A Novel Adaptive Distribution Distance-Based Feature Selection Method for Video Traffic Identification

Licheng Zhang[1,3], Shuaili Liu[1,3], Qingsheng Yang[1,3], Zhongfeng Qu[2], and Lizhi Peng[1,3](✉)

[1] Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan, China
plz@ujn.edu.cn
[2] School of Mathematical Sciences, University of Jinan, Jinan, China
[3] Quancheng Laboratory, Jinan, China

Abstract. The rapid proliferation of video applications in recent years has triggered an unprecedented surge in Internet video traffic, which in turn has presented substantial challenges for effective network management. However, existing methods for extracting features from video traffic primarily focus on conventional traffic attributes, resulting in suboptimal identification accuracy. Furthermore, the challenge of handling high-dimensional data is a common hurdle in video traffic identification, necessitating a robust approach to select the most pertinent features crucial for accurate identification. Despite the abundance of studies utilizing feature selection to enhance identification performance, there exists a notable lack of research that addresses the quantification of feature distributions with small or no overlap. This study proposes, firstly, the extraction of features relevant to videos, thereby assembling an expansive feature repertoire. Secondly, in the pursuit of forming an effective subset of features, the current research introduces the adaptive distribution distance-based feature selection (ADDFS) methodology. Using the Wasserstein distance metric to quantify the differences between feature distributions. To gauge the efficacy of this proposal, a dataset comprising video traffic from various platforms within a campus network environment was collected, and a series of experiments were conducted using these datasets. The experimental results indicate that the proposed method can achieve highly accurate identification performance for video traffic.

Keywords: Feature selection · Feature extraction · Video traffic identification · Network management

1 Introduction

At present, most Internet users watch videos daily, resulting in a rapid increase in video traffic. According to Ericsson's mobility report, video traffic is anticipated

L. Zhang and S. Liu—Contributed equally to this paper.

X. Yang et al. (Eds.): ADMA 2023, LNAI 14179, pp. 226–240, 2023.
https://doi.org/10.1007/978-3-031-46674-8_16

to comprise 80% of the complete mobile data traffic by the year 2028. Therefore, effective identification and management of video traffic, particularly game video traffic, have become an important research topic for network management.

Some researchers have explored video traffic identification over the past decade. In the early research of computer vision, video content identification often uses image shapes, textures and other features to complete [1]. However, this method is not applicable from the network traffic perspective. The rapid development of network traffic identification is helpful in solving this problem. Most existing researchers have extracted traffic features related to video transmission, such as application data unit, burst, etc., and used them thereafter to complete the prediction of video QoS and QoE, identification of video application type. Unfortunately, they did not focus on identifying video scene traffic. Additionally, cloud game, as an emerging game mode, is essentially a way of video flow transmission, which is potentially harmful to teenagers. As far as we know, there is no research reported about identifying cloud game traffic. Some researchers have also begun to focus on improving the QoE and QoS of cloud game traffic, but they have not approached it from the perspective of network traffic. Thus, to extract effective features for video identification becomes an urgent concern.

A key issue is that current research studies mainly focus on traditional features, these features can not achieve the ideal video identification effect, and further research on video scene traffic feature extraction is needed. Besides, another key problem that should be further discussed is that the quality of extracted or selected features can significantly and directly impact the performance of identification. Irrelevant or redundant features can cause unnecessary cost and time overhead, even negative impact for the model identification. Thus, a high-performance feature selection method is crucial for traffic recognition.

In response to the challenges outlined above, we present the following contributions.

- A novel method for adaptive distribution distance-based feature selection (ADDFS) is introduced.
- A new feature extraction method based on video traffic peak point is proposed, which can be used as an effective supplement of traditional packet and flow level features.
- Different kinds of video traffic data are collected, including cloud game video traffic and video scene traffic.

Roadmap: Sect. 2 introduces the related research. Sect. 3 reviews the video traffic identification method. Sect. 4 presents the experimental results. Lastly, this paper is concluded in Sect. 5.

2 Related Research

2.1 Video Traffic Identification

Three kinds of traffic identification methods have been used for video traffic identification: port-based, deep packet inspection, and machine learning algorithms.

