DOI:10.16511/j.cnki.qhdxxb.2013.05.019

ISSN 1000-0054 清华大学学报（自然科学版）2013 年 第 53 卷 第 5 期 26/26
CN 11-2223/N J Tsinghua Univ (Sci & Tech), 2013, Vol. 53, No. 5 734-740

基于在线排序逻辑回归的垃圾邮件过滤

孙广路[1]， 齐浩亮[2]

（1. 哈尔滨理工大学 计算机科学与技术学院，哈尔滨 150080；2. 黑龙江工程学院 计算机科学与技术学院，哈尔滨 150050）

摘 要：垃圾邮件过滤是网络信息处理中的重要问题，基于机器学习方法的垃圾邮件过滤技术是目前的研究热点。现有研究一般将过滤问题视为二值分类问题进行解决，存在着模型优化目标和性能评价指标 1-AUC 不一致的问题，导致模型优化结果产生偏差，过滤性能受到很大影响。该文通过直接优化评价指标 1-AUC 来提升过滤器性能，将垃圾邮件过滤问题转化成排序问题进行建模，提出了在线排序逻辑回归学习算法，解决了在线学习中的邮件得分偏移问题；综合应用 TONE 算法和重采样技术，提出参数权重更新算法，解决模型学习中在线调整模型参数时的处理速度问题，满足垃圾邮件实时过滤的要求。在垃圾邮件过滤公开评测数据集上的实验结果表明，基于在线排序逻辑回归模型的过滤结果全面优于在线逻辑回归模型的过滤结果。

关键词：垃圾邮件；判别模型；排序学习；在线排序逻辑回归；1-AUC 指标

中图分类号：TP 391.3 **文献标志码**：A

文章编号：1000-0054(2013)05-0734-07

Spam filtering based on online ranking logistic regression

SUN Guanglu[1], QI Haoliang[2]

(1. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China;
2. College of Computer Science and Technology, Heilongjiang Institute of Technology, Harbin 150050, China)

Abstract: Spam filtering is an important issue in Web information processing. Many machine learning methods are utilized to filter spam. Current researches transform the filtering problem into binary classification, in which the optimization target of the classification model is not consistent with 1-AUC, the usual evaluation measurement. The inconsistence results in the deviation of model optimization, which makes a bad influence on filtering results. In this study, spam filtering was transformed into the ranking model through the optimization oriented to 1-AUC with online ranking logistic regression model then proposed to tackle the deviation of the model's score in the online learning module. TONE (train on or near error), re-sampling and weights update methods were used to promote the learning speed in online adjustment of model's parameters. Experiments on open evaluation datasets show that the developed method is better than the traditional online logistic regression model with statistical significance.

Key words: spam; discriminative model; learning to rank; online ranking logistic regression; 1-AUC measurement

随着互联网技术的发展，电子邮件方便了人们的工作和生活。但是，随之而来的大量垃圾邮件已严重影响了电子邮件的正常使用。为了解决垃圾邮件泛滥问题，人们提出了许多解决方案。其中，基于垃圾邮件的内容进行过滤是一种主要手段。

垃圾邮件的内容过滤目标是基于邮件内容将邮件区分成两种形式：垃圾邮件(spam)或正常邮件(ham)；因此，将其视为二值分类问题是自然的想法。目前，机器学习模型在垃圾邮件过滤问题上取得很好的效果[1]。这些分类算法从模型原理上可以分为两种：生成模型和判别学习模型。朴素 Bayes 模型是生成模型的典型代表，著名的 Bogofilter 系统依据朴素 Bayes 模型构建，其在国际文本检索会议(TREC)的垃圾邮件过滤评测中作为基准(baseline)系统[2-3]。在基于判别学习模型的过滤系统中，典型的模型有支持向量机模型和逻辑回归模型。它们在 TREC 数据中取得了很好的性能[4]。对于垃圾邮件过滤，判别学习模型的分类效果通常要好于生成模型，在国内外垃圾邮件过滤评测中，判别模型都取得了最好的结果[5-6]。

