A FUZZY-BASED ROUGH SETS CLASSIFIER FOR FORECASTING QUARTERLY PGR IN THE STOCK MARKET (PART I)

YOU-SHYANG CHEN¹, CHING-HSUE CHENG² AND DA-REN CHEN¹

¹Department of Information Management
Hwa Hsia Institute of Technology
111, Gong Jhuan Rd., Chung Ho, Taipei 235, Taiwan
{ys_chen; danny}@cc.hwh.edu.tw

²Department of Information Management
National Yunlin University of Science and Technology
123, Section 3, University Road, Touliu, Yunlin 640, Taiwan
chcheng@yuntech.edu.tw

Received September 2009; revised April 2010

ABSTRACT. For practitioners, a PGR (profit growth rate) is an effective evaluation indicator in the short- to midterm to see how big the potential power of future development is and measures the growth of future development for a target company that may be selected into investment portfolios by them. Although the PGR is one of the core financial ratios, it is used little in an academic study for researchers. Hence, the study aims to fill this knowledge gap. The study focuses mainly on solving the practical problems of forecasting quarterly PGR and offers three intelligent hybrid models to generate decision rules as knowledge-based systems for helping investors select appropriate investment portfolios. These proposed hybrid models are constituted differently by six basic components: experiential knowledge (EK), feature selection method (FSM), discretization method (DM), fuzzy set theory (FST), rule filter (RF), and rough set theory (RST), and the presentation of them are organized into two parts (papers). Part I groups the concepts, principles and expressions of them into this paper. Part II is intended to evaluate the proposed hybrid models, and an empirical case study is implemented, which will be presented in the next paper. A conclusion of the proposed hybrid models is given finally in this paper.

Keywords: Rough set theory, Fuzzy-rule similarity, Profit growth rate (PGR), Feature selection, Discretization method, Minimize entropy principal approach (MEPA)

1. Introduction. The stock market plays a major role in the price determination of the financial instrument for an existing financial product. More meaningfully, the stock market always reflects basic economic energies of a country; therefore, the flourishing of a country is bound, of course, to the growth and boom of stock market. Over the past two decades in Taiwan, high economic development has motivated a wave of prosperity from this market; at the same time, economic growth is related closely and positively to a booming stock market and boosts market development. Restated, the stock market has supported a key role and has been a trigger in the process of Taiwan’s economic development [1] from the 1960s. Unfortunately, in recent years, it is a serious challenge for Taiwan to thrive in the globally keener, competitive and dynamic financial climate. On the one hand, economic depression occurs by a specific environment of politics; on the other hand, a global financial crisis, originally caused by the USA, was encountered by Taiwan’s financial industry. Obviously, increasing the needs for searching a better way that helps investors to survive in such a serious environment is an increased trend today.
For an investment of stock investors, the most effective way in practice to assess operating performance of a specific company is to conduct a profitability analysis, which is the necessary tool for determining how to allocate the investor’s resource to ensure profits for themselves best. Thus, proper profitability analyses provide invaluable evidence concerning the earnings potential of a company and the effectiveness of management. A profitability analysis is the most significant of financial ratios, which provide a definitive evaluation of the overall effectiveness of management, based on the returns generated on sales and quarterly and/or annually investment. Financial ratios are widely used for modeling purposes by both practitioners and researchers. Many interested parties (i.e. stakeholders) involve the owners, customer, management, personnel, investors, competitors, suppliers, creditors, media, regulatory agencies, academicians, and special interest groups, each having their views in applying financial statement analysis, such as financial ratios, in their evaluations [2], particularly in investors. With the view of stock investors, profit growth rate (PGR) is one of core financial ratios and is an effective evaluation indicator for them to see how big the potential power of future development is and measures the growth of future development for a target company that may be selected to investment portfolios [3]. Generally, the growth of revenue should always be contingent on a growth of operating profit, and it is a good outcome that the former (RGR- revenue growth rate) is smaller than the latter (PGR), because that refers to the economic scale for a firm’s operation; otherwise, it refers to the poor performance of management and illness of cost control. Hence, the study aims to further explore impacts on growth for PGR on a quarterly financial statement. Basically, it is used in this study for two reasons. Firstly, the PGR is a usual and appropriate tool for measuring a firm for practitioners, such as stock investors, and it is always seen in the practice. Secondly, however, academicians rarely explore it in an academic study to forecast stock price for a short run or mid run.

