Modeling for Prediction of Tomato Yield and Its Deviation using Artificial Neural Network

Syed Abou Iltaf Hussain*1, Diganta Hatibaruah*2

1ME student, Department of Mechanical Engineering, Jorhat Engineering College, Jorhat, Assam, India
2Associate Professor, Department of Mechanical Engineering, Jorhat Engineering College, Jorhat, Assam, India

I. INTRODUCTION

Prediction is the way things will occur in the future. Prediction is a difficult job especially in agriculture predicting the crop yield. As the yield is dependent upon various factors varying from weather to the amount of fertilizers required. Out of the various factors some factor changes from day to day like the temperature of the day, rainfall, sunshine etc. On the other hand some factors remains constant like the pH of soil but some factors like the amount of fertilizers used depend upon the will of farmer how much they are applying.

According to FAOSTAT, India is the second largest producer of tomatoes producing approximately 12 million tonnes in 2010 but the value rose to 17.5 million metric tonnes of tomatoes annually in 2014. Out of all the states growing tomatoes, Andhra Pradesh leads the tomato growth in India covering approximately 35% of the total production followed by Karnataka. Assam is one of the lowest tomato growing states of the country producing approximately 402.49 thousand tonnes in 2013.

Basically Tomato cultivators of Phesual, Assam are not economically strong so they cannot invest large amount of money in their farming. Moreover the farmers of Assam have small land holdings. Their main motive is to increase their production by putting minimum efforts, involving less labours and investing minimum money. Hence a need for a predicting tool arises that could predict or give a rough idea about the amount of the tomato yield at the end of the season so that a decision could be made weather to cultivate or not. Artificial Neural Network is one of the many computing models that could predict the live in situation accurately.

II. LITERATURE SURVEY

In 2005 Kaul et al used ANN to develop model to predict the yield of corn and soybeans by considering various environmental factors as inputs. In 2006 Ushadaa and Murase used artificial neural network to develop a model that showed relationship of minimum temperature \( T_{\text{min}} \), maximum temperature \( T_{\text{max}} \), optimum temperature \( T_{\text{opt}} \) and ambient temperature \( T_{\text{amb}} \) with the as heat unit accumulation, relative rate of growth, leaf area index, height of moss, mass of moss and temperature stress factor. The specific leaf area and the ground area had the best experimental values of 1.498 m\(^2\)/plant respectively in the leaf area index sub-model. In 2007, B. J I et al have developed a better model than the multiple linear regression-based yield models to predict the Fujian rice yield of the Fujian province of China from the location-specific rainfall data, soil fertility data and the weather variables such as sunshine hours per day, solar radiation per day and temperature sum per day. The values of \( R^2 \) and RMSE are comparatively higher and lower respectively in case of ANN rice yield model than multiple linear regression-based yield models. In 2010, Rahman and Balamodeled a network to predict the jute production from i) Julian day, ii) solar radiation, iii) maximum temperature, iv) minimum temperature, v) rainfall and vi) type of biomass. Jiří ŠTASTNÝ et al in 2011 has predicted the crop yield level using artificial neural network. The input values were density of nurslings per meter square and average onion yield. The model was less complex than the other existing models and higher accuracy so that the model is easy to use and the prediction is more accurate. In 2014, SaisuneeJabjone and SuraWannasang developed a model that could predict the rice production of Phimai district, Thailand from the technique used for irrigation, rice breed, region, field area and characteristics, cultivation technique and damage area.

The output from the model was compared stepwise with the linear regression models and the result obtained from the neural network was better than the linear regression method.

III. OBJECTIVE

The main objective of this paper is to develop a model to predict tomato yield and its deviation from the maximum possible amount of production from the amount of fertilizers used by the farmers, pH of soil and land available for tomato cultivation for each farmer.

IV. METHODOLOGY

Data collected were based on the interviewed survey of random tomato cultivators from the Phesual region of Jorhat district.

A. Artificial Neural Network

Artificial Neural Networks are a family of statistical learning models inspired by the central nervous system of animals particularly brain. The brain learns from the past experience. Brain is a complex network of neurons which process signals as
received by the sensory organs and asked to react according to the situation. Similarly artificial neural networks are a system of interconnected neurons which exchange messages with each other. There are some numeric weights at the connections of the neurons. These numeric weights can be readjusted through training. Training is the process of learning new jobs by repeatedly doing a particular job. When an artificial neural network model is trained the predicted output obtained is compared with the actual output and the numeric weights are updated. When a particular artificial neural model is trained repeatedly the numeric weights are updated until the predicted output and actual output are similar or the error between the two is the least.

