

# 基于模糊单类支持向量机的丢包区分算法

麻海圆<sup>1,2</sup>, 孟相如<sup>1</sup>, 李哲<sup>3</sup>, 温祥西<sup>1</sup>, 朱子健<sup>1</sup>

(1. 空军工程大学信息与导航学院, 陕西西安, 710077; 2. 94303 部队, 山东潍坊, 261051; 3. 空军工程大学, 陕西西安, 710051)

**摘要** 针对现有丢包区分算法难以获取先验知识和不具有推广性的问题, 通过对模糊单类支持向量机的改进提出一种新的丢包区分算法。该算法根据无线误码丢包与按序到达包的时延特征分布一致的特点, 由按序到达包的时延特征构成训练集, 从而将区分误码丢包和拥塞丢包的二分类问题转化为判断丢包是否为误码丢包的单分类问题。由于无需采集两类丢包样本, 解决了难以获取先验知识的问题, 使新算法能实现在线的模型训练和丢包区分, 具有很好的推广能力。仿真结果显示, 新算法区分效果良好, 提高了无线网络的传输效率。

**关键词** 拥塞控制; 丢包区分; 支持向量机; 模糊隶属度

**DOI** 10.3969/j.issn.1009-3516.2012.03.015

**中图分类号** TP393.04 **文献标识码** A **文章编号** 1009-3516(2012)03-0070-05

在无线网络中存在 2 种丢包, 一种是由拥塞引起的, 一种由链路误码引起的。传统的拥塞控制机制不对 2 种丢包加以区分, 在发生误码丢包时盲目地启动拥塞控制策略, 严重影响传输效率。丢包区分算法(Loss Differentiation Algorithm, LDA)就是实现对 2 种丢包的区分, 为拥塞控制机制提供决策依据。

早期的丢包区分算法通过对相关网络参数的统计和分析, 依据经验设定阈值实现丢包的区分<sup>[1-5]</sup>。由于拥塞丢包是由队列缓存溢出引起的, 因此这些参数大都与时延或相对单项传输时延(Relative One-trip Time, ROTT)有关。在这些凭经验设计的 LDA 算法中, 需要为不同网络设置不同的区分阈值。文献[6]使用支持向量机(Support Vector Machine, SVM)设计 LDA 算法, 当训练集和测试集是在同一仿真场景下得到的, 因此能取得很好的区分效果。然而实际的网络环境、业务及流量是复杂多变的, 仅通过一个仿真场景获得的模型难以适用于其它网络, 因此该方法不具有推广性。

本文提出一种基于模糊单类支持向量机的丢包区分算法, 将区分误码丢包和拥塞丢包的二分类问题转化为判断丢包是否是误码丢包的单分类问题, 实现在线的模型训练和丢包区分, 具有很好的推广性。

## 1 模糊单类支持向量机

### 1.1 思想来源

拥塞丢包是由队列溢出产生的, 队列长度的变化体现在端系统就是时延的变化; 而误码丢包的发生是随机的, 与队列长度无关, 因此其时延分布特征应与按序到达包相同。图 1 绘制了丢包发生前最后一个按序到达包的 ROTT 和时延偏差样本值, 同时也绘制了按序到达包的样本值。从图中可以看出, 误码丢包和按序到达包的时延特征分布基本一致, 而拥塞丢包后的时延特征与前两者存在较大差别。因此可以使用按序到达包的样本集替代误码丢包样本集, 通过训练该样本集得到一个分类器。当有丢包发生时使用该分类器对丢包进行判别, 如果丢包的特征符合样本集特征, 则判定为误码丢包, 否则判定为拥塞丢包, 从而将区分误码丢包和拥塞丢包的二分类问题转化为判别丢包是否为误码丢包的单分类问题。单类支持向量机(One-Class

\* 收稿日期: 2011-10-15

基金项目: 陕西省自然科学基金基础研究计划资助项目(SJ08F14; 2009JQ8008)

