A Prediction System Based on Fuzzy Logic

Vaidehi .V ,Monica .S , Mohamed Sheik Safeer .S, Deepika .M[4], Sangeetha .S

ABSTRACT: The main objective of the paper is to build a prediction system to predict the future occurrence of an event. Fuzzy logic, among the various available Artificial Intelligence techniques, emerges as an advantageous technique in predicting future events. Subjective and Objective modeling are two types of fuzzy modeling. Objective type fuzzy modeling is used to build the prediction system. It is a combination of a clustering algorithm and fuzzy system identification which proves effective in improving the efficiency of the prediction. To train the prediction system, historical data is obtained from the web. Data specific to the desired application is obtained and is recorded. This recorded information is subjectively reasoned to develop containing only the necessary inputs to the prediction system. The subtractive clustering algorithm is used for its computational advantages and fuzzy rules are formed using system identification technique. Stock markets are excellent examples where this prediction system can be applied and the possibility of a rise or a fall in the market prices is predicted. The entire prediction system is realized using Java.

Keywords: Prediction, Data modeling, Subtractive clustering, System identification, Fuzzy logic.

I. INTRODUCTION

Prediction of an event requires vague, imperfect and uncertain knowledge [9]. Complexity in a prediction system is its intrinsic characteristic. Various Artificial Intelligence (AI) techniques have been utilized in realising a prediction system [2]. The AI based prediction models can be classified into four groups: models based on neural networks, fuzzy logic, genetic algorithm and expert systems. Such prediction systems play important roles in several organisational decisions, of which the stock market is a vivid example. Rules which determine market behaviour have been elicited from raw data by AI methods. As stock market prediction involves imprecise concepts and imprecise reasoning, fuzzy logic is a natural choice for knowledge representation [2]. The Web, with its boundless information, acts as a source of historical happenings of events. Relevant data concerning the application is parsed and filtered and used to train the prediction system.

Manuscript received Jul 12, 2008.

V.Vaidehi is a faculty and Monica.S, Mohamed Sheik Safeer .S, Deepika .M, Sangeetha.S are students of Department of Electronics Engineering, Madras Institute of Technology, Anna University, Chennai, Tamil Nadu, India

(email id : vaidehivijay@gmail.com)

Earlier, prediction systems were built with rules formed manually. Rules became complicated with increase in the number of inputs and predicting an event grew tedious. Engineering helps build a prediction system that could adapt to the increasing number of inputs and frame rules accordingly. The accuracy and speed obtained is superior to manual prediction schemes.

Objective fuzzy modelling [1] used to build the prediction system requires numerical inputs. The IF-THEN rules formed have vague predicates in their antecedent part while the consequent part is a linear or quadratic combination of the antecedent variables. Since the consequent parts of rules are crisp values rather than vague and fuzzy ones, there is no need to defuzzify the output. This characteristic of the objective fuzzy modeling technique favours it over several other fuzzy modeling techniques and is utilized to improve the prediction efficiency.

The objective type fuzzy modelling has excellent learning capabilities and requires less computational effort. Subtractive clustering technique used operates on raw numerical data. Increasing the number of inputs affects the prediction system only to a small extent. Further this clustering technique provides similar degree of accuracy and robustness together with lesser computational complexity as compared to various other clustering techniques. These advantages along with the characteristic that no separate defuzzification is required, makes this prediction faster than several previous systems.

Section II presents an overview of the prediction system and the inherent concepts. Section III discusses the fuzzy modeling technique in detail with the underlying mathematical foundations. Section IV discusses the implementation results of the technique used and analyses the results. Section V gives a brief conclusion.

II. PREDICTION SYSTEM

Predicting a system is usually done by learning from the past for which historical data is obtained and analyzed to study the resulting pattern in the market [3]. The architecture of the prediction system based on fuzzy logic is given in Fig 1.

Predicting any event requires knowledge about past performance. Data from the past is used mainly to learn the patterns that existed. Historical data provides information on the specific pattern of learning the data. Learning from the past provides knowledge about future to some extent.

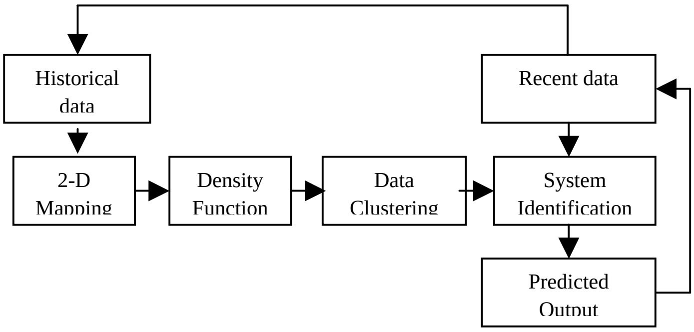

Fig 1 Block Diagram for Fuzzy Based Prediction System

Web feeds provide users with frequently updated content. The block diagram to obtain inputs to the prediction system from web feeds is shown in fig 2. Web pages relating to the specific application are identified. In this case, stock market related web pages are all identified. The payload from the chosen web pages is obtained using a feed reader. The payload could be obtained in any desired format.

