Development and Performance Evaluation of PSO based Single Layer Nonlinear ANN Classifiers of Indian Online Shoppers

Ritanjali Majhi
School of Management
National Institute of Technology, Warangal, India
e-mail:ritanjalimajhi@gmail.com

Bijayalaxmi Panda
Dept. of CSE
ITER, S 'O'A University, Bhubaneswar, India
e-mail: bijaya_003@gmail.com

Babita Majhi
Dept. of IT
ITER, S 'O'A University Bhubaneswar, India
e-mail: babita.majhi@gmail.com

T. Rahul
School of Management
National Institute of Technology, Warangal, India
e-mail:rahulkitss@gmail.com

Ganapati Panda
School of Electrical Sciences
Indian Institute of Technology Bhubaneswar
e-mail : ganapati.panda@gmail.com

Abstract— **In this paper an in depth study is made on identifying primary factors on which the Indian online shoppers are influenced. Based on the dominating factors extracted from the internet shoppers' replies they are clustered. A novel and efficient single layer nonlinear ANN model with PSO based weight adaptation (SNANNP) is developed by taking the demographic input of the shopper and clustering results as the desired class. Close observation of experimental results indicate that the proposed method exhibits superior classification performance compared to that obtained by discriminant analysis and conventional single layer ANN (SNANN) model.**

Keywords-Single layer nonlinear ANN (SNANN); Consumer Classification; Online Shopping; Factor Analysis; Discriminant Analysis and Particle swarm optimization(PSO) based training

I. Introduction

In recent years, internet marketing has gained popularity. The competition among the retailers have increased due to introduction of several online services. The literature survey reveals that nearly 67% shoppers use internet and half of them do shopping online. The advantages for choosing online shopping are: reduction in the cost by employing less manpower, saving of shopping time, less fear of loss of money in online shopping, increase in the bargaining power of shoppers, increase of rivalry among the competitors and improvement in security and ease of delivery.

Gradually the interest towards online shopping is increasing during last few years. The consumers mostly prefer online services to purchase air/railway tickets, movie tickets, consumer electronic goods, audio/video files, software packages like an operating system, web browser, audio/video player etc. Usually the Indian consumers discuss with their friends and relatives before buying a product.

The neural network has been used to classify consumers in choosing hospitals and marketing implications [1]. This model is also useful for market strategy planning. In another paper [2], a new approach has been proposed for direct marketers targeting potential consumers new to their categories. It is reported that the ANN model provides improved classification accuracy. A recent study [3] deals with conjoint analysis to create health plans that optimize value for consumers.

A systematic study is made in this paper to group the consumers based upon their online shopping behaviour. The choice of people to prefer a service is diverse. A questionnaire was made on internet shopping based upon which 14 variables pertaining to the theme are chosen. Considering the correlation between the variables these are reduced into a number of factors. Using loading of the factors internet shoppers are grouped into clusters. Each cluster represents group of consumers with similar behaviour.

The consumers are then identified into number of classes depending on the demographic attributes such as age, gender, level of education, amount spent on shopping and amount spent on online shopping. The results of the cluster analysis is used as the training class of the classification model. Three types of classification methods such as discriminant analysis, single layer nonlinear artificial neural network(SNANN) and single layer ANN with PSO based training are used. The rest of the paper is organized as follows :

Section II develops a unique intelligent single layer nonlinear ANN model used for consumer classification. The basic of particle swarm optimization algorithm is outlined in Section III. Section IV deals with issues relating to data collection and data reduction using factor analysis. Cluster analysis for consumer grouping is dealt within Section V. The classification operation using discriminate analysis, SNANN and SNANNP models is carried out in Section VI. The results obtained from simulation studies are also presented and discussed in the same section. Finally Section

VII deals with the conclusion of the study.

II. SINGLE LAYER NONLINEAR ANN(SNANN) CLASSIFIER

The SNANN possesses the simplicity of single layer ANN and performance capability multilayer ANN. Hence it is chosen here to develop an efficient adaptive classifier. The enhanced performance is achieved by introducing nonlinearity through trigonometric functional expansion [6]. The block diagram of an SNANN structure is shown in Fig.1. Let the input signal vector be represented as