Stock market climates and gains/losses can be vastly changed within seconds; thus, the accuracy of information for investment planning is very crucial. Researchers had proposed statistical methods to study related financial markets and provided satisfactory results [4] with these markets. However, their performances may depend strongly on the field of application [5], the study goal [6], or the user experience [7]. Besides, the noise caused by changing market conditions and environments requires investors to consider more market variables. Conventional forecasting methods, like statistical methods, rely on the restrictive assumption on linear separability, multivariate normality, and independence of the predictive variables [8]. Recently, due to rapid changing of information technology (IT) today, the most common tools [9] of artificial intelligence (AI) techniques for classification (i.e., prediction or forecast), such as rough set theory (RST), fuzzy set theory (FST), decision trees (DT), and artificial neural networks (ANN), have become significant research trends to both academicians and practitioners [10,11]. Particularly in intelligent hybrid systems, they are composed of several models [12] for processing the classification problems in practice. Moreover, it has been proven that such ensemble classifiers outperform the stand-alone models [8,10]. With the limitations of AI techniques in the stand-alone models, it is worthwhile to propose an intelligent hybrid system in order to amplify the advantages of the individual models and minimize their limitations. Thus, it can be observed that the interest to design and employ a variety of intelligent hybrid systems has increased significantly over the past decade [8]. Clearly, it is a field of growing importance for increasing demand for AI, fast advance in IT techniques, and processing huge amounts of digital data [13]. Furthermore, in view of the literature [14,15], the study notices that a significant trend of knowledge discovery is to build a rule-based model that can provide a reasonable and powerful explanation for interested parties. Therefore, it
is a valuable issue to find more suitable ways to construct a rule-based model based on some AI techniques that are suitably applied in the context of stock markets.

Due to having ever-skilful experiences and professional knowledge of the authors working in the financial industry for over 14 years, it is on the expected lines that the study considers improving upon models of solving quarterly PGR problems faced by stock investors. As such, the study proposes some intelligent hybrid models based on rough sets, which include experiential knowledge, feature selection methods, discretization methods, rule filter, fuzzy set theory, and rough set theory, for classifying quarterly PGR problems faced by stock investors. A trustworthy forecasting model is extremely desired by them.

Rough sets are a predictive data mining and mathematical tool, which are especially helpful in dealing with vague, incomplete, and uncertain data in decision situations and have become an important tool in AI algorithm [16]; concurrently, they have been successfully applied in many different fields, such as stock market prediction, financial data analysis, medical data analysis, voice recognition, information retrieval systems, and image processing [17]. In the meanwhile, fuzzy set theory is the core discipline of ‘soft computing’ and provides a new impetus for research in the field of AI. Thus, using it to solve quarterly PGR classification problems is a newer testing method in practice. Following that, two types of groups for the discretization method are expert discretization and automatic discretization. The former refers to a domain expert following his judgment, knowledge, or intuition, or uses norms established in the subject domain specified by the subintervals for the discretization; restated, the latter is the subintervals of the discretization defined automatically [18]. In feature selection, its significant advantages include the ability to improve classification accuracy by eliminating noise-inducing features, the risk of reducing overfitting, and the use of designing faster and more cost-effective-based models [19]. Finally, a rule filter is implemented for reducing a number of rules with low support and then refines the extracted rules by reducing the number of rules without (or rare) compromising accuracy.

In general, the four research objectives involved in the study are described in the following: (1) Conduct some suitable hybrid models, which use fewer selected attributes, fewer generated rules, higher accuracy, and more stability on forecasting, based on some refined enhanced rough sets classifiers for forecasting quarterly PGR in the stock market; (2) Examine the determinants (core attributes) of influencing the quarterly PGR and list their ranking order of importance; (3) Evaluate the effects of the discretization method and rule filter for the proposed hybrid models depending on some criteria; and (4) Generate comprehensive decision rules that are applied in knowledge-based systems by the rough sets LEM2 algorithm and provide a powerful explanation for investors.

The rest of this paper is organized in the following. First, we discuss related studies of classifying the PGR of a financial holding company. Second, the proposed hybrid models with their algorithms are introduced. Verification and comparisons of experimental results for Taiwan’s financial holding companies are described in the next section. Finally, discussion and findings are formulated, and they are followed by some conclusions.

2. Related Works. This section mainly explores the related issues of the financial holding company, which consist of profitability analysis and profit growth rate, rough set theory - rule extraction-LEM2 algorithm and rule filter, fuzzy logic and similarity function, feature selection method, and discretization method - minimize entropy principle approach and decision trees C4.5 algorithm.