虽然利用二值分类器的判别学习方法解决垃圾邮件过滤问题在国内外的评测中取得不错的成绩，但从问题分析和建模的角度，用分类模型解决垃圾

收稿日期：2012-10-12

基金项目：国家自然科学基金资助项目（60903083）；
黑龙江省新世纪人才项目（1155-ncet-008）；
教育部博士点新教师基金资助项目（20092303120005）

作者简介：孙广路(1979—)，男(汉)，黑龙江，教授。
E-mail: guanglu_sun@163.com

邮件过滤存在不足。在分类模型的训练过程中，分类器的优化目标是寻求一组带权重的参数，或者一个最优分类面，并在此基础上进行一定程度上的泛化，以求最小化邮件分类错误的个数。也就是说，它们的优化目标是降低 spam 被错误划分为 ham 和 ham 被错误划分为 spam 的错误数总和。

然而，1-AUC 指标是垃圾邮件过滤性能的核心评价指标，被 TREC、CEAS、SEWM 评测一致使用。而邮件分类错误的个数与 1-AUC 并不直接相关，导致现有分类模型的优化目标和过滤评价指标的不一致。换言之，将分类错误数总和降至最低并不能保证过滤器的性能也达到最优。由此可见，垃圾邮件过滤的性能尚有提升的空间和更好的解决方法。

到目前为止，尚未见以 1-AUC 为优化目标的垃圾邮件过滤器。在整个机器学习领域，以 1-AUC 为优化目标的研究也较少，在二值分类的相关研究中，据本文作者所知，只有文[7-9]进行了一定程度的研究。文[7]直接根据 1-AUC 的定义对该指标进行优化。文[8-9]指出 Wilcoxon's Rank Sum Statistic 与 1-AUC 相关。由于直接计算 1-AUC 计算量大，文[9]采用近似算法进行计算，但模型优化存在偏差。文[8]改进 SVM 模型使其适合于排序方法，并直接通过降低错误的样本序对来达到优化 1-AUC 的目的，但由于 SVM 模型的复杂度较高、计算量偏大，在实际过滤系统中无法应用。因此，这些相关的研究和方法都不能直接应用到垃圾邮件过滤的解决中[10]。

本文针对现有垃圾邮件过滤模型中存在的优化目标和垃圾邮件过滤问题评价指标不一致、模型优化结果产生偏差、性能受到制约的问题，并基于对核心评价指标 1-AUC 直接优化来提升过滤器性能的思想，提出基于排序学习的垃圾邮件过滤建模新方法，以取代传统分类模型的方法，并提出在线排序逻辑回归算法解决在线排序时邮件得分偏移问题。同时，与支持向量机模型相比，它大大减少模型计算代价。此外，运用了 TONE(train on or near error)算法和重采样技术，以解决排序模型计算量过大的问题，满足垃圾邮件实时过滤的要求。

本文结构如下。第 1 章主要介绍基于 1-AUC 指标的排序学习策略、邮件过滤模型及学习算法；第 2 章详细介绍提升基于排序的过滤模型的两种方法，并给出了模型参数权重更新算法；第 3 章是论文的实验及讨论部分；最后是论文的结论。

1 基于在线排序学习的垃圾邮件过滤模型

为了使垃圾邮件不被过滤，垃圾邮件发送者不停改变垃圾邮件的发送方法、技术，导致垃圾邮件随着时间的推移不断变化。这种变化在在线学习中被称之为概念迁移(concept drift)[11]。这就要求过滤器在使用过程中能够及时调整，适应不断变化的垃圾邮件。具体而言，过滤器根据用户的反馈(即过滤器是否正确分类)情况对当前的模型进行更新，调整模型的参数，以适应垃圾邮件过滤的需要。

本文的过滤模式采用在线学习方式(如图 1 所示)，以过滤器和训练器为中心分为过滤和训练(即学习)两部分。过滤器根据特征库，对输入的邮件流进行过滤，即判断每个邮件的属性。训练器根据用户的反馈对邮件的过滤结果进行学习，并进一步调整特征权重库中相应特征及其权重，提高过滤器的适应能力与性能。

