VEGETABLE PRICE PREDICTION BASED ON TIME SERIES ANALYSIS

M.Subhasree
M.Phil Research Scholar
Department of computer science
Psgr krishnammal college for women
Coimbatoure, Tamilnadu

Mrs.C.Arun priya
Assistant Professor
Department of computer science
Psgr krishnammal college for women
Coimbatoure, Tamilnadu

ABSTRACT

Predicting the price vegetable is vegetable is essential in agriculture sector for effective decision making. This forecasting task is quite difficult. Neural network is self adapt and has excellent learning capability and used to solve variety of tasks that are intricate. The two machine learning algorithms namely back propagation neural network and genetic based neural network are compared in this work. The models are assessed and it is concluded from the derived accuracy that the performance of genetic based neural network is better than back propagation neural network percentage of prediction is derived.

Keywords:

Back propagation neural network, genetic algorithm, price prediction,

1. INTRODUCTION

1.1. Data mining

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified.

Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Data

Data are any facts, numbers, or text that can be processed by a computer. Today, organizations are accumulating vast and growing amounts of data in different formats and different databases. This includes:

- operational or transactional data such as, sales, cost, inventory, payroll, and accounting
- nonoperational data, such as industry sales, forecast data, and macro economic data
- meta data - data about the data itself, such as logical database design or data dictionary definitions

1.2. Vegetable prize

As vegetable is one of the most important consumer goods for residents, the fluctuation of the prices has a direct impact on people's daily life. Vegetable prices are affected by several factors such as climate, supply, demand, and festival etc. so the prediction is more difficult than ordinary commercial products. It is very difficult to collect data based on these factors.

Ensuring stability in the prices of essential commodities is an area of major concern for policy makers. Price instability

affects both producers and consumers and has macroeconomic implications as well. A steep rise in the prices of primary commodities spills over to other sectors of the economy and leads to an increase in the overall rate of inflation.

- Data mining provides the methodology to transform these data into useful information for decision making. Vegetable price changes fast and unstable which makes great impact in our daily life. Vegetable price has attributes such as high nonlinear and high noise. So, it is hard to predict the vegetable price.

Machine learning technique

Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. Machine learning explores the construction and study of algorithms that can learn from and make predictions on data.

2. RELATED WORK

Amit jain et.al [1], proposed a technique for the short term load forecasting problem.fuzzy logic is used to obtain next day load forecast .Swarm optimization technique is used on the training data set and Euclidean norm is used with weight factors for the selection of similar days.This provides a reasonable accuracy for the predicted details.

Mukesh,Rohini T.V et.al [5], used multilayer perceptron neural network to solve the problem of atock market prediction.Least mean square algorithm(LMS) and sigmoid delta algorithm are compared to calculate RMS error.Least mean squre algorithm has lower RMS error than sigmoid delta algorithm.Map reduce programming model is used for the large amount of data to provide rapid process.This provides better accuracy for the predicted data and also shows the effectiveness in reducing the time of prediction process.

Chauhan,Bhargwant et.al [12], implemnented neural networks with back propogation algorithm for stock market.Err error rate is reduced using this algorithm.The data's are predicted easily through artificial neural networks

Jaan peralta donate et. al [7], used evolving artificial neural networks [EANN] in forecasting application. Two methods are used to evolve neural networks architectures.They are Genetic Algorithm and differential evolution Algorithm.By comparing these algorithm genetic algorithm gives improved accuracy system in final forecasting by using artificial neural networks.

K.K Sureshkumar et, al [15], implemented predict tools that is used to predict the future stock prices and their performance statistics is evaluated. This would help the investor to analyze better in business decision such as buy or sell a stock.

Alionue Dieng et. al [8], used two forecasting approaches is obtained from the methods and it is evaluvated using quantitative and qualitative criteria.Jenkins autogressive model is used for generating vegetable price forecasts for producers and consumers.

Ozgur kisi et,al [16], presents a comparison of different artificial neural networks (ANNs) algorithms for short term daily streamflow forecasting. Four different ANN algorithms, namely, backpropagation, conjugate gradient, cascade correlation, and Levenberg–Marquardt are applied to continuous streamflow data of the North Platte River in the United States. The models are verified with untrained data. The results from the different algorithms are compared with each other.

3. METHODOLOGY

3.1. Existing system

Vegetable price changes fast and unstable which makes great impact in our daily life. Vegetable price has attributes such as high nonlinear and high noise. So, it is hard to predict the vegetable price. Based on the complexity of vegetable price prediction, making use of the characteristics of data mining classification technique like neural networks such as self adapt, self study and high fault tolerance, to build up the model of back propagation neural networks to predict vegetable price. BPNN is usually based on the error back propagation to the multi-layer Neural Network. In this system, former three week data of tomato price are taken as input and later one week data as output for weekly price prediction. So three input neurons for weekly price prediction consider. Three layer feed forward network structure is used for weekly vegetable price prediction.

