高技术通讯 2012 年 第22卷 第9期:889~896

doi:10.3772/j.issn.1002-0470.2012.09.001

# 基于用户关系挖掘和信誉评价的垃圾邮件识别算法①

王 巍② 苘大鹏 玄世昌 杨 武 邱文真

(哈尔滨工程大学信息安全研究中心 哈尔滨 150001)

**摘 要** 采用基于行为的垃圾邮件识别思想,提出了一种基于用户关系挖掘与信誉评价的垃圾邮件识别算法。首先在分析邮件用户群体历史通信关系的基础上,建立邮件用户关系描述模型。进行邮件信誉评价时,利用用户关系描述模型对潜在用户通信关系进行挖掘,并根据邮件接收用户和邮件指纹建立邮件通信路径集,根据路径集中的用户对该邮件对应指纹的历史评价结果得到其综合信誉值,以识别垃圾邮件。在收件人进行反馈时,实时更新用户和邮件指纹信誉值。实验结果表明,该算法识别垃圾邮件的准确率达到95%以上。在算法设计中考虑了实用性,可用于实际的垃圾邮件识别系统。

**关键词** 垃圾邮件,行为分析,用户关系挖掘,信誉评价

## 0 引 言

基于行为的垃圾邮件分析技术是当前的研究热点,该技术根据邮件的发送行为特征来判断其合法性[1],还可以与黑白名单、可追查性检查、发送频率、内容分析、发送方策略框架(sender policy framework,SPF)验证、信誉机制等策略相结合,更有效地识别垃圾邮件。在诸多基于行为的垃圾邮件分析方法中,基于信誉机制的垃圾邮件分析方法已经有了若干应用案例,如 DCC[2]、Vipul's Razor[3]及 TES[4]等。此类方法通过邮件用户之间的通报与合作来确认可疑邮件,更为灵活和有效。本文在分析和总结相关研究成果的基础上,提出了一种基于用户关系挖掘和信誉评价的垃圾邮件识别算法,即根据用户之间的通信关系建立用户关系模型,在此基础上挖掘特定的用户关系路径,根据路径上用户的邮件历史评价信息对邮件信誉进行评价,根据评价结果识别垃圾邮件。实验表明,该算法有较高的识别率。

## 1 研究背景及相关工作

基于行为的垃圾邮件分析方法可分为基于邮件头的方法和基于发送方信誉的方法。

基于邮件头的方法是通过提取和分析邮件头部的信息特征来识别垃圾邮件。张耀龙等[1]通过提取邮件头中的发件人 IP、域名、收发件人地址长度等信息特征形成规则,建立决策树模型,用以判别垃圾邮件。Leiba 等人提出了一种 SMTP 路径分析方法[5],该方法通过提取邮件头 Received 字段中邮件服务器的 IP 地址,根据收到的正常邮件及垃圾邮件的历史记录建立邮件服务器的信誉值,并根据信誉值识别垃圾邮件。当邮件发送者伪造 Received 字段信息时,该方法误判率较高。张尼等人提出了一种基于地理路径分析的垃圾邮件行为分析方法[6],即利用邮件头中 Received 字段信息构建邮件传输路径,并结合地理拓扑关系来识别垃圾邮件,该方法侧重于解决骨干网络或者边界路由环境下的垃圾邮件识别问题。

基于发送方信誉的方法可分为基于发送方 IP-信誉的方法[2]、基于发送方域名-信誉的方法[7]以及基于邮件指纹信誉的方法[8]。由于恶意用户可以伪造 IP 或域名,或采用动态 IP,因此基于发送方 IP 或域名信誉的方法难以对垃圾邮件溯源。基于邮件指纹信誉的方法则不存在该问题,而且可以通过精确地识别内容相似的邮件识别垃圾邮件的群发行为。Prakash[4]等提出的方法中采用用户对邮件的评价来识别垃圾邮件。该方法根据用户的信誉将用户分为可信和不可信两类,通过可信用户的评价来

① 863 计划(2007AA01Z473)和中央高校基本科研业务费专项资金(HEUCF100601)资助项目。

② 男,1974 年生,博士,副教授;研究方向:网络与信息安全;联系人,E-mail:w_wei@hrbeu.edu.cn

(收稿日期:2011-09-20)

识别垃圾邮件,同时更新用户信誉。该方法的不足是只利用用户信誉值评价邮件,而忽略了邮件指纹的信誉;此外,其没有考虑邮件用户间通信关系,可能导致误判率较高。Zheleva[8] 等人根据用户的评价更新自身和邮件指纹的信誉,但恶意用户可以通过大量虚假的正确评价获得很高的信誉,使得该方法在大量恶意用户出现的情况下难以正确处理垃圾邮件。

综上所述,目前基于行为的垃圾邮件识别方法的研究都没有考虑到邮件接收者之间的关系对垃圾邮件识别的影响。某些邮件对于某些用户来讲是垃圾邮件,而对另外一些用户来说则可能是正常邮件。在实际的邮件系统中,某一邮件群体内用户之间或多或少地会存在一定的联系,即相互间会发生邮件收发行为,因此针对目前已有方法存在的问题,本文将邮件信誉评价与邮件接收用户的通信关系结合,提出了基于用户关系挖掘与信誉评价的垃圾邮件识别算法,更为准确和有效地识别垃圾邮件。

