International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org, editorijaiem@gmail.com
Volume 2, Issue 7, July 2013 ISSN 2319 - 4847

A PROBABILISTIC NEURAL NETWORK APPROACH FOR CLASSIFICATION OF VEHICLE

Madhusmita Mishra[1], Amrut Ranjan Jena[2] and Dr. Raja Das[3]

[1&2] Asst. Prof.,Department of Computer Sc. & Engineering
C V Raman College of Engineering, Bhubaneswar-752054, Odisha,

[3]Dr.Raja Das, Asst.Prof., School of Advanced Science, VIT University, Vellore

Abstract

The vehicle classification, which consists of determining the vehicle of different company, is very important for a customer because a vehicle may fit into multiple categories. Before buying a vehicle we consult reviews, ratings from numerous agencies. Some agencies perform rigorous testing on the vehicles and quantize the vehicle features like acceleration, braking, fuel economy etc, while other relies on the consumer reviews, awards won by particular vehicle model. Therefore mathematically we can say that: Vehicle success = f (Vehicle features). In this work, we propose a new approach for vehicle classification based on a Probabilistic Neural Network and feature selection. Our goal is to classify a customer liking vehicle among the number of different vehicles available in the market on the basis of market survey. For this purpose, the datasets are extracted from a genuine source and are used as training datasets. The proposed work provides a soft computing-based tool capable of classifying vehicles, as closely as possible to classifications performed by skilled operators. Such a tool is capable of extracting a number of numerical parameters characterizing the vehicles in areas like Value for Money. For data analysis in this paper the vehicle is a car type of different company.

Keywords: Probabilistic Neural Network, Probability Density Function, Vehicle Classification, Bayes's decision rule.

1. INTRODUCTION

Car can be categorized into various types on the basis of their design, functional specifications, performance, safety, cost etc. We can't classify all the cars on the basis of one single formula. A particular car model can be categorized as a sports car as well as a convertible. Cars are classified on the basis of their manufacturing specifications like: Medium Cars, Family Cars, People Movers, Sports Cars, Luxury SUVs, etc.
The model evaluates cars by considering the following features of the car:

Features	Descriptions
CAR	car acceptability
PRICE	overall price
BUYING	buying price
MAINT	price of the maintenance
TECH	technical characteristics
COMFORT	comfort
PERSONS	capacity in terms of persons to seat
SAFETY	estimated safety of the vehicle
DOORS	number of doors
LUG BOOT	the size of luggage boot

Each feature is weighted as following manner:

Buying- v-high, high, med, low.
Maintenance- v-high, high, med, low.
Doors- 2, 3, 4, 5.
Persons- 2, 4, 5.
lug boot- small, med, big.

Safety- low, med, high.
Each criterion is given an importance weighting score. An individual can purchase a car with its own satisfaction in all areas. Each car's overall weighting reflects its position within the class and should only be compared within its category.

Classification is a variable technique referred with assigning data cases to one of a fixed number of possible classes. Here a Probabilistic Neural Network is used for vehicle classification.

Probabilistic Neural Network:
A Probabilistic Neural Network (PNN) is defined as an implementation of statistical algorithm called Kernel discriminate analysis in which the operations are organized into multilayered feed forward network with four layers: input layer, pattern layer, summation layer and output layer. A PNN is predominantly a classifier since it can map any input pattern to a number of classifications. Among the main advantages that discriminate PNN is: Fast training process, an inherently parallel structure, guaranteed to converge to an optimal classifier as the size of the representative training set increases and training samples can be added or removed without extensive retraining. Accordingly, a PNN learns more quickly than many neural networks model and have had success on a variety of applications. Based on these facts and advantages, PNN can be viewed as a supervised neural network that is capable of using it in system classification and pattern recognition. The main objective of this paper is to describe the use of PNN in vehicle classification.

The architecture of PNN:
The PNN was first proposed in [11]. A probabilistic neural network is built with four layers as shown in the figure.1. PNN is one of the types of neural networks with a one pass learning algorithm. . The PNN architecture is composed of many interconnected processing units or neurons organized in successive layers. The input layer unit does not perform any computation and simply distributes the input to the neurons in the pattern layer. On receiving a pattern from the input layer, the neuron of the pattern layer computes its output as the probability density function (pdf) for a single sample.

i.e., $$f_k(X)=\frac{1}{(2\pi)^{p/2}.\sigma^p}e^{-\frac{\|X-X_k\|^2}{2\sigma^2}}$$

