set approach is the discovery of the dependences between attributes. Another important concept is the reduction of attributes. This process is performed for finding the smallest subset assuring the same quality of classification. The attributes in the information table are composed of condition attributes and decision attributes.

The reduct in rough set theory is the set which contains the same elements as other sets but possesses the fewest attributes (Ruggiero, 1997). Elimination of superfluous attributes helps to extract strong and non-redundant classification rules (Jagielska et al., 1999). The intersection of all the reducts is called the core. The core is a collection of the most relevant and strong attributes (Dimitras et al., 1999).

2.2. A formal treatment

As mentioned earlier, data are often presented in a table using columns, rows and entries. Such tables are known as information tables or information systems (Pawlak, 1997). The information table is composed of 4-tuples, $S = (U, Q, V, f)$, where $U = \{x_1, x_2, \ldots, x_n\}$ is a finite set of objects, $Q$ is a finite set of attributes, $V = \bigcup_{q \in Q} V_q$ is a set of attribute values and $V_q$ is the set of values of attribute $q$, and $f: U \times Q \rightarrow V$ is a total function such that $f(x_i, q) \in V_q$ for every $q \in Q, x_i \in U$, called an information function (Pawlak, 1991). Any pair $(q, v)$ from the sets $q \in Q, v \in V_q$ is a descriptor in $S$. The attributes in $Q$ are composed of two disjoint subsets, condition attributes $C$ and decision attributes $D$, such that $Q = C \cup D$ and $C \cap D = \emptyset$.

Given an information table $(U, Q, V, f)$, let $P$ be a subset of $Q$, and let $x_i$ and $x_j$ be members of $U$. A binary relation $R(P)$, called a $P$-indiscernible relation, is defined as $R(P) = \{(x_i, x_j) \in U | \forall q \in P, f(x_i, q) = f(x_j, q)\}$. If $f(x_i, q) = f(x_j, q)$ for every $q \in P$, then $x_i$ and $x_j$ are said to be indiscernible by the set of attributes $P$.

Two equivalence relations $R(C)$ and $R(D)$ can be defined for an information table $S$. Let a concept $Y$ be an equivalence class of the relation $R(D)$; then the partition of $U$ with respect to the concept $Y$ is defined as $R^*(D) = \{Y, U - Y\} = \{Y, \neg Y\}$ (An et al., 1996).

In addition, an object $x_j$ specifies the equivalence class $[x_j]_E$ of the relation $R(C)$ based on the set of conditional attributes $C$ as follows:

$$[x_j]_E = \{x_i \in U | \forall q \in C, f(x_i, q) = f(x_j, q)\}$$

The conditional probability of a concept $Y$ on the equivalence class $[x_j]_E$ can be defined as

$$Pr(Y|[x_j]_E) = \frac{\text{card}(Y \cap [x_j]_E)}{\text{card}([x_j]_E)}$$

where card denotes cardinality of the set.

Let $Y \subseteq U$ be a subset of objects specifying a concept, and $R^*(C) = \{X_1, X_2, \ldots, X_n\} = \{[x_1]_E, [x_2]_E, \ldots, [x_n]_E\}$ be the collection of equivalence classes induced by an indiscernibility relation $R(C)$. The lower and upper approximations of a set $Y$ are formally defined as follows:

$$\text{lower approximation} = R(C)(Y) = \bigcup_{Pr(Y|X_j) > 0} \{X_j \in R^*(C)\}$$

$$\text{upper approximation} = \overline{R(C)(Y)} = \bigcup_{Pr(Y|X_j) = 1} \{X_j \in R^*(C)\}$$

In addition, the boundary region of $Y$ is defined as

$$\text{boundary region} = R(C)(Y) - \overline{R(C)(Y)}$$

If the boundary region of $Y$ is the empty set, then the set $Y$ will be called crisp with respect to $C$, and otherwise the set $Y$ will be referred to as rough with respect to $C$ (Pawlak, 1997).
2.3. The variable precision rough set model

The central problem of the original theory of rough sets is classification analysis but the classification must be fully certain. However, partially incorrect classification should be taken into account because real-world data are almost always polluted with noise and other imperfections. The variable precision rough set (VPRS) model is a generalized model of rough sets (Ziarko, 1993). The generalization is aimed at handling uncertain information and is directly derived from the original rough set model without any additional assumptions. This model extends upon the original rough set model by making use of the statistical information inherent in the data to handle incomplete and ambiguous data (An et al., 1996).

