Hans Journal of Data Mining 数据挖掘, 2012, 2, 16-20
http://dx.doi.org/10.12677/hjdm.2012.22004 Published Online April 2012 (http://www.hanspub.org/journal/hjdm)

Research on Spam SMS Detection and Prevention on Android Platform

Yin Li[1,2], Mingyu Fan[1,2], Guangwei Wang[1,2]

[1]Department of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu
[2]Research Center of Information Security, University of Electronic Science and Technology of China, Chengdu
Email: vince-32@126.com

Received: Feb. 3rd, 2012; revised: Feb. 28th, 2012; accepted: Mar. 2nd, 2012

Abstract: Mobile terminal is currently a fast developing area, among which Android holds quite a big market share. But Android faces huge security threats due to its platform openness, especially threats from spam SMS. This dissertation researches on spam SMS problems and proposes a method by combining blacklist and Bayesian Algorithm on the detection of spam SMS. By testing the implemented system, the purpose of filtering and effectively resisting harassment from spam SMS has been achieved.

Keywords: Mobile Terminal; Android; Spam SMS Detection; Bayesian Algorithm

Android 平台垃圾短信检测及防范研究

李　寅[1,2]，范明钰[1,2]，王光卫[1,2]

[1]电子科技大学计算机科学与工程学院，成都
[2]电子科技大学信息安全研究中心，成都
Email: vince-32@126.com

收稿日期：2012 年 2 月 3 日；修回日期：2012 年 2 月 28 日；录用日期：2012 年 3 月 2 日

摘　要：当前移动终端的发展十分迅速，其中 Android 平台以其开放性占据了很大的市场份额。然而 Android 平台由于开放性，其面临的安全威胁十分突出，尤其是来自垃圾短信的威胁。本文研究了现阶段 Android 平台面临的垃圾短信骚扰问题，并针对垃圾短信检测提出了结合黑名单和贝叶斯算法的检测方法。本文实现的系统经测试能达到过滤垃圾短信的目的，并有效抵御垃圾短信的骚扰。

关键词：移动终端；Android；垃圾短信检测；贝叶斯算法

1. 引言

目前，智能移动终端的普及十分迅速，Android 平台占据了很大的市场份额，但伴随而来的是日趋严重的安全问题。尤其是来自垃圾短信的威胁，使得用户面临着经济损失、隐私泄露及被骚扰的危险[1]。目前传统的解决方案普遍大多局限在电信运营商层面，提供的仅是针对非法服务提供商(SP)的监控，无法对个人的恶意、非法行为进行有效监控，用户端的垃圾短信骚扰问题始终没有得到很好的解决。因此本文的写作目的在于研究、总结用户端 Android 平台垃圾短信的检测方法，并编程进行实验及测试，为今后垃圾短信的防治提供理论和数据支持。

本文的组织结构如下：第一章介绍 Android 平台面临的垃圾短信威胁；第二章分析垃圾短信的危害和现有检测技术；第三章着重阐述垃圾短信检测方法及系统实现；第四章对提出的检测系统进行测试并提供测试结果分析；第五章是结束语，对本文作一个回顾和展望。

2. 研究现状

2.1. 垃圾短信的定义及危害

垃圾信息是指未经接收者同意、包含违法或具有广告性质内容，或恶意报复他人的侵害通信自由的信息[2]。垃圾短信的泛滥，具有十分大的危害，具体表现在以下几个方面。

1) 违法内容的短信为违法犯罪活动提供了活动的渠道，潜在地威胁到了用户的生命财产安全，并且，违法内容的短信实际上为犯罪分子提供了一个沟通的平台，对社会治安起到了很大的负面影响。

2) 具有广告性质的短信对用户的影响存在两个方面首先，用户不能甄别短信的真实性，任何非法的商业行为都能通过该平台进行，如销售伪劣产品等。其次，广告短信往往针对特定用户会多次重复发送，对用户形成一种骚扰，这种危害主要体现在对用户精神层面的伤害。

