Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Forecasting electronic industry EPS Using an integrated ANFIS model

Ching-Hsue Cheng[1]
Department of Information Management
National Yunlin University of Science & Technology Douliou, Yunlin, Taiwan
chcheng@yuntech.edu.tw

Jia-Wei Hsu[2]
Department of Information Management
National Yunlin University of Science & Technology Douliou, Yunlin, Taiwan
g9623740@yuntech.edu.tw

Sue-Fn Huang[3]
Department of Information Management
National Yunlin University of Science & Technology Douliou, Yunlin, Taiwan
g9623801@ yuntech.edu.tw

***Abstract*—The process of buying stock, the major indicator is earning per share (EPS). It is the earning return on original investment; it represents the profit ability of common stock, and the final result of company performance. Therefore, in this study integrates financial-statement related indicators to predict the future EPS. Base on literatures, the relationship of EPS and related attributes is nonlinear, and the nonlinear model can predict well in EPS, so we propose an integrated Adaptive Network-Based Fuzzy Inference System (ANFIS). It combines with the decision tree which is the pre-process for enhancing predicting ability, and there are three stages in study (1) use feature selection to reduce attributes, and the attributes are discretized by decision tree, then encoding the dicretization value (2) take fuzzy inference system (FIS) to fuzzify the encoding value, and use adaptive network to tune optimal parameters. (3) employ an integrated ANFIS model to predict EPS. We collect nine-quarter EPS data for predicting, and then the proposed method surpasses in accuracy these conventional data mining models**

***Keywords*—Earning Per Share, Financial Statement, Decision Tree, Discretization, Adaptive Network-Based Fuzzy Inference System Introduction**

I. INTRODUCTION

Recently the development of global e-commerce and the domestic business-to-business (B2B) and business-to-consumer (B2C) e-commerce market is also a substantial growth year on year. It promotes the development of the electronics industry. The electronic industry in Taiwan, it's the crucial part of economic, especially in semiconductor industry. So to find out the profitability of the company is an important task. Because the importance of EPS, many researchers take attention to investigate. For predicting EPS, they always propose EPS related attributes to construct model.

The relationship between the EPS and the related attributes is nonlinear. In recent years, many researchers have applied the suitable data mining technique - neural network to financial analysis. And in fuzzy inference system (FIS) involves uncertainty, which takes human knowledge into if-then rules and analyzes reasoning process, but neural network and FIS have their own drawbacks. Based on the concepts above, this paper proposes an integrated ANFIS model to forecast the EPS in the Taiwan electronic stock. It combines the advantages of neural network and FIS.

The rest of this paper is organized in the following. Sec. 2 describes related works; Sec. 3 presents briefly the proposed model; Sec. 4 describes experiments and comparisons. Finally, the conclusions of the study are made in Sec.5.

II. RELATED WORKS

This section reviews related studies of the EPS, Decision Tree and ANFIS.

A. EPS

According to the importance of EPS, there are many researches in EPS-related issues. Brown [2] uses the time series method to predict the quarterly EPS in a twenty-three-firm sub sample of the fifty-firm sample data set. In the long-term investment, EPS reflects the future estimate value of a company [1].

So earnings forecasts play an important role in the analytical and empirical literature in Finance and Accounting. And the issue of EPS-predicting is picked up the suitable attributes so that can construct the appropriate forecasting model.

Callen [3] discusses the relationship between EPS and related indicators; he points out that cash flow-related indicators have some degree of predict ability. Charitou [5] shows that a positive relationship between cash flows and dividend which has a strong connection with EPS. Messod [12] takes fundamental analysis for the prediction of extreme stock returns. Zhang [20] also uses cash flow information to predict EPS based on the research by Jegadeesh [10].

In the data mining method, Kenneth [11] shows that logit-based financial statement analysis can predict abnormal returns on investments in equity securities. Qi [13] uses the neural network to predict the stock return.

B. Decision Tree

Artificial Intelligence induction methods offer an alternative to such knowledge acquisition methods for business application. Simplest One of the methods named decision tree, and a decision tree base on collection of rules. Decision tree is a tree branch used to generate the structure of rules. [16].