The first two methods have become ineffective owing to the dynamic port and encryption techniques, making machine learning-based method a widely used technology. In 2012, Ameigeiras et al. [2] analyzed YouTube's video traffic generation pattern to predict the quality of video watching experience. Given that early YouTube videos were based on Flash, which is no longer used, this method is no longer effective for current video traffic. Reed et al. [3] proposed a new bit per peak feature extraction method, and used these features for classifying video stream titles.

At present, only a few researches focus on cloud gaming video traffic identification, and the existing study has primarily concentrated on the analysis and modeling of cloud gaming traffic and improving the cloud gaming experience. Suznjevic et al. [4] collected cloud gaming video samples to calculate video indicators from the time and space dimensions. Thereafter, they analyzed the relationship among game types, cloud gaming video traffic features, and video indicators. In 2015, Amiri et al. [5] proposed a paradigm for SDN controller to reduce cloud gaming delay. These studies rarely focus on identifying cloud game video traffic and video scene traffic, and this paper will focus on it.

2.2 Feature Selection

Feature selection is vital for traffic identification because all types of features are extracted from raw traffic data. Many of these features are redundant or with no contribution for identification. Therefore, researchers have attempted to develop effective methods to evaluate and select traffic features in recent years. Zhang et al. [6] and Mousselly et al. [7] used KL and JS divergence respectively to analyze the correlation and redundancy of different class labels, which can effectively deal with the fluctuation of feature samples. Nevertheless, their research did not address the issue of small overlap or no overlap between feature distributions.

Recently, certain researchers have started employing feature selection techniques for video traffic identification. Dong et al. [8] combined ReliefF and PSO to solve the excessive dimensionality problem in network traffic classification. Wu et al. [9] used a linear consistency-constrained method to select features for multimedia traffic classification and completed instance purification in the selection process. As far as we know, no study using has been conducted on distribution distance to measure the similarity between video traffic feature distributions. Therefore, this paper overcomes this drawback, by using Wasserstein distance to adaptively measure the similarity between feature distributions, and build an effective feature selection algorithm thereafter.

3 Methodology

This section describes the framework for video traffic identification, as shown in Fig. 1.

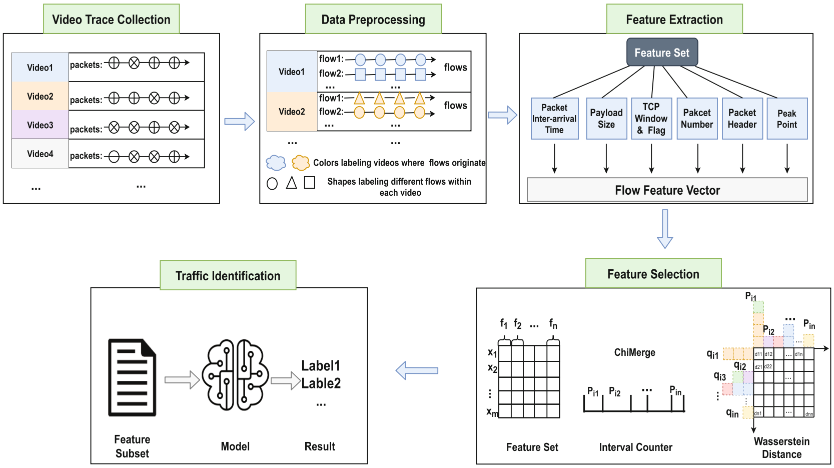


Fig. 1. The framework of the proposed video traffic identification method.

3.1 Data Collection

Only a few public video traffic data sets are available for video traffic identification research. Thus, a cloud gaming video traffic data set (CG-UJN-2022) and video scene traffic data set (VS-UJN-2022) in a controlled campus environment was collected.

Video Scene Traffic Data Collection. The collected video scene traffic data can be divided into two categories: static and action scene videos. The action scene video mainly consists of fragments from science fiction action films, such as *Pirates*, *Transformers*, *The Avengers*, etc. However, static scene videos have a simple scene, such as light music video, natural views, and class scenes. We collected both types of data from YouTube and Bilibili.

Videos from the mentioned categories will be initially downloaded to the client computer, followed by using FFmpeg to segment the original video into clips with a consistent duration of 120 s. We regard a 120 s video segment as a scene because such a segment can provide sufficient network features for coarse-grained video scene identification.