2.1. Profitability analysis and profit growth rate (PGR). Since gaining profit is major topic aims of operating for a specific company, profitability analysis regarded as
one of the financial analyses is an important tool for judging whether is capable of value on investment portfolios by investors. Profitability represents the ability that company can raise earnings in a periodic time (may be monthly, quarterly, or yearly). Thus, to best allocate the investor’s investment portfolios to ensure maximal profits, to implement profitability analysis is needed for assessing operating performance of a specific company. Interestingly, financial ratios are used as the most common way of profitability analysis for academicians and practitioners [20]. Briefly, financial ratios provide a clear picture of the financial information of a company that is otherwise unclear from raw financial data. The ratios that are of the most interest to existing stockholders and potential stockholders (investors) include those ratios that focus on net income, dividends, and stockholders’ equities, and they mainly include the firm’s accounting variable and basic financial status, such as price to earning ratio, dividend yield, current ratio, earnings per share, price to book ratio, gross sales, book to market ratio, return on net worth, return on equities, etc. [21,22]. Generally, they are categorized as profitability, stability, activity, cash flow, and growth. In the study, the growth is focused and it includes RGR (revenue growth rate), PGR, sales increase ratio (SIR), year-over-year (YoY) change on growth rate, quarter-over-quarter (QoQ) change on growth rate, month-over-month (MoM) change on growth rate, etc. [23]. Consider an example of YoY; it reports the changes in a year’s worth of data, in comparison with the previous year. Although many figures are released monthly, many important indicators are released on a quarterly basis, and they are therefore reported QoQ. QoQ will tend to be more volatile than Year-over-Year figures.

Simply, PGR refers to the proportion of periodic variation quarterly of operating profits for a specific company in the study, and there are four quarters, named first, second, third and fourth quarter in a year. The equation of quarterly PGR can be calculated in the following: Quarterly PGR (%) = ((Current quarter Profit – Last quarter Profit) \times 100)/Last quarter Profit. For example, if the net profit of a specific company is $20 Millions in the first quarter of 2007 and the profit is $25 Millions in the next quarter of 2007. Then the PGR is equal to 25% in the second quarter of 2007. However, the PGR can be either positive or negative depended on which PGR is bigger between this quarter and last quarter. Furthermore, the PGR has no limit with respect to its range. The positive PGR represents good and optimistic with company for future, whereas negative PGR is bad and pessimistic. With its application, the PGR is a critical determinant of a firm’s success [24]. The higher the PGR of firms is, the better their future is [3]. The higher PGR represents increasing the operating profit, which may be motivated by a growth of sale amount of products, contingent with enlarging the market share rate, and it is regarded as optimistic with future development for a firm; at the same time, a better future of firms will trigger stock prices higher particularly. Meaningfully, this is an important implication for investment decision-making of investors in the stock market.

2.2. Rough set theory (RST). Rough set theory, first proposed by Pawlak [25], employed mathematical modeling to deal with class data classification problems, and then turned out to be a very useful tool for decision support systems, especially when hybrid data, vague concepts and uncertain data were involved in the decision process. To use the rough set process, one begins with a relational database, a table of objects with attributes, and attributes values for each object. One attribute is chosen as the decision attribute, then the rest of the attributes are the condition attributes [25]. Rough sets address the continuing problem of vagueness, uncertainty and incompleteness by applying the concept of equivalence classes to partition training instances according to specified criteria. Two partitions are formed in the mining process. The members of the partition can be formally described by unary set-theoretic operators or by successor functions for
lower approximation and upper approximation spaces from which both possible rules and certain rules can be easily derived. Thus, the RST approach is based on refusing certain set boundaries, implying that every set will be roughly defined using a lower and an upper approximation [26].

Let \( B \subseteq A \) and \( X \subseteq U \) be an information system. The set \( X \) is approximated using information contained in \( B \) by constructing lower and upper approximation sets, respectively: 
\[
\underline{BX} = \{ x | [x]_B \subseteq X \} \quad \text{(lower)}
\]
\[
\overline{BX} = \{ x | [x]_B \cap X \neq \emptyset \} \quad \text{(upper)}.
\]
The elements in \( \underline{BX} \) can be classified as members of \( X \) by the knowledge in \( B \). However, the elements in \( \overline{BX} \) can be classified as possible members of \( X \) by the knowledge in \( B \). The set \( BN_B(x) = \overline{BX} - \underline{BX} \) is called the \( B \)-boundary region of \( X \) and it consists of those objects that cannot be classified with certainty as members of \( X \) with the knowledge in \( B \). The set \( X \) is called ‘rough’ (or ‘roughly definable’) with respect to the knowledge in \( B \) if the boundary region is non-empty. Rough sets theoretic classifiers usually apply the concept of rough sets to reduce the number of attributes in a decision table [17] and to extract valid data from inconsistent decision tables. Rough sets also accept discretized (symbolic) input.

2.2.1. Rule extraction – rough sets LEM2 algorithm. A popular method is to induce a decision table transformed into rules that are focused on a minimal set of rules. In RST, decision rules are often induced from a given decision table. Rough set rule induction algorithms were implemented for the first time in a LERS (Learning from Examples) [27] system. The learning system LERS induces a set of rules from examples and classifies new examples using the set of rules induced previously by LERS. A local covering is induced by exploring the search space of blocks of attribute-value pairs which are then converted into the rule set. The LEM2 (Learning from Examples Module, version 2) [28] method is used in generation of decision rules in the study. The LEM2 algorithm for rule induction is based on computing a single local covering for each concept from a decision table.

