Journal of Computer Applications
计算机应用,2013,33(1):80-82,87
ISSN 1001-9081
CODEN JYIIDU
2013-01-01
http://www.joca.cn

文章编号:1001-9081(2013)01-0080-03
doi:10.3724/SP.J.1087.2013.00080

基于互信息选择聚类集成的网络流量分类方法

丁要军[1,2*],蔡皖东[1]
(1.西北工业大学 计算机学院,西安 710129; 2.咸阳师范学院 信息工程学院,陕西 咸阳 712000)
(*通信作者电子邮箱 dingyaojun@yahoo.cn)

摘 要:针对互联网流量标注困难以及单个聚类器的泛化能力较弱,提出一种基于互信息(MI)理论的选择聚类集成方法,以提高流量分类的精度。首先计算不同初始簇个数K的K均值聚类结果与训练集中流量协议的真实分布之间的规范化互信息(NMI);然后基于NMI的值来选择用于聚类集成的K均值基聚类器的K值序列;最后采用二次互信息(QMI)的一致函数生成一致聚类结果,并使用一种半监督方法对聚类簇进行标注。通过实验比较了聚类集成方法与单个聚类算法在4个不同测试集上总体分类精度。实验结果表明,聚类集成方法的流量分类总体精度能达到90%。所提方法将聚类集成模型应用到网络流量分类中,提高了流量分类的精度和在不同数据集上的分类稳定性。

关键词:聚类集成;K均值;流量分类;互信息
中图分类号:TP393.06 **文献标志码**:A

Internet traffic classification method based on selective clustering ensemble of mutual information

DING Yaojun[1,2*], CAI Wandong[1]
(1. *School of Computer Science, Northwestern Polytechnical University, Xi'an Shaanxi* 710129, *China*;
2. *School of Information Engineering, Xianyang Normal University, Xianyang Shaanxi* 712000, *China*)

Abstract: Because it is difficult to label Internet traffic and the generalization ability of single clustering algorithm is weak, a selective clustering ensemble method based on Mutual Information (MI) was proposed to improve the accuracy of traffic classification. In the method, the Normalized Mutual Information (NMI) between clustering results of K-means algorithm with different initial cluster number and the distribution of protocol labels of training set was computed first, and then a serial of K which were the initial cluster number of K-means algorithm based on NMI were selected. Finally, the consensus function based on Quadratic Mutual Information (QMI) was used to build the consensus partition, and the labels of clusters were labeled based on a semi-supervised method. The overall accuracies of clustering ensemble method and single clustering algorithm were compared over four testing sets, and the experimental results show that the overall accuracy of clustering ensemble method can achieve 90%. In the proposed method, a clustering ensemble model was used to classify Internet traffic, and the overall accuracy of traffic classification along with the stability of classification over different dataset got enhanced.

Key words: clustering ensemble; K-means; traffic classification; Mutual Information (MI)

0 引言

互联网流量分类是网络管理和网络安全的基础,也是了解用户上网行为,提高网络服务质量的重要途径。近几年基于机器学习的流量分类方法成为国内外研究的热点,这类方法无需对应用层负载进行检测,保护了用户隐私,而且可以对加密流量进行分类。目前的成果中大多使用有监督学习的分类模型[1-3],需要大量标注准确的训练样本来训练模型,随着应用协议的不断升级以及加密技术的广泛应用,准确标注变得非常困难。如果无法获得一定数量的标注准确的训练样本,所有的监督学习方法的识别精度都无法保障。无监督的聚类方法对标注样本的依赖大大降低,并且可以识别新的未知协议流量,Bernaille等[4]和Erman等[5]等率先将聚类方法应用到流量分类中,取得了很好的实验效果。Bernaille等[4]使用规范化互信息(Normalized Mutual Information, NMI)来确定聚类的初始簇个数,训练样本中协议的分布与聚类结果中簇的分布比较相似度,协议的真实分布作为最优聚类,与不同的聚类个数K产生的聚类结果进行比较,但最终还是确定了一个K值,使用单一的聚类器实现分类,鲁棒性偏低。Erman等[5]将K-means算法和DBSCAN(Density-Based Spatial Clustering of Applications with Noise)算法应用到流量分类中,并与已经应用于流量分类的AutoClass算法进行了比较,发现K-means和DBSCAN的时间效率较高。刘琼等[6]对现有的流量分类算法进行了比较,并指出单个算法无法实现对所有协议类别的准确分类和识别,多分类器的结合是流量分类的发展方向。

