Research on Multi-layer Classification Method of Network Traffic Based on Machine Learning

Bin Wang

1 College of Electronics Engineering, Naval University of Engineering, Wuhan, China

*Corresponding author e-mail: wbnavy@163.com

Abstract. According to the network traffic characteristic group, this paper designs a kind of network traffic classification with coarse and fine two layers. The first layer based on the experience attribute selection and the fast multilayer clustering method, constructs the unsupervised classifier, which realizes the coarse classification of the network traffic dataset and forms the data subset; the second layer, based on the selected subset of data, constructs a strong supervised classifier, which realizes the fine classification of the network traffic data set, according to the ensemble learning method. Theoretical analysis and experiments show that this method can not only improve the accuracy of network traffic classification, but also reduce the time of classification greatly, which is important to improve the accuracy of small class classification of imbalanced network traffic data.

1. Introduction

To realize the quantitative management of network and accurate operation and maintenance need to grasp the distribution of network flow accurately; network traffic classification technology plays a very important role in modern network operation and maintenance management system. In computer networks, network traffic is also known as data flow, which is a sequence of messages with the same five-tuple (consisting of five elements: source IP, source port, destination IP, destination port, transport layer protocol), including one-way flow and two-way flow.

According to a classification model, from the collection of data streams generated by various application layer protocols, the process of identifying a data stream belonging to a certain class of application layer protocol is the classification of network traffic. there are three main classification methods, such as port-based classification, deep packet inspection (DPI), and deep flow inspection (DFI). The port-based classification method relies on the analysis of port numbers in TCP or UDP packets, because the well-known port number (IANA [1]) is mapped to identify different types of applications. The experiments in the literature 2 show that the classification method based on well-known ports is only 69.27% accurate. The DPI method can only identify the known non-encrypted traffic, but cannot classify other unknown traffic and encrypted traffic, and the directly analyses to the content of the application layer will also bring privacy violations and security issues. The DFI method based on machine learning can classify high accuracy of network traffic with dynamic port and content encryption, and is easy to realize the network real-time traffic classification. As in literature 3 using Bayesian classification method, literature 4 using improved C4.5 classification method, literature 5 using PSO-mixed kernel function of SOM, literature 6 using iterative tuning SVM, literature 7 using random forest algorithm, not only have achieved the classification of network traffic, and but also obtained a high overall classification accuracy rate.
The distribution of each protocol flow in a network is usually imbalanced. The number of flow samples of one or several network protocols is far more than the number of streams in other network protocols, this protocol flows are called "large class" flows, and other protocol flows are called "small class" flows. The uneven distribution of network data is ubiquitous, which is one of the main problems that plague the method of traffic classification based on machine learning. The network traffic for P2P application account for 60%~80% of the total ISP business, and even more than 90% late at night. In the Environment of network protocol flow imbalance, the classifier is easily drowned by the large class stream in the stream sample, which makes the recognition rate of small class flow low and even cannot be discovered. However, the actual application value of small and medium-class flow is no less than the large class of flow, such as the flow of intrusion in the proportion of the total traffic is very small, but extremely important. Therefore, it is very important to study the flow classification algorithm based on machine learning in the imbalanced environment of network protocol flow.

2. Related Research

Currently there are two main ways to solve imbalanced data classification based on machine learning method including the solutions from data level or algorithm aspects. From the data aspects, the imbalance of sample distribution can lead to the scarcity of rare sample, and the under-sampling or over-sampling method can enhance the characteristic attribute of small class data in the dataset. The under-sampling method can reduce the number of large class training samples, and the over-sampling method can increase the number of small class training samples. Although the over-sampling method has achieved good results on some datasets, such as the borderline-SMOTE method proposed in the literature, which is through the over-sampling of the boundary small classes to obtain the improve of the boundary recognition rate. However, there are some drawbacks to this approach, because the method does not add any new data, but just to repeat some samples or add some artificially samples, which has increased the training time. More dangerously, the over-sampling copies the small class samples, or generates some artificial ones around them, which has caused the classifier to focus too much on the small class samples and lead to overfitting. Some samples of the large class will be removed in the under-sampling, which can reduce the imbalance. Such as the random under-sampling method presented in literature, the method randomly selects some large class samples as the training data set, which is possible to remove the important information. Although there are some heuristic under-sampling methods that only remove redundant samples or noise samples, in most cases these samples are only a small part, so the adjust to the imbalance dataset by the under-sampling method is quite limited.