2.2.2. Rule filter (RF). Rough set-mediated rule sets usually contain large numbers of distinct rules [29]. The large number of rules limits the classification capabilities of the rule-set as some rules are redundant or of ‘poor quality;’ therefore, some filtering-rule algorithms can be used to reduce the number of rules [30]. For example, a filtering rule solution may be based on the computed quality indices of rules in a rule-set. The quality index of each rule is computed using a specific rule quality function, which determines the strength of a rule, based on the measure of support, consistency, and coverage. The upper approximation of minimal rules is determined by removing some rules from the input rule set \( \mathcal{R} \). The heuristic is based on the assumption that the strongest rules are preferred to form the minimal coverage set. The heuristic algorithm proposed here does not seek the minimal solution for efficiency reasons. First, in the initialization step all rules are marked as ‘unused.’ For each object, \( supp \) strongest rules that cover it are identified and marked to join the resulting set. The remaining rules not used in the covering process are filtered out.

2.3. Fuzzy logic and similarity function. In the traditional crisp set, the degree of an element belonging to a set is either 0 or 1. To deal with the uncertain and imprecise data in real-world applications, Zadeh [31] first proposed fuzzy set theory (also called fuzzy logic). This theory was developed based on the concept of traditional crisp sets, and it is that an element belongs to a fuzzy set with a certain degree of membership. The FST has been successfully applied to many fields, such as fuzzy controls and fuzzy expert systems.

Furthermore, fuzzy number, linguistic variable, linguistic value, and defuzzification are introduced accordingly. The special significance of fuzzy sets are that, membership
functions are defined on the set $\mathbb{R}$ of real numbers, which have the form $\hat{A} : \mathbb{R} \to [0, 1]$ clearly have a quantitative meaning and may be viewed as fuzzy number or fuzzy interval. Zadeh [32] indicated that a linguistic variable differs from a numerical variable in that its values are not numbers, and they are words or sentences in a natural or artificial language. For example, linguistic variable is described by linguistic terms in nature, such as ‘small,’ ‘medium,’ and ‘large.’ Each linguistic variable may be assigned one or more linguistic terms (i.e., linguistic values), which are connected to a numeric value through the mechanism of membership function. Defuzzification is the last step in generating an output from a fuzzy inference system.

The definitions of related FST described by the literature [33] were introduced:

**Definition 2.1.** Let $X$ be a universe of discourse and $x$ be a generic element of $X$. A fuzzy set $A$ in $X$ is defined as a set of ordered pairs:

$$A = \{(x, \mu_A(x)) \mid x \in X\},$$

where $\mu_A(x) : X \to [0, 1]$ is called the membership function for fuzzy set $A$. The membership function maps each element of $X$ to a membership grade (membership value) between 0 and 1.

**Definition 2.2.** Let $A$ and $B$ be two fuzzy sets of the universe of discourse $X$ with membership functions $\mu_A$ and $\mu_B$, respectively. The union of the fuzzy sets $A$ and $B$, denoted as $A \cup B$, is defined by

$$\mu_{A\cup B}(x) = \max(\mu_A(x), \mu_B(x)).$$

The intersection of fuzzy sets $A$ and $B$, denoted as $A \cap B$, is defined by

$$\mu_{A\cap B}(x) = \min(\mu_A(x), \mu_B(x)).$$

**Definition 2.3.** Let $A$ and $B$ be two fuzzy sets defined on the universe of discourse $X$ with membership functions $\mu_A$ and $\mu_B$, respectively. The fuzzy subsethood $S(A, B)$ [34] measures the degree in which $A$ is a subset of $B$

$$S(A, B) = \frac{M(A \cap B)}{M(A)} = \frac{\sum_{x \in X} \min(\mu_A(x), \mu_B(x))}{\sum_{x \in X} \mu_A(x)},$$

where $S(A, B) \in [0, 1]$.

**Fuzzy-rule similarity function:** The generated value of this function is between 0 and 1 that is determined by calculating the minimum similarity of all of the fuzzy matches that were made on the right hand side (RHS) of a rule. A value ‘0’ is always returned if not called from the RHS of a rule. If there were no fuzzy matches on the left hand side (LHS) of a rule, then a value ‘1’ is returned. The similarity function provides the maximum value of the intersection of two fuzzy values, and a higher value indicates greater similarity.

