2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

A Minimax Model of Portfolio Optimization Using Data Mining to Predict Interval Return Rate

Meng YUAN [1], and Junzo WATADA [2], *IEEE*

[1,2] Graduate School of Information, Production & Systems, Waseda University,
2-7 Hibikino, Wakamatsu, Kitakyushu 808-0135 JAPAN
E-mails: [1] irisy28@gmail.com , [2] junzow@osb.att.ne.jp

***Abstract*—In 1950s, Markowitzs first proposed portfolio theory based on a mean-variance (MV) model to balance the risk and profit of decentralized investment. The two main inputs of MV are expected return rate and the variance of expected return rate. The expected return rate is an esti- mated value which is often decided by experts. Various uncertainty of stock price brings difficulties to predict return rate even for experts. MV model has its tendency to maximize the influence of errors in the input assumptions. Some scholars used fuzzy intervals to describe the return rate. However, there were still some variables decided by experts. This paper proposes a classification method to find the latent relationship between the interval return rate and the trading data of a stock and predict the interval of return rate without consulting any expert. Then this paper constructs the portfolio model based on minimax rule with interval numbers. The evaluation results show that the proposed method is reliable.**

Keywords- Portfolio, Minimax, Interval number, Classification

I. INTRODUCTION

In finance, portfolio is a collection of investment pursued by an institution or a private individual. Profit cannot be separated from risk in financial investment. This implies that investor should balance investment risk and profit. The portfolio selection problem is to decide which assets and in which proportion will better respect the investor's declared preferences. It is difficult to decide which securities should be selected because of the existence of uncertainty of their returns, so a balance between maximizing expected return and minimizing the risk is needed. Investors and scholars concerned about how to measure the expected return and risk and how to balance the return and risk. In 1950s, Markowitzs portfolio theory first gave answers to the questions.

According to Markowitz [1,2,3], for a given specific return rate, one can derive the minimum investment risk by minimizing the variance of a portfolio; or for a given risk level which the investor can tolerate, one can derive the maximum returns by maximizing the expected returns of a portfolio. The main input data of the Markowitz mean variance model are expected returns and variance of expected returns of these securities. Simplifying the number and types of the input data has been one of the main research topics in this field for the last four decades. Although some breakthroughs, such as the Index Model, have been implemented to all of these methods have some drawbacks. As Levy and Markowitz [4] have noted, the exact choice of the efficient portfolio which maximizes $E_u(R)$ is only possible if the returns from all securities are normally distributed, or if the utility function $u(R)$ is quadratic. But the normal distribution of returns is only a hypothesis, which has not been empirically corroborated, and quadratic utility functions present many logical flaws.

The meaning of various kinds of uncertainties (ambiguity and vagueness) is clarified in Inuiguchi and Ramik [6], which highlights also the advantages and disadvantages of the use of a fuzzy approach with respect to a stochastic programming in a portfolio selection context. One of the advantages of fuzzy and possibilistic programming approaches is that they are generally more tractable than those based on stochastic programming and allow the inclusion of the knowledge of the experts in the model. Indeed, possibility portfolio models are based on a possibility distribution that is constructed by using experts judgments. With regard to this, see Tanaka et al. [7]. In the decision-making literature , based on interval analysis, another approach allows us to handle imprecise input data. This approach consists in assuming that the data of a decision-making problem are not well defined but may vary in given intervals.

During the past decade, the computational intelligence researchers have proposed a sufficient number of data mining algorithms to solve real world classifica-

Corresponding author (Meng YUAN).

tion and clustering problems. Since many of the current modeling techniques are based on linear assumptions, data mining a new kind of financial analysis under consideration of the nonlinear analysis of integrated financial markets. Even though there exists a number of non-linear regression techniques, most of these techniques require that the non-linear model must be specified before the estimation of parameters can be determined. One non-linear modeling to solve these problems involves the use of classification[8]. In fact, classification offers a novel technique that does not require a pre-specification during the modeling process because they independently learn the relationship inherent in the variables. This is especially useful in security investment and other financial areas where much is assumed, and little is known about the nature of the processes to determine asset prices (Burrell and Folarin, 1997)[10].