## 2 相关定义与模型描述

介绍模型之前,首先给出用户集合的定义。

**定义 1** 邮件用户群体 $U$: 对于一个邮件用户群体 $U$, 其中邮件用户相对固定,并且彼此之间存在着邮件发送和接收的关系。此外, $U$ 内会存在恶意用户。

$U$ 内的合法用户之间会以邮件的方式进行信息交流,每个用户可能是邮件发送者或接收者。从长期看来,彼此间的通信行为是双向的;恶意用户的行为是分发垃圾邮件,并且不期待相关回复,与其他用户间的通信行为表现为单向。图 1 描述了某邮件用户群体 $U$ 内邮件通信拓扑结构。节点 $u_A$ 和 $u_B$ 是恶意用户,他们与其它用户间的通信关系多为单向,而合法用户间的通信关系多为双向。用户之间相互通信次数越多,其关系越紧密。下面将根据用户间通信关系的紧密程度建立用户关系模型。

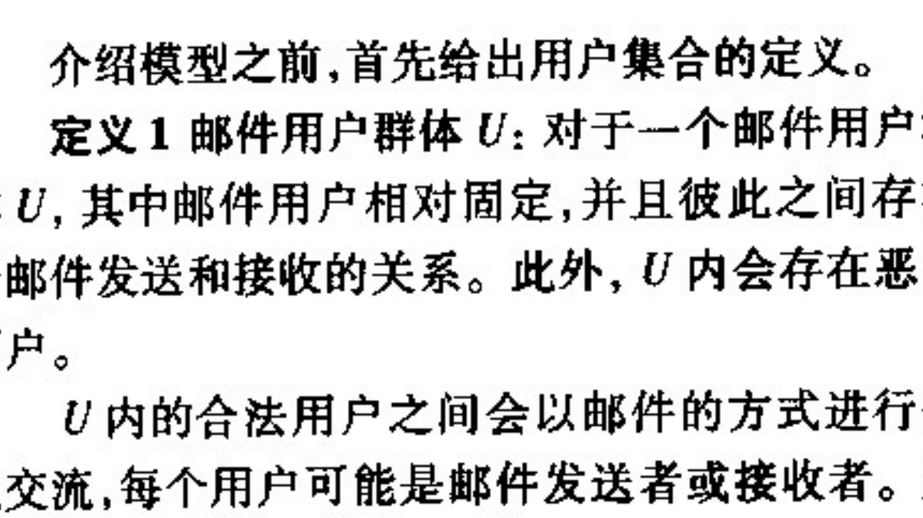

图 1 邮件用户群体 $U$ 的邮件通信拓扑图

**定义 2** 邮件用户通信关系图:邮件用户通信关系图 $G = < U, E >$, $G$ 为有向图,其中设 $U$ 是结点集合,即邮件用户群体集合, $E$ 是 $U$ 中用户边的集合,代表邮件用户间通信关系,则图 $G$ 的邻接矩阵表示如下:

$$A[i,j] = \begin{cases} w_{i,j} & < u_i, u_j >, < u_j, u_i > \in E(G) \\ 0 & \text{否则} \end{cases} \tag{1}$$

其中 $u_i, u_j \in U$, $w_{i,j}$ 表示从 $u_i$ 到 $u_j$ 的双向通信次数, $w_{i,j}$ 越大,表示 $u_i$ 与 $u_j$ 通信关系越紧密。如果 $u_i$ 到 $u_j$ 不存在双向通信, $w_{i,j}$ 的值 0。

**定义 3** 用户通信关系紧密程度图:用户通信关系紧密程度图 $G_1 = < U, E >$, 其中 $U$ 和 $E$ 同定义 2。$G_1$ 为无向图,表示用户之间的紧密程度。则根据图 $G$, 得到 $G_1$ 的邻接矩阵表示如下:

$$B[i,j] = \begin{cases} L - \min(w_{i,j}, w_{j,i}) & < u_i, u_j >, < u_j, u_i > \in E(G) \\ \infty & \text{否则} \end{cases} \tag{2}$$

其中 $L$ 的取值根据 $U$ 的规模动态调整。$u_i$ 与 $u_j$ 相互通信的次数越多, $B[i, j]$ 的值越小,反之其值越大。若 $u_i$ 与 $u_j$ 之间从未发生过通信,则 $B[i, j]$ 的值为 $\infty$。

根据邮件用户关系模型定义,给出可信用户和可信路径的定义。

**定义 4** 可信用户:对任意 $u \in U$, 如果 $trust(u) > \theta_{turst}$, 则称 $u$ 可信用户, $trust(u)$ 为 $u$ 的信誉, $\theta_{turst}$ 为可信用户判别阈值。

**定义 5** 可信路径:对任意 $u_i, u_j \in U$, 如果在 $G_1$ 中存在一条从 $u_i$ 到 $u_j$ 的路径 $P_{i,j}$, 可表示为 $(u_i, u_{k_1}, u_{k_2}, \cdots, u_{k_n}, u_j)$ (其中 $u_{k_1}$ 可以为 $u_i$, $u_{k_n}$ 可以为 $u_j$), 总存在一个正整数 $M$, 满足下式成立:

$$L_p = B(u_i, u_{k_1}) + \sum_{t=1}^{n} B(u_{k_t}, u_{k_{t+1}}) + B(u_n, u_j) < M \tag{3}$$

且 $P_{i,j}$ 中任何一个用户都是可信用户,则称 $P_{i,j}$ 就是一条可信路径, $L_p$ 即为可信路径长度。

## 3 算法思想描述

### 3.1 用户关系挖掘

下面分 3 种情况介绍如何基于邮件用户通信关系图 $G$ 和邮件用户通信紧密程度图 $G_1$ 来挖掘关系密切的邮件用户群体。

(1)当接收邮件用户为可信用户时,利用 Yen 的算法[9]对图 $G$ 和 $G_1$ 实施路径挖掘,得到前 $K$ 短路径,并从中提取可信路径,得到可信路径集 $P$。将 $P$ 中的路径按照长度进行排序,取其前 $N$ 条最短路径,形成路径集 $P_1$。这 $N$ 条路径中所包含的邮件用户即为与收件人关系最为紧密的用户子群体。其中 $N$ 应随着可信路径集中路径数量进行变化,以合理控制子群体规模,保证信誉评价效果。