From the algorithm aspects, ensemble learning, cost-sensitive matrix or feature selection method is mainly adopted. By synthesizing different learning methods on different data sets, the ensemble learning method improves the detection accuracy of the imbalanced datasets. Different subsets is trained by extracting different features to obtain different classifiers, the different classifiers take the absolute majority and the relative majority to decide the best one, finally the best classifier is tested by imbalanced datasets. By increasing the penalty weight of the small class data in the imbalanced data, the cost-sensitive matrix is used to reduce the probability of the wrong division in the small class data. Using the adaptive optimization algorithm in machine learning, the wrong cost of imbalanced data is applied to the sample weight adjustment, so that the following classifier learns to focus on the samples of the few important classes that have been wrongly divided. By means of empirical method or automatic comparison of information gain, the feature selection method is used to obtain the characteristic attribute group which is advantageous to improve the small class data recognition rate. Because of the deletion of the feature that interferes with small class data recognition, not only the detection accuracy is improved, but also the detection efficiency is improved.

3. Algorithm Principle & Implementation

Due to the ubiquitous existence of imbalanced data distribution in network traffic, this paper will synthesize the advantages from the data & algorithm aspects, and form a combinatorial algorithm to solve the network traffic classification, especially to improve the classification precision of small class
data. From the data aspects, this paper uses the clustering method to construct several subsets of imbalanced data, and by comparing the distance between data and centroid, automatically optimizes the number of data subsets. Before that, the feature selection is performed by experience, and redundancy and disturbance features are removed to improve the speed and precision of data subset generation. From the algorithm aspects, the feature attribute can be further optimized by using the information gain sort method. The ensemble learning methods are mainly relied on to combine the weak learners to the strong learner, which can improve the classification accuracy. The cost-sensitive matrix is configured, which can increase the error-cost of small class flow and improve the detection precision of small class flow in imbalanced network traffic.

3.1. Coarse Classification

3.1.1. Principle. The coarse classification consists of "Experience Feature Selection" and "XMeans".

A. Experience Feature Selection

Experience feature selection is based on the characteristic attributes of network traffic. According to different applications, different network features to analyze and excavate can be selected by the experience method. The classification of network traffic is based on the different features of various streams. The main features include:

a) Protocol, source/destination port.
b) Number of packets in a flow.
c) Total size of a flow.
d) Average, median, maximum, minimum, and standard deviation of packet size in a flow.
e) The duration of the connection.
f) Number of FIN, SYN, ACK, URG flag in a flow.

B. XMeans

Based on the specified range of data subsets number, XMeans selects the optimal number K of data subsets, through the calculation of the BIC score, and the iterative algorithm uses the two-class clustering. The principle is: with a given range of K value, starting from the Kmin Value, KMeans operator is repeated call for two classification. Each time the cluster is completed, its BIC score is calculated, and if the BIC score of the parent and child class differ greatly, the two classes are separated, then there is k=k+1. In the iteration, the new class centroid is the last (K-1) class centroid and the divided two subclass class centroid. When K=Kmax, the preferred subset of data is returned. The XMeans clustering process is described in Fig. 1.

![Figure 1. XMeans clustering process.](image)

The calculation method of the BIC Score is described in: \[ \text{BIC}(M_j) = \hat{\ell}_j(D) - \frac{p_j}{2} \log n \]

\( \hat{\ell}_j(D) \) indicates the possibility that the data is divided into j class, \( p_j \) indicates the number of free parameters of the model.
3.1.2. Implementation. The pseudo-code of the coarse classification is:

a) According to the network traffic characteristic, experience is used to select the features of the dataset, such as A1, A2, ..., Ak, ...

b) Estimate the maximum Kmax and the minimum Kmin of clusters’ number.

c) Remove interference and irrelevant feature attributes, leaving related feature attributes Akeys;

d) Carry out XMeans clustering, and generate data subsets C1, C2, ..., Ck, ...

e) Each data subsets are result of the coarse classification, and the “experience & clustering” model is the coarse classification model of the flow data.

f) The coarse classification can be used for initial filtering of the data.