Generally, classification is a data mining function that describes and distinguishes data classes or concepts. The goal of classification is to accurately predict class labels of instances whose attribute values are known, but class labels are unknown[8]. Some researchers made a progress of using data mining technique in portfolio sector.

This paper proposes a minimax portfolio selection model with interval expected return rate. The interval number was predicted by the classification. Then the interval return rate is used in minimax portfolio selection model. The purposes of this paper are to describe the return and risk of a portfolio more accurately and have better strategy than Markowitzs model.

This paper is organized as follows: In section two, a survey and research background of portfolio with interval number and minimax model is presented. In section three, predicting the interval of return rate is stated by classification. Also the minimax portfolio model is discussed by using the predicted interval in . Section four provides a result of experiment with six Japanese stocks. In section five, concluding comments are presented.

II. BACKGROUND

2.1 Portfolio approaches

Let us consider n securities with return rate $c_i(i = 1, , n)$ and denote by x_i (with $x_i \geqslant 0$) the proportion of total amount of funds invested in the i-th security. A portfolio selection problem is to find the investment rate vector $\mathbf{x} = (x_1, x_2, , x_n)$ which maximizes the total portfolio return.

Since investors usually make their decisions under an uncertain environment, the traditional approaches to portfolio selection treat the return rates as a random variables vector following a probability distribution with mean Vector(**m**) and a covariance Matrix(**V**). In particular, the well-known model proposes by Markowitz [1], results in minimizing the portfolio variance, while reaching a given portfolio mean return rate.

In the decision-making literature, interval analysis based approach allows handling imprecise input data. This approach consists in assuming that the data of a decision-making problem are not well defined but may vary in given intervals.

On this subject, in Alefeld and Mayer (2000)[10] both theory and some applications of interval analysis are presented. Moreover, many papers deal with interval linear programming problems (see, for example, Chinneck and Ramadan, 2000[11] and Inuiguchi and Sakawa, 1995[12]).

In particular, Inuiguchi and Sakawa (1995)[12] analyze a linear programming problem with interval objective function coefficients and give a new solution concept, based on the minimax regret criterion.

It is interesting that the minimax regret criterion has been also adopted to formulate possibilistic portfolio selection problems (see Inuiguchi and Ramik, 2000[6] and Inuiguchi and Tanino, 2000). The idea is that an investor, who is supposed to know the value of the return rates after making its investment decision, wants to minimize the worst (maximum) regret. The worst regret represents the maximum deviation between the return that the investor could receive if he/she invests in the optimal portfolio and the portfolio return that he/she actually realizes.

Recently many applications of interval programming to portfolio selection can be found in Wang and Zhu (2002)[20]; among them Lai et al. (2002)[13] extend the Markowitzs model to an interval-programming model by quantifying the expected return and the covariance as intervals. Moreover, Ida (2003)[14] solved a multi objective portfolio selection problem with interval coefficients, in a Markowitz framework.

2.2 Minimax portfolio approaches

Young (1998)[15] formulated an LP for maximizing the minimum return to select a diversified portfolio based on historical returns data. He referred to the LP as a minimax model because of its greater familiarity and this convention will be followed. The performance of the model was compared to other similar linear and nonlinear models and statistical analyses and simulation were employed to find that the minimax approach outperformed the meanvariance approach with respect to mean square estimation error under the widely used log-normal distribution. He showed the minimax modeling approach to be compatible with expected utility maximization and explored the incorporation of fixed transaction charges. Though the minimax portfolio model seems similar to value at risk(VaR) portfolio model, they are quite different to each other. Value-at- Risk (VaR) is a general measure of risk developed to equate risk across products and to aggregate risk on a portfolio basis. VaR is defined as the predicted worst-case loss with a specific confidence level (for example, 95%) over a period of time (for example, 1 day). The goal of VaR portfolio model is to minimize the maximum loss[23]. And the loss is calculated by loss function[23]. However, Young's minimax model treats minimum return of historical data as the risk, the objective is to maxmize the minimum return.