作者简介: 麻海圆(1982-), 男, 河南南阳人, 博士生, 主要从事网络拥塞控制研究. E-mail: haiyuanma@live.cn

SVM, OCSVM) 是一种先进的解决单分类问题的机器学习方法。

OCSVM 是在异类样本难以获取的情况下, 通过训练目标样本库得到对目标样本的正确描述。对待测样本  $x$  的分类就是检验  $x$  是否服从该分布, 如果服从, 则接受  $x$  为目标样本, 否则判为异类样本<sup>[6-8]</sup>。其基本思想是用一个核函数将输入样本映射到高维空间  $F$  中, 在  $F$  中构建一个超球面, 通过确定球心  $O \in R^d$  和半径  $R$ , 使得超球面能以尽量小的半径包含尽可能多的样本<sup>[8]</sup>。依据待测样本到球心的距离与  $R$  的大小关系判断该样本是目标样本还是异类样本。

### 1.2 模糊单类支持向量机

传统单类 SVM 在训练过程中对所有的数据点采用相同的惩罚因子  $C$ , 从而分类模型对野点相当敏感。模糊支持向量机 (Fuzzy SVM, FSVM)<sup>[9]</sup> 通过惩罚系数区分样本的不同贡献, 在构造分类模型时对野点赋予较小的权值 (即模糊隶属度), 减小它们对分类模型的影响, 提高了泛化性。经模糊化的训练样本集合为  $X_f = \{(x_1, \mu_1), (x_2, \mu_2), \dots, (x_n, \mu_n)\}$ , 其中,  $\mu_i \in [0, 1]$  为样本点  $x_i$  的模糊隶属度。

模糊隶属度的设置是 FSVM 的关键, 由于现有的方法不适用于基于超球面的单分类支持向量机, 因此本文提出一种新的模糊隶属度设计方法: 先令  $\mu_i = 1 (i = 1, 2, \dots, n)$ , 使用一种智能方法快速求出一个近似超球面, 根据点到近似球心的距离设计模糊隶属度, 距离越远隶属度越小。支持向量机的求解实质上是一个凸二次规划问题, 计算量大。为加快求解速度, 可以利用其局部最优解就是全局最优解的特点, 用进化算法快速求解近似超球面。粒子群优化算法<sup>[10]</sup> (Particle Swarm Optimization, PSO) 计算简单、收敛速度快且不易陷入局部最优。因此本文选用 PSO 算法训练原始目标样本库获得近似超球面, 然后以样本点  $x_i$  到球心的距离  $d_i$  为度量, 用降半高斯函数作为隶属度函数:

$$\mu_i(d_i) = \begin{cases} 1 & , d_i \leq R' \\ \exp\left[-\frac{(d_i - R')^2}{2\varphi^2}\right] & , d_i > R' \end{cases} \quad (3)$$

式中  $R'$  是近似超球面的半径。当样本点距离超球面较近时, 该样本是正常样本的可能性较大, 因此隶属度函数下降较为缓慢。随着样本点远离超球面, 属于正常样本的可能性变小, 而属于野点的可能性变大, 隶属度快速减小并趋近于 0。本文设置参数  $\varphi = 0.2$ , 则当  $d_i > 1.5R'$  后, 隶属度将小于 0.05, 这些样本点对分类面的影响几乎可以忽略不计。

### 1.3 SV 预选集

SVM 的训练时间随样本数的增长呈指数增长, 当样本数很大时, 其计算量是难以承受的。文献<sup>[11]</sup> 证明了支持向量机的训练过程及训练结果仅与支持向量 (SV) 有关。因此可以在求解最优超球面之前对样本集进行筛选得到一个 SV 预选集, 如果该预选集中样本数量较少且包含所有 SV, 则训练 SV 预选集就能获得与训练原样本集完全相同的分类模型, 而训练速度会得到很大提高<sup>[12]</sup>。在基于超球的单支持向量机中, SV 是那些处于超球面上或超球外部的极少数样本点, 因此一定在近似超球面附近。基于此特性, 本文提出以下 SV 预选方法: 在使用 PSO 算法求出近似超球面后, 将  $d_i$  小于  $R' - \zeta$  的样本剔除, 留取  $d_i$  大于  $R' - \zeta$  的样本, 得到的 SV 预选集为  $\{x_i \mid d_i \geq R - \zeta, i = 1, 2, \dots, n\}$ 。

