

基于 SU 和 AMB 的网络流量特征选择算法^①



庞玉林^{1,2}, 李喜旺¹

¹(中国科学院 沈阳计算技术研究所, 沈阳 110168)

²(中国科学院大学, 北京 100049)

通信作者: 庞玉林, E-mail: pangyulin19@163.com

摘要: 在基于网络流量分析, 被动式的网络设备识别研究中, 网络流量数据中往往存在许多高维数据, 其中的部分特征对设备识别贡献不大, 甚至会严重影响分类结果和分类性能. 所以针对这个问题本文提出了一种将 Filter 和 Wrapper 方式相结合, 基于对称不确定性 (SU) 和近似马尔可夫毯 (AMB) 的网络流量特征选择算法 FSSA, 本文提出的方法首先利用对称不确定性算法选择出对于各个类别具有分类贡献的特征, 去除不相关的特征属性; 然后在候选特征子集中利用近似马尔可夫毯算法删除冗余特征, 最后采用 Wrapper 方式基于 C4.5 分类算法, 进行最后的特征优选. 实验表明, 该方法下选择出的特征对网络设备操作系统类型识别的精确率相较于经典的特征选择方法有了一定的提高, 在小类别数据上的召回率也得到了提升.

关键词: 特征选择; 网络流量; 对称不确定性 (SU); 近似马尔可夫毯 (AMB); 网络设备识别; 机器学习

引用格式: 庞玉林, 李喜旺. 基于 SU 和 AMB 的网络流量特征选择算法. 计算机系统应用, 2022, 31(4): 281-287. <http://www.c-s-a.org.cn/1003-3254/8410.html>

Feature Selection Algorithm of Network Traffic Based on SU and AMB

PANG Yu-Lin^{1,2}, LI Xi-Wang¹

¹(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

²(University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: In the research of passive network device identification based on network traffic analysis, much high-dimensional data often appears in the network traffic data, and some of these features do not contribute much to device identification and even can seriously affect the classification results and performance. Therefore, this study proposes a network traffic feature selection algorithm FSSA that combines Filter and Wrapper approaches based on symmetric uncertainty (SU) and approximate Markov blanket (AMB). Specifically, the proposed method in this study first uses the SU algorithm to select the features with classification contributions for each category and remove irrelevant feature attributes. Then, the AMB algorithm is adopted to delete redundant features in the subset of candidate features. Finally, the Wrapper approach based on the C4.5 classification algorithm is employed to determine the final feature preference. The experimental results show that the accuracy of the features selected under this method for type identification of the network device operating system has been improved compared with classical feature selection methods, and the recall rate on small class data has also been raised.

Key words: feature selection; network traffic; symmetric uncertainty (SU); approximate Markov blanket (AMB); network devices identification; machine learning

① 基金项目: 兴辽英才计划 (XLYC2019019)

收稿时间: 2021-06-07; 修改时间: 2021-07-07, 2021-07-20; 采用时间: 2021-07-27; csa 在线出版时间: 2022-03-22

随着联网设备的增多,越来越多的网络设备暴露在公网内,网络空间扫描工具 Shodan 甚至可以轻而易举获取到隐私家用摄像头的内容.在工业网络中,原本封闭的环境变得开放,联网设备版本难以升级,不易管理,这使得工业网络内的专业设备有着更容易被入侵的风险^[1],如果我们能够及时获取联网设备信息,掌握网络中设备之间的拓扑关系,有效的进行资产管理和资产分析,了解设备的脆弱点,及时将漏洞信息和设备脆弱点结合起来,在发生故障时能够及时进行故障定位和故障处置分析.目前现有的网络设备识别研究中大多数都是通过主动探测的技术来发现设备,这种方法会干扰或者中断设备的运行状态,但在实际环境中,大部分设备过时或计算能力较低,并不支持中断或干扰其系统^[2-4].基于网络流量特征的被动式识别方法具有干扰性小、自动化高的特点,所以其目前成为学者们的研究热点,也更符合未来的发展要求.