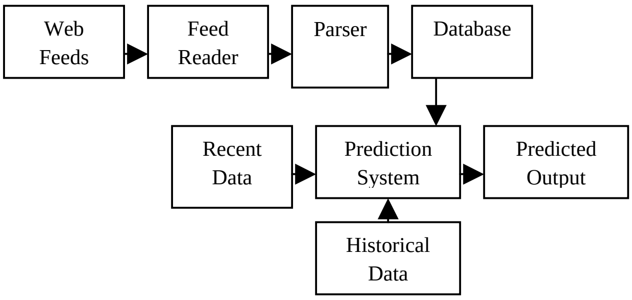

Fig 2 Block diagram showing web inputs to prediction system

XML format is widely used to create most web pages [10]. The payload could hence be obtained in the XML format. The Document Object Model (DOM) is the foundation of Extensible Mark-up Language, or XML. XML documents have a hierarchy of informational units called nodes [11]. The XML DOM (Document Object Model) defines a standard way for accessing and manipulating XML documents. The DOM presents an XML document as a tree structure, with elements, attributes, and text as nodes. Information from the web pages is obtained by parsing the xml document. A database is formed with the parsed data. An analysis of the database provides a picture of the variations in the market due to the numerous available factors ranging from economical to political factors. All these factors can be finally distilled into one market variable, the stock market price.

Stock prices for a day are of various categories like opening price, high, low, closing price, etc [7]. Of these, open and close prices are considered and used to produce a 2-D mapping [6]. The past values of open and close prices of a particular stock is recorded for a sequence of days and stored in database to train the prediction system. A definite number of such pairs of values corresponding to a set of continuous days train the system to learn the pattern of behaviour of the market over a definite period.

III. FUZZY MODELING

The problem of fuzzy system identification is the problem of eliciting IF-THEN rules from raw input and output data. This proceeds through two steps:

1) Clustering

2) Specification of input-output relations (IF-THEN rules)

Clustering algorithms are used extensively not only to organize and categorize data, but are also useful for data compression and model construction [4]. By finding similarities in data, one can represent similar data with fewer symbols.

The density function for a data point is defined as the measure of potential for that data point. It is estimated based on the distance of this data point from all other data points, Therefore, a data point lying in a heap of other data points will have a high chance of being a cluster centre, while a data point which is located in an area of diffused and not concentrated data points will have a low chance of being a cluster centre.

The process of cluster centre determination involves determining the potentials of every data point considered. The data point with the highest potential is chosen as the first cluster centre. The subsequent cluster centres are found by first revising the potential of data points to cancel the effect of the previous cluster centres found. This process is a self-terminating one, that is, when the revised potentials of data points are not sufficient for the particular data point to become a cluster centre, the cluster centre determination terminates. In our case, a set of two clusters is formed - one to denote the higher range in the price values and the other to denote the lower range of prices. The system is now said to be trained.

If a set of recent data values is now presented to this system, the pattern is studied and the possibility of a rise or a fall is predicted along with the next possible value for the market variable, the price. The recent data values are initially processed, their membership with each of the clusters formed earlier is determined. Each of the data points received recently is placed in the cluster where its membership with that cluster is the highest.

ISBN: 978-988-98671-0-2

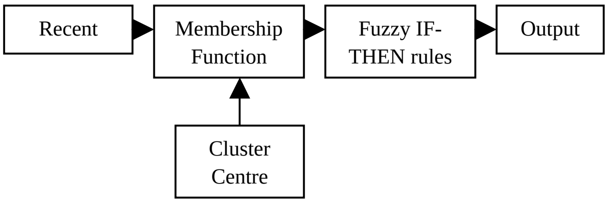

Fig 3 Block Diagram for System Identification

Each cluster centre corresponds to fuzzy rule and the cluster identified by the exponential membership function represents the antecedent of this rule. The rule checks if the input is the same as the exponential membership function of cluster I and if so, then the output is set to the quadratic combination of the input variables.

A higher number of data points placed in one cluster increases the favourability for that cluster, that is, if a set of recent inputs has had a larger number of prices in the higher range, then the possibility is that the prices are likely to rise in the near future. The next immediate value of the price is found using equation 7.

