© IJCIRAS | ISSN (O) - 2581-5334
August 2019 | Vol. 2 Issue. 3

SOIL TYPE CLASSIFICATION BASED ON NEURAL NETWORK

Khin May Yee[1], Thu Zar Aung[2], Thida San[3]

[1]Faculty of Computer Systems and Technologies, Computer University, Myitkyina and 1011, Myanmar
[2]Information Technology Dept, Government Technological High School, Meiktila and 05181, Myanmar
[3]Information Technology Support and Maintenance Dept, University of Computer Studies, Meiktila and 05181, Myanmar

Abstract

This paper is aimed to implement classification of soil type by using neural network. For soil type classification, there are a great variety of different abundant algorithms for training neural networks. Backpropagation (BP) is the most well-known procedure to handle for supervised training multilayer neural networks. In this paper, Backpropagation (BP) algorithm is implemented for the training of multilayer neural networks employing in soil type recognition system. This implementation is worked by neural network architecture is a totally linked three layer network. The network can train over 7 soil type since the 5-element output vector is used as output units.

Keyword: Artificial Neural Networks (ANNS), Data Mining, Soil Type

1.INTRODUCTION

The soil, the ordinary substance that provisions entirely constructions and investing such as dams, buildings, roads, bridges and tunnels. To determine variable physical and chemical characteristic of soil, which can differ from place to place and from layer to layer even within the limits of the projected construction is aided by soil test. A failed dam is not only a waste of effort; it loses future income and security. For profound base to confirm the behavior bulk of the soil, soil test is necessary to support the loads from profound substance [2]. Soil type classification is not easy regularly to determine because the features of soil are too much. According to this condition, it is difficult to produce the correct outputs. In data mining techniques, Artificial Neural Network is the greatest technique for solving soil type classification problem. Dam construction is a crucial practice in several agriculture and irrigation systems and also useful water applications and management are key concerns. The productivity and consistency of irrigation can be preserved from the verities of information based systems by taking into account weather, water and soil type [4].

An artificial neural network (ANN) has been carried by a precondition of intelligent system to develop a new technology which offers varied solution for the complicated difficulties in construction researches. Instead of linear system, ANN can solve many problems and becomes commonly used type of ANN, which is the multilayer feed-forward network [8]. The basic principle of ANN architecture, application of ANN in classifying soil type by using various types of soil performance factors as the input critical particularly in inventing and evolving better goods for society. In this proposed system is offered procedures for choosing ANN method and future expansion and current trends to classify soil.

In this system, ANN receives in eleven input parameters, processes the soil data and then gives five types of soil. The data of soil test results for dams have been collected from the Assistant Director Office (Soil, Concrete Laboratory & Construction Quality Control), Irrigation Department (Patheingyi).

The proposed system is analyzed with five sections. In section 1, introduction is described. In section 2, background theory is explained. Sample dataset of soil test are described in section 3. In section 4, test and result are implemented. Finally, in section 5, the paper has been concluded.

2.BACKGROUND THEORY

Artificial Neural networks (ANNs) methods in data mining have enhance a very significant method for a wide diversity of functions such as pattern classification, character recognition system, prediction system etc [1]. ANNs are models of biological neural structures. Basic neural network is comprised of three layers: an input layer, at least one hidden layer and an output layer. Each node from input layer is linked to a node from hidden layer [3].

ANN operation consists of training and testing phases. In training phase, feedforward is used to calculate output and backpropagation is used to check error and adjust weights. In testing phase, the weights and biases values from the trained network are applied [3]. Backpropagation in testing is not done as in training phase. Inputs of neural network need to be normalized. Backprogagation is a Neural Network Algorithm for classification that employs a method of gradient descent. The interconnection weights and threshold values in each neuron is adjusted to minimize the mean square distance between the network's class prediction and the actual class label of data tuples. When all Δw_ij in the preceding period are so small as to be under some listed threshold, the training is stopped. When an portion of the neural network miscarries, it can remain deprived of several problem by their equivalent nature. Artificial Neural networks are able to estimated multiplex non-linear plottings and actual adaptable with regard to unfinished, missing and noisy facts [5].

The input parameters of Soil Type Classification for Dams are Clay, Silt, Sand, Gravel, Liquid Limit (L.L), Plastic Limit (P.L), Plastic Index (P.I), Specific Gravity (Sp: Gr), Optimum Moisture Content (O.M.C), Maximum Dry Density (M.D.D) and ECN and output parameters are Clayey Gravels (GC), Clayey Sands (SC), Lean Clay (CL), Silt (ML), and Elastic Silts (MH). The structure of multilayer neural network is shown in Figure 1.

