GLOBAL JOURNAL OF ENGINEERING, DESIGN & TECHNOLOGY
(Published By: Global Institute for Research & Education)
www.gifre.org

Performance Analysis of Radial Basis Function Neural Network for Internet Traffic Classification

P.Kumaraguru, V.R.Prakash & M.Abraham
Assistant Professor, Hindustan University, Chennai, India

Abstract

The area of internet traffic classification has advanced rapidly over the last few years due to a dramatic increase in the number and variety of applications running over the internet. These applications include www, e-mail, P2P, multimedia, FTP applications, Games etc. Since traditional internet traffic classification techniques become ineffective for certain complicated applications which use dynamic port number instead of well-known port number and various encryption techniques to avoid detection by unauthorised third party, therefore Machine Learning (ML) techniques are developed to handle such problems in internet traffic classification. Neural Networks are also one of the important ML techniques. In this paper, Radial Basis Function Neural Networks (RBFNN) are employed for internet traffic classification which is a type of multilayer feed forward neural network. In this research work, Performance of RBFNN is analysed based upon accuracy, recall, number of hidden layer neurons and training time of network using large feature data set and reduced feature data set . This experimental analysis shows that RBFNN is an efficient technique for offline internet traffic classification for reduced feature data sets too.

Keywords:- Internet traffic classification, Radial basis function neural network, Machine Learning, Accuracy, Recall, Training Time, Features.

1. Introduction

Internet Traffic Classification is a rapidly spreading field for solving difficult network management problems for Internet Service Providers (ISPs) and other important issues [1], [2]. Variety of applications running over internet are www, e-mail, p2p, multimedia, FTP applications, interactive services, Games etc which lead to rapid increase in internet traffic. Classification of this internet traffic is necessary in order to solve ISPs network management and monitoring problems such as available bandwidth planning and provisioning, measure of QoS, identification of customer's use of particular application for billing, detection of indicators of denial of service attacks and any severe problem degrading the performance of network etc. Now a day, it is also being utilized by various governmental intelligence agencies from security point of view. Internet Traffic classification can be either offline or online. In online classification, analysis is performed while data packets flowing through the network are captured; but in case of offline classification technique, firstly data traces are captured and stored and then analysed later [2]. Traditionally, various internet traffic classification techniques have been based upon direct inspection of packets flowing through the network [1]. These techniques are payload based and port number based packet inspection techniques. In payload based technique , payload of few TCP/IP packets are analysed in order to find type of application which is not possible today because of use of cryptographic techniques used to encrypt data in packet payload and privacy policies of governments which do not allow any unaffiliated third party to inspect each packets payload. In port number based packet inspection technique, well-known port numbers are provided in header of IP packets which are reserved by IANA (Internet Assigned Numbers Authority) for particular applications e.g. port number 80 is reserved for web based applications. Unfortunately, this method also becomes ineffective due to the use of Dynamic port numbers instead of Well-known port numbers for various applications.

Now a day, Machine learning (ML) techniques are utilized effectively for internet traffic classification which is based upon supervised and unsupervised learning techniques [1]. Neural Network which is a massively parallel distributed network consisting of number of information processing units (Neurons) inspired by the way human brain work, also comes under the category of ML techniques.

This paper is based upon use of Radial Basis Function Neural Network (RBFNN) for internet traffic classification [10], [11]. RBFNN is a type of supervised multilayer feed forward neural network which consists of basis functions for each of its hidden layer and its output is weighted linear superposition of all these basis functions. In this paper, performance of RBFNN is analysed for two different training and testing data sets having different number of features contained in input samples. Performance of this network is evaluated on the basis of accuracy, recall, number of hidden layer neurons and training time [7], [8]. This paper shows that as the Radial Basis Function neural network gives good classification accuracy even by reducing the number of features of input samples in training and testing data sets to much extent and it also leads to reduced complexity and reduction in training time of RBFNN.

The rest of the paper is organized as follows: Section II gives introductory information about RBFNN for readers who are new to this field. Section III reveals certain information about data sets of internet traffic. Implementation and result analysis is given in Section IV. Some conclusions are given in Section V.

2. RADIAL BASIS FUNCTION NEURAL NETWORK

Forward neural network which uses radial basis functions $\Phi m(.)$ as activation functions for each hidden layer neuron [10]. The output of RBFNN is weighted linear superposition of these basis functions. Consider training data set of input-target pairs T = {Xk,dk} k=1,2,3…n. In this network shown in figure 1,weights are fixed for input-hidden layer interconnections. While weights for hidden-output layer interconnections are trainable. Each hidden layer neuron have a basis function Φm(.).For input vector n-dimensional input vector X^T = [x1, x2, x3, …....., xn] , the output of this network is given by following interpolation function as:

$$\sum_{i=0}^{m} Wi\ \Phi(\| X - Xi \|)$$

(1)

Where □ $(\| X - Xi \|)$ are M basis functions consisting of Euclidean distance between applied input X and training data point Xi (also known as Center point in Gaussian Basis Functions).