Secondly, with the Selenium library and Xpath Helper, fixed video clips are automatically uploaded to YouTube and Bilibili. With t-shark, we achieved automatic on-demand delivery of targeted videos and automatic collection of video traffic while playing the videos. During video playback on the client's computer, all other network applications are shut down to prevent the generation of extraneous traffic.

**Cloud
Gaming Video Traffic Data Collection.** YOWA cloud gaming, Tencent

Start, MiguPlay, and Tianyi cloud gaming are the four cloud gaming platforms we visited. To compare with other data features, wireshark was set to automatically save collected data as a.pcap file every 120 s. Similar to video scene traffic data collection, other applications were closed while collecting target traffic. Segments of the background traffic were also captured, primarily encompassing the most prevalent application categories. Detailed information about the collected traffic data is presented in Tables 1 and 2.

3.2 Data Preprocessing

First, We group the collected traffic data into flows based on five tuple information: {src IP, src port, dst IP, dst port, protocol (TCP/UDP)}. Since that YOWA, MiguPlay, and TianyiPlay use UDP as the transport layer protocol, we focus on UDP packets when analyzing the three platforms and TCP packets for the rest of the traffic.

Second, elephant flows are selected from the mice flows. Elephant flows is an important focus in this study, as video traffic is mostly elephant flows. The number of non zero payloads is used to eliminate mice flow. According to experience, those flows with under 500 packets are considered mice flows to be eliminated.

Lastly, the SNI extension field within the Client Hello packet serves the purpose of identifying whether the captured flow corresponds to the intended target flow.

Table 1. The details of video scene traffic data

Data Set	Platform	Flows	Bytes
VS-UJN-2022	YouTube	Action	4,994,062,797
		Static	3,294,957,584
	Bilibili	Action	5,956,426,782
		Static	7,672,791,660
CG-UJN-2022	START	83	3,401,167,754
	YOWA	58	2,116,354,048
	MiguPlay	61	2,548,042,936
	TianyiPlay	63	3,132,444,936

3.3 Feature Extraction

A total of 89 statistical features are extracted from preprocessed data in this study. We analyze the packet sequence features of each flow from three directions, namely upstream, downstream, and all packets. The traditional traffic features mainly include packet inter-arrival time (IAT), payload size, TCP window size, TCP flag, packet number, packet header. The detail are shown in Table 3.

Table 2. The details of background traffic data

	Flows	Bytes
Online Meeting	5	21,501,967
Chatting	56	134,386,569
PC Video	42	704,530,546
PC Live Steaming	40	397,123,864
Web Browsing	57	73,440,492
Online Shopping	42	142,277,293
File Download	40	57,409,424

Additionally, different video styles will lead to different traffic behavior patterns. Therefore, the maximum data transmission amount over a period of time will be defined as the peak point in this study.

Payload peak point (PPP). Assume there are d packets in a flow and the packets is Pkt_1,Pkt_2, ...,Pkt_d. Payload size of the sth packet is presented as pay_s. If $pay_s \geqslant pay_{s-1}$ and $pay_s \geqslant pay_{s+1}$ $(1 < s \leq d-1)$, then payload reaches a peak in a certain period of time, which is defined as the PPP. A set of counters $c_1^l, c_2^l, ..., c_\theta^l$ was used to count the number of PPP every α s in the first β s of the lth flow, then the θ is calculated as follows:

$$\theta = \frac{\beta}{\alpha}. \tag{1}$$

Then, the count matrix CT can be obtained by traversing the entire flow sequence.

$$CT = \begin{bmatrix} c_1^1 & c_2^1 & \cdots & c_\theta^1 \\ c_1^2 & c_2^2 & \cdots & c_\theta^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_1^t & c_2^t & \cdots & c_\theta^t \end{bmatrix} \tag{2}$$

Based on CT, the std and mean of the PPP of the tth flow is obtained as follows:

$$M_t = \frac{1}{\theta}\sum_{a=1}^{\theta} c_a^t, \tag{3}$$

$$Std_t = \sqrt{\frac{1}{\theta}[(c_1^t - M_t)^2 + (c_2^t - M_t)^2 + ... + (c_\theta^t - M_t)^2]}. \tag{4}$$

In a similar vein, the standard deviation and mean of PPP for all flows can be derived. Additionally, we extracted the maximum, minimum, and aggregate count of PPPs in three orientations. Nevertheless, in cases where certain scene videos are being consistently transmitted, alterations in packet payload remain insignificant. Hence, we introduce the concept of byte rate peak point (BRPP).