2.4. **Feature selection method (FSM).** In order to completely remove redundant information and enable quicker and more accurate training, a number of preprocess were done against the dataset. KDD and the process of mining knowledge rules from data are important for some fields. One of the critical aspects of any knowledge discovery process is feature selection [35]. Feature selection is one of special optimization techniques, which will help to evaluate usefulness of attributes, to select relevant attributes and then to reduce the dimensionality [36] of datasets by deleting unsuitable attributes that may degrade the performance of classification, and hence improves the performance of data
A FUZZY-BASED ROUGH SETS CLASSIFIER FOR FORECASTING QUARTERLY PGR

mining algorithms [37]. There are three important reasons for feature selection: (1) Simpler model, (2) easier to interpret, and (3) faster model induction and structural knowledge. Working with fewer attributes can make the problem of finding classification rules or prediction model much more tractable [11]. Feature selection techniques can be roughly categorized into three broad types, the filter model, the wrapper model, and the hybrid model [38]. The filter model relies on the general characteristics of the training data to select some features independently in any learning algorithm, and therefore it does not inherit any bias of a learning algorithm [38]. The study applied it as feature selection.

From Witten and Frank [11], the five FSMs of the filter model, InfoGain, Chi-square, Gain Ratio, Consistency, and Cfs are well known in academic work and are adopted in common usage. Thus, the study introduces the five-subset evaluator methods, as follows:

- **Chi-squared:** Chi-squared method evaluates the worth of an attribute by computing the value of the chi-squared statistic with respect to the class.
- **InfoGain:** InfoGain is one of the simplest attribute ranking methods and is often used in text categorization applications. The method evaluates the worth of an attribute by measuring the information gain with respect to the class.
- **Gain Ratio:** The method evaluates the worth of an attribute by measuring the gain ratio with respect to the class.
- **Consistency:** This method values an attribute subset by the level of consistency in the class values when the training instances are projected to the attribute subset.
- **Cfs (Correlation based feature selection):** The method evaluates subsets of attributes that are highly correlated with the class having the low intercorrelation is preferred.

2.5. Discretization method (DM). Although there is many unsupervised automatic discretization methods studied in machine learning community, the study mainly introduces minimize entropy principle approach and decision trees C4.5 as representative in the following two subsections.

2.5.1. **Minimize entropy principle approach (MEPA).** A key goal of entropy minimization analysis is to determine the quantity of information in a given data set. The entropy of a probability distribution is a measure of the uncertainty of the distribution [39]. To subdivide the data into membership functions, the threshold among classes of data must be established. A threshold line can be determined with an entropy minimization screening method. The segmentation process starts with two classes. Thus, repeating partition with threshold value calculations allows further partition the data set into a number of fuzzy sets.

Assume that a threshold value is being sought for a sample ranging between \(x_1\) and \(x_2\). An entropy equation is written as the regions \([x_1, x]\) and \([x, x_2]\); the first region is denoted \(p\), and the other is denoted \(q\). Entropy with each value of \(x\) is expressed as [40]:

\[ S(x) = p(x)S_p(x) + q(x)S_q(x), \]

\[ S_p(x) = -[p_1(x) \ln p_1(x) + p_2(x) \ln p_2(x)], \]

\[ S_q(x) = -[q_1(x) \ln q_1(x) + q_2(x) \ln q_2(x)], \]

\[ p(x) + q(x) = 1. \]

where \(p_k(x)\) and \(q_k(x)\) = conditional probabilities that the class \(k\) sample is in the region \([x_1, x_1 + x]\) and \([x_1 + x, x_2]\), respectively, and \(p(x)\) and \(q(x)\) = probabilities that all samples are in the region \([x_1, x_1 + x]\) and \([x_1 + x, x_2]\), respectively.
A value of $x$ that gives the minimum entropy is the optimum threshold value. The entropy estimates, $p_k(x)$ and $q_k(x)$, and $p(x)$ and $q(x)$, are listed, as follows:

$$p_k(x) = \frac{n_k(x) + 1}{n(x) + 1},$$  \hspace{1cm} (8)

$$q_k(x) = \frac{N_k(x) + 1}{N(x) + 1},$$  \hspace{1cm} (9)

$$p(x) = \frac{n(x)}{n},$$  \hspace{1cm} (10)

$$q(x) = 1 - p(x).$$  \hspace{1cm} (11)

where

- $n_k(x)$ = number of class $k$ samples located in $[x_1, x_1 + x]$,
- $n(x)$ = the total number of samples located in $[x_1, x_1 + x]$,
- $N_k(x)$ = number of class $k$ samples located in $[x_1 + x, x_2]$,
- $N(x)$ = the total number of samples located in $[x_1 + x, x_2]$,
- and $n$ = total number of samples in $[x_1, x_2]$.

