一种增量式的加权集成学习流量分类方法

王 艳， 程 光， 潘吴斌 ， 郭晓军

(1.东南大学计算机科学与工程学院，南京，210096；2.计算机网络和信息集成教育部重点实验室（东南大学），南京，210096)

摘 要： 由于网络流量特征随着时间和环境的变化而发生改变，机器学习分类方法很难保持稳定的分类性能。如果仅仅根据过去或当前流量建立的分类器存在过时或丢失先验知识的问题，结合两者流量建立的分类器将会影响分类器的性能。因而，本文提出一种增量式的加权集成学习流量分类方法，使用先前流量建立分类器，然后使用增量学习方法引入新环境流量更新并学得自适应分类器，再根据精度加权的集成学习方法综合分类结果。实验结果表明该算法在处理流量的概念漂移问题上表现出较好的分类性能和泛化能力，分类效率满足分类实时性要求。

关键词： 增量学习，加权集成学习，流量分类，机器学习

An Incremental Traffic Classification Approach based on Weighted Ensemble Learning

Wang Yan, Cheng Guang, Pan Wubing , Guo Xiaojun

(1.School of Computer Science and Engineering, Southeast University, Nanjing 210096;
2.Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 210096)

Abstract: As the characteristics of network traffic change with time and environment, machine learning classification methods are difficult to maintain stable classification performance. Classifier trained based on previous or current traffic exist the problem of outdated or missing prior knowledge, however, a combination of both traffic would affect the efficiency. In this paper, we propose a traffic classification system with incremental weighted ensemble learning. We firstly use previously flow to establish the classifiers. New environmental traffic is introduced to update and learn adaptive classifiers based on incremental learning schema, and then integrate the classification results based on accuracy-weighted ensemble classifier. Experimental results show that the algorithm possesses better classification performance and generalization ability to handle the traffic with concept drift, and classification efficiency meet the real-time requirements.

Key words: Incremental learning, Weighted ensemble learning, Traffic classification, Machine learning

随着 Internet 的快速发展，P2P、流媒体及网络游戏等新应用已经占据了 60%以上的网络流量[1]，鉴于每种应用都有独特的流量行为特征，统计不同应用的流量可以获知用户使用网络的行为，发现影响网络资源分布的新应用。近年来，基于传输层的统计特征的机器学习分类方法研究广泛[2-5]，但很少关注到分类过程中普遍存在的概念漂移问题[6]，即流量特征和分布随着时间的推移或者环境的变化而发生改变。概念漂移导致分类模型适用性降低，在实际应用中存在以下问题：1)只在新的流量上训练新的分类器将导致一些历史知识丢失，而且重新标记样本成本过高；2)结合不同时期收集的所有流量训练分类器会导致性能问题[7]。此外，如果某个特定时期具有较大的数据量，将对流量分类起主导作用。3)随着 P2P 和流媒体等新应用的不断出现，无法收集和分析完整的训练样本，使得现有的流量分类方法在有些网络环境下分类准确率较低。4)基于监督学习的方法具有较好的分类性能，但标记样本

基金项目： 国家高技术研究发展计划（863 计划）（2015AA010201）；

作者简介： 王艳，（1991-），女，硕士研究生，E-mail: yanwang@njnet.edu.cn;程光，（1973-），男，教授，博导，E-mail: gcheng@njnet.edu.cn.

难以收集，而且无法识别未知应用[8]。而基于无监督学习的方法不需要标记样本，分类速度快，可以识别未知应用，但其分类精度较低，训练难度较大。因此，构建自适应复杂多变的网络环境中的分类模型是一个巨大的挑战。

本文借鉴集成学习和增量学习思想，提出一种自适应的流量分类系统，借助增量学习和集成学习有效的更新分类器，集成学习保留先前流量训练的分类器，而增量学习不断引入新环境流量。本文采用该方法测试发生概念漂移的网络流数据，实验结果表明该方法可以有效的应对网络流量分类中的概念漂移问题，且与常用的流量识别机器学习算法相比具有较高的识别精度和较低的分类错误率。

