

用户生成内容中的垃圾意见研究综述 *

杨风雷, 黎建辉

(中国科学院 计算机网络信息中心, 北京 100190)

摘 要: 随着互联网的快速发展, 互联网用户更倾向于在博客、购物网站等场所借助于发表文章、评论等手段表达自己对感兴趣话题的关注、意见和相互交流。对这些用户生成内容进行挖掘可以得到很多宝贵的可应用于很多领域的知识, 但其中存在的不请自来、与话题无关(广告等)或虽与话题相关但是虚假性的内容严重影响了挖掘的效果, 并产生了误导用户、浪费资源等不良影响。因此, 作为意见挖掘研究工作重要组成部分的垃圾意见研究工作开始得到空前的重视。从垃圾意见特征、垃圾意见检测方法两个方面对目前垃圾意见的研究状况进行了综述。

关键词: 用户生成内容; 意见挖掘; 垃圾意见; 检测方法

中图分类号: TP391

文献标志码: A

文章编号: 1001-3695(2011)10-3601-05

doi:10.3969/j.issn.1001-3695.2011.10.001

Survey on research of opinion spam in user-generated-content

YANG Feng-lei, LI Jian-hui

(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: With the rapid development of Internet, Internet users are more inclined to express their concerns, views for the interest topics and interact with each other in blog sites, ecommerce sites and other places by means of publishing articles, reviews and other means. And a lot of valuable knowledge mined from the user-generated-content can be applied to many fields. However, the contents which are uninvited, unrelated to the topics (e. g. advertising) or related to the topics but false in the user-generated-content have serious impact on the results of mining, and produce bad results such as misleading internet users and wasting resources and so on. Therefore, the research of opinion spam which is an important part of opinion mining research is receiving unprecedented attention. This paper summarized the research of opinion spam from the features of the opinion spam and the detection method for the opinion spam.

Key words: user-generated-content; opinion mining; opinion spam; detection method

0 引言

据中国互联网络信息中心(CNNIC)发布的中国互联网络发展状况统计调查报告,截至2010年12月31日,中国的网民数达4.57亿,网民渗透率已达34.3%,互联网发展速度惊人^[1]。随着互联网的快速发展,它已经显著地影响了人们的工作、生活、学习,网民的上网设备和使用互联网的应用也进一步呈现多元化的趋势,此状况下人们表达意见的模式也在发生着显著的变化。当前,互联网媒体已被公认为是“第四媒体”,其表现出了与传统媒体极其不同的特点。互联网环境下,互联网用户可以通过BBS论坛、博客、微博、社交网站等场所,借助于互联网信息的浏览、转发、发表评论、发布博客文章及评论等手段来表达自己对于感兴趣话题的关注和意见,也可以与相关的人员进行交流,这种模式已经成为互联网用户表达意见、相互交流的重要模式。在这种模式下用户所产生的观点、意见等内容被称为用户生成内容,这与传统的由网站所有者提供内容的模式完全相反。目前,人们已经认识到,对这些用户生成内容进行挖掘可以得到很多宝贵的可应用于很多领域的知识。比如通过用户针对某一产品的评论,潜在的用户可以在购买产品之前对其进行更为切实的了解;产品厂商也可以通过了解用户对产品的评论改进其产品从而更好地满足用户的需要。

由于越来越多的用户在互联网上发布自己的观点或意见,这类用户生成内容越来越多,仅靠人工的方法难以应对网上的海量信息。为了使用计算机帮助用户快速获取和整理(挖掘)这些用户生成内容,就需要用到意见挖掘(opinion mining)技术。意见挖掘也称为观点挖掘、情感分析等,是指对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程。这是一个虽然新兴但很活跃的多学科交叉的研究领域,与文本挖掘、信息检索、机器学习、本体学、自然语言理解等均有着较强的相关性。之前,意见挖掘研究领域中的研究工作主要集中在采用自然语言处理、数据挖掘等技术完成情感信息的抽取(意见所有者、意见对象、意见词语等)、情感信息的分类(主客观、倾向性)以及情感信息结果的呈现等^[2~15]。但目前,一些研究者在研究工作中发现,由于互联网上的用户生成内容没有质量控制机制,网民可以在互联网上发布任何观点,用户生成内容中存在很多噪声内容或者虚假的内容,这在很大程度上影响了意见挖掘的效果,即用户意见的有用性、可信性问题上影响了意见挖掘的质量和效果^[16,17]。为此,垃圾意见研究工作开始得到关注并成为意见挖掘的一个重要组成部分。垃圾意见指的是用户生成内容中不请自来的、和话题无关的(广告等)或者虽然与话题相关但是虚假性的内容,因为这些内容一般出现在论坛、博客、社交网站等所谓的Web 2.0应用