At the first step of decision tree, an attribute is selected on which to branch that means to partition to each class. Use entropy measure to decide on which attribute to branch. In decision method, C4.5 has great ability to construct mode. C4.5 builds decision trees from a set of training data, using the concept of information entropy. C4.5 uses the fact that each attribute of the data can be used to make a decision that splits the data into smaller subsets. The attribute with the highest normalized information gain is the one used to make the decision. [17].

C. ANFIS: Adaptive-Network-based Fuzzy Inference System

Adaptive Network-Based Fuzzy Inference System (ANFIS) is a fuzzy Sugeno model that adopts the adaptive systems framework to facilitate learning and adaptation [9][19]. And a generalized model of ANFIS has five layers; the whole architecture is as follow:

Layer 1:

Input layer, it takes input variables mapping to fuzzy sets with a node function, such as triangle or bell shaped function. With a membership function, the output should be a degree of membership in linguistic labels. The formula can be expressed as below:

$$O_{1,ij} = \mu_j(X_i) \quad \text{for i=1, 2,…, N j=1,2,…,M} \tag{1}$$

Where X_i is the input variable and μA_j denotes the membership function, $\mu_j(X_i)$

is the membership function of ith input variable to jth fuzzy set,

$$\mu_j = \exp((-(x_i - c_i)/(a_i))^2) \tag{2}$$

Where a_i and c_i are the parameter set which are referred as premise parameters or nonlinear parameters.

Layer 2:

Product layer. In this layer, using permutation and combination to previous result, this defines as follow:

$$O_{2,k} = W_k = \prod_{i=1}^{N} \mu_{ji}(X_i) \quad \text{for ji=1,2,…M k=1,2,…,K} \tag{3}$$

Where $\prod$ denoted the multiplier, in this layer it consists of K nodes that receives the incoming signals and send the product out.

Layer 3:

Normalization layer, The kth node calculates the ith rule's firing strength and output range is [0,1].

$$O_{3,k} = \overline{W_k} = \frac{W_k}{\sum_{k=1}^{K} W_k} \quad \text{for k=1,2,…,K} \tag{4}$$

Layer 4:

Defuzzification layer, f_k is the linear function and r_{ki} is the consequent parameters, it's also the linear parameters we want to simulate.

$$O_{4,k} = \overline{W_k} \; f_k = \overline{W_k} \left(\sum_{k=1}^{K} r_{ki} x_i \right), \quad x_0 = 0$$

$$f_k = r_0 + r_1 x_1 + \ldots + r_k f_k \tag{5}$$

Layer 5:

Output layer, There is a single node in the layer; the overall output is the total of incoming signal. In this layer, the defuzzification process transforms each

fuzzy input into a crisp output.

$$O_{5,1} = \sum_{k=1}^{K} \overline{w_k} f_k = \frac{\sum_{k=1}^{K} w_k f_k}{\sum_{k=1}^{K} w_k} \tag{6}$$

III. PROPOSED MODEL

ANN has been accepted as a potentially useful tool for modeling complex non-linear systems and widely used for prediction [8]. But ANN is a black box method and the rules mined from ANN are not easily understandable. And in fuzzy inference system (FIS) involves uncertainty, which takes human knowledge into if-then rules and analyzes reasoning process, but it's short of accurate quantitative analysis.

So we employ the Adaptive-network-based fuzzy inference system (ANFIS) model to handle the financial data. ANFIS integrates the advantages of ANN and FIS. Furthermore, the ANFIS model employing fuzzy if-then rules can model the qualitative aspects of human knowledge and can be applicable for human to use, and can deal with uncertainty and nonlinear problem.

In this study, we take nonlinear model: the integrated ANFIS model to predict EPS, it takes decision tree for data dicretization to enhance performance, and there are three stages in the proposed model: (1) use feature selection to reduce attributes, and the attributes are discretized by decision tree, then encoding the dicretization value. (2) take fuzzy inference system (FIS) to fuzzify the encoding value, and using adaptive network to tune optimal parameters. (3) employ integrated ANFIS model to predict EPS. Then, the overall procedures of the proposed model are shown as Fig 1.