BRPP. Assuming that the summation of packet payloads (SPP) during period T is calculated in the following manner:

$$SPP = \sum_{b=1}^{H} pay_b, \tag{5}$$

where H is the total number of packets within T s, and pay_b is the size of the bth packet payload. Thereafter, the definition of byte rate (BR) in T seconds is as follows:

$$BR = \frac{SPP}{T}. \tag{6}$$

Similarly, If BR satisfies the criteria of being a peak point, it is labeled as BRPP. In this study, T is configured to be 1 s.

Table 3. The details of the extracted traditional features

Feature Name	Description
UIAT_*	Mean, Min, Max, Std of upstream IAT interval
DIAT_*	Mean, Min, Max, Std of downstream IAT interval
IAT_*	Mean, Min, Max, Std of all packets IAT interval
UWindow_*	Sum, Mean, Min, Max, std of upstream TCP window sizes
DWindow_*	Sum, Mean, Min, Max, std of downstream TCP window sizes
Window_*	Sum, Mean, Min, Max, std of all packets TCP window sizes
*_pnum	Number of packets for three directions
*_pnum_s	The rate of packet number for three directions
UDpnum_s	packets downstream to/packets upstream
*_cnt	TCP flag count
UDPSH,UDURG_cnt	Upstream and downstream PSH and URG count
Uhdr,Dhdr,hdr	Sum of packet header length for three directions
*_hdrR	the packet header length sum/the packet payload sum
Upay_*	Mean, Min, Max, Std of upstream payload
Dpay_*	Mean, Min, Max, Std of downstream payload
pay_*	Mean, Min, Max, Std of all packets payload

BRPP with sliding windows (BRPPSW). To catch continuous video information more accurately, we design sliding windows to extract the size of peak points as feature vectors based on BRPP. Length of the sliding window is L and offset factor is denoted by Z. In this study, L and Z are set to 3 and 0.5, respectively. For a packet sequence (Pkt_1,Pkt_2, ...,Pkt_d), we calculate the sum of packet size under in time window L and use the offset factor thereafter to move the window to calculate the total packet size in turn. The sum of packet size in the zth window can be calculated as follows:

$$r_z = \sum_{pt=0}^{L} pktLen_{pt}, \tag{7}$$

where pt is the arrival time of the packet and $pktLen_{pt}$ is the packet size at ptth s. The processed sequence R= (r_1,r_2, ... ,r_n) is obtained, where n is the number of sliding windows. If the value in the sequence meets the definition of the preceding peak point, then the point is defined as BRPPSW. Therefore, we will obtain the sequence R_F=(r_1,r_2, ... ,r_u) of BRPPSW, which is a subset of R.

We calculate the mean, std, maximum and minimum values of BRPPSW from three directions. The first, second, and third quartile of BRPPSW are also extracted as features.

3.4 Feature Selection

By the previous step, a comprehensive feature set is obtained. However, note that we do not consider whether these features are redundant or useless at the extracting process. In order to choose a feature subset that is both effective and concise, we introduce an approach called Adaptive Distribution Distance-Based Feature Selection (ADDFS).

Assuming a dataset $X = \{X_1, X_2, ..., X_n\}$, where X_i (1$\leqslant$ i $\leqslant$ n) represents the ith sample data, and m denotes the total number of samples. Moreover, x_{ij} denotes the value of the jth feature for the ith sample.

First,we employ Min-Max scaling to standardize all feature values across the dataset into the [0,1] interval. The formula for Min-Max scaling is as delineated below:

$$X_{ij} = \frac{X_{ij} - min(X_{.j})}{max(X_{.j}) - min(X_{.j})}, \tag{8}$$

here, $max(X_{.j})$ represents the maximum value of the jth feature, while $min(X_{.j})$ corresponds to the minimum value of the jth feature.

