

Probabilistic Fuzzy Logic based Stock Price Prediction

V. Govindasamy

Research Scholar

Department Of Computer Science and
Engineering

Pondicherry Engineering College

P. Thambidurai, Ph.D

Principal and Professor, Computer Science and
Engineering

Perunthalaiver Kamarajar Institute of Engineering
and Technology , Karaikal

ABSTRACT

In a real-time application scenario, the proposed Probabilistic Fuzzy Logic (PFL) approach can be implemented in various applications such as health care, stock trading, click stream analysis, retail and supply chain management. This work analyses stock trading due to its high non-linear, uncertain and dynamic data over time. Therefore, this paper presents an innovative probabilistic approach for stock price prediction that minimizes the investors risk while investing money in the stock market. We implemented this approach in a publisher/subscriber middleware system, where the crucial Complex Event Processing (CEP) technology processes the large number of incoming stock quotes with the deployment of probabilistic framework. This methodology identifies the event patterns subscribed by the stock traders/brokers over the incoming event stock quotes of market data. This approach triggers an appropriate output event to notify the opportunities to buy and to sell share in real-time based on event patterns of price movements. Experimental evaluation is carried out based on the published data to demonstrate the effectiveness of the proposed approach.

Keywords

Event Processing, Probabilistic Fuzzy Logic, Stock Trading, Data Uncertainty, Complex Event Processing.

1. INTRODUCTION

Nowadays, the national economy is strongly influenced by its performance in stock trading where enormous amounts of capital are traded through the stock market all around the world. The stock market is an inter-network, where significant economic transactions are performed around the world at a dynamic rate. The stock price or value is based on the state of market equilibrium [1]. Recently, the stock market has become as an easily accessible investment tool, not only for strategic investors but also for ordinary people, as well. As a result, it is interlinked to various economic factors. The stock market can influence the everyday life of the common people in a more straight way. However, the main feature of the stock market is uncertainty because the movement of the stock market is highly non-linear and dynamic. This movement is dictated by various numbers of factors such as political events, general economic conditions and trader's demands. This uncertain feature is undesirable for the stock investor and this feature makes the stock price prediction very difficult, but at the same time, it is also unavoidable whenever the stock trading is preferred as an investment tool. Stock market prediction [2] is the act of determining the future stock value of a company on a stock exchange. Hence, the best way to predict the stock price is to reduce the level of uncertainty by analyzing the movement of the stock price. The main motivation of our work is the successful prediction of stock future value that can yield enormous capital profits and can

avoid potential market risk. In order to predict the future stock value, current and previous conditions of the market are analyzed based on economic data. This process involves the gathering of historical data from the internet and then analyzing patterns based on fundamental and technical indicators to predict future events. Several classical approaches have been evolved based on linear time series models, but the patterns of the stock market are not linear. These approaches lead to inaccurate results that may be susceptible to highly dynamic factors such as macro-economic conditions and political events. In the proposed approach, the stock prediction is performed using the Complex Event Processing (CEP) where incoming events are processed based on the event queries. The main objective of this work is

- ♦ To implement the proposed dynamic Probabilistic Fuzzy Logic (PRL) model on stock market prediction. It improves the accuracy of prediction and reduces the risk to take the right investment decision.

The outline of the paper is arranged as follows: Section 2 analyzes the related works. Section 3 proposes an overview of the proposed Probability Fuzzy Logic (PFL) for stock price prediction under uncertainty. Section 4 accounts on experimental set up, simulation results and performance analysis. The conclusion of the paper is given in section 5.

2. RELATED WORK

Prediction of stock price variation is gazed as an extremely challenging task due to the uncertainties in the market movement. The stock market behaves merely like a random walk. In recent times, many researchers proposed various approaches to predict the stock price with the help of fundamental and technical analysis by deploying various stochastic models, machine learning and data mining algorithms. Here, we present a literature appraisal of some of the significant existing approaches.

