2017 13th International Conference on Computational Intelligence and Security

A method of improved CNN traffic classification

Huiyi Zhou[1,3,4], Yong Wang[2,3,4], Xiaochun Lei[2,3,4], Yuming Liu[3,4]

1. School of Information and Communication,
Guilin University of Electronic Technology, Guilin Guangxi 541004, China
2. School of Computer Science and Information Security,
Guilin University of Electronic Technology, Guilin Guangxi 541004, China
3. Guangxi Cooperative Innovation Center of Cloud Computing and Big Data,
Guilin University of Electronic Technology, Guilin Guangxi 541004, China
4. Guangxi Colleges and Universities Key Laboratory of Cloud Computing and Complex Systems,
Guilin University of Electronic Technology, Guilin Guangxi 541004, China
(E-mail: 1276129997@qq.com; ywang@guet.edu.cn; lxc8125@guet.edu.cn; suiyanlovetian@foxmail.com)

***Abstract*—A traffic classification algorithm based on improved convolution neural network is proposed in this paper. It aims to improve the traditional traffic classification method. Firstly, the min-max normalization method is used to process the traffic data and map them into gray image, which will be used as the input data of convolution neural network to realize the independent feature learning. Then, an improved structure of the classical convolution neural network is proposed, both of the parameters of the feature map and the full connection layer are designed to select the optimal classification model to realize the traffic classification. Compared with the traditional classification method, the experimental results show that the proposed CNN traffic classification method can improve the accuracy and reduce the time of classification.**

Keywords-network traffic classification; convolutional neural network; normalized; feature selection component

I. INTRODUCTION

With the arrival of big data era and the rise of new network applications, the composition of the network becomes more and more complex. There is an increasing demand for businesses to make development planning evaluation from the information which is carried out from the growing flow of traffic data. As one of the key technologies in network management and network security, network traffic classification can not only optimize the network configuration, to reduce the network security risks, but also provide better service quality according to user behavior analysis. Currently, network traffic classification based on machine learning has become the hotspot, which includes wake learning algorithm and deep learning algorithm [1]. Wake learning algorithm are mainly used by researchers such as Cao J et al. [2] proposed a real-time accurate SVM training model called the SPP-SVM and Tong D et al. [3] proposed online traffic classification algorithms and architecture based on C4.5 decision tree algorithm and entropy MDL (Minimum Description Length) discrete algorithm, etc. Deep learning has achieved good results in image recognition [4], speech recognition [5], the audio processing [6] and natural voice processing [7], etc. but in network traffic classification it is rarely mentioned. In this paper, we propose an algorithm based on convolutional neural network called Min-Max Normalization Convolutional Neural Networks(MMN-CNN). This algorithm can extract feature implicitly by self-learning from the training data set which can avoid the artificial feature extraction and solve the problem of difference that comes from feature selection in of different classification algorithms. To this end, the contributions of this paper are as follows: (1) A method of Min-Max Normalization Convolutional Neural Networks applied to traffic classification; (2) A method of normalized data pre-processing. The rest of this paper is organized as follows. In Section II, it will cover a method of data preprocessing according to the characteristics of convolutional neural network learning data, and convert the traffic data to grayscale images; Section III will explain the details of the MMN-CNN architecture. In Section IV, the testing and evaluation results of the algorithm are displayed. Conclusion and future work are drawn in Section V.

II. ALGORITHM OF MMN-CNN

Convolution neural network can extract feature implicitly by self-learning from the training data set. According to this characteristic, the data pre-processing is as follows.