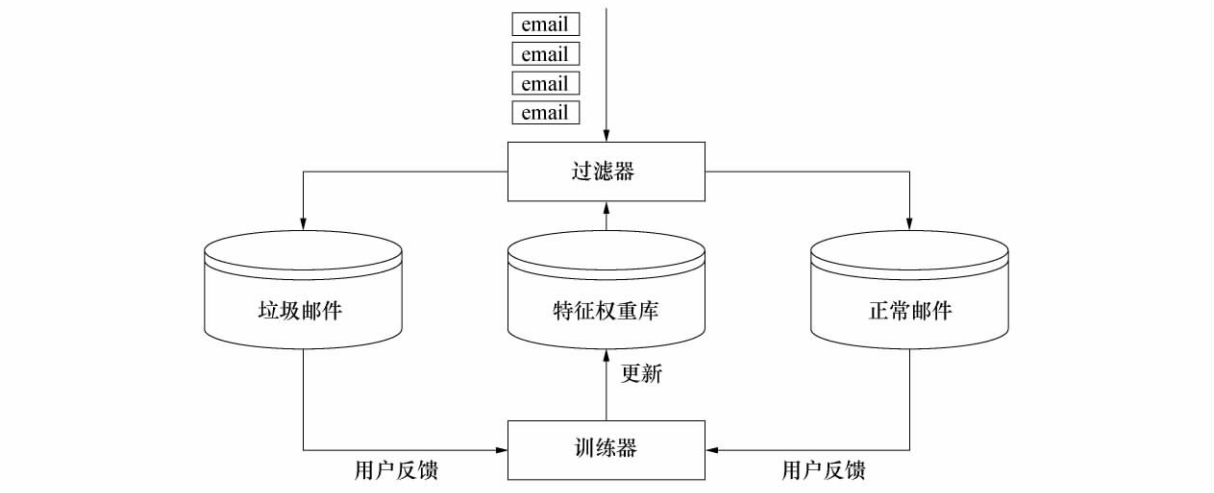

图 1 垃圾邮件过滤在线模式

1.1 基于1-AUC指标的排序学习策略

垃圾邮件过滤器会为每封邮件计算分值,来预测该封邮件为垃圾邮件的可能性。已知每封邮件的分值后,可以把sm%计算为hm%的函数(sm%表示spam的误判率,hm%为ham误判率),这一函数的图形表示就是一个ROC曲线(查全率-错检率曲线)。ROC曲线下方的区域是过滤器在所有可能值上有效性的累积度量。由于hm%和sm%衡量过滤器的失效性而不是有效性,为保持一致,垃圾邮件评测时,以ROC曲线上方的区域作为垃圾邮件过滤器的评价指标,即(1-AUC)%。

ROC区域也有概率学解释:即任一封ham比任一封spam获得更低分值的概率。由此可见,1-AUC值越小,表示过滤器效能越好。因此,构建垃圾邮件过滤模型时应以1-AUC为优化目标。

但是,直接计算1-AUC比较复杂,文[12]分析一个有限样本集的ROC曲线是一个阶梯函数,并给出了1-AUC的计算方法如式(1)所示,该方法被TREC、CEAS和SEWM的评测工具一致采用,也在文[8-9]中被采用。

$$1-\mathrm{AUC}=\frac{\mathrm{SwappedPairs}}{mn}. \tag{1}$$

其中,m和n分别为标准标注的正例spam和反例ham的个数。在任意的ham和spam组成的〈ham,spam〉序对集合中,spam的得分低于ham、排序出现错误的序对被定义为不一致序对(swapped pairs)。

为了最小化不一致序对,必须尽量排除任何错误的序对,以代替仅仅把精力集中在错误分类数上。因此,1-AUC指标的优化问题更适合利用排序模型来解决,而不是分类模型。在这种考虑下,面向评价指标1-AUC来提升垃圾邮件过滤性能,可以转化为邮件序对的排序问题来解决。

排序学习是信息检索的一个核心问题,其主要方法是通过特征构造有效的排序函数。目前已有的排序算法有3种:Pointwise、Pairwise和Listwise方法。Pointwise方法采用一个查询对应的一个文档作为训练样本,其主要思想是将排序问题转化为多分类问题或回归问题,如Pranking with ranking算法[13];Pairwise方法是采用查询对应的文档对作为训练样本,其主要思想是将排序问题形式化为二元分类问题,如Ranking SVM算法[14];Listwise方法是采用查询对应的文档序列作为训练样本,直接对样本的排序结果进行优化,如ListNet算法[15]。由于邮件可被分为ham和spam两类,Pairwise方法更符合邮件过滤问题的定义形式,本文的排序策略将ham和spam作为邮件对进行训练。