After Young, many other researchers proposed other minimax model. Some dealt with immunization problems [16], some minimized the maximum loss over all past observation periods for a given level of return, some are based on a game theoretical approach, some used scenario analysis method [17], and some minimized the maximum risk of the individual assets [18].

Xia[19] proposed a new model for portfolio selection in which the expected returns of securities are considered as variables rather than as the arithmetic means of securities. However, the weight and the change tendency are both decided by experiences and expertise, while it is not guaranteed on the basis of the experts decisions are always right.

III. Methodology

3.1. Markowitz MV Portfolio Optimization Model

Markowitz MV portfolio optimization theory assumes that investors are risk averse, meaning that given two portfolios that offer the same expected return, investors will prefer the less risky one. Thus, an investor will take on increased risk only if compensated by higher expected returns. Conversely, an investor who wants higher expected returns must accept more risk. The exact trade-off will be the same for all investors, but different investors will evaluate the trade-off differently based on individual risk aversion characteristics. The implication is that a rational investor will not invest in a portfolio if a second portfolio exists with a more favorable risk-expected return profile. Markowitz(MV) portfolio optimization models is written as follows:

$$max f(x) = \sum_{i=1}^{n} E(r_i)x_i,$$

$$s.t. \qquad \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij}x_i x_j \leqslant \omega,$$

$$\sum_{i=1}^{n} x_i = 1,$$

$$0 \leqslant x_i \leqslant \mu_i, i = 1, 2, ..., n, j = 1, 2, ..., n,$$

where the objective function $f(x)$ is to maximize the mean of return of the portfolio for a given upper bound ω for the variance of the portfolio return. r_i is the return on asset i, and x_i is the proportion of asset i in the portfolio.

$$min f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij}x_i x_j,$$

$$s.t. \qquad \sum_{i=1}^{n} E(r_i)x_i \geqslant r_0,$$

$$\sum_{i=1}^{n} x_i = 1,$$

$$0 \leqslant x_i \leqslant \mu_i, i = 1, 2, ..., n, j = 1, 2, ..., n,$$

where the objective function $f(x)$ is to minimize the variance of the portfolio return for a given lower bound r_0 for the mean of return. r_i is the return on asset i, and x_i is the proportion of asset i in the portfolio.

3.2 Interval number in portfolio

Xia[19] proposed a new model for portfolio selection in which the expected returns of securities are considered as variables rather than as the arithmetic means of securities. Xia used the past returns of a security to estimate an approximate value, which is the interval of expected return rate.

To give an order of the expected returns of securities in a portfolio and to determine the change ranges of the expected returns of securities, Xia considered the following three factors:

1. Arithmetic mean: Although the arithmetic means of securities should not be put as the expected returns directly, they are a good approximation.

2. Historical return tendency: If the recent historical returns of a security remain increasing, the expected returns of the security is greater than the arithmetic mean based on the historical data. However, if the recent historical returns of the security decrease along time, the expected returns of the security is smaller than the arithmetic mean based on the historical data.

3. Forecast of the future returns of a security: The third factor influencing the expected return of a security is its estimated future returns. Based on the financial report of a corporation, if we believe that the returns of this corporation's stock will increase, then the expected returns of this security should be larger than the arithmetic mean based on the historical data. On the contrary, if we think that the future returns of this corporation's stock will decrease, the expected returns

of this security will be smaller than the arithmetic mean.

In Xia's paper the expected return rate is decided by:

$$R_i = h_1 * R_{ai} + h_2 * R_{to} + j_3 * R_{fb}, i = 1, 2, \cdots, n$$

where R_{ai} is the arithmetic mean of security i, R_{t0} expresses the change tendency of the returns of security i, and R_{fi} is an approximation of the future expected returns of security i. R_{ai} can be calculated with the historical data. Denote the arithmetic mean of the recent historical data as R_{ti}. If there is no obvious change tendency, R_{ti} can be equal to zero. The computation of derivation of R_{fi} requires some forecasts based on the financial report and individual experience. Also different professionals with different experience should give the weights of these three factors.

However, it is not guaranteed that experts decisions are correct. To avoid too much relying on experience, this paper used classification technique in data mining. According to historical data, we can know the return rate in the past. Mark each return rate using an interval as a label. The interval of return rate is the classification label, based on this train the past data to generate a classification model. When new data comes, use the classification model to predict the label, which is the interval of expected return rate.