The VPRS model deals with partial classification by introducing a probability value \( \beta \) which measures the size of the largest group of objects as a proportion of the total number of objects in an equivalence class (Beynon et al., 2000). \( \beta \) is a real number in the range \((0.5, 1]\). If the value of \( \beta \) is 1, then the VPRS model is the standard model of rough sets. Thus, the standard model of rough sets becomes a special case of the VPRS model.

The standard model of rough sets does not make use of the statistical information in the boundary region. For this reason, an attempt is made to overcome this limitation by introducing a \( \beta \)-approximation space (An et al., 1996). The set approximations are defined as follows.

\[
\begin{align*}
\text{POS}_\beta(Y) &= \bigcup_{Pr(Y|X_i)=\beta} \{X_i \in R^*(C)\} \\
\text{NEG}_\beta(Y) &= \bigcup_{Pr(Y|X_i)<1-\beta} \{X_i \in R^*(C)\} \\
\text{BND}_\beta(Y) &= \bigcup_{1-\beta<Pr(Y|X_i)<\beta} \{X_i \in R^*(C)\}
\end{align*}
\]

The \( \beta \)-positive region of the set \( Y \subseteq U \) corresponds to all those elementary sets of \( U \) which can be classified into the concept \( Y \) with conditional probability \( Pr(Y|X_i) = \beta \). Similarly, the \( \beta \)-negative region of the set \( Y \) corresponds to all those elementary sets of \( U \) which can be classified into the concept \( Y \) with conditional probability \( Pr(Y|X_i) = 1 - \beta \). The difference between these two sets is referred to as the \( \beta \)-boundary region. We seek an empty boundary region for as high a value of \( \beta \) as possible (Beynon et al., 2000).

A probabilistic reduct, \( RED_\beta(C, D) \), referred to as a \( \beta \)-reduct, is a maximal independent subset of condition attributes with respect to \( Y \). The procedure for finding a single reduct is very straightforward. Let \( q \in C \) be a condition attribute. If the \( \beta \)-positive region \( POS^\beta_{C,q}(Y) \) of the set \( Y \) is the same as \( POS^\beta_{(Y)} \), then the condition attribute \( q \) is regarded as a superfluous factor and is removed. The remaining set of condition attributes is a reduct. If there is more than one reduct, then we select a ‘best’ reduct that depends on the optimality criterion associated with the attributes (An et al., 1996). Some researchers select the reduct with the smallest number of attributes or the maximum number of objects classified (the strength) (Wakulicz-Deja & Paszek, 1997), and others select the reduct on the basis of the increase in the quality of classification by successive augmentation of the subset of attributes (Moraczewski et al., 1996). Dimitras et al. (1999) select the ‘best’ reduct by the recommendation of domain experts.

If the reduct is selected, then rules are generated from it. Let \( R^*(D) = \{ Y, \neg Y \} \) be the partition induced by the set of decision attributes \( D \), and \( R^*(RED) = \{ X_1, X_2, \ldots, X_n \} \) be the collection of equivalence classes of the relation \( R(RED) \) where \( RED \) is a reduct which is a reduced set of conditional attributes \( C \). Each equivalence class \( X_i \) of the equivalence relation \( R(RED) \) is associated with a unique combination of values of attributes belonging to the reduced set of condition attributes \( RED \). This notion is referred to as the description of the equivalence class \( X_i \in R^*(RED) \). The description of the equivalence class \( X_i \) is formally represented as

\[
\text{Des}(X_i) = \bigwedge_{q \in RED} (q = f(x_i, q))
\]

where \( \bigwedge \) denotes the conjunction operator, and \( x_i \) is an object in the equivalence class \( X_i \). Thus, the descriptions of \( Y \) and \( \neg Y \) are represented as follows:

\[
\begin{align*}
\text{Des}(Y) &= (d = f(x_i, d)) \\
\text{Des}(\neg Y) &= (d \neq f(x_i, d))
\end{align*}
\]

where \( d \) is the decision attributes in \( D \) and \( x_i \in Y \).