Dimensional relationship is the root of comparability. In this way, in order to eliminate the dimensional relationship of network traffic data, this paper normalizes the corresponding features in the data set. The specific process is as follows:

It assumed that the dataset may be represented as a matrix of n rows and m columns:

$$T = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} \end{bmatrix} \tag{1}$$

According to the features, the division can be expressed as (2) matrix:

0-7695-6341-4/17/31.00 ©2017 IEEE
DOI 10.1109/CIS.2017.00046

$$B_i = [A_{1i}, A_{2i}, A_{3i}, \cdots A_{ni}]^T \tag{2}$$

where B_i represents the all data value of the i feature, then the matrix T can be expressed as:

$$T = [B_1, B_2, B_3, \cdots, B_m] \tag{3}$$

Each column of the normalized matrix yields B_i' :

$$B_i' = B_i \Big/ \max\{A_{1i}, A_{2i}, A_{3i}, \cdots A_{ni}\} \tag{4}$$

Then the quantized matrix can be expressed as:

$$T' = [B_1', B_2', B_3', \cdots, B_m'] \tag{5}$$

After normalized the traffic data set, we construct the proper matrix dimension for MMN-CNN according to the number of features in the data set. In this paper, we used the dataset provided by Moore which includes 249 statistical characteristics of network flow. So we define the matrix dimension as 16, while the number of statistical characteristics in Moore dataset is less than the number of the constructed matrix elements. According to the number of features in the dataset, we construct the matrix dimension, which is suitable for the study of convolution neural network. As the feature dimensions are less than the number of matrix elements, so the corresponding fill operation is last 0 matrix.

Then, each element in the matrix is treated as a pixel, and the value in the matrix is the gray scale of the pixel. The gray value of the picture is proportional to the value. As the image become whiter the grayscale vale will get bigger, on the contrary, an image will get blacker as the grayscale value get smaller. Each network traffic will be converted into a corresponding gray image; Fig I stands for one of the flow data corresponding to the gray image.

Figure I Network traffic visualization

III. CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE

As an important model of deep learning, convolutional neural network was originally designed to solve the problem of image recognition and had achieved good results, but few people use it in the field of network traffic classification. Both the image and the traffic data are made up of numerical values, this paper proposes a method of feature learning and classification based on convolutional neural network for traffic classification.

One of the most representative experimental systems in early convolution neural networks is the LeNet-5 model [8], designed by Yann LeCun in 1998 for handwritten numeral recognition. The LeNet-5 model has seven layers, the input layer, the first convolutional layer, the first pooling layer, the second convolutional layer, the second pooling layer, the full-connected layer and the output layer. In view of the characteristics of network data flow, this paper improves the traditional LeNet-5 model as follows:

(1) The data set is analyzed to obtain the appropriate grayscale image, and then the input layer of the network is designed as the pixel matrix of 16* 16. The specific details are described in the second part of this article.

(2) The purpose of network traffic classification is to classify different traffic application types in the data set, including 12 different application types. In this paper, the output layer of the traditional LeNet-5 model has changed the number of neurons from 10 to 12. So the neuron's number of the output layer of the traditional LeNet-5 model changes from 10 to 12.

(3) The performance of traffic classification is mainly related to the hierarchical structure of the convolutional neural network. We design six different types of convolutional neural network based on the basic hierarchical structure of LeNet-5 network to choose which performs better. As it is shown in Table I below:

TABLE I. SIX DIFFERENT NETWORK STRUCTURES OF CONVOLUTION NEURAL NETWORK

NUM	Convolution		Subsampling		Convolution		Subsampling		Full connection	
	Filter size	Output	Pool stride	Output	Filter size	Output	Pool stride	Output	Filter size	Output
1	32*(3*3)	32*(16*16)	2*2	32*(8*8)	64*(5*5)	64*(8*8)	2*2	64*(4*4)	256*(4*4)	256*1
2	32*(3*3)	32*(16*16)	2*2	32*(8*8)	64*(5*5)	64*(8*8)	2*2	64*(4*4)	128*(4*4)	128*1
3	16*(3*3)	16*(16*16)	2*2	16*(8*8)	32*(5*5)	32*(8*8)	2*2	32*(4*4)	256*(4*4)	256*1
4	16*(3*3)	16*(16*16)	2*2	16*(8*8)	32*(5*5)	32*(8*8)	2*2	32*(4*4)	128*(4*4)	128*1
5	8*(3*3)	8*(16*16)	2*2	8*(8*8)	16*(5*5)	16*(8*8)	2*2	16*(4*4)	256*(4*4)	256*1
6	**8*(3*3)**	**8*(16*16)**	**2*2**	**8*(8*8)**	**16*(5*5)**	**16*(8*8)**	**2*2**	**16*(4*4)**	**128*(4*4)**	**128*1**