收稿日期: 2011-04-18; 修回日期: 2011-05-29 基金项目: 中国科学院“十一五”信息化建设专项资助项目(INFO-115-C01)

作者简介: 杨风雷(1973-),男,副研究员,博士,主要研究方向为数据挖掘、分布式计算等(flyang1999@yahoo.com.cn);黎建辉(1973-),男,正研级高工,博导,博士,主要研究方向为海量数据管理、处理与挖掘分析的理论、方法与关键技术。

中,所以也称为垃圾 2.0(或者 Web 2.0 垃圾)^[18,19]。与意见挖掘中的其他研究内容相比,之前垃圾意见研究的受关注度相对较低,但作为意见挖掘研究的一个不可或缺的重要组成部分,其对意见挖掘的意义极为重大,将成为意见挖掘步骤中的重要一环,目前已经开始得到空前的重视^[16,18~20]。因此,有必要对目前垃圾意见研究的总体状况作一综述,使得业界对此有一个系统的了解。

1 垃圾意见特征

针对互联网上的垃圾页面(Web spam)、垃圾邮件(e-mail spam)的研究已经进行了很多年,其中垃圾页面的主要目的是欺骗搜索引擎得到比较高的排名,从而吸引人们去点击访问。其所采用的方式主要是内容作弊、链接作弊,前者主要通过网页的标题、元数据、正文、锚文本、URL 中重复一些关键词或者大量添加不相关的关键词,从而提高页面和特定查询的相关度以得到较高的搜索排名或者加大相关查询的量以得到更大的访问量;后者主要通过增加指向重要网站的链出链接和采取欺骗措施增加有效的链入链接等方式提高页面的中心性和权威性,从而得到更高的搜索排名。而垃圾邮件中更多涉及的是一种不请自来的、带有商业性质的广告^[21~33]。

下面对用户生成内容中的垃圾意见的特征进行叙述。为使叙述更有条理及理解方便,首先对用户生成内容中的垃圾意见从其产生方式、表现形式、性质、目的等纬度进行分类,如图 1 所示。下面分纬度进行叙述^[10,16,18~20,34~42]。



图1 垃圾意见分类

1.1 按垃圾意见的产生方式分类

从垃圾意见的产生方式上看,主要可分为人工产生垃圾意见、采用自动机产生垃圾意见两种方式。前者是采用人工的方式发布内容,这里面又可以分为单独一个人发布垃圾意见和多人组成的一个团体协作发布垃圾意见两种方式。后者主要采用一些信息技术的手段形成一些工具自动地发布垃圾内容,这里面也可以分为两种形式:a)一些相对较简单的对特定网站自动发布垃圾意见的软件;b)与爬虫类似的一种代理软件,称为垃圾机器人(spambot),它可以在网络上进行爬行,发现合适的可以发布垃圾意见的场所(论坛、博客等)时即可发布相应的垃圾内容。

1.2 按垃圾意见的表现形式分类

如前述,用户生成内容中的垃圾意见存在于博客、论坛等诸多应用中,下面按照其主要存在的场所、表现形式等分别叙述。

1) 博客中的垃圾意见

在博客中垃圾意见的主要表现形式为建立垃圾博客日志(splog)、发布垃圾评论(comment spam)、建立垃圾链接提醒(trackback spam)等。在这些日志、评论、链接提醒中,垃圾意见作者通过注入包含垃圾信息或者链接(链向特定网页或网站)的内容,以实现其误导用户或者搜索引擎、推广特定对象等特定目的。

2) 在线购物网站上针对产品的垃圾意见

此领域的垃圾意见主要表现为在一些在线购物网站上,针对特定产品的不可信的评论(不恰当抬高或者不切实际地贬低特定产品)和一些与产品无关的言论(如广告、无关评论或者仅仅与品牌有关的评论)等。

3) 社交网站中的垃圾意见

此领域的垃圾意见主要表现为在个人信息(profile)中添加一些虚假或无关的个人信息、发布垃圾内容或评论来误导用户或者推广特定网页。

4) 维基网站中的垃圾意见

此领域的垃圾意见主要表现为通过修改维基网页,从而将特定的垃圾链接或内容添加到维基网站中。

5) 推荐系统中的垃圾意见

此领域的垃圾意见主要表现为通过加入一些不恰当的推荐信息(attacks),从而欺骗推荐系统,使得特定对象得到不应得的推荐排名。

6) 论坛中的垃圾意见

此领域的垃圾意见主要表现为通过发布垃圾贴或者回复垃圾评论的方式将垃圾内容、链接或附件展示在用户面前以误导用户(宣传或贬低特定对象)或推广特定网页、对象。

1.3 按垃圾意见性质分类

1) 不可信意见

不可信意见也就是欺骗性的意见,一方面表现为对特定的对象、事件、人物等给出不符合实际情况的超高评价、恭维等;另一方面也可能表现为对特定的对象、事件、人物等给出不符合实际情况的超低评价、谩骂、攻击等。