本文内容的组织结构如下：第 2 节综述网络流量分类中概念漂移问题的相关研究。第 3 节描述了集成学习流量分类方法。第 4 节给出实验数据集、简要说明实验环境，并分析增量集成学习流量分类算法的性能；第 5 节总结全文并展望未来的工作。

1 相关研究

自“概念漂移”(concept drift)在 1986 年由 Schlimmer 和 Granger[9]首次提出后，国内外研究人员对流量分类过程中产生的概念漂移问题进行了不少的研究，并取得了一定的成果。Zhong 等人[10]在 CVFDT 的基础上提出了解决 P2P 流量概念漂移问题的算法 iCVFDT，该算法在叶节点可能会在概念漂移时产生一棵备选子树，并且在新概念的子树变得更精确时用新子树替代原先的子树，从而解决概念漂移所导致的预测性能下降，但并未解决多种应用情况下的概念漂移。Zhang [11]等提出了一种采用加权对称不确定性和 ROC 曲线下面积度量的混合特征选择算法，不需要改变类别分布就能提高少数类的查全率和查准率、以及分类的字节准确率，有效解决类别不平衡性，但没有解决动态数据流引起的概念漂移。Fahad[12]等提出一种将多种特征选择方法集成的混合式特征选择方法，该方法有利于简化分类模型，减少模型建立和分类时间，但是该方法特征选择过程耗费时间长，且没有考虑到概念漂移问题。Li[13]等考虑到不同时间域和空间域对流量分类效果的影响，采用 FCBF 和对称不确定性度量选择特征子集，由于 FCBF 特征选择方法对于多个数据集很难保持较高的分类性能，没有很好的解决概念漂移问题。由于监督学习分类方法需要重新标记样本代价高，Erman[14]首次将半监督学习分类方法用于网络应用分类，但该方法只用了一种聚类算法，聚类方法本身分类准确率低，并缺乏与其它算法进行比较。Raahemi[15]提出基于 tri-training 的 P2P 流量分类，采用滑动窗口技术周期性评价窗口样本的分类准确率，以此检测概念漂移，该方法周期性性能评价和重新训练模型需要耗费大量时间和资源，无法用于大规模网络及多类识别。Li[16]采用协同训练 semi-SVM 方法进行流量分类，由于采用 CBF 和 IG 提取特征子集，特征子集差异性不强，协同训练分类准确率低于监督学习，导致协同训练无法起作用。

与前述工作相比，本文针对流量分类中存在的概念漂移问题提出一种增量式的加权集成学习流量分类模型，该模型具有以下特点：1)与常见的单分类器算法相比，基于多分类器的识别方法增加了决策力度；2)采用混合式特征选择方法选取稳定的特征子集，以维持较高的分类准确率；3)根据增量学习策略引入新环境流量更新分类器，并剔除集成学习中分类性能下降的分类器，充分发挥各个分类器的优势。4)根据分类精度加权的集成学习综合分类结果，能更好的提高分类准确率，有效应对概念漂移问题。

2 集成流量分类方法

为了解决概念漂移带来的不利影响，本文提出一种集成流量分类(Accuracy-weighted Ensemble learning, AWE)模型。如图 1 所示，该系统主要包括特征选择和基于精度权重的集成学习分类两部分。前者提出混合式特征选择方法选取稳定的特征子集，使得分类器能在很长一段时间维持稳定的分类准确率;后者根据增量学习引入新流量更新分类器，并且剔除分类性能下降的分类器，最后使用精度加权的集成学习综合分类结果，构建出适应新环境的分类模型。