在基于网络流量分析的网络设备识别研究中,利用 Wireshark 在特定环境中抓取的网络流量数据包中的特征属性集往往是高维的,这些属性集中存在许多冗余或不相干的属性,这些冗余信息不但会降低数据集的数据质量,也会影响后续网络设备识别效率和精确率,妄加无效的工作量,所以为了保证数据集的数据质量和设备识别效率及精确率,有效地对数据包中的数据特征属性进行降维是一项非常重要的工作.

1 相关工作

目前网络流量特征选择算法在网络流量分类的研究中应用比较广泛,大多数的文献主要是针对网络流量分类的效率和精确率来对网络流量特征选择算法进行研究和改善^[5].刘雪亚等人^[6]、唐宏等人^[7]和赵择等人^[8]的工作主要是针对网络流量不均衡,面向网络流量数据中单个类别的分类,刘雪亚等人^[6]提出了一种基于卡方算法和 SU (对称不确定性) 算法的结合来提高小类别的网络流量数据的分类精确率;唐宏等人^[7]提出一种基于 FCBF 的特征选择算法更容易地选择出来与小类别流量具有强相关性的特征;赵择等人^[8]提出一种基于改进一对一算法的集成学习方式把网络流量分类的多分类任务拆解为多个相互独立的二分类任务,其思想的重点不在于特征选择的工程如何改善,而是把重点放在了利用集成学习的策略上来充分挖掘每类流量的数据特征来提高分类性能;曹杰^[9]在工作中

提出了一种 Filter-Wrapper 混合的特征选择模型,其先利用信息增益对有分类贡献的特征进行性能评估排序,在候选子集上再利用 Wrapper 方式进行二次特征选择.

但本文的研究目的是选择出能够更高效、更精确地对网络设备进行识别的网络流量特征,Nguyen-An 等人^[10]提出了一种通过计算网络流量特征的信息熵来对网络设备进行识别的方法;Jeon 等人^[11]提出了把网络端口和一系列 TCP/IP 协议栈的相关特征字段作为设备识别的指纹特征,但是这类基于人为分析选取的网络流量特征的被动式设备识别方法在通用性,精确度方面还有待提高.

2 算法设计

2.1 FSSA

本文主要是针对网络设备的操作系统类型进行识别,所以针对这个问题本文提出了一种将 Filter 和 Wrapper 方式相结合,基于 SU^[12]和 AMB^[13]的网络流量特征选择算法 FSSA (feature selection based on symmetric uncertainty and approximate Markov Blanket),该算法首先在 Filter 方式中利用 SU 算法选择出对于各个类别具有分类贡献的特征,去除不相关的特征属性;然后在候选特征子集中利用马尔可夫毯算法删除冗余特征,最后再采用 Wrapper 方式,基于 C4.5 决策树分类算法选择出使分类器效果最好的最优特征集,这样不仅可以对特征集进行降维,也可以提高分类器的性能.

2.2 SU 算法

假设样本总数为 N , 样本中有 j 个类别的样本,类别用 C 表示,特征属性集 $F = \{f_1, f_2, \dots, f_i\}$, SU 可以用来描述特征属性 f_i 和特征属性所属类别 C 之间的相关性,也可以描述特征属性集 F 中特征属性与特征属性的相关性^[12,14].

首先我们引入信息熵 $H(F)$ 和条件熵 $H(F|C)$ 的定义,如式 (1) 和式 (2),其中 $p(f_i)$ 表示 f_i 在特征属性集 F 中出现的概率, $p(c_j)$ 表示类别 c_j 的出现概率, $p(f_i|c_j)$ 表示特征 f_i 在类别 c_j 条件下的条件概率.

$$H(F) = - \sum_{i=1}^n p(f_i) \log_2 p(f_i) \quad (1)$$

$$H(F|C) = \sum_{j=1}^m p(c_j) \sum_{i=1}^n p(f_i|c_j) \log_2 p(f_i|c_j) \quad (2)$$

由式(1)和式(2)可以得到特征属性与类别的信息增益 $IG(F,C)$,如式(3).