As for the problems found in the design process above, this paper mainly improved in the following aspects: Firstly, the size of the data samples of the input layer is 16 * 16, and the calculated amount is small. To avoid the image of the edge of the information loses too fast, the boundary of the current matrix is filled with 0, which guarantees that the size of the matrix propagated by the convolution layer is consistent with the current matrix, and the moving step of the convolution kernel is set as 1. When the feature map of convolution kernel is larger than 64, it is easy to get the local optimum for a seven layers convolutional neural network. So the feature map of the convolution kernel should not be more than 64.

Two different pooling layers were used to compare the performance of six convolutional neural network designed in this paper. The two kinds of pooling layer commonly used, one is max-pooling, and the other is mean-pooling. The results of the experiment are shown in Table II and Table III.

TABLE II. TEST RESULTS OF THE MAX-POOLING METHOD

Num	1	2	3	4	5	6
Overall Accuracy (%)	86.9	98.9	99.18	98.98	99.32	**99.30**
Testing Time(s)	20.13	18.41	12.01	11.30	6.29	**6.03**

TABLE III. TEST RESULTS OF THE MEAN-POOLING METHOD

Num	1	2	3	4	5	6
Overall Accuracy (%)	98.30	98.33	98.42	98.40	98.22	**98.58**
Testing Time(s)	20.69	18.67	11.55	10.77	6.22	**5.99**

.

The traffic classification method is divided into two stages based on MMN-CNN: training and testing. For the real-time traffic classification task to be carried out in the later period, which the training analysis can be offline, that it does not affect the final classification effect. Therefore, only the testing time of each algorithm is compared and analyzed in this paper.

The experimental results show that with the increase of feature map number, the accuracy of traffic classification varies little, but the testing time of traffic classification increases obviously. When the max-pooling method is adopted, the model 1 shows the local optimal, while the classification accuracy is improved greatly with the decrease of the feature map. Compared with the two test results, the model 6 has achieved the best results in classification accuracy and testing time. Therefore, this paper selects it as the optimal convolutional neural network. The overall accuracy and testing time of the convolutional neural network model are 99.30% and 6.03s, respectively. The structure of the model 6 is shown in Figure II.

In convolution layer C1, the input data, which is filled with 0 to get a 16-dimensional square matrix, are convoluted with 8 filters of size 3*3 to get 8 feature maps of size 16*16. In subsampling layer S2, the feature maps of size 16*16 is subsampled by a 2*2 window and 8 feature maps with size 8*8 is obtained. In convolution layer C3, the output data from S2 is filled with 0 and convoluted with 16 filters of size 5*5. The output data in C3 are 16 feature maps of size 8*8. In convolution layer S4, the output data from previous layer is subsampled by a 2*2 window to generate the 16 new feature maps with size 4*4, which are mapped into a vector of size 128*1 in full connection layer C5. The soft-max classification is used in the output layer, which exports 12 types of applications.