3.2. Fine Classification

3.2.1. Principle. Based on the statistical characteristics of flow data, the fine classification is used to classify the subsets generated by the coarse classification. The fine classification consists of feature selection algorithm and ensemble learning algorithm based on cost-sensitive matrix. In general, in the coarse classification process, through the experience selection of feature parameters, redundant features have been removed, if further using the feature selection algorithm to choose, it is possible to destroy the original data characteristics, therefore, the general feature selection algorithm is only used without the experience selection. The cost-sensitive matrix can enhance the classification sensitivity of the small class flow, but at the same time, it reduces the precision of the small class flow, because many large class flows are divided into small class flows, which reduces the classification accuracy of the whole sample. The cost-sensitive matrix is used only in the cases where the recall of small class flows that need to be highlighted (e.g., the small class flow is the network attack flow, and so on).

The ensemble classifier accomplishes the classification task by ensemble multiple machine learning algorithms, which can obtain the better performance than the single classifier. There are two kinds of generative methods, one is boosting, where there is the strong dependence among the individual learners, and is generated by the serialization Method. Another algorithm is bagging and random forest, where there is loose coupling between individual learners, and is generated by parallelization Method.\footnote{24} \footnote{25} \footnote{26} The boosting algorithm first trains a base learner from the initial training set, then adjusts the distribution of training samples according to the performance of the base learner, which makes the training samples of the previous base learners’ mistakes be more concerned. The bagging algorithm generates multiple training sets and then trains a base learner based on each sample, using the voting method, and then combining the base learner. In this paper, the boosting algorithm is used as the basis of the ensemble learning, which is mainly based on its ability to form strong learners through multiple iterations, which is helpful to improve the accuracy of data classification, especially to improve the classification accuracy of imbalanced data, and to avoid partial overfitting. The boosting algorithm is segmented according to different features in the dataset, using different feature subsets to construct different classifiers, so that the training of different classifiers and the recognition process of new samples are based on the distribution of the data in different feature subspace, and the various classifiers constructed will have a great difference. Therefore, the various classifiers for a new sample will probably get different outputs, which can be ensemble with the fusion method to form a robust classifier ensemble system.

3.2.2. Implementation. The pseudo-code of the fine classification is:

\[
\text{Input: training set } D = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}
\]

Base learning algorithm S;

Number of training wheels L

for i=1, 2, ..., n do

\[ D_1(x_i) = 1/n; \] (1)

end

for t=1, 2, ..., L do

\[ h_t = S(D, \ D_t) \] (2)
\begin{equation}
E_t = \sum_{i=1}^{n} \frac{\text{cost}_t(x_i) \| (h_t(x_i) \neq y_i)}{\text{cost}_t(x_i)} \quad (\text{cost}_t(x_i) \text{is punitive factor for misjudging } y_i \text{ as } y_t)
\end{equation}

\begin{equation}
\alpha_t = \frac{1}{2} \ln \left( \frac{1 - E_t}{E_t} \right) \quad n 
\end{equation}

for \( i = 1, 2, \ldots, n \) do

\begin{equation}
D_{t+1}(x_i) = \frac{D_t(x_i)}{z_t} \times \begin{cases} 
eq y_i & \text{if } e^{\alpha_t h_t(x_i)} = y_i \\ = y_i & \text{otherwise} \end{cases} \quad (Z_t \text{ is normalization factor})
\end{equation}

end

\begin{equation}
\text{output: } H(x) = \arg \max_{y \in Y} \sum_{t=1}^{T} \alpha_t \| (h_t(x) = y)
\end{equation}

4. Experimental Analysis & Evaluation

4.1. Evaluation Indicators

The evaluation indicators of classification are based on the confusion matrix, in which TP and TN represent the correct classification of the number of positive and negative classes, and the FN and FP represent the error classification of the number of positive and negative classes.

A. True Positive Rate (TPRate): is the proportion of the correct sample number that is judged as the positive class to all positive class samples.

\begin{equation}
\text{TPRate} = \frac{TP}{TP + FP}
\end{equation}

B. False Positive Rate (FPRate): is the proportion of the error samples that is judged as the negative class to all negative class samples.

\begin{equation}
\text{FPRate} = \frac{FN}{FN + TN}
\end{equation}

C. The Recall Rate(recall): is the proportion of the samples correctly categorized to all samples.

\begin{equation}
\text{recall} = \frac{TP}{TP + FN}
\end{equation}

D. The Precision (precision): is the proportion of the positive samples correctly categorized to all positive samples.

\begin{equation}
\text{precision} = \frac{TP}{TP + FP}
\end{equation}

The next classification indicators that can reflect the performance more comprehensively and consider the recognition rate of large class and small class are as follows:

E. ROC curve, and AUC: is the area covered by the ROC curve.

ROC curves and AUC can treat large and small classes samples fairly. Like precision and recall, ROC curves can balance the recognition rate of small class and large class. The x-axis of the ROC curve indicates the FPR, the y-axis of the ROC curve indicates the TPR. The points of the ROC curve are adjusted by the threshold of the classifier, and the ROC curve is more convex, and the points closer to the upper left, which indicate that the classifier has a higher generalization ability. AUC refers to the area under the ROC curve, which can be quantified to represent the generalization ability of the classifier corresponding to the ROC curve.