至此, 我们得到了本文改进的模糊单类支持向量机 (FOCSVM) 训练过程:

- Step 1 通过 PSO 训练原始目标样本得到近似分类超球面;
- Step 2 选择到近似球心距离大于  $R' - \zeta$  的样本点作为 SV 预选集;
- Step 3 使用式 (3) 为 SV 预选集内样本点赋予相应的隶属度, 构成最终的训练集;
- Step 4 对 SV 预选集中的样本进行训练得到最终的分类模型。

## 2 基于 FOCSVM 的丢包区分算法

### 2.1 特征选取

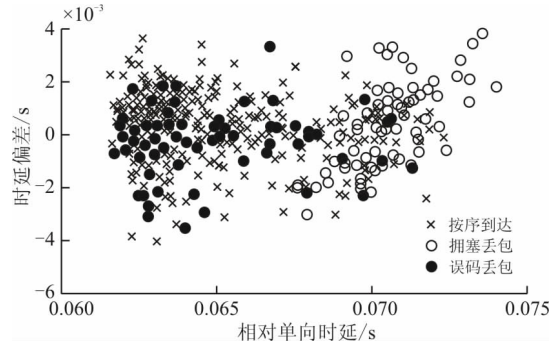


图 1 两类丢包样本的分布图

Fig. 1 Distribution of the two types of packet loss

现有丢包区分算法一般使用 ROTT 和包到达间隔区分丢包的特征。研究表明,基于包到达间隔的算法仅在瓶颈链路上无背景流的网络环境中能够达到较高的区分精度<sup>[3]</sup>,而时延偏差能更好地反映网络的拥塞状态<sup>[13]</sup>。时延偏差为相邻 2 个数据包到达间隔与发送间隔之差。如果相邻 2 个包的发送时刻分别为  $t_{\text{send}}(i-1)$  和  $t_{\text{send}}(i)$ ,接收时间分别为  $t_{\text{recv}}(i-1)$  和  $t_{\text{recv}}(i)$ ,则时延偏差为:

$$D(i) \triangleq [t_{\text{recv}}(i) - t_{\text{send}}(i)] - [t_{\text{recv}}(i-1) - t_{\text{send}}(i-1)] \quad (4)$$

路由器缓存溢出之前有一个队列的累积过程,因此往往在丢包发生前 ROTT 呈现增加趋势。因此,本文选取丢包前 5 个包和丢包后 3 个包的 ROTT 和时延偏差构成 16 维的特征向量。

## 2.2 算法流程

本节将第 1 节的模糊单类支持向量机用于无线网络的丢包区分,提出基于模糊单类支持向量机的丢包区分算法(FOCSVM-LDA)。该算法的流程见图 2。初始化阶段用于统计足量的按序到达包的特征样本。在本文中,当训练集样本数达到 500 个后,使用 FOCSVM 训练分类模型,随后进入丢包区分阶段。在丢包区分阶段,使用分类模型对丢包进行区分。如果满足式(2),则判为误码丢包;否则判为拥塞丢包,执行相应的拥塞控制策略。主成分分析(Principle Component Analysis,PCA)是对样本进行降维降噪的预处理技术。