3.3 Use classification to predict the interval of expected return rate

Classification is a two-step process. First, it builds classification model using training data. Every object of the dataset must be pre-classified i.e. its class label must be known. Second step is model usage. The model generated in the preceding step is tested by assigning class labels to data objects in a test data set. Each tuple/sample is assumed to belong to a predefined class, as determined by the class label attribute. The model is represented as classification rules, decision trees, or mathematical formulae. Accuracy rate is the percentage of test set samples that are correctly classified by the model. If the accuracy is acceptable, use the model to classify data tuples whose class labels are not known. In this paper, we proposed a method using the trading data and stock price to predict the stock price changing.

Figure 1 shows the flow of classification based portfolio selection model to find the hidden rule of return rates. All factors that affect stock priceexterior factor or interior factor will eventually reflect in the change of stock price and the trading data. It is possible for investors not to consider why the stock price changes. It will be much more convenient if investors just analyze the trading data and make a prediction of

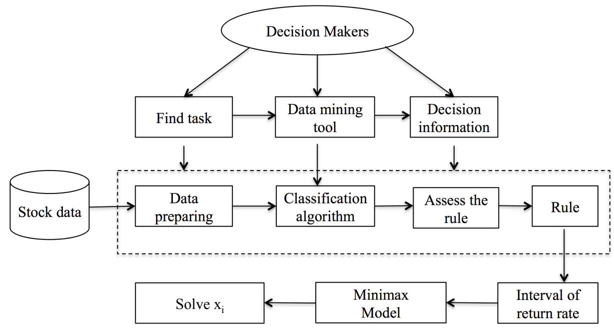


Fig. 1. Flow of using classification to acquire knowledge

the future stock performance. This paper uses an open source software Weka as the data mining tool. There are major four steps in Weka. First, prepare the data and load the data. The data can be in .arff format or .csv format. After loading the data choose classify then choose classification algorithm and generate the rules. Then analyze the result and evaluate the model.

This paper uses stock data from Yahoo Finance. Then compute the return rate and mark it with the corresponding interval as Table 1 shows. For example, the return rate of a stock is -0,0077519, then it is a stock has the return rate in interval of (-0.01,0) on January 4th.The trading data includes open price, highest price, lowest price, close price, trading volume and trading amount. When all intervals of stcoks are computed, the data are prepared. We expect to find some relationship between trading data and the interval. So we use Weka as the data mining tool to train the classification model with training data. The training data removes return rate to avoid over-fitting as Table 2 shows. Next predict the interval of future data using trained classification model. Finally solve x_i based on minimax rule. Classification and prediction are the process of identifying a set of common features and models that describe and distinguish data classes or concepts. The models are used to predict the class of objects whose class label is unknown.

This paper used multiclass classifier in Weka to generate the classification model. Classification consists of predicting a certain outcome based on a given input. In order to predict the outcome, the algorithm processes a training set containing a set of attributes and the respective outcome, usually called goal or prediction attribute. The algorithm tries to discover relationships between the attributes that would make it possible to predict the outcome. Next the algorithm is given a dataset not seen before, called prediction set, which contains the same set of attributes, except for the prediction attribute not yet known. The algorithm analyses the input and produces a prediction. The

TABLE I
TRAINING SET

Open	High	Low	Close	Vol	Amt	Interval
258	260	255	256	1427600	366540300	(0.01,0)
256	251	252	252	334900	834258800	(-0.03, -0.01)
253	252	256	256	2435800	618750300	(0,0.029

TABLE II
PREDICTION SET

Open	High	Low	Close	Vol	Amt	Interval
261	264	261	263	2214200	581034900	?
265	265	262	263	1257300	331315000	?
263	264	262	264	857800	225632700	?