Then we will show the core concept of proposed algorithm step by step.

Step 1: Data collection and feature selection

The data are collected from TEJ database, and we use feature selection method (decision tree) to select the key attributes. For feature selection, the class EPS need to partitioned to three equal-frequency classes

Step 2: Data discretization

This study uses decision tree (C.45) for data discretization, the advantage of it is minimize entropy concept, keep splitting attribute well into several parts. And the splitting position will be the cut point of each attribute. The step segments the numerical attributes into nominal value. But for input format of ANFIS, it only accepts the numerical data, so the nominal values need to be encoded to ordinal value. It presents the nominal value will be labeled with digit 1, 2, 3, etc.

Step 3: Set the parameters of membership function

In this step, need to initialize membership function to determine the membership degree of each input in linguistic variable, and need to decide the number for membership function, and the type of input and output membership function.

Step 4: Generate FIS

The encoding value will be the input of ANFIS. After deciding the type of membership function for attributes and the number of membership function, the six attributes are partitioned as linguistic values by grid partition. Then, the ANFIS model will generate fuzzy if-then rules, where the linguistic values from input membership functions are used as the if-condition part, and the output membership functions is the then part.

Step 5: Train parameters of FIS from training datasets

In training phase, we employ the least-squares method and the back-propagation gradient descent method for training the forecasting model. The purpose of process is to optimize the parameters of if- then rules.

Step 6: Select optimum FIS model

The choice of FIS model depends on the evaluation of error, the selected FIS structure of the system is considered as the best model for which the root mean squared error (RMSE) is the minimum. The RMSE formulation is defined as:

$$\sqrt{\frac{\sum (F(t) - P(t))^2}{n}} \tag{7}$$

Step 7: Forecast EPS

Use the if-then rules in Step 6 to forecast the EPS in testing data. Then, we can acquire the predicted EPS base on former actual one.

Step 8: Evaluate performance

To evaluate the performance, the final output which means the class (actual and predicted EPS) will be partitioned to three equal-frequency classes for classification (the same amount in each class). Each partition of EPS can be described as low, medium, and high as linguistic attribute. The discretization value of predicted EPS in testing data is used for classification accuracy, which is determined by whether the actual discrete value matches the predicted one.

IV. Experiments and Comparisons

In verifications and comparisons, we get the six attributes by using feature selection from [11][14][15], the six attributes is (1) Diar: Days in accounts receivable (2) Ddiar: % change in attribute Diar (3) Dops: % change in op. profit margin (4) ROTA: Return on total assets (5) STI: Sales to inventory (6) OPITA: Operating profit/total assets

And the quarterly EPS data is collected from 2006Q1 to 2008Q1 (nine subsets) to the proposed model and operate experiment. Each quarter are separated to two parts: training set and testing set, each training set is 66% of one quarter and the remainder data of one quarter 34% is used for testing.

The verification phase takes five data mining techniques as comparison models, to contrast original ANFIS, proposed model (Integrated ANFIS), and Normalized ANFIS. In Normalized ANFIS, make each attribute value need to be standardized in the range [0-1].

V. Conclusions

The modified ANFIS models include proposed model (Integrated ANFIS) and Normalized ANFIS create a great predicting ability in a specific quarter of first experiment and only in 2006Q1 and 2007Q1 present weak performance,

In future work, the time spanning could be extended backward and forward to acquire the boarder data set, more experiments is needed for verification.

References

[1] Abarbanell, J. S., & Bushee, B. J., 1997. Fundamental analysis, future EPS, and stock prices. *Journal of Accounting Research*, 35(1), 1–24.

[2] Brown, L. D., Rozeff, M. S., 1979 Univariate time-series models of quarterly accounting earnings per share: A proposed model. *Journal of Accounting Research*, 17(Spring), 179–189.

