ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

FORECASTING VEGETABLE PRICE USING TIME SERIES DATA

M.Subhasree[1], Mrs.C.Arun Priya[2].

1. M.Phil Research Scholar, Department of computer science. Psgr krishnammal college for women, Coimbatore, Tamil Nadu.
2. Assistant Professor, Department of computer science, Psgr krishnammal college for women, Coimbatore, Tamil Nadu.

Manuscript Info

Manuscript History:

Received: 14 December 2015
Final Accepted: 19 January 2016
Published Online: February 2016

Key words:
Back propagation neural network, genetic algorithm, radial basis function.

Corresponding Author

M.Subhasree.

Abstract

Predicting the vegetable price 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. This model is used to predict the next day price of vegetable using the previous price of time series data. The three machine learning algorithms are incorporated in this work namely Radial basis function, back propagation neural network and genetic based neural network are compared. 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 and radial basis function and improves the accuracy percentage of vegetable price prediction.

Introduction:-

a. Existing system:-

Vegetable price changes fast and unstable which makes great impact in our daily life. 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. Back Propagation Neural Network 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.

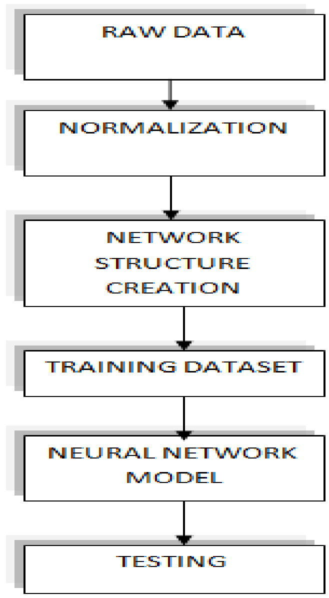
b. Proposed system:-

In the proposed work, vegetable price prediction using time series data is carried out by using genetic based neural network. The proposed work consists of four steps

- Normalization
- Network structure creation
- Training
- Testing

c. 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 work, taken only the most perishable vegetables Tomato, Ladies finger, Broad beans, Small onion and Brinjal as experimental data. Most important point in network design is determining the data size and frequency. This is mostly depends on the final output. Taking previous three weeks daily price of five vegetable is taken for simulating the model and later one week daily price as test data for the model. Here time series data are collected manually from tirupur new bus stand ulavar market.



The data preparation involves analysis of collected time series data and identification of the required features to carry out the research work efficiently. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. The features that derived are weather, lack of availability (LA), over production (OP), transportation cost (TC), supply, demand. These features along with the

Data mining:-

price are considered for further processing. The value for weather feature is assigned as 1 for summer, 2 for winter, 3 for spring, and 4 for autumn. The LA, OP, supply and demand feature values are 1 for yes and 0 for no. If there is over production (1), there will be no lack of availability (0). If there is supply (1), there will be no demand (0). The transportation cost values derived from the diesel price and the transportation distance.

d. 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 here.

Here *X*' is normalized input data, *Xi* is Actual Input, and 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 work 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 NN and each data set will consist of the following:·

Y input values that correspond to y normalized previous values of period t: *N, N... N,*] One output value: *N*

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

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

e. 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. In this work we construct three models of neural network RBFN, BPNN, GANN.

Construction of RBF Neural Network Model:-

Radial basis function network is an artificial neural network that uses radial basis functions as activation functions. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. Radial basis function networks have many uses, including approximation, time, classification, and system control Radial basis networks can be used to approximate functions.

The experiment is carried out using using the daily price data for all the five vegetables. The data is normalised using the max -min normalization. The network structure is created using the function

net = newrb(P,T);

adds neurons to the hidden layer of a radial basis network until it meets the specified mean squared error goal. newrb creates a two-layer network. The first layer has radbas neurons, and calculates its weighted inputs with dist and its net input with netprod. The second layer has purelin neurons, and calculates its weighted input with dotprod and its net inputs with netsum. Both layers have biases. Initially the radbas layer has no neurons. The following steps are repeated until the network's mean squared error falls below goal.

1. The network is simulated.
2. The input vector with the greatest error is found.
3. A radbas neuron is added with weights equal to that vector.
4. The purelin layer weights are redesigned to minimize error.