$$IG(F,C) = H(F) - H(F|C) \quad (3)$$

由式(1)和式(2)推算式(3)可以得到如下关系,此处不再赘述.

$$IG(F,C) = IG(C,F) \quad (4)$$

最后,由上述4个公式我们可以得到特征 f_i 与类别 C 之间的对称不确定性 $SU(f_i,C)$,如式(5).

$$SU(f_i,C) = \frac{2IG(C,f_i)}{H(C)+H(f_i)} \quad (5)$$

首先通过计算特征 f_i 与类别 C 的对称不确定性 $SU(f_i,C)$,对所有结果进行一个降序排序,排名越靠前的值其对应的特征对类别分类的贡献就越大,通过选取合适的阈值 δ ,对于 SU 值大于阈值 δ 的特征,将其放入候选特征子集内.

2.3 AMB 算法

马尔可夫毯是进行特征冗余性分析的一种常用的工具,在一个特征空间中,目标特征的马尔可夫毯包含了它的所有信息,所以非马尔可夫毯就可以看成是目标特征的冗余特征.因此通过发现目标特征的马尔可夫毯就可以精确确定目标特征的冗余特征,从而降低特征空间的维数,达到特征选择的目的^[13].

首先引入马尔可夫毯的概念:假设在随机变量的全集 U 中,对于给定的变量 $X \in U$ 和变量集 $MB \subset U$ 且 $X \notin MB$,若满足式(6),则称能满足上述条件的最小变量集 MB 为 X 的马尔可夫毯.

$$X \perp \{U - MB - \{X\} | MB \quad (6)$$

由式(6)我们可以给出特征的近似马尔可夫毯:若存在特征集合 F ,对于给定特征 f_i ,使得特征子集 $MB_i \subset F$ 且 $f_i \notin MB_i$,满足式(7)时,称 MB_i 是特征 f_i 的马尔可夫毯.

$$P(F - MB_i - f_i, C | f_i, MB_i) = P(F - MB_i - f_i, C | MB_i) \quad (7)$$

由式(6)和式(7)我们可以定义冗余特征:对于上述特征集合 F ,如果满足式(8),即对于给定特征 f_i 和分类 C 是弱相关的,并且可以在 F 内找到它的近似马尔可夫毯 f_j ,那么特征 f_i 就是冗余特征,应该在 F 中移除.

$$\begin{cases} SU_{j,C} \geq SU_{i,C} \\ SU_{i,j} \geq SU_{i,C} \end{cases} \quad (8)$$

2.4 算法流程描述

本文提出的算法伪代码如算法1所示,主要分为

两大模块,第一模块是Filter式特征选择,分别利用SU(代码1-5行)和AMB(代码6-16行)进行特征初选,第二模块是Wrapper式(代码17-27行)特征选择,利用C4.5分类器,在第一模块选取出的候选特征子集种,依次选取排名靠前的特征放入最优特征集中,并根据这个最优特征集预处理数据集,一边记录分类器训练测试效果,一边进行最后的最优特征子集搜索,最后选择出分类器测试效果最好的特征子集作为最优特征子集.

算法1. 算法流程代码

输入: $F(f_1, f_2, \dots, f_N, C), \delta$ //training set and predefined threshold
输出: F_{best}, PR //the best feature set and precision rate

```

1. for  $f_i \in F$ 
2.   calculate  $SU(f_i, C)$ 
3.   if  $SU(f_i, C) \geq \delta$ 
4.     put  $f_i$  into  $F'$ 
5.   sort  $SU(f_i, C)$  by descending of  $F'$ 
6.   while  $F' \neq \emptyset$ 
7.      $f_j = \text{getFirstElement}(F')$ 
8.     while  $f_j \neq \emptyset$ 
9.        $f_i = \text{getNextElement}(F', f_j)$ 
10.      while  $f_i \neq \emptyset$ 
11.         $f'_i = f_i$ 
12.        if  $SU(f_i, f_j) \geq SU(f_i, C)$ 
13.          remove  $f_i$  from  $F'$ 
14.           $f_i = \text{getNextElement}(F', f'_i)$ 
15.        else  $f_i = \text{getNextElement}(F', f_j)$ 
16.       $f_j = \text{getNextElement}(F', f_j)$ 
17.  $F_{top} = F'$ 
18.  $F_{best}^* = \emptyset, PR_0 = 0$ 
19. while  $F_{top} \neq \emptyset$ 
20.   for  $f_i \in F_{top}$ 
21.     put  $f_i$  into  $F_{best}^*$ 
22.     Preprocessed training set  $S$  and testing set  $D$  using  $F_{top}$ 
23.     Training C4.5 model using training set  $S'$ 
24.     Testing C4.5 model using testing set  $D'$ 
25.     calculate  $PR_i$ 
26.     if  $PR_i \geq PR_{i-1}$ 
27.        $F_{best} = F_{best}^*$ 