[3] Callen, J. L., Kwan, C. C. Y., Yip, P. C. Y., & Yuan, Y., 1996 Neural network forecasting of quarterly accounting earnings. *International Journal of Forecasting*, 12(4), 475–482.

[4] Cessie, L. S., and Houwelingen, V. J.C., 1992 Ridge Estimators in Logistic Regression. *Applied Statistics,* Vol. 41, No. 1, pp. 191-201.

[5] Charitou, A., Vafeas, N., 1998 ´The Association of Earnings and Cash Flow Information with Dividend Changes: An Empirical Investigation, *Journal of Business, Finance and Accounting*, 1998, Vol. 25, No. 1&2, January-March, pp. 225-249.

[6] Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of online learning and an application to boosting. J. *Computer System Science* 55(1), 119–139.

[7] George, F., 1984 The generation of random, binary unordered trees *Journal of Classification*, vol. 1, issue 1, pages 187-233

[8] Ham FM., Kostanic I., 2001 Principles of neurocomputing for science and engineering. *New York: McGraw Hill*

[9] Hilde. V., 2005 Comparison of data-driven Takagi–Sugeno models of rainfall–discharge dynamics, *Journal of Hydrology* 302 173–18

[10] Jegadeesh, N. and Titman, S., 1993 "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance* 48, 65–91.

[11] Kenneth S. Lorek, G. Lee. Willinger, 1998 A Multivariate Time-Series Prediction Model for Cash-Flow Data. *The Accounting Review*

[12] Messod, D., 2001 Contextual Fundamental Analysis Through the Prediction of Extreme Returns, *Review of Accounting Studies*, 6, 165–189

[13] Qi, M., 1999. Nonlinear predictability of stock returns using financial and economic variables. *Journal of Business and Economic Statistics,* 17(4), 419–429.

[14] Ou J. A., 1990, "The Information Content of Non-earnings Accounting Numbers as Earnings Predictors," *Journal of Accounting Research*, Vol. 28, No.1, pp. 44-163

[15] Ou, J. A. and Penman S. H., 1989 "Financial Statement Analysis and the Prediction of Stock Returns." *Journal of Accounting and Economics* 11, 295–329.

[16] Quinlan, J. R., 1993 C4.5: Programs for Machine Learning. *Morgan Kaufmann Publishers*

[17] Quinlan, J. R., 1996 Improved use of continuous attributes in c4.5. *Journal of Artificial Intelligence Research*, 4:77-90

[18] Rumelhart, DE, Hinton GE, Williams RJ.,1986 Learning internal representations by error propagation. Parallel distributed processing: explorations in the microstructure of cognition, vol. I. Cambridge, MA: The MIT Press, 31962.

[19] Takagi. T, Sugeno M., 1985 Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans Syst, Man,Cybernet*;15:116–32.

[20] Zhang, W., Cao, Q., Schniederjans, MJ, 2004 Neural Network Earnings per Share Forecasting Models: A Comparative Analysis of Alternative Methods, *Decision Sciences* Volume 35 Number 2.

