Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2320-7051.6405

ISSN: 2320 – 7051

Int. J. Pure App. Biosci. **6 (2):** 1408-1414 (2018)

Research Article

International Journal of Pure & Applied Bioscience

OPEN ACCESS

Multilayer Perceptron Method of Artificial Neural Network for Classification of Farmers Based on Adoption of Drought Coping Mechanisms

Halagundegowda G. R.* **and Abhishek Singh**

Section of Agricultural Statistics, Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi-221 005, India

*Corresponding Author E-mail: hgowda8127@gmail.com

Received: 12.03.2018 | Revised: 18.04.2018 | Accepted: 25.04.2018

ABSTRACT

The study was carried out to develop the machine learning algorithm to predict and classify the farmers into adopters and non-adopters in Kolar district of Karnataka for the year 2013. Multi-layered perceptron method of artificial neural network was computed by considering the various socio-economic characteristics of farmers as input vectors and adoption behaviour of the farmers as output vector in order to assess the socio economic factors influencing on adoption of drought coping mechanisms. The MLP method of ANN has trained by considering input layer with 12 input nodes, single hidden layer with 5 hidden nodes and an output layer with 2 output nodes. The hyperbolic tangent function used as activation function in hidden layers and the error is the cross entropy error because softmax activation function is applied to the output layer. The result shows that the cross entropy error is 33.17 and the following variables such as Crop insurance (100%), followed by Education (59.2%), Extension visits (58.5%), Income status (53.4%), Crop Diversification (47.4%), Animal Husbandry (28.6%) and Farm Size (24.0%) were significantly influencing on adoption process.

Key words: *Animal Husbandry, Crop Diversification,* Drought

INTRODUCTION

The State of Karnataka has 114 lakh hectare cultivable lands and 72 per cent of the cultivable area is rainfed; only 28 per cent is under irrigation. The State is the second largest in terms of arid region and it ranks second, next only to Rajasthan in India, in terms of total geographical area prone to drought. Drought is a common phenomenon in State of Karnataka. The State faced consecutive droughts during the years 2001-02, 2002-03 and 2003-04 resulted in sharp decline of agricultural output[9]. Drought stress is the major limiting factor for rice production and yield stability under rainfed crop eco system.

Cite this article: Halagundegowda, G.R. and Singh, A., Multilayer Perceptron Method of Artificial Neural Network for Classification of Farmers Based on Adoption of Drought Coping Mechanisms, *Int. J. Pure App. Biosci.* **6(2):** 1408-1414 (2018). doi: http://dx.doi.org/10.18782/2320-7051.6405

Karnataka faces high risk of moisture stress at maximum tillering and reproductive stages of crop, which may lead to yield loss of 25 to 100 per cent[6].

Adoption of Drought coping mechanisms

Drought is defined as "when a region receives below average precipitation, resulting in prolonged shortages in its water supply, whether atmospheric, surface or ground water. It can have a substantial impact on the ecosystem and agriculture of the affected region". As drought occurs in a particular area obviously its affects the crop and livestock production, in order to reduce the effect of drought on farm production and to stabilize the farm income, farmers have to take some systematic measures such measures are called drought coping mechanisms.

Research Problem:

One of the problems of classification lies in the use of appropriate methods to fit the model depending on the nature of data. It is well known that, most of data related to adoption of any agriculture technology (Agriculture Extension Survey data) are having qualitative response variable with two or more categories, which is a problem when using the traditional statistical methods, such as linear regression analysis because of not satisfying the assumptions of quantitative regresend in classical linear regression model. In such case we can think of qualitative response models such as logit model, probit model, tobit model, poison regression and multivariate techniques like linear discriminant analysis. Sometimes the results of prediction using these methods are inaccurate and may not give an appropriate picture of what could be the future events and there are other major problems trapped to use these classical models such as,

1. More stringent assumptions to satisfy by these statistical models, such as no multicolinearity among predictors, assumptions of multivariate normal distributions, assumptions of equality of variance covariance matrix, etc.
2. Cannot apply due to noisy data, contaminated data, outlier data, dirty data or incomplete data in nature and sometimes insufficient sample will not make generalisation of parameters, because of restriction of minimum sample size and will make cumbersome to use these models.
3. Most of Agriculture extension survey data having nonparametric in nature like nominal or ordinal scale of measurement, which will not support the analysis by using these classical models.

Therefore, in order to make more comfortable analysis, meeting very few assumptions and robust in any condition it is necessary to look for other machine learning methods, which are having following useful characteristics such as:

1. Parallel processing of information with high speed in distributed manner and they possess the capability to generalise the result and can predict the new outcome from past trends.
2. Robust systems and are fault tolerant. They can recall full patterns from incomplete, partial or noisy patterns and exhibit adaptability, they can adopt the free parameter to the changes in is surrounding environment.
3. Exploit nonlinearity, most of real life problem are highly nonlinear and exhibit input output mapping, that is, they can map input patterns to their associated output patterns.
4. Learn by examples, these architectures can be 'trained' with known examples of a problem before they are tested for their 'inference' capability on unknown instances of the problem. Hence, they can identify new objects previously untrained.