$$\underline{X}(k) = [x(k)\ x(k-1)....................x(k-m+1)]^T \quad (1)$$

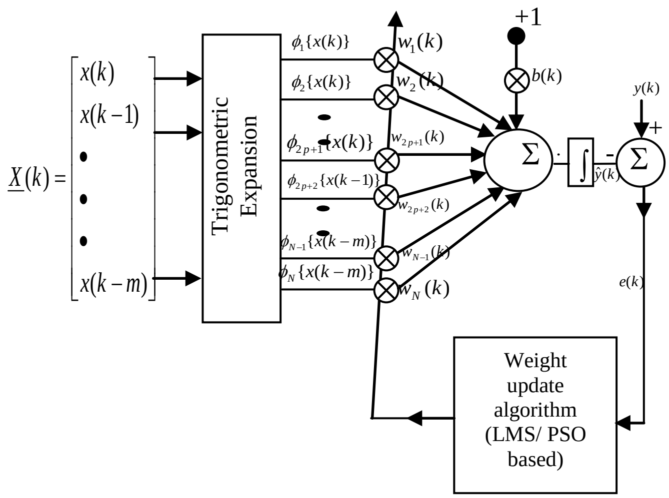

Fig. 1 Proposed block diagram of SNANN/SNANNP model with or without PSO based training

which is the demographic input of an online shopper. Then the functional expansion (FE) block maps each element $x(k)$ into $(2p+1)$ nonlinearly expanded independent components. For trigonometric expansion $\underline{\phi}[x(k)]$ is given by

$$\underline{\phi}[x(k)] = [x(k), \cos\{\pi x(k)\}, \sin\{\pi x(k)\}.......$$
$$\cos\{p\pi x(k)\}, \sin\{p\pi x(k)\}]^T = \quad (2)$$
$$[\phi_1\{x(k)\}, \phi_2\{x(k)\},......\phi_{2p+1}\{x(k)\}]^T$$

where p is an integer. When each element of $\underline{X}(k)$ is expanded then the expanded vector is represented as

$$\underline{\phi}[\underline{X}(k) = [\phi_1\{x(k)\}.......\phi_{2p+1}\{x(k)\}, \phi_{2p+2}\{x(k-1)\},.....$$
$$\phi_{2(2p+1)}\{x(k-1)\}..........\phi_N\{x(k-m)\}]^T \quad (3)$$

where $N = m(2p+1)$ and m is the number of signal samples fed into the SNANN. The output $y\hat{}(k)$ of Fig. 1 is then given by

$$\hat{y}(k) = \underline{f}\{\phi^T(\underline{X}(k))\underline{W}(k) + b(k)\} \quad (4)$$

where $b(k)$ is the bias weight and $f\ \{.\}$ denotes tanh function. $W(k)$ represents the weight vector given by

$$\underline{W}(k) = [w_1(k), w_2(k)..........w_N(k)]^T \quad (5)$$

The weights of the SNANN model are trained using the algorithm

$$\underline{W}(k+1) = W(k) + \mu\underline{\phi}\{\underline{X}(k)\}.e(k)(1-\hat{y}^2(k)) \quad (6)$$

where the error term is given by

$$e(k) = y(k) - y\hat{}(k) \quad (7)$$

The convergence coefficient is represented by μ and its value lies between 0 and 1. Equations (2), (3), (4) and (7) represent the key equations of SNANN algorithm.

III PARTICLE SWARM OPTIMIZATION (PSO)FOR TRAINING OF CLASSIFIER WEIGHTS

The PSO is a simple but efficient population based swarm intelligence based optimization algorithm. The PSO algorithm emulates swarm behavior and each particle represents a point in the D-dimensional solution space. The swarm, a collection of particles, initially contains a population of random solutions. Each particle is given a random velocity to fly through the problem space. Each particle keeps track of its previous best position, called *pbest* and its fitness value. Each swarm remembers its best position called *gbest* . The velocity $v_i(d)$ and the position $x_i(d)$ of the d th dimension of ith particle are adapted [4]-[5] according to

$$v_i(d) = wv_i(d) + c_1 * rand1_i * (p_i(d) - x_i(d)) +$$
$$c_2 * rand2_i(d) * (p_g(d) - x_i(d)) \quad (8)$$

$$x_i(d) = x_i(d) + v_i(d) \quad (9)$$

where $p_g(d)$ and $p_i(d)$ are the d th dimensional positions corresponding to the *gbest* and *pbest* respectively, $rand1_i(d)$ and $rand2_i(d)$ represent random numbers in the range [0, 1] and are different in different dimensions, c_1 and c_2 are acceleration coefficients and $w$is the inertial weight which plays the crucial role of balancing between the global search and local search.

IV DATA COLLECTION AND FEATURE REDUCTION

(a)Data collection

The data is collected through a questionnaire from Indian online shoppers. Each answer is weighted in a 5 point-scale in which '1' indicates "strongly disagree" and '5' indicates "strongly agree".