1.2 基于在线排序策略的垃圾邮件过滤模型

下面利用基于排序的模型框架将垃圾邮件过滤模型形式化。

令x_i表示正例,x_j表示反例,$\overline{X}_{i,j}=(x_i,x_j)$表示一致的序对,其目标值为$y'_{ij}=1$;$\overline{X}_{j,i}=(x_j,x_i)$表示不一致的序对,其目标值为$y'_{ij}=-1$。排序模型目标是在假设空间$H$中找到一个$h\in H$,使其满足最小化不一致序对$\overline{X}_{j,i}$:

$$h'_w(\overline{X})=\mathrm{argmax}\left\{\sum_i\sum_j y'_{ij}\cdot\Psi(w,x_i,x_j)\right\}. \tag{2}$$

式(2)可以进一步变换为

$$h'_w(\overline{X})=\mathrm{argmax}\left\{\sum_i\sum_j y'_{ij}\cdot\Psi(w,x_i-x_j)\right\}. \tag{3}$$

根据式(3)得到最优的参数w后,对于一个类别未知的邮件x,$\Psi'(w,x)$就是模型对它预测的分值。

令$\Psi(w,x_i,x_j)=\mathrm{sgn}[\Psi'(w,x_i)-\Psi'(w,x_j)]$。其中$\mathrm{sgn}(x)$为符号函数,当$x\geq 0$时,$\mathrm{sgn}(x)=1$;否则,$\mathrm{sgn}(x)=-1$。式(3)可以改写成

$$h'_w(\overline{X})=\mathrm{argmax}\left\{\sum_i\sum_j y'_{ij}\cdot\mathrm{sgn}[\Psi'(w,x_i)-\Psi'(w,x_j)]\right\}. \tag{4}$$

由此,建立了应用排序回归学习解决排序问题的框架。在建立基于排序策略的垃圾邮件过滤模型框架后,将深入研究如何利用机器学习方法解决排序模型的学习算法问题。

1.3 在线排序逻辑回归学习算法

在已有的垃圾邮件过滤研究中,逻辑回归(logistic regression,LR)模型是性能表现最好的模型之一[16]。与同样性能极佳的SVM模型相比,LR模型的时间和空间复杂度明显低于SVM模型。鉴于以上因素,改进逻辑回归模型,将$\Psi(w,x_i,x_j)$定义为$\Psi'(w,x_i)-\Psi'(w,x_j)$,使得传统的逻辑回归模型能够解决垃圾邮件过滤的排序问题。

$$\Psi(w,x_i,x_j)=\frac{\exp(w\cdot x_i)}{1+\exp(w\cdot x_i)}-\frac{\exp(w\cdot x_j)}{1+\exp(w\cdot x_j)}. \tag{5}$$

将式(5)的方法称为在线排序逻辑回归学习算法,此算法利用两封邮件的差值为优化目标。

令 $f(w,x)=\frac{\exp(w\cdot x)}{1+\exp(w\cdot x)}$，则有

$$\frac{\partial\Psi}{\partial w}=\frac{\partial f(w,x_i)}{\partial w}-\frac{\partial f(w,x_j)}{\partial w}=$$
$$f(w,x_i)\cdot(1-f(w,x_i))\cdot x_i-$$
$$f(w,x_j)\cdot(1-f(w,x_j))\cdot x_j. \quad (6)$$

参数向量权重 w 的更新可以根据式(6)和梯度下降方法解决：

$$\Delta w=(y'_{ij}-\text{difference})\text{TRAIN\_RATE}\frac{\partial\Psi}{\partial w}. \quad (7)$$