集成学习是机器学习领域研究的热点,集成学习可以有效地实现多分类器融合,本文将聚类集成方法[7-9]引入到流量分类中,提高了流量分类的鲁棒性和精确度。Strehl等[7]首次提出了基于互信息理论的聚类集成思路,但并未直接给出计算一致聚类结果的方法;Topchy等[8]给出了完整的基于互信息的聚类集成方法;唐伟等[9]提出一种基于互信息加权的选择聚类集成方法,提高了集成效率;罗会兰[10]对聚类集成中的一致函数进行了分析、比较,并指出基于互信息的方法

收稿日期:2012-08-02;**修回日期**:2012-08-29。 **基金项目**:国家863计划项目(2009AA01Z424);陕西省教育厅专项(12JK0933)。
作者简介:丁要军(1980-),男,河南许昌人,讲师,博士研究生,主要研究方向:网络与信息安全; 蔡皖东(1955-),男,陕西西安人,教授,博士生导师,主要研究方向:网络安全、信息对抗。

效率更高,因此比较适合流量分类。

本文在上述文献的基础上提出一种基于 NMI 的基聚类器选择方法,并基于平方互信息(Quadratic Mutual Information, QMI)生成一致聚类结果。

1　相关概念和原理

1.1　流量的抽象表示

在一段时间间隔内,具有相同源 IP、目的 IP、源端口、目的端口、传输层协议的网络报文序列称为网络流,这 5 个属性也称为五元组。

流统计特征是以流为单位计算出来的统计特征,包括流的报文大小的期望和方差、报文到达时间间隔的期望和方差等。

已知 1 组网络流即标注样本集 $X=\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n\}$,其中 $\boldsymbol{X}_i=(x_{i1},x_{i2},\cdots,x_{ij},\cdots,x_{im})$,$x_{ij}$ 是网络流的第 j 个统计特征。n 个网络流样本对应的协议类别为 $\{C_1,C_2,\cdots,C_i,\cdots,C_n\}$,$C_i$ 取值属于集合 $C=\{c_1,c_2,\cdots,c_r,\cdots,c_k\}$,$C_i$ 与 C_j 的取值相同时表示 2 条网络流的协议相同。

流量分类的目的是训练分类模型 $f:X\rightarrow C$,判断网络流 $\boldsymbol{X}_i$ 的协议类别 C_i。

1.2　聚类集成的原理

假设给定一个包含 n 个样本的数据集 $X=\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n\}$,一组由 H 种不同的聚类方法对数据集 X 产生的不同聚类划分集合 $\Pi=\{\pi_1,\cdots,\pi_H\}$,使用 H 种聚类方法对于样本 $\boldsymbol{X}_i$ 进行聚类划分的结果可以表示为:$\boldsymbol{X}_i\rightarrow\{\pi_1(\boldsymbol{X}_i),\pi_2(\boldsymbol{X}_i),\cdots,\pi_H(\boldsymbol{X}_i)\}$,最终通过一致函数产生一个一致聚类结果。

聚类集成主要考虑两个问题:一是如何产生不同的聚类划分;二是生成一致的聚类结果。本文算法将考虑使用不同 K 值的 K-means 算法产生不同的聚类划分,基于互信息理论来生成一致聚类结果。因为流量分类中的数据量较大,K-means 算法是聚类算法中效率较高的[4-5]。