(2)当接收邮件用户为不可信用户时,在形成路径集 $P$ 后,将不再提取可信路径,直接提取路径集 $P_1$。

(3)接收邮件用户信誉的变更对用户关系挖掘的影响。可信用户会由于过多的恶意评价而导致其信誉值降低,可能会蜕变为不可信用户;反之亦然。因此,在挖掘用户关系时要及时更新邮件用户的信誉状态,保证信誉评价的正确性。

### 3.2　邮件信誉评价计算

下面介绍邮件信誉评价的具体过程。

令 $MailID$ 表示邮件的唯一标识,$record(MailID)$ 为每封邮件的判定记录,对于 $P_1$ 中的任一用户 $u'$,$record(MailID)$ 存储了 $u'$ 对该邮件对应指纹 $s$ 的历史评价结果,记为 $report(u', s)$,用 $count(u', MailID)$ 表示 $u'$ 在路径集 $P_1$ 中出现的次数,则当前用户群体对邮件指纹 $s$ 的信誉评价计算方法如下式所示:

$$score(u,s,MailID) = \frac{\sum_{u' \in P_1} report(u',s) \times count(u',MailID)}{\sum_{u' \in P_1} count(u',MailID)} \tag{4}$$

其中 $report(u', s)$ 的取值为建立 $record(MailID)$ 时,用户 $u'$ 评价该邮件指纹 $s$ 为正常指纹的次数与其评价 $s$ 总次数的比值。由于用户 $u'$ 可能出现在 $P_1$ 中不同的路径上,其在 $record(MailID)$ 出现的次数是不定的,因此在计算信誉评价时要将 $count(u', MailID)$ 考虑进去。

### 3.3　邮件指纹与用户信誉更新

邮件指纹信誉值以及用户信誉值的更新策略如下。

(1)邮件指纹信誉更新

(i) $repu(s)$ 表示邮件指纹 $s$ 的信誉,如果用户 $u$ 评价 $s$ 为正常邮件,则 $repu(s)$ 的更新策略如下:

$$repu(s) = \varepsilon(1 + \alpha)repu(s) \tag{5}$$

其中 $\varepsilon$ 为更新因子,$\alpha$ 是指纹信誉提高因子。

(ii) 如果 $u$ 评价邮件为垃圾邮件,$repu(s)$ 的更新策略如下:

$$repu(s) = \varepsilon(1 - \beta)repu(s) \tag{6}$$

其中,$\beta$ 是指纹信誉衰减因子。

(2)用户信誉更新

对于邮件 $m$,首先从 $record(m)$ 得到用户 $u'$ 对其指纹 $s$ 的历史评价 $report(u', s)$,如果该值小于垃圾邮件判别阈值,则 $u'$ 评价 $m$ 为垃圾邮件;如果该值大于正常邮件阈值,则 $u'$ 评价其为正常邮件。进行信誉更新时,$u'$ 在 $record(m)$ 出现的次数越多,表示 $u'$ 参与评价的次数越多,更新时要用评价次数进行加权。

(i) 如果 $u'$ 与邮件接收用户评价结果相同,则 $u'$ 的信誉为

$$trust(u') = trust(u') + \varepsilon\eta count(u', MailID) \tag{7}$$

其中 $\eta$ 是用户信誉提高因子。

(ii) 如果 $u'$ 与邮件接收用户评价结果相反,则 $u'$ 的信誉为

$$trust(u') = trust(u') - \varepsilon\delta count(u', MailID) \tag{8}$$

其中 $\delta$ 是用户信誉衰减因子。

在式(5)至式(8)中更新因子 $\varepsilon$ 的取值规则如下:如果收件人评价结果与本方法评价结果一致,则根据邮件判定记录,更新因子取值为 1,使得邮件指纹及相应用户信誉值的更新幅度大些。否则,更新因子取值为 0.5,降低更新幅度。

$\alpha$ 和 $\eta$ 取值较小,$\beta$ 和 $\delta$ 取值较大,使得指纹和用户信誉提高速度较慢,信誉降低速度较快,以保证信誉更新策略的健壮性。

### 3.4　邮件指纹提取

有代表性的指纹计算方法包括 Border 等提出的 DSC 算法[10]和 Chowdhurd 提出的 I-Match 算法[11]。与 DSC 算法相比,I-Match 算法有识别准确率高、对输入关键词顺序不敏感等优点,因此本文采用 I-Match 算法计算邮件的指纹。

### 3.5　算法描述

下面将以形式化描述的方式对本文算法进行描述。用户关系挖掘与信誉评价算法描述如下。

**算法:用户关系挖掘与信誉评价算法**

输入:邮件用户群体 $U$、子群体 $U_1$、$U_2$,邮件指纹 $s$,邮件标识 $MailID$,邮件接收用户 $u'$,路径挖掘长度 $K$、$N$

输出:邮件指纹 $s$ 的信誉评价

(1) $U_1 = U_2 = \varnothing$,$P_1 = P_2 = \varnothing$

(2) 在 $U$ 中寻找收到为指纹为 $s$ 的邮件的用户 $u$,形成用户

集合 $U_1$

(3) if 邮件用户 $u'$ 为可信用户

(4) 对于任意 $u \in U_1$

(5) if $u$ 为可信用户

(5) then $U_2 = U_2 \cup \{u\}$

(6) 对于 $U_2$ 中的其它用户 $u'' \neq u$，根据图 $G$ 的邻接矩阵，计算 $u''$ 到 $u$ 的前 $K$ 短路径，如果该路径为可信路径，则将其加入 $P$。转步骤(9)