TABLE III
ORIGINAL STOCK DATA

Date	Open	High	Low	Close	Vol	Amt	Return rate	Interval
1-4	258	260	255	256	1427600	366540300	-0.007751938	(0.01,0)
1-5	256	251	252	252	334900	834258800	-0.011764706	(-0.03, -0.01)
1-6	253	252	256	256	2435800	618750300	0.011857708	(0,0.02)
...	...	...	...	...	...	...	...	...
12-30	266	266	263	264	650900	172074800	-0.007518797	(-0.01,0)

prediction accuracy defines how good the algorithm is.

3.4 A minimax model of portfolio selection

Because expected return rate varies within an interval, it is possible that it reaches the smallest value in the interval. Even in that case, it is reasonable that investors still want to gain profit as much as possible. So our goal is to maximize the return even when the return rate is the smallest value in its interval. We consider a capital market with n risky assets and without riskless assets. The notation and assumptions in this section are the same as those in Deng's paper [20]. Therefore, the minimax portfolio model is written as follows:

$$(P_\omega)\begin{cases} max_y min_r f(\vec{y},\vec{r}) \\ s.t. \sum_{i=1}^{n} y_i = 1 \\ a_i \leqslant r_i \leqslant b_i, i = 1, 2, \cdots, n \end{cases}$$

where

$$f(\vec{y},\vec{r}) = (1-\omega)\sum_{i=1}^{n} r_i y_i - \omega \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij} y_i y_j$$

The following notations are employed:

$\tilde{r_i}$	the holding period rate of return on risky asset $i(i = 1, 2, 3, \cdots, n)$, equal to the ratio of the value of the asset at the end of the period over its current value.
r_i	the expected value of $\widetilde{r_i}(i = 1, 2, 3, \cdots, n)$
r_{n+1}	the holding period rate of return on the risk-less asset, which is a constant
σ_{ij}	$conv(\widetilde{r_i}, \widetilde{r_j})$, the covariance between $\widetilde{r_i}$ and $\widetilde{r_j}$, $(i, j = 1, 2, 3, \cdots, n)$
p_i	the price of per share asset $i, (i = 1, 2, 3, \cdots, n)$
x_i^0	the number of shares of asset i owned by the investor before the transaction, $(i = 1, 2, 3, \cdots, n)$, and
x_i	the number of shares of asset i owned by the investor after the transaction, $(i = 1, 2, 3, \cdots, n)$

Here, all σ_{ij} are assumed to be given and fixed while each r_i $(i = 1, 2, 3, \cdots, n)$ is assumed to vary over the range

$$a_i \leqslant r_i \leqslant b_i \tag{1}$$

where a_i and b_i are known from the classification step. For a new investor, it can be taken that

$$x_i^0 = 0, i = 1, 2, 3, \cdots, n \tag{2}$$

As usual, throughout the paper we will assume that the variance-covariance matrix $\mathbf{V} = (\sigma_{ij)_{n\times n}}$ is positive definite.

Let

$$M^0 = \sum_{i=1}^{n} p \tag{3}$$

be the total value of initial portfolio in terms of the price system $(p_1, p_2, \cdots, p_n)^T$, which is here assumed to be positive. Because no money is added into or withdrawn from the portfolio during the transaction, we have

$$\sum_{i=1}^{n} p_i x_i = M^0 \tag{4}$$

Let

$$y_i = \frac{p_i x_i}{M^0}, \quad \text{and} \quad y_i^0 = \frac{p_i x_i^0}{M^0}, \tag{5}$$

be the proportions of the wealth invested in asset asset $i(i = 1, 2, 3, \cdots, n)$ after and before the transaction, respectively. Then, from (3) and (4),

$$\sum_{i=1}^{n} y_i = 1, \quad \sum_{i=1}^{n} y_i^0 = 1,$$

The random return of portfolio $\mathbf{x} = (x_1.x_2, \cdots, x_n)$ is

$$\sum_{i=1}^{n} \widetilde{r}_i p_i x_i = M^0 \sum_{i=1}^{n} \widetilde{r}_i y_i$$

Hence,

$$R(\vec{y}) = \sum_{i=1}^{n} \tilde{r}_i y_i + r_{n+1} y_{n+1}$$

is the rate of return of a portfolio. The expected value of the rate of return is

$$E[R(\vec{y})] = \sum_{i=1}^{n} r_i y_i, \tag{6}$$

and the variance of the rate of return is

$$Var[R(\vec{y})] = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} y_i y_j, \tag{7}$$