2.1 混合式特征选择

当前网络上新应用不断出现，每个应用都有其独特的流统计特征。流统计特征和分布随时间和环境变化发生概念漂移，使得特征选择方法难以获得

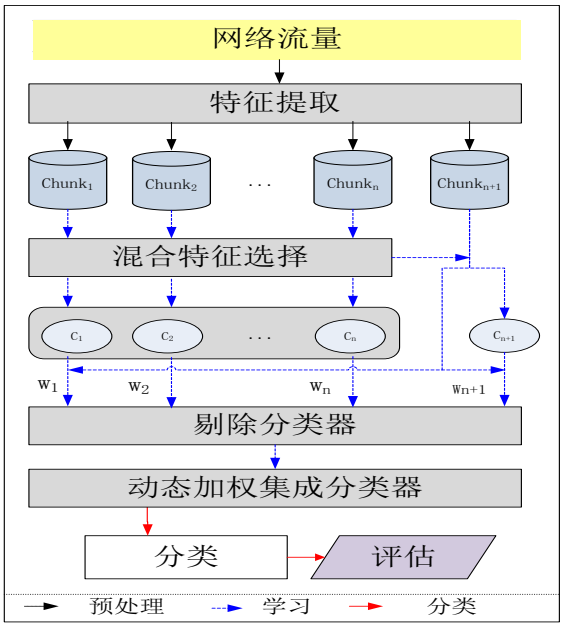

图 1 集成流量分类模型

稳定的特征子集。此外，流特征属性中包含的冗余和不相关特征会增加模型复杂度、降低模型可信度，导致分类效果和效率同时下降[17]。比如，端口号、流持续时间和平均包大小的区分性明显，而平均包数，实际接收到的字节数等特征针对某些应用可能区分性不足。因此，有必要选择稳定的特征子集使得分类模型能在很长一段时间维持稳定的分类准确率。

混合式特征选择方法先对流统计特征进行 FCBF（fast correlation-based filter）特征选择，再采用 Wrapper 方式[18]根据 C4.5 决策树二次搜索最优特征子集，该方法可以获得特征较少且稳定的最优特征子集。FCBF 采用对称不一致性指标 SU[19]，

$$SU(X,Y)=2\left[IG(X\mid Y)/(H(X)+H(Y))\right] \quad (1)$$

其中，$H(X)$ 和 $H(Y)$ 是信息熵，$IG(X\mid Y)$ 是信息增益，$SU\in[0,1]$。混合式特征选择方法选择最优特征子集的过程如下：

(1) 所有特征根据 FCBF 计算获得的重要性排序；

(2) 识别最有价值的特征，选取独立的测试集并采用 C4.5 决策树测试不同数量特征的分类性能；

(3) 在不同数量特征上训练 C4.5 决策树，并评估分类器在测试集的分类性能；

(4) 选择最高分类准确率的特征数目作为最优值，与此同时，最小化特征数目。

2.2 基于精度权重的集成学习方法

将训练集分成大小相同的块 $S_1,S_2,...,S_n$，其中 S_n 是最新的块。从每一个 S_i 中学习一个分类器 C_i，$i\geq 1$。对于给定的测试集 T，给每一个分类器 C_i 一个权重，这个权重与 C_i 分类测试集 T 的期望误差成反比。获取分类器 C_i 的权重通过评估其在测试集的期望预测误差[20]。假设最近的训练数据集 S_n 的类别分布是最接近当前测试集的类别分布，因而分类器的权重可以通过计算分类器在 S_n 的分类误差来近似估计。具体来说，假设 S_n 是由 (x,c) 的数据格式构成，其中 c 是当前记录的真实标记，C_i 对实例 (x,c) 的分类错误率为 $1-f_c^i(x)$，$f_c^i(x)$ 是分类器 C_i 判断 x 所属类别标记为 c 的概率。因而，分类器 C_i 的均方差为：

$$MSE_i=\frac{1}{|S_n|}\sum_{(x,c)\in S_n}(1-f_c^i(x))^2 \quad (2)$$

分类器 C_i 的权重应该反比于 MSE_i。另一方面，分类器预测的随机性（x 被分类为类别 c 等于 c 的类别分布 $p(c)$）将会产生均方差：

$$MSE_r=\sum_c p(c)(1-p(c))^2 \quad (3)$$

由于随机模型并不包含数据的有用信息，我们使用 MSE_r，随机分类器的错误率作为加权分类器的阈值。换句话说，如果分类器的错误率等于或者大于 MSE_r，则丢弃该分类器。为了使得计算简单，我们使用下面的公式来衡量分类器 C_i 的权重 w_i。

