Consolidation Strategies based on Fuzzy Clustering

Yen-Tseng Hsu[1] Ming-Chung Liu[2] Hui-Fen Hung[2] Chin-Tu Huang[1] Dian-Lin Wu[1]

[1]Department of Computer Science and Information Engineering,

[2]Department of Electrical Engineering,

National Taiwan University of Science and Technology, Taipei, Taiwan

Abstract

In this paper, we propose a new method, called FCC(Fuzzy Consolidation Clustering), for finding the right trading points to gain stable profit with low risk. In addition, there are some technical indicators ,such as RSI, Bias, UPP and DNP, used to determine risk levels according to the fuzzy membership functions. The experiments show us the performance of FCC would be better than that of SCC(Statistical Consolidation Clustering). Finally, some trading strategies would be designed to make profit steadily with low risk at all trading points and then beat the market.

Keywords: FCC, SCC, RSI, Bias, UPP, DNP.

1. Introduction

1.1 Research Background and Motivation

In General, people are highly interested in forecasting future tendency of some events, such as investment in stock market, which is necessary to be forecasted for obtaining high profit and reducing the investment risk. Since the prediction is mainly used to reduce the uncertainty or risk in making decision, therefore prediction accuracy is necessary. So far, there are many researchers proposes the different methods to predict the tendency of the stock market in the academic and industry world, respectively [1-4]. Although the above forecasting performance is good, belong to the long-term tendency forecasting mostly, and unwell be used for the consolidation market, so this thesis will regard AiSM (Artificial Intelligence to Stock Market) system [5] as the base, then utilize the fuzzy theory to make the further study analysis, and combine the technical indexes, constructing a decision system which is suitable for short-term operation.

1.2 Fuzzy Sets and Fuzzy Clustering

The term "fuzzy" used here seems to have been first introduced in 1962 by Zadeh[6]. The central concept of fuzzy set theory is the membership function which represents numerically the degree to which an element belongs to a set. Fuzzy set theory proposes to be a tool for handling imprecision, uncertainty and vagueness without undue simplifications, and for giving a consistent representation of linguistically formulated knowledge which allows the use of precise operators and algorithms.

Human decision analysis[7-9] is essentially linguistic. The concepts of linguistic variables were proposed by Zadeh[10] and it is reasonable that we view each attribute as a linguistic variable. Each attribute can be partitioned by various linguistic values by using simple fuzzy partition methods. For instance, k various linguistic values are defined in each quantitative attribute. The value of k is also pre-specified before executing the proposed method. Triangular or trapezoidal membership functions are usually used for the linguistic values. We can classify or cluster our training data using fuzzy clustering approach. The final cluster will be determined according to the highest membership degree among these fuzzy linguistic variables. For example, the attribute "Width" (denoted by x1) that ranges from 0 to 60 are shown as Figure 1. That is, three (i.e., Small, Medium and Large) various linguistic values are defined in Figure 1.

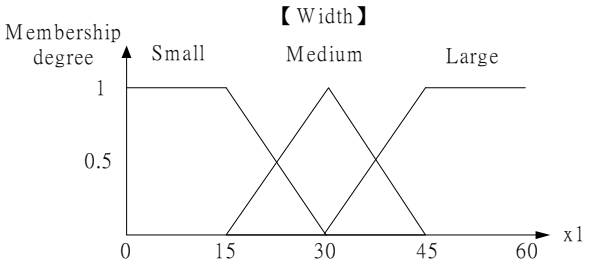

Fig. 1: The fuzzy linguistic variables of attribute "Width".

2. Methodology

2.1 Research Flow

We focus on technical analysis and use some useful indicators as feature vector. First, four kinds of short-

term index, such as UPP,DNP, BIAS, RSI, will be used in the FCC and SCC system, among them, the UPP,DNP is used to judge the amplitude of upward/downward, BIAS is used to judge the degree that bias, as for the RSI is applied to support the BIAS. Second, an AI-based model was proposed through the integration of statistical analysis and fuzzy clustering. Next, we'll introduce some short-term operation strategies for FCC and SCC system. Finally, there are some quantitative metrics are used to estimate the performance of each trading strategy. The research architecture of this thesis is shown as Figure 2.