The relationship between the partition \( R^*(RED) \) and the partition \( R^*(D) \) can be represented by the following rules:

\[
\begin{align*}
\text{Des}(X_i) \rightarrow c_i \text{ Des}(Y) & \quad \text{if } Pr(Y|X_i) = \beta \\
\text{Des}(X_i) \rightarrow c_i \text{ Des}(\neg Y) & \quad \text{if } Pr(Y|X_i) = 1 - \beta
\end{align*}
\]

where \( X_i \in R^*(RED) \) and \( c_i \) is a certainty factor. The certainty factor \( c_i \) is \( Pr(Y|X_i) \) in the first case and is \( 1 - Pr(Y|X_i) \) in the second (An et al., 1996).

2.4. Business applications of rough sets

The applications of rough sets in business were mainly for the evaluation and prediction of bankruptcy and business failure. One of the earliest applications was done by Słowinski and Zopounidis (1995). They used rough sets to evaluate bankruptcy. Słowinski et al. (1997) compared the rough set approach with discriminant
analysis for the prediction of company acquisition in Greece. In addition, Dimitras et al. (1999) compared the rough set approach with discriminant analysis and the logit model to predict business failure. Recent research tends to hybridize rough sets with ANNs. Hashemi et al. (1998) proposed a composite model of ANN and rough set components to predict a sample of bank holding patterns. They suggested the two-dimensional reduction approach. This approach reduced noise by potentially removing both attributes (horizontal reduction) and redundant records (vertical reduction). Ahn et al. (2000) also applied a similar approach to that of Hashemi et al. (1998). They predicted the failure of firms through the two-dimensional reduction approach. They compared this approach with the conventional ANN model and discriminant analysis.

The rough set approach, however, may not be proper for the evaluation of bankruptcy and business failure because it generates too many rules for the predicted cases. In Slowinski et al. (1997), 24 rules were generated from 30 cases. In addition, some cases in the holdout data may not have matching rules. This problem is crucial in the field of bankruptcy prediction because all the judgement is to be given to all new problems to come.

<table>
<thead>
<tr>
<th>Attribute number</th>
<th>Name of attribute</th>
<th>Formula</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>Stochastic %K</td>
<td>( \frac{C_t - LL_{t-5}}{HH_{t-5} - LL_{t-5}} \times 100 )</td>
<td>Achelis, 1995</td>
</tr>
<tr>
<td>A₂</td>
<td>Stochastic %D</td>
<td>( \frac{\sum_{i=0}^{n-1} %K_{t-i}}{n} )</td>
<td>Achelis, 1995</td>
</tr>
<tr>
<td>A₃</td>
<td>RSI (relative strength index)</td>
<td>( 100 - \frac{100}{1 + \frac{\sum_{i=0}^{n-1} Up_{t-i}/n}{\sum_{i=0}^{n-1} Dw_{t-i}/n}} )</td>
<td>Achelis, 1995</td>
</tr>
<tr>
<td>A₄</td>
<td>Momentum</td>
<td>( C_t - C_{t-4} )</td>
<td>Chang et al., 1996</td>
</tr>
<tr>
<td>A₅</td>
<td>ROC (rate of change)</td>
<td>( \frac{C_t}{C_{t-5}} \times 100 )</td>
<td>Murphy, 1986</td>
</tr>
<tr>
<td>A₆</td>
<td>A/D oscillator</td>
<td>( \frac{H_t - C_{t-1}}{H_t - L_t} )</td>
<td>Chang et al., 1996</td>
</tr>
<tr>
<td>A₇</td>
<td>CCI (commodity channel index)</td>
<td>( \frac{M_t - SM_t}{0.015 \times D_t} )</td>
<td>Chang et al., 1996</td>
</tr>
<tr>
<td>A₈</td>
<td>OSCP (price oscillator)</td>
<td>( \frac{MA_3 - MA_{10}}{MA_5} )</td>
<td>Achelis, 1995</td>
</tr>
<tr>
<td>A₉</td>
<td>Disparity 5 days</td>
<td>( \frac{C_t}{MA_5} \times 100 )</td>
<td>Choi, 1995</td>
</tr>
</tbody>
</table>