The fundamental unit of neural network is known as neuron. Artificial neural network are represented by a set of interconnected neurons which exchanges messages between each other. The artificial neuron receives one or more inputs and sums them to produce an output. The inputs of each node are summed and then weighted. The sum is passed through an activation function or transfer function which is a non-linear function. Mathematically, for neuron k:

\[ u_k = \sum_{j=1}^{m} (w_{kj} x_j) \]  

and 

\[ y_k = \varphi(u_k + b_k) \]

where \( x_j \) are input signals and \( w_{kj} \) are synaptic weights of neuron k, \( u_k \) is the linear combiner output due to input signals, \( b_k \) is the bias, \( \varphi(\cdot) \) is the activation function and output signal of the neuron is \( y_k \). The use of bias \( b_k \) has the effect of applying an affine transformation to the output \( u_k \) of the linear combiner in the model of Figure 1 as shown by the following equation \( v_k = u_k + b_k \).

With respect to the weights in the network the backpropagation methods calculates the gradient of loss function. 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.

**Feedforward backpropagation algorithm**

A feedforward neural network consists of three layers. The first layer is the input layer, second or the middle layer consists of one or more hidden layers and the third is the output layer. Each of the input layer, hidden layer and the output layer consists of a number of neurons. Every neuron of the input layer is connected to the neurons of the first hidden layer and every neuron of first hidden layer is connected to the neurons of the second hidden layer and so on. The neurons of the last hidden layer are connected to the neurons of the output layer.

**Cascade-forward backpropagation algorithm**

A cascade-forward neural network consists of three layers. The first layer is the input layer, second or the middle layer consists of one or more hidden layers and the third is the output layer. Each of the input layer, hidden layer and the output layer consists of a set of neurons. Every neurons of the input layer is connected to each neuron of the hidden layer and the output layer. Every neuron of first hidden layer is connected to the neurons of the second hidden layer and so on. The neurons of the last hidden layer are connected to neurons of the output layer.

**Backpropagation algorithm**

Backpropagation is a form of supervised learning. When using a supervised learning method, the network is provided with both sample inputs and predicted outputs. The predicted outputs are compared against the actual outputs for given input.
1. Input layer  
2. Hidden layer  
3. Output layer  

Figure 3: Typical cascade-forward neural network  

B. Preprocessing of data  

It is one of the many steps in data mining. Data pre-processing is done before the actual processing of data. It prepares the raw data for further processing.  

Data Normalization  

Data pre-processing is also known as Data Normalization. The reason for using feature scaling method is that the gradient descent converges faster. Mathematically, the normalized data is given by:

\[ x' = \frac{x - \text{min}(x)}{\text{max}(x) - \text{min}(x)} \]

where \( x \) is the input data of a parameter, \( \text{min}(x) \) is the minimum input data of the parameter, \( \text{max}(x) \) is the maximum input data of the parameter.

Deviation of tomato yield  

The tomato cultivators are cultivating the variety Avinash-2. The main characteristic of this variety is high yield under controlled conditions. Yield of this variety is about 1200 quintal per hectare. Deviation of tomato yield = 1200*(\( L_i \)) – \( y_i \)

Where \( (L_i) \) is land available for tomato cultivation with each farmer  

\( y_i \) is the tomato yield from that particular land  

V. ANALYSIS AND DISCUSSIONS  

Data were collected through an interview survey amongst the professional tomato cultivators of the Phesual region of Jorhat district. The collected data were normalized using feature scaling method so that the network converges faster.

The ANN network consists of 3 layers. The first layer is the input layer. It consists of five neurons viz. i) pH of soil, ii) amount of superphosphate used, iii) amount of potash used, iv) amount of urea used and v) land available for tomato cultivation. The second layer is the hidden layer. It consists of single hidden layer. The hidden layer consists of 10 neurons. The final layer is the output layer and it consists of two output neurons viz. i) tomato yield and ii) deviation of tomato yield from the maximum. Table-I and Table-II represents the total predicted tomato yield and its deviation obtained from the feed-forward neural network and cascade-forward neural network respectively.