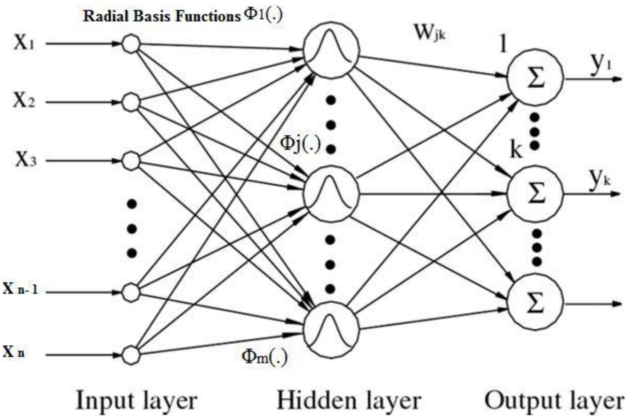


Fig. 1. Structure of Radial Basis Function Neural Network There are many basis functions such as Multiquadrics, Inverse Multiquadrics and Gaussian Function. Butcommonly used basis function in RBFNN is Gaussian function which is given as follows:

$$\square\ (X) = \exp\left(- \frac{\|\tilde{x}-\mu\|}{2\sigma^2}\right)$$

(2)

Where μ is the Center and σ is spread constant which have direct effect on the smoothness of interpolating mapping function F(X).

RBFNN can overcome some of the limitations of Back Propagation Neural Network and some other neural networks because it can use a single hidden layer for modelling any nonlinear function. Therefore, it is able to train data faster. While RBFNN has simpler architecture, it still maintains its powerful mapping capability. Due to the benefits of these characteristics, RBFNN is an interesting alternative technique for classification problem.

In this research work RBFNN with reduced number of hidden layer neurons i.e. centres of radial basis functions are used as compared to number of input because numbers of internet traffic inputs for training are very large. This makes RFBNN more generalized in order to give more accurate classification of internet traffic data.

3. Internet Traffic Data Set

In this research work, we have used a data set of 10,193 samples for two different cases of input features. [5], [6]. Each data sample belongs to a particular internet application. These applications for present data set are of twelve types which are given as: WWW, Mail, FTPCONTROL, FTP-PASV, Attack, P2P, Database, FTPData, Multimedia, Services, Interactive and Games.

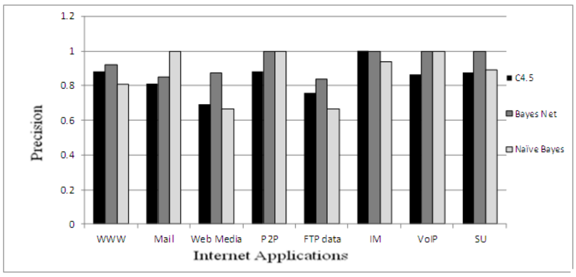
For first case, we have trained RBFNN using a data set of 10,193 samples, where each input sample consists of 248 features to characterize each particular application. These features mainly include inter packet arrival times (max., min., median, mean, variance, first quartile, third quartile etc), total number of packets (server to client and client to server), total number of bytes on the wire and in IP packet (max., min., median, mean, variance, first quartile, third quartile etc), control signals, bandwidth, duration, FFTs of various features, server port, client port and many other features. After the network is being trained, then out of the data set of 10,193 data samples, 2548 input data samples are being used for testing purpose and to obtain the classified outputs.

For second case, we have trained RBFNN using a data set of 10193 samples again, where each input sample consists of only 48 features to characterize each particular application. These features mainly include Inter packet arrival times (min., max., mean and variance), total number of packet byes on the wire and in IP packet (max., min.,mean, variance), bandwidth features, duration etc. After the network is being trained usind this reduced feature data set, 2548 data samples of same data set are used for testing purpose and to obtain classified outputs.