Figure 1 illustrates the partitioning process for a sample ranging between $x_1$ and $x_2$. While moving $x$ in this region $[x_1, x_2]$, the values of entropy for each position of $x$ are calculated. The value of $x$ in the region holding the minimum entropy is called the primary (PRI) threshold value. By repeating the process, secondary threshold values denoted as SEC1 and SEC2 can be determined. Developing seven partitions requires tertiary threshold values denoted as TER1, TER2, TER3, and TER4.
2.5.2. Decision trees (DT) C4.5 algorithm. Classification is an important data mining function. Many classification models have been proposed, for instance, neural networks base, decision trees base, statistical base, and distance base. DT is a flow-chart-like tree structure where each internal node denotes a test on an attribute; each branch represents an outcome of the test, and leaf nodes represent class or class distributions. The ID3 [41] is a DT algorithm based on information theory. The basic strategy used by ID3 is to choose splitting attributes with the highest information gain. The concept used to quantify information is called entropy. Entropy is utilized to measure the information in an attribute.

Assume that a collection set $S$ includes $c$ outcomes, and then the entropy is defined as:

$$H(S) = \sum (-p_i \log_2 p_i),$$  \hspace{1cm} (12)

where $p_i$ is the proportion of $S$ belonging to class $i$.

$Gain(S, A)$ is information gain of example set $S$ on attribute $A$ and is defined as:

$$Gain(S, A) = H(S) - \sum \frac{|S_v|}{|S|} H(S_v),$$  \hspace{1cm} (13)

where $v$ is a value of $A$, $S_v = a$ subset of $S$, $|S_v| = $ number of elements in $S_v$, and $|S| = $ number of elements in $S$.

C4.5 is a software extension of the ID3 algorithm designed by Quinlan [42] to address the issues not dealt by ID3: Avoiding overfitting the data, determining how to grow a decision trees, reducing error pruning, ruling post-pruning, handling continuous attributes, choosing an appropriate attribute selection measure, handling training data with missing attribute values, handling attributes with differing costs, and improving computational efficiency.

3. The Proposed Hybrid Models. For solving the practical problems faced by investors, the study proposes a set of hybrid models for forecasting quarterly PGR and determining the quality of a rough set classification system. The proposed hybrid models are originally based on experiential knowledge (EK) of the authors in the financial industry and stock market, followed by various combinations of five additional components (treatments): FSM, DM (including MEPA and DT-C4.5), RF, FST (similarity function), and RST. They include the three differently integrated models: (1) EK+FST+MEPA+RST+RF (model A), (2) EK+FSM+MEPA+RST+RF (model B), and (3) EK+DT-C4.5+RST+RF (model C). For a convenient view of the above, the three hybrid models are integrated in Table 1. Table 1 shows the major differences among the three hybrid models. For example, the model A is organized by five treatments, including experiential knowledge, discretization method, a similarity of FST, a rough sets classifier, and rule filter method.

### Table 1. An overview for different types of hybrid models based on rough sets

<table>
<thead>
<tr>
<th>Processes</th>
<th>Models</th>
<th>Model A</th>
<th>Model B</th>
<th>Model C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiential knowledge</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>Feature selection</td>
<td></td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discretization (MEPA or DT)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>FST (Similarity)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>Rough sets</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>Rule filter</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
</tr>
</tbody>
</table>
Afterwards, the concept of the proposed hybrid models is firstly stated in Subsection 3.1, and Subsection 3.2 is the introduction of the three proposed hybrid models, respectively.

3.1. Concept and research framework of the proposed hybrid models. In the real world, the stock market is one of the most greedy and covetous battlefields of the money war. More seriously, the winner takes all, whereas the loser gains nothing. Market climates and gains/losses can be changed within seconds; thus, the accuracy information for forecasting quarterly PGR of a specific firm is of extreme importance. Besides, due to too many noises that are caused by changes in market conditions and environments, in this market, the performances of forecasting models are highly dependent on the different context and/or the used data. Therefore, to find a suitable way that is applied into the context of the stock markets is very crucial for investors.

In view of the earlier literature [10], the study notices that a significant trend is to build soft computing architectures for solving real-world problems; particularly in hybrid intelligent systems, they are always composed of several models to solve the problems in practice. Moreover, it is proven that such ensemble classifiers outperform the standalone models [8]. In this respect, in order to overcome the current existing drawbacks of traditional forecasting methods in the stock market for investors, the study offers a hybrid model based on ensemble rough sets classifiers from an intelligent perspective and provides an alternate method for forecasting the quarterly PGR of the stock market in Taiwan. Afterwards, the proposed hybrid models briefly involve three different types of hybrid models, constituted differently by six components: EK, FSM, DM, FST, RF, and RST.

