Efficient Training of Self Organizing Map Network for Pattern Recognition

Preksha Pareek
Assistant Professor
Nirma University, Ahmedabad

Bhaskar Bissa
Sr. Software Engineer
Creative-Era, Ahmedabad

ABSTRACT
Pattern recognition is the science which helps in getting inferences from input data, usage of tools from machine learning and other algorithm designing. Neural networks techniques are popular in the field of pattern recognition. The importance of Neural Network is that it provides very powerful framework for representing mappings from several input variables to output variables. Self Organizing Map(SOM) technique has been applied in this work where implementation of one-D, two-D SOM has been done and modified algorithm of SOM has been proposed. In SOM unsupervised learning is employed where targets are not specified. Implementation of this has been done in C++. As a result of this modified algorithm of SOM performs better than using architecture of one-D map and two-D map networks for some sets of patterns.

Keywords
SOM ,unsupervised learning, machine learning.

1. INTRODUCTION
Small children are easily able to recognize digits and letters. Small characters, large characters, handwritten, machine printed characters. This ability is taken for granted until this task of teaching a machine to do the same is faced. Pattern recognition is the study of how machines can observe the environment, learn to distinguish among the patterns of interest from their background, and make good and finally making reasonable decisions about the categories of the patterns. Design of a general purpose machine pattern recognizer is an elusive goal and computers can surely be taught to recognize patterns[1]. In present scenario, computers are taught to recognize patterns by using Artificial Neural Network. The main characteristics of neural networks are that they can learn complex nonlinear input-output relationships, use training procedures, and adapt themselves, so they make task of pattern recognition automated. The increasing popularity of neural network models to solve pattern recognition problems has been primarily due to the availability of efficient learning algorithms to use. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object. Many problems in business, science, industry, and medicine can be treated as classification problems [7]. In these work artificial neural networks is used to classify a person as belonging to Below Poverty Line (BPL) or not.

In this work different learning algorithms of neural networks have been applied. For determining classification performance for Below Poverty Line dataset elusive task remains to find the appropriate network architecture so that classification accuracy can be maximized. This work comprises of reducing dimensionality of input dataset using Self Organizing Map or Kohonen network, which utilizes unsupervised learning method. In Kohonen network two commonly used maps- (1-d and 2-d) have been considered. Dimensionality reduction of the dataset is done using kohonen network and training performances of 1-d and 2-d maps have also been compared and a modified algorithm for SOM has been given and effect of training performance for classification by using these maps has been observed.

2. SELF ORGANIZING MAP
SOM is a grid like artificial neural network, whose cells or neurons becomes specifically tuned to various input signal patterns or class of patterns thorough an unsupervised learning process[2] . SOM groups the input data into clusters which are, commonly used for unsupervised learning. The self-organizing map (SOM) can project high-dimensional data into a visibly low-dimensional space[3]. SOMs are a type of unsupervised artificial neural network used to map *n*-dimensional data onto a low-dimensional space.

In engineering, the most straightforward application of SOM is in the identification and abstraction of hidden information from high-dimensional into low dimension, which is otherwise very difficult to perceive and interpret[10].

3. ARCHITECTURE
Method for implementing feature mapping process is to use architecture of an competitive learning network, but in addition to competitive network, at each stage weights are updated not only for the winning unit, but also for the units in the neighbourhood. The neighbourhood region is generally progressively reduced during learning .This is called self-organization network with Kohonen's learning. The data used during the training phase or new data can be mapped onto the network. Such a mapping reduces the dimensionality of the data and provides a means for identifying clusters and patterns[22]. The Kohonen network with *one-dimensional map* has two layers, an input layer and a Kohonen output layer as shown in figure 1. The input layer size is determined by user and must match with the size of entire row (pattern) in the input data file[23]. Each input neuron is connected to all other neurons in Kohonen output layer. These connections are weighted. Kohonen output layer is collection of nodes arranged in 1-dimension.

Two- dimensional map consists of 2 layers as shown in figure 2[27], first is input layer of neurons arranged linearly and second Kohonen layer which consists of neurons arranged in 2-dimensional array. Each neuron in the input is connected to all other neurons in the Kohonen layer where each connection is weighted.