Notes: C, closing price; L, low price; H, high price; LLₙ, lowest price in the last n days; HHₙ, highest high price in the last n days; MA, moving average of price; Mt, \( (H_t + L_t + C_t)/3 \); SMt, \( \sum_{i=1}^{n} M_{t+i}/n \); Dt, \( \sum_{i=1}^{n} |M_{t+i} - SM_t|/n \); Up, upward price change; Dw, downward price change.
Table 2: Discretizing threshold

<table>
<thead>
<tr>
<th>Attribute number</th>
<th>Name of attribute</th>
<th>Threshold 1</th>
<th>Threshold 2</th>
<th>Threshold 3</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_1</td>
<td>Stochastic %K</td>
<td>[0, 25]</td>
<td>(25, 75]</td>
<td>(75, ∞)</td>
<td>Chang et al., 1996; Edwards and Magee, 1997</td>
</tr>
<tr>
<td>A_2</td>
<td>Stochastic %D</td>
<td>[0, 25]</td>
<td>(25, 75]</td>
<td>(75, ∞)</td>
<td>Chang et al., 1996; Edwards and Magee, 1997</td>
</tr>
<tr>
<td>A_3</td>
<td>RSI</td>
<td>[0, 30]</td>
<td>(30, 70]</td>
<td>(70, 100]</td>
<td>Murphy, 1986; Achelis, 1995; Choi, 1995</td>
</tr>
<tr>
<td>A_4</td>
<td>Momentum</td>
<td>(-∞, 0]</td>
<td>(0, ∞)</td>
<td></td>
<td>Murphy, 1986; Chang et al., 1996</td>
</tr>
<tr>
<td>A_5</td>
<td>ROC</td>
<td>(-∞, 100]</td>
<td>(100, ∞)</td>
<td></td>
<td>Murphy, 1986; Choi, 1995</td>
</tr>
<tr>
<td>A_6</td>
<td>A/D oscillator</td>
<td>(-∞, 0.5]</td>
<td>(0.5, ∞)</td>
<td></td>
<td>Chang et al., 1996</td>
</tr>
<tr>
<td>A_7</td>
<td>CCI</td>
<td>(-∞, 0]</td>
<td>(0, ∞)</td>
<td></td>
<td>Murphy, 1986; Choi, 1995; Chang et al., 1996</td>
</tr>
<tr>
<td>A_8</td>
<td>OSCP</td>
<td>(-∞, 0]</td>
<td>(0, ∞)</td>
<td></td>
<td>Achelis, 1995; Choi, 1995</td>
</tr>
<tr>
<td>A_9</td>
<td>Disparity 5 days</td>
<td>(-∞, 100]</td>
<td>(100, ∞)</td>
<td></td>
<td>Choi, 1995</td>
</tr>
</tbody>
</table>

3. Trading rule extraction and market timing

Market timing is an investment strategy which is used for the purpose of obtaining excessive return. Traditionally, excessive return is achieved by switching between asset classes in anticipation of major turning points in the stock market (Waksman et al., 1997). Detecting market timing in this study means determining when to buy and sell to get an excess return from trading. It is an alternative way of detecting market timing to extract trading rules from the stock market. The market timing system usually employs a profitable rule-base to capture the turning points.

The rough set approach generates trading rules of the general form shown in Figure 1. There are \( n \) conditions that are evaluated for each trading day. If all conditions are satisfied, then the model will generate a ‘Bull market’ signal or a ‘Bear market’ signal. \( X_n \) denotes the value of elementary conditions. The ranges of values for elementary conditions differ according to the attribute.

4. Research data and experimental design

The research data used in this study are the KOSPI 200 from May 1996 to October 1998. KOSPI 200 is the underlying index of the KOSPI 200 futures which is the first derivative instrument in the Korean stock market. The KOSPI 200 futures market started trading on 3 May 1996, and it had a daily trading volume of more than 60,000 contracts in October 1999. The underlying index, the KOSPI 200, is a market value weighted index which consists of 200 stocks selected by criteria on liquidity and their status in the industry. The 200 stocks were selected from six industrial groups including manufacturing, construction, communication, electricity and gas, distribution and services, and financing. The total number of samples is 660 trading days, from May 1996 to October 1998. Since it is different from the spot market, the futures market does not have continuity of price data. This is because the futures market has price data set by contract. The nearby contract data method is used in this research because this method is popular in futures market analysis.