1.3　聚类划分的表示方式

如何将 H 种不同的聚类算法对数据集 X 的聚类结果用一种唯一的表示方式来表示非常重要,参照文献[7]的方法规定如下:

1) 保证样本 $\boldsymbol{X}_1$ 的簇标签为 1;

2) 保证后面样本的簇标签大于等于前面已经出现的簇标签。

假设两个聚类算法 π_1 和 π_2 对一个简单数据集 $S=\{S_1,S_2,S_3,S_4,S_5\}$ 进行聚类划分,划分结果可以用样本所属的簇标签组成的向量表示,分别用 $\boldsymbol{\lambda}^{(1)}$ 和 $\boldsymbol{\lambda}^{(2)}$ 表示,$\boldsymbol{\lambda}^{(1)}=(1,1,2,2,3)^{\mathrm{T}}$,$\boldsymbol{\lambda}^{(2)}=(2,2,1,1,3)^{\mathrm{T}}$,虽然两个划分结果的表示方式不一样,但聚类划分的结果是一致的。按照上面的要求就可以把划分结果一致的向量用唯一的方式来表示,方便集成阶段的计算。

2　基于聚类集成的流量分类

2.1　基于 NMI 的 K 值序列选择

训练集由一组标注好协议类型的流量样本组成,将训练集中协议的真实分布作为最优聚类结果,用 π_C 表示,不同的 K 值产生的不同的 K-means 算法对训练集的聚类划分用 π_i 表示,则 π_C 与 π_i 的 NMI 计算如下:

$$NMI(\pi_C,\pi_i)=\frac{I(\pi_C,\pi_i)}{\sqrt{H(\pi_C)H(\pi_i)}} \tag{1}$$

$$I(\pi_C,\pi_i)=\sum_{r=1}^{K}\sum_{j=1}^{K(i)}p(C_r,L_j^i)\lg\left(\frac{p(C_r,L_j^i)}{p(C_r)p(L_j^i)}\right) \tag{2}$$

其中:K 表示 π_C 的划分中簇的个数;$K(i)$ 表示 π_i 的聚类划分中簇的个数;$p(C_r)$ 表示在 π_C 的划分中,样本被划分到簇 r 的概率;$p(L_j^i)$ 表示在 π_i 的划分中,样本被划分到簇 j 的概率;$p(C_r,L_j^i)$ 表示在 π_C 的划分和 π_i 的划分中,样本同时被划分到簇 r 和簇 j 中的概率。

假设训练集中样本总数为 n,在 π_C 的划分中,被划分到簇 r 的样本数为 n_r;在 π_i 的划分中,被划分到簇 j 的样本数为 $n_j^{(i)}$;在 π_C 的划分和 π_i 的划分中,被同时划分到簇 r 和簇 j 中的样本数为 n_{rj},则式(2)可以表示为:

$$I(\pi_C,\pi_i)=\sum_{r=1}^{K}\sum_{j=1}^{K(i)}\frac{n_{rj}}{n}\lg\left(\frac{n\cdot n_{rj}}{n_r\cdot n_j^{(i)}}\right) \tag{3}$$