3) 恶意骚扰、报复他人的短信对用户造成的危害主要在精神层面，对用户造成很大的困扰，虽然用户基本不会面临经济损失的风险，但其危害也不容小视。

垃圾短信很有可能与 Android 平台的恶意程序相结合，如在短信中包含一些非法广告、骚扰以及带有病毒的链接地址，一旦用户收到这种类型的短信，很有可能误将其中包含的链接打开，致使手机中毒[3]。

2.2. 垃圾短信检测技术

垃圾短信检测技术大多都借鉴了垃圾邮件的检测技术。由于手机功能十分有限，因此早期的检测方法主要集中部署在电信运营商这一端，而随着智能手机的普及以及手机功能的日渐强大，基于客户端垃圾短信检测研究也逐渐增多。

文献[4]分析了基于状态的检测方法，状态检测是指电信运营商对某一手机或 SP 接收或发送的短信数量进行检测，一旦超过某个阈值则向用户报警。如果能检测到发送垃圾短信的个人或 SP 则可立即对其进行过滤。状态检测包括两种方法，第一种是检测单位时间内发送的短信数量，达到阈值则自动报警。该方法对缓慢攻击无能为力。第二种是由运营商检测每两条短信发送或接收的间隔，如果过于频繁，超过了设定的频率则自动报警。此方法适用于运营商，缺点是会对电信服务器带来较大的额外负担[4]。

智能手机的普及以及手机功能的日渐强大为客户端垃圾短信检测创造了可行性。根据近几年国内外的研究成果，一些新的方法和检测手段被引入了垃圾短信检测领域。

文献[5]基于随机映射提出了一种实时垃圾短信检测方法，它在 k-means (k = 2)聚类算法的基础上增加了额外的降维操作，将向量点映射到随机选择的低维空间。该方法利用了大量垃圾短信在分类期间更易被分类为大样本容量的类别的原理，在 SMSC(短消息服务中心)实现了转发短信的空档期，实现了对垃圾短信的过滤。该方法能在高维空间基于相似度分类短信，但它不能区分大量发送的正常短信和垃圾短信[5]。

文献[6]提出了一种基于 CAPTCHA 验证机制的垃圾短信检测系统。CAPTCHA 是一种自动区分计算机和人类的验证机制，我们平时使用的验证码即其一种常见应用。该文献在 SMSC 插入了 CAPTCHA 验证系统，当 SMSC 收到短信时，其首先随机选择图片和文本，生成验证短信发送给发件人。如果发件人正确回复了验证短信，则被认为是人类发送者，否则判定为计算机程序[6]。该方法能有效区分人类和计算机，对短信群发器有很好的检测效果。但该方法每次都要求发件人回复，过程十分繁琐，且只能检测通过群发器发送的垃圾短信。

José María Gómez Hidalgo 等[7]分析了现有各种垃圾邮件的检测方法，指出在垃圾邮件检测领域，机器学习算法是比较有效的检测方法，其中贝叶斯聚类算法非常适用于引入到垃圾短信检测领域[7]。但作者并未进一步深入探索如何利用贝叶斯聚类算法检测垃圾短信。

根据对现有 Android 平台的部分垃圾短信过滤工具(如 360 手机卫士、金山手机卫士等)的研究，目前普遍采用的做法是基于黑名单和关键字的过滤，需要用户的人工输入黑名单和过滤关键字，自动化程度较低。

3. Android 平台垃圾短信检测系统

文献[7]已经指出基于机器学习的方法(特别是贝叶斯聚类算法)是很有效的检测方法，并且现有的智能手机平台和手机强大的硬件资源为客户端的垃圾短信检测系统的提供了可行性。另外，黑名单机制也能