从式(7)可以看出，以两类目标值均衡的方式进行特征权重调整，从理论上保证了两个类目标值的对称性，解决了邮件得分偏移问题。

2 排序逻辑回归模型的提升方法

虽然上述排序逻辑回归学习算法相对于其他排序模型(如 Ranking SVM 模型)在计算的时间和空间复杂度上已有很大程度上的降低，但从过滤模型的排序框架本身来看，其不足之处在于任意两个训练样本组成的序对数量庞大，直接计算所有序对导致计算量过大，计算效率很低。在 TREC 的评测报告中显示，垃圾邮件过滤器不必对所有的样本对进行训练就能达到最佳的性能[17]。在这个结论的基础上，考虑如何提升排序逻辑回归算法的训练速度。本文提出了减小模型问题的大小和减少训练邮件的序对数来降低模型训练的计算量。下面将详细介绍这两种方法。

2.1 减少模型问题的大小

对于一个包含 n 封邮件的邮件集合，其中 n_1 封 ham，n_2 封 spam，则组成邮件对的个数为 n_1n_2。排序逻辑回归算法要对所有邮件对进行计算，其理论计算复杂度是 $O(n^2)$。垃圾邮件过滤问题中包含了庞大的邮件数以及在线处理的需求，模型难以支持如此巨大的计算量。考虑到最新出现的垃圾邮件反映出垃圾邮件发送技术的特点，更适用于模型的训练，本文采用贪心算法，利用时间序列对过滤样本进行采样，只将最新的 m 封 spam 和 m 封 ham 作为选择训练样本，即组成邮件对的数目为 m^2，以避免邮件集合中所有邮件参与计算。虽然降低采样邮件规模后，模型的时间复杂度仍旧是 $O(m^2)$，但是当 $m\ll n$ 时，计算量明显下降。随着 m 值逐渐增大至一定值之后，过滤器的性能几乎没有变化。关于 m 的取值将在实验中进行讨论。

2.2 减少训练邮件的序对数

在模型训练中，每出现一封邮件时，对这封邮件组成的所有邮件序对都要进行训练。这种方法在测试时已经取得了很好的效果，但是产生了两个问题。第一，内容相近的邮件对可能被多次训练，增加了资源消耗；第二，邮件对中会出现过度训练，当邮件对中某些特征已经被训练多次，过多地进行训练会导致准确率的下降。

本文提出 TONE 方法选择部分邮件对进行训练。TONE 方法是在 TOE(train on error)方法的基础上改进得来的。除了 TOE 中被过滤器错误判别的样本需要参与训练外，过滤器正确判别但邮件对分值小于设定阈值的样本，即判别比较模糊和容易出错的样本也要参与训练。这就克服了 TOE 方法参与训练的样本数过少带来的问题。例如：当过滤值为 0.5 时，TONE 算法设定阈值为 0.1，则分值为 0.4～0.6 之间的邮件对都被作为训练样本。

将 TONE 算法引入排序逻辑回归算法中，仅使用错误或在正确边缘的邮件序对进行训练，使模型的训练速度进一步提高。在 TONE 策略下，对于某一封邮件，首先根据分类器的输出和 TONE 阈值，决定是否对该邮件进行训练。其次，在对该邮件进行训练时，需要高效的算法找出需要训练的序对。

2.3 模型的训练算法

在降低采样邮件数的同时，根据式(6)和式(7)对在线排序逻辑回归算法进行权值更新，设计了新的参数权重更新算法。其中：η 代表算法学习速率，gap 代表 TONE 的设定阈值。

算法 1 基于 TONE 的参数权重更新算法

```
Initialize: w=0
Parameters: η, TONE
For each spam-ham pair x_i and x_j:
{
    Calculate gap=Ψ(w, x_i, x_j)
    if gap<TONE then
    {
    Δw=(1−gap)η ∂Ψ/∂w
    w+=Δw
    }
}
```

算法具有以下两个优点：一是训练次数的降低带来了模型参数更新速度的提升；二是只选择错误的或者不确定的样本进行训练，有效地增强模型的适应性。

3 实验及讨论

3.1 实验设置

垃圾邮件过滤的测试数据来自 TREC、CEAS 和 SEWM 提供的公开评测数据集。

TREC(Text Retrieval Conference):国际文本信息检索评测会议,于 2005—2007 年进行垃圾邮件过滤评测子任务,并在 2006 年进行了中文垃圾邮件过滤评测。