Second, the supervised ChiMerge algorithm [10] is used to divide each feature into multiple consecutive intervals. For each feature, we first sort all values in ascending order. Thereafter, we group the data with the same feature value into the same interval, and calculate the chi-square value of the interval. Each adjacent chi-square value is calculated and the smallest pair of intervals are merged. This step is repeated until the set maximum binning interval or chi-square stopping threshold is reached. Lastly, the chi-square binning interval of each feature is obtained. According to empirical values, the maximum binning interval and stop confidence threshold in this paper are set to 15 and 0.95, respectively. The chi-square calculation formula is as follows:

$$\chi^2 = \sum_{\gamma=1}^{G} \sum_{\psi=1}^{C} \frac{(A_{\gamma\psi} - E_{\gamma\psi})^2}{E_{\gamma\psi}}, \tag{9}$$

$$E_{\gamma\psi} = \frac{N_\gamma}{N} \times C_\psi, \tag{10}$$

where G stands for the number of intervals, and C represents the number of classes, $A_{\gamma\psi}$ represents the quantity of samples from the ψth class within the γth interval, $E_{\gamma\psi}$ is the expected frequency of $A_{\gamma\psi}$, and N, N_γ, and C_ψ denotes the overall sample count, the sample count within the γth interval, and the sample count within the ψth class, respectively.

For each feature, the number of samples of a particular feature within the chi-square binning interval in each class is counted. Take the feature F_j as an example. For class $C1$, the distribution of feature F_j within the chi-square binning intervals $(p_{11}, p_{12}, ..., p_{1k})$ can be acquired by tallying the occurrences of feature F_j across each interval, in which k is the number of chi-square binning intervals for this feature. For class $C2$, the distribution of feature F_j can be calculated as $(p_{21}, p_{22}, ..., p_{2k})$. On this basis, we can obtain the feature distribution matrix P of feature F_j on n classes. In the same manner, the feature distribution of other features on different classes can also be obtained.

$$P_{n\times k} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nk} \end{bmatrix} \tag{11}$$

The Wasserstein distance (EMD) is employed to quantify the distribution disparity between every pair of classes. A higher EMD value for a given feature across two classes indicates a more discerning characteristic. The computation of EMD for each class pair is conducted as follows:

$$W(P_U, P_V) = inf_{\gamma \sim \Pi(P_U, P_V)} E_{(U,V)\sim\gamma} \; [\|x - y\|], \tag{12}$$

where P_U and P_V are the feature distribution of a feature on two classes, $\Pi(P_U, P_V)$ denotes the set of all potential joint distributions P_U and P_V, while $W(P_U, P_V)$ signifies the mathematical lower bound of the expected value of $\gamma(x, y)$. The calculation of EMD for multi-class is detailed as follows:

$$EMD = \sum_{\kappa=1}^{C} \sum_{\lambda=\kappa+1}^{C} W(P_\kappa, P_\lambda), \tag{13}$$

Finally, calculate the EMD value for each feature. Subsequently, sort features in descending order based on their respective EMD values. The pseudo-code of ADDFS is shown in Algorithm 1.

Algorithm 1. Adaptive distribution distance-based feature selection algorithm

```
Require: Feature set $F$, classes $C$
Ensure: The selected feature subset $S$
BEGIN
compute $F$ according to euqation (8);
for each $f \in F$:
    Interval=ChiMerge($f$);
  / / The ChiMerge algorithm is used to divide each feature into multiple consecutive intervals
    for each $c \in C$:
        Pc=count($f$,$c$,Interval);
      / / Calculate the number of samples of the feature in the chi-square binning interval of each class
    end for;
    for each $p \in P$:
        for each $p' \in P$, $p' \neq p$:
            compute W($p$,$p'$) according to equation; (12)
            $EMD_f$ += W($p$,$p'$);
        end for;
    end for;
end for;
$S$= sort($EMD$);
END;
```

3.5 Machine Learning Model

This study employs six machine learning models for identification. Noted that we do not focus on the actual machine learning model but on the effect of our proposed method combined with the machine learning model on video traffic identification. By comparing the identification results of different models, we can choose the model with superior performance for video traffic identification.