以较低的系统开销保证垃圾短信的过滤[8]。因此本文结合黑名单机制，提出了基于机器学习的 Android 平台垃圾短信检测方法，并实现了相应程序。

3.1. 系统总体设计

本文提出的基于机器学习的 Android 平台垃圾短信检测系统结合了黑名单机制和机器学习的方法，进行垃圾短信检测。

黑名单的实现较简单，通过用户输入黑名单号码(包括直接输入号码和选择联系人两种方式)，在收到短信时提取发件人号码进行匹配，匹配成功则过滤短信并提示用户。黑名单主要用于快速过滤来自已知联系人的短信。

基于机器学习的检测方法实现比较复杂，本文采用了机器学习算法中的贝叶斯聚类算法，将短信分类为垃圾短信和正常短信两类。贝叶斯算法的实现过程是首先收集垃圾短信样本作为训练集训练分类器，训练完成后使用另外一组样本作为测试集进行测试。

根据黑名单机制和基于机器学习的检测方法的设计思路，系统总体设计为四个模块：操作界面模块、数据库操作模块、Service 模块以及 Broadcast Receiver 模块。

操作界面模块分为手动输入和导入通讯录两种输入方式；数据库操作模块接收请求，完成黑名单的匹配并反馈结果；Service 模块按功能分为黑名单和贝叶斯分类器两大部分：黑名单部分按照上述方法进行匹配判定。贝叶斯分类器采用离线模式进行训练，训练好后置入系统，完成对短信的分类；Broadcast Receiver 模块由收到短信事件触发，根据判定结果决定是否过滤短信。

上述模块相互配合完成整体功能，当收到一条短信时按如下流程检测：首先触发 Broadcast Receiver 根据操作界面设置判定是否启动过滤服务；若过滤服务已开启，则运行 Service 模块进行黑名单匹配和短信分类过程。其中黑名单匹配在短信分类前完成，并只对不属于黑名单的短信才进行第二步的分类过程；最后，Broadcast Receiver 模块根据检测结果过滤对应短信。系统总体结构图如下图 1 所示。

3.2. 系统详细设计

本文提出的基于机器学习的 Android 平台垃圾短信检测系统结合采用了黑名单机制和贝叶斯分类算法，其中贝叶斯分类器的设计是关键。另外，本系统在黑名单输入方面也考虑了多种方式。下面我们分别阐述操作界面以及贝叶斯分类器的详细设计。

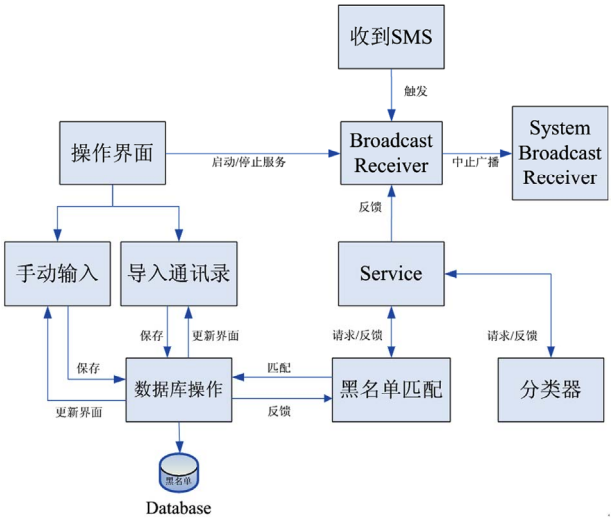

Figure 1. System overall architecture
图 1. 系统总体结构

3.2.1. 操作界面

目前常用的输入方式包括手动输入、导入短信记录、导入通讯录以及按地区过滤等方式。由于垃圾短信通常会被删除，因此导入短信记录的方式并不实用。另外，按地区过滤的方式会造成无法接收同一地区其它短信，误过滤率很高。

经综合分析，手动输入陌生号码和导入通讯录两种方式比较适宜，其中前者包括单个添加和批量添加两种方式。设置好的黑名单通过提交请求到数据库操作模块进行保存。导入通讯录在程序启动时从通讯录中读取并显示联系人信息，用户只需选择联系人并保存黑名单即可。