1. *Subtractive Clustering*

Clustering is a process in which data are placed into groups or clusters, such that data in a given cluster tend to be similar to each other, and data in different clusters tend to be dissimilar. When the clustering estimation is applied to a set of input-output data, each cluster centre can be considered as a fuzzy rule that describes the characteristic behaviour of the system. Each cluster centre corresponds to fuzzy rule, and the cluster identified represents the antecedent of this rule. This step forms the system identification.

Subtractive clustering is a technique for automatically generating fuzzy inference systems by detecting clusters in input-output training data. Subtractive clustering, unlike mountain clustering which considers intersection of grid Lines, considers each data point as a potential cluster centre. The measure of potential for a data point is estimated based on the distance of this data point from all other data points. Therefore, a data point lying in a heap of other data points will have a high chance of being a cluster centre, while a data point which is located in an area of diffused and not concentrated data points will have a low chance of being a cluster centre.

After measuring the potential of every data point, the data point with the greatest potential value is selected as the first cluster centre. To find the next cluster centre, potentials of data points must be revised. For each data point, an amount proportional to its distance to the first cluster centre will be subtracted. This reduces the chance of a data point near the first cluster being selected as the next cluster centre. After revising the potential of al1 data points, the data point with the maximum potential will be selected as the next cluster centre. The potential of data points in the first step is measured as [8]:

$$p_i = \sum_{j=1}^{n} e^{-\alpha \|x_i - x_j\|^2} \qquad (1)$$

where $\alpha = \dfrac{4}{r_a^2}$

and x_i is the i^{th} data point and r_a is a vector which consists of positive constants and represents the hyper sphere cluster radius in data space. The constant r_a is effectively the radius defining a neighbourhood; data points outside this radius have little influence on the potential.

The potential which has been calculated through Equation 1 for a given point, is a function of that point's distance to all other points, and the data point which corresponds to maximum potential value is the first cluster centre. Let p_1^* denotes the maximum potential, if x_1^* denotes the first cluster centre corresponding to p_1^*,

$$p_1^* = \bigcup_{i=1}^{n} p_i \qquad (2)$$

where $\bigcup$ denotes the maximum of al1 p_i s

To revise the potential values and select the next cluster, the following formula is suggested.

$$p_i = p_i - p_1^* e^{-\beta \|x_i - x_j^*\|^2} \qquad (3)$$

where $\beta = \dfrac{4}{r_b^2}$

and r_b is a vector which consists of positive constants and is called the hyper sphere penalty radius. The constant r_b is effectively the radius defining the neighbourhood which will have measurable reductions in potential.

To avoid cluster centres being close to each other, r_b must be greater than r_a. A desirable relation is as follows [8]:

$$r_b = 1.5 r_a \qquad (4)$$

Subtractive clustering can be used as a stand-alone approximate clustering algorithm in order to estimate number of clusters and their locations.

2. *Fuzzy Rule Formation*

When the clustering estimation is applied to a set of input-output data, each cluster centre can be considered as a fuzzy rule that describes the characteristic behaviour of the system. Theoretically, a system with multiple inputs and multiple outputs can be reduced to several multiple inputs but single output systems (MISO). Therefore, the fuzzy rule of a MIMO system can also be presented as a set of rules with multi-antecedent and single-consequent such that for a system with *n* output, each multi-consequent rule is broken into *n* single-consequent rules.

Considering data in an n-dimensional space, where the first k dimensions correspond to input variables and (n-k) dimensions correspond to output variables, the clustering estimation on this data space divides the data into fuzzy clusters that overlap with each other. The dependency of each data vector can be defined by a membership grade in [0, 1]. The data vector with membership grade, one, is called the cluster centre. The membership grade of each data vector is defined as follows:

$$\mu_i(x) = e^{-\alpha \|x - x_i^*\|^2} \qquad (5)$$

where x is the input vector.

Each cluster centre c_i corresponds to i, and the fuzzy rule i cluster identified above by the exponential membership function represents the antecedent of this rule. If A_i notifies the exponential membership function of cluster i, then rule i can be represented as:

IF X is A_i THEN Y_i is B_i

where X is the input variables vector, Y_i is the i[th] output variable and B_i is a singleton defined as a linear or quadratic combination of input variables. When *B* is defined as a linear combination, the model is called a first order model and when B is a quadratic combination, the model is called a second order model. For the first order model that we are concerned about in this work, B_i is given as follows:

$$B_i = \sum_{j=1}^{N} p_{ij} x_j + p_{i0} \qquad (6)$$

where, p_{ij} is the coefficient of x_j in rule i. The fuzzy IF-THEN rules for the first order model would be as follows (generically):

IF X is A_i THEN $Y_i(X) = \sum_{j=1}^{N} p_{ij} x_j + p_{i0}$

For a given X_0 the output of the model y_0, is computed as:

$$y_0 = \frac{\sum_{i=1}^{s} \mu_i(x_0) Y_i(x_0)}{\sum_{i=1}^{s} \mu_i(x_0)} \qquad (7)$$

The system is thus formed. And the predicted output of the system is given by equation 7.