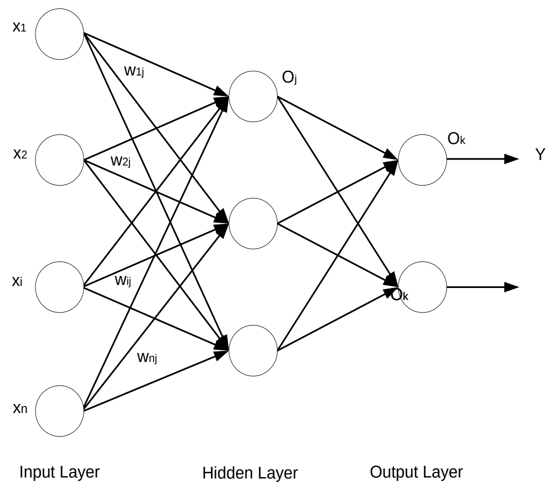

Figure 1. Backpropagation Neural Network with one hidden layer

2.1. Calculation Steps with Backpropagation Algorithm

Backpropagation: Neural network learning for classification or prediction, using the backpropagation algorithm [1].

- Input
  - D defines as a dataset of the guiding tuples and their related values;
  - l means the learning rate
  - Momentum
  - Network is a multilayer feed-forward network
- Initialize the weights
  - Select initial weights in the networks from small random numbers.
- Initialize bias
  - Each unit has a bias related with it.
  - Small random numbers initialize the bias.
- Data Normalization
  - All values of attributes in the dataset has to be changed to contain values in the interval [0,1], or [-1, 1] normalize all data.
  - Data normalization has many methods such as min-max normalization, z-score normalization, and normalization by decimal scaling [6]. A is a numeric element with n observed values v_1 ,v_2, ….v_n.

- Performs a linear transformation on the original data using Min-max normalization as follow:

$$v' = \frac{V-min_A}{max_A-min_A}(new\_max_A - new\_min_A) + new\_min_A$$

where v = original value

v′= normalized value

min_A= minimum value of attribute A

max_A= maximum value of attribute A

❖ Circulate the inputs onward

- The training tuple is fed to the input layer of the network.
- For an input unit, j, its output, O_j, is equal to its input value, I_j.
- To calculate the net input to the unit j, each input linked to the units is increased by its equivalent weights in order to form a weighted sum, which is enhanced to the bias allied with unit j.
- Each unit j in a hidden or output layer, the net input I_j to unit j is

$$I_j = \sum_i w_{ij} O_i + \theta_j$$

where w_{ij} = the weight of linking from unit i in the preceding layer to unit j

$I_j =$ sum of the weighted input

O_i = output of unit i from the previous layer

θ_j = bias of the unit

- The activation function, sigmoid function is applied to the net input of the output layer.
- For the net input I_j to unit j, then O_j, the output of unit is calculated as the following equations:

$$O_j = \frac{1}{1+e^{-I_j}}$$

- The activation function, hyperbolic tangent function which is used for the net input of hidden layer.
- For the net input I_j to unit j, then O_j, the output of unit is computed as

$$O_j = \tanh(I_j)$$

- Compute the output values, O_j, for each hidden layer, up to and including the output layer, which gives the network's prediction.

❖ Backpropagate the error

- To reproduce the error of the network's prediction, the error is circulated rearward by updating the weights and biases.
- For a unit j in the output layer, the error Err_j is computed by

$$Err_j = O_j(1 - O_j)(T_j\text{-}O_j)$$

where $Err_j =$ error

O_j = actual output

T_j = known target value

- The error of the hidden layer unit j is

$$Err_j = O_j(1 - O_j).\sum(Err_k.W_{jk})$$

where Err_k = error of unit k

w_{jk} = the weight of the association from element j to a unit k in the next layer

- The weight and biases are updated to reflect the propagated errors.
- The following equations restructure biases

$\Delta\theta_j$ = (l) Err_j

$\theta_j = \theta_j + \Delta\theta_j$

- The following equations restructure weights:

Δw_{ij} = (l) $Err_j O_i * \text{momentum}$

$w_{ij} = \Delta w_{ij} + w_{ij}$

Where, Δw_{ij} = the change in weight w_{ij}.

❖ Terminating condition: Training stops when

- The percentage of tuples misclassified in the preceding period is below some threshold, or

- A pre identified number of period have ended.