Rough sets classifiers usually apply the concept of rough sets to reduce the number of attributes in a decision table; furthermore, extracting decision rules based on the rough sets LEM2 algorithm is superior to some traditional methods, because they deduce rule sets directly from data with symbolic and numerical attributes [17]. Unfortunately, one of the drawbacks of traditional rough sets is that data must be discretized first for improving classification accuracy. Besides, another major limitation is the too large a number of decision rules generated to reduce the complexity of knowledge-based systems [29]. The large number of rules limits the classification capabilities of the rule-set, as some rules are redundant or of ‘poor quality.’ Consequently, some filtering-rule methods are needed to reduce the number of rules. Moreover, the DM of attributes is a valuable aspect of data mining, particularly for rough set classifiers and classification problems. In addition, the FSM is used to select the most influential attributes from the original attribute set and then constructs a classifier with superior performance. Therefore, the proposed hybrid models are dedicated to improving the accuracy of a rough sets classification system.

3.2. Three different types of hybrid models based on rough sets. Overall, the proposed hybrid models include three different types of hybrid models A-C, and computation processes of each are presented step by step in the next subsection, respectively.

3.2.1. The algorithms of proposed hybrid models. First, to introduce the proposed hybrid model A, it consists of five treatments assigned into two parts. The first part of such a soft computing system focuses on the professional knowledge of relating the stock market in the financial industry to select attributes in advance; the second uses the AI techniques, such as rough sets and fuzzy logic, for classifying quarterly PGR and generating decision rules. Figure 2 illustrates the flowchart of model A, including eight steps.

Accordingly, in supporting more clear definitions, the steps of the computing process for the proposed hybrid model A are detailed systematically, as follows:

The rough steps of the algorithm for the proposed hybrid model A
Step 1: Collect the attributes of a practical dataset by EK. At first, based on professional knowledge, select the dataset as experimental data for verification with setting some pre-conditions. For example, collect data in a period or in an industry. After that, collect the practical data from the selected dataset.

Step 2: Preprocess the dataset. Real-world data are generally incomplete (lacking attribute values, lacking certain attributes of interest, or containing only aggregate data), noisy (containing errors or outliers), or inconsistent (containing discrepancies in codes or names). To minimize redundant information and enable quicker and more accurate training, several preprocessing operations were made against the dataset. Firstly, change the collected data into an EXCEL file for convenience operations. Accordingly, records with missing values are deleted, redundant attributes are eliminated, and data are transformed into an easily processed format used in classifying a specific decision attribute.

Step 3: Build the membership function by MEPA. Before building the membership function, it is needed to partition the continuous attributes firstly. According to the study of Chen et al. [43], MEPA discretization performs well in rule-based classification systems. Thus, this study applies MEPA to partition continuous attributes. First, iteration number r is determined to produce the linguistic number m, where

\[ m = \begin{cases} 2^r - 1 & r \geq 2 \\ 2 & o.w. \end{cases} \]

The study adopts r = 2. Therefore, the linguistic number is 3. Next, the entropy value of each continuous attribute is computed using the entropy equation proposed by Christensen [40]. Repeating this procedure to subdivide the data obtains the thresholds. If \( n(x) \) and \( N(x) \) equal zero, then data within the range are not further subdivided. Using the computed thresholds as the midpoint of triangular fuzzy numbers, the first and last membership functions are used to build trapezoid membership functions. Whether the
attribute value is more or less than SEC1, the membership degree always equals 1. The membership function of the minimal entropy approach can be established accordingly.

**Step 4:** Fuzzify the data of continuous attributes into unique corresponding linguistic values. Using the membership function mentioned above, the membership degree of each data is calculated, and the maximal membership degree of attributes is used to determine its linguistic value. A linguistic value dataset is then mapped by MEPA to convert discretized continuous data into a unique corresponding linguistic value, such as L_1 (small), L_2 (medium), and L_3 (large).

**Step 5:** Select the similarity threshold by FST. To find the closeness between the decision attribute and each subgroup of condition attributes, calculate the subsethood of them. According to the Equation (4), the subsethood is defined as follows:

$$S(C_s, L_i) = \frac{M(C_s \cap L_i)}{M(C_s)}.$$  \hspace{1cm} (14)

where $S(C_s, L_i)$ denotes the subsethood and $S(C_s, L_i) \in [0, 1]$; $C_s$ is the class label; $L_i$ is the linguistic value of attribute; $M(C_s)$ is the sum of membership degree in the same class label; and $M(C_s \cap L_i)$ denotes the sum of min (membership degree of decision attribute, membership degree of condition attribute) for each record.

Furthermore, the similarity threshold is defined as follows:

$$S_{T_i} = \text{Max}\{S(C_s, L_i)\}.$$  \hspace{1cm} (15)

where $S_{T_i}$ denotes the similarity threshold; it is selected by the maximal value from $S(C_s, L_i)$.