```

3 实验分析

3.1 实验环境与数据集

本文使用Weka平台作为仿真环境,Weka是基于Java环境下开源的机器学习以及数据挖掘软件,该软件集合了大量能承担数据挖掘任务的机器学习算法,包括对数据进行预处理,分类、回归、聚类、关联规

则以及在新的交互式界面上的可视化^[15,16]。

本文的操作系统指纹数据集分别来自 Network-Miner、p0f、Nmap 的指纹文件^[17-19]设计的指纹特征库, 本文对这些指纹特征库进行整理, 形成本次实验的数据集。数据集中包含的操作系统类别详细信息如表 1 所示, 所有类别包含的特征和特征代表的涵义如表 2 所示。

表 1 数据集详细信息

操作系统类型	样本数量	占比 (%)
Windows 7	3562	20.756
Windows 8	1842	10.734
Windows 10	4938	28.775
Windows XP	1395	8.129
Linux 2.4	1423	8.292
Linux 2.6	2132	12.424
Linux 3	1869	10.891

表 2 数据集特征属性详细信息

特征属性	含义
MSS	最大报文长度
WIN	窗口大小
TTL	数据包的生存时间
WS	窗口扩大因子
LEN	IP报文头长度
D	短延迟位设定标志
S	选择性确认技术允许标志
T	时间戳
N	无操作
F	SYN、ACK位设定标志

3.2 实验流程

本实验根据操作系统的不同类别将操作系统指纹数据集分为 7 个数据子集, 每个子集中训练集 Training 与测试集 Test 的比例为 7:3。

本文提出算法实验流程如图 1 所示, 在实验过程中, 分别将未处理的全部特征数据集 (10 维特征)、本文提出的算法 FSSA、FCBF 特征提取算法^[12]、基于改进 SU 的特征选择算法^[1], 通过实验得到的特征子集数目、网络设备识别精确率、网络设备识别召回率、算法时间复杂度、实际算法分类消耗时间进行对比分析, 说明本文算法的优势; 此外, 在实验的基础上, 将本文算法的冗余特征分析部分的近似马尔可夫毯算法换成 mRMR 算法^[20] (后文中称之为 FSSM), 与本文提出的算法做对比实验, 验证马尔可夫毯算法的冗余特征分析能力。关于 SU 特征选择算法部分对于阈值 δ 的设

定参考了文献 [1] 中的参数设定, 本文算法阈值 δ 设为 0.02, 关于 mRMR 实验部分实验参数参考文献 [20], 此处不再赘述。

3.3 评价指标

本实验采用整体识别精确率和召回率来衡量本文特征选择算法对网络设备操作系统识别的精确性。其中精确率和召回率根据表 3 的定义如下:

