Proceedings of the International Conference on
Pattern Recognition, Informatics and Medical Engineering , March 21-23, 2012

Vegetable Price Prediction Using DataMining Classification Technique

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

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

Abstract— **Each and every sector in this digital world is undergoing a dramatic change due to the influence of IT field. The agricultural sector needs more support for its development in developing countries like India. Price prediction helps the farmers and also Government to make effective decision. Based on the complexity of vegetable price prediction, making use of the characteristics of neural networks such as self-adapt, self-study and high fault tolerance, to build up the model of Back-propagation neural network 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 absolute error percentage of monthly and weekly vegetable price prediction and analyze the 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. Government can also take decision whether to allow or not to export and import of respective commodities. 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. 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. The aim of this paper is to develop Neural Network model to predict the price of tomato in Coimbatore market.

II. 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 layer 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.

978-1-4673-1039-0/12/$31.00 ©2012 IEEE

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

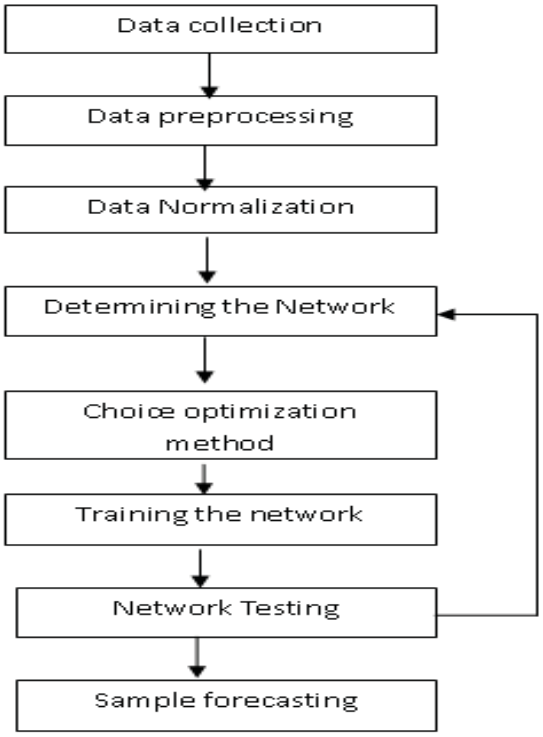

Figure 1. A Flow chart for developing ANN

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 the 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 normally daily data are not available. So, weekly and monthly data are used for forecasting because it has less noise. In this study three years' of tomato price data from 2009 to 2011 are taken for creating the model. Taking previous monthly and weekly price of tomato from Coimbatore market for simulating the model and later few monthly and 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 we dropped it for 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.[5].

Minimax normalization is used in this paper.

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

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

The time series data will be transformed into a data set depending on the x input nodes of a particular ANN and each data set will consist of the following:

- X input values that correspond to x normalized previous values of period t: $N_{t-1}, N_{t-2}, \ldots, N_{t-k}$
- One output value : N_t (designed target)

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.

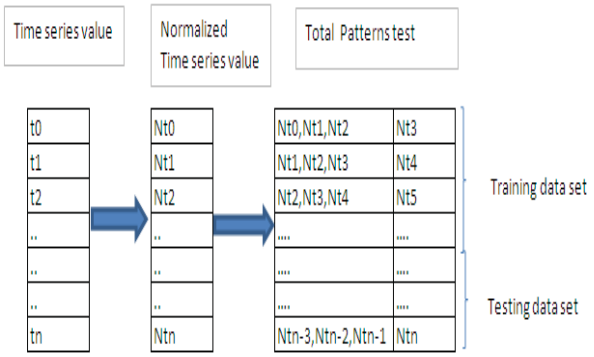

Figure 2. Train and validation dat set

In this paper, former three week data of tomato price are taken as input and produced later one week data as output for weekly price prediction. And former four month data of tomato price are taken as input and later one month data as output for monthly price prediction. So three input neurons for weekly price prediction and four input neurons for monthly price prediction are considered.

C. Model Establishment

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.[8] 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.

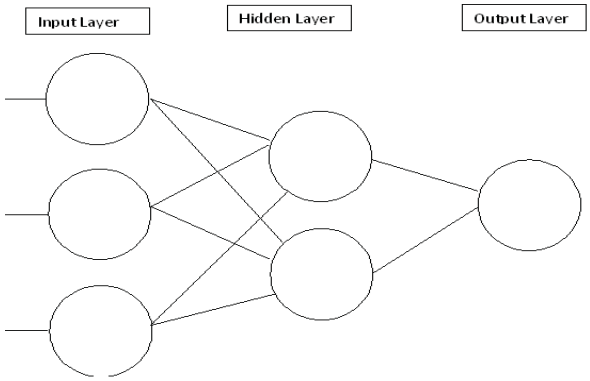

Figure 3. ANN structure for weekly price prediction

Three layer feed forward network structure is used for monthly and 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 a certain period, is the price of input in the successive 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.