For a rational investor, he/she attempts not only to maximize the expected value but also to minimize the variance (as a measure of risk) of rate of return on a portfolio. So, he/she must make a trade-off between the two objectives. Then the investor attempts to maximize

$$f(\vec{y}, \vec{r}) = (1-\omega) \sum_{i=1}^{n} r_i y_i - \omega \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} y_i y_j$$

But on the other hand, because r_i are not known exactly, the investor may choose to find a strategy which is the best for the worst case. That is , it makes sense for him or her to act so as to maximize under the worst possible expected rates of return on portfolios. As a consequence, he/she must solve the following minimax problem P_ω. with parameter ω varying in $(0, 1)$. We denote

$$\mathbb{Y} = y : \sum_{i=1}^{n} y_i = 1$$

$$\mathbb{R} = r_i \geqslant r_{i+1}, a_i \leqslant r_i \leqslant b_i, i = 1, 2, ..., n.$$

So we can rewrite it as the following standard form of the minimax problem:

$$max_{y \in \mathbb{Y}} min_{r \in \mathbb{R}} f(y, r).$$

From the known result of Deng[20], to solve Problem (8), one needs only first to solve the problem

$$max_{y \in \mathbb{Y}} f(y, r).$$

. Clearly, it can be described as

$$\begin{aligned} max_y (1-\omega) \sum_{i=1}^{n} r_i y_i - \omega \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} y_i y_j) \\ s.t. \sum_{i=1}^{n} y_i = 1 \end{aligned} \tag{8}$$

Defining $\vec{y} = (y_1, y_2, ..., y_n)^T, \vec{r} = (r_1, r_2, ..., r_n)^T$, $\vec{e} = (1, 1, ..., 1)^T$, $\mathbb{V} = (\sigma_{ij})_{n \times n}$,then the above constrained optimization problem ; can be formulated as the following unconstrained optimization problem:

$$max_y (1-\omega) r^T y - \omega y^T V y$$

$$s.t. e^T y = 1.$$

According to Deng[20], the solution of P_ω is

$$y(\omega) = \frac{(1-\omega)}{2\omega} \mathbb{V}^{-1} (r^* - r_{n+1} 1) \tag{9}$$

where $\vec{r^*} = (r_1{}^*, r_2{}^*, ..., r_n{}^*)$ is the optimal return for P_ω. Hence, by (5), the optimal portfolio is

$$x_i(\omega) = \frac{y_i^*(\omega) M^0}{p_i} i = 1, 2, ..., n \tag{10}$$

and by (6) and (7), the expected return rate and risk of the optimal portfolio is

$$E(\omega) = r_n + 1 + \frac{(\vec{1} - \omega)}{2\omega} V^{-1} (r^* - r_{n+1} \vec{1}) \tag{11}$$

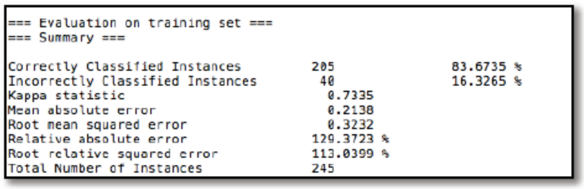
and

$$\sigma^2(\omega) = \frac{(\vec{1} - \omega)^2}{4\omega^2} \mathbb{V}^{-1} (r^* - r_{n+1} \vec{1}) \tag{12}$$

IV. RESULTS

This paper used a year data of five stocks from yahoo finance. We can see from figure 2 that for 83.6735 % of instances have been correctly predicted. The value of the Kappa statistic is 0.7335 which means that statistical significance of the model is rather high. Because Kappa statistic is a chance-corrected measure of agreement between the classifications and the true classes. A value greater than 0 means that your classifier is doing better than chance.Therefore, it can be applied in prediction of interval of return rate.