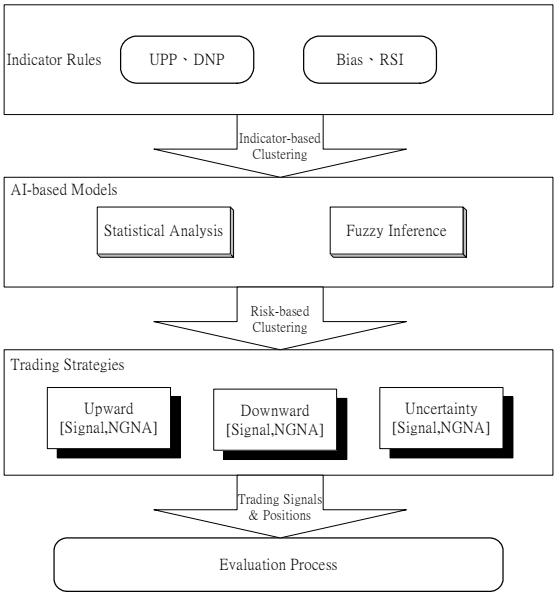

Fig. 2: The research architecture of this thesis.

2.2 Research Data

The sample data in this paper is the past data of TAIEX, which period is from 1987/5/10 to 2005/3/7, be divided into 15 sectors. The front 10 sectors are learning data sets (LDS) and the back 5 ones are testing data sets (TDS), respectively. We will establish the membership functions according to the LDS technical indicator number, after, then can substitute TDS into the membership functions to determine the risk levels. The risk levels are determined based on the time duration and price fluctuation of the consolidatiaon market, including low risk, meddle risk and high risk.

Table 1 Testing Data Set

sector	Date (begin)	Date (end)
TDS(1)	911024	920526
TDS(2)	920704	920814
TDS(3)	920822	921003
TDS(4)	921104	930102
TDS(5)	930517	930714

Table 2 Learning Data Set

sector	Date (begin)	Date (end)
LDS(1)	860313	860613
LDS(2)	861018	861202
LDS(3)	870227	870410
LDS(4)	870828	871222
LDS(5)	880416	880603
LDS(6)	880813	881222
LDS(7)	890429	890824
LDS(8)	891023	891220
LDS(9)	900201	900402
LDS(10)	901207	910502

2.3 Research Method

Since the technical analysis can reflect the market current conditions objectively, and point out clearly that time of the buy/sell, and the approach of fuzzy clustering[11] would obtain outstanding results. Therefore, this paper combines these two methods— Statistical Analysis and Fuzzy Clustering, to analyze stock market for obtaining the high profit in the consolidation market.

2.3.1 Statistical Analysis

Descriptive statistics refers to the graphical presentation of data and the analysis of its statistics and describes a phenomenon by using the historical statistics of samples such as mean, mode, variance, and so on. Chart analysis is just another form of time series analysis, based on a study of the past, which is exactly what is done in all forms of time series analysis. The only type of data anyone has to go on is past data. We can only estimate the future by projecting past experiences into that future. So it seems that the use of past price data to predict the future in technical analysis is grounded in sound statistical concepts. If anyone were to seriously question this aspect of technical forecasting[12], he or she would have to also question the validity of every other form of forecasting based on historical data, which includes all economic and fundamental analysis.

In this paper, we are interested in estimating the expected deadline of consolidation market and the expected price fluctuation by using the statistics of large sample observations , such as mean and mode , in order to determine risk degrees by these heuristic values. The histograms about time duration and price fluctuation as well as their related risk levels are illustrated as Figure 3 and Figure 4.

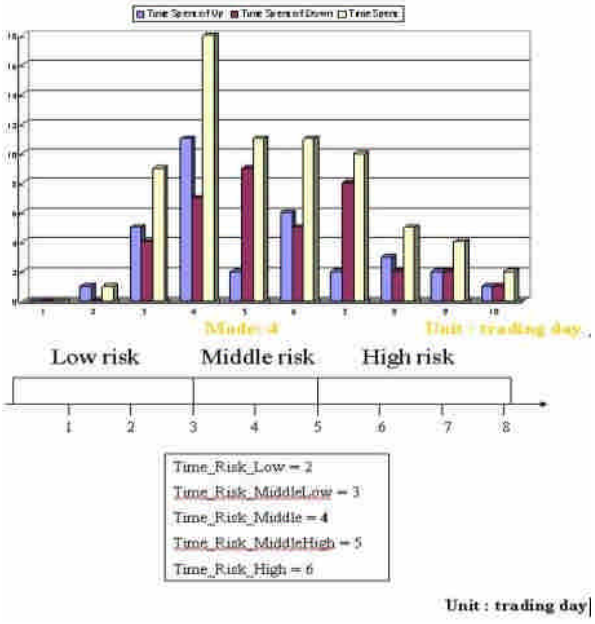

Fig. 3: The histogram of time duration and risk level.