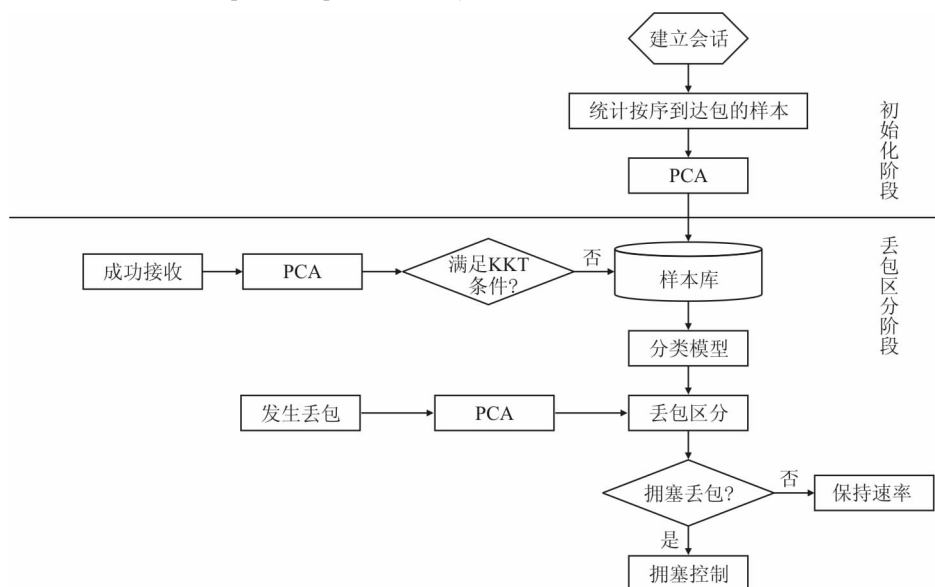


图2 FOCSVM-LDA 算法流程

Fig.2 Flowchart of FOCSVM-LDA

为了适应网络的动态性,分类模型应随着会话的进行不断更新。在丢包区分阶段,每按序到达一个数据包,就判断该包的特征是否满足 KKT 条件,如果满足,则忽略该包,否则就将该样本加入 SV 预选集中重新训练分类模型,实现分类模型的在线更新。由于使用了 SV 预选策略,模型训练和更新的速度都得到很大提高。在本文仿真中,初始化阶段的平均训练时间为 3 ms,模型更新时间不到 1 ms,完全满足实时性的要求。

## 3 性能分析

本节使用 NS2 仿真工具分析所提算法性能,并与 ZBS 和文献[5]的算法 SVM-LDA 进行比较。网络拓扑和仿真参数与文献[3]相同。在 FOCSVM-LDA 中, $C=0.5$ , $\sigma=0.1$ , $\varphi=0.2$ , $\zeta=0.9$ 。在使用 PCA 降维降噪过程中,按照贡献值从大到小的顺序选取主成分,使得所选择取分量和累积贡献值大于 85%,最终得到三维无成分分量。衡量丢包区分算法的度量包括瓶颈链路归一化吞吐率( $T_{\text{nom}}$ )、拥塞丢失率( $L_c$ )、拥塞丢包错判率( $P_{\text{wic}}$ )和误码丢包错判率( $P_{\text{clw}}$ )<sup>[3]</sup>。

仿真结果见图 3。从图中可以看出,当瓶颈链路上仅有一个流时,3 个 LDA 算法都能将绝大多数的误码丢包区分出来,因此都能有效提高瓶颈链路的吞吐率。3 个算法的拥塞丢包错判率都为 0,且对误码丢包的错判率都较低,其中 SVM-LDA 错判率最低,仅为 2.3%。这是因为该算法的分类模型是在完全相同的网络场景中训练得到的,因此能够达到很高的区分精度,从而其吞吐量略高于 ZBS 和 FOCSVM,也具有略高的拥塞丢包率。由于原始 TFRC 使用 TCP 吞吐量公式,对丢包的响应较为强烈,因此瓶颈链路吞吐率仅为

78.8%。

当2个流竞争瓶颈链路时,队列抖动加剧导致  $P_{clw}$  增加,也使吞吐率低于单个流时的吞吐率。3个算法的  $P_{clw}$  都随着流数目的增加而持续变大,但是吞吐率也不断提高。

SVM-LDA 算法的各项指标都是最优的。这是因为训练和测试使用的是相同网络环境,因此具有很好的区分精度。如果将某一个环境下得到的分类模型用于其它网络环境,将难以得到满意的结果。图中的 SVM-2 指的是将流数目为2时训练得到的分类模型用于其它场景时的性能。随着流数目的增加,  $L_e$ 、 $P_{clw}$  和  $P_{wic}$  都不断增大,最终超过 ZBS 和 FOCSVM-LDA。可见,SVM-LDA 算法推广性不强。