4 Experiment

4.1 Performance Measures

In this paper, accuracy (ACC) and F1 score can be derived as the evaluation criteria in our experiment. The accuracy (ACC) in a binary classification task can be defined as follows:

$$ACC = \frac{TP + TN}{TP + FN + TN + FP}, \tag{14}$$

Precision and recall can be defined as follows:

$$Precision = \frac{TP}{TP + FP}, \tag{15}$$

$$Recall = \frac{TP}{TP + FN}. \tag{16}$$

With precision and recall, F1 score, a widely used performance measure, can be derived as follows:

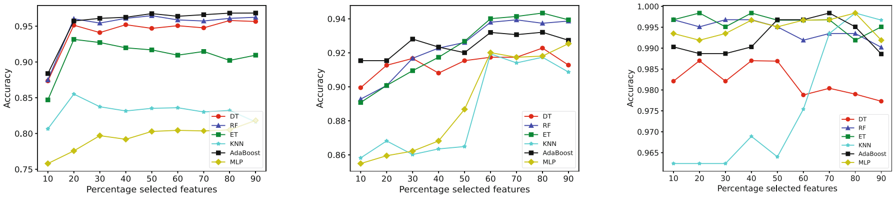
$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \tag{17}$$

4.2 Evaluation of ADDFS with Video Traffic Identification

The overall identification performance of video traffic is first evaluated by using the selected learning models and proposed feature selection algorithms. ADDFS is utilized to choose feature subsets comprising 10%, 20%, 30% ..., and 90% of the complete feature set. Thereafter, all selected learning models are used to identify both types of video traffic. The results are presented in Fig. 2.

From the perspective of the number of selected feature set, for YouTube and Bilibili, the identification effects of most of the learning models hit the optimum at 20% and 60%, respectively, of the feature set and reach a steady state thereafter. For cloud games, the recognition effect of the learning model maintains a small range of fluctuations on different feature subsets.

From a learning model perspective, Random Forest (RF), Extremely Randomized Trees (ET), and Adaptive Boosting (AdaBoost) perform well. In a stable state, RF and AdaBoost achieve accuracy levels exceeding 0.95 on YouTube. Furthermore, the accuracies of RF, ET, and AdaBoost on Bilibili and cloud gaming are above 0.92 and 0.99, respectively.



(a) YouTube video scene traffic data

(b) Bilibili video scene traffic data

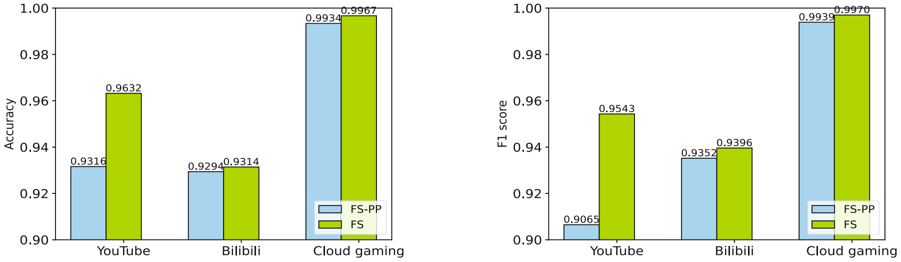
(c) Cloud gaming video scene traffic data

Fig. 2. Accuracy results with varying feature number percentage selected by ADDFS

4.3 Assessment of the Efficacy of Peak Point Features

This subsection assesses the influence of various sliding window sizes and offset factors on video flow identification in cloud gaming. The Random Forest (RF) classifier is employed, and a 10-fold cross-validation approach is once again implemented.

Figure 3(a) and (b) shows the results of the comparison, in which FS is the complete feature set with peak point features, and FS-PP is the feature set without peak point features. The results of FS are observed to be better than those of FS-PP, particularly for data of video scene traffic on the YouTube platform. That is, ACC and F1 increased by over 3%. For the other two cases, the two evaluation measures also improved slightly with the joining of peak point features. Hence, the experimental outcomes unequivocally demonstrate the efficacy of the proposed peak point feature for video traffic identification.



(a) The comparative results of accuracy (ACC)

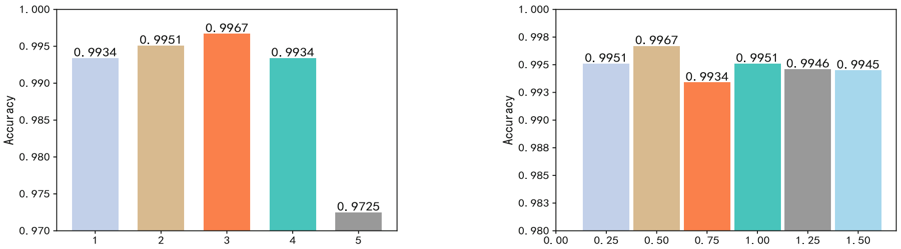
(b) The comparison results of F1 score

Fig. 3. The comparison results with/without peak point features

4.4 Evaluation of the Impact of Sliding Windows

This subsection evaluates the impact of different sliding window sizes and offset factors on video flow identification on cloud gaming. RF is used as a classifier, and 10-fold cross-validation is again applied.