上述两种输入方式既保证了良好的垃圾短信过滤功能，又避免了对部分短信的误过滤。

3.2.2. 贝叶斯分类器

贝叶斯算法是一种常用的聚类算法，其最基本的形式是朴素贝叶斯方法。其原理是计算文本 s 属于某个类别的条件概率 $P\left(c_j\middle|s\right)$，概率值最大的类即文本 s 所属的类别[9]。计算概率值时利用了贝叶斯公式：

$$P\left(c_j\middle|s\right)=P\left(c_j\right)P\left(s\middle|c_j\right)\Big/P\left(s\right) \quad (1)$$

其中 $P(c_j)$ 是类别的先验概率，$P(s|c_j)$ 是类别条件概率。s 由一组特征向量表示($t_1, t_2,\ldots, t_n$)，假设各特征彼此独立，则：

$$P(s|c_j) = P(t_1|c_j)P(t_2|c_j)\cdots P(t_n|c_j) \quad (2)$$

其中 $P(c_j)$ 和 $P(t_1|c_j)$ 可用训练集估计。

为提高自动化程度和检测速度，本文设计的贝叶斯分类器采用离线分析的方式获取关键字，训练完成后再置入整体系统中，实时分析短信内容。

垃圾短信具有长度短、发送频率大的特征，且其内容通常是恶意骚扰或广告等。因此本文选择短信长度、发送频率及内容作为特征向量，采用朴素贝叶斯算法，按如下步骤计算特征：

1) n 维特征向量 $s = (t_1, t_2,\cdots, t_n)$代表了一条短信样本，用于描述对 n 个属性 $a_1, a_2,\cdots, a_n$ 样本的 n 个关键字。

2) 根据朴素贝叶斯算法，分别计算短信样本 s 属于 c_0 和 c_1 两个类别(分别代表正常短信和垃圾短信集合)的后验概率值，选择概率值最大的类别作为 s 的分类，该类被称为最大后验假定。

3) 由公式(1)，$P(s)$为常量，要最大化 $P(c_j|s)$，则需最大化 $P(c_j)P(s|c_j)$。$P(c_j)$ 是类别的先验概率，先验概率可以用 c_j 类中的样本数 m_j 除以样本总数 m 求得。

4) 上述计算的复杂度随特征向量长度急剧增大。为降低计算复杂度，我们假设特征向量各属性彼此独立，根据公式(2)，我们可利用训练集估计 $P(t_1|c_j)$，其中 $i = 1, 2,\cdots, n$。

5) 分别计算 $P(c_j)P(s|c_j)$，将短信样本 s 归类到对应后验概率值最大的类。

4. 系统测试

根据系统检测垃圾短信的两种机制，我们分别对黑名单过滤、贝叶斯分类器以及系统性能进行了测试。

4.1. 黑名单过滤测试

我们将人工收集的 613 条垃圾短信输入本系统检测，统计其平均响应时间、过滤准确率等数据，得到了如下表 1 测试结果。

其中连续过滤准确率是指在连续收到多条短信，且每两条短信间隔不超过 2 s 的情况下，能正确过滤并显示的短信条数占测试短信总量的比例。

Table 1. Blacklist filter test
表 1. 黑名单过滤测试

测试指标	结果
平均响应时间	0.38 s
过滤准确率	99.7%
连续过滤准确率	99.3%

测试结果表明黑名单过滤的响应时间和准确率都较理想，能快速过滤来自已知联系人和部分陌生号码的短信。

4.2. 贝叶斯分类器测试

在构建分类器时，我们使用上述 613 条垃圾短信对分类器进行了训练。测试期间我们另外收集了 1121 条短信作为分类器测试集，其中垃圾短信 217 条(恶意骚扰类 26 条、广告类 134 条、诈骗类 57 条)，占 19.4%。同时仿照黑名单测试检测了响应时间及准确率等指标。