3.2. Proposed system

Data Collection and Data Preparation

Vegetable prices are affected by several factors such as climate, supply, demand, and festival etc. so the prediction is more difficult than ordinary commercial products. It is very difficult to collect data based on these factors. Therefore in this paper, we take only the most perisible vegetable price (tomato) as experimental data. Most important point in network design is determining the data size and frequency. This is mostly depends on the final output. For short time forecasting, daily frequency data is preferred. But in this paper weekly data are used for forecasting because it has less noise. Taking previous weekly price of tomato for simulating the model and later few weekly price as test data for the model.

Data Normalization

Normalization is an important issue in Neural Network. Normalization is to transfer the data to fit within the limit of transfer function. Data normalization used to speed up training time by stating the training process for each feature within the same scale. There are many types of data normalization are available, they are Zscore normalization, Minimax sigmoid etc.[12]. Minimax normalization is used in this paper.

$$X' = (X_{max} - X_{min}) * ((X_i - X_{min}) / (X_{max} - X_{min})) + X_{min}$$

Here *X*' is normalized input data, *Xi* is Actual Input, X_{min} and X_{max} are boundary values of the old data range, they are 0 and 1.

Time series is a sequence of data which depend on time. In this paper predict the price *Y* at some future time Y[t+1] = f(Y[t],Y[t-1],…). The time series data will be transformed into a data set depending on the *Y* input nodes of a particular ANN and each data set will consist of the following:·

Y input values that correspond to y normalized previous values of period t: N_{t-1}, N_{t-2}, …, N_{t-2},]

One output value : N_t

This data set will be used to train validate each ANN. The data set will be split into two subset one for network training and another for network validation.

Structure Construction

The structure of the network affects the accuracy of the prediction. Configuration on the network depends on the number of hidden layers number of neurons in each hidden layer, and activation function.

There is no clear cut guideline for deciding the architecture of ANN. It is problem dependent, and there is no formula to determine number of neurons in hidden layers. If number of neurons in the hidden layer is increased then the computation time will be more. The exact number of neurons in the hidden layer determined is based on experience.

The number of neurons in the hidden layer can be selected by one of the following thumb rules:

a) (n – 1) neurons, where n is number of input neurons.

b) (n + 1) neurons, where n is number of input neurons.

c) For every input neuron, 8 hidden neuron can be taken

d) Number of input neuron / number of output neuron

e) Half the sum of input and output neuron

f) P / n neuron, were n is the number of neurons and P represents number of training sample

In this paper, former three week data of tomato price are taken as input and later one week data as output for weekly price prediction. So three input neurons for weekly price prediction consider. Three layer feed forward network structure is used for weekly vegetable price prediction. The network structure includes input layer, hidden layer and output layer. The connection from one nerve cell to all nerve cells in the next layer. But there is no connection among nerve cells at the same layer. Because the price of the vegetable which is output under certain period, is the price of input in the previous period. Choice of activation function, learning rate and optimization target were determined by experiment. In this paper activation function from input layer to hidden layer is tansig() and hidden to output layer is purelin(). The optimization algorithms were compared and Levenberg-Marquardt algorithm was chosen, which leads to fast convergence and higher hit rate compared to gradient decent algorithm.

Construction of Neural Network Model Based on Genetic Algorithm

The neural network based on GA is constructed. Its structure is similar to BPNN. The main process of using GA for optimization of neural network model is as follow:

(1) Gene Encoding. According to the BPNN, gain its weight number. Every weight is on behalf of a gene. All of them structure a chromosome.

(2) Initial chromosome group generation. Choose the number of chromosome in initial population. To each chromosome, generate weights randomly in the given range to construct the initial group.

(3) Individual fitness computation. Use training samples to train the individual chromosome which is on behalf of an ANN, and then, calculate the individual learning error E. The formula is as follow:

$$E = \sum_{i=1}^{N} E_i$$, there into, $$E = \sum_{i=1}^{N} \sum_{l} (y_l^i - c_l^i) \frac{\square^2}{2}$$

Here, n is the number of training sample, m is the number of output unit. $(y_l^i - c_l^i)$ is the difference between actual value and expected value of l-output when it takes i-sample to train. Fitness function fs is as bellow:

fs = 1/ E

(4) Selecting operation. Select the individual taking roulette wheel and retain the best individuals.

(5) Crossover operation. Assume x1 and x2 is parents, its children y1 and y2 after crossover is gained by the formula as bellow:

$$Y_1 = \alpha X_1 + (1-\alpha)x_2, y_2 = \alpha X_2 + (1-\alpha)x_1$$

Here, α is a parameter which changed with the evolution algebra.