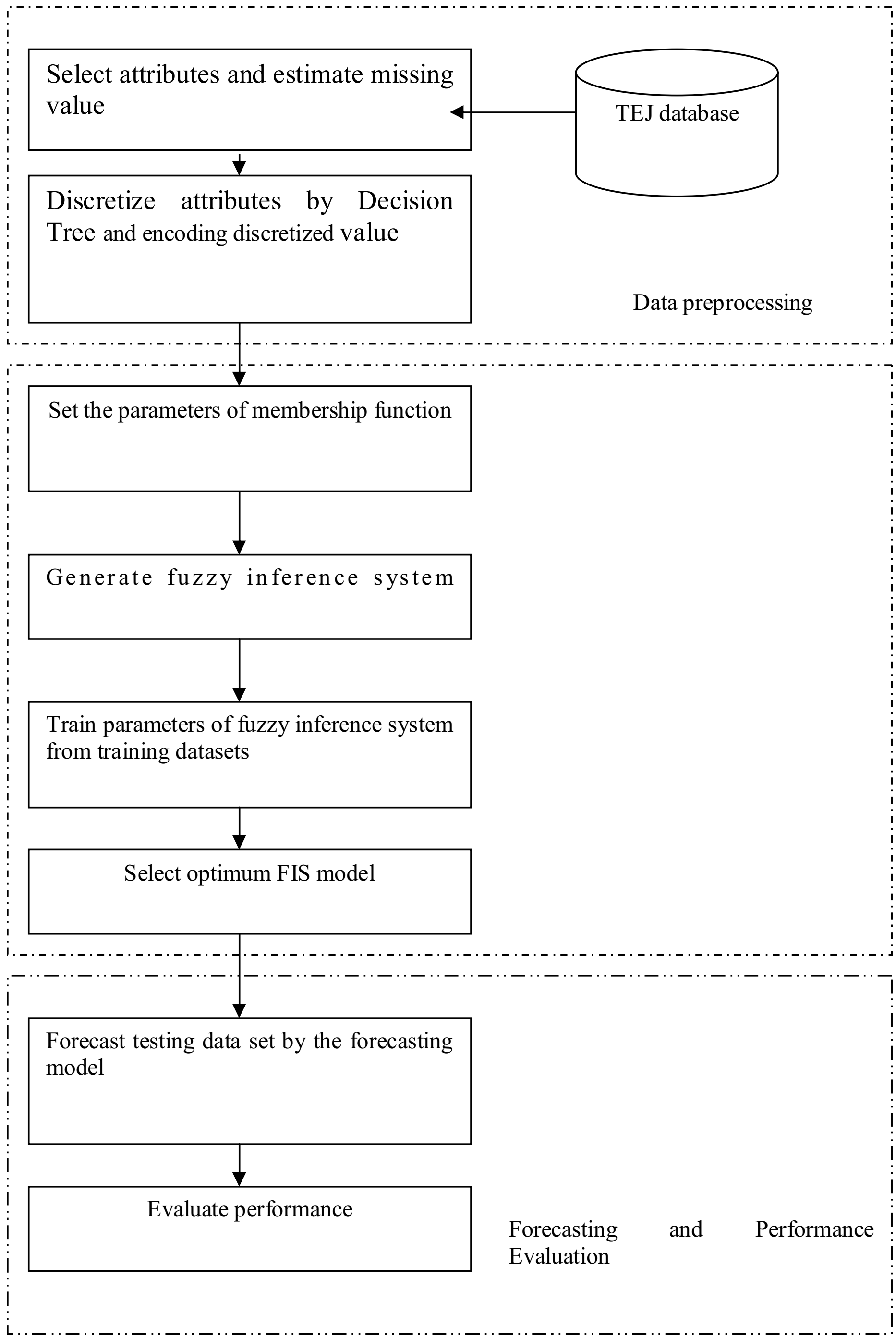

Fig. 1 The proposed procedures

TABLE I. ACCURACY IN SPECIFIC QUARTER

Data Set	ANFIS	Normalized_ ANFIS	Logistic [4]	C4.5 [16]	MLP [18]	Adaboost [6]	RandomTree [7]	Proposed Model
2006Q1	73.39	74.68	74.79	72.24	74.15	69.10	61.89	69.10
2006Q2	63.64	80.95	78.12	77.31	77.14	78.21	67.20	70.56
2006Q3	77.82	76.98	75.82	74.26	75.17	69.31	65.10	77.41
2006Q4	75.0	73.33	75.28	75.28	76.35	74.05	64.95	77.5
2007Q1	76.37	75.95	75.78	75.78	77.48	71.84	65.78	73.31
2007Q2	74.37	74.27	73.20	73.20	72.13	71.22	62.51	74.48
2007Q3	73.27	73.73	73.44	73.44	75.09	71.05	64.39	75.85
2007Q4	78.72	78.30	75.86	75.86	75.49	76.07	64.21	72.77
2008Q1	78.39	78.39	80.22	79.43	79.18	80.13	67.25	81.28