CEAS(Conference on Email and Anti-Spam):电子邮件及反垃圾邮件技术大会,于 2007—2008 年进行专门针对垃圾邮件过滤问题的评测。

SEWM(Search Engine and Web Mining):中国搜索引擎和网上信息挖掘学术研讨会,于 2007—2012 年举办了 5 次中文垃圾邮件过滤评测。

各数据集具体情况如表 1 所示。依照上述评测,本文使用(1-AUC)%作为过滤器的评估指标,逻辑平均误判率(lam%)也被用于参考。

表 1 测试数据集

数据集名称	语言	正常邮件数	垃圾邮件数	邮件总数
TREC05p	英文	39 399	52 790	92 189
TREC06p	英文	12 910	24 912	37 822
TREC07p	英文	25 220	50 199	75 419
CEAS08	英文	167 989	41 285	209 274
TREC06c	中文	21 766	42 854	64 620
SEWM07	中文	15 000	45 000	60 000
SEWM08	中文	20 000	50 000	70 000
SEWM10	中文	15 000	60 000	75 000
SEWM11	中文	15 000	45 000	60 000

本文实验在 Ubuntu 10.10 环境下,使用 C 语言编写。在特征模型中,采用字节级 n-grams(n=4)特征提取方法从每封邮件的前 3 000 个字符中提取邮件的特征,并建立特征空间向量。

3.2 实验结果及讨论

使用在线逻辑回归模型过滤器作为基准过滤器(baseline LR),基准过滤器的参数为:TONE=0.45,TRAIN_RATE=0.003,实验结果见表 2。

表 2 排序 LR 模型和基本 LR 模型的实验结果

测试集	baseline LR		排序 LR	
	Lam%	(1-AUC)%	Lam%	(1-AUC)%
TREC05p-1	0.42	0.012 4	0.38	0.012 0
TREC06p	0.55	0.029 5	0.51	0.028 2
TREC07p	0.14	0.005 8	0.12	0.005 7
CEAS08	0.11	0.002 1	0.08	0.001 5
TREC06c	0.07	0.001 0	0.05	0.000 6
SEWM07	0.00	0.000 0	0.00	0.000 0
SEWM08	0.01	0.000 0	0.02	0.000 0
SEWM10	0.00	0.000 0	0.00	0.000 0
SEWM11	0.01	0.000 0	0.01	0.000 0

在在线排序逻辑回归模型中,采用了贪心算法,利用时间序列对邮件进行采样,仅选择当前邮件及其前 m 邮件形成的邮件序对作为选择训练样本,并非选择所有样本。随着选择样本数 m 值的变化,过滤器的性能变化情况如图 2 所示。从图 2 可以看出,m 值在 2 到 10 之间逐渐增大时,过滤器性能急剧变好。m 值大于 10 时,过滤器的性能变化缓慢;m 值达到 100 时,并在此之后,过滤器性能几乎没有变化。然后从图 3 可以看出,随着 m 的值逐渐增大,过滤器消耗的时间逐渐增大。因此,随着 m 的增大,过滤性能趋近于稳定状态,但是过滤器消耗的时间在一直增大,对于过滤器系统来说,并非 m 的值越大越好。在本文的实验中,使用的参数 m 为 100。