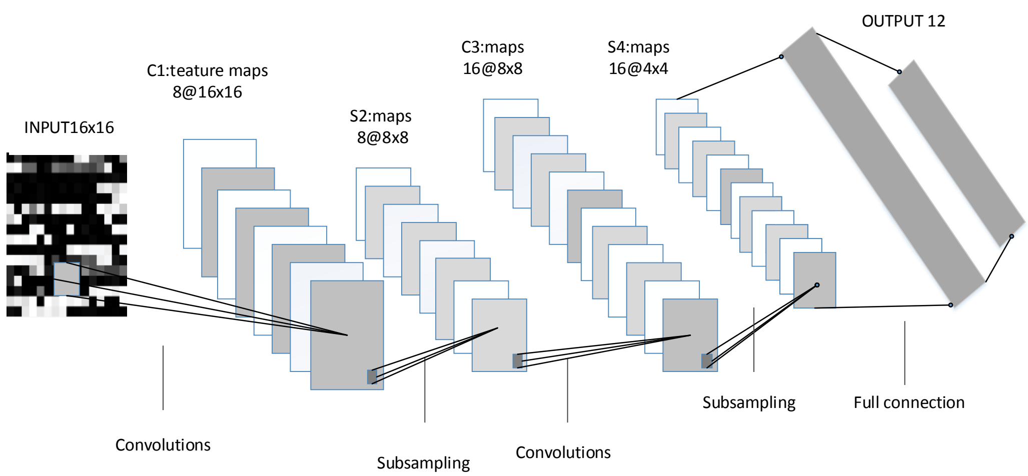

Figure II The convolution neural network structure is designed in this paper

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. *Experimental data*

In this paper, the traffic data used in this experiment is the Moore dataset, which consists of 12 application types. And each flow contains 249 kinds of flow feature, in which the last feature corresponds to the category that it belongs to. The number and proportion of each class in network traffic of the Moore dataset are summarized in Table IV.

TABLE IV. DISTRIBUTION OF THE MOORE DATASET

Category	Amount	Proportion (%)
WWW	328092	86.906
MAIL	28567	7.567
FTP-DATA	5797	1.536
FTP-PASV	2688	0.712
FTP-CONTROL	3054	0.809
SERVICES	2099	0.556
DATABASE	2648	0.701

P2P	2094	0.555
ATTACK	1793	0.475
MUITIMEDIA	576	0.153
INTERACTIVE	110	0.029
GAME	8	0.002
Total	377526	100.000

The Moore dataset is divided into two sets: training dataset and testing dataset. The proportion of each class in the two datasets is consistent with the Moore dataset. Then we randomly select 100000 data as the testing dataset and the others as the training dataset.

B. Evaluating indicator

There are two main indexes for evaluating the experimental results. One is the accuracy of the classification, including the accuracy and reliability of each type of class and the overall accuracy of all class; the other is the testing time of the convolutional neural network model.

Specifically, the accuracy of the class i in this experiment is: $A_i = \frac{TP_i}{TP_i + FN_i}$; the credibility of the class i is: $T_i = \frac{TP_i}{TP_i + FP_i}$; the overall accuracy is: $OA = \frac{\sum_{i=1}^{m} TP_i}{\sum_{i=1}^{m} (TP_i + FH_i)}$.

Where TP_i (True Positive) refers to the number of samples correctly predicted by the classification model in the actual type i . FN_i (False Negative) refers to the number of misjudging as other types of samples for the classified model. FP_i (False Positive) are misjudged as the sample number of i , actually they are non-i in the classified model.

C. Experimental results and analysis

The method of dimensionality reduction commonly used include Principal Component Analysis(PCA), Gauss Random Projection(GRP) and Sparse Random Projection (SRP), etc. Compared with Gauss Random Projection, Sparse Random Projection can guarantee the quality of dimension reduction and improve the efficiency of memory. So that we chose the methods of PCA and SRP to reduce the feature of each flow data in Moore dataset from 249 to 16 firstly. Then the data preprocessing method adopted by Kou [9] is used to normalize the flow feature in vector. Finally, the flow feature vector is transformed into Euler matrix by using the Euler distance between the features and visualized as grayscale images. In this paper, the optimal convolution neural network model selected from part III is used to compare the two-different gray scale image mapping methods according to each parameter of the evaluation index. The experimental results are shown in Table V, Table VI and Table VII.

TABLE V. THE OVERALL ACCURACY AND TEST TIME IN DIFFERENT ALGORITHMS

Method	PCA	SRP	MMN-CNN
Overall Accuracy (%)	96.09	97.23	**99.30**
Testing Time(s)	6.03	5.92	**6.03**

TABLE VI. THE ACCURACY OF CLASS IN DIFFERENT ALGORITHMS

Method	WWW	MAIL	F-D	F-P	F-C	SERV	DB	P2P	ATT	MULT	INT	GAME
PCA(%)	99.57	91.13	73.15	2.81	75.34	91.2	0	17.45	65.13	0	0	0
SRP(%)	99.74	94.79	93.04	48.53	75.59	96.05	0.43	23.92	**67.23**	0	0	0
MMN-CNN(%)	**99.83**	**99.39**	**99.8**	**96.63**	**95.92**	**98.38**	**82.48**	**90.45**	66.32	**54.25**	0	0