The questionnaire comprises of 5 sections. First section contains 5 questions. In sections 2, 3 and 4, 14 variables are used. The last section deals with the questions related to the

ISBN: 978-1-61208-101-4

demography of the consumers. Demographic attributes include age, gender, education, average amount spent on shopping and average amount spent on online shopping. A total of 163 complete data are collected each having 14 variables excluding the demographic variables. The 14 variables used are listed in Table 1.

Table 1
List of variables affecting behavior of internet shoppers

Information Design (ID)	Internal Norm (IN)
Visual Design (VD)	Convenience(C)
Navigation Design (ND)	Online Innovativeness(OI)
Communication (Comm)	Enjoyment (E)
Self Efficacy (SE)	External Norm (EN)
Privacy (P)	Security (S)
Vender repute (VR)	Social Presence (SP)

(b)Data reduction using factor analysis

Factor Analysis [7] is a statistical tool to find interrelationships between the variables. This is a technique through which the prominent factors hidden in the data are extracted. This is essentially a data reduction method used to reduce a set of observed variables to a set of latent variables.

This analysis is an ideal method for extracting factors. There are several methods to decide how many factors has to be extracted. The most widely used method for determining the number of principal factors is based on eigen value consideration [8]. In the present study five factors influencing consumer behavior are obtained using factor analysis. The eigen values, percentage of variance and cumulative percentage of variance of five prominent variables obtained from factor analysis are listed in Table 2. The factor loadings obtained are provided in Table 3.

V. CLUSTER ANALYSIS

Clustering mainly groups unlabeled datasets. It is an unsupervised classification technique. Mainly two types of clustering techniques are used : hierarchical and partitioning/ non-hierarchical [9]. Each one has its own merits/demerits depending on the applications in which it is used. In this study hierarchical clustering is chosen because it is suitable when the number of possible clusters are not known apriori. Cosine distance is used as the similarity measure to evaluate the closeness among the data points. The dendrogram of the factor scores computed is shown in Fig.2.

The dendrogram shows that the entire data are grouped into three clusters. It is found that 41, 67 and 55 online shoppers belong to clusters 1, 2 and 3 respectively. The factor scores with major contribution to each cluster are shown in Table 4. The demographic variables of the consumers in each cluster are listed in Table 5.

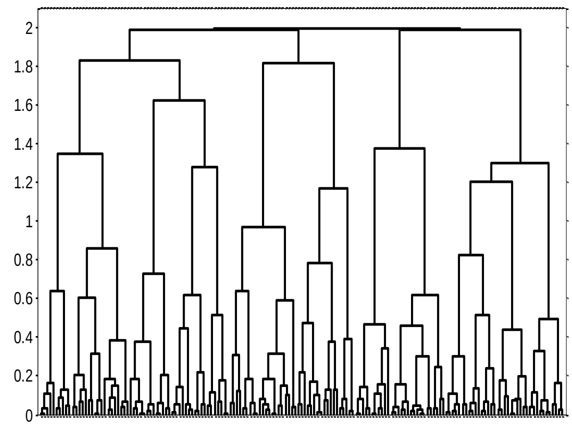

Fig. 2 Dendrogram of the factor scores using hierarchical clustering

Table 2
Eigen values, percentage of variance and cumulative percentage of variance of variables obtained from factor analysis

Eigen values	% of variance	Cumulative % of variance
4.4286	31.6330	31.6330
1.8162	12.9728	44.6057
1.6309	11.6494	56.2551
1.1081	7.9153	64.1704
1.0542	7.5303	71.7007

VI. CONSUMER CLASSIFICATION

Clustered online shoppers need to be classified. The classification is based upon their behavior towards online shopping. The demographic attributes of the consumers are used as the inputs to the classifier. The results obtained from the cluster analysis are used as the training class for the classifier. The SNANN, SNANNP as well as discriminate classifiers are employed to classify these consumers.

(a)Discriminant analysis based classification

Discriminant analysis is employed to classify the objects into various groups. This is a supervised classification technique [10, 11] in which the dependent variable is the class/group and the independent variables are the features/attributes. In this, a number of discriminate