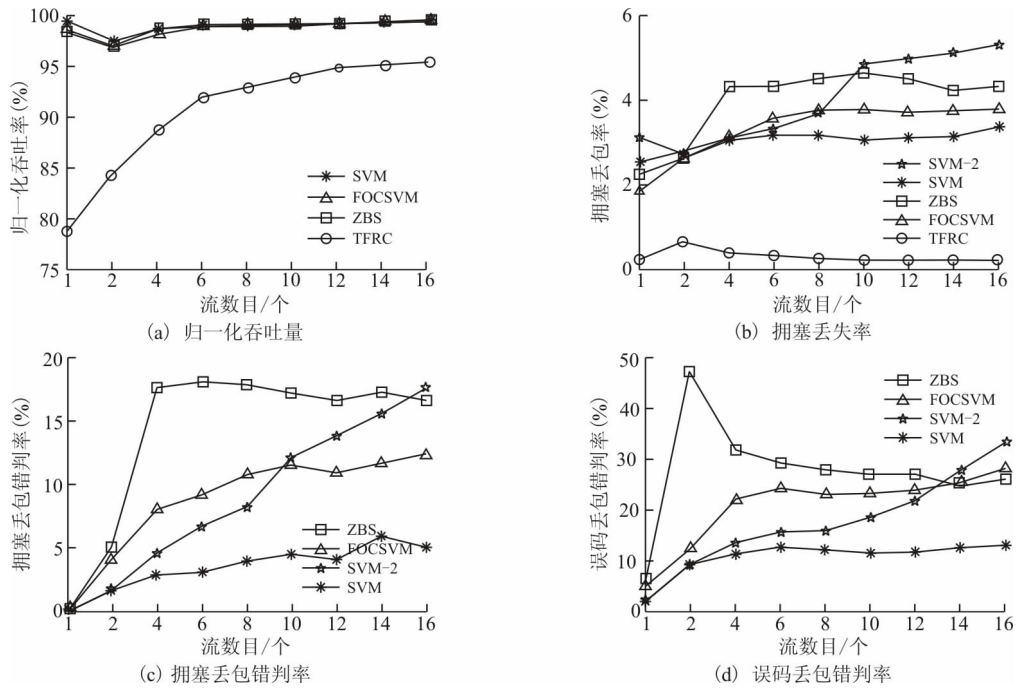


图3 3个LDA算法的性能比较

Fig. 3 Performances comparison of thress LDA algorithm

FOCSVM-LDA 各项指标都优于 ZBS,体现了支持向量机的优势。由于使用的是单分类器,区分精度劣于二分 SVM。但是 FOCSVM-LDA 的所有样本都是在会话过程中直接获取,并且在线更新分类模型,因此可以推广到各种网络环境。

### 4 结束语

本文针对现有丢包区分算法难以获取先验知识和不具有推广性的问题,提出了基于模糊单类支持向量机的丢包区分算法,取得了良好的区分效果。该算法与现有算法的最大区别在于,使用按序到达包的特征替代误码丢包的特征,在会话过程中建立训练集,从而解决了现有算法难以在实际网络中获取先验知识的难题,使丢包区分算法能广泛应用于各种网络环境。改进的模糊单类支持向量机使用 PSO 算法快速求解一个近似分类超球面,依据样本点到球心的距离设置模糊隶属度,减轻了野点对最优分类面的影响,从而保证了较小的拥塞误码错判率;通过预选样本集,减小了训练时间,保证了丢包区分算法训练和更新模型的实时性。

### 参考文献 (References):

- [1] Biaz S,Vaidya N. Discriminating congestion losses from wireless losses using inter-arrival times at the receiver[C]//IEEE symposium on application-specific systems and software engineering and technology. Richardson, TX:IEEE press,1999:10-17.
- [2] Tobe Y,Tamura Y,Molano A, et al. Achieving moderate fairness for UDP flows by path-status classification[C]//25th Annual IEEE conference on local computer networks (LCN 2000). Tampa, FL:IEEE press,2000:252-261.
- [3] Song Cen, Cosman Pamela C, Voelker Geoffrey M. End-to-end differentiation of congestion and wireless losses[J]. ACM/IEEE transaction on networking,2003, 11(5):703-717.