Output: A trained neural network

2.2. Estimating Classifier Accuracy

Evaluating the performance of learning algorithms is a primary feature of machine learning. Estimating classifier correctness is significant in that it lets one to evaluate how exactly a given classifier will label future data, that is, data on which the classifier has not been trained and also help in the comparison different classifiers. An estimate of classifier accuracy on new instances is the most common performance evaluation criterion, although others based on information theory have been suggested. The accuracy of each interval is the percentage of documents correctly classified out of the number of documents in that interval.
Using training data to derive a classifier and then to estimate the accuracy of the resulting learned classifier can result in misleading overoptimistic estimates due to overspecialization of the learning algorithm to the data. The accuracy of a classifier on a specified test set is the percentage of test set tuples that are correctly classified by the classifier.

$$\text{Accuracy} = \frac{\text{No.of correct documents}}{\text{No.of total documents}}$$

The accuracy of each interval is the percentage of documents correctly classified out of the number of documents in that interval. The bootstrap, holdout, random-sub sampling and cross-validation are common techniques for accessing accuracy based on randomly sampled partitions of the given data.

3. SAMPLE DATASET OF SOIL TEST

In this work, data are collected the concrete laboratory and construction quality control of dams from the soil test results. The spectral analysis was sufficiently sensitive to capture the variation in soil fertility between the different soil natures. The soil test dataset which contains the physical and chemical characteristics of soil are shown in Table 1.

No.	Grain Size Distribution				Atterberg Limit			Sp: Gr	Standard Proctor Compaction		E C N	Soil Type
	Clay %	Slit %	Sand%	Gravel%	L.L %	P.L %	P.I %		O.M.C %	M.D.D PcF		
1	10.5	55.0	17.5	17.0	37.9	25.9	12.0	2.66	14.6	115.6	5	ML
2	13.5	67.5	16.5	2.5	42.0	24.7	17.3	2.64	24.6	101.2	5	CL
3	10.0	25.5	16.5	48.0	42.1	25.3	16.8	2.67	15.3	110.9	6	GC
4	10.5	38.0	31.5	20.0	43.0	26.4	16.6	2.66	19.0	108.0	5	SC
5	8.5	36.0	22.0	33.5	38.9	22.5	16.4	2.66	18.2	110.0	5	GC
6	15.5	65.5	14.0	5.0	40.7	25.4	15.3	2.65	20.8	105.8	5	CL
7	8.0	32.0	52.5	7.5	36.2	23.5	12.7	2.66	16.2	111.3	6	SC
8	6.0	21.0	28.5	44.5	33.2	23.2	10.0	2.66	16.3	110.0	5	GC
9	10.0	41.9	46.4	1.7	46.5	28.6	17.9	2.68	20.4	109.6	5	ML
10	27.0	66.0	6.0	1.0	52.2	30.6	21.6	2.67	27.6	94.0	5	MH