ISBN: 978-1-61208-101-4

Table 3
Factor loading of different variables under different factors

Variables	Factor1	Factor2	Factor3	Factor4	Factor5
ID	0.2130	**0.4800**	0.4023	0.0602	-.0392
VD	0.1677	**0.8720**	0.0570	0.1575	-.0754
ND	0.1703	**0.7237**	0.1424	0.2232	0.0167
Comm	0.0640	0.1671	0.0258	**0.8239**	0.0458
SP	0.0789	0.1102	0.0618	**0.6818**	0.0351
SE	0.3715	0.3907	**0.4150**	-.1339	0.0158
Privacy	**0.8460**	0.1897	0.1807	0.0597	-.0419
Security	**0.7416**	0.1726	0.2853	0.0779	0.0043
VR	**0.5191**	0.2751	0.2846	0.2947	0.0616
EN	-.1596	-.1216	0.0916	0.2442	**0.6016**
IN	0.1571	0.0578	0.0655	-.1265	**0.9730**
Enjoyment	0.1318	0.1939	**0.7547**	0.2457	0.0246
Convenience	0.3304	0.0320	**0.6090**	-.1239	0.0122
Online innovativeness	0.0650	0.0423	**0.3019**	0.0332	0.0716

Table 4
Factor scores under different clusters

Factors	Cluster 1	Cluster 2	Cluster 3
Factor 1	-0.3974	**0.3371**	-0.1145
Factor 2	**0.2489**	0.1696	-0.3922
Factor 3	-0.9219	0.2284	**0.4090**
Factor 4	-0.0025	-0.6127	**0.7483**
Factor 5	-0.7892	0.1706	**0.3804**

functions are formed which is equal to the number of classes in the data. These functions represent the classification rules for each class. The actual data is applied to each function and the corresponding output is obtained.

The data belongs to a group that provides maximum output. Linear Discriminant analysis is used for this purpose. 80 percent of the total data is used for training and remaining 20 percent isused for testing the performance. The classification results obtained from the tests are shown in Table 6.

(b)SNANN based classification

In this model the input signal $x(k)$ represents the five demographic data of the online shoppers, each of which is expanded to three trigonometric terms. These are then fed to the SNANN model. The three outputs obtained from the model using (4) are compared with those obtained from the cluster analysis to produce error signal

$$e(k) = y(k) - y\hat{}(k) \qquad (10)$$

The mean square error is used as the performance index given by

$$MSE(k) = \frac{\sum_{k=1}^{K} e^2(k)}{K} \qquad (11)$$

The epoch based training of weights is carried out using the LMS algorithm. The classification results obtained from test are shown in Table 7.

(c) Proposed SNANNP based classification

The training rule for updating the weights of the proposed SNANNP model is outlined in the following steps:

1.The model is provided with K input patterns each consisting of five demographic features of the consumers. Each feature is expanded to three trigonometric terms and then given to the SNANN model.

Table 5
Demographic characteristics of the consumer under each cluster

		Cluster1	Cluster2	Cluster3	Cumulative
Gender	**Male**	**35**	**56**	**47**	**138**
	Female	**6**	**11**	**8**	**25**
Age	**<=20**	**2**	**2**	**2**	**6**
	21-30	**29**	**55**	**45**	**129**
	31-40	**8**	**9**	**6**	**23**
	> 40	**2**	**1**	**2**	**5**
Education	**undergraduates**	**2**	**3**	**3**	**8**
	graduates	**0**	**2**	**4**	**6**
	postgraduates	**36**	**59**	**46**	**141**
	doctoral	**3**	**2**	**2**	**7**
	others	**0**	**1**	**0**	**1**
Amount spent on shopping	**<=25,000**	**24**	**31**	**22**	**77**
	25,001-50,000	**7**	**12**	**15**	**34**
	>50,000	**10**	**24**	**18**	**52**
Amount spent on online shopping	**<=20,000**	**35**	**43**	**39**	**117**
	20,001-40,000	**2**	**11**	**5**	**18**
	>40,000	**4**	**13**	**11**	**28**
					163

 ISBN: 978-1-61208-101-4

2. Each new term is multiplied with the corresponding weight and then summed to give an output and in this way K numbers of estimated outputs are computed.
3. Each desired output is compared with the corresponding model output and K errors are produced.
4. The mean square error (MSE) (corresponding to n_{th} particle) is determined by using the relation defined in (11). This is repeated for M times, where M is the number of particles.
5. Since the objective is to minimize MSE (n), n = 1 to M the PSO based optimization method is used.
6. The velocity and position of each particle is updated using (8) and (9).
7. For each iteration the minimum MSE, MMSE is stored which indicate the learning efficiency of adaptive model.
8. When the MMSE reaches the pre-specified value the optimization process is stopped

Table 6
Classification results obtained from testing using discriminant analysis

Classified observations	Cluster 1	Cluster 2	Cluster 3
Class 1	2	3	3
Class 2	6	10	5
Class 3	0	0	3
Cumulative	8	13	11

Table 7
Classification results obtained from testing using SNANN

Classified observations	Cluster 1	Cluster 2	Cluster 3
Class 1	3	0	0
Class 2	5	13	7
Class 3	0	0	4
Cumulative	8	13	11

(b) At this stage all the particles attains almost the same positions, which represent the desired solution of the given SNANN model.