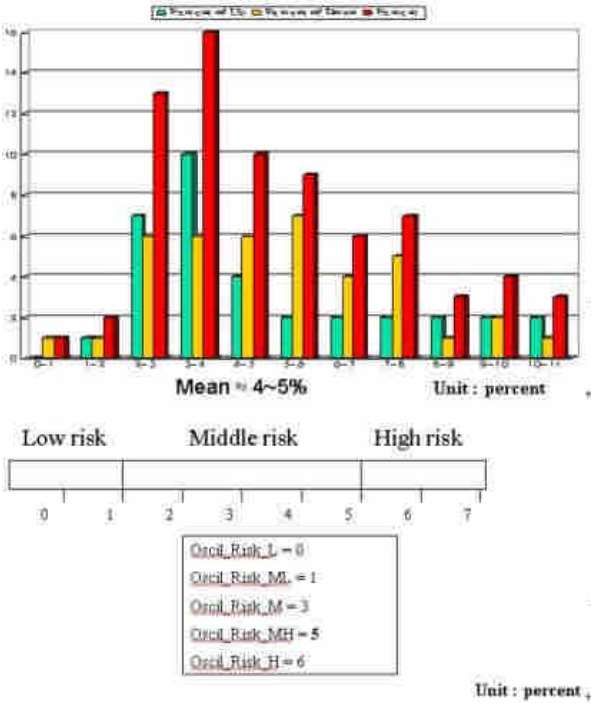

Fig. 4: The histogram of price fluctuation and risks.

2.3.2 Fuzzy Clustering

We would define the membership functions of the fuzzy linguistic terms such as "Low risk", "Middle risk" and "High risk" in this paper for analyzing the status of the consolidation. The membership functions of risk levels about time duration and price fluctuation are shown in Figure 5 and 6, respectively.

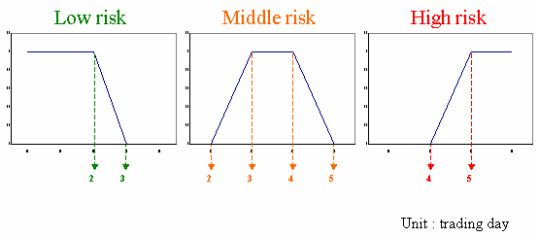

Fig. 5: Membership functions of time duration risk.

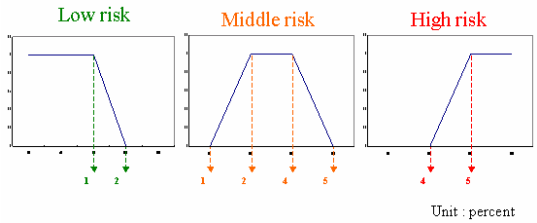

Fig. 6: Membership functions of price fluctuation risk.

3. Experiment Results

The experiments focus on the TAIEX from 1997/3/13 to 2004/7/14. The range of the training data is from 1997/3/13 to 2002/5/2 and the testing region is from 2002/10/24 to 2004/7/14. The experiment results are shown as figure 7, figure 8, table 3 and table 4, respectively.