$$H(\pi_C)=\sum_{r=1}^{K}\frac{n_r}{n}\lg\frac{n_r}{n} \tag{4}$$

$$H(\pi_i)=\sum_{j=1}^{K(i)}\frac{n_j^{(i)}}{n}\lg\frac{n_j^{(i)}}{n} \tag{5}$$

根据不同 K 值的 K-means 算法聚类结果与最优聚类结果在不同训练集上计算出的 NMI 值的分布情况来确定 K 值序列。

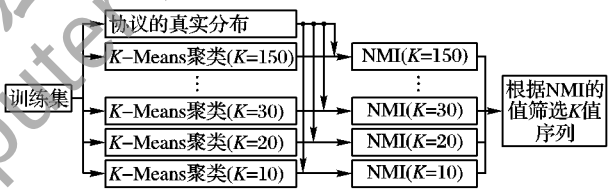

图 1　基于 NMI 的 K 值序列选择示意图

2.2　基于 QMI 的一致聚类生成

假设待分类数据集中样本总数为 n,一组由 2.1 节中选择的不同 K 值的 K-means 聚类算法产生的聚类划分 $\Pi=\{\pi_1,\cdots,\pi_i,\cdots,\pi_H\}$,其中:$\pi_i=\{L_1^i,\cdots,L_j^i,\cdots,L_{K(i)}^i\}$,$L_j^i$ 表示聚类划分 π_i 下的第 j 个簇,共有 $K(i)$ 个簇。对任意两个聚类划分 π_s 和 π_t,定义函数 $U(\pi_s,\pi_t)$[8],如式(6)所示:

$$U(\pi_s,\pi_t)=\sum_{r=1}^{K(s)}p(L_r^s)\sum_{j=1}^{K(t)}p(L_j^t\mid L_r^s)^2-\sum_{j=1}^{K(t)}p(L_j^t)^2 \tag{6}$$

其中:L_r^s 表示聚类划分 π_s 下的第 r 个簇,共有 $K(s)$ 个簇;L_j^t 表示聚类划分 π_t 下的第 j 个簇,共有 $K(t)$ 个簇。则有:

$$p(L_r^s)=|L_r^s|/n \tag{7}$$

$$p(L_j^t)=|L_j^t|/n \tag{8}$$

$$p(L_j^t\mid L_r^s)=\left|L_j^t\cap L_r^s\right|/|L_r^s| \tag{9}$$

其中:$|L_r^s|$ 表示簇 L_r^s 中包含的样本个数,$|L_j^t|$ 表示簇 L_j^t 中包含的样本个数,$\left|L_j^t\cap L_r^s\right|$ 表示同时属于簇 L_j^t 和簇 L_r^s 的样本个数。

函数 $U(\pi_s,\pi_t)$ 表示了两个聚类划分的相关性,聚类划分 π_s 与聚类划分集合 $\Pi=\{\pi_1,\cdots,\pi_H\}$ 中的所有划分的相关性的和用 $U(\pi_s,\Pi)$ 表示,如式(10):

$$U(\pi_s,\Pi)=\sum_{i=1}^{H}U(\pi_s,\pi_i) \tag{10}$$

则最优的聚类划分:

$$\pi_s^{\text{best}} = \arg\max_{\pi_s} U(\pi_s, \Pi) \tag{11}$$

最优聚类划分也是聚类集成的一致聚类结果,表示了与其他聚类划分互信息最大的聚类结果。

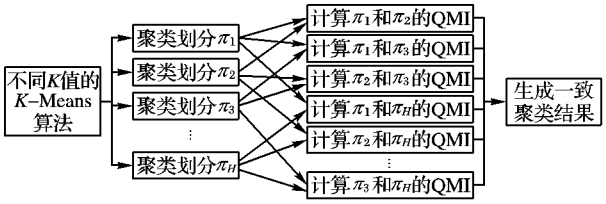

图 2 基于 QMI 的一致聚类生成示意图

2.3 聚类结果标记

聚类集成的结果只是为每一个样本指定了簇号,还需要对每一个聚类簇所属的协议类别进行标注,这也是无监督学习算法的缺点。最优的聚类结果是每一个簇只包含一种协议的样本,但通常都会存在聚类误差。本文参考文献[11]的方法,提出一种半监督策略的聚类结果标注方法,实现过程如下。

1) 将聚类集成算法在已标注的训练集与未标注的测试集组成的混合数据集上进行聚类,生成一致聚类结果。

2) 假设一致聚类结果中第 i 个簇中共包含 n_i 个样本,其中:已标注样本的个数为 n_{il},未标注样本的个数为 n_{iu},n_{il} 个样本中属于协议类别 c_r 的样本个数记为 n_{il}^r。