Construction of BP Neural Network Model:-

Back propagation, an abbreviation for "backward propagation of errors", is a common method of training networks used in conjunction with an optimization method such as gradient descent. The method calculates the gradient of a loss function with respect to all the weights in the network. The gradient is fed to the optimization method which in turn uses it to update the weights, in an attempt to minimize the loss function

Back propagation requires a known, desired output for each input value in order to calculate the loss function gradient. It is therefore usually considered to be a supervised learning method, although it is also used in some unsupervised networks such as auto encoders. It is a generalization of the delta rule to multi-layered feed forward networks, made possible by using the chain rule to iteratively compute gradients for each layer.

Back propagation requires that the activation function used by the artificial neurons be differentiable. Before back-propagation, most networks used non differentiable hard-limiting binary nonlinearities such as step functions and there were no well-known general methods for training multilayer networks. To train a network, it is necessary to have a set of input patterns and corresponding desired outputs, plus an error or cost function that measures the cost of differences between network outputs and desired values. The simplest implementation of standard back-propagation learning updates the network weights and biases in the direction in which the performance function decreases most rapidly.

Construction of Neural Network Model Based on Genetic Algorithm:-

A genetic algorithm tries to stimulate the natural evolution process. Its purpose is to optimize a set of parameters. Each parameter string represents a possible solution to the examined problem. For the GANN problem, it contains information about the contruction of a neural network. The quality of the solution is stored in the fitness value. The main idea of GA is to start with a population of solutions to a problem, and attempt to produce new generations of solutions which are better than the previous ones. GA operates through a simple cycle consisting of the following four stages: initialization, selection, crossover, and mutation.

- **Selection**

The selection of individuals for cross-over and mutation is biased towards good individuals. In the classical fitness-based roulette-wheel, the chance of an individual to be selected is based on its relative fitness in the population.

- **Genetic Operators**

The next step is to generate a second generation population of solutions from those selected through a combination of genetic operators: crossover and mutation. For each new solution to be produced, a pair of "parent" solutions is selected for breeding from the pool selected previously. By producing a child solution using the above methods of crossover and mutation, a new solution is created which typically shares many of the characteristics of its parents.

New parents are selected for each new child, and the process continues until a new population of solutions of appropriate size is generated.

Crossover simulates performed by taking parts of the bit-string of one of the parents and other parts from the other parent and combining both in the child. There three basic kinds of crossover: one-point, two-point and uniform.

In genetic algorithm we have to create a fitness function. This function uses the classify function to measure how well mass spectrometry data is grouped using certain masses. The input argument the Population is a vector of row indices from the mass spectrometry data Y. Classification performance is a linear combination of the error rate and the posteriori probability of the classifier.

- Validation performance has increased more than max_fail times since the last time it decreased (when using validation).

The 70% of the normalized dataset is used for the training phase. The train and sim function are used for training the model.

```
net= train(net,Pseq,Tseq);
Pseq = con2seq(P);
Y = sim(net,Pseq);
```

The 30% of the dataset is used for the training phase. The train and sim function are used for training the model.

```
net= train(net,Pseq,Tseq);
Pseq = con2seq(P);
Y = sim(net,Pseq);
```

f. Training and testing

Training set is the data set used to adjust the weights on the neural network. A set of examples used for learning that is to fit the parameters [i.e., weights] of the classifier. Validation set is the data set used to minimize over fitting. If the accuracy over the training data set increases, but the accuracy over then validation data set stay the same or decreases, then you're over fitting your neural network and you should stop training. A set of examples used to tune the parameters [i.e., architecture, not weights] of a classifier, for example to choose the number of hidden units in a neural network. Testing set is the data set used only for testing the final solution in order to confirm the actual predictive power of the network. A set of examples used only to assess the performance [generalization] of a fully specified classifier.

```
net.trainParam.epochs = prm(1);
net.trainParam.goal = prm(2);
net.trainParam.lr = prm(3);
```

Each epoch, if performance decreases toward the goal, then the learning rate is increased by the factor lr_inc. If performance increases by more than the factor max_perf_inc, the learning rate is adjusted by the factor lr_dec and the change that increased the performance is not made.

Training stops when any of these conditions occurs:

- The maximum number of epochs (repetitions) is reached.
- The maximum amount of time is exceeded.
- Performance is minimized to the goal.
- The performance gradient falls below min_grad.

Training set can be made easily directly from the time series. Certain number of measured values is used as inputs and the value to be predicted (i.e., the value in the future, in some chosen distance after these input measured values) is used as required output. Input part of the time series is called window, the output part is the predicted value. By shifting the window over time series the items of training set are made. It is advised to left part of time series for testing, i.e., to not use this part during learning, but to use it to test how successfully the network learned to predict our data. The training set obtained in this way can be then adjusted for the needs of a particular neural network.