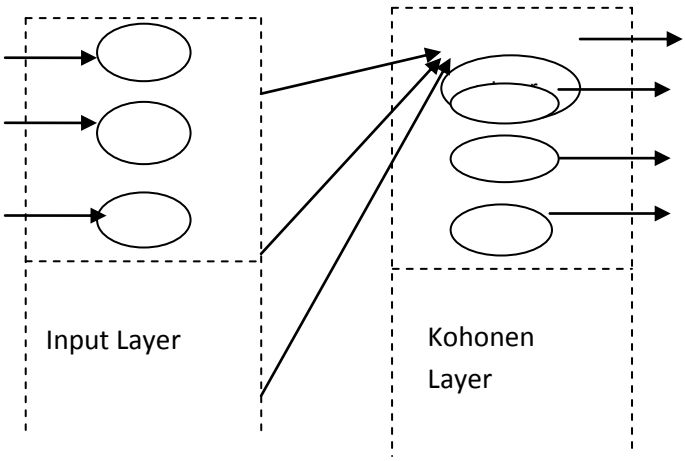

Figure 1: A Kohonen network with 1- dimensional map.

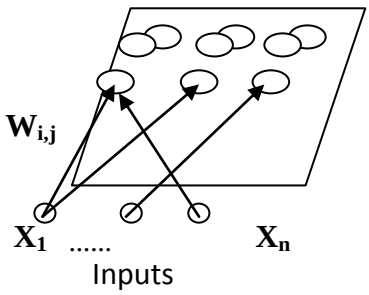

Figure 2*:* A Kohonen network with 2-dimensional map

4. WORKING

SOM is a ordered, smooth mapping of high dimensional data input manifold into the elements of low dimensional array where mapping is implemented as:-

During learning, those nodes that are topographically close in the given array up to a certain distance will activate each other to learn something(according to problem) from the same input x.

5. SOM ALGORITHM

Original algorithm of SOM is given as below[4]:

Randomize node weight W_{ij} from each input neuron to output neuron.

While (cycle<max_limit)
Grab the input vector x
For j<-1 **to** num_outputs
For i<-1 **to** num_inputs
o_j <-$\sum w_{ij} x_i$
end
Find winner node using maximum of o_j value
end
For neighbourhood size n_z
Update the node weight in the neighbourhood(n_z) of the winner neuron:
W_{new}<- $W_{old}+\alpha(x_i - W_{old})$ where α:gain value.
Decrement n_z , α at specified time.
Repeat until only winning neuron is left
Increment cycle
end

5.1 Whereas Modified Algorithm of SOM is given below

Randomize node weight W_{ij} from each input neuron to output neuron.

While *(cycle<max_limit)*
Grab the input vector x
For *j<-1* ***to*** *num_outputs*
For *i<-1* ***to*** *num_inputs*
o_j <-$\sum w_{ij} x_i$
end
Find winner node using maximum of o_j *value .*
end
if *number of patterns<threshold*
Update the node weight in the neighbourhood(n_z) of the winner neuron:
W_{new}<- $W_{old}+\alpha(x_i - W_{old})$ *where α:gain value.*
Decrement n_z *, α at specified time.*
else
Update the node weight in the new_nei (where new_nei=neighbourhood-index) of the winner neuron :
W_{new}<- $W_{old}+\alpha(x_i - W_{old})$ *where α:gain value.*
Decrement new_nei , α at specified time.
Repeat until only winning neuron is left .
Increment cycle.
end

5.2 Comparative Analysis of self Organizing Map and its Variants Dataset Used

Below Poverty Line (BPL) : BPL dataset consists of 13 feature values which are

(1)Land holdings:- (i)nil holding (ii) less than1 ha (iii) more than 1 and less than 2 ha

(2)Types of house: (i)nil house (ii)non-cemented house(iii)cemented house .

(3)Availability of clothing :(i)Less than 2 pairs(ii)More than 2 but less than 4 pairs(iii)More than 4 but less than 6 .

(4)Food security: (i)Less than 1 meal per day in major part of year(ii)Normal 1 meal but sometimes less(iii)Normal 1 meal throughout the year.

(5)Sanitation: (i)open(ii)group bathrooms with irregular water supply (iii) group bathrooms with regular water supply.

(6)Consumable durables(T.V., electric fan, kitchen appliances): (i)nil (ii)Any one item (iii) Any 2 item (7)literacy status of highest literate: (i) illiterate (ii) upto primary(iii) completed secondary.

(8)Status of household labour:(i) bonded labour(ii) women and child labour(iii) adult males.

(9)Means of livelihood:(i)Casual labour(ii) subsistence cultivation(iii) artisan.

(10)Status of children:(i) not going to school and working(ii) going to school and working (iii)going to school and not working.

(11) Types of indebtedness: (i)daily consumption purpose from normal sources(ii) production purpose from normal sources(iii)for other purpose from normal sources.

(12) Reason for migration :(i)casual work(ii) other form of livelihood (iii) non migrant .