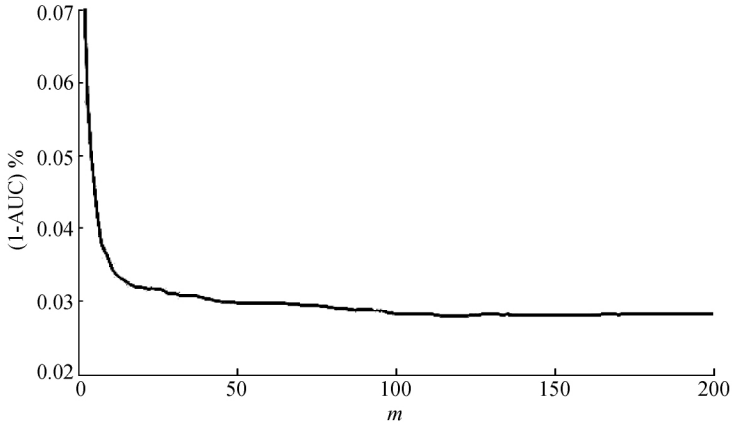

图 2 随着 m 值的变化过滤器性能(1-AUC)%变化情况

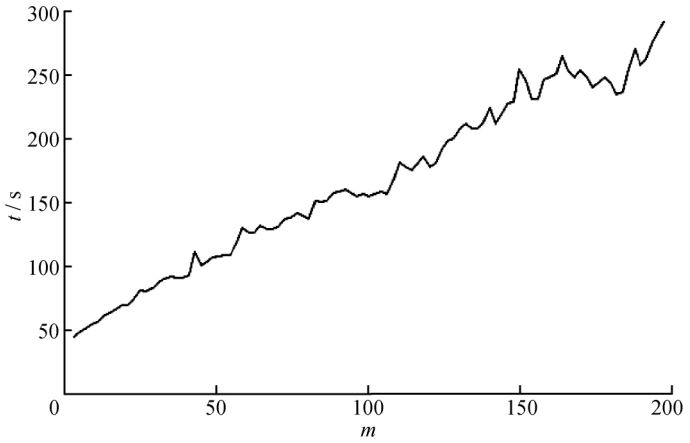

图 3 随着 m 值的变化过滤器消耗时间变化情况

在线排序逻辑回归模型在 3 类数据集上的测试结果如表 2 所示。其中，baseline LR 模型的参数 TONE 和学习速率 η，基于特征的模型的参数 TONE、学习速率 η 以及 m 值如表 3 所示。

表 3 排序 LR 模型和基本 LR 模型的参数设置

模型	TONE	η	m
baseline LR	0.45	0.003	—
排序 LR	0.99	0.005	100

根据表 2 的结果可以看出，本文提出的在线排序 LR 模型以邮件序对形式训练，符合 1-AUC 的优化目标，从理论上对邮件过滤问题进行了优化，使得性能变得更好。本文模型训练时控制了两封邮件得分均衡机制，解决了排序模型中出现的得分偏移问题。因此基于排序的 LR 模型的性能取得了预期的结果。

4 结 论

垃圾邮件过滤模型性能由评价指标 1-AUC 体现，而基于分类的模型优化目标与 1-AUC 不一致，导致模型优化结果产生偏差，过滤性能受到制约。

本文提出的基于在线排序策略的过滤新模型和在线排序逻辑回归学习算法很好地解决了这一问题。同时，TONE 算法和重采样技术的运用，解决了排序模型存在的计算量过大的问题，满足了邮件实时过滤的要求。用国内外公开评测数据进行过滤实验，结果表明新方法具有很好的过滤性能，优于已有的逻辑回归模型的过滤结果。此外，本文还针对排序算法中的参数选择问题进行了实验和详细的讨论。

参考文献 (References)

[1] Cormack G V，Bratko A. A batch and on-line spam filter evaluation [C]//The 3rd Conference on Email and Anti-Spam (CEAS 2006). Mountain View，2006.

[2] Cormack G V，Lynam T. TREC 2005 spam track overview [C]//The Fourteenth Text REtrieval Conference (TREC 2005) Proceedings. 2005

[3] Cormack G V. TREC 2006 spam track overview [C]//The Fifteenth Text REtrieval Conference (TREC 2006) Proceedings. 2006.

[4] Goodman J,Yih W. Online discriminative spam filter training [C]//Proceedings of the Third Conference on Email and Anti-Spam (CEAS). 2006.

[5] Cormack G V. TREC 2007 spam track overview [C]//The Sixteenth Text REtrieval Conference (TREC 2007) Proceedings. 2007.

[6] Sculley D,Wachman G M. Relaxed online SVMs for spam filtering [C]//The 30th Annual International ACM SIGIR Conference (SIGIR'07). 2007.

[7] Park L,Moon J. A learning method of directly optimizing classifier performance at local operating range [C]// Proceedings of International Conference on Intelligent Computing (ICIC-05). 2005.

[8] Joachims T. A support vector method for multivariate performance measures [C]//Proceedings of the 22nd International Conference on Machine Learning (ICML-05). 2005.

[9] Yan L,Dodier R,Mozer M C,et al. Optimizing classifier performance via an approximation to the Wilcoxon-Mann-Whitney Statistic [C]//Proceedings of the 20th Annual International Conference on Machine Learning. 2003.