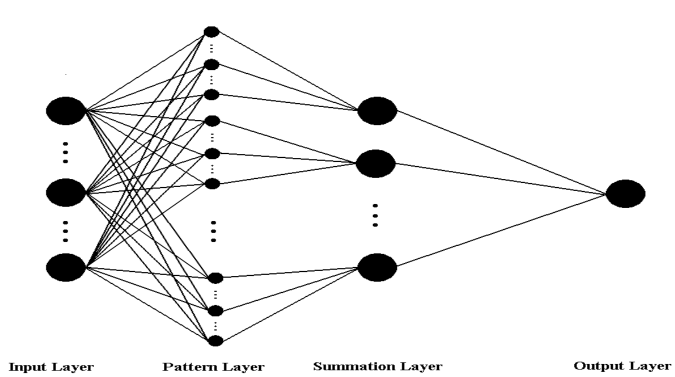
where X is unknown input vector, X_k is *k*th sample input vector, i is the smoothing parameter and *p* is the dimension of the input vector. The neuron of the summation layer computes its output as the pdf for a single population or pattern.

i.e., $$g_i(X)=\frac{1}{(2\pi)^{p/2}.\sigma^p}\frac{1}{n_i}\sum_{k=1}^{n_i}e^{-\frac{\|X-X_K\|^2}{2\sigma^2}}$$

Where n_i is the total number of samples in the *i*th population.

If the a priori probabilities for each class are the same, and the losses associated with making an incorrect decision for each class are the same, the decision layer unit classifies the pattern in accordance with the Bayes's decision rule based on the output of all the summation layer neurons.

i.e, $h(x)=\arg\max\{g_i(x)\},\quad i=1,2,3,\cdots,m$ Where $h(x)$ denotes the estimated class of the pattern *x* and *m* is the total number of classes in the training samples. This includes determining the network size, the pattern layer neurons and an appropriate smoothing parameter.



(Fig-1)

Training Set
The training set must be thoroughly representative of the actual population for effective classification:
- More demanding than most NN's
- Sparse set sufficient
- Erroneous samples and outliers tolerable

Adding and removing training samples simply involves adding or removing "neurons" in the pattern layer Minimal retraining required, if at all. At the training set increases in size, the PNN asymptotically converges to the Bayes optimal classifier. The estimated pdf approaches the true pdf, assuming the true pdf is smooth. The training process of a PNN is essentially the act of determining the value of the smoothing parameter, sigma.
Training is fast because Orders of magnitude faster than back propagation.

Determining Sigma
Educated guess based on knowledge of the data .It estimate a value using a heuristic technique. Systematic testing of values for sigma is done over some range. Bound the optimal value to some interval and shrink the interval. Jackknifing is used to grade the performance of each "test" sigma.

2. METHODOLOGY

The datasets that were used in the paper were taken from a valid database. Here the total numbers of instances are 1900 in number. All the instances are divided into six attribute and again each attribute has different features as described in the given table:

Attribute	Number of Features
Buying	4
Maintenance	4
Doors	3
Persons	3
Lug-boot	3
Safety	3

After identifying the attributes, the total numbers of instances are divided into four classes. The number of instances and the percentage of individual classes are given in the following table:

Class	Number of Instances(I)	(I %)
Unacc	1400	73.684
Acc	400	21.052
Good	60	3.157
V-good	40	2.105

I% is calculated as: (I/N × 100), where N is the total number of instances and I is the number of instances under a class. Now applying PNN on the above data the error can be checked.

3. IMPLEMENTATION

Spread Factor of Probabilistic Neural network (SFPNN) is of the range starting from 0.1 to 0.9.Here all the instances are checked through SFPNN and it was found that the output matched 100% up to 0.6 spread factor values. But for spread factor value 0.7, 0.8, 0.9 few values are not matched and it is given in the following table:

Class	Number of Instances(I)	(I%) In PNN	No. of Spread factor error		
			0.7	0.8	0.9
Unacc	1400	73.684			3
Acc	400	21.052	3	10	22
Good	60	3.157	2	24	33
V-good	40	2.105	1	9	21

4. RESULTS

Now the percentage of error in PNN for spread factors 0.7, 0.8 and 0.9 are given in the following table:

Class	(I %)	(I%)In PNN	(%) of Error
Unacc	73.684	73.510	0.174
Acc	21.052	19.779	1.273
Good	3.157	1.847	1.310
V-good	2.105	0.889	1.216

Training and Testing data size:

Input Data Size	Training Set	Testing set	Number of Classes	Attribute Type
6 × 1900	6 × 200	6 × 200	4	Real