F. F-Measure: can measure the overall performance of the classifier, the higher the value of which, the better the overall performance of the classifier.

\begin{equation}
f - \text{measure} = \frac{(1 + \beta)^2 \times \text{recall} \times \text{precision}}{\beta^2 \times \text{recall} \times \text{precision}}
\end{equation}

F1-Measure is \( F1 = 2 \times \text{precision} / (\text{precision} + \text{recall}) \). P stands for precision, R stands for recall. \( \beta \) is the regulatory factor, usually \( \beta = 1 \).
4.2. Experimental Data & Tool

To explain and evaluate the classification, the Moore dataset\textsuperscript{[27]} is used, which comes from the network exit shared by three biological institutes and is the two-way traffic data captured through the network outlet at ten different periods of one day. The Moore dataset contains 377,526 network traffic samples, divided into 12 types, the name, and scale of each type are shown in Table 1. The Moore dataset is a typical unbalance flow sample set, and the number of each type varies greatly. The largest number of WWW application samples accounted for more than 85% of the total data set, considered as the large class flow, while the sum of the other eleven application samples accounted for less than 15%, considered as small class flow. Because the number of INTERACTIVE and GAMES application samples is too sparse (0.029% and 0.002% of the total, respectively), there are not enough training sets, so the two types of data is ignored.

| Category       | Number of flows | Percentage (%) | Note        |
|----------------|-----------------|----------------|-------------|
| WWW            | 328092          | 86.906         |             |
| MAIL           | 28567           | 7.567          |             |
| FTP-CONTROL    | 3054            | 0.809          |             |
| FTP-PASV       | 2688            | 0.712          |             |
| ATTACK         | 1793            | 0.475          |             |
| P2P            | 2094            | 0.555          |             |
| DATABASE       | 2648            | 0.701          |             |
| FTP-DATA       | 5797            | 1.536          |             |
| MULTIMEDIA     | 576             | 0.153          |             |
| SERVICES       | 2099            | 0.556          |             |
| INTERACTIVE    | 110             | 0.029          | No consideration |
| GAMES          | 8               | 0.002          | No consideration |
| Total          | 377526          | 100            |             |

The experimental tool uses Weka (Waikato Environment for Knowledge Analysis), which is a java open source data mining platform researched by Waikato university. Weka integrates many machine learning algorithms for data-mining, which can be used to preprocess data, mine association rules, classify, cluster and so on. Weka also provides many visualization functions. Based on the API provided by Weka, this experiment uses java to implement the coarse and fine classification, and to complete the algorithm verification and comparison.

4.3. Validation & Evaluation

4.3.1. Analysis of Coarse Classification. Each sample in the Moore dataset contains 249 attributes, nearly half of which are derived from the Fourier transform. To improve the speed of the coarse classification, 26 features is selected in the Moore data according to the experience feature selection algorithm, namely: 1, 2, 3, 5, 6, 8, 9, 12, 13, 15, 16, 17, 33, 34, 45, 46, 47, 48, 61, 62, 63, 64, 75, 76, 212, 249, including the port, number of groups, packet length, time characteristics and the flag. After removing the other features, the data is loaded into the Weka, and the XMeans is used for cluster analysis, where Kmax is set to 8, Kmin to 2.