**Step 6:** Transform the fuzzied attributes. After that, according to the membership degrees derived from Step 4 and the similarity thresholds derived from Step 5 for corresponding linguistic value of continuous attributes, transform the fuzzied dataset further. The transformation method is defined as follows:

$$I'_i = \begin{cases} 
1, & I_i \geq S_{T_i} \\
0, & \text{otherwise} 
\end{cases}$$  \hspace{1cm} (16)

where $I_i$ is the membership degree of each linguistic value in condition attributes, and $I'_i$ is a new value after the comparison processes for the same linguistic value of the same attribute. If the membership degree $I_i$ is greater than the similarity threshold $S_{T_i}$, then transform the membership degree to ‘1’, otherwise denoted as ‘0’. By the way, all the transformed values of continuous attributes are generated by the Equation (16).

**Step 7:** Extract the decision rules by the RST LEM2 algorithm. A rule extracted by the LEM2 algorithm is superior to those extracted by traditional methods, because they deduce rule sets directly from data with symbolic and numerical attributes. As such, the LEM2 algorithm is used to construct decision rules of a set from the discretized dataset.

**Step 8:** Improve the rule quality by RF. Since more rules complicate prediction, the study employs RF to diminish decision rules. The rules set extracted in Step 7 undergoes a filtering process in which rules below the support threshold are eliminated to further improve rule quality.

As for the proposed hybrid models B-C, the flowchart of all of them is partially similar to that of the model A; therefore, for the shortened study in an easily understood manner, only the steps of different processes among them are illustrated in the following roughly.

**The rough steps of the algorithm for the proposed hybrid model B**

**Steps 1-2:** They are the same as the Steps 1-2 of model A.

**Step 3:** Select the core attributes by five FSMs. A method to simplify knowledge discovery is needed. Feature selection, a special optimization technique that removes noisy features and reduces the dimensionality of datasets by deleting unsuitable attributes,
improves the performance of data mining algorithms [37]. The irrelevant attributes may degrade the performance of data mining algorithms; therefore, this Step 3 identifies more relevant and discriminating attributes by using five FSMs: Cfs, Chi-square, Consistency, Gain Ratio, and InfoGain. The five methods above are well known in academic work and widely used as attribute evaluators. Thus, this study applies the five FSMs for diminishing the selected attributes and selecting the core attributes.

Steps 4-5: They are the same as the Steps 3-4 of model A. Before building the membership function, partitioning of the continuous attributes by MEPA is needed firstly.

Steps 6-7: They are also the same as the Steps 7-8 of model A.

The rough steps of the algorithm for proposed hybrid model C

Steps 1-2: They are the same as the Steps 1-2 of model A.

Step 3: Determining the attribute-granularity by EK and the DT-C4.5 algorithm. In order to improve the accuracy of classification effectively, it is needed that specific attributes should be pre-granulated. Firstly, granulate the classes of the decision attribute based on EK. Next, generate decision trees by the DT-C4.5 algorithm with selected condition attributes in advance by EK. Finally, the condition attributes are granulated according to the cutoff points of the generated decision trees by the C4.5 algorithm.

Step 4: Build the decision table. From the condition attributes selected in Step 3 above and a class (decision attribute), which are discretized beforehand, build a decision table.

Steps 5-6: They are similar to the Steps 7-8 of model A.

4. Conclusions. Three hybrid models (A, B, and C) based on rough sets has proposed for solving the classification problems of quarterly PGRs faced by investors in the financial industry of the stock market in this study (Part I). The three proposed hybrid models are constituted differently by six basic components: experiential knowledge, feature selection methods, discretization methods, rule filter, fuzzy-rule similarity, and rough sets. The effect of components for proposed hybrid models is focused particularly, including FSM, DM, FST (Similarity functions), and RF. Hence, the performance of them will be then evaluated and described conclusively in the next paper (Part II) accordingly. Mainly, they are compared with other AI techniques. Some evaluation criteria, such as the accuracy with standard deviation, the number of attributes, and the number of rules with its standard deviation, will be adopted as the standard of comparisons. Additionally, the comparison is spilt into two-fold: internal comparison and external comparison. Finally, in order to further compare the performance of Traditional Rough Sets with the models A-C, an improved accuracy rate and an improved rule rate will be calculated accordingly. Furthermore, the determinants (core attributes) of influencing the quarterly PGR are examined, and comprehensive decision rules applied to knowledge-based systems by the rough sets LEM2 algorithm and provide a powerful explanation for investors are extracted.

Acknowledgment. The authors would like to thank the Editor-in-Chief, associate editor, and anonymous reviewers for their very useful comments and suggestions, which are very helpful in improving this study.

REFERENCES


