International Journal of Computational Intelligence and Informatics, Vol. 2: No. 2, July - September 2012

Forecasting Model for Vegetable Price Using Back Propagation Neural Network

G.M.Nasira
Asst. Professor
Department of Computer Science
Govt. Arts College (Autonomous)
Coimbatore, India
nasiragm99@yahoo.com

N.Hemageetha
Asst. Professor
Department of Computer Science
Govt. Arts College for Women
Salem, India
geekani2010@gmail.com

***Abstract*- The Agricultural sector needs more support for its development in developing countries like India. Price prediction helps the farmers and also the Government to make effective decision. 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 network (BPNN) to predict vegetable price. A prediction model was set up by applying the neural network. Taking tomato as an example, the parameters of the model are analyzed through experiment. At the end of the result of Back-propagation neural network shows accuracy percentage of the price prediction.**

Keywords- data mining, neural networks, back-propagation (BP), vegetable price

I. INTRODUCTION

Data mining is the process of extracting important and useful information from large sets of data Abello. J. Pardalos PM, Resende M [1]. Data mining in agriculture is a novel research field. Farmers are not only harvesting vegetables and crops but also harvesting large amount of data. 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. Data mining classification techniques can be used to develop an innovative model to predict the market price of respective commodity. Price prediction is highly useful in agriculture for forecasting the market price for the respective commodities and also useful for farmers to plan their crop cultivation activities so that they could fetch more price in the market. Government can use the market forecast price for planning and implementation of agriculture development programs to stabilize the market price for the respective commodity. Consumers can use this price prediction for their daily lifestyle planning. This innovative application is not only useful for farmers and consumers but also useful for agriculture planning; framing polices and schemes in agriculture and market planning. Time series forecasting takes an existing series of data to predict future value. Data mining classification technique such as Neural Network plays an important role in non-linear time series prediction [2, 3, 4]. There are many kinds of prediction method on basis of Neural Network, among them the application of BP Neural Network algorithm is most important one.

II. LITERATURE REVIEW

There has been large number of studying on forecasting of vegetable price. This section presents a very brief review of the related and recent studies. Alionue Dieng [5] investigated the performance of parametric models for forecasting vegetable prices and to make recommendation to potential user. The author used two forecasting approaches. The forecasting methods used consist of three alternative parameter models and a non parametric model. The parametric models consist of the naïve model, exponential smoothing models and box-Jenkins interacted moving average model (ARIMA). The non paramedic model uses the spectral analysis. The author collected monthly average price of tomato, potato and onion for the year 1980 to 2003. In this study both parametric model and non parametric model were used to generate forecasting of potato, tomato and onion price. Based on the results the parametric models would be recommended for forecasting vegetable price. Among these ARIMA model receives high priority. Koffi N.Amegbeto [6] presented a study of examined the dynamics of selected vegetable prices and the quantities supplied to the main fruit and vegetable market in Kabul, Afghanistan. Forecasting models were developed to aid in vegetable marketing decisions using data collected from Aug 2004 to Dec 2005. The results show that prices and supplies of certain vegetables were erratic and

negatively correlated in most cases and vegetables were mainly supplied to Kabul market between November and April, which generally corresponds to low supplies and high prices. Manish Shukla and Sanjay Jharkharia [7] investigated the ability of ARIMA model in wholesale vegetable market; models were built for sales data of one perishable vegetable for Ahmadabad wholesales market in India. Goal of this paper is studying the application of ARIMA models on vegetable wholesale data and to forecast the future demand with accuracy. In [8] presented a prediction model of vegetables price was set up by applying the neural network based genetic algorithm. Taking mushrooms price as an example. Chang shou Luo, Qingfeng wei, liying zhou, fwnf zhang and sufen sum [9] present a integrated method to forecast monthly vegetable prices. In this paper four models were constructed for the prediction. The authors collected lentinus edodes price from 2003 to 2009 for Beijing xinfudi market. The results showed that the integrated prediction model was the best one. The aim of this paper is to develop Neural Network model that can be used to predict the price of tomato in Coimbatore market.