Table 1: Soil Test Dataset Sample

4. SYSTEM FLOW DIAGRAM AND RESULTS

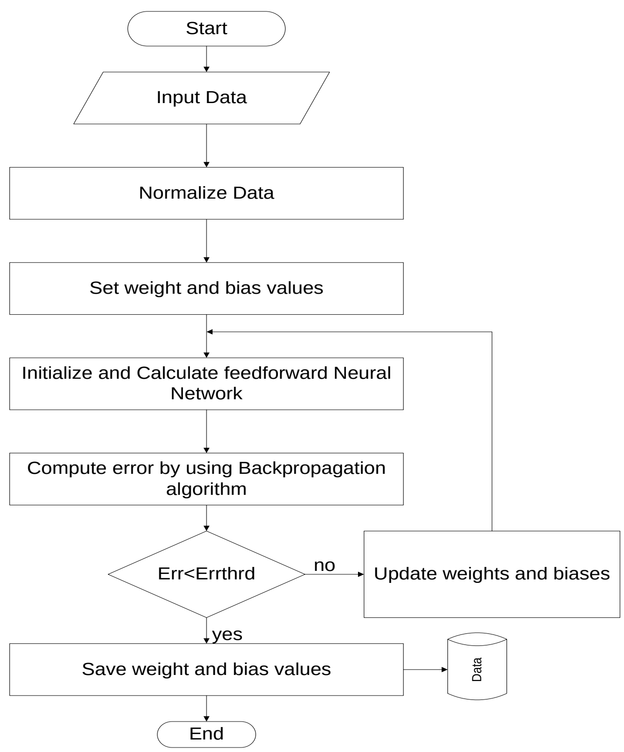

Figure 2. System Flow Diagram of Training Phase

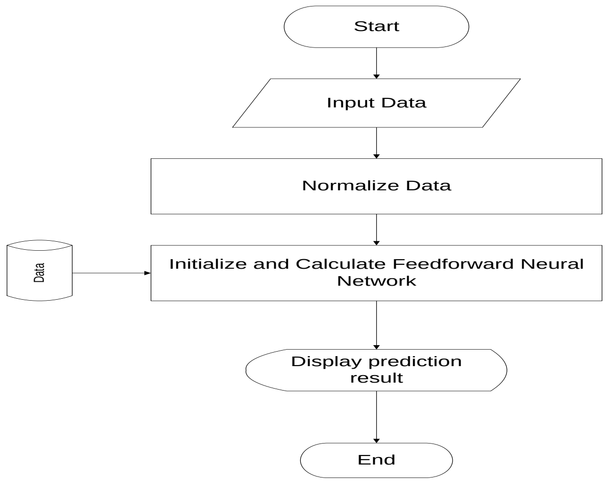

Figure 3. System Flow Diagram of Testing Phase

The functions of the system are load data, insert, update, delete data, normalize data, train data, accuracy test and get soil type. In testing phase, final weight and bias values stored in database are used. Inputs are accepted and then preprocessed. Then inputs are manipulated with final weight and bias values using feedforward neural network and displays a result. If users want to get suitable soil type to build in their construction, users must enter eleven attributes in textbox. When the users enter data, the system calculates the user given with final weight and bias value obtained from training by using Feedforward Neural Network. And then, the system displays the soil type result to user.

5. CONCLUSIONS

Soil type classification system is implemented using Artificial Neural Network. The system helps geotechnical engineers with general guidance about engineering properties of the soils through the accumulated experience. According to the system flow diagram, the system gives the verified result as given by ANN test and the result is satisfactory. The system can classify soil type based on various parameters. This paper is useful for geotechnical engineers who are economically and cannot afford the lab soil test. The objectives of the system are to give the effective decisions for classification of soil type for dams.

A neural network must be trained on certain input data. The two main difficulties in applying this training considered in the following sectors are:

- Stating the set of input to be operated (the learning environment)
- Determining on an algorithm to train the network

Training a network includes performing input forms in a way so that the system reduces its error and increases its performance. The training algorithm may differ reliant on the network architecture, but the most common training algorithm used when manipulative financial neural networks is the Backpropagation algorithm.

Neural networks have several advantages and disadvantages. The benefits of neural networks are:

- generalization ability and robustness,
- mapping of input/output,
- no assumptions of model has to be made and
- flexibility.

The drawbacks of neural networks are:

- Black-box property,
- over fitting,
- expertise for choice of input and
- training takes a lot of time.

Some of the limitations of ANNs are:

- NNs need very huge number of previous events
- The finest network architecture (topology) is still unspecified
- For more complicated network, reliability of results may decrease
- Statistical relevance of the results is needed and
- A more carefully data design is needed.

Default problems like financial market prediction can "actuate memorization" of idiosyncratic patterns in the training data that will not be of help in out-of-sample data. If a network cannot minimize error by learning significant relationships between input variables and the output variables, it tends to do so by memorizing trivial relationships.

Too much of unnecessary information not only extends the learning time, the output will possibly hurt. Neural networks cannot be used to authorize ex-post the uniqueness of causative reasons. Also, ex-ante identification of factors does not provide strong grounds to assert causality even when the users have a good empirical fit. Few statistical concepts have been applied in the development of neural network.

6. ACKNOWLEDGEMENT

The authors are very grateful to the editors and reviewers for their valuable comments and suggestions. This work was supported by the Assistant Director Office (Soil, Concrete Laboratory & Construction Quality Control), Ministry of Agriculture and Irrigation. Other contributors for developing and maintaining the papers which have been used in the preparation of this template.

REFERENCES

[1] J. Han and M. Kamber, "Data Mining (Second Edition) Concepts and Techniques".

[2] Standard Practice for Classification of Soils for Engineering Purposes "Unified Soil Classification System", Designation: D 2487 – 06.

[3] Dr. Yashpal Singh, A. Singh Chauhan, "Neural Networks in data mining", Journal of Theoretical and Applied Information Technology.

[4] S. Kumar Lenka, Ambarish G. Mohapatra "Gradient Decent with Momentum Based Neural Network Pattern Classification for the Prediction of Soil Moisture Content in Precision Argriculture", DOI10.1109/iNIS.2015.56 IEEE

[5] S. Khairunniza-Bejo, S. Mustaffha and Wan Ishak Wan Ismail "Application of Artificial Neural Network in Predicting Crop Yield: A Review", Journal of Food Science and Engineering 4 (2014). 1-9, Department of Biological and Agricultural Engineering, Faculty of Engineering, University Putra Malaysia, Serdang43400, Selangor, Malaysia

[6] V. Rajeswari* and K. Arunesh, "Analysing Soil Data using Data Mining Classification Techniques", Indian Journal of Science and Technology, ISSN(Print): 0974-6846, ISSN (Online): 0974-5645, Vol 9(19), DOI: 10.17485/ ijst / 2016/v9i19/93873, May 2016.

[7] W. Pantanahiran, S. Kriengkraipetch, "The Classification of Suitable Areas for Agriculture in Sakaeo Province, Thailand", 2015 IEEE.

[8] Y. Eindray, "Crop Suggestion System using Artificial Neural Network", University of Computer Studies, Mandalay, December 2016.