[10] Waegeman W,Baets B,Boullart L. ROC analysis in ordinal regression learning [J]. *Pattern Recognition Letters*,2008, **29**(1): 1-9.

[11] Widmer G,Kubat M. Learning in thepresence of concept drift and hidden contexts [J]. *Machine Learning*,1996,**23**: 69-101.

[12] Mann H B,Whitney D R. On a test whether one of two random variables is stochastically larger than the other [J]. *Ann Math Statist*,1947,**18**(1): 50-60.

[13] Crammer K,Singer Y. Pranking with ranking [C]//NIPS 2002. 2002.

[14] Joachims T. Optimizing search engines using click through data [C]//KDD 2002. 2002: 133-142.

[15] Cao Z,Qin T,Lin T,et al. Learning to ranking: From pairwise approach to listwise approach [C]//ICML 2007. 2007,**227**: 129-136.

[16] Sculley D. Onlineactive learning methods for fast label efficient spam filtering [C]//Proceedings of CEAS-07. 2007.

[17] Assis F. OSBF-Lua: A text classification module for Lua: The importance of the training method [C]//Proceedings of the Fifteenth Text REtrieval Conference (TREC-06). 2006.

(**上接第** 733 **页**)

[3] Tobias S A, Fishwick W. Theory of regenerative machine tool chatter [J]. *The Engineer*, 1958, **205**(7): 199-203.

[4] Shi H M, Tobias S A. Theory of finite amplitude machine instability [J]. *International Journal of Machine Tool Design and Research*, 1984, **24**(1), 45-69.

[5] Hanna N H, Tobias S A. A theory of nonlinear regenerative chatter [J]. *Journal of Engineering for Industry*, 1974, **96**(1): 247-255.

[6] Deshpande N, Fofana M S. Nonlinear regenerative chatter in turning [J]. *Robotics and Computer—Integrated Manufacturing*, 2001, **17**(1-2): 107-112.

[7] Balachandran B. Non-linear dynamics of milling process [J]. *Philosophical Transactions of Royal Society London A*, 2001, **359**(1781): 793-819.

[8] **石莉, 陈尔涛, 姜增辉. 正交车铣铝合金薄壁回转体振动信号的试验分析** [J]. **兵工学报**, 2009, **30**(3): 356-360.
SHI Li, CHEN Ertao, JIANG Zenghui. Test analysis on vibration signal of thin aluminium-alloy cylinder machined with orthogonal turn-milling [J]. *ACTA Armamentarii*, 2009, **30**(3): 356-360. (in Chinese)

[9] **朱立达, 王宛山, 李鹤, 等. 正交车铣偏心加工三维颤振稳定性的研究** [J]. **机械工程学报**, 2011, **47**(23): 190-196.
ZHU Lida, WANG Wanshan, LI He, et al. Research on 3D chatter stability of orthogonal and eccentric turn-milling [J]. *Journal of Mechanical Engineering*, 2011, **47**(23): 190-196. (in Chinese)

[10] Tobias S A. **机床振动学** [M]. **天津大学机械制造系, 译. 北京: 机械工业出版社**, 1965.
Tobias S A. Machine-Tool Vibration [M]. Faculty of Mechanical Engineering of Tianjin University, Trans. Beijing: Machinery Industry Press, 1965.

[11] Faassen R. Chatter Prediction and Control for High-Speed Milling: Modeling and Experiments [D]. Eindhoven, Holland: Eindhoven University of Technology, 2007.

[12] Insperger T. Stability Analysis of Periodic Delay-Differential Equations Modeling Machine Tool Chatter [D]. Budapest, Hungary: Budapest University of Technology and Economics, 2002.

[13] Skubachevskii A L, Walther H O. On the floquet multipliers of periodic solutions to non-linear functional differential equations [J]. *Journal of Dynamics and Differential Equations*, 2006, **18**(2): 257-355.

[14] Insperger T, Stépán G. Semi-discretization of delayed dynamical systems [C]// Proceedings of the ASME. 2001 Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Pittsburgh, USA: ASME Press, 2001: 1-6.

CONTENTS