The training data (modelling period) includes observations from May 1996 to November 1997 and the holdout data (validation period) includes observations from January 1998 to October 1998. About 65% of the data are used for modelling and 35% for validation. The training data are used to model trading systems using rough sets while the holdout data are used to validate the performance and generalizability of the proposed model.

Many stock market analysts have used technical and fundamental indicators. In general, fundamental indicators are used for long-term analysis while technical indicators are used for short-term analysis. This study uses technical indicators as input variables because it focuses on short-term analysis. Nine technical indicators were selected as input variables through a review of domain experts and past research. Table 1 gives the attributes selected.

In this study, the rough set approach is used to find profitable trading rules. The ROSETTA software system (Öhrn et al., 1998a, 1998b) is used to perform experiments. The data must be discretized to extract rules from the
Table 3: Summary statistics

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Max</th>
<th>Min</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A&lt;sub&gt;1&lt;/sub&gt;</td>
<td>151.02</td>
<td>0.00</td>
<td>43.69</td>
<td>33.77</td>
</tr>
<tr>
<td>A&lt;sub&gt;2&lt;/sub&gt;</td>
<td>118.57</td>
<td>0.00</td>
<td>43.65</td>
<td>28.70</td>
</tr>
<tr>
<td>A&lt;sub&gt;3&lt;/sub&gt;</td>
<td>100.00</td>
<td>0.00</td>
<td>43.58</td>
<td>29.21</td>
</tr>
<tr>
<td>A&lt;sub&gt;4&lt;/sub&gt;</td>
<td>10.90</td>
<td>-11.35</td>
<td>-0.44</td>
<td>5.24</td>
</tr>
<tr>
<td>A&lt;sub&gt;5&lt;/sub&gt;</td>
<td>129.14</td>
<td>81.51</td>
<td>99.55</td>
<td>5.86</td>
</tr>
<tr>
<td>A&lt;sub&gt;6&lt;/sub&gt;</td>
<td>1.71</td>
<td>-0.10</td>
<td>0.48</td>
<td>0.32</td>
</tr>
<tr>
<td>A&lt;sub&gt;7&lt;/sub&gt;</td>
<td>229.14</td>
<td>-212.44</td>
<td>-12.89</td>
<td>81.59</td>
</tr>
<tr>
<td>A&lt;sub&gt;8&lt;/sub&gt;</td>
<td>8.14</td>
<td>-9.09</td>
<td>-0.39</td>
<td>99.72</td>
</tr>
<tr>
<td>A&lt;sub&gt;9&lt;/sub&gt;</td>
<td>114.72</td>
<td>87.25</td>
<td>99.72</td>
<td>3.16</td>
</tr>
</tbody>
</table>

Figure 2: Example of derived reducts and their strengths.

The following criteria are used to generate rules from the several reducts in this paper. First, this study selects reducts with the largest strength. The largest strength means that many objects of historical data support the extracted rules. Examples of the derived reducts and their strength are presented in Figure 2. Twelve reducts (Reduct #1–#12 in Figure 2) were selected by this criterion. The value of strength for these reducts is 60.

Trading rules are extracted from the selected reducts. Dimitras et al. (1999) summarized methods for the induction of decision rules from decision tables. The first method is the generation of a minimal set of rules covering all objects. The second is the generation of an exhaustive set of rules consisting of all possible rules. The third is the generation of a set of strong decision rules, even partly discriminant but covering relatively many objects. In this study, we use the third method because stock market data are too noisy to generate a rule set satisfying the criteria of the first and second methods. This study permits 51% of the minimal level of discrimination because it specifies the minimum value of \( \beta \) as 0.5. If a rule with level of discrimination below 50% is used for the analysis, the rule may confuse upward patterns with downward patterns. In addition, 21 of the minimum strength for each rule is permitted for the criteria of rule generation from each reduct.