5. DISCUSSION

In the research work, by applying PNN in the data we got 0.174% of error for the first output i.e “Unacc”, 1.273% error for the second output i.e “Acc”, 1.310% of error for the third output i.e “Good” and 1.216% of error for the fourth output i.e “V-good”. For 0.9 spread factor, we got 3 number of errors in “Unacc”, 22 number of errors in “Acc”, 33 number of errors in “Good” and 21 number of errors in “V-good”. For 0.8 spread factor, we got zero errors in “Unacc”, 10 number of errors in “Acc”, 24 number of errors in “Good” and 9 number of errors in “V-good”. For 0.7 spread factor, we got zero errors in “Unacc”, 3 number of errors in “Acc”, 2 number of errors in “Good” and one of error in “V-good”. But for spread factor 0.1 to 0.6 there was no error. We had also trained the model by taking 300 training set data and test the result. In testing we found that the result of the test data was 100% for spread factor value 0.1 to 0.6 and it gave error in spread factor value 0.7 to 0.9.Therefore we conclude that the testing was done successfully. The main advantage of PNN is fast learning as it is a one-pass training algorithm. It does not require an iterative training process. The training time is just the loading time of the training matrix. Thus, it requires many training samples to span the variation in the data and all these to be stored for the future use. However, there is only one disadvantages that there is no intuitive method for choosing the optimal smoothing factor. For the best performance of the PNN, the value of spared parameter is determined through trial and error method.

6. CONCLUSION

We have presented a new Probabilistic Neural Network approach for VEHICLE classification based on their feature studied from the market. The model was also tested successfully by using training set data. The result has shown very less amount of error to the actual output data. Therefore A PNN using probability density function and Bayes’s decision rule can be used as a remedy for classification problems with full compound of Training and Learning. Further other soft computing tools can be used for error comparison.

References

[1.] [1]PV NAGESWARA RAO, T UMA DEVI, DSVGK KALADHAR, GR SRIDHAR, ALLAM APPA RAO” A PROBABILISTIC NEURAL NETWORK APPROACH FOR PROTEIN SUPERFAMILY CLASSIFICATION” in Journal of Theoretical and Applied Information Technology,Vol-6,No-1@2009, pp.101-105.

[2.] Rojalina Priyadarshini; “Functional Analysis of Artificial Neural Network for Dataset Classification”, Special Issue of IJCCT Vol. 1 Issue 2, 3, 4; 2010 for International Conference [ACCTA-2010], 3-5 August 2010.

[3.] Sucheta Chauhan*, Prof. Prema K. V. “Vehicle Classification Using Artificial Neural Network” in International Journal of Scientific and Research Publications, Volume 2, Issue 12, December 2012 1 ISSN 2250-3153

[4.] Guoqiang Peter Zhang “Neural Network for Classification- A Survey 2000” IEEE Transactions on systems, man and cybermetics- part c: applications and reviews, Vol 30, No 4.

[5.] R. Furferi, L. Governi; “Neural Network based classification of vehicle seat fabrics” *International Journal of Mathematical Models and Methods in Applied Sciences*; Issues 3, Vol. 5, 2011.

[6.] V. Arulmozhi, "Classification task by using Matlab Neural Network Tool Box- A Beginner's View" *International Journal of Wisdom Based Computing*, Vol. 1(2), August 2011.
[7.] Prema K.V.,"A Multilayer Neural Network Classifier", *Journal of Computer Society of India*, Volume 35, Issue no: 1, January-March 2005.
[8.] Jerzy Jackowski, ICSENG '05 Proceedings of the 18th *International Conference on System Engineering*, IEEE Computer Society Washington, DC, USA ©2005, pp. 212-217.
[9.] Linhui Liu, "Realization and Application Research of BP Neural Network Based on MATLAB", International Seminar on Future Bio Medical Information Engineering (FBIE 2008), pp. 130- 133.
[10.] Florin Gorunescu, Benchmarking Probabilistic Neural Network Algorithms,*Proceedings of International Conference on Artificial Intelligence and DigitalCommunications, 2006.*
[11.] D. F. Specht, "Probabilistic neural networks," Neural Networks, vol. 3, no. 1, pp. 109–118, 1990

AUTHORS

Madhusmita Mishra received the B.E. in CSE from BPUT and M.Tech. degree in IT from Jadavpur University in 2006 and 2010, respectively. Her research interest includes ANN, Data mining and Fuzzy classification. She has several publications in national/international journals and conference proceedings. She is now associated with C.V. Raman College of Engineering, Bhubaneswar, Odisha, India.

Amrut Ranjan Jena received the B.E. in CSE from Utkal University and M.Tech. degree in IT from Jadavpur University in 2004 and 2008, respectively. His research interest includes ANN , Data mining and Fuzzy classification. He has several publications in national/international journals and conference proceedings. He is now associated with C.V. Raman College of Engineering, Bhubaneswar, Odisha, India.

Dr.Raja Das is an Assistant Professor in the School of Advanced Sciences, VIT University, Vellore, Tamil Nadu, India. He received his MSc in Applied Mathematics from National Institute of Technology, Rourkela, India and PhD from Sambalpur University, Odisha. His area of research interest includes Mathematical Modelling, Optimization and Soft Computing. He has published more than ten research papers in the international journal and presented more than four articles at different international/national conferences.