(7) else

(8) $U_2 = U_1$

(9) 对于 $U_2$ 中的其它用户 $u'' \neq u$，根据图 $G$ 的邻接矩阵，计算 $u''$ 到 $u$ 的前 $K$ 短路径，加入 $P$

(10) 对 $P$ 中的路径进行排序，取前 $N$ 短路径，加入到 $P_1$ 中

(11) 根据邮件 $m$ 的标识 $MailID$，建立判定记录 $record(MailID)$

(12) 对于 $P_1$ 中出现的每个用户 $v$，如果 $v$ 对邮件 $m$ 的指纹 $s$ 进行过评价，则将 $v$ 的评价值 $report(v, s)$ 以及 $v$ 在 $P_1$ 中的出现次数 $count(v, MailID)$ 插入到 $record(MailID)$ 中

(13) 利用 $report(v, s)$ 和 $count(v, MailID$ 计算邮件指纹 $s$ 的信誉评价，具体方法见式(4)

基于用户关系挖掘与信誉评价的垃圾邮件识别算法描述如下。

**算法：基于用户关系挖掘与信誉评价的垃圾邮件识别算法**

输入：待判别的邮件集合 $M$，邮件接收用户 $u$，邮件指纹集合 $S$，垃圾邮件指纹判别阈值 $\theta_{sfin}$，正常邮件指纹判别阈值 $\theta_{hfin}$，垃圾邮件判别阈值 $\theta_{spam}$，正常邮件判别阈值 $\theta_{ham}$。输出：邮件 $m$ 是否为垃圾邮件的判别结果

(1) 将集合 $S$ 初始化为空集

(2) While ($M$ 非空)

(3) 从 $M$ 中取出邮件 $m$，计算邮件 $m$ 指纹，记为 $s$

(4) if $s \notin S$

(5) $S = S \cup \{s\}$

(6) 将 $s$ 置为疑似垃圾邮件

(7) 将 $s$ 的信誉值 $repu(s)$ 置为 0.5

(8) else if ($repu(s) < \theta_{sfin}$)

(9) 将 $s$ 置为垃圾邮件

(10) else if ($repu(s) > \theta_{hfin}$)

(11) 将 $s$ 置为正常邮件

(12) else

(13) 利用图 2 描述的用户关系挖掘与信誉评价算法，计算邮件指纹 $s$ 的信誉值，记为 $Score(< s, MailID >)$

(14) If ($Score(< s, MailID >)) < \theta_{spam}$

(15) 将 $s$ 置为垃圾邮件

(16) If ($Score(< s, MailID >)) > \theta_{ham}$

(17) 将 $s$ 置为正常邮件

(18) else

(19) 将 $s$ 置为疑似垃圾邮件

(20) if $u$ 对 $s$ 进行了评价

(21) 采用 3.3 节介绍的方法对 $s$ 的信誉和已经对评价 $s$ 过的用户信誉进行更新

(22) end of while

步骤(6)至(12)是计算已经接收过的邮件对应指纹信誉值，并与垃圾邮件和正常邮件的信誉判别阈值进行比较，形成判别结果。如果无法进行判别，则执行步骤(13)得到该邮件指纹的信誉值，并进行判别。如果执行步骤(14)至(18)后仍无法确定其是否为垃圾邮件，则将该邮件列入可疑邮件，并提交给邮件用户，对其进行评价和反馈。当用户再接收到该邮件时，可根据其反馈结果对其进行准确判别。

## 4 实验过程与结果分析

### 4.1 实验环境与数据集

实验的邮件样本来自于 The TREC 2006 Chinese Public Corpus(TREC'06C)语料库[12]，其中正常邮件 21766 封、垃圾邮件 42854 封。

实验收集了 50 个实际电子邮箱形成邮件用户群体，0 个账户编号为 0-49。当用户接收邮件并进行评价时，0-47 号用户模拟正常用户行为，48-49 号用户模拟恶意用户行为。

### 4.2 实验过程

本实验过程可分为数据去噪、邮件指纹计算、用户关系模型建立、算法参数确定和算法执行 5 个部分，下面分别进行介绍。

(1) 数据去噪

由于语料库保存的是邮件原始数据，不能直接使用，需要对其进行去噪，具体包括以下几个步骤：

(i) 去除每封邮件的邮件头部分。

(ii) 对于正文格式为超文本的邮件，则先将其解析成纯文本内容。

(2) 邮件指纹计算

采用 I-Match 算法，具体步骤如下

(i) 用 TREC'06C 语料训练得到原始特征词词库。

(ii) 去除反向文档频率(inverse document frequency，IDF)值较小的 25% 原始特征词得到最终词库。

(iii) 根据特征词库，利用 I-Match 算法计算每封邮件的指纹。

(3) 用户关系模型建立

利用0-49号邮件用户之间的历史通信关系建立用户关系模型,形成用户通信关系图以及用户通信关系紧密程度图,并建立邻接矩阵。

(4)算法参数确定

算法中涉及到的相关参数取值如表1所示。

**表1　算法参数取值**

| 相关参数 | | 取值 |
|---|---|---|
| 名称 | 符号 | |
| 垃圾邮件指纹判别阈值 | $\theta_{sfin}$ | -0.5 |
| 正常邮件指纹判别阈值 | $\theta_{hfin}$ | 1.5 |
| 正常邮件判别阈值 | $\theta_{ham}$ | 0.7 |
| 垃圾邮件判别阈值 | $\theta_{spam}$ | 0.3 |
| 可信用户判别阈值 | $\theta_{trust}$ | 0.3 |
| 指纹信誉提高因子 | $\alpha$ | 0.005 |
| 指纹信誉衰减因子 | $\beta$ | 0.01 |
| 用户信誉提高因子 | $\eta$ | 0.0001 |
| 用户信誉衰减因子 | $\delta$ | 0.001 |