MATERIAL AND METHODS

The specific reason for choosing this study area was that, Kolar district belongs to Eastern Dry Zone of Karnataka and most of farmers were involved in rainfed agriculture because of shortage of rainfall and drought affected area. Hence adoption of certain coping strategies against drought is the major solution to stabilize the farm income during the drought period. The specific reason for choosing this study was to know the factors influencing on adoption of any strategies against drought and its impact of agriculture policy on Karnataka agricultural cropping pattern and how it's fluctuating from period to period and area to area when the drought occurs.

Nature and source of data

The current study utilizes both classification and prediction techniques. The household secondary data was used to fit the classificatory statistical models and the data were recorded on Socio- characters of farmers of Kolar districts of Karnataka (India). The data is mainly related to coping strategies implemented against drought by the farmers of this region and was collected by employing the multi stage sampling design during the year 2013-14, the department of Agricultural Economics (CARDS), Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu (India).

Table1: Variables Encoding Summary

Code	Variables	Measurement
Y	Adoption behaviour	Y= 0 for Non-Adopters = 1 for Adopters
X1	Age of the farmer	Number of years
X2	Education of the farmer	Formal Years of Education
X3	Household Size	Number of family members
X4	Farm Size	Number of acre's
X5	Farming Experience	Number of years
X6	Animal Husbandry	Number of farm animals and poultry birds
X7	Media Exposure	Number of sources exposed frequently
X8	Extension Visits	Number of Visits made to an research organisations
X9	Crop Diversification	Number of Crops Grown in that year
X10	Income Status	In Rupees (Rs.)
X11	Worth of Liquidating Assets	In Rupees (Rs.)
X12	Crop Insurance got by the government	In Rupees (Rs.)

Artificial Neural Network (ANN)

Artificial neural networks can be defined as information processing tools which mimic or copy the learning methodology of the biological neural networks. It derives its origin from human nervous system, which consists of massively parallel large interconnection of large number of neurons, which activate different perceptual and recognition task in small amount of time.

Multilayer Feed-forward Neural Networks:

Networks that contain more than one layer of artificial neurons, which allow unidirectional forward connections of inputs and outputs, are called multi-Layered Perceptron's (MLP) or multi-layered Feed-forward Neural Networks. A multi-Layered Perceptron's consists of a set of input terminals, an output neural layer, and a number of layers of hidden nodes between the input terminals and the output layer (Fig.1).

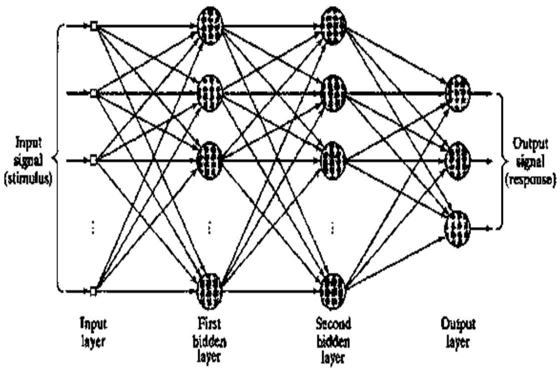

Fig. 1: Architectural graph of multilayer perceptron

In multi-layer feed forward network, information is transmitted from input layer to output layer, as in the case in the human brain where signals go in one direction. Feed forward networks use any Boolean function and are guaranteed to reach stability provided that the number of hidden neurons is sufficiently large. In fact, multi-layered Perceptron's can be considered as a special case of non-linear regression techniques. In economics, finance and agriculture, not all relations are always direct. Hidden layers grab the indirect relations between input and output variables.

Due to lack of information about the variables represented in the hidden layers, ANN is called by most researchers a "Black Box". The number of hidden layers and units in each hidden layer depends on the ability of network to approximate more complex functions. Most networks with complex structures do not perform necessarily better.

RESULTS AND DISCUSSION

A Multilayered perceptron neural network was fitted to the data with the help of SPSS 22.0 statistical package, the number of hidden nodes from 2 to 10. Thus, different numbers of neural network models are tried before arriving at the final structure of the model. Out of all neural network structures a neural network model with 12 input nodes and 5 hidden nodes performed better than other competing models in respect of out-of sample prediction and classification of adoption behaviour of farmers.