Experimental result:-

Experiment of Radial Basis Function Neural Network:-

Daily prices for five vegetables such as Brinjal, tomato, ladies finger, broad beans, small onion along with their factors weather, lack of availability, over production, transportation, supply and demand are taken as the feature vector inputs. The number of hidden layers is set to 10. The Optimization goal value is fixed to 0.1. The algorithm is code in Matlab. The data of first three weeks in twelve months are used as the training data to train the model. After training, the model predicts the next day price of Fourth week of twelve months. The results are as follows:

Table 5.1 Actual and Predicted Values of Five Vegetables in RBFN

Month	Brinjal		Ladies finger		Tomato		Broad Beans		Small onion	
	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Jan	48	30	36	20	12	10	50	23	26	20
Feb	23	10	17	20	25	10	55	25	21	10
Mar	48	20	28	19	18	11	34	24	30	20
Apr	40	20	18	19	12	14	35	25	25	11
May	40	20	14	16	16	15	45	21	35	25
June	55	30	24	17	24	14	45	25	35	20
July	32	23	24	15	12	18	32	12	36	40
Aug	50	25	18	18	13	15	20	22	40	45
Sep	20	15	12	18	7	10	20	22	30	20
Oct	30	25	39	20	10	11	40	20	20	17
Nov	24	17	27	20	22	12	45	16	24	23
Dec	40	25	29	21	34	18	50	29	35	21

Experiment of back propagation neural network:-

One year daily price data of five vegetables such as Brinjal, tomato, ladies finger, broad beans, small onion along with their factors such as whether, lack of availability, over production, transportation, supply and demand are taken as the feature inputs. The number of hidden layers is set to 10. The Optimization goal value is fixed to 0.1. The algorithm is code in Matlab. The data of first three weeks in twelve months are used as the training data to train the model. After training, the model predicts the next day price of Fourth week of twelve months. The results are as follow.

Table 5.2 Actual and Predicted Values of Five Vegetables in BPNN

Month	Brinjal		Ladies finger		Tomato		Broad Beans		Small onion	
	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Jan	48	40	36	38	12	17	50	43	26	30
Feb	23	19	17	10	25	20	55	45	21	10
Mar	48	45	28	35	18	15	34	23	30	20
Apr	40	35	18	21	12	10	35	23	25	14
May	40	38	14	13	16	15	45	31	35	25
June	55	54	24	20	24	15	45	35	35	20
July	32	25	24	20	12	10	32	20	36	40
Aug	50	40	18	21	13	11	20	12	40	45
Sep	20	18	12	14	7	5	20	12	30	20
Oct	30	34	39	30	10	10	40	30	20	17
Nov	24	25	27	20	22	18	45	35	24	23
Dec	40	44	29	28	34	20	50	40	35	21

Experiment of genetic algorithm based neural network:-

One year daily price data of five vegetables such as Brinjal, tomato, ladies finger, broad beans, small onion along with their factors such as whether, lack of availability, over production, transportation, supply and demand are taken as the feature inputs. In this model, the chromosome group is 30, optimization goal is 0.5. Optimization algebra is 3,000. Its input value and output value are the same as BP neural network. The number of hidden layer is 10. The algorithm is code in Matlab. The data of first three weeks in twelve months are used as the training data to train the model. After training, the model predicts the next day price in Fourth week of twelve months. The results are as follows:

Table 5.3 Actual and Predicted Values of Five Vegetables in GA

Month	Brinjal		Ladies finger		Tomato		Broad Beans		Small onion	
	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Jan	48	49	36	33	12	15	50	53	26	24
Feb	23	29	17	15	25	25	55	55	21	20
Mar	48	45	28	29	18	23	34	33	30	31
Apr	40	40	18	18	12	12	35	33	25	24
May	40	47	14	10	16	13	45	41	35	35
June	55	54	24	23	24	22	45	45	35	33
July	32	31	24	27	12	13	32	30	36	34
Aug	50	40	18	21	13	11	20	21	40	34
Sep	20	18	12	14	7	9	20	19	30	27
Oct	30	32	39	35	10	10	40	20	20	17
Nov	24	22	27	27	22	21	45	45	24	23
Dec	40	38	29	29	34	31	50	51	35	31