$$w_i=MSE_r-MSE_i \quad (4)$$

算法 1 描述了基于权重的集成学习算法的流程，第 1~2 行从最近的数据集 S 中获取分类器 C'，并计算分类器 C' 的权重 w'，第 3~7 行将 S 作为测试集计算 $C=\{C_1,C_2,...,C_k\}$ 每个初始分类器的权重 $w_i\,(1\leq i\leq K)$，淘汰 $w_i\leq 0$ 的分类器，最终从 $C\cup\{C'\}$ 中返回权重前 K 个分类器。

算法 1 的伪代码为 AWE 的训练过程，分类过程比较明确在此省略，即给定一个测试实例 y，用 K 个分类器分类 y，其中实例 y 的分类结果是将 K 个分类器的输出按照权重取均值作为最终的输出结果。

2.3 算法复杂度分析

假设在大小为 s 的数据集上构建一个分类器的

算法 1　　基于精度权重的集成学习方法

输入：

S: 从标记文件中获取最近的数据块
K: 分类器的数目
C: K 个预先训练的分类器

输出：

C:带有更新权重的 K 个分类器的集合

1　从训练样本集 S 训练分类器 $C^{'}$；
2　计算 $C^{'}$ 错误率，通过公式(3)获取的 $C^{'}$ 权重 $w^{'}$；
3　for　$C_i \in C$　do
4　　compute　MSE_i　based on (2);　/*将 C_i 应用于 S*/
5　　compute　$w_i = MSE_r - MSE_i$ based on (2) and (3);
6　　if　$w_i \le 0$　/*淘汰权重 $w_i \le 0$ 的分类器*/
7　　从 C 中移除 C_i；
8　$C = TopK(C \cup \{C^{'}\})$；　/*获取权重前 K 的分类*/
9　return C;

复杂度为 $f(s)$，为了获取分类器的权重 w，需要每个分类器分类测试集 S，而分类测试集的复杂度与测试集的大小成线性关系。假设整个数据流被分成 n 份，则算法 1 的时间复杂度为 $O(n*f(s/n)+Ks)$，其中 n >> K。另一方面，在数据集 s 上构建单分类器需要 $O(f(s))$。对于大多数的分类算法，$f(\cdot)$ 是超线性的，因而集成方法更高效。

3　实验与分析

3.1　数据集

本文采用的数据集是从 WIDE[21]获取的去掉有效载荷的匿名网络数据，我们根据常用的端口号标记网络流，主要分为 6 类，分别为 HTTP, SSH, SMTP, DNS, SSL, POP3，选取连续 4 年的网络流数据进行测试，每个数据集包含连续一周（每天持续 15 分钟）的网络流，数据集的具体分布如表 1 所示。

本文利用 tcptrace 获取 WIDE 数据集的 10 种统计特征，采用混合特征选择后最后选择了服务器和客户端两个方向的初始窗口发送字节大小及客户端方向最小报文长度三个特征作为特征子集，流统计特征及混合特征选择方法最后选取的特征子集如表 2 所示。

表 2　统计特征及特征子集

特征	描述	特征子集
Push_pkts_serv	TCP 头部设置 push 位的总数------服务器	否
Init_win_bytes_clnt	初始窗口发送字节大小------客户端	是
Init_win_bytes_sev	初始窗口发送字节大小------服务器	是
Avg_seg_size_serv	平均报文长度------服务器	否
IP_bytes_med_clnt	IP 报文的平均总字节------客户端	否
Act_data_pkt_clnt	TCP 负载至少一字节的报文总数------客户端	否
Data_bytes_var_sev	报文总字节方差------服务器	否
Min_seg_size_clnt	最小报文长度------客户端	是
RTT_samples_clnt	RTT 时间内样本总数------客户端	否
Push_pkts_clnt	TCP 头部设置 push 位的总数------客户端	否