- [4] Papadimitriou Panagiotis, Tsaoussidis Vassilis, Zhang Chi. End-to-end loss differentiation for video streaming with wireless link errors[J]. Telecommunication system, 2010, 43: 295-312.
- [5] Deng Qian hua, Cai An ni. SVM-based loss differentiation algorithm for wired-cum-wireless networks[J]. Journal of China universities of posts and telecommunications, 2009, 16(4): 104-111.
- [6] 廖勇, 王晓丹, 齐俊杰. 一种新的动态 SUM 选择集成算法[J]. 空军工程大学学报: 自然科学版, 2010, 11(5): 26-30. LIAO Yong, WANG Xiaodan, QI Junjie. A new dynamic SUM selected ensemble algorithm[J]. Journal of air force engineering university: nature science edition, 2010, 11(5): 26-30. (in Chinese)
- [7] 李应红, 尉询楷, 刘建勋. 支持向量机的工程应用[M]. 北京: 兵器工业出版社, 2004: 21-22. LI Yinghong, WEI Xunkai, LIU Jianxun. Engineering application of support vector machine[M]. Beijing: Armament industry press, 2004: 21-22. (in Chinese)
- [8] Schölkopf B, Platt J C, Shawe-Taylor J, et al. Estimating the support of a high-dimension distribution[J]. Neural computation, 2001, 7(13): 1443-1471.
- [9] Lin Chun Fu, Wang SH D. Fuzzy support vector machine[J]. IEEE transaction on neural network, 2002, 13(2): 464-471.
- [10] Kennedy J, Eberhart R C. Particle swarm optimization[C]//IEEE international conference on neural networks. Piscataway, NJ: IEEE press, 1995: 1942-1948.
- [11] 焦李成, 周伟达, 张莉, 等. 智能模式识别与分类[M]. 北京: 科学出版社, 2010: 69-70. JIAO Licheng, ZHOU Weida, ZHANG Li, et al. Intelligent object recognition and classification[M]. Beijing: Science press, 2010: 69-70. (in Chinese)
- [12] Li Yuhua. Selecting training points for one-class support vector machines[J]. Pattern recognition letters, 2011, 32: 1517-1522.
- [13] Wang Sheng-Shuen, Hsiao Hsu-Feng. TCP-friendly congestion control for the fair streaming of scalable video[J]. Computer communications, 2010, 33(14): 1578-1588.

(编辑: 徐楠楠)

## Loss Differentiation Algorithm Based on Fuzzy One-class Support Vector Machine

MA Hai-yuan<sup>1,2</sup>, MENG Xiang-ru<sup>1</sup>, LI Zhe<sup>3</sup>, WEN Xiang-xi<sup>1</sup>, ZHU Zi-jian<sup>1</sup>

(1. School of Information and Navigation, Air Force Engineering University, Xi'an 710077, China; 2. Unit 94303, Weifang 261051, Shandong, China; 3. Air Force Engineering University, Xi'an 710051, China)

**Abstract:** The use of the existing Loss Differentiation Algorithms (LDA) cannot get prior knowledge from actual networks, so the algorithms can hardly be generalized to various environments. To solve this problem, a new LDA is proposed by enhancing the Fuzzy One-class Support Vector Machine (FOCSVM). The new LDA converts the differentiation packet loss between congestion and bit error on wireless links into judging that whether the loss is because of bit error. The training set consists of only the in-sequence packets received by the receiver, so the classification model can be got and updated during the session, without the prior knowledge about the two types of loss, which makes the LDA have fine generalization. Simulation results show that the proposed LDA has good differentiation accuracy and can enhance the utilization of resources in wireless networks.

**Key words:** congestion control; loss differentiation; support vector machine; fuzzy membership