| Table-I | Predicted Tomato Yield |
|---|---|---|---|---|---|
| Farmer | Actual yield | Predicted yield | Farmer | Actual yield | Predicted yield |
| | Feed-forward | Cascade-forward | | Feed-forward | Cascade-forward |
| 1 | 45 | 46.48 | 51.88 | 26 | 52.5 | 52.39 | 51.46 |
| 2 | 200 | 179.77 | 185.83 | 27 | 126 | 107.53 | 114.82 |
| 3 | 190 | 178.36 | 169.90 | 28 | 44 | 34.62 | 33.87 |
| 4 | 26.25 | 31.77 | 33.13 | 29 | 44 | 34.64 | 32.84 |
| 5 | 90 | 95.22 | 88.66 | 30 | 150 | 162.00 | 169.36 |
| 6 | 227.5 | 228.77 | 219.79 | 31 | 75 | 67.05 | 74.76 |
| 7 | 70 | 66.93 | 65.86 | 32 | 10 | 33.95 | 30.79 |
| 8 | 90 | 105.03 | 111.72 | 33 | 26.25 | 31.72 | 32.82 |
| 9 | 85 | 82.25 | 86.80 | 34 | 25 | 34.37 | 30.79 |
| 10 | 105 | 107.53 | 114.82 | 35 | 65 | 62.83 | 62.17 |
| 11 | 112.5 | 113.21 | 115.78 | 36 | 26.25 | 32.73 | 21.45 |
| 12 | 165 | 164.65 | 154.61 | 37 | 175 | 162.00 | 169.36 |
| 13 | 27.5 | 31.42 | 25.83 | 38 | 225 | 233.58 | 238.42 |
| 14 | 37.5 | 32.20 | 34.31 | 39 | 165 | 192.79 | 181.82 |
| 15 | 40 | 34.67 | 41.98 | 40 | 50 | 41.81 | 42.65 |
| 16 | 78.75 | 76.57 | 75.52 | 41 | 61.25 | 68.85 | 66.96 |
| 17 | 100 | 97.33 | 109.18 | 42 | 37.5 | 60.22 | 68.58 |
| 18 | 82.5 | 68.68 | 76.09 | 43 | 80 | 86.22 | 84.64 |
| 19 | 35 | 33.92 | 27.36 | 44 | 30 | 32.47 | 28.33 |
| 20 | 25 | 36.04 | 44.80 | 45 | 255 | 254.74 | 234.19 |
| 21 | 90 | 91.83 | 87.26 | 46 | 45 | 34.38 | 37.89 |
| 22 | 325 | 193.64 | 317.50 | 47 | 68.75 | 48.51 | 60.95 |
| 23 | 162.5 | 185.28 | 190.76 | 48 | 105 | 101.07 | 95.16 |
| 24 | 157.5 | 157.67 | 160.63 | 49 | 78.75 | 74.91 | 74.01 |
| 25 | 127.5 | 106.28 | 113.25 | 50 | 225 | 194.95 | 183.65 |

ISSN: 2231-5381  http://www.ijettjournal.org  Page 104
A scatter plot diagram is plotted for tomato yield by taking actual values along the abscissa and predicted along the ordinate. The equation of best fit is given by:

A. For feed-forward neural network:

\[ y = 0.8476x + 11.76 \]

Where \( y \) is the predicted tomato yield
\( x \) is the actual tomato yield

From the scatter plot diagram following points are observed:
1. The predicted output showed a linear relation with the actual output.
2. The coefficient of determination for the line of best fit is 0.911 i.e. the predicted output from the model is having 91.1% accuracy.

B. For cascade-forward neural network:

\[ y = 0.9476x + 4.6771 \]

where \( y \) is predicted tomato yield
\( x \) is actual tomato yield

From the scatter plot diagram following points are observed:
1. The predicted output showed a linear relation with the actual output.
2. The co-efficient of determination for the line of best fit is 0.9685. Hence we can conclude that the predicted output from the model is having 96.85% accuracy.
R² = 0.985

A scatter plot diagram is plotted for deviation of tomato yield by taking actual values along the abscissa and predicted along the ordinate. The equation of best fit is given by:

a. For feed-forward neural network:

\[ y = 0.8083x + 67.813 \]

Where \( y \) is the predicted deviation of tomato yield
\( x \) is the actual deviation of tomato yield.

![Predicted vs Actual](http://www.ijettjournal.org)

Figure 6: Line of best fit between the actual and predicted deviation of tomato yield from the feed-forward backpropagation algorithm

From the scatter plot diagram following points are observed:
1. The actual output is showing linear relation with the predicted output.
2. As the co-efficient of determination for the line of best fit is 0.9341. Hence we can conclude that the predicted output from the model is having 93.41% accuracy.

b. For cascade-forward neural network:

\[ y = 0.9835x + 7.2749 \]

Where \( y \) is the predicted deviation of tomato yield
\( x \) is the actual deviation of tomato yield

\[ R^2 = 0.934 \]

The predicted tomato yield and its deviation showed 98.53% accuracy.

In case of cascade-forward neural network following points are observed:
1. The predicted tomato yield and its deviation are varying linearly with the actual tomato yield and its deviation respectively.
2. From the scatter plot of tomato yield it was found that the predicted value of tomato yield showed 93.41% accuracy and the predicted deviation showed 98.53% accuracy.

Cascade-forward neural network can predict the tomato yield and its deviation more accurately than the feed-forward neural network as in case of cascade-forward neural network the input layer is directly linked with the output layer. As a result of which the synaptic weights could be better updated this is not possible in case of feed-forward neural network. Hence it can be concluded that Cascade-forward neural network of Artificial Neural Network can be used efficiently and effectively for prediction of tomato yield and its deviation.

**VI. CONCLUSION**

The purpose of the study was to model a network that would predict the tomato yield and its deviation in a specific region. Two different networks were created where one was feed-forward and the other was cascade-forward neural networks. Both the neural networks were trained and learned by backpropagation algorithm.

In case of feed-forward neural network following points are observed:
1. The predicted tomato yield and its deviation are varying linearly with the actual tomato yield and its deviation respectively.
2. From the scatter plot of tomato yield it was found that the predicted value of tomato yield showed 91.1% accuracy and the predicted deviation showed 96.85% accuracy.

**REFERENCES**