3.2　评估策略

为了评价流量分类方法的性能，本文选用 4 种常用的评价指标，包括正确率(Accuracy)、查准率（Precision）、查全率（Recall）和综合评价（F-measure），通过对分类器的评价，能客观的了解几种分类器各项性能的优劣。查准率和查全率体现了识别方法在每个单独协议类别上的识别效果，整体准确率体现了识别方法的总体准确率。一个好的方法不仅要求具有较高的总体准确率，还应该在各个类别上具有较高的查准率和查全率，特别当样本类别分布不均匀时，查全率和查准率可以准确获知每个类别的分类情况。F-Measure 是综合查准率 Precision 和查全率 Recall 给出的一个综合评价指标，当 F-Measure 较高时则比较说明方法比较理想。

假设 N 为流量样本数，m 为应用类型数。n_{ij} 表

表 1　WIDE 流统计信息

数据集	年份	http	ssh	smtp	dns	ssl	pop3	total
WIDE1	2009	45484	10358	2368	1497	1076	90	60873
WIDE2	2010	37385	1148	2095	4340	2102	74	47144
WIDE3	2011	67616	33	805	518	2363	90	71425
WIDE4	2012	35558	395	1144	427	2188	32	39744

示实际类型为 i 的应用被标记为类型 j 的样本数。真正 TP 代表实际类型为 i 的样本中被正确标记的样本数，$TP_i = n_{ii}$。假负 FN 代表实际类型为 i 的样本中被误标识为其他类型的样本数，$FN_i = \sum n_{ij}$。假正 FP 代表实际类型为非 i 的样本中被误标识为类型 i 的样本数，$FP_i = \sum n_{ji}$。根据这些概念，给出衡量分类模型准确率、查准率、查全率和综合评价（F-measure）的形式化描述。

整体准确率：

$$accuracy = \sum_{i=1}^{m}(TP_i) / \sum_{i=1}^{m}(TP_i + FN_i) \quad (5)$$

查准率：

$$precision = TP_i / (TP_i + FP_i) \quad (6)$$

查全率：

$$recall = TP_i / (TP_i + FN_i) \quad (7)$$

综合评价：

$$F\text{-}Measure = \frac{2 \times presion \times recall}{presion + recall} \quad (8)$$

3.3 算法分类效果分析

流量产生环境改变导致的概念漂移，使得分类器很难获得稳定的分类性能，因此，有必要建立自适应分类器能在很长一段时间维持稳定的性能。本文选用 4 种常用的评价指标分别统计算法的分类性能，包括准确率(Accuracy)、查准率（Precision）、查全率（Recall）和综合评价（F-measure）。由于在单分类器中决策树 C4.5 的分类效果较好[22]，所以本文选用 C4.5 作为基分类器，将五种基分类器采用基于权重的集成学习方法集成，并与基于投票的集成学习分类算法 Ensemble 和 3 种常用的分类器算法（Adaboost、RandomForest 和 C4.5）进行对比，Adaboost 采用 C4.5 作为基分类器，Adaboost 和 RandomForest 都是集成学习算法，分类性能如表 3 所示。

从表 3 可见，AWE 算法具有较高且稳定的分类准确率，分类效果明显高于其他分类器。由于 AWE 分类器引入新环境的样本学得自适应分类器，有效应对网络流概念漂移，相对于 Ensemble 集成学习只是基于简单的投票策略，AWE 对分类器根据当前的分类精度进行加权投票，可以更好的自适应当前样本环境。

分类准确率只能综合评价整个数据集的识别精度，一个好的算法不仅要有较高的识别准确率，还应该在每个待识别的应用上都具有较高的查准率和查全率，特别当各个应用的样本分布不均匀时，对每个应用的查准率和查全率特别重要，查准率和查全率体现了识别方法在每个单独应用类别上的识别效果。F-Measure 是综合查准率 Precision 和查全率 Recall 给出的一个综合评价指标，当 F-Measure 较高时则比较说明方法比较理想。各个算法的查准率 Precision、查全率 Recall 如表 4 及其综合评价 F-Measure 如图 2 所示。