For classification purpose 80 percent of the total data is used for training and remaining 20 percent is used for testing. Various parameters chosen in the simulation study for PSO are : number of particles =10, inertia weight, $w = 0.5$, acceleration constants $c_1 = 0.5$ and $c_2 = 0$, no of generations = 1000 and no. of ensampling average used =20. The classification results obtained from test data are shown in Table 8.

VII. DISCUSSION

Five key factors are obtained from fourteen variables selected based on eigen values greater than one. The factor scores obtained are used in the cluster analysis. Clustering, operation provides 41, 67and 55 consumers in clusters 1, 2 and 3 respectively. From Table 4 it is observed that cluster 1 is heavily loaded by the factor 2. So all the consumers seeking website design belong to cluster 1. Cluster 1 represents website design. Factor1 load heavily on cluster 2.Consumers who are particular about privacy and security belong to cluster 2 and hence is named as privacy and security. Factors which load heavily on cluster 3 are factors 3, 4 and 5. In this cluster the consumers are interested in ease, enjoyment, communication and norms. So cluster 3 is named as ease, enjoyment, communication and norms. The statistics of the demography of the consumers are listed in Table 5.

The accuracy classification for three classifiers using demographic features as inputs are shown in the Tables 6, 7, and 8. The comparative results of the three classifiers are listed in Table 9.

Table 8
Classification results obtained from testing using SNANNP model

Classified observations	Cluster 1	Cluster 2	Cluster 3
Class 1	5	0	0
Class 2	3	13	3
Class 3	0	0	8
Cumulative	8	13	11

Table 9
Comparative results of efficiency obtained during testing

Model	No. of correct classifications	Classification accuracy
Discriminant analysis	15	46.87%
SNANN	20	62.50%
SNANNP	26	81.25%

The results presented in Table 9 demonstrate that the SNANNP model provides best classification performance compared to those obtained by other two methods.

VIII. CONCLUSION

In this paper, a novel method has been proposed for classification of Indian consumers based on their behavior towards online shopping. The factor analysis is carried out on the data to achieve reduced number of factors. The hierarchical based clustering is used to group the consumers. Three different classifiers : discriminant analysis, SNANN and SNANNP are used for classification. The simulation results indicate that the proposed SNANNP based model provides best classification performance compared to its statistical counterpart. The proposed research can also be extended for classification of other real life data. The classification accuracy can further be enhanced by adding physcographic and cultural inputs to the proposed classifier.

REFERENCES

[1] Wan-I Lee, Bih-Yaw Shih and Yi-Shun Chung, "The exploration of consumers" behavior in choosing hospital by

the application of neural network", Expert systems with applications, vol. 34, pp. 806-816, 2008.

[2] F. Kaefer, C. M. Heilman and S.D. Ramenofsky, "A neural network application to consumer classification to improve the timing of direct marketing activities", Computers and Operations Research, vol. 32, pp. 2595-2615, 2005.

[3] R. Gates, C. McDaniel and K. Braunsberger, "Modeling consumer health plan choice behavior to improve customer value and health plan market share", Journal of Business research, vol. 48, pp. 247-257, 2000.

[4] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory", in Proc. of 6th Int. symp. Micro machine Human Sci., Nagoya, Japan, 1995, pp. 39-43, 1995.

[5] J. Kennedy and R. C. Eberhart, "Particle swarm optimization", in Proc. of IEEE Int. Conf. Neural Networks, 1995, pp. 1942-1948.

[6] J. C. Patra, R. N. Pal, B. N. Chatterjee and G. Panda, "Identification of nonlinear dynamic systems using functional link artificial neural networks", IEEE Trans. on Systems, Man and Cybernetics – Part B, vol. 29, no. 2, pp. 254-262, April 1999.

[7] S.Sharma and A. Kumar, *Cluster analysis and factor analysis*, University of South Carolina, Arizona State University.

[8] Glover, Jenkins and Doney, "Principal Component and Factor Analysis, Modeling Methods for Marine Science, pp.81-117, 2008.

[9] N. Jardine and R. Sibson, "The construction of hierarchic and non-hierarchic classifications", the Computer Journal, pp.11-177, 1968.

[10]S.Balakrishnama,A.Ganapathiraju and J. Picone,"Linear Discriminant Analysis for Signal Processing Problems", IEEE, pp. 78-81, 1999.

[11]S.Balakrishnama and A. Ganapathiraju,, Linear discriminant analysis-as brief tutorial, Department of Electrical and Computer Engineering, Mississippi State University.