Figure 4(a) demonstrates the impact of different sliding window sizes on identification accuracy. Offset factor is set to 0.5. As window size grows, identification accuracy increases initially. Thereafter, it reaches the highest when window size is set to 3. Accuracy decreases thereafter as window size increases. Therefore, we obtain the empirical optimal window size of 3. Figure 4(b) shows the results with the varying offset value. Note that when offset factor is 0.5, accuracy of video traffic identification hits the highest value. When offset factor increases, recognition accuracy tends to be stable. Thus, we set window size L to 3 and the offset factor Z to 0.5 in our studies.



(a) The accuracy of different sliding window sizes (s)

(b) The accuracy of different offset factor sizes (s)

Fig. 4. The impact of sliding window

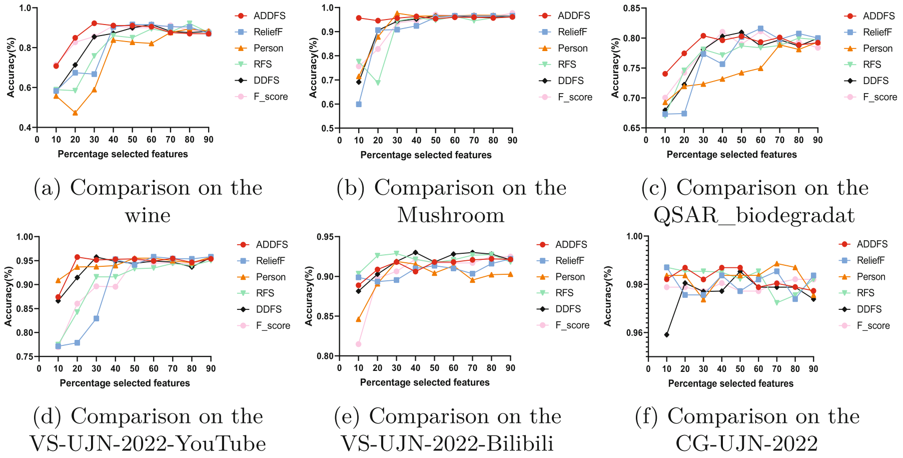
4.5 Evaluation of the ADDFS Performance

To further verify the effectiveness of the feature selection algorithm ADDFS, we conduct comparative experiments on 3 public datasets (wine, Mushroom and QSAR_biodegradat, the first one is from KEEL, the last two are from UCI) and 3 private traffic datasets (VS-UJN-2022-YouTube, VS-UJN-2022-Bilibili and CG-UJN-2022) with 5 feature selection algorithms. The five compared feature selection methods are Relief [11], Person [12], RFS [13], DDFS [14], and F-score [15]. We employ Decision Tree (DT), as the classifier and compare the ACC of the evaluated methods using 10-fold cross-validation. The classification ACC outcomes are illustrated in Fig. 5.

As shown in Fig. 5, all compared methods will receive increasing accuracy as the number of selected features increases for most data sets, and reach a relatively steady state thereafter. In cases where the count of chosen features is limited, ADDFS demonstrates superior accuracy when compared to the alternative methods. Note that it has consistently maintained efficient and stable performances for the cases of the wine, Mushroom, QSAR_biodegradat, and VS-UJN-2022-Bilibili datasets. Although there are numerous redundant and irrelevant features in the CG-UJN-2022 dataset, ADDFS can still obtain a relatively stable classification accuracy in the early stage.

5 Conclusion

A comprehensive feature set is constructed in this study for identifying video traffic. In order to obtain an efficient feature subset, a novel ADDFS method is introduced. Moreover, we collected video traffic data from different platforms in a campus network environment and used these data to conduct a set of experiments. The experimental findings demonstrate a significant enhancement in identification performance through the utilization of the proposed peak point feature. The proposed ADDFS can also be considerably applied to the task of video traffic identification.