Ching Long Su et al. used an integrated Neuro-Fuzzy model to estimate the dynamics of the stock market using technical indicators [3]. The neural network identifies patterns and adapts to manage with the stock market movements using the human knowledge incorporated on the fuzzy inference logic. This approach integrates the advantages in both neural and fuzzy model to facilitate reliable intelligent stock value forecasting. However, feature selection is carried out only by fundamental attributes such as momentum and volatility of stock, but their paper does not consider the fractional deviation within a day. To overcome this limitation, Aditya et al. in [4] introduces Hidden Markov models (HMMs) to predict the stock price based on the daily fractional change in the stock share value of intra-day high and low. In order to optimize the HMM model, it is trained using Baum-Welch

algorithm. Further, a maximum posterior approach predicts the stock price using the trained HMM. This algorithm accurately predicts the stock price for more than one day in the future. Moreover, it predicts the stock price using the technical indicators of stock to be predicted rather than other stocks in the same market because of lack of correlations between the stocks in the same market. It leads to poor performance. Therefore, in order to derive the correlation between the technical indicators and reduce the large dimensionality space, the approach in [5] deploys Principal Component Analysis (PCA) to select the most effective technical indicators among the large number of highly correlated variables. It linearly transforms the original large set of input variables into a smaller set of uncorrelated variables to reduce the large dimensionality space. Furthermore, the selected technical indicators are inputted to the stochastic model to predict the stock price. Thus, this approach outperforms other HMM model without PCA by obtaining Mean Absolute Percentage Error (MAPE) less than 1.77%. Like this, Rohit et al. introduces the approach in [6] where Genetic Algorithm selects the most optimal subset of features among the large number of input features, and then selected features are given as an input to the SVM Light software package. Technical analysis is carried out based on technical indicators from the stock to be predicted and also from other stocks that are highly correlated with it. Here, it can accurately predict the stock price using the liner scaling that normalizes each feature component within a specified range. The main drawback is that decision is carried out only based on the input feature variables of technical indicators. It leads to prediction error due to the lack of precise domain knowledge and not consideration of various political and economic factors which affect the stock market other than the technical indicators. In order to acquire knowledge about the domain for stock prediction, a semantic Complex Event Processing is proposed to predict the stock price using the knowledge base that consists of background knowledge of the specific stocks in the market [7]. Here, semantic reasoning is carried out to detect the events based on their hierarchy or temporal/spatial relationships of a stock price to the company attributes in a domain rather than simple syntactic processing with a stock price and volume. Furthermore, expressive SPARQL query language expresses the event queries in a rule language that can process the data fusion of continuously arriving live event streams and knowledge base. It provides real-time processing and high scalability. But, it is so difficult to integrate the external knowledge base for event processing. Furthermore, it would lead to poor performance when the throughput of the incoming event streams is increasing rapidly, and the size of background knowledge is high. Aforementioned approaches are base level standards that have been adopted only to predict the market direction, but none of the approaches has been capable to consistently decide the accurate stock market value. It remains doubtful for many traders and analysts to predict the stock price due to the uncertainties and momentum of the market in real-time. The approach deploying Probabilistic Fuzzy Logic inference in CEP technology must outperform other approaches for command relevance in the stock market prediction.

3. PROPOSED PROBABILITY FUZZY LOGIC (PFL) FOR STOCK PRICE PREDICTION UNDER UNCERTAINTY

The proposed stock prediction approach is modeled as a publisher/subscriber middleware [8] as represented in the Fig.

1. It can predict the stock value for stocks of companies across various sectors of the economy like Automobiles, IT and pharmaceuticals.