(6) Variation. Take Gaussian approximate variance to improve the local search performance of GA on key search area. During the variation, using a random number of normal distribution which average is *P* and variance is P2 instead of original gene.

(7) Fitness value of chromosomes group is calculated again.

(8) If it meets the stop search criteria, output the result. Otherwise, go to step (4).

4. EXPERIMENTAL RESULT

The existing Back-propagation neural network (BPNN) and proposed genetic algorithm based neural network (GANN) model is used to predict vegetable price in terms of accuracy and Mean square error (MSE).

4.1. Estimation of Accuracy

The Accuracy of the system is measured by following,

Accuracy = 100 – MSE

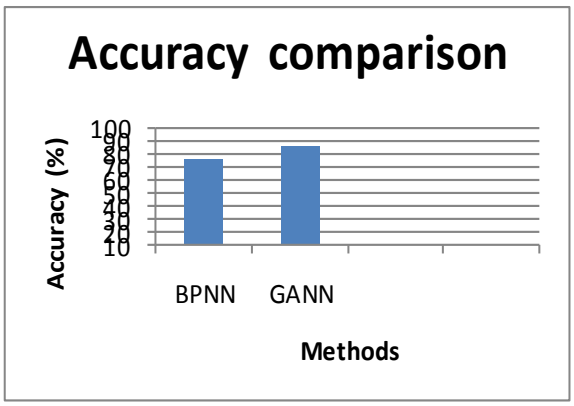


Figure 1. Accuracy graph

The Accuracy is shown in this graph. In the X-axis comparison of BPNN and GANN is taken. Y-axis Accuracy is taken. The Accuracy of proposed system is high compared to the existing Back-propagation neural network (BPNN) method.

4.2. Mean square error (MSE)

There are many measuring of predictor error, such as the mean square error.

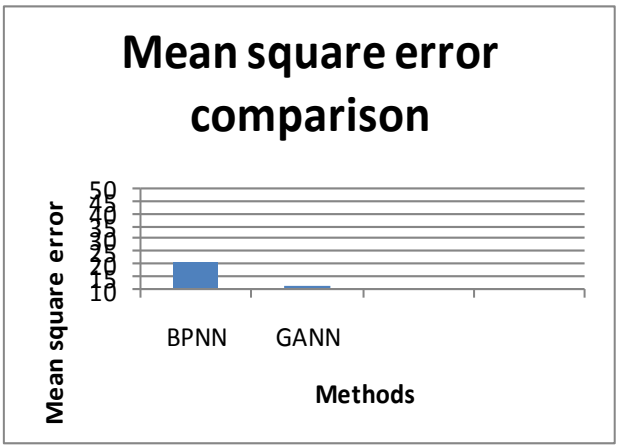
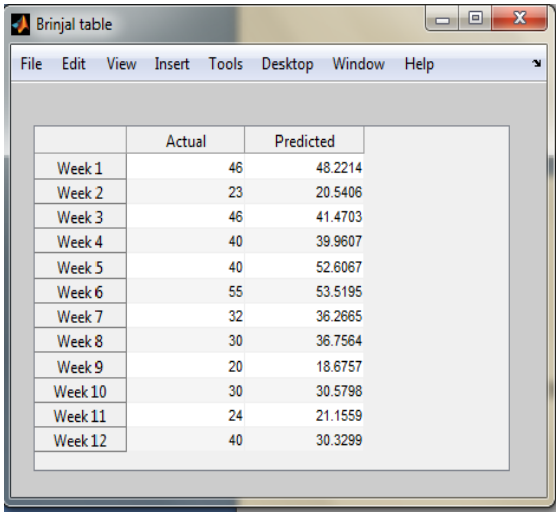


Figure 2. Mean square error graph

$$MSE = [1/2 \sum_{i=1}^{N} (T_i - O_i)^2]/N$$

Where, *Ti* = target value, *i* = actual value, *N* = number of pattern The Mean square error is shown in this graph. In the X-axis comparison of BPNN and GANN is taken. Y-axis Mean square error is taken. The Mean square error of proposed system is low compared to the existing Back-propagation neural network (BPNN) method.



	Actual	Predicted
Week 1	46	48.2214
Week 2	23	20.5406
Week 3	46	41.4703
Week 4	40	39.9607
Week 5	40	52.6067
Week 6	55	53.5195
Week 7	32	36.2665
Week 8	30	36.7564
Week 9	20	18.6757
Week 10	30	30.5798
Week 11	24	21.1559
Week 12	40	30.3299

Figure. 3 Predicated price result

Figure.3 shows the sample result of predicated price of a vegetable.