$$(1) \text{ 精确率 (precision rate): } PR = \frac{TP}{TP+FP}$$

$$(2) \text{ 召回率 (recall rate): } RR = \frac{TP}{TP+FN}$$

3.4 实验结果与分析

3.4.1 算法识别精确率对比

按照实验方案进行实验, 可以得到各特征选择算法获得到的最优特征个数, 如表 4, 可以看出本文提出的算法在各类型样本的特征属性集上都有不错的降维效果, 在特征选择的数目上看, 对于个别操作系统类型, 其它 3 种算法和本文提出的算法大致相同, 但是本文提出的算法在各类型的数据集上得到的特征数目是基本持平的, 其他 3 类算法均有波动, 说明对于不同类别的数据集, 这 3 类算法的特征选择结果不稳定; 表 5 展示了各特征选择算法在各类操作系统类型上的整体识别精确率, 可以明显看出本文提出的算法在各个类型的设备识别精确率都较高, 其它算法虽然在个别类别上的识别精确率也不错, 但是在小类别的数据集上, 本文提出的算法明显优于其它 3 类算法, 这是由于在特征选择的过程中, 本文算法加入了 Wrapper 式的特征选择过程, 直接将分类器的性能作为特征选择的评价标准, 这大大提高了本文算法在设备识别上的精确率; 表 6 展示了各特征选择算法在各类操作系统类型上的召回率, 在特征选择的过程中, 本文算法能够在小类别的数据 (比如 Windows XP 和 Linux 2.4) 的特征属性中选择出更具有强相关性的特征, 能够减少小类别数据被错分为其他类别的数量, 所以本文算法无论是在小类别数据还是大类别数据上表现都优于其他算法。

FSSA 算法和 FSSM 算法整体来看特征选择数目相对来说比较平均, 但是在小类别数据上可以看出, FSSM 算法选择出的特征明显多于本文算法; 在设备识别准确率和召回率上看, 本文算法明显优于 FSSM, 这是因为 mRMR 算法原理是用互信息来衡量特征于类别、特征与特征之间的相关性, 而本文算法中的马尔可夫毯算法部分利用的是对称不确定性来衡量特征

与类别、特征与特征之间的相关性,实验数据也表明,在相同的条件下进行特征相关性分析、不同的条件下

进行特征冗余性分析,本文算法选用的马尔可夫毯在冗余特征分析上效果更好.

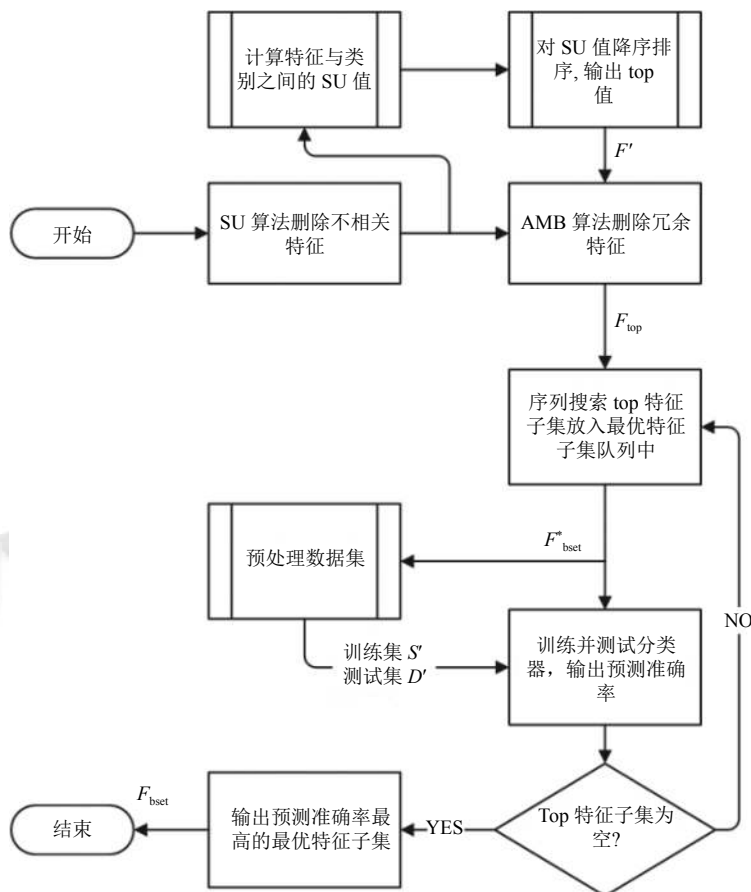


图 1 算法实验流程图

表 3 二分类矩阵

	实际正类	实际负类
预测正类	TP	FP
预测负类	FN	TN

3.4.2 算法复杂度对比

假定当前特征总数为 n , m 是数据集的实例总数. 表 7 定性分析了各特征选择算法的时间复杂度^[7,12,21], 表 8 展示了实际实验过程中,应用各特征选择算法后对数据进行分类的时间,从表中可以看出,本算法由