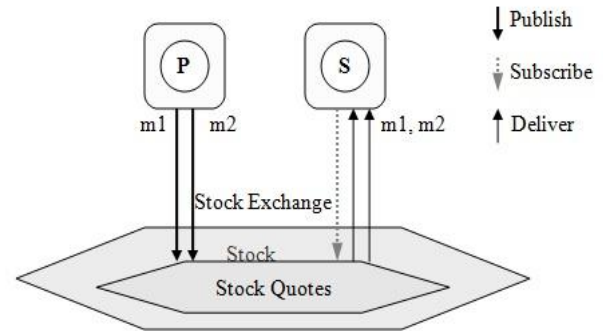


Fig. 1: Stock exchange in pub/sub middleware system

In Fig.1, Publisher (stock exchange) publishes stock quotes to a large number of subscribers (stockbroker, traders, and investors). The stock quotes consist of events related to the various sectors that have five attributes such as a global identifier, the name of the company, the volume of stocks, the price and the identifier of the selling trader. Subscribers who are interested in stock trading can express their interest over the incoming events to buy or to sell a share in accordance to the event patterns available in the stock market. The Probabilistic Fuzzy Logic (PFL) approach predicts the future price of stock value by extracting the higher level knowledge from the large number of incoming complex events from the stock quotes based on the stock request.

3.1 Stock Exchange Scenario

In the stock market, all companies in the different domains have some business dependencies to each other. For example, a company X produces raw material 'm1' for manufacturing car and the business of the car company Y relies on this raw material for its company car production. So, company Y may get into a big dilemma if the company X does not provide the raw material for production. If another finance company Z can provide finance for the car company Y, then Z may get economic problems if the Y does not get raw material properly from the company X. Thus, the business of these three companies is strongly inter-correlated to each other. Let's consider that stock broker/trader wants to invest money in a stock exchange event stream of automobile domain. He knows about this dependency chain and also can define a required complex pattern for the stock exchange stream of complex events. His subscription is to know when the price of the car company Y has started falling to a low /stock price of Rs.90.

3.1.1 Stock Quote

Stock trading consists of two types of stock events such as stock quote and stock request. The stock exchange disseminates the large number of continuously arriving stock data as a Stock Quote and the stock brokers or traders use Stock Request to express their interest in buying/selling stock. Publishers publish the continuously arriving stock exchange stream as follows

```

{
  ...
  {(Name,"TATAMOTORS") (price, 174.88) (volume, 2,735)},
  {(Name,"FORD") (price, 69.31) (volume, 8,991)},
  {(Name,"FIAT INDIA LTD") (price, 189.31) (volume, 1,481)},
  {(Name,"HONDA MOTORS") (price, 87.84) (volume, 6,165)},
  {(Name,"TOYOTA MOTORS") (price, 91.36) (volume, 5,742)},
  {(Name,"MARUTI SUZUKI") (price, 150.36) (volume, 3,354)},
  ...
}

```

Fig. 2: Stock Exchange Event stream

3.1.2 Stock Request

The stockbrokers or traders will like to start a query on the event stream as a Stock Request similar to the query as represented in Fig. 3. They express their interest as event queries, which are represented in the form of Complex Event Pattern Subscription Language (CEPSL) [9]. Suppose a broker wants to invest share in a large car company for the product of a monotonic decrease of stock price less than 90 with volume of 5,000. Due to the decrease in stock price, the stock abruptly rebounds through increasing up to at least 5% in value in the last 15 minutes.

3.2 Stock Prediction using Dynamic Probabilistic Fuzzy Relation Model

From the web, historical stock (open, close, low, high) prices of different companies are collected and constructed in the form of event hierarchy called as a Dynamic Probabilistic Fuzzy Relational model (DPFRM) [10]. It consists of a set of event classes with its associated attributes and then the Joint Probability Distribution is computed using the conditional probabilistic dependencies between the event sequences in accordance of the stock request. The constructed graphical DPFRM model is used to learn the non-linear and dynamic functional relationships between the incoming events to predict the future trends of the market. This model has the capability to perform fundamental and technical analysis of the incoming stock quotes based on a variety of different parameters. In addition, the model will be able to predict the future as to whether to buy or sell the share of the stock. Therefore, this model is used for evaluating the opening price, closing price, highest price and the lowest price of the company stock based on the technical indices, namely Relative Strength Index (RSI), Gross domestic Product (GDP), Manufacturing Index (MI) and Interest Rate (IR) whereas fundamental indices, namely Price-to-Earnings ratio (P/E), Price-to-Sales Ratio (PSR), Return On Equity (ROE), Earnings Growth (EG) and Debt-to-Asset Ratio (D/A).