After XMeans classification, four subsets are generated. Overall, the data are mainly concentrated in datasets 1, 2, 4, accounting for 39.252%, 32.413% and 24.205%, in total of 95.87%. The distribution of the samples in the Moore dataset and four subsets is described in the Table 2, the change column describes the ratio of samples per type in the Moore dataset and four subsets of each type sample. Therefore, by the coarse classification, the distribution of the data can be estimated roughly. All kinds of data in its large proportion subset, the data accounted for the proportion of the subset has increased, and therefore further inhibit the imbalance of the data distribution.
Table 2. Analysis results of the coarse classification.

| Category   | Moore     | Subset 1 | Subset 2 | Subset 3 | Subset 4 | Change |
|------------|-----------|----------|----------|----------|----------|--------|
| WWW        | 86.906    | 80.869   | 95.216   | 71.511   | 88.194   |        |
| MAIL       | 7.567     | 11.003   | 2.243    | 0        | 0.198    |        |
| FTP-CONTROL| 0.809     | 1.357    | 0.704    | 0        | 0.194    |        |
| FTP-PASV   | 0.712     | 0.587    | 0.521    | 1.764    | 0.991    |        |
| ATTACK     | 0.475     | 0.236    | 0.075    | 0.385    | 1.414    |        |
| P2P        | 0.555     | 1.001    | 0.168    | 0.321    | 0.387    |        |
| DATABASE   | 0.701     | 0.429    | 0.031    | 0.564    | 2.064    |        |
| FTP-DATA   | 1.536     | 3.565    | 0        | 1.68     | 0.276    |        |
| MULTIMEDIA | 0.153     | 0.364    | 0        | 0.064    | 0.03     |        |
| SERVICES   | 0.556     | 0.574    | 1.01     | 0.026    | 0.009    |        |

4.3.2. Analysis of Fine Classification. Based on the AdaBoost, one of ensemble learning algorithm which chooses the J48 algorithm as the weak learning algorithm, and iterates 10 times, the fine classification, respectively uses the Moore data, the subset 1, the subset 2, the subset 3 and the subset 4, with the features selected by experience selection, to train the classification model. The classifications based on the ensemble learning compare the classifications of the above five datasets trained by the J48 algorithm with the F-measure and AUC (ROC area). The results are shown in Table 3.

Table 3. Analysis results of the fine classification.

| Dataset   | Instance | Attributes | Classification | F-Measure | ROC Area |
|-----------|----------|------------|----------------|-----------|----------|
| Moore     | 377526   | 26         | J48            | 0.998     | 0.996    |
| Moore     | 377526   | 26         | AdaBoost       | 0.998     | 0.998↑   |
| Subset1   | 148188   | 26         | J48            | 0.998     | 0.998    |
| Subset 1  | 148188   | 26         | AdaBoost       | 0.998     | 0.999↑   |
| Subset 2  | 122367   | 26         | J48            | 0.998     | 0.992↑   |
| Subset 2  | 122367   | 26         | AdaBoost       | 0.998     | 0.992↑   |
| Subset 3  | 15592    | 26         | J48            | 0.993     | 0.994    |
| Subset 3  | 15592    | 26         | AdaBoost       | 0.994↑    | 0.995↑   |
| Subset 4  | 91379    | 26         | J48            | 0.998     | 0.994    |
| Subset 4  | 91379    | 26         | AdaBoost       | 0.998     | 0.996↑   |

It is not difficult to find, the fine classifier, based on the ensemble learning, can improve the overall classification accuracy compared with the single classification, while the fine classification can also improve the recognition ability on the imbalanced data. The multimedia data of the subset 1, which belongs to the small class stream, accounts for 0.364% of the subset and 93.576% of this type data, and, the f-measure of which is improved higher by the AdaBoost than the J48. The attack data of the subset 4, which also belongs to the small class stream, accounts for 1.414% of the subset and 72.058% of this type data, and, the AUC of which is improved higher by the AdaBoost than the J48. The analysis results shown in Table 4.
### Table 4. Analysis results of small class flow recognition.

| Dataset | Protocol       | Proportion of subset | Proportion of type | Algorithm | F-Measure | ROC Area |
|---------|----------------|----------------------|-------------------|-----------|-----------|----------|
| Subset 1 | MULTIMEDIA     | 0.364                | 93.576            | J48       | 0.966     | 1.000    |
| Subset 1 | MULTIMEDIA     | 0.364                | 93.576            | AdaBoost  | 0.977†    | 1.000    |
| Subset 4 | ATTACK         | 1.414                | 72.058            | J48       | 0.950     | 0.961    |
| Subset 4 | ATTACK         | 1.414                | 72.058            | AdaBoost  | 0.950     | 0.973†   |

#### 4.3.3. Comprehensive Analysis.
This paper is based on two-layer classification model to verify the multi-layer classification of network traffic. The clustering process of the coarse classification supports iteration, it is from the data level, realizes the network traffic classification. It uses the feature selection based on the experience, compared with the information gain-based feature selection, can make the classification more generalization, which reduces the training sample's influence on the classification greatly, and highlights the regularity characteristic of the network traffic itself. Before performing a fine classification that requires a large amount of computational force, the dataset is divided into several subsets by the coarse classification, which is not only beneficial to improve the imbalance of sample distribution in each subset, but also reduces the resource requirements and calculation time required for the fine classification.