于加入了 Wrapper 式特征选择方法,所以在特征选择的时间复杂度上比较大,在实际运行的特征选择过程中时间也较长,但是由于本文算法选择出的特征相关性高,冗余小,所以在分类时运行的时间大大缩小. 本文所采用的算法和 FCBF 算法均使用马尔可夫毯算法进行特征冗余性分析,在实际分类时间上均比使用 mRMR 算法进行冗余性分析的 FSSM 要短,所以本文选用的马尔可夫毯算法在进行冗余性分析时效果更好.

表 4 各特征选择算法所选特征数目(个)

操作系统类型	Windows 7	Windows 8	Windows 10	Windows XP	Linux 2.4	Linux 2.6	Linux 3	平均值
未处理特征集	10	10	10	10	10	10	10	10
FSSA算法	6	6	6	5	5	6	6	6
FSSM算法	6	7	6	6	6	5	6	6
FCBF算法	7	6	7	5	5	6	7	6
改进SU算法	7	6	7	5	5	7	6	6

表5 各特征选择算法的精确率 (%)

操作系统类型	Windows 7	Windows 8	Windows 10	Windows XP	Linux 2.4	Linux 2.6	Linux 3	平均值
未处理特征集	90.125	83.21	84.653	80.256	82.369	84.109	84.065	84.112
FSSA算法	93.012	86.963	93.103	90.136	89.331	89.562	90.101	90.315
FSSM算法	91.763	90.025	91.106	88.659	86.56	89.013	88.635	89.394
FCBF算法	92.865	86.568	91.24	88.14	85.124	88.661	87.154	88.536
改进SU算法	93.069	86.005	93.885	89.33	84.721	89.625	89.256	89.413

表6 各特征选择算法的召回率 (%)

操作系统类型	Windows 7	Windows 8	Windows 10	Windows XP	Linux 2.4	Linux 2.6	Linux 3	平均值
未处理特征集	75.121	68.316	64.213	39.562	46.386	66.115	59.35	59.866
FSSA算法	81.121	75.301	70.142	60.124	68.75	76.863	69.352	71.664
FSSM	80.236	73.356	68.342	52.031	59.892	72.301	60.364	66.646
FCBF算法	79.254	72.356	69.386	40.265	59.22	69.163	59.561	64.171
改进SU算法	76.32	68.253	65.896	41.352	49.319	65.389	56.354	69.413

表7 各算法时间复杂度

算法名称	时间复杂度
FSSA	$O(mn \log_2 n + mn)$
FSSM	$O(mn \log_2 n + mn^2)$
FCBF	$O(mn \log_2 n)$
改进的SU	$O(mn^2)$

表8 各算法实际分类时间 (s)

算法名称	分类时间
FSSA	2.36
FSSM	3.42
FCBF	2.81
改进的SU	3.51

4 结论与展望

本文主要是在面向基于网络流量分析的网络设备识别应用中,网络流量特征选取的问题进行了研究,针对这个问题提出了一种将 Filter 和 Wrapper 方式相结合,基于 SU 和 AMB 的网络流量特征选择算法 FSSA. 实验表明,该方法下选择出的特征对网络设备操作系统类型识别的精确率相较于经典的特征选择方法有了一定的提高,在小类别数据上的召回率也得到了提升,但是在进行 Wrapper 式特征搜索时性能代价也较高. 所以针对实验过程中的问题,如何对网络设备更多版本的操作系统类型进行识别,如何更新完善目前的数据集,如何提高模型的性能,以及除了操作系统信息之外,如何获取网络设备其它信息是我们下一步的工作重点^[22].