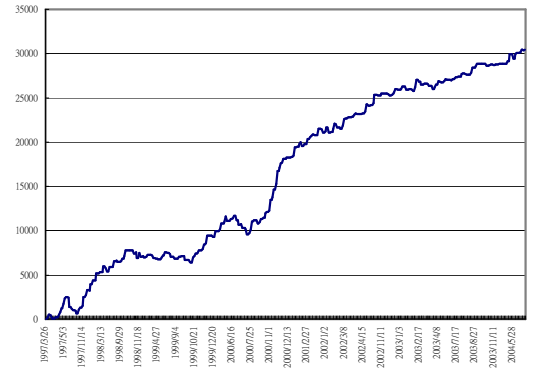

Fig. 7: Profit curves based on FCC method

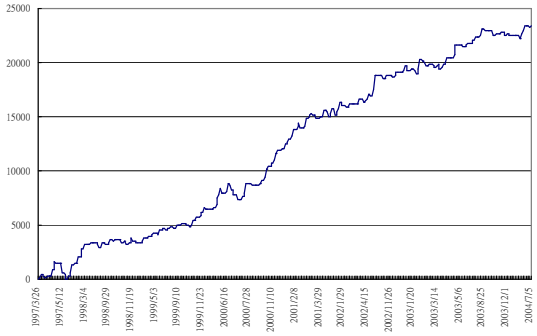

Fig. 8: Profit curves based on SCC method

Table 3. The performance evaluation of FCC

Simulation Date Range	19961019 ~ 20041207		
Total Net Profit	30485	Ratio AvgWin/AvgLoss	1.5
Gross Gain	46141	Profit Factor	2.95
Gross Loss	-15657	Total Net Profit%	1044%
Percent profitable	66%	# of Trades/Month	2.5
Max Conse. Gains	9198	Average Trade	240
Max Conse. Losses	-1862	Gross Gain%	781%
Total # of Trades	258	Gross Loss%	-263%
Total # of winning Trades	170	Average Winning Trade	271
Total # of losing Trades	88	Average Losing Trade	-178
Max DrawDown(3 Month)	-2089	Max consec. Winners	24
Max DrawDown(6 Month)	-2089	Max consec. Losers	6
Max Gain(3 Month)	7904	Max Winning Trade	1475
Max Gain(6 Month)	10025	Max Losing Trade	-1079
Account size required	850000	Ideal Profit	0.4933

Table 4. The performance evaluation of SCC

Simulation Date Range	19961019 ~ 20041207		
Total Net Profit	23395	Ratio AvgWin/AvgLoss	1.6
Gross Gain	37283	Profit Factor	2.68
Gross Loss	-13888	Total Net Profit%	888%
Percent profitable	63%	# of Trades/Month	2.3
Max Conse. Gains	5667	Average Trade	214
Max Conse. Losses	-1508	Gross Gain%	638%
Total # of Trades	239	Gross Loss%	-250%
Total # of winning Trades	150	Average Winning Trade	249
Total # of losing Trades	89	Average Losing Trade	-156
Max DrawDown(3 Month)	-1508	Max consec. Winners	21
Max DrawDown(6 Month)	-1508	Max consec. Losers	6
Max Gain(3 Month)	4931	Max Winning Trade	1399
Max Gain(6 Month)	6637	Max Losing Trade	-844
Account size required	780000	Ideal Profit	0.4572

The experimental results show us the FCC method would be performed better than the SCC one.

4. Conclusions

In this paper we employed two approaches— SCC and FCC, to obtain risk levels for determining the trading points. However, the FCC method is better than the SCC one. The FCC method would be proposed because of its good performance and profitability. In our further works, it is necessary to combine even more technical indicators for inductive analysis and integrate some feasible strategies in order to meet our expectations.

References

[1] Nie Zurong, and Chen Junning, "Grey Analyzing and Predicting Stock Market via Grey Model, " The Journal of Grey System.,Vol 15(3), p. 287-290, 2003.

[2] Hsu,Y.T.,Chen,H.C. and Lin,C.B., "A Lont-term Prediction Using GMs," The Journal of Grey System., Vol 12(1), p.41-54, 2000.

[3] Baba, N. and Kozaki, M., "An Intelligent Forecasting System of Stock Price Using Neural Networks," IJCNN, I:371-377,1992.

[4] Sheng-Chai Chi, Hung-Pin Chen, and Chun-Hao Cheng, "A Forecasting Approach for Stock Index Future Using Grey Theory and Neural Networks," International Joint Conference on,Volume: 6 , 10-16 July 1999.

[5] Hsu,Y.T., "AI-based stock forecasting system (AiSM), " National Taiwan University of Science and Technology, 2001.

[6] L.A. Zadeh, "Fuzzy Sets," Information and Control 8, pp. 338-353, 1965.

[7] R.E. Bellman and L.A. Zadeh, "Decision-making in a Fuzzy Environment," *Management Science* 17, pp. B141-B164, 1970.

[8] L.A. Zadeh, "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," *IEEE Transactions on Systems, Man & Cybernetics SMC-3* , pp. 28-44, 1973.

[9] H.J. Zimmermann and P.Zysno, "Latent Connectives in Human Decision Making," *Fuzzy Sets and Systems* 4, pp. 37-51, 1980.

[10] L.A. Zadeh, "The concept of linguistic variable and its application to approximate reasoning," *Inform. Sci*., 1975

[11] E.H. Ruspini, "A new approach to fuzzy clustering ,"Information and Control. Vol. 15, pp.22-32, 1969.

[12] John J. Murphy "Technical Analysis of the Futures Markets – A Comprehensive Guide to Trading Methods, " 1986.