III. ARTIFICIAL NEURAL NETWORK(ANN)

ANN is an emulation of biological neural network which is composed of many interconnected neurons. The most used kind of ANNs is the multilayer perception, in which neurons are organized in layers. The input layer neurons receive the input signal, which is then fed in the network. These neurons do not perform any task. The neurons on the output layers are active and the result they provide is considered as the output provided by the network. There are some hidden layers between the input and output layer. Each neuron can receive input signal from the neuron belongs to previous layer and it can send its output to neurons belong to the successive layer. BPNN is usually based on the error back propagation to the multi-layer Neural Network. It is designed by D.E.Rumelhart and J.L.McColland and research team in 1986.

The main steps in BP algorithm as follows:

Step 1: Feed the normalized input data to the network and Compute the corresponding results.

Step 2: Compare the error between calculated result and actual result.

Step 3: The connection weight and membership functions are adjusted based on the error.

Step 4: If error greater than the tolerance then go to Step 1 else stop the process.

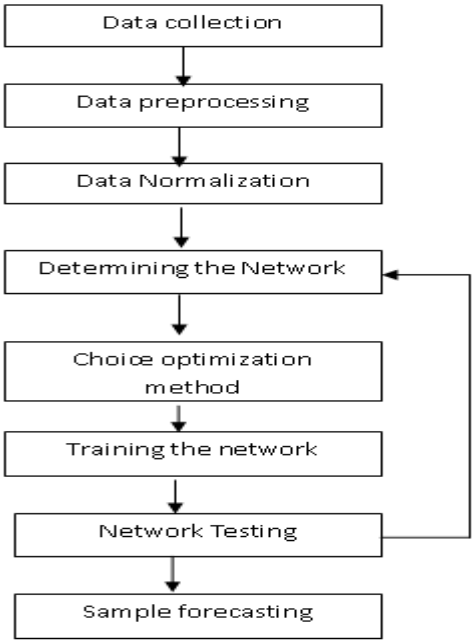

Figure 1. A flow chart for developing ANN

A. 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. In this study tomato price data from JAN 2009 to MARCH 2012 are taken for creating the model. Taking previous weekly price of tomato from Coimbatore market for simulating the model and later few weekly price as test data for the model. The data are collected from the web site www.tnau.ac.in. Sunday price is not available, so Sunday price has been dropped for price prediction.

B. 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 Z-score normalization, Minimax sigmoid etc.[12].

Minimax normalization is used in this paper.

$$X' = (X_{max} - X_{min}) * ((X_i - X_{min}) / (X_{max} - X_{min})) + X_{min} \quad (1)$$

Here X' is normalized input data, X_i 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 $Y[t+1] = f(Y[t],Y[t-1],\ldots)$. 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-k}
- 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.

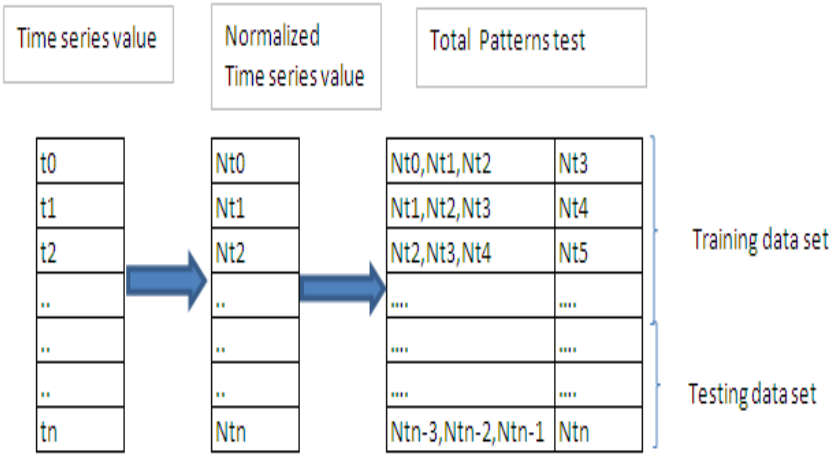

Figure 2. Train and Validation Data set

C. Strucutre 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. Gowri T. M. and Reddy V.V.C. [10] suggest guideline for a 3 layer ANN.

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.