参考文献

- 刘翔元. 基于网络流量分析的网络设备类型识别关键技术研究 [硕士学位论文]. 南京: 南京邮电大学, 2019.
- Aluthge N. IoT device fingerprinting with sequence-based features [Master's thesis]. Helsinki: University of Helsinki, 2018.
- Guo H, Heidemann J. IP-based IoT device detection. Proceedings of the 2018 Workshop on IoT Security and Privacy. ACM, 2018. 36–42. [doi: 10.1145/3229565.3229572]
- Salman O, Elhadj IH, Chehab A, et al. A machine learning based framework for IoT device identification and abnormal traffic detection. Transactions on Emerging Telecommunications Technologies, 2019. [doi: 10.1002/ett.3743]
- Ji SY, Jeong BK, Choi S, et al. A multi-level intrusion detection method for abnormal network behaviors. Journal of Network and Computer Applications, 2016, 62: 9–17.
- 刘雪亚, 姜志侠, 徐轩, 等. 基于卡方方法及对称不确定性的网络流量特征选择方法. 长春理工大学学报 (自然科学版), 2019, 42(2): 74–78.
- 唐宏, 刘丹, 姚立霜, 等. 面向类不平衡网络流量的特征选择算法. 电子与信息学报, 2021, 43(4): 923–930.
- 赵择, 徐佑宇, 唐亮, 等. 基于改进一对一算法的网络流量分类. 中国科学院大学学报, 2020, 37(4): 570–576.
- 曹杰. 基于 SVM 的网络流量特征降维与分类方法研究 [博士学位论文]. 长春: 吉林大学, 2017.
- Nguyen-An H, Silverston T, Yamazaki T, et al. Entropy-based IoT devices identification. 2020 21st Asia-Pacific Network Operations and Management Symposium (APNOMS). Daegu: IEEE, 2020. 73–78.
- Jeon S, Yun JH, Choi S, et al. Passive fingerprinting of SCADA in critical infrastructure network without deep

- packet inspection. arXiv: 1608.07679, 2016.
- 12 Yu L, Liu H. Feature selection for high-dimensional data: A fast correlation-based filter solution. Proceedings of the Twentieth International Conference on International Conference on Machine Learning. Washington, DC: ACM, 2003. 856–863.
- 13 何宪. 基于贝叶斯网络的马尔可夫毯发现算法研究 [硕士学位论文]. 成都: 电子科技大学, 2012.
- 14 Lee CY, Lin WC. Induction motor fault classification based on FCBF-PSO feature selection method. Applied Sciences, 2020, 10(15): 5383. [doi: [10.3390/app10155383](https://doi.org/10.3390/app10155383)]
- 15 刘纪伟, 赵月显, 赵杨. 一种基于统计排序的网络流量特征选择方法. 电子技术应用, 2018, 44(1): 84–87.
- 16 皮寿熹, 李富合, 缪磊, 等. 基于分类算法的网络设备识别方法. 舰船电子工程, 2019, 39(12): 129–132, 157.
- 17 李良盛, 段海新, 郑晓峰. 基于 HTTP User-Agent 标记的被动操作系统识别指纹库自动生成方法. 计算机应用与软件, 2020, 37(5): 309–314, 326.
- 18 黄引翔. 网络流量分类中特征工程的研究 [硕士学位论文]. 南京: 南京邮电大学, 2017.
- 19 NETRESEC. NetworkMiner. <https://www.netresec.com/>. (2015-10-11).
- 20 张俐, 王枳, 郭文明. 利用近似马尔可夫毯的最大相关最小冗余特征选择算法. 西安交通大学学报, 2018, 52(10): 141–145.
- 21 Cvitić I, Peraković D, Periša M, *et al.* Ensemble machine learning approach for classification of IoT devices in smart home. International Journal of Machine Learning and Cybernetics, 2021. [doi: [10.1007/s13042-020-01241-0](https://doi.org/10.1007/s13042-020-01241-0)]
- 22 Chowdhury RR, Aneja S, Aneja N, *et al.* Network traffic analysis based IoT device identification. Proceedings of the 2020 the 4th International Conference on Big Data and Internet of Things. Singapore: IEEE, 2020. 79–89.