Experiment computer, configured as CPU i7-7700, the main frequency 3.6G, memory 32G. With the non-preprocessing Moore dataset, AdaBoost (weak learning algorithm is the J48, iterative 10 times, same below) is directly used to train the classification, which due to memory overflow and timeout.

With the Moore dataset having 26 features selected by experience, AdaBoost is directly used to train the classification, which takes 940.63 seconds. With the four subsets of the Moore dataset having 26 features which is divided by the coarse classification based on the XMeans, AdaBoost is separately used to train the classification, which take 365.04 seconds totally, accounted for 38.8% of the last time-consuming. In the above two cases, the f-measure and AUC values of the trained classifier are similar. The comparison is shown in Table 5.

### Table 5. Comparison of resource requirements and calculation time.

| Dataset | Instance | Attributes | Feature Selection | Clustering | F-Measure | AUC | Time   |
|---------|----------|------------|-------------------|------------|-----------|-----|--------|
| Moore   | 377526   | 249        | None              | None       | 0.998     | 0.998| timeout |
| Moore   | 377526   | 26         | Experience        | None       | 0.998     | 0.999| 940.63 |
| Subset1 | 148188   | 26         | Experience        | XMeans     | 0.998     | 0.999| 145.69 |
| Subset2 | 122367   | 26         | Experience        | XMeans     | 0.998     | 0.992| 100.54 |
| Subset3 | 15592    | 26         | Experience        | XMeans     | 0.994     | 0.995| 7.58   |
| Subset4 | 91379    | 26         | Experience        | XMeans     | 0.998     | 0.996| 111.23 |
| Moore   | 377526   | 26         | Experience        | XMeans     | 365.04    |     |        |

Based on the ranker-infogainattribute algorithm, the attribute is selected automatically, which is 1, 43, 44, 47, 81, 82, 83, 84, 86, 90, 93, 94, 95, 96, 99, 156, 158, 163, 165, 177, 179, 180, 184, 186, 187, 249. Based on the experiences, the attribute is selected, which is 1, 2, 3, 5, 6, 8, 9, 12, 13, 15, 16, 17, 33, 34, 45, 46, 47, 48, 61, 62, 63, 64, 75, 76, 212, 249. With the Moore dataset having 26 features separately selected by ranker-infogainattribute or experiences, AdaBoost is separately used to train the classification. Compared the classification based on ranker-infogainattribute with experiences, the f-measure and AUC values of the trained classifier are similar, but the classification based on experiences have better data universality, shown in Table 6.
Table 6. Comparison of feature selections based on experience with information gain.

| Dataset | Instance | Attributes | Feature Selection       | F-Measure | AUC  |
|---------|----------|------------|-------------------------|-----------|------|
| Moore   | 377526   | 26         | Ranker-infogainattribute| 0.998     | 0.998|
| Moore   | 377526   | 26         | Experience              | 0.998     | 0.998|

5. Conclusion

The classification method of "coarse" and "fine" two layers can realize the classification of network traffic quickly and well, which considers the traffic characteristics and data characteristics of network datasets, improves the recognition ability of small class flow, and has better adaptability. It can be embedded into network operation and maintenance platform to carry out network traffic data analysis quickly, which can flexibly carry out the application of different level.

The quality of network operation and maintenance management, ultimately relying on the "human" and "machine" intelligence fully integrated can be perfected, relying solely on either side cannot reach the optimal. In the era of artificial intelligence, the rational realization of the combination of "human" and "machine" intelligence, according to the goal of network operation and maintenance management, appropriate and effective use of various algorithms tools is proper.