III. Experimental Results and Analysis

A. *Monthly Simulation and Prediction Analysis of BPNN*

Taking former four month data of tomato as input and later one month data as output the code is developed using MATLAB. BPNN is constructed using Jan 2009 to may 2011 monthly price data and later month's data are used to test the model. The number of hidden neurons is 5. Optimization target is 0.001. The results are as follows:

TABLE I. Actual value , predicted value and error(%) of monthly report

Month	Actual	Predicted	Error %
1	14	15.0036	7.168571
2	13	13.2024	1.556923
3	13	12.458	4.169231
4	9	9.1115	1.238889
5	13	12.4633	4.128462
6	18	17.4878	2.845556
7	26	25.5212	1.841538

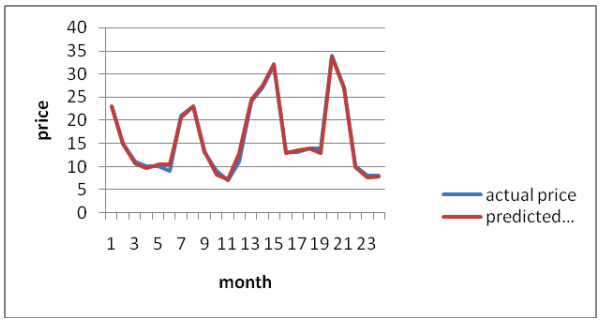

Figure 4. Simulation Result

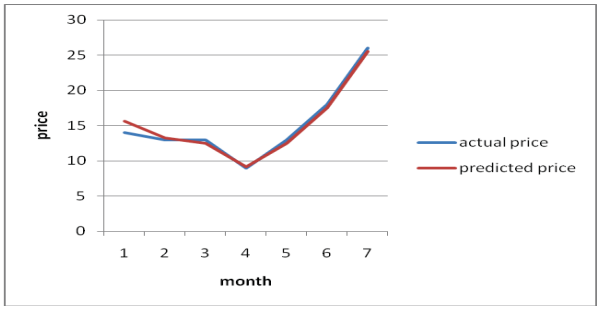

Figure 5. Prediction Result

For calculating the absolute error between measured value and predicted value. We use the error percentage formula,

(Actual price – Predicted price) /Actual price) * 100 (2)

In this model we observed that the absolute error of monthly price prediction is within 10%, So the accuracy is up to 90%.

B. *Weekly Prediction Analysis of BPNN*

Taking three week's data of tomato as input and later one week data as output, we have written program in MATLAB. BPNN is constructed using previous 135 weeks from Jan 2009 price data and later week's data are used to test the model. The number of hidden neurons is 4. Optimization target is 0.001. The results are as follows,

TABLE II. ACTUAL VALUE , PREDICTED VALUE AND ERROR(%) OF WEEKLY REPORT

Week	Actual price	Predicted price	Error %
1	26	20	23.07692
2	31	26	16.12903
3	34	28	17.64706
4	25	24	4
5	18	20	11.11111
6	18	18	0
7	14	17	21.42857
8	11	13	18.18182
9	9	10	11.11111
10	10	10	0

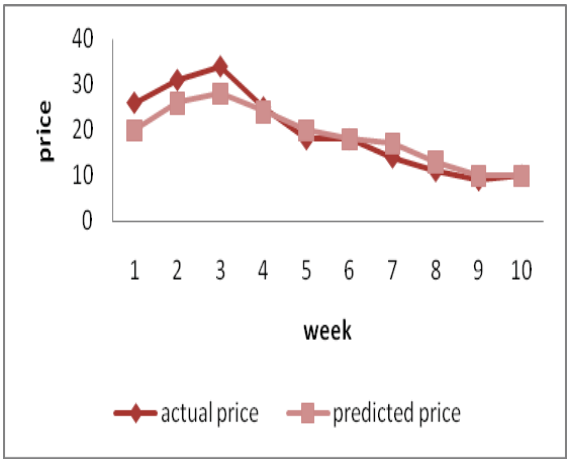

Figure 6. Weekly Price Prediction

Calculate the absolute error between measured value and predicted value. In this model the absolute error of weekly price prediction is within 25%, So the accuracy is up to 75%. Performance of ANN models can be measured by mean relative percentage error. We trained several combinations of number of neurons in hidden layer, learning rate epochs and so on. Time series forecasting nature is need point data to predict future, this network can be used to predict one period ahead.

IV. CONCLUSION

In this paper, the BP neural network prediction model of vegetable market price is established. We have taken three years Coimbatore market price of tomato as an example and simulated the result using MATLAB and predict. The prediction results of monthly and weekly are discussed in this paper. 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 PM, Resende M, Hand book of massive datasets. Vol. 3 of the Kluwer Series on Massive Computing, 2002Kluwer, New York.

[2] 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.

[3] 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

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

[5] T.Jayalskshmi, Dr.A.Santhakumar "Statistical Normalization and Back propogation for classification" International Journal of computing Theory nad Engineering, Vol 3- No1-Feb2011.

[6] 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.

[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] 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.