IV. RESULTS

Stock prices for five organisations are considered and the prediction system is applied to each organisation. The predictions made for the organisations have an accuracy of about 80%. The results for one of the organisations considered are elaborated. The xml file obtained from the feed reader is given as input to the DOM parser. Historical data comprising of opening price and closing price of a specific organization as collected from the data sheet of a company for a specific period is shown in Table 1.

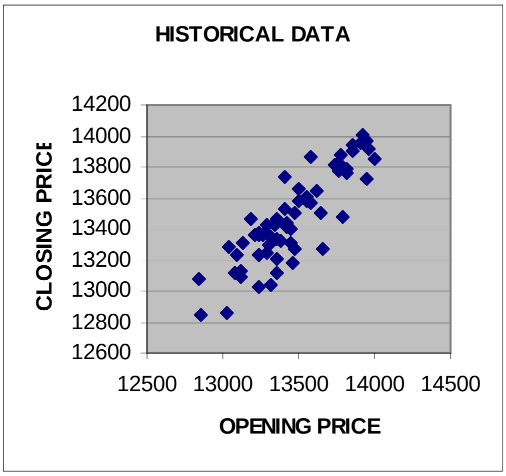

Fig 4 Historical Data

Cluster centres are calculated using subtractive clustering algorithm. The two cluster centres that were obtained for the past values that were considered are shown in Fig 5. The data point at the lower range in Fig 5 indicates the cluster centre for a FALL in the price values.

Table 1 Data Sheet

Date	Open	Close	Date	Open	Close
9/26/2007	13779.3	13878.15	8/9/2007	13652.33	13270.68
9/25/2007	13757.84	13778.65	8/8/2007	13497.23	13657.86
9/24/2007	13821.57	13759.06	8/7/2007	13467.72	13504.3
9/21/2007	13768.33	13820.19	8/6/2007	13183.13	13468.78
9/20/2007	13813.52	13766.7	8/3/2007	13462.25	13181.91
9/19/2007	13740.61	13815.56	8/2/2007	13357.82	13463.33
9/18/2007	13403.18	13739.39	8/1/2007	13211.09	13362.37
9/17/2007	13441.95	13403.42	7/31/2007	13360.66	13211.99
9/14/2007	13421.39	13442.52	7/30/2007	13266.21	13358.31
9/13/2007	13292.38	13424.88	7/27/2007	13472.68	13265.47
9/12/2007	13298.31	13291.65	7/26/2007	13783.12	13473.57
9/11/2007	13129.4	13308.39	7/25/2007	13821.4	13785.79
9/10/2007	13116.39	13127.85	7/24/2007	13940.9	13716.95
9/7/2007	13360.74	13113.38	7/23/2007	13851.73	13943.42
9/6/2007	13306.44	13363.35	7/20/2007	14000.73	13851.08
9/5/2007	13442.85	13305.47	7/19/2007	13918.79	14000.41
9/4/2007	13358.39	13448.86	7/18/2007	13955.05	13918.22
8/31/2007	13240.84	13357.74	7/17/2007	13951.96	13971.55
8/30/2007	13287.91	13238.73	7/16/2007	13907.09	13950.98
8/29/2007	13043.07	13289.29	7/13/2007	13859.86	13907.25
8/28/2007	13318.43	13041.85	7/12/2007	13579.33	13861.73
8/27/2007	13377.16	13322.13	7/11/2007	13500.4	13577.87
8/24/2007	13231.78	13378.87	7/10/2007	13648.59	13501.7
8/23/2007	13237.27	13235.88	7/9/2007	13612.66	13649.97
8/22/2007	13088.26	13236.13	7/6/2007	13559.01	13611.68
8/21/2007	13120.05	13090.86	7/5/2007	13576.24	13565.84
8/20/2007	13078.51	13121.35	7/3/2007	13556.87	13577.3
8/17/2007	12848.05	13079.08	7/2/2007	13409.6	13535.43
8/16/2007	12859.52	12845.78	6/29/2007	13422.61	13408.62
8/15/2007	13021.93	12861.47	6/28/2007	13427.48	13422.28
8/14/2007	13235.72	13028.92	6/27/2007	13336.93	13427.73
8/13/2007	13238.24	13236.53	6/26/2007	13352.37	13337.66