TABLE VII. THE CREDIBILITY OF CLASS IN DIFFERENT ALGORITHMS

Method	WWW	MAIL	F-D	F-P	F-C	SERV	DB	P2P	ATT	MULT	INT	GAME
PCA(%)	96.97	90.56	93.83	29.41	73.79	90.07	0	71.85	92.81	0	0	0
SRP(%)	98.07	92.91	94.33	55.18	92.18	96.75	23.08	70.00	94.96	0	0	0
MMN-CNN (%)	**99.61**	**98.61**	**98.90**	**92.24**	**96.16**	**96.81**	**98.81**	**83.11**	**99.06**	**80.58**	0	0

As shown in the above experimental results, compared with the preprocessing of PCA and SRP, the MMN-CNN algorithm change dataset into gray-scale pictures directly, which can achieve a better performance in accuracy and credibility from both the overall and individual perspective.

Therefore, the method of MNN-CNN algorithm can reduce the workload and improve the accuracy of classification. From the TABLE VI and the TABLE VII, it finds that the accuracy of INTERACTIVE and GAME is zero, indicates that the convolution neural network is not ideal for the data types which number is less. However, with the rapid development of network technology and the number of network traffic increases with EB level, this problem will be solved well.

V. CONCLUSION

This paper first introduces the method commonly used in network traffic classification. Then we proposed an algorithm called MMN-CNN to extract more abstract features from traffic dataset. And we designed six different types of convolutional neural network to choose which is the best one in network traffic classification. Compared with the other two methods, the MMN-CNN can avoid the accumulation of errors caused by artificial feature extraction and improve the effect of traffic classification.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (Nos. 61163058, 61662018, 61661015). Guangxi Natural Science Foundation of China (No. 2016GXNSFAA380153), Guangxi Colleges and Universities Key Laboratory of cloud computing and complex systems (No. 14103)

REFERENCES

[1] Y. Lecun, Y. Bengio and G. Hinton, "Deep learning," Nature, vol. 512, May. 2015, pp. 436-444, doi: 10. 1038/nuture14539.

[2] J. Cao, Z Y. Fang, G N. Qu, H Y. Sun and D. Zhang, "An accurate traffic classification model based on support vector machines," Networks, vol. 27, Jan-Feb 2017, doi: 10. 1002/nem. 1962.

[3] D. Tong, Y R. Qu and V K. Prasanna," Accelerating Decision Tree Based Traffic Classification on FPGA and Multicore Platforms," IEEE Transactions on Parallel & Distributed Systems, June. 2017, pp. 1-1, doi: 10. 1109/TPDS. 2017. 2714661.

[4] K M. He, X Y. Zhang and S Q. Ren, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," IEEE International Conference on Computer Vision, 2015, pp. 1026-1034, doi: 10. 1109/ICCV. 2015. 123

[5] S N. Sun, B B. Zhang, L. Xie and Y N. Zhang, "An unsupervised deep domain adaptation approach for robust speech recognition," Neurocomputing, vol. 257, Sep. 201and7, pp. 79-87, doi: 10.1016/j.neucom.2016.11.063

[6] D S. Williamson, D L. Wang, "Time-Frequency Masking in the Complex Domain for Speech Dereverberation and Denoising," IEEE/ACM Transactions on Audio Speech & Language Processing, vol. 25, Jul. 2017, pp. 1492-1501, doi: 10.1109/TASLP.2017.2696307

[7] Kim, Yoon, "Convolutional Neural Network for Sentence Classification," Association for Computational Linguistics (ACL), 2014, pp. 1746-1751, doi: 10.3115/v1/d14-1181

[8] Y. Lecun, L. Bottou, Y. Bengio, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, Nov. 1998, pp. 2278-2324, doi: 10.1109/5.726791

[9] G. Kou, G M. Tang, S. Wang, H T. Song and Y. Bian, "Using deep learning for detecting BotCloud," Journal on Communications, vol. 37, Nov. 2016, pp. 114-128, doi: 10.11959/j.issn.1000-436x.2016228