5. Experimental results

The proposed approach produces the following results. First, the core of attributes is empty. This implies that no single attribute perfectly explains the characteristics of decision classes. This may be due to the complex character-
Table 4: The profit of buying and selling simulations (assume investment of 10,000 won)

<table>
<thead>
<tr>
<th>Modelling period</th>
<th>Validation period</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Buy &amp; hold’ follower</td>
<td>4167 won</td>
</tr>
<tr>
<td>(−58.3%; 0%)</td>
<td>(−16.8%; 0%)</td>
</tr>
<tr>
<td>Rule set #1 follower</td>
<td>10,872 won</td>
</tr>
<tr>
<td>(Reduct #1: {A_1, A_2, A_6, A_8})</td>
<td>(8.7%; 67.0%)</td>
</tr>
<tr>
<td>Rule set #2 follower</td>
<td>9539 won</td>
</tr>
<tr>
<td>(Reduct #2: {A_1, A_2, A_8})</td>
<td>(−4.6%; 53.7%)</td>
</tr>
<tr>
<td>Rule set #3 follower</td>
<td>10,830 won</td>
</tr>
<tr>
<td>(Reduct #3: {A_1, A_3, A_6, A_8})</td>
<td>(8.3%; 66.6%)</td>
</tr>
<tr>
<td>Rule set #4 follower</td>
<td>10,177 won</td>
</tr>
<tr>
<td>(Reduct #4: {A_1, A_6, A_8})</td>
<td>(1.8%; 60.1%)</td>
</tr>
<tr>
<td>Rule set #5 follower</td>
<td>9881 won</td>
</tr>
<tr>
<td>(Reduct #6: {A_3, A_6, A_8})</td>
<td>(−1.2%; 57.1%)</td>
</tr>
<tr>
<td>Rule set #6 follower</td>
<td>9689 won</td>
</tr>
<tr>
<td>(Reduct #7: {A_4, A_7, A_8})</td>
<td>(−3.1%; 55.2%)</td>
</tr>
<tr>
<td>Rule set #7 follower</td>
<td>9430 won</td>
</tr>
<tr>
<td>(Reduct #10: {A_4, A_8, A_9})</td>
<td>(−5.7%; 52.6%)</td>
</tr>
<tr>
<td>Rule set #8 follower</td>
<td>9689 won</td>
</tr>
<tr>
<td>(Reduct #11: {A_4, A_7, A_8})</td>
<td>(−3.1%; 55.2%)</td>
</tr>
<tr>
<td>Rule set #9 follower</td>
<td>9430 won</td>
</tr>
<tr>
<td>(Reduct #12: {A_5, A_8, A_9})</td>
<td>(−5.7%; 52.6%)</td>
</tr>
</tbody>
</table>

Accumulated profit (the accumulated rate of return; the excess rate of return over following ‘buy and hold’ strategy)

Note: US$1 = 1200 won (approximate).

istics of the decision classes. Second, 280 reducts are obtained from the experiment.

In this study, the extracted rules were applied to our trading strategy which is described in Figure 3. In the process of applying the rules, the reducts which do not produce a profit are excluded because these reducts do not produce any profit or loss. Finally, nine reducts (Reduct #1, #2, #3, #4, #6, #9, #10, #11 and #12 in Figure 2) were selected and they were applied to the holdout data in order to validate the generalizability. Trading profits earned from simulation are presented in Table 4.

While the underlying index decreased more than 58% during the modelling period, in this study we found rules that would yield a high level of profit if they were used to trade stocks on a given set of historical data. In the derived rule sets, the rule set from Reduct #3 (Rule set #3) in Table 4 produces the best performance for the validation period. While the underlying index decreased about 17% during the validation period, the trading strategy followed by the derived rules gets 31.5% of trading profit for Reduct #3. This means that the investor who follows the rules from Reduct #3 can earn 48.3% more profit during the validation period than the investor who follows a ‘Buy and hold’ strategy.

The description of the rules of Reduct #3 is shown in Table 5. The other rule sets in Table 4 are good alternative.

‘Bull market’ signal are excluded because these reducts do not produce any profit or loss. Finally, nine reducts (Reduct #1, #2, #3, #4, #6, #9, #10, #11 and #12 in Figure 2) were selected and they were applied to the holdout data in order to validate the generalizability. Trading profits earned from simulation are presented in Table 4.

While the underlying index decreased more than 58% during the modelling period, in this study we found rules that would yield a high level of profit if they were used to trade stocks on a given set of historical data. In the derived rule sets, the rule set from Reduct #3 (Rule set #3) in Table 4 produces the best performance for the validation period. While the underlying index decreased about 17% during the validation period, the trading strategy followed by the derived rules gets 31.5% of trading profit for Reduct #3. This means that the investor who follows the rules from Reduct #3 can earn 48.3% more profit during the validation period than the investor who follows a ‘Buy and hold’ strategy.