从表 4 和图 2 可见，AWE 分类器在单个类别的查准率，查全率以及综合评价均高于其他分类器，特别是在 dns 和 pop3 应用。另外，结合表 1 可知，由于训练样本各类别比例的不均衡性，各个类别的样本数目对分类结果有很大影响，HTTP 的样本数目充足，各个分类器的分类的准确率较高；而 pop3 的样本数目稀少，在各分类器的分类效果相对较差，尤其是基于投票的 Ensemble，pop3 的查全率和查准率及综合评价均为 0。而且，Ensemble 在样本数目较少的 ssl 和 dns 分类效果均最差，该方法不适用于识别样本数目稀少的应用类别。

表 3　分类准确率

数据集	C4.5		Adaboost		RandomForest		Ensemble	AWE
	Initial	Final	Initial	Final	Initial	Final		
WIDE 1	96.03	94.4	96.38	94.98	97.83	96	94.2	98.75
WIDE 2	98.25	88.76	98.46	92.16	98.65	89.62	95.15	98.75
WIDE 3	95.57	92.86	95.93	93.78	97.42	94.67	94.98	97.12
WIDE 4	98.65	94.88	98.65	95.63	98.8	96.62	95.61	98.81

注：Initial:未发生概念漂移时的分类精度　　Final:发生概念漂移后的分类精度

表 4　WIDE1 查准率和查全率

类别	Precision					Recall				
	AWE	Ensemble	C4.5	Ada	RF	AWE	Ensemble	C4.5	Ada	RF
dns	0.92	0.83	0.89	0.89	0.74	1.00	0.40	0.50	0.52	0.45
ssh	0.98	0.97	0.66	0.69	0.91	1.00	1.00	0.99	0.99	0.98
smtp	1.00	1.00	0.98	0.98	0.99	0.95	0.92	0.65	0.67	0.84
http	1.00	0.99	0.99	0.99	0.99	0.99	0.95	0.96	0.97	0.97
pop3	1.00	0.00	0.57	0.57	0.57	1.00	0.00	0.31	0.31	0.17
ssl	0.94	0.63	0.77	0.79	0.80	0.98	0.97	0.96	0.97	0.97