用户关系挖掘中前$K$短路径的$K$值对算法影响较大,其值过大会降低路径挖掘过程性能;其值过小会导致用户关系挖掘结果不正确。本文采用具体的实验来确定$K$的取值,以避免人为设定的误差。

取垃圾邮件和正常邮件样本各150封,将该样本集持续重复发送,共发送980封垃圾邮件与1500封正常邮件,并模拟用户对接收的邮件评价,评价统计结果是用户评价邮件的概率约为80%,发生的评价中错误评价约占10%。通过调整$K$的取值,并重复上述过程,统计本文算法的垃圾邮件判别准确率及针对每封邮件的路径挖掘平均用时,得到了如表2所示的结果。

**表2　取不同$K$值时,算法的准确率以及路径挖掘平均用时**

| $K$ | 准确率(%) | 路径挖掘评价用时(ms) |
|---|---|---|
| 1 | 97.879 | 148.873 |
| 2 | 98.96 | 275.049 |
| 3 | 98.943 | 368.645 |
| 4 | 98.921 | 463.343 |
| 5 | 99.117 | 558.446 |
| 6 | 99.123 | 647.849 |

从表2中看出,路径挖掘平均用时随着$K$值的增大而增加,而准确率并没有相应提高,因为$K$值越大,挖掘得到的路径数也越多,从而导致其中包含恶意用户的概率增加,进而影响了评价结果的准确性。可见$K$的取值并不是越大越好,因此在后续实验中,$K$的取值为2,可兼顾算法执行效率和准确性。

(5)算法执行

利用实验过程步骤(1)至(4)确定的数据集、模型和算法参数执行本文的垃圾邮件识别算法。从数据集中随机提取垃圾邮件和正常邮件样本各200封,利用I-Match算法得到87个垃圾邮件指纹和200个正常邮件指纹。循环发送邮件集5次,产生2000条记录。用户对2000条记录进行评价,评价时用户只能对其收到的邮件进行评价。该过程重复8次。

### 4.3　实验结果分析

图2描述了正常用户和恶意用户的信誉变化趋势。在图2中,正常用户信誉值会不断提升,而恶意用户的信誉值会下降,但在达到$\theta_{trust}$后基本保持稳定。这是由于初始化时邮件用户群体中的用户都被认为是可信用户,用户间的通信关系是紧密的,恶意用户做出的评价基本上都是错误的。

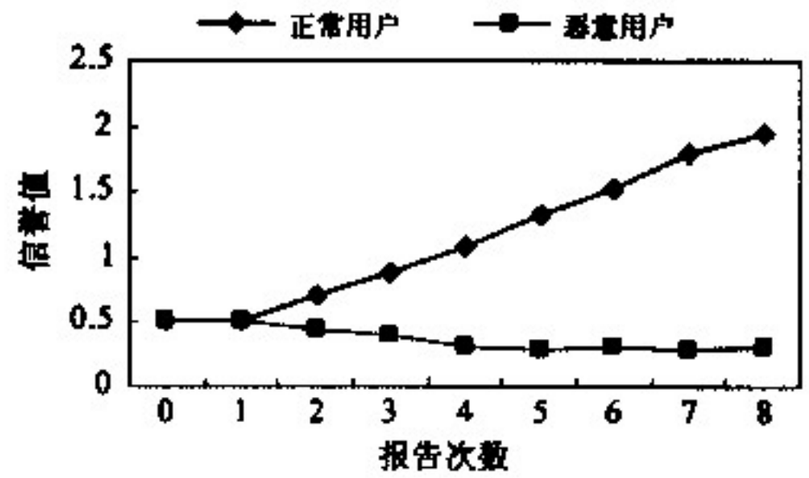


**图2　正常与恶意用户信誉趋势**

该实验说明,在考虑邮件用户关系的基础上,对用户信誉进行评价的策略是正确的。

图3描述了正常邮件指纹和垃圾邮件指纹的信誉变化趋势。正常邮件指纹的信誉值不断提升,而垃圾邮件指纹的信誉值不断下降,且下降的速度要比正常指纹提升的速度要快,这和理论分析的结果是相符的。当指纹阈值低于$\theta_{sfin}$时,该指纹所对应的邮件被判为垃圾邮件;当指纹阈值高于$\theta_{hfin}$时,该指纹所对应的邮件被判为正常邮件。

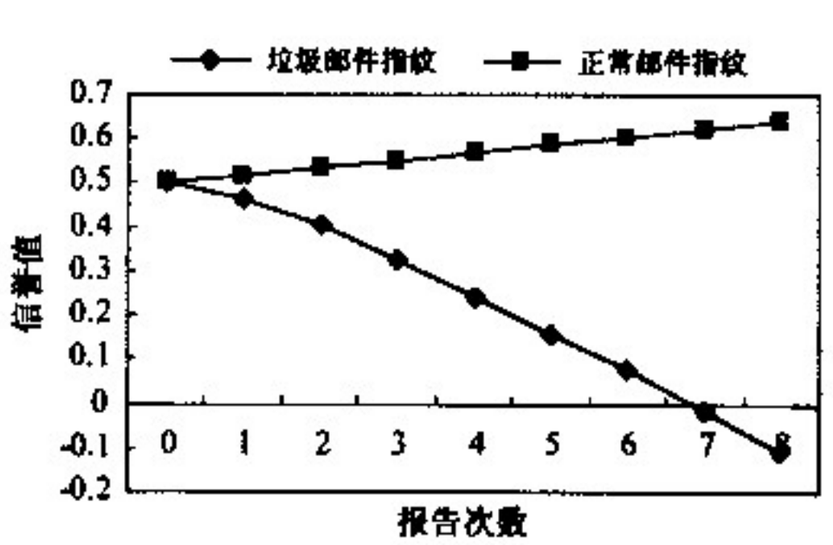


**图3　邮件指纹信誉走势**

图 4 给出了所有评价结束后两类用户对正常邮件指纹和垃圾指纹信誉评价。从图中可以清晰地辨别出可信用户和恶意用户的判断行为，其中 0 – 47 为正常用户，48 – 49 为恶意用户。正常用户的评价与事实相符合，恶意用户的评价与事实相反。