Comparative study:-

The performance of the radial basis function, back propagation and genetic algorithm based neural network are shown Figure 4.3.7. The comparative result shows that the predictive accuracy of genetic algorithm is better than the back propagation and radial basis function

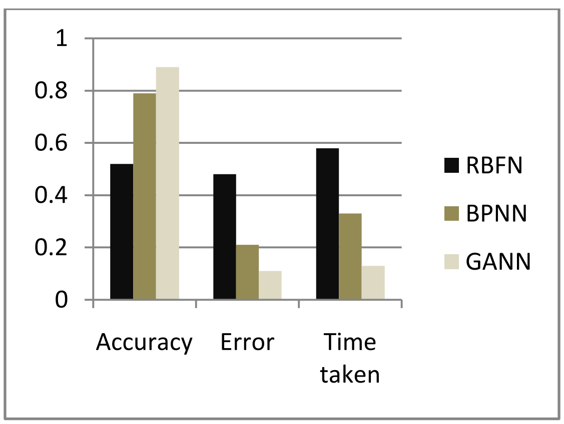


Figure 5.4 Comparative Results of Three Algorthims

The table shows the result of Genetic algorithm based neural network model is equal or better than other three models. And then, the neural network based on back propagation is followed by. The performance of RBFN is worse than the neural network based on back propagation.

Table 4. Comparison of Accuracy, Time Taken and Error Rate Prediction

Algorithm	**RBFN**	**BPNN**	**GANN**
Accuracy	52%	79%	89%
Error Rate	0.48	0.21	0.11
Time taken	0.58	0.33	0.13

Conclusion:-

The vegetable price prediction is essential for common people to recognize the price of vegetable in advance. In this research work, vegetable prices are predicted using Back propagation neural network, Radial basis function and Genetic algorithm based Neural network. Daily price data of vegetables price data and six features are considered to construct model. The genetic algorithm based neural network model utilizes the advantage which is provided by single predicting method that obtains the best prediction accuracy than other models. Further research can be done by adding more vegetables. More influential factors need to be considered. The work can be further extended using other supervised learning techniques to increase the performance accuracy.

References:-

[1] Dieng, A,. "Alternative Forecasting Techniques for Vegetable Prices in Senegal", Revue sénégalais des richerches et agricolesetagroallementalress, 2008 Vol. 1, No. 3, PP 5-10.

[2] 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,.

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

[4] 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.

[5] philippe binard, general delegate of freshfe, "understanding price fluctuations of fresh fruit and vegetables helps consumers in qualified shopping decisions", European Fresh Produce Association, 8 march 2011.

[6] M. M. Gupta, L. Jin, and N. Homma, Static and Dynamic Neural Networks, New York: Wiley, 2003 [19] K. I. Funahashi, "On the approximate realization of continuous mappings by neural networks", Neural Netw, 1989, Vol. 2, PP 183–192.

[7] ZabirHaider Khan, TasnimSharminAlin 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.

[8] BhagwantChauhan, UmeshBidave, AjitGangathade, Sachin Kale," Stock Market Prediction Using Artificial Neural Networks", International Journal of Computer Science and Information Technologies", 2014, Vol. 5, No.1, PP 904-907.

[9] AlionueDieng, "Alternative Forecasting Techniques And Vegetable Prices In Senegal" Revue senegalaise des recherché agroalimentaires, 2008, Vol. 1, No. 3, PP 6-10.

[10] ChangshouLuo, "Prediction of vegetable price based on Neural Network and genetic algorithm", International Federation for Information Processing, 2011, Vol. 102, PP 672-681.

[11] 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.

[12] 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.

[13] Celso Correia de Souza1*, José Francisco dos Reis Neto1, Daniel Massen Frainer1, Francisco de Assis Rolim Pereira1 and Rafael Gabriel "Forecast of prices for horticultural products with the use of artificial neural networks" African Journal of Agricultural Research,2015, Vol. 10, No. 30, PP 2919-2927.

[14] Md. SyedulAmin,Md. Mamun, FazidaHanimHashim,JubayerJalil 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.

[15] Ozgur Kisi "Streamflow Forecasting Using Different Artificial Neural Network Algorithms" Journal of Hydrologic Engineering, 2007, Vol. 12, Issue. 5, PP 532-539.

[16] K.K.Sureshkumar ,Dr.N.M.Elango "Performance Analysis of Stock Price Predictionusing Artificial Neual Network", Global Journal of Computer Science and Technology, 2012, Vol.12 No. 1, PP 19-25.