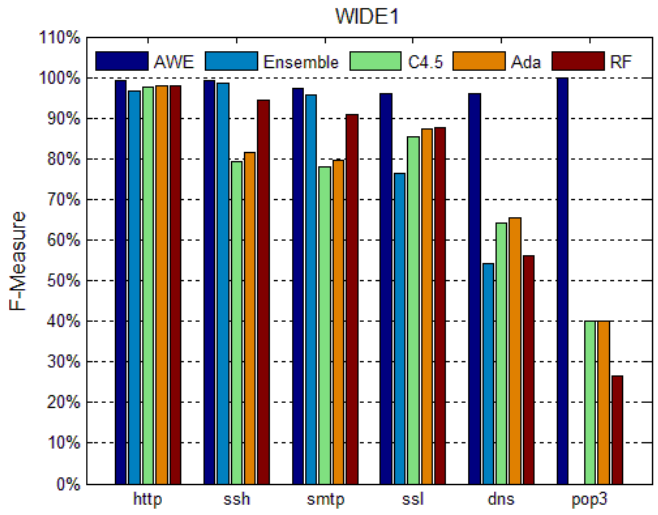

图 2　综合评价 F-measure

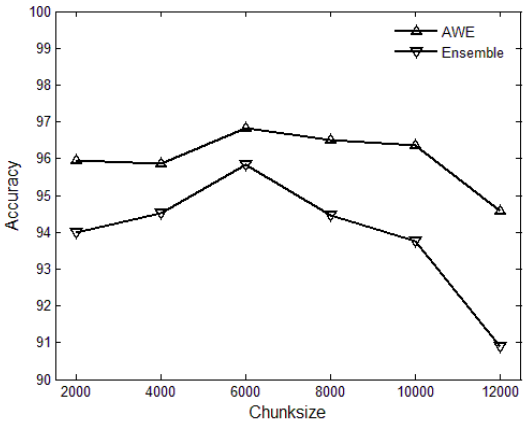

图 3　初始训练集数目对总体准确率的影响

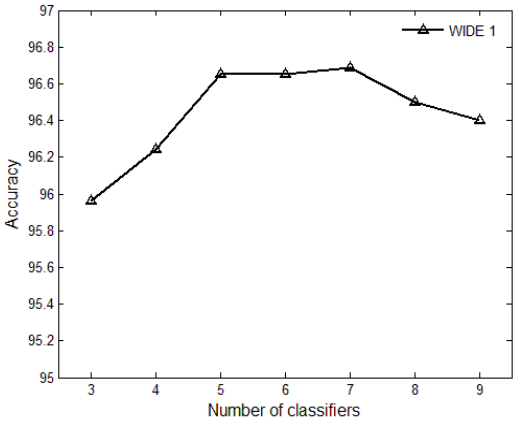

图 4　分类器数目对总体准确率的影响

3.4 算法影响因素分析

本文采用 C4.5 决策树作为基分类器，并且比较 AWE 方法与 Ensemble 集成方法在初始训练数据集大小不同的情况下的分类精度，图 3 描述了初始训练集样本数从 2000 至 12000 时，两种集成学习方法分类效果的差异。结果显示，当初始训练集数目相同时，AWE 方法的分类精度明显高于 Ensemble 集成学习方法，约 2%。另外，图 4 描述了不同集成分类器数目对分类效果的影响，分类器数目变化从 3 至 9。结果显示，初始时增加集成分类器的数目可以提高分类结果的精确度，当初始分类器的数目为 5-7 时，分类效果较好，最好可以达到 96.7%。如果继续增加集成分类器的数目，反而会降低分类准确率。

3.5 时间复杂度

分类系统的及时反馈可以更好的预判网络异常行为，采取及时准确的应对措施。为了检验该算法的分类效率，统计算法的时间开销，包括模型建立时间和分类时间，实验重复 30 次取平均值。表 5

表 5　不同算法的时间开销

方法	模型建立时间(ms)						分类时间(ms)					
	2000	4000	6000	8000	10000	12000	2000	4000	8000	16000	32000	64000
AWE	594	842	975	1149	1379	1649	92	151	240	490	950	1806
Ensemble	442	597	728	860	956	1184	46	70	97	155	273	521
C4.5	191	412	379	487	512	676	14	12	18	22	36	62
Adaboost	1134	2066	2476	3834	4248	5430	20	28	55	73	187	335
RandomForest	474	755	967	1264	1489	1721	22	16	31	43	80	152

描述了 AWE 算法的模型建立和分类时间开销，样本集范围分别介于（2000-12000）和（2000-64000）。

从表中可以看出，时间花费随样本数量增长而增加。AWE 算法的模型建立时间介于 0.59-1.6 秒，明显高于 C4.5，因为 AWE 集成算法需要建立多个 C4.5 模型。另外，AWE 算法在分类时间上明显高于其他算法，因为 AWE 算法集成多个分类器的结果。因此，可以动态调整样本数量达到较快的模型建立和分类速度，也可以采用并行计算来提高效率。

4 结论

随着流量产生环境的改变，分类模型产生概念漂移，使得分类器很难维持较高的分类性能。针对此问题，本文提出了增量式加权集成分类方法。首先，采用混合特征选择方法选取稳定的特征子集。其次，利用增量学习策略引入新环境样本，学得自适应分类器，同时剔除分类性能较差的分类器。最后，根据精度加权集成学习方法综合多个加权分类器的分类结果。与常用的流量识别机器学习算法相比，该算法可以有效应对流量分类中出现的概念漂移问题，实现分类效果和效率的最优平衡。

下一步将主要研究代价敏感学习，通过将本文提出的增量式加权集成分类方法与代价敏感学习方法相结合来进一步解决概念漂移引起的类别不平衡问题。

参考文献

[1] CAIDA: traffic-analysis: classification-overview: Internet Traffic Classification http://www.caida.org/research/traffic-analysis/classification-overview

[2] Dainotti A, Pescape A, Claffy K C. Issues and future directions in traffic classification[J]. Network, IEEE, 2012, 26(1): 35-40.