5. CONCLUSION

Vegetable price changes fast and unstable which makes great impact in our daily life. Vegetable price has attributes such as high nonlinear and high noise**.** BPNN is good at simulation, but relatively poor at prediction. The Genetic Algorithm based neural network is constructed for price prediction to increase the accuracy percentage. Here five vegetable prizes are predicted through Genetic Algorithm based neural network. The experimental results show that, the proposed system achieves high prediction performance compared with existing systems in terms of accuracy and mean square error.

REFERENCE

[1] Amit Jain and M. Babita Jain," A Novel Hybrid Method for Short Term Load Forecasting using Fuzzy Logic and Particle Swarm Optimization", international conference on power system technology IEEE, 2010, Vol. 2, No. 3, PP 01-07.

[2] Kunwar Singh Vaisla ,Dr. Ashutosh Kumar Bhatt,"An Analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting", International Journal on Computer Science and Engineering, 2010, Vol. 02, No. 06, pp 2104-2109.

[3] Dieng, A.,. Alternative Forecasting Techniques for Vegetable Prices in Senegal, Revue sénégalais de richerches agricoles etagroallementalress, 2008 Vol. 1, No. 3, pp 5-10.

[4] Z.Tang and P.A.Fishwick, "Backpropagation neural nets as models for time series forecasting," ORSA journal on computing, 1993, vol. 5, No. 4, pp 374 - 384,.

[5]Mukesh and Rohini T V, "Market Price Prediction Based on Neural Network Using Hadoop MapReduce Technique", International Conference on Innovative Technology and research, 2015, Vol. 2, No. 6, pp 69-73.

[6] Zabir Haider Khan, Tasnim Sharmin Alin and Sylhet, Bangladesh, "Price Prediction of Share Market using Artificial Neural Network (ANN)", International Journal of Computer Application, 2011, vol. 22, no. 2, pp 42-47.

[7]Juan Peralta Donate, Xiaodong Li, German Gutierrez Sanchez,& Araceli Sanchis de Miguel., "Time series forecasting by evolving Artificial Neural Networks with Genetic algorithm, Differential evaluation and Estimation of distribution algorithm", Springer-verlag London limited, 2013, Vol. 22, Issue. 1, pp 11-20.

[8]Alionue Dieng ., "Alternative Forecasting Techniques And Vegetable Prices In Senegal" Revue senegalaise des recherché agroalimentaires, 2008, Vol. 1, No. 3, pp 6-10.

[9]Changshou Luo, "Prediction of vegetable price based on Neural Network and genetic algorithm", International Federation for Information Processing, 2011, Vol. 102, pp 672-681.

[10]Gu,Q.W.,Chen,G.,Zhu ,L.L "Short –term marginal price forecasting based on genetic algorithm and radial basis function neural network". Power System Technology, 2006, vol.25, No.7, pp 18-21.

[11] Manish Shukla and Sanjay Jharkharia "Applicability of ARIMA Models in Wholesale Vegetable Market" International Conference on Industrial Engineering and Operations Management, 2011, Vol. 102, No. 2, pp 1125-1130.

[12] Bhagwant Chauhan, Umesh Bidave, Ajit Gangathade, Sachin Kale," Stock Market Prediction Using Artificial Neural Networks " , / (IJCSIT) International Journal of Computer Science and Information Technologies", 2014, Vol. 5, No.1, pp 904-907.

[13] Md. Syedul Amin,Md. Mamun, Fazida Hanim Hashim,Jubayer Jalil and HazahHusain, "Design

and Implementation of Novel Artificial Neural Network Based StockMarket Forecasting System on Field-Programmable Gate Arrays", American Journalof Applied Sciences, 2011, Vol. 8, No. 10, pp 1054-1060, ISSN 1546-9239.

[14] Zabir Haider Khan, Tasnim Sharmin Alin and Sylhet, Bangladesh," Price Prediction of Share Market using Artificial Neural Network (ANN)", International Journal of Computer Applications, 2011, Vol. 22, No. 2, pp 0975 – 8887.

[15] K.K.Sureshkumar , Dr.N.M.Elango "Performance Analysis of Stock Price Prediction using Artificial Neual Network", Global Journal of Computer Science and Technology, 2012, vol-12 No 1 pp 19-25.

[16] Ozgur Kisi "Streamflow Forecasting Using Different Artificial Neural Network Algorithms" Journal of Hydrologic Engineering 2007 vol 12 issue 5 pp 532-539.

Authors Details

1.M.Subhasree, subhasreem03@gmail.com, contact no:9629301750

2.C.Arun Priya, arunpriya.bs@gmail.com, contact no: 9095014100