(a) Comparison on the wine

(b) Comparison on the Mushroom

(c) Comparison on the QSAR_biodegradat

(d) Comparison on the VS-UJN-2022-YouTube

(e) Comparison on the VS-UJN-2022-Bilibili

(f) Comparison on the CG-UJN-2022

Fig. 5. Results of the compared feature selection methods

Acknowledgment. This research was partially supported by the National Natural Science Foundation of China under Grant No. 61972176, Shandong Provincial Natural Science Foundation, China under Grant No. ZR2021LZH002, Jinan Scientific Research Leader Studio, China under Grant No. 202228114, Shandong Provincial key projects of basic research, China under Grant No. ZR2022ZD01, Shandong Provincial Key R&D Program, China under Grant No. 2021SFGC0401, and Science and Technology Program of University of Jinan (XKY1802).

References

1. Brezeale, D., Cook, D.J.: Automatic video classification: a survey of the literature. IEEE Trans. Syst. Man Cybern. Part C (Applications and Reviews) **38**(3), 416–430 (2008)
2. Ameigeiras, P., Ramos-Munoz, J.J., Navarro-Ortiz, J., Lopez-Soler, J.M.: Analysis and modelling of YouTube traffic. Trans. Emerg. Telecommun. Technol. **23**(4), 360–377 (2012)
3. Dubin, R., Dvir, A., Pele, O., Hadar, O.: I know what you saw last minute-encrypted http adaptive video streaming title classification. IEEE Trans. Inf. Forensics Secur. **12**(12), 3039–3049 (2017)
4. Suznjevic, M., Beyer, J., Skorin-Kapov, L., Moller, S., Sorsa, N.: Towards understanding the relationship between game type and network traffic for cloud gaming. In: 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 1–6. IEEE (2014)
5. Amiri, M., Al Osman, H., Shirmohammadi, S., Abdallah, M.: An SDN controller for delay and jitter reduction in cloud gaming. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 1043–1046 (2015)
6. Zhang, Y., Li, S., Wang, T., Zhang, Z.: Divergence-based feature selection for separate classes. Neurocomputing **101**, 32–42 (2013)
7. Moussellу-Sergieh, H., Döller, M., Egyed-Zsigmond, E., Gianini, G., Kosch, H., Pinon, J.M.: Tag relatedness using Laplacian score feature selection and adapted

Jensen-Shannon divergence. In: Gurrin, C., Hopfgartner, F., Hurst, W., Johansen, H., Lee, H., (eds.) MultiMedia Modeling. MMM 2014. LNCS, vol. 8325. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-04114-8_14

8. Dong, Y., Yue, Q., Feng, M.: An efficient feature selection method for network video traffic classification. In: 2017 IEEE 17th International Conference on Communication Technology (ICCT), pp. 1608–1612. IEEE (2017)
9. Wu, Z., Dong, Y., Wei, H.L., Tian, W.: Consistency measure based simultaneous feature selection and instance purification for multimedia traffic classification. Comput. Netw. **173**, 107190 (2020)
10. Kerber, R.: ChiMerge: discretization of numeric attributes. In: Proceedings of the Tenth National Conference on Artificial Intelligence, pp. 123–128 (1992)
11. Robnik-Šikonja, M., Kononenko, I.: Theoretical and empirical analysis of ReliefF and RReliefF. Mach. Learn. **53**(1), 23–69 (2003)
12. Hall, M.A.: Correlation-based feature selection for machine learning, Ph. D. thesis, The University of Waikato (1999)
13. Nie, F., Huang, H., Cai, X., Ding, C.: Efficient and robust feature selection via joint $\ell 2$, 1-norms minimization. In: Advances in Neural Information Processing Systems 23 (2010)
14. Liu, S., Zhang, L., Sun, P., Bao, Y., Peng, L.: Video traffic identification with a distribution distance-based feature selection. In: 2022 IEEE International Performance, Computing, and Communications Conference (IPCCC), pp. 80–86. IEEE (2022)
15. Jaganathan, P., Rajkumar, N., Nagalakshmi, R.: A Kernel based feature selection method used in the diagnosis of Wisconsin breast cancer dataset. In: Abraham, A., Lloret Mauri, J., Buford, J.F., Suzuki, J., Thampi, S.M. (eds.) ACC 2011. CCIS, vol. 190, pp. 683–690. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22709-7_66