3) 第 i 个簇中的任意一个样本 $\boldsymbol{X}_j^i$ 属于协议 c_r 的概率 $P(c_r)$ 为:

$$P(c_r) = n_{il}^r / n_i \tag{12}$$

4) 最终第 i 个簇协议类别 c 为:

$$c = \arg\max_{c_r}(P(c_r)) \tag{13}$$

3 实验结果与分析

3.1 实验数据集

使用英国剑桥大学计算机实验室提供的 08simple 数据集[12]进行实验,数据集中包含了三天的采样数据,数据集中包含的流统计特征共有 13 个,具体特征请参考文献[12]。本文删除了原始数据集中流个数较少的协议数据,最后数据集中剩余 5 类常用协议,具体包含的协议种类如表 1 所示。

表 1 实验数据中包含的协议类别

类别	协议
BULK	FTP
DATABASE	Postgres,SQLNet Oracle,Ingres
MAIL	IMAP,POP2/3,SMTP
WWW	WWW
P2P	KaZaA,BitTorrent,GnuTella

将 Day1 的数据分成 2 个训练集,Day2 和 Day3 的数据分成 4 个测试集,每类协议包含的网络流条数如表 2 所示。

表 2 各种协议包含的网络流的条数

协议	训练集 1	训练集 2	测试集 1	测试集 2	测试集 3	测试集 4
WWW	16 392	49 719	45 034	56 049	66 893	52 220
MAIL	3 632	4 141	5 817	770	1 890	1 211
BULK	1 669	1 301	4 559	1 327	1 862	1 323
P2P	203	526	995	4 268	783	5 560
DATABASE	190	804	808	1 478	715	1 965
共计	22 086	56 491	57 213	63 892	72 143	62 279

3.2 基聚类器参数训练

本文使用不同 K 值的 K-means 算法来产生基聚类器,K-means 算法的距离参数选择欧氏距离[5]。将训练集中协议的真实分布作为最优分类结果[4],计算不同 K 值的 K-means 算法在训练集上的聚类结果与最优分类结果的 NMI 值,在训练集 1 和训练集 2 上的实验结果如图 3 所示。

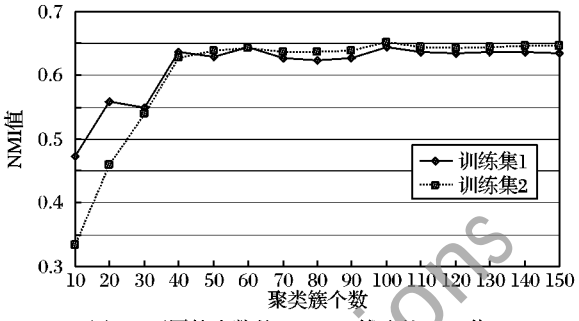

图 3 不同簇个数的 K-means 算法的 NMI 值

文献[4]中只确定了一个 K 值来实现了单聚类器的流量聚类,单个分类器的鲁棒性较差,本算法选取一组 K 值序列来生成多个基聚类器,最后通过集成的方法来确定最优聚类结果。从实验结果可以看出,当簇个数大于 40 时,NMI 的值趋于稳定,并且 NMI 的值较高。考虑到更大的簇个数导致计算复杂度更高,选取[40,50,60,70,80,90,100]这 7 个取值为最终的 K 值序列。