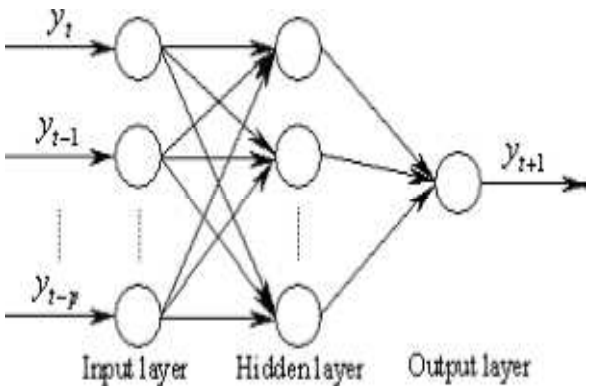

Figure 3. Architecture of the network

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.

IV. EXPERIMENTAL RESEARCH AND ANALYSIS

A. Weekly Prediction Analysis of BPNN

Jan 2009 to Dec 2010 weekly prices data have been taken for prediction in paper [11] so, consider this data as first data set. In this paper Jan 2009 to Mar 2012 weekly price data of tomato has been taken for prediction. This new data set as second data set. Taking former three week price data as input and later one week price data as output. For model construction and simulation coding MATLAB is used. BPNN is constructed using previous 155 weeks from Jan 2009 price data and later week's data are used to test the model. Several combination of number of neurons in the hidden layer, learning rate, epochs are tried and finally the various parameters are determined.

The various parameters used in this paper are:

Network	: 3-4-1
No. of hidden Layer	: 04
No. of input neurons	: 03
Transfer function	: sigmoid
Error calculation	: MSE
Learning rate	: 0.06
Network tolerance	: 0.001

A Part of MATLAB code as follows:

```
t=[ 0 1; 0 1; 0 1]
nnet = newff( t,[ 4, 1 ],{ 'tansig', 'purelin' }, 'trainlm' );
nnet.trainParam.show = 5;
nnet.trainParam.epochs = 30000;
nnet.trainParam.goal = 1e-3;
nnet.trainParam.lr = .06;
[ nnet, ttr ]=train( nnet, x1, y1 )
A= sim( nnet, x2 )
```

The results are as follows,

TABLE I. ACTUAL VALUE, PREDICTED VALUE

week	Actual value	Predicted value for dataset 1	Predicted value for dataset 2
1	26	20	23
2	31	26	25
3	34	28	26
4	25	24	27
5	18	20	20
6	18	18	17
7	14	17	18
8	11	13	13
9	9	10	11
10	10	10	9

B. Estimation of Accuracy

There are many measuring of predictor error, such as the mean square error, the mean absolute error, relative square error and relative absolute error [12]. Mean square error is used in this paper.

Mean square error (MSE) = $[\ ½\ \sum_{i=1}^{N} (Ti - O_i)^2\ ]\ /\ N$

Ti = target value, Oi = actual value, N = number of patterns

Accuracy = 100 – MSE.

The back propagation neural network was used in this work which gives accuracy of 89.2%. The corresponding mean square error for each data set is shown in the table.

TABLE II. MEAN SQUARE ERROR AND ACCURACY

Data set	Mean square error	Accuracy
First Data set	11.6	88.4
Second dataset	10.8	89.2

In the first dataset 135 weeks price data are used but in the second data set 160 weeks data are used. The result shows that the second dataset have produced more accuracy result than the first one. So ANN will produce more accurate result for larger dataset compared with smaller data set.

V. CONCLUTION

In this paper, the BP neural network prediction model of vegetable market price is established. Three years and three months Coimbatore market price of tomato as an example and simulated the result using Matlab and predict the result. The prediction results of weekly are discussed. The result shows that the larger dataset produced more accuracy result than the smaller data set. The result shows that neural network is one way of predicting the market price of vegetable with the non-linear time series. In future the Genetic Algorithm based neural network will be constructed for price prediction to increase the accuracy percentage.

REFERENCES

[1] Abello, J, Pardalos, P.M, and Resende, M., Hand book of massive datasets. Vol. 3 of the Kluwer Series on Massive Computing, 2002Kluwer, New York.