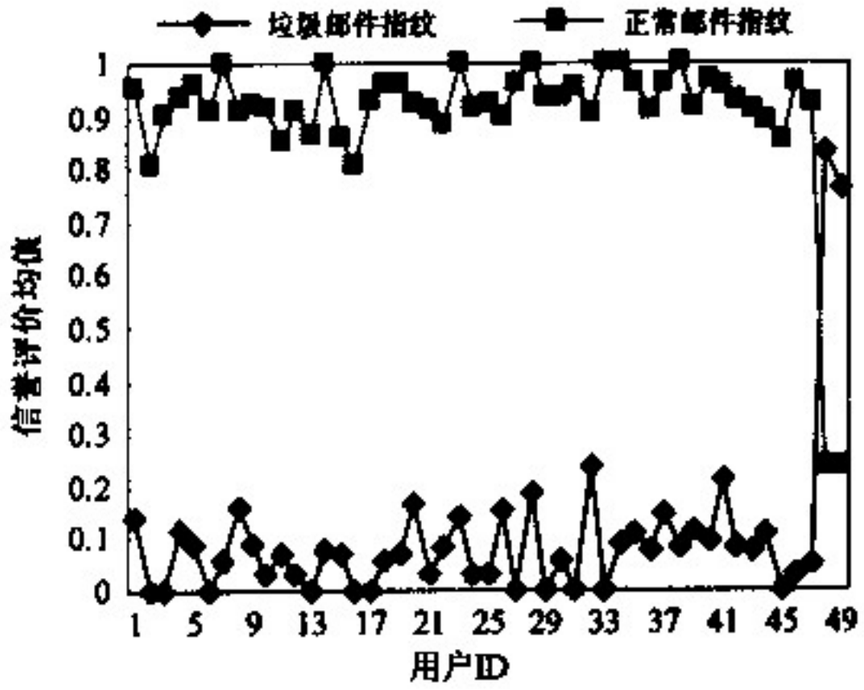


图 4 用户对两类指纹评价均值

该实验结果说明，本文方法可利用用户对邮件信誉评价结果，识别邮件用户群体中的恶意用户。

图 5 显示 8 轮实验中每一轮实验后算法的准确率和召回率。

从图中可知算法准确率较高，多轮发送后稳定于较为理想的数值，而对应的召回率有所下降。需要指出的是该实验没有对第一轮的结果进行记录，

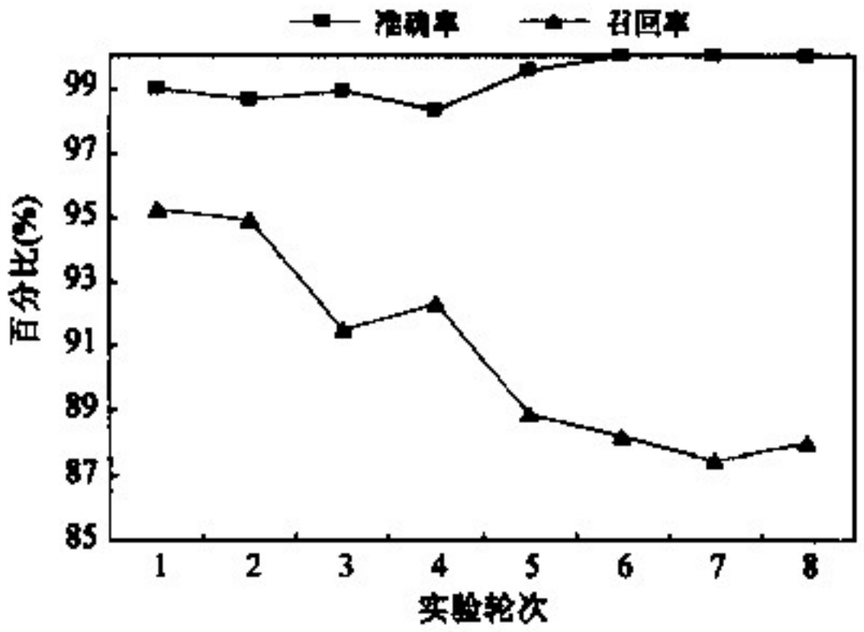


图 5 本方法的准确率和召回率

这是由于在进行第一次之前计算用户评价所需的历史评价信息不存在。

实验过程中也模拟了大量恶意用户情况下对本方法的影响。实验中将邮件用户群体分为两类，其中 0 – 24 为正常用户，25 – 49 为恶意用户，并重新构建用户关系矩阵。

向邮件用户群体随机发送 50 封正常邮件（邮件编号为 1 – 50）和 50 封垃圾邮件（邮件编号为 51 – 100），并重复 5 次。进行邮件评价时，正常用户对邮件指纹评价基本与事实相符，恶意用户则相反。经过统计，用户评价率在 80% 左右，其中错误的评价约占 25%。接下来将 50 封正常邮件和 50 封垃圾邮件分别发送给正常用户和恶意用户，并对评价结果进行统计，得到如图 6 和图 7 的实验结果。