[3] Grimaudo L, Mellia M, Baralis E, et al. Self-learning classifier for Internet traffic//Proceedings of the IEEE INFOCOM 2013. Turin, Italy, 2013: 3381-3386

[4] Jin Y, Duffield N, Erman J, et al. A modular machine learning system for flow-level traffic classification in large networks. ACM Transactions on Knowledge Discovery from Data(TKDD), 2012, 6(1): 4

[5] Lee S, Kim H, Barman D, et al. Netramark: a network traffic classification benchmark. ACM SIGCOMM Computer Communication Review, 2011, 41(1): 22-30

[6] Senthamilarasu S, Hemalatha M. Ensemble Classifier for Concept Drift Data Stream[M]//Informatics and Communication Technologies for Societal Development. Springer India, 2015: 127-137.

[7] Liu Q, Liu Z, Wang R, et al. Large traffic flows classification method[C]//Communications Workshops (ICC), 2014 IEEE International Conference on. IEEE, 2014: 569-574.

[8] Nguyen T T T, Armitage G. A survey of techniques for internet traffic classification using machine learning[J]. Communications Surveys & Tutorials, IEEE, 2008, 10(4): 56-76.

[9] Schlimmer J C, Granger R H. Beyond Incremental Processing: Tracking Concept Drift[C]//AAAI. 1986: 502-507.

[10] Zhong W, Raahemi B, Liu J. Classifying peer-to-peer applications using imbalanced concept-adapting very fast decision tree on IP data stream[J]. Peer-to-Peer Networking and Applications, 2013, 6(3): 233-246.

[11] Zhang H, Lu G, Qassrawi M T, et al. Feature selection for

optimizing traffic classification. Computer Communications, 2012, 35(12): 1457-1471

[12] Fahad A, Tari Z, Khalil I, et al. Toward an efficient and scalable feature selection approach for internet traffic classification. Computer Networks, 2013, 57(9): 2040-2057

[13] Li W, Canini M, Moore A W, et al. Efficient application identification and the temporal and spatial stability of classification schema. Computer Networks, 2009, 53(6): 790-809

[14] Erman J, Mahanti A, Arlitt M, et al. Semi-supervised network traffic classification[C]//ACM SIGMETRICS Performance Evaluation Review. ACM, 2007, 35(1): 369-370.

[15] Raahemi B, Zhong W, Liu J. Exploiting unlabeled data to improve peer-to-peer traffic classification using incremental tri-training method[J]. Peer-to-peer networking and applications, 2009, 2(2): 87-97.

[16] Li X, Qi F, Xu D, et al. An internet traffic classification method based on semi-supervised support vector machine[C],Communications (ICC), 2011 IEEE International Conference on. IEEE, 2011: 1-5.

[17] Dash M, Liu H. Feature selection for classification[J]. Intelligent data analysis, 1997, 1(3): 131-156.

[18] Kohavi R, John G H. Wrappers for feature subset selection[J]. Artificial intelligence, 1997, 97(1): 273-324.

[19] Yu L, Liu H. Feature selection for high-dimensional data: A fast correlation-based filter solution[C], ICML. 2003, 3: 856-863.

[20] Wang H, Fan W, Yu P S, et al. Mining concept-drifting data streams using ensemble classifiers[C], Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2003: 226-235.

[21] Borgnat P, Dewaele G, Fukuda K, et al. Seven years and one day: Sketching the evolution of internet traffic[C], INFOCOM 2009, IEEE. IEEE, 2009: 711-719.

[22] Ruggieri S. Efficient C4. 5 [classification algorithm][J]. Knowledge and Data Engineering, IEEE Transactions on, 2002, 14(2): 438-444.