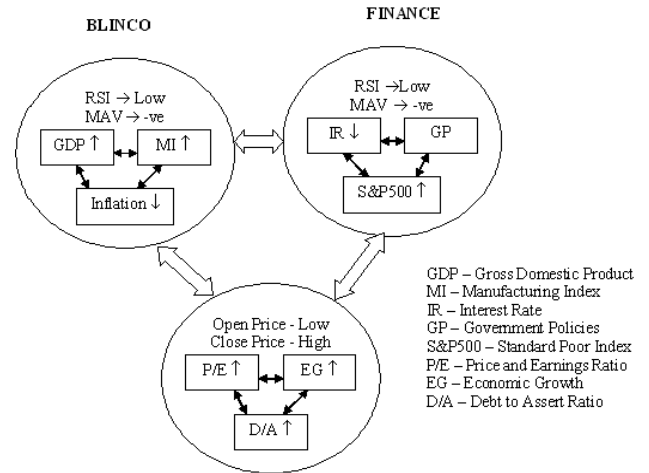


Fig. 4: Dynamic Probabilistic Fuzzy Relational Model

Fig. 4 shows the DPFRM for the event sequences E with a set of events $E = \{e_1, e_2, \dots, e_n\}$ and each event e_i is associated with a set of descriptive attributes and reference slots. The sample space of an event sequence E is represented as the conjunction of a set of possible stock events with its associated probability measure from stock event history.

3.2.1 Conditional Probability Computation of Stock Events

The novelty of the work is that the constructed probabilistic model is used to compute the probability of future stock price based on the conditional probabilistic independencies between the events in stock history. The probability space is represented as triples $\{W_T, \Omega_T, \theta_T\}$ where W_T is a set of possible stock values, Ω_T is a history associated with each possible stock and θ_T is a probability measure of stock value. The Conditional Probability Distribution of event e_i is determined using the probability distribution over the values of events given each combination of stock values in its stock history $P(e_i)$ [11]. It is computed as follows

$$P(e_i/e_{i+1}) = \prod P(e_i / e_{\Omega_i, \Phi_i})$$

Here, Ω_i is the set of stock values in stock history of e_i and Φ_i is the parameter vector associated with e_i . The overall probability associated with the stock events is factorized by aggregating the product of local conditional dependence of the event given their stock history events.

3.2.2 Probabilistic Fuzzy Logic based Inference Engine

In order to make a better decision whether to buy or sell the share for making a profit in the stock market, Probabilistic Fuzzy Logic (PFL) [12] estimates the fuzzy linguistic variables from the computed Conditional Probability Distributions (CPD) in the large probability space. Fuzzy partitioning scheme reduces the large sample space by partitioning the number of possible worlds into a certain number of pre-determined classes using membership function. A fuzzy system consists of a set of rules defined by the characteristic function called as a membership function, which is represented as $\mu_F: P(E) \rightarrow [0, 1]$. It is used to define the certainty that element $P(E)$ belong in that fuzzy set F . The performance of fuzzy logic is strongly influenced by the selection of membership functions in input/output. The fuzzy logic rules are fashioned by domain experts to define a real world situation and also guides how to behave in each case.