Table 4 contains detail prediction value and the actual value of the interval. When using test data to judge the accuracy of trained classification model, we can see the actual and predicted value of each instance. If the instance is classified to wrong class, a + mark will be added to error column.



```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances         205               83.6735 %
Incorrectly Classified Instances        40               16.3265 %
Kappa statistic                          0.7335
Mean absolute error                      0.2138
Root mean squared error                  0.3232
Relative absolute error                129.3723 %
Root relative squared error            113.0399 %
Total Number of Instances              245
```

Fig. 2. Classification result

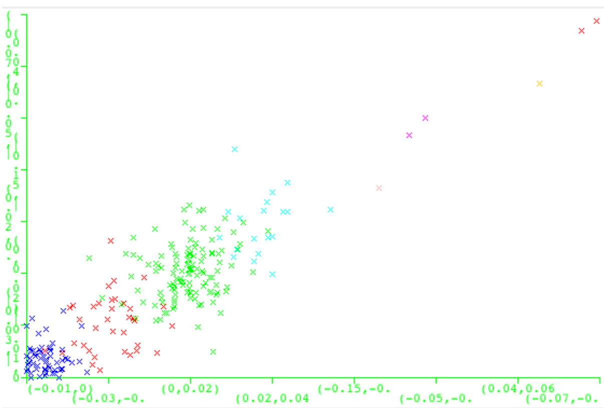


Fig. 3. Visualization of the classes

Table 3 is the evaluation of the classification model. The correctly and incorrectly classified instances show the percentage of test instances that were correctly and incorrectly classified. The raw numbers are shown in the confusion matrix, with abc, d, e, f, g representing the class labels.

TP/True positive: examples in the positive class that got predicted as positive (= true/correct prediction) FP/False positive: examples that got predicted as positive but are negative actually (= false/wrong prediction; also Type I error). Precision is proportion of the examples which truly have class x / Total classified as class x. The first two columns are the TP Rate (True Positive Rate) and the FP Rate (False Positive Rate). TP Rate is the ratio of predicted correctly cases to the total of positive cases (eg: 8 out of 9 is predicted correctly =8/9=0.88). ROC Area means Area Under roc Curve(AUC). It shows if the clasifier is good. When the value is 0.5, the instances are classified randomly. When the value is 1, each instance is classified correctly.

V. Conclusions

A new way of determining the interval of expected return of portfolio is proposed. Interval of return rate was treated as the classification label. The correctness of prediction by classification was 83%. According to Accuracy assessment index like ROC, precision, TP rate, FP rate, Kappa statistic mentioned above, it shows that the classification model is reliable. Compared to expert prediction, the proposed method has the advantage of reducing the rely on expert decision.

Furthermore, this paper uses a minimax portfolio selection model with interval expected return rate to solve proportion of each security in a portfolio. Because the return rate is a variable, it is possible that the return rate is the minimum one. In that case, the minimax model offered an optimal strategy to make the investment reasonable.

VI. Acknowledgement

This work was supported partly by Grants-in-Aid for Science Research MEXT (No23500289) and also in part under a Waseda University Grant for Special Research Projects (Project number 2014A-040).

References

[1] H. Markowitz, Portfolio selection, *Journal of Finance*, Vol.7, No.60, 77-91, 1952

[2] J. C. T. Mao, “Models of capital budgeting: EV vs ES,” *Journal of Financial and Quantitative Analysis*, Vol. 4, No. 5, 657-675, 1970

[3] H. Konno, H. Yamazaki, “Mean-absolute deviation portfolio optimization model and its application to Tokyo Stock Market,” *Management Science*, Vol.37, No.5, 519-531, 1991

[4] H. Konno, K. Suzuki, “A mean-variance-skewness optimization model,” *Journal of the Operation Research Society of Japan*, Vol. 38, No.2, 137-187, 1995

[5] H.Levy and H.M.Markowitz, Approximating Expected Utility by a Function of Mean and Variance. *The American Economic Review*, Vol. 69, No. 3, 308-317, 1979.

[6] M. Inuiguchi, J. Ramik, Possibilistic linear programming: A brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem, *Fuzzy Sets and Systems*, Vol.111,Issue 1, 3-28, 2000.