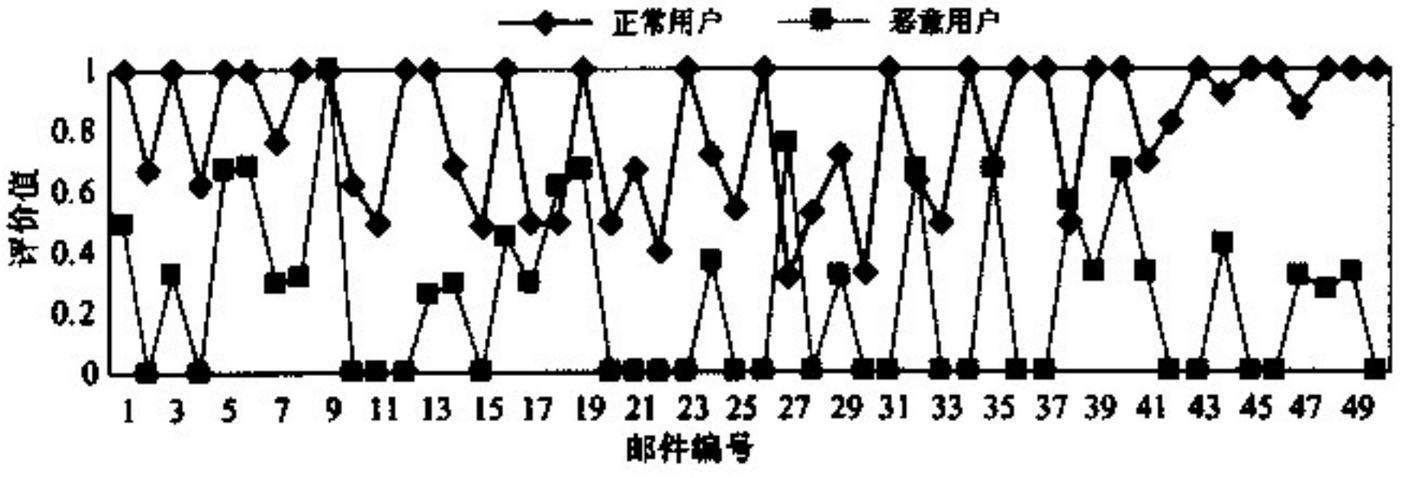


图 6 本文方法对正常邮件指纹评价

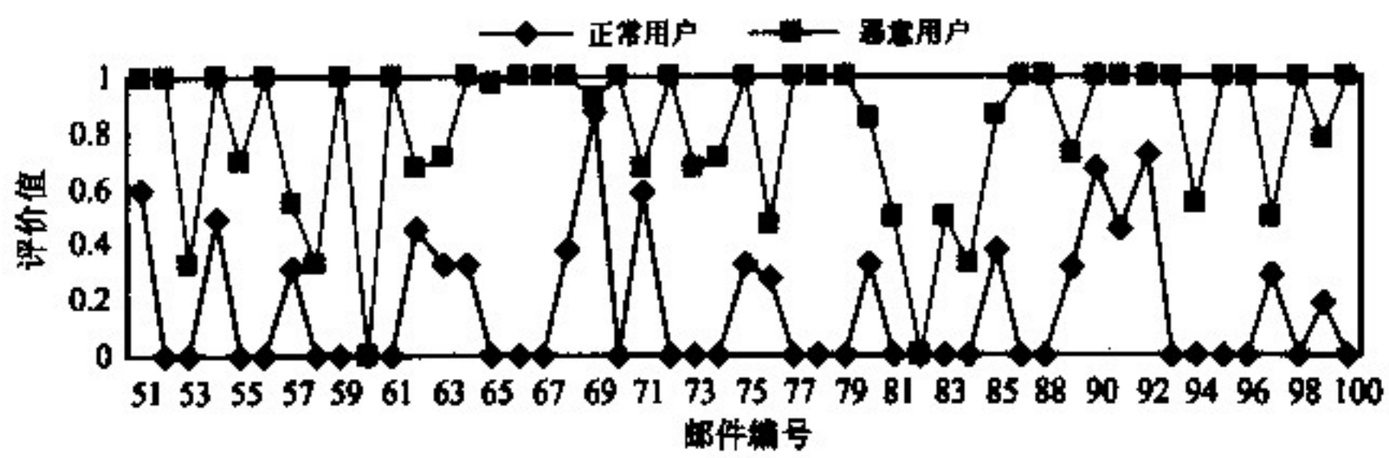


图 7 本文方法对垃圾邮件指纹评价

从图6和图7中可以看到,本文方法对同一封邮件发往正常用户与恶意用户的识别结果是不同的。对于发往正常用户的正常邮件,本文方法大部分判断为正常,少部分判断为可疑。对于发往正常用户的垃圾邮件,大部分判为垃圾,少部分判为可疑,存在极少的误判,总体上判断结果与邮件的真实属性相符。理论上,当邮件发往正常用户时,所挖掘的路径中不会存在恶意用户节点,这样恶意用户对邮件的评价不会影响本文算法的识别结果。当邮件发往恶意用户时,无论判断结果如何,对于正常用户来说并没有影响。

当恶意用户对邮件评价时,由于正常用户和恶意用户间不存在双向通信关系,所以路径集中一般不会存在正常用户节点,所以恶意用户的评价不会对正常用户的信誉造成影响或影响很小,其评价只会导致其他恶意用户信誉值发生变化。另一方面,恶意用户的评价可能导致邮件指纹信誉的变化,但只要指纹的信誉没有超过正常或垃圾指纹的阈值,就不会影响本文算法的识别结果。

本文还对不使用用户关系挖掘的信誉评价与本文算法在垃圾邮件识别效果方面进行了对比。前者仅仅根据所有用户对邮件指纹的综合评价来判断邮件的属性,而忽略用户之间的通信关系。