Table 2: Case Processing Summary

Data Set		N	Percentage
Sample	Training	120	80.0%
	Testing	30	20.0%
Valid		150	100.0%
Excluded		0	
Total		150	

Table 2 shows that 120 cases were assigned to the training sample and 30 sample as testing sample, this is in conformity with the rules that 80% of data set as training sample and 20% of data set as testing sample. The training sample comprises the data records used to train the neural network; some percentage of cases in the dataset must be assigned to the training sample in order to obtain a model. The testing sample is an independent set of data records used to track errors during training in order to prevent overtraining. Specify a numeric variable that assigns each case in the active dataset to the training and testing data set. Cases with a positive value on the variable are assigned to the training sample and cases with a value of 0, to the testing sample.

Table 3 shows that the information about the detail neural network architecture. Here 12 input variables used as covariates in analysis and the Standardized Rescaling Method is used for adjusting the Covariates. Scale dependent variables and covariates are rescaled by default to improve network training. All rescaling is performed based on the training data, even if a testing sample is defined. The network has an input layer with 12 input nodes; the number of units in the input layer is the number of covariates. A single hidden layer with 5 hidden nodes and an output layer with 2 output nodes.

Table 3: Network Information Summary

Input Layer	Covariates	1	Age
		2	Education
		3	Household Size
		4	Farm Size
		5	Farming Experience
		6	Animal Husbandry
		7	Media Exposure
		8	Extension Visits
		9	Crop Diversification
		10	Income Status
		11	Liquidating Assets
		12	Crop Insurance
	Number of Units[a]		12
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		5
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Adoption
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit

The hyperbolic tangent function used as activation function in hidden layers and it takes real valued arguments then transforms them to the range (–1, 1). The error is the cross entropy error because softmax activation function is applied to the output layer. It takes a vector of real valued arguments and transforms it to a vector whose elements fall in the range (0, 1) and sum to 1. Softmax is available only if all dependent variables are categorical.

Table 4 displays information on the result of training and applying the final network to the testing sample. Cross entropy error is displayed because the output layer uses the softmax activation function. This is the error function that the network tries to minimize during training. Cross-entropy error will have a predicted value for each category, where each predicted value is the probability that the case belongs to the category.

Table 4: Model Summary

Training	Cross Entropy Error	33.178
	Percent Incorrect Predictions	10.8%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.10
Testing	Cross Entropy Error	5.816
	Percent Incorrect Predictions	6.7%

a. Error computations are based on the testing sample.

In the above table the cross entropy error is 33.17 which is tolerable level and can continue the analysis for further steps also. The percentage of incorrect predictions is taken from the classification table and there is 10.8% of predictions are miss match with the original observed samples. Here there is one step to allow before checking for a decrease in error. The estimation algorithm stopped because the maximum number of epochs was reached. Ideally, training should stop because the error has converged. The cross entropy error is 5.8 and 6.7% incorrect predictions for testing data. The declining in both entropy error and incorrect predictions is mainly due to the effect of sample size.

Fig.2 present the importance of an independent variable is a measure of how much the network's model predicted value changes for different values of the independent variable. Normalized importance is simply the importance values divided by the largest importance values and expressed as percentages. The results are dominated by the variable Crop insurance (100%), followed by Education (59.2%), Extension visits (58.5%), Income status (53.4%) and then followed distantly by other predictors such as Crop Diversification (47.4%), Animal Husbandry (28.6%) and Farm Size (24.0%). These variables have the greatest effect on classification of farmers; it is the direction of the relationship between these variables and the predicted probability of adoption. We would guess that a larger amount of these variables indicates a greater likelihood of adoption and it needs to use a model with more easily interpretable parameters.