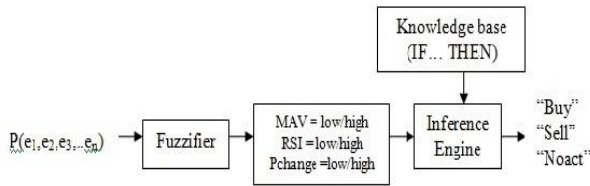


Fig. 5: Probabilistic Fuzzy logic for Stock Prediction

Fig. 5 shows the Probabilistic Fuzzy Logic used in the DPFRM model in which Fuzzifier converts the input probability of the possible worlds (non fuzzy) of a sample space into fuzzy values to train the DPFRM model. Moreover, the knowledge base consists of fuzzy if-then rules. The set of statements below give an idea about the rules that were implemented:

→ If (MAV is negative) and (%price change is Positive) and (Low RSI) THEN “BUY” (Tomorrow close price value > than today’s price)

→ If (MAV is negative) and (%price change is Positive) and (High RSI) THEN “SELL” (tomorrow close price < than today’s price)

→ If (MAV is null) and (%price change is null) and (Stable RSI) THEN “NO ACTION” (tomorrow close price value remains same as today’s price).

The inference engine executes fuzzy logic to map any one of the linguistic variable from the fuzzy sets consists of three linguistic variables {“Buy”, “Sell”, “No Action”} according to the rule base. Finally, the satisfied condition throws the corresponding linguistic variable as the output of the model whether to buy or sell the share to gain profit. In addition to daily stock prediction, it is capable of predicting the open, high, low and close prices of desired stock weekly and monthly.

4. PERFORMANCE EVALUATION

We tested the proposed approach with the automobile stocks of TATA motors, Honda motors, Toyota motors, Fiat India Ltd, Ford as mentioned as Stock Quote in 3.1.1. The data used for evaluating this approach was obtained from the website www.moneycontrol.com that provided the stock prices prevailing at NASDAQ stock quotes. We collected data for the period from July 2012 to April 2013. From the collected data, we obtained the opening, highest, lowest and closing values of the stock price for each day with in this period. To test the proposed approach, we implemented in a publisher/subscriber system using Java Message Service based subscription API in the environment of Java Enterprise Edition/NetBeans/Apache/JMS. It takes incoming events as inputs of various company stocks. And then, it generates suitable output that indicates the type of decision will be suitable for the investor to make a profit in a highly dynamic stock environment.

4.1 Experimental Results

The performance of the proposed PFL approach is evaluated using two common evaluation metrics such as Deviation in predicted stock price versus actual price and Mean Absolute Percentage Error (MAPE).

4.1.1 Accuracy of Stock Prediction

The proposed approach can perform effective prediction where the predicted stock price closely resembles the actual price in the stock market. It is demonstrated in Fig.6. The Fig.

6 shows that the accuracy of stock prediction in the proposed PFL approach over a period from July 2012 to April 2013.

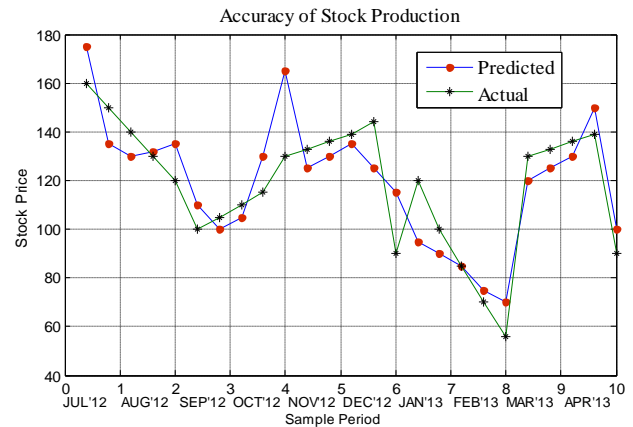


Fig. 6: Accuracy of Stock Prediction

4.1.2 Error Rate Analysis

The error rate of PFL approach of NASDAQ stock is calculated in terms of mean square error which is represented as follows

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{[Actual\ value]_t - [Predicted\ Value]_t}{[Actual\ value]_t} \right|$$