3.3 分类评价标准及分类结果

3.3.1 分类评价标准

总体分类精度 所有类别中被正确分类的样本数占所有样本总数的百分比。

鲁棒性 分类模型在不同测试集上分类精度的变化幅度,变化幅度较小则鲁棒性较好。

分类时间 分类模型完成对测试集的分类所需要的时间。

3.3.2 分类精度对比

在 4 个测试集上分别运行了单个聚类器和集成聚类器,总体分类精度如图 4 所示。

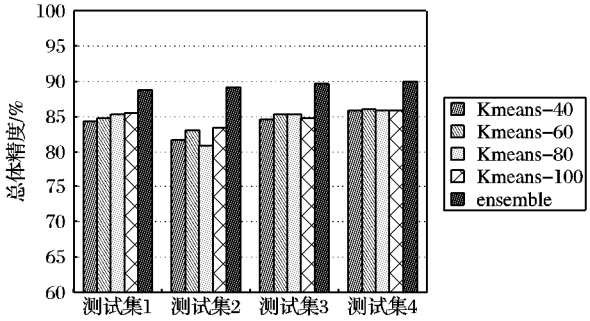

图 4 集成聚类算法与单个聚类器的总体精度对比

其中:Kmeans-40、Kmeans-60、Kmeans-80、Kmeans-100 分别表示 K 值为 40,60,80,100 的单聚类器;ensemble 表示由 7 个基聚类组成的集成聚类器。从图 4 可以看出,在 4 个不同的测试集上集成聚类算法的总体分类精度均高于不同 K 值的单聚类器。单聚类器在不同测试集上的总体分类精度存在一定的波动,而集成聚类器的总体分类精度始终保持在 90% 左右,鲁棒性较强。

3.3.3 时间效率对比

实验中记录了 5 种模型在不同测试集上的分类时间,如表 3 所示。

(下转第 87 页)

2004, 22(4): 747 - 756.

[3] YAMANEGI K, HASEGAWA G, MURATA M. Achieving predictable throughput of TCP based on inline network measurement [J]. IEICE Technical Report, 2006, 106(237): 7 - 12.

[4] SHIMONISHI H, HAMA T, MURASE M. TCP congestion control enhancements for streaming media [C]// CCNC 2007: Proceedings of the 4th IEEE Consumer Communications and Networking Conference. Piscataway: IEEE Press, 2007: 303 - 307.

[5] HASHIMOTO M, HASEGAWA G, MURATA M. Trade-off evaluation between fairness and throughput for TCP congestion control mechanisms in a wireless LAN environment [C]// 2010 International Symposium on Performance Evaluation of Computer and Telecommunication System. Ottawa: Society for Modeling and Simulation International, 2010: 22 - 29.

[6] LAI C D, LEUNG K C, LI V O K. Enhancing wireless TCP: a serialized-timer approach [C]// Proceedings of the 2010 IEEE International Conference on Computer Communications. Piscataway: IEEE Press, 2010: 391 - 395.

[7] SHIN K, KIM J, CHOI S B. Loss recovery scheme for TCP using MAC MIB over wireless access networks [J]. IEEE Communications Letters, 2011, 15(10): 1059 - 1061.

[8] MAHMOODI T, FRIDERIKOS V, HOLLAND O, *et al.* Cross-layer design to improve wireless TCP performance with link-layer adaptation [C]// VTC-Fall 2007: IEEE 66th Vehicular Technology Conference. Piscataway: IEEE Press, 2007: 1504 - 1508.

[9] MENDES L D P, BRITO J M C. Some analysis of a cross-layer design for a wireless TCP network [C]// ICWMC'09: Proceedings of the Fifth International Conference on Wireless and Mobile Communications. Piscataway: IEEE Press, 2009: 64 - 69.

[10] ANDO R, MURASE T, OGUCHI M. Characteristics of QoS-TCP on real mobile terminal in wireless LAN [C]// 2011 IEEE International Workshop Technical Committee on Communications Quality and Reliability. Piscataway: IEEE Press, 2011: 1 - 6.

[11] CHUA K C, MALCOLM J A, ZHANG Y. Theoretical analysis of TCP throughput in Ad Hoc wireless networks [C]// IEEE Global Telecommunications Conference 2005. Piscataway: IEEE Press, 2005: 2714 - 2719.

[12] GEETHA V, AITHAL S, CHANDRASEKARAN K. Effect of mobility over performance of the Ad Hoc networks [C]// Proceedings of the 2006 International Symposium on Ad Hoc and Ubiquitous Computing. Piscataway: IEEE Press, 2006: 138 - 141.