V. Table : Classification Accuracy and Training Time of Five ML Classifiers for Dataset $2_{consistency}$

ML Classifiers	**MLP**	**RBF**	**C4.5**	**Bayes Net**	**Native Bayes**
Classification Accuracy (%)	34.375	75	81.875	85	70
Training Time (Seconds)	3.97	35.98	0.31	0.06	0.01

Table show the performance of five ML classifiers in terms of classification accuracy and training time. It is clear from these tables and figure 5 that classification accuracy of all ML algorithms is maximum and training time is reduced to much extent in case of both reduced feature dataset. On the basis of comparative analysis of two FS algorithms, it is revealed from table II and III that Consistency based FS algorithm have only 8 features characterizing each application sample which make it inefficient for effective IP traffic classification irrespective of its smaller training time for all ML algorithms as compared to that of Correlation based FS algorithm. So, from these results it is obvious that Correlation based FS algorithm is a better choice for feature reduction in order to make IP traffic classification real time compatible.



Conclusion

In this paper, we have first designed a RBF Neural Network. Then this neural network is trained by using given full feature and reduced feature data sets of training data samples. Then for both the cases, the performance of RBFNN is evaluated based upon increase in number of hidden layer neurons and training time of RBFNN. Results show that for producing better accuracy using RBFNN, number of features in training and testing inputs should not be very high. It should include important and meaningful features to represent various internet applications. With increase in number of hidden layer neurons in RBFNN, it gives very high classification accuracy. But this high degree of accuracy is obtained at the expense of increase incomplexity of RBFNN due to increase in number of hidden layer neurons and very large time taken by this network for training. Thus this research work shows that RBFNN is an effective machine learning technique for offline internet traffic classification.

References

[1] T. Nguyen and Grenville Armitage. "A Survey of Techniques for Internet Traffic Classification using Machine Learning," IEEECommunications Survey & tutorials, Vol. 10, No. 4, pp. 56-76, Fourth Quarter 2012.

[2] Arthur Callado, Carlos Kamienski, Géza Szabó, Balázs Péter Ger˝o, Judith Kelner,Stênio Fernandes ,and Djamel Sadok. "A Survey on Internet Traffic Identification," IEEE
Communications Survey & tutorials, Vol. 11, No. 3, pp. 37-52, Third Quarter 2012.

[3] Runyuan Sun, Bo Yang, Lizhi Peng, Zhenxiang Chen, Lei Zhang, and Shan Jing. "Traffic Classification Using Probabilistic Neural Network," in Sixth International Conference on Natural Computation (ICNC 2010), 2010, pp. 1914-1919.

[4] Luca Salgarelli, Francesco Gringoli, Thomas Karagiannis. "Comparing Traffic Classifiers,"ACM SIGCOMM Computer Communication Review, Vol. 37, No. 3, pp. 65-68, July 2007.

[5] Andrew W. Moore, Denis Zuev, Michael L. Crogan, "Discriminators for use in flow-based classification," Queen Mary University of London, Department of Computer Science, RR-05-13, August 2005.

[6] Wei Li , Marco Canini , Andrew W. Moore ,Raffaele Bolla, "Efficient application identification and the temporal and spatial stability of classification schema" Elsevier Journal, Computer Networks, Vol. 53, pp. 790– 809,23 April, 2009.

[7] Feng Qian, Guang-min Hu, Xing-miao Ya,"Semi-supervised internet network traffic classification using a Gaussian mixture model," International journal AEU. of Electronics & Communication, Vol. 62,pp. 557-554, 2008.

[8] Murat Soysal a, Ece Guran Schmidt, "Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison," Elsevier Journal, Performance Evaluation, Vol.67, pp. 451-467, 2010.

[9] Masatsugu Ichino, Hiroaki Maeda, Takeshi Yamashita, Kentaro Hoshi, Naohisa Komatsu,Kei Takeshita, Masayuki Tsujino, Motoi Iwashita and Hideaki Yoshino, "Internet Traffic Classification Using Score Level Fusion of Multiple Classifie,"
9th IEEE/ACIS International Conference on Computer and Information Science, 2009, pp.105-110.

[10] Satish Kumar, Neural Networks: A Classroom Approach, 6th edition, Tata McGraw – Hill Publishing Company Limited, New Delhi, 2008.

[11] Sheng Chen, Andreas Wolfgang, Chris J. Harris, and Lajos Hanzo, "Symmetric RBF Classifier for Nonlinear Detection in Multiple-Antenna- Aided Systems," IEEE Transactions on Neural Network, Vol. 19, No. 5, pp. 737-745, 2008.

[12] Kuldeep Singh and Sunil Agrawal, Internet traffic classification using RBF Neural Network, in International Conference on Communication and Computing technologies(ICCCT-2011), Jalandhar, India, February 25-26, 2011, paper 10, p.39-43.

[13] Kuldeep Singh and Sunil Agrawal. Comparative Analysis of five Machine Learning Algorithms for IP Traffic Classification, In International Conference on Emerging Trends in Networks and Computer Communications (ENCTT-2011), Udaipur, Rajasthan, India, April 22-24, 2011.