[2] Zabir Haiderkan, Tasnim Sharmin Alin, & Md. Akter Hussain., "Price prediction of share market using Artificial Neural Network". International Journal of Computer Applications Vol 22. No2. may2011

[3] Akintola,K.G , Alese,B.K and A.F. Thompson., "Timeseries forecasting with neural network –a case study of stock price of intercontinental bank Nigeria" IJRRAS Dec2011.

[4] Juan Peralta Donate, Xiaodong Li, German Gutierrez Sanchez, & Araceli Sanchis de Miguel., "Time series forecasting by evolving Artificial Neural Networks with Genetic algorithms, Differential evoluation and Estimation of distribution algorithms" Springer-verlag London limited 2011.

[5] Alionue Dieng .,"Alternative Forecasting Techniques and vegetable prices in Senegal" Revue senegalaise des recherches agricoies et agroalimentaires- Vol. 1. No 3,jan-juin 2008.

[6] Koffi N.Amegbeto., "vegetable price analysis and marketing windows for protected agriculture technology adopters in Afghanistan" Protected agriculture project in Afghanistan FEB 2006.

[7] Manish Shukla, Sanjay Jharjharia., "Applicability of ARIMA models in Wholesale vegetable market, An investigation" Preceeding of the 2011 "International Conference on Industrial Engineering and Operations Management" KualaLumour, Malasia, Jan22-24,2011.

[8] Guo Qiang, LUO Chang-shou, WEI Qing-feng .,"prediction and research on vegetable price based on genetic algorithm and Neural network model" Asia Agricultural Research 2011, 3(5):148-150.

[9] Chapgshou Luo, Qingfeng Wei, Liying Zhou, Jungeng Zhang and R. Suien Sun "Prediction of vegetable price based on Neural Network and Genetic Algorithm". IFIP AICT 346, pp. 672-681 © Springer link 2011.

[10] Gowri,T.M and Reddy,V.V.C, 2008. Load Forecasting by a Novel Technique using ANN. ARPN Journal of Engineering and Applied Sciences. 3(2): 19-25.

[11] Nasira,G.M and Hemageetha,N., "Vegetable price prediction using data mining classification technique" Preceedings of the International Conference on pattern Recognition, Informatics and Medical Engineering (PRIME 2012), PP. 99-102 ISBN No:978-1-4673-1038-3. © 2012 IEEE

[12] Jiawei Han and Micheline Kumber., "Data mining Concept and Techniques" ELSEVIR second Edition.

Dr. G M Nasira is working as Assistant Professor, Department of Computer Science, at Government Arts College (Autonomus), Coimbatore. She has around 15 years of experience in teaching at college level in various positions like Lecturer, Assistant Professor and Professor. She has published so far 12 research papers in referred journals and presented 40 papers in various conferences and seminars in addition she has also authored two books titled Fundamentals of Middleware Technologies and Web Technologies. Her research papers have won the Best paper award in two International conferences. She is a Recognized Supervisor to guide Ph.D in the area of Computer Applications for various Universities and also a Member of Board of Examiner to evaluate the Ph.D Thesis of various Universities. At present, 20 research scholars are doing their Ph.D. under her guidance and as many as 38 students have completed their M.Phil thesis under her guidance. She has also guided more than 100 Masters' degree projects so far. She has also acted as a reviewer in the International Conferences and various journals. She has been as a Resource Person for National Workshops/Seminars and Symposiums. Her specialization includes Soft computing, Artificial Neural Networks, Artificial Intelligence, Genetic Algorithms, Data Mining & Data Warehousing , Optimization Techniques. She has excellent track record in the administration of academic institutions in the capacity of Head of the Department and Vice-Principal. She got Best teacher award and her department got Best Department award twice under her headship.

.

N. Hemageetha. Received her MCA degree from University of Madras, Chennai, India in 1996, the M.Phil degree in computer Science from Bharathidasan University, Trichy, India in 2005 and is currently pursuing the PhD degree in the research area of Data Mining under the guidance of Dr. G.M. Nasira. Currently she is working as Assistant Professor in the Department of Computer Science, Govt. Arts College for Women, Salem 636008.