References

[1] IANA. Internet assigned numbers authority [EB/OL]. on http://www.iana.org/assignments/port-numbers.
[2] Moore A, Papagiannaki K. Toward the accurate identification of network applications, Proceedings of Passive and Active Measurement Workshop (PAM2005), Boston (USA), 2005.
[3] Andrew Moore, Denis Zuev. Internet traffic classification using Bayesian analysis techniques, in Proc. ACM Sigmetrics, 2005, pp.50–60.
[4] Xu Peng, Lin Sen, Internet Traffic Classification Using C4.5 decision Tree, Journal of Software, 20 (2009) 2692-2704.
[5] Li Tao, Li Juan, Research on SOM Network Traffic Classification based on Pso-Mixed Kernel Function, Computer Applications and Software, 11 (2015) 117-125.
[6] Lin Haitao, Chen Yuan, Iterative Tuning Support Vector Machine for Network Traffic Classification, Journal of Naval Engineering University, 28 (2016) 10-14.
[7] Chaoxiao, Xia Jingpo, Li Minghui, Research on Classification of Network Traffic Based on Random Forests Algorithm, Journal of CAEIT, 04(2013) 184-190.
[8] Li Linlin, Zhang Xiaoyi, Li Xia, Li Qing, Network Traffic Classification Based on Ensemble Learning, Journal of Information Engineering university, 04(2015) 240-244.
[9] Wang Tao, Yu Shunzheng, Advances in Machine Learning Based Network Traffic Classification, Journal of Chinese Computer System, 33(2012) 1034-1040.
[10] Chen Yunjing, Research on P2P Traffic Identification Technology, Yangzhou university, 2009.
[11] Zhen Liu,Ruoyu Wang,Ming Tao, SmoteAdaNL:a learning method for network traffic classification, J Ambient Intell Human Comput, 07 (2016) 121–130
[12] Jiang Jie, Wang Zhuofang, Gong rong-sheng, Chen Tieming, Imbalanced Data Classification Method and its Application Research for Intrusion Detection, Computer Science, 40(2013) 131-135.
[13] Ye Zhifei, Wen Yimin, Lu Baoliang, A Survey of Imbalanced Pattern Classification Problems, CAAI Transactions on Intelligent Systems, 04(2009) 148-155.
[14] Han H. Wang W Y. Mao B H, Borderline—SMOTE: a new over—sampling method in imbalanced data sets learning, in International Conference on Intelligent Computing, 3644(2005) 878－887.
[15] Batista G E, Prati R C, Monard M C, A study of the behaviour of several methods for balancing machine learning training data, ACM Sigkdd Explorations Newsletter, 06(2004) 20－29.
[16] Li Yong, Liu Zhandong, Zhang Haijun, An overview of the integrated classification algorithm for imbalanced data, Computer Application Research, 05(2014) 1287-1291.

[17] Li Yong, An Ensemble Algorithms Based Voting for Imbalanced Data Classification, Science Technology and Engineering, 07(2014) 275-279.

[18] Yang Jieming, Gao Chong, Qu Chaoyang, Kan Zhongfeng, Gao Zhi, Random Forest Classification Algorithm Based on Cost-sensitive for Imbalanced Data, Science Technology and Engineering, 02(2018) 285-290.

[19] Jiang Shengyi, Xie Zhaoqing, Yu Wen, Naive Bayes Classification Algorithm Based on Cost Sensitive for Imbalanced Data Distribution, Journal of Computer Research and Development, 48(2011) 387-390.

[20] Li Xia, Wang Lianxi, Jiang Shengyi, Ensemble Learning based Feature Selection for Imbalanced Problems, Journal of Shandong University (Engineering Science), 06(2011) 7-11.

[21] Zhao Ying, Chen Junjun, Network Traffic Classification Based on Correlation of Flows, Computer Engineering and Applications, 51(2015) 25-29.

[22] Liu Kang, Feature Selection Study in the Internet Traffic Classification, Yangzhou university, 2013.

[23] PELEG D, MOORE A, X-means: Extending K-means with Efficient Estimation of the Number of Clusters, Proceedings of the Seventeenth International Conference on Machine Learning, San Francisco: Morgan Kaufmann Publishers, 2000, 89-97.

[24] Wang Yisen, Xia Shutao, A Survey of Random Forests Algorithms, Information and Communication Technology, 01 (2018) 49-55.

[25] Yu Ling, Wu Tiejun, Assemble learning: A Survey of Boosting Algorithms, PR & AI, 3 (2004) 52-59.

[26] Shen Xuehua, Zhou Zhihua, Wu Jianxin, Chen Shaoqian, Survey of Boosting and Bagging, Computer Engineering and Applications, 12 (2000) 31-33.

[27] Moore AW, Zuev D, Crogan M, Discriminators for use in flow-based classification. Technical Report, RR-05-13, Queen Mary University of London, 2005.