The description of the rules of Reduct #3 is shown in Table 5. The other rule sets in Table 4 are good alternative.

If today’s signal is ‘Bull market’

And if the previous day’s decision was BUY,

THEN SELL stocks;

Else BUY stocks;

Else if today’s signal is NONE (if any rules do not match with today’s market condition) or ‘Bear market’,

And if the previous day’s decision was BUY,

Then SELL stocks;

Else do nothing (no trading)

Figure 3: Trading strategy.
trading rules although there are minor differences in simulated performance. These experimental results indicate that the rough set approach is a promising method for extracting profitable trading rules.

### 6. Concluding remarks

In this study we wished to find profitable rules using rough sets for the stock price index futures. As mentioned earlier, the rough set approach is quite valuable for extracting trading rules because it can be used to discover dependences in data while eliminating the superfluous factors in noisy data. An additional advantage of rough sets is that the trading rules generated produce trading signals only when the rules are fired. The rough set approach does not generate trading signals when the pattern of the market is uncertain because the selection of reducts and the extraction of rules are controlled by the strength of each reduct and rule. This is very important for detecting market timing because market timing is detected by capturing the major and certain turning points in data. In addition, this is also valuable because investors in the stock market generally do not trade everyday owing to the very high fee for trade. Each generated rule produces trading signals for less than 20% of all holdout data in this study.

The rules derived from the experiments are good alternative trading rules although there are minor differences in simulated performance. The experimental results are quite encouraging and prove the usefulness of rough sets for stock market analysis.

This study has some limitations. We have validated the experimental results for an 11 month period. However, a more extensive validation process is needed to generalize the experimental results. In addition, domain knowledge has been used to discretize continuous data because rough sets do not have a general guidance for discretization. It may be expected that a different performance is produced with other discretizing methods including the percentile method (Scott et al., 1997; Buhlmann, 1998), the clustering method (Kontkanen et al., 1997; Scott et al., 1997) and the entropy minimization heuristic method (Fayyad & Irani, 1993; Martens et al., 1998). Additional research is needed to compare the effects on performance of different methods for discretization.

### Acknowledgements

The authors would like to thank the Korea Science and Engineering Foundation for supporting this work under Grant 98-0102-08-01-3.

### References


---

**Table 5:** Description of derived rules from Reduct #3 (Rule set #3)

<table>
<thead>
<tr>
<th>Rule number</th>
<th>Elementary conditions</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A_1$, $A_2$, $A_6$, $A_8$, $A_{10}$, Strength</td>
<td></td>
</tr>
</tbody>
</table>

- **Rule #1:** $A_2$ $A_6$ $A_8$ ‘Bull market’ 29
- **Rule #2:** $A_2$ 1 ‘Bear market’ 21
- **Rule #3:** $A_6$ $A_8$ ‘Bull market’ 47
- **Rule #4:** $A_2$ 1 ‘Bear market’ 21
- **Rule #5:** $A_6$ 1 ‘Bear market’ 21
- **Rule #6:** $A_8$ ‘Bear market’ 33
- **Rule #7:** $A_{10}$ ‘Bear market’ 104


**The authors**

**Kyoung-jae Kim**

Kyoung-jae Kim is a researcher at the Techno-Management Research Institute in the Korea Advanced Institute of Science and Technology. He received his MS and Ph.D degrees in management information systems from the Graduate School of Management of the Korea Advanced Institute of Science and Technology and his BA degree from Chung-Ang University. He has published in *Expert Systems with Applications, Journal of Knowledge Management Research* etc. His research interests include data mining, knowledge management and intelligent agents.

**Ingoo Han**

Ingoo Han is an associate professor at the Graduate School of Management in the Korea Advanced Institute of Science and Technology. He received his PhD degree from the University of Illinois at Urbana-Champaign. He has published in *Decision Support Systems, Engineering Economists, Expert Systems with Applications, Information and Management, International Journal of Electronic Commerce, International Journal of Intelligent Systems in Accounting, Finance and Management*, and other journals. His research interests include data mining, knowledge management, and information systems audit and security.