[7] H. Tanaka, P. Guo, I.B. Trksen, Portfolio selection based on fuzzy probabilities and possibilities distributions, *Fuzzy Sets and Systems*, Vol.111, Issue 3, 387397, 2000

[8] David Enke, Suraphan Thawornwong, The use of data mining and neural networks for forecasting stock market returns, *Expert Systems with Applications*, Vol.29, Issue 4, 927-940, 2005

[9] Burrell, P. R., Folarin, B. O.,The impact of neural networks in finance, *Neural Computing and Applications* Vol.6, Issue 4, 193-200,1997

[10] Alefeld, G., Mayer, G.,Composite measures for the evaluation of investment performance. *Interval analysis: Thoery and applications, Journal of Computational and Applied Mathematics*, Vol.121,Issue 02, 421-464, 2000

[11] J.W. Chinneck, K. Ramadan, Linear programming with interval coefficients, *Journal of the Operational Research Society*, Vol.51,No.2, 209-220,2000

TABLE IV
EVALUATION OF CLASSIFICATION METHOD

TP Rate	FP Rate	Precision	Recall	ROC Area	Class
0.644	0.048	0.089	0.644	0.835	(0.01,0)
0.778	0.005	0.966	0.778	0.964	(-0.03,-0.01)
0.992	0.25	0.805	0.992	0.867	(0,0.02)
1	0	1	1	1	(-0.15,-0.13)
1	0	1	1	1	(-0.05,-0.03)
1	0	1	1	1	(0.04,0.06)
1	0	1	1	1	(-0.07,-0.05)

TABLE V
PREDICTION IN WEKA

Instance	Actual	Predicted	Error
1	1:(-0.01,0)	3:(0,0.02)	+
2	2:(-0.03,-0.01)	1:(-0.01,0)	+
3	3:(0,0.02)	3:(0,0.02)	
...	...	...	...
245	1:(-0.01,0)	1:(-0.01,0)	

[12] M. Inuiguchi, M. Sakawa, Minimax regret solution to linear programming problems with an interval objective function ,*European Journal of Operational Research*, VOl.86 ,Issue 3, 526-536, 1995

[13] K.K. Lai, S.Y. Wang, J.P. Xu, S.S. Zhu, Y. Fang, A class of linear interval programming problems and its application to portfolio selection *IEEE Transactions on Fuzzy Systems*, Vol.10 , Issue 6, 698-704,2002

[14] M. Ida, Portfolio selection problem with interval coefficients *Applied Mathematics Letters*, Vol.16, Issue 5, 709-713,2003

[15] Martin R. Young, A minimax portfolio selection rule with linear programming solution,Management Science, Vol. 44, No. 5 (May, 1998), pp. 673-683

[16] X.Q. Cai, K.L. Teo, X.Q. Yang, X.Y. Zhou, Portfolio optimization under a minimax rule, *Management Science*, Vol.46, 957-972.

[17] Henk Grootveld, Winfried Hallerbach, Variance vs downside risk: is there really that much difference, *European Journal of Operational Research*, 144,Issue 7,304-319, 1999

[18] Y. Simaan, Estimation risk in portfolio selection: the mean variance model versus the mean absolute deviation model, *Management Science*, Vol.43, No.10 ,143-1446, 1997

[19] Yusen Xia, Baoding Liu, Shouyang Wang, K.K. Lai, A model for portfolio selection with order of expected returns, *Computers and operations research*, Vol.27, Issue 5, 409-422, 2000

[20] Xiao-Tie Deng, Zhong-Fei Li, Shou-Yang Wang, A minimax portfolio selection strategy with equilibrium, *European Journal of Operational Research* ,Vol. 166, Issue 1, 278-292, 2005

[21] S. Wang, S. Zhu, On fuzzy portfolio selection problem,*Fuzzy Optimization and Decision Making*, Vol.1 ,Issue 4, 361-377, 2002

[22] M. Inuiguchi, T. Tanino, Portfolio selection under independent possibilistic information, *Fuzzy Sets and Systems*, Vol.115,Issue 1, 83-92,2000

[23] S. Alexander, T.F. Coleman, Y. Li, Minimizing CVaR and VaR for a portfolio of derivatives, *Journal of Banking and Finance*, Vol. 30, Issue 2, 583-605, 2006