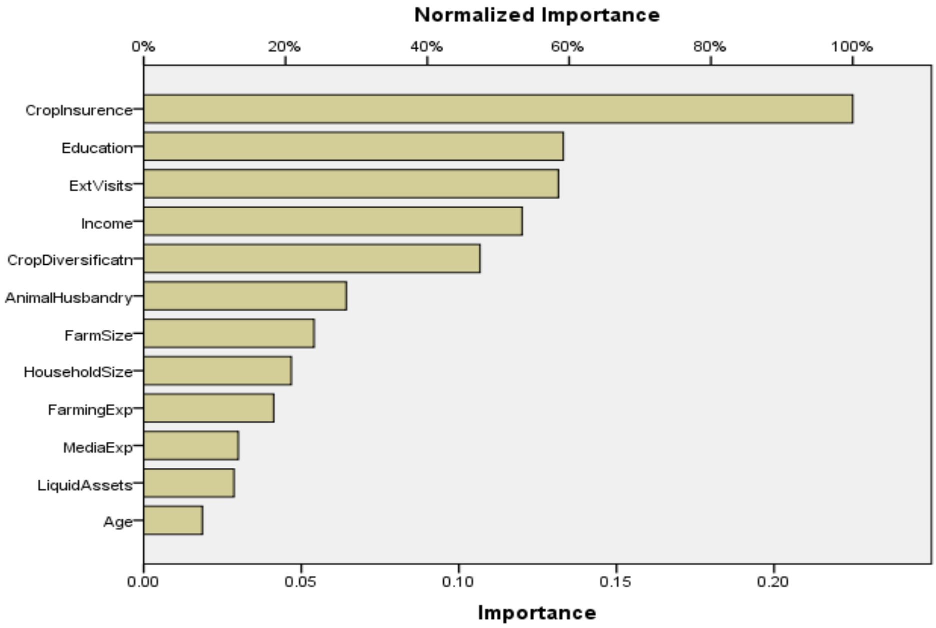

Fig. 2: Normalized variable Importance

Table 5: Classification Summary

Sample	Observed	Predicted		
		Non Adopters	Adopters	Percent Correct
Training	Non Adopters	43	7	86.0%
	Adopters	6	64	91.4%
	Overall Percent	40.8%	59.2%	89.2%
Testing	Non Adopters	11	1	91.6%
	Adopters	2	16	88.9%
	Overall Percent	46.7%	53.3%	90.2%

Table 5 shows that the cells on the diagonal of the cross classification of cases are correct predictions for each sample. The cells off the diagonal of the cross classification of cases are incorrect predictions of the cases used to create the model, 64 of the 70 farmers who previously adopted the drought coping strategies are classified correctly. 43 of the 50 non-adopters are classified correctly. Overall, 89.2% of the training cases are classified correctly, corresponding to the 10.8% incorrect shown in the model summary table. A better model should correctly identify a higher percentage of the cases.

Classifications based upon the cases used to create the model tend to be too "optimistic" in the sense that their classification rate is inflated. The testing sample helps to validate the model; here 90.2% of these cases were correctly classified by the model. This suggests that overall our model is in fact correct and efficient in prediction and classification.

REFERENCES

1. Alipasandi, A., Hosein, G and Saman, Z. A., Classification of three Varieties of Peach Fruit Using Artificial Neural Network Assisted with Image Processing Techniques. *International Journal of Agronomy and Plant Production*. **4 (9):** 2179-2186 (2013).

2. Fang, J., Wang, S and Zhang, C., Application of Genetic Algorithm（GA） Trained Artificial Neural Network to Identify Tomatoes with Physiological Diseases. *Nature and Science.* **3 (2):** 52-57 (2005).
3. Folorunso, O., Rebecca O. V., Ogunde, A.O and Agbool, B.A., Knowledge Sharing Adoption Model Based on Artificial Neural Networks. *International Journal of E-Adoption*. **2(4),** 1-14 (2010).
4. Funes, E., Allouche, Y., Beltrán, G and Jiménez, A., A Review: Artificial Neural Networks as Tool for Control Food Industry Process. *Journal of Sensor Technology*. **5:28**-43 (2015).
5. Ghamari, S. Classification of chickpea seeds using supervised and unsupervised artificial neural networks. *African Journal of Agricultural Research*. **7(21):** 3193-3201 (2012).
6. Hanamaratti, N.G., Prashanthi, S.K., Salimath, P.M., Hanchinal, R.R., Mohankumar, H.D., Parameshwarappa, K.G., and Raikar, S.D., Traditional Land Races of Rice in Karnataka: Reservoir of valuable traits, *Current Science*, **94(2):** pp.242-247 (2008).
7. John, N., Iheanacho, A.C and Irefin, D., Effects of Scio-Economic Characteristics of food crops farmers on the selection of coping strategies against drought in Bourno state, Nigeria. *Lincoln University Journal of Science.* **2 (1):** 13-18 (2011).
8. Mohammad, S.A. and Seyyed, A.N., Agro-Economic Factors Determining on Adoption of Rice-Fish Farming: An Application for Artificial Neural Networks. *Journal of Advanced Agricultural Technologies.* **1(2):** 151-157 (2015).
9. Nagaratna Biradhar and Sridhar, A Study on Consequences of 2003 Drought in Karnataka with reference to Livestock and Fodder, *Journal of Human Ecology*, **26(2):** pp.123-130 (2009)
10. Olaleye, O.L., Drought Coping Mechanisms: A Case Study of Small Scale Farmers in Motheo District of the Free State Province. *Master of Science thesis, University Of South Africa. SA.* (2010).
11. Pazoki, A.R., Farokhi, F and Pazoki, Z., Classification of Rice Grain Varieties Using Two Artificial Neural Networks (MLP and Neuro-Fuzzy). *The Journal of Animal & Plant Sciences.* **24(1):** 336-343 (2014).
12. Rakgase, M and Norris, D., Factors that Influence Choice of Drought Coping Strategies in Limpopo Province, South Africa. *Journal of Human Ecology*. **47(2):** 111-116 (2014).