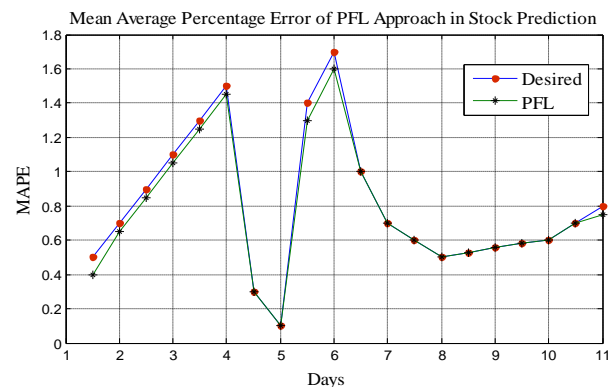


Fig. 8: MAPE of PFL Approach in Stock Prediction

Fig. 8 shows the MAPE of PFL approach in stock prediction versus the desired error rate. It clearly illustrates that the proposed approach achieves the satisfactory error rate as desired under reasonable qualitative decisions.

5. CONCLUSION

This paper demonstrates the outstanding capabilities of the Probabilistic Fuzzy Logic approach for implementing decision making tasks in the stock market prediction. It is mainly proposed for predicting the future direction of stock prices using empirical datasets. Therefore, the PFL approach enhances the effective decision making for investors in the stock market by offering more accurate stock prediction compared to existing approaches. Experimental evaluation is carried out to test the approach under real-time financial data.

The evaluation has shown that the proposed methodology is a very good decision making and a planning tool.

6. REFERENCES

- [1] Hung Pham, Andrew Chien, Young whan Lim, “A Framework for Stock Prediction”, <http://cs229.stanford.edu>, 2013.
- [2] Warrington, Philadelphia, Baltimore, Ogden, ”Stock Market prediction using Natural Language processing”, United States Patent Applications, US20140030981.
- [3] Ching Long Su, Chuen Jyh Chen and Shih Ming Yang, “A Self-Organized Neuro-Fuzzy System for Stock Market Dynamics Modeling and Forecasting”, *International Journal of Education and Information Technologies*, Vol 4, No 3, pp.174-186, 2010.
- [4] Aditya Gupta and Bhuwan Dhingra, “Stock Market Prediction Using Hidden Markov Models”, *Proceedings of the International Conference on Engineering and systems*, 2012.
- [5] Jyoti Badge, “Forecasting of Indian Stock Market by Effective Macro- Economic Factors and Stochastic Model”, *Journal of Statistical and Econometric Methods*, Vol.1, No.2, pp. 39-51, 2012.
- [6] Rohit Choudhry and Kumkum Garg, “A Hybrid Machine Learning System for Stock Market Forecasting”, *World Academy of Science, Engineering and Technology*, 2008.
- [7] Kia Teymourian, Malte Rohde and Adrian Paschke, “Knowledge-Based Processing of Complex Stock Market Events”, *Proceedings of the fifteenth International Conference on Extending Database Technology*, pp. 594-597, USA, 2012.
- [8] Patrick TH. Eugster, Pascal A. Felber, Achid Guerraoui, “The Many Faces of Publish/Subscribe”, *ACM Computing Surveys*, Vol. 35, No. 2, pp. 114–131, 2003.
- [9] Alan Demers, Johannes Gehrke and Biswanath Panda, “Cayuga: A General Purpose Event Monitoring System”, *Proceedings of the Conference on Innovative Data Systems Research*, 2007.
- [10] Sumit Sanghai Pedro Domingos Daniel Weld, “Dynamic Probabilistic Relational Models”, *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pp. 992-997, USA, 2003.
- [11] Daisy Zhe Wang, Eirinaios Michelakis, and Liviu Tancau, “Probabilistic Complex Event Triggering”, *Technical Report, EECS-2009-114*, EECS Department, University of California, Berkeley, 2009.
- [12] Isabel L. Nunes, “Handling Human-Centered Systems Uncertainty Using Fuzzy Logics – A Review”, *The Ergonomics Open Journal*, Vol 3, pp., 38-48, 1010.