(13)Preference of assistance: (i)wage employment (ii)self employment iii)Housing[6].

Whole experiment has been performed in C++.

Input file: Size of input file is 13*24. Here 13 features of Below Poverty Line dataset are taken where 24 patterns of such samples is considered. All the parameters are given and then training of 1-D SOM is performed.

Here below table 1 shows the performance of different architectures of SOM by fixing neighbourhood size m, alpha (learning rate), network size(i/p and o/p) and performance in terms of correctly classified samples have been taken, same thing has been performed in table 2 with change in parameters i.e. neighbourhood size m, alpha(learning rate),network size(i/p and o/p) .This comparison has been plotted and shown in figure 3.

Table 1: Training performance of 1-D ,2-D and modified algorithm map using neighborhood size 8.

Algorithm of Self Organizing Map	Neighbourhood size and alpha	Size of input layer	Size of kohonen layer	Performance (correctly classified samples)
1-D	0.83,8	13	14	11
2-D	0.83,8	13	7,2	14
Modified Algorithm	0.83,8	13	7,2	18

Table 2 : Training performance of 1-D ,2-D and modified algorithm map using neighbourhood size 10.

Algorithm of Self Organizing Map	Neighbourhood size and alpha	Size of input layer	Size of kohonen layer	Performance after dimensionality Reduction(correctly classified samples)
1-D	0.53,10	13	12	15
2-D	0.53,10	13	6,2	19
Modified Algorithm	0.53,10	13	6,2	21

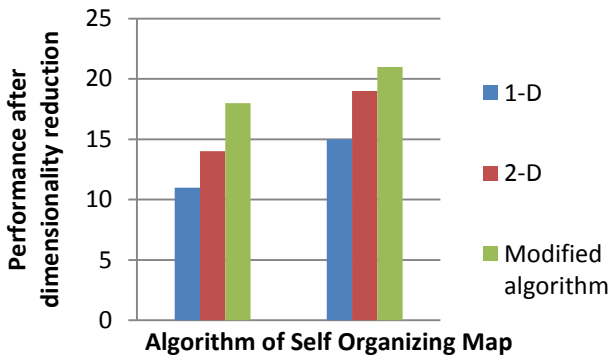

Figure 3: Comparison of training performances of map of 1-D, 2D and modified algorithm

6. CONCLUSION

This work aims to explore different learning algorithm of Self Organizing Map available with neural networks to classify a person as belonging to Below Poverty Line or not. Self Organizing Map (1-D and 2-D) and the modified SOM algorithm are compared to show the training performance after dimensionality reduction and as a result of this modified algorithm resulted in most efficient after reduction of dimensionality.

7. REFERENCES

[1] Anil K. Jain , Robert P.W. Duin, Jianchang Mao, January 2000 ,"Statistical Pattern Recognition : a Review", IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 1.

[2] Teuvo Kohonen," march 1990 The Self Organizing Map", IEEE .

[3] Chung-Chian Hsu, Shu-Han Lin, January 2012" Visualized analysis of mixed Numeric and Categorical data via Extended Self-Organizing Map", IEEE transactions on neural networks and learning systems, vol. 23, no. 1.

[4] Valluru Rao, Hayagriva Rao, "C++ NEURAL NETWORKS AND FUZZY LOGIC",M&T Publishing, Edition 2nd.

[5] Kelvin L. Preedy, Paul E.Keller, "ARTIFICIAL NEURAL NETWORKS AN INTRODUCTION" ,SPIE press edition 2005.

[6] Indira Hirway,2003," Identification of BPL Households for Poverty Alleviation Programmes" Economic and Political Weekly, vol 38,no 45,pp 4803-4808.

[7] Guoqiang Peter Zhang ,2000 "Neural Networks for Classification: A Survey" IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 30, no. 4,.

[8] R. Zurita-Milla, J. A. E. van Gijsel, N. A. S. Hamm, P. W. M. Augustijn, and A. Vrieling,, August 26, 2012, "Exploring Spatiotemporal Phenological Patterns and Trajectories Using Self-Organizing Maps", IEEE transactions on geoscience and remote sensing, accepted,.

[9] Jayanta Kumar Basu, Debnath Bhattacharyya, Tai- Hoon Kim, April 2010," Use of Artificial Neural Network in Pattern Recognition", International Journal of Software Engineering and Its Applications, Vol. 4, No. 2.

[10] Abhinandan De, Nirmalendu Chatterjee, April 2002," Recognition of Impulse Fault patterns in transformers using Kohonen's Self-Organizing Feature Map", IEEE transactions on power delivery, vol. 17, no. 2.