将以上测试集输入贝叶斯分类器检测，共过滤出垃圾短信 211 条，占所有短信的 18.8%。经人工统计，其中包括恶意骚扰类 23 条、广告类 133 条，以及诈骗类 55 条。如下表 2 所示，贝叶斯分类器的平均响应时间为 0.33 s，过滤准确率和连续过滤准确率分别达到了 97.2%和 96.3%，能够满足客户端的需要。

4.3. 性能测试

本小节性能测试的内容主要包括源文件大小、运行期间内存占用，以及后台过滤功能等。如下表 3 所示本文提出的系统在非运行期间仅占用 21.6 KB 的内存空间，运行期间的内存占用为 63 KB，几乎不会对手机造成负担。

接下来我们测试后台过滤功能。退出整个程序，我们向模拟器发送了一条短信，利用贝叶斯分类器进行分类。如下图 2 所示，该系统仍能够过滤短信，完全不影响用户的正常使用。

5. 结束语

本文提出了基于机器学习的 Android 平台垃圾短信检测系统。该系统结合了黑名单和朴素贝叶斯方

Table 2. Bayesian classifier test
表 2. 贝叶斯分类器测试

测试指标	结果
平均响应时间	0.33 s
过滤准确率	97.2%
连续过滤准确率	96.3%

Table 3. Performance test
表 3. 性能测试

测试指标	结果
源文件大小	21.6 KB
运行期间内存占用	63 KB

Figure 2. Filter spam SMS test in background
图 2. 后台过滤功能测试

法，其中黑名单适用于过滤来自已知联系人的恶意骚扰类垃圾短信，而贝叶斯分类器则基于内容进行过滤，自动化程度较高，适用于过滤广告或诈骗类的垃圾短信。根据系统测试结果，该系统是一款轻量级的过滤系统，非常适用于手机用户，但贝叶斯分类器的检测准确率方面还有提升的空间。因此今后的工作重点将围绕提高提高贝叶斯分类器的检测准确率以及增加后续处理功能进行。

6. 致谢

本论文撰写期间，范明钰和王光卫老师提供了很全面的指导，帮助我顺利完成了该论文撰写，在此表示衷心地感谢！

参考文献 (References)

[1] 戴沁芸, 王允非. 网络安全迎来 3G 时代[J]. 信息安全与通信保密, 2010, 11(5): 34-37.
[2] 何培舟, 温向明, 郑伟. 垃圾短信的防治方法研究[J]. 通信技术, 2008, 41(12): 340-341.
[3] 落红卫, 孙萌. 移动终端安全威胁和防护措施[J]. 现代电信科技, 2009, 39(11): 24-25.
[4] 张尼, 张智江, 宋建, 李晓宇. 垃圾短消息过滤技术综述[J]. 移动通信, 2009, 33(6): 17-21.
[5] S. Dixit, S. Gupta and C. V. Ravishankar. Lohit: An online detection & control system for cellular SMS spam. Proceedings of the IASTED International Conference Communication, Network and Information Security, Phoenix, 2005: 52-54.
[6] M. H. Shirali-Shahreza, M. Shirali-Shahreza. An anti-SMS-spam using CAPTCHA. ISECS International Colloquium on Computing, Communication, Control and Management, Guangzhou, 2008: 320.
[7] J. M. G. Hidalgo, G. C. Bringas, E. P. Sánz and F. C. García. Content based SMS spam filtering. Proceedings of the ACM Symposium on Document Engineering, New York, 2006: 107.
[8] 吴文俊. 一种垃圾短消息过滤与举报系统 Mobile 客户端的设计与实现[D]. 北京大学, 2009.
[9] 潘文锋. 基于内容的垃圾邮件过滤研究[D]. 中国科学院计算技术研究所, 2004.