[13] LIU C, WU J. Adaptive routing in dynamic Ad Hoc networks [C]// 2008 IEEE Wireless Communications and Networking Conference. Piscataway: IEEE Press, 2008: 2603 - 2608.

[14] PERKINS C, BELDING-ROYER E, DAS S. Ad Hoc On-Demand Distance Vector (AODV) routing, RFC3561 [S]. IETF, 2003.

[15] HAMAD S, NOUREDDINE H, RADHI N, *et al.* Efficient flooding based on node position for mobile Ad Hoc network [C]// IIT 2011: 2011 International Conference on Innovations in Information Technology. Piscataway: IEEE Press, 2011: 162 - 166.

(上接第 82 页)

表 3　5 种模型的分类时间(CPU-seconds)

模型	测试集 1	测试集 2	测试集 3	测试集 4
Kmeans-40	2.47	2.58	2.86	2.63
Kmeans-60	2.55	2.64	2.94	2.65
Kmeans-80	2.58	2.75	2.98	2.78
Kmeans-100	3.32	2.96	3.22	2.92
ensemble	4.12	4.27	4.62	4.31

因为聚类集成模型要实现对单聚类器聚类结果的集成,所以分类时间要略高于单个聚类器。由于采用了 QMI 的集成方法,集成效率比较高,总体的分类时间与单聚类器差异不大。考虑到在分类精度上的优势,基于聚类集成模型的流量分类方法要优于单个聚类器模型。

4　结语

与监督学习算法相比,聚类算法对标注样本的依赖较小,更适合于标注困难的网络流量分类,但单个聚类器在不同的聚类参数影响下聚类结果差异较大。集成学习是机器学习发展的新方向,本文提出一种基于聚类集成的流量分类方法,提高了聚类器的鲁棒性和总体分类精度。

参考文献:

[1] MOORE A W, ZUEV D. Internet traffic classification using Bayesian analysis techniques [C]// Proceedings of the 2005 ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems. New York: ACM Press, 2005: 50 - 60.

[2] 徐鹏, 刘琼, 林森. 基于支持向量机的 Internet 流量分类研究[J]. 计算机研究与发展, 2009, 46(3): 407 - 414.

[3] 胡婷, 王勇, 陶晓玲. 混合模式的网络流量分类方法[J]. 计算机应用, 2010, 30(10): 2653 - 2655.

[4] BERNAILLE L, TEIXEIRA R, SALAMATIAN K. Early application identification [C]// Proceedings of the 2006 ACM Conference on Emerging Networking Experiments and Technologies. New York: ACM Press, 2006: 70 - 82.

[5] ERMAN J, ARLITT M, MAHANTI A. Traffic classification using clustering algorithms [C]// Proceedings of the 2006 SIGCOMM Workshop on Mining Network Data. New York: ACM Press, 2006: 281 - 286.

[6] 刘琼, 刘珍, 黄敏. 基于机器学习的 IP 流量分类研究[J]. 计算机科学, 2010, 37(12): 35 - 40.

[7] STREHL A, GHOSH J. Cluster ensembles - a knowledge reuse framework for combining multiple partitions [J]. Journal of Machine Learning Research, 2003, 3(3): 583 - 617.

[8] TOPCHY A, JAIN A K, PUNCH W. Clustering ensembles: models of consensus and weak partitions [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(12): 1866 - 1881.

[9] 唐伟, 周志华. 基于 Bagging 的选择性聚类集成[J]. 软件学报, 2005, 16(4): 496 - 502.

[10] 罗会兰. 聚类集成关键技术研究[D]. 杭州: 浙江大学, 2007.

[11] ERMAN J, MAHANTI A, ARLITT M. Semi-supervised network traffic classification [C]// Proceedings of the 2007 ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems. New York: ACM Press, 2007: 369 - 370.

[12] LI W, CANINI M, MOORE A W, *et al.* Efficient application identification and the temporal and spatial stability of classification schema [J]. Computer Networks, 2009, 53(6): 790 - 809.