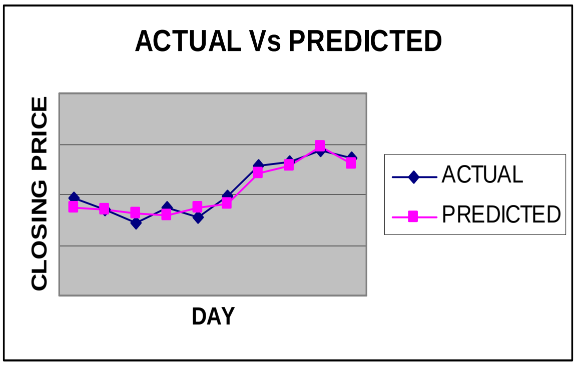

Fig 6 Actual Vs Predicted Curves

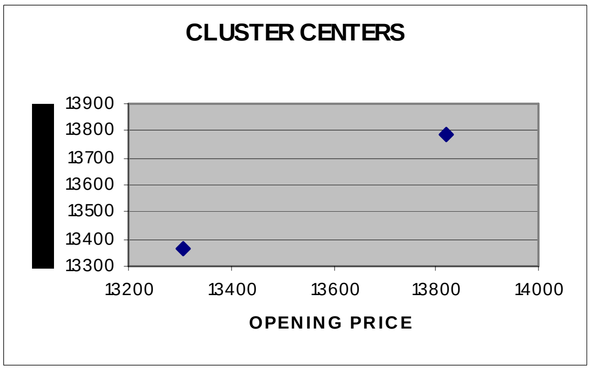

Fig 5 Cluster Centres

Similarly, the data point at the higher range indicates the cluster centre for a RISE in the price values. We have considered 64 values of opening and close prices and have plotted the data points as shown in Fig 4.The recent data given to the trained system produces a predicted output which is indicated as the PREDICTED curve in Fig 6. This curve is plotted in comparison with the actual output, indicated by the ACTUAL curve in Fig 6.

V. CONCLUSION

A subtractive clustering based fuzzy system identification method is used to successfully model a general prediction system that can predict future events by taking samples of past events. Historical data is obtained and is used to train the prediction system. Recent data are given as input to the prediction system. All data are specific to the application at hand. The system, that is developed using Java, is tested in one of the many areas where prediction plays an important role, the stock market. Prices of previous sessions of the market are taken as the potential inputs. When recent data are given to the trained system, it predicts the possibility of a rise or a fall along with the next possible value of data. The accuracy obtained is about 80%. The prediction model that we have designed is trained by daily market price data. It can also be used as a weekly or a monthly predictor. This serves as one possible area of

ISBN: 978-988-98671-0-2

future work. Further, the inputs from the XML document and the fuzzy rules can be integrated to serve a real time application.

REFERENCES

[1] Alaa Sheta, “Software Effort Estimation and Stock Market Prediction Using Takagi-Sugeno Fuzzy Models” IEEE International Conference on Fuzzy System, pp.171-178, July 2006.

[2] Chu S.C. And Kim H.S, "Automatic knowledge generation from the stock market data", Proceedings of 93 Korea Japan joint conference on expert systems, pp. 193-208, 1993.

[3] Justin Wolfers, Eric Zitzewitz, “Prediction markets in theory and practice”, national bureau of economic research, pp.1-11, March 2006.

[4] Khaled Hammouda, Prof. Fakhreddine Karray, “A comparative study of data clustering techniques” pp.1-11, March 2006.

[5] R. Babuska, J. A. Roubos, and H. B. Verbruggen, “Identification of MIMO systems by input-output TS fuzzy models,” in Proceedings of Fuzzy-IEEE’98, Anchorage, Alaska, 1998.

[6] R. J. Van Eyden, “Application of Neural Networks in the Forecasting of Share Prices” Finance and Technology Publishing, 1996.

[7] Ypke Hiemstra, “A Stock Market Forecasting Support System Based on Fuzzy Logic”, Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System, Sciences, IEEE,pp.281-287,1994.

[8] Chiu, S. L.; 1994, "Fuzzy model identification based on cluster estimation", Journal of Intelligent and Fuzzy Systems, 2, John Wiley & Sons, pp. 267-278.

[9] Available online (Process modelling) http://www.itl.nist.gov/div898/handbook/pmd/pmd.htm.

[10] Available online: www.wc3.com

[11] Available online: www.ibm.com/developerWorks