实验过程中将相同的50封垃圾邮件样本发送给正常用户。图8描述了未使用用户关系挖掘方法与本文算法的评价结果对比。可以看到,建立用户关系挖掘之上的判别结果在准确性上有很大的提高,而在不进行用户关系挖掘的情况下判断结果几乎是不正确的。

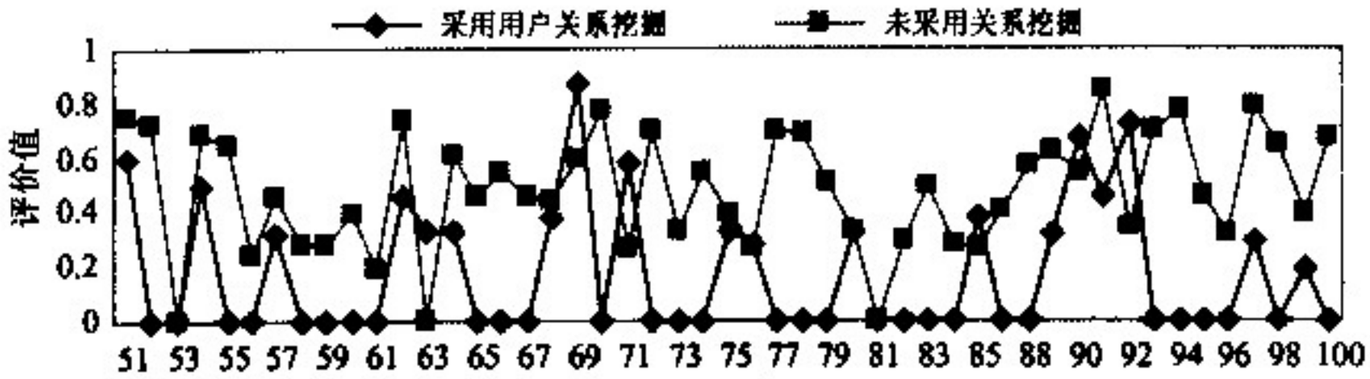


图8　采用与未采用用户关系挖掘时的垃圾邮件识别结果比较

## 5　结　论

本文通过挖掘用户历史邮件通信关系,并将其与邮件评价相结合,提出了一种基于用户关系挖掘和信誉评价的垃圾邮件识别算法,即基于邮件接收用户和邮件指纹形成邮件通信路径集,在此基础上形成邮件用户和邮件指纹的信誉评价结果,并依托该结果识别垃圾邮件。实验结果表明,本算法可以准确、有效地识别垃圾邮件,可应用于企业网邮件网关,以识别垃圾邮件。下一步的研究包括改进邮件指纹提取算法,以更好地防范未知垃圾邮件,此外,还要研究更为高效的路径挖掘算法,以提高垃圾邮件识别效率。

**参考文献**

[1] 张耀龙. 行为识别技术在反垃圾邮件系统中的研究与应用. 北京:北京邮电大学,2006

[2] DCC: Distributed CheckSum Clearinghouses, Version 1.3.115 [EB/OL]. (2009-08-26)[2009-07-25]. http://www.dcc-servers.net/dcc/

[3] Razor. Vipul's Razor: Reference Description [EB/OL]. (2007-05-10)[2009-08-06]. http://razor.sourceforge.net/.

[4] Prakash V, ODonnell A. A reputation-based approach for efficient filtration of spam: Reference Description [EB/OL]. 2005[2009-08-20] http://www.cloudmarK.com/releases/docs/wp_reputation_filtration_10640406.pdf.

[5] Leiba B, Ossher J, Rajan V, et al. SMTP path analysis. In: Proceedings of the 2nd Conference on E-mail and Anti-Spam, California, USA, 2005

[6] 张尼,姜誉,方滨兴等. 基于邮件路径地理属性分析的垃圾邮件过滤算法. 通信学报, 2007, 28(12):90-95

[7] Taylor B. Sender reputation in a large webmail service. In: Proceedings of the 3rd Conference on E-mail and Anti-Spam, California, USA, 2006

[8] Zheleva E, Kolcz A, Getoor L. Trusting spam reporters: A reporter-based reputation system for email filtering. *ACM Trans on Information System* (*TOIS*), 2008, 27(1): 65-92

[9] Yen J. Finding the *K* shortest loopless paths in a network. *Management Science*, 1971, 17(11): 712-716

[10] Border A, Glassman S, Manasse M, et al. Syntactic clustering of the web. In: Proceedings of the 6th International World Wide Web Conference, Essex, UK, 1997

[11] Chowdhury A, Frieder O, Grossman D, et al. Collection statistics for fast duplicate document detection. *ACM Trans on Information System*, 2002, 20(2): 171-191

[12] Cormack G. TREC 2006 spam track overview. In: Proceedings of the 15th Text Retrieval Conference, Gaithersburg, USA, 2006

# A spam identification algorithm based on user relationship mining and reputation evaluation

Wang Wei, Man Dapeng, Xuan Shichang, Yang Wu, Qiu Wenzhen

(Information Security Research Center, Harbin Engineering University, Harbin 150001)

**Abstract**

According to behavior-based spam identification ideas, a spam identification algorithm based on user relationship mining and reputation evaluation is proposed. First, a mail user relationship description model is established on the basis of analyzing historical communication in the mail user group. When mail reputation is evaluated, the potential user communication relationship is mined using the model proposed, and the mail communication path collections are established according to mail receiving users and mail fingerprints. The comprehensive reputation value is calculated to identify spam according to historical evaluation of mail fingerprints given by users in the path collections. When a mail receiver gives his feedback, the reputation of user and mail fingerprints is updated. The experimental results show that the proposed algorithm can identify spam correctly and effectively. The availability is considered in the design of the algorithm, which makes it work in actual spam identification systems.

**Key words**: spam, behavior analysis, user relationship mining, reputation evaluation