2) 低质量意见

此种意见内容一般长度较短,其内容可能是有用的,也可能是无用的,但是由于其内容对特定的话题/产品描述不详细,不能非常确定其对特定话题/产品的意见挖掘的意义,因此也认为是一种垃圾意见(针对计算机而言)。

3) 无关意见

此种意见主要表现为广告或者与话题无关的内容。

1.4 按垃圾意见目的分类

除低质量的垃圾意见无法特别确定其倾向外,其余垃圾意见的目的主要可分为两种情况:a)为了推广特定的对象如特定的网页、网站或产品等,从而给出一些不符合实际情况的正面评价、恭维或者广告链接等;b)为了贬低某些特定的对象如网页、网站或产品等,从而给出一些不符合实际情况的负面评价、谩骂、攻击等。当然,也可能二者兼有。

相比较而言,用户生成内容中的垃圾意见和传统的垃圾邮件、垃圾页面之间最大的不同在于其传播不是通过自己建立的网站或邮件服务器进行,而主要是通过一些合法的、正常网站(如博客、论坛、社交网站、维基网站等合法的网站,甚至包括一些政府、学校、商业企业的网站)来进行。正由于此,用户生成内容中的垃圾意见生成更容易、影响更大。据不完全统计,全球平均每天产生评论性垃圾意见 3 000 万^[43]。这些垃圾意见的存在造成了极大的社会、经济影响,诸如误导用户、破坏特定对象(产品、网站等)的声望、非法提高特定对象(产品、网站等)的品牌形象、造成网络带宽和其他资源的浪费等。有研究表明,用户生成内容中的垃圾意见所造成的损失已经超过垃圾邮件所造成的损失^[44]。因此,对用户生成内容中的垃圾意见特征、检测方法等进行研究自然就变成一件受到领域研究者非常重视的重要工作。

2 垃圾意见检测方法

如上述,传统的互联网垃圾检测研究工作主要侧重在针对垃圾页面和垃圾邮件的检测研究。其中垃圾页面所采用的欺骗手段包括在网页的特定位置重复一些关键词、大量添加不相关的关键词、增加指向重要网站的链出链接和采取欺骗措施增加有效的链入链接等,通过这些方式可以提高页面的中心性、权威性或网页在搜索引擎中和特定查询的相关度、排名等,从而得到更大的用户访问量。垃圾邮件中更多的内容是一些和用户关注点无关的广告。而在用户生成内容的垃圾意见中,除了与传统的互联网垃圾类似的无关内容、广告等外,还存在众多的对特定的人、事件等对象给予不符合实际情况的超高或贬低的评价(不可信意见),从而误导受众、挖掘系统等。总体上,用户生成内容中的垃圾意见检测问题可以看做是一个分类问题,更准确的可以看做是一个二值分类问题(是垃圾意见、不是垃圾意见),原则上这可以通过各种基于规则或机器学习的方法完成。但相对传统的垃圾页面、垃圾邮件检测工作而言,考虑到不可信意见的识别难度问题(传统的垃圾页面、垃圾邮件中的垃圾特征相对比较明显,而不可信意见的特征不明显),用户生成内容中的垃圾意见(尤其是不可信意见)的检测相对更难。

为进行用户生成内容中的垃圾意见检测,有必要首先了解一下其所存在的部分相对明显的分布特点^[16,20,45~48]。a)内容相似性是指垃圾意见在不同的网站或者网站的不同频道等不同位置出现,但内容却存在极大的相似性甚至完全相同。b)作者相似性是指在一定的时间范围内可以观察到垃圾意见的作者(通过ID标志)会在一定的范围内保持相对集中。c)目标集中性是指特定时间段内,垃圾意见所指目标对象(产品、网站链接等)会保持一个相对有限的集合。d)时间集中性,如上述,针对特定目标对象的垃圾意见会相对集中出现在一个特定的时间段内。e)IP地址集中性是指在特定时间段内,垃圾意见的作者所使用的IP地址会出现在一个有限的IP地址集合内。这些分布特点可以在垃圾意见的检测工作中加以利用。当然,随着技术发展,垃圾意见的作者也在不停地改进其策略,使得垃圾意见在内容、作者、目标、时间和作者IP地址等方面尽可能地模仿正常的用户生成内容,从而进一步增加了垃圾意见的检测难度。下面按照垃圾意见存在的不同场所、不同表现形式等分别叙述垃圾意见检测方法的研究状况。

2.1 博客中垃圾意见检测方法研究

博客领域的垃圾意见研究是当前用户生成内容中垃圾意见研究工作的一个热点,目前国内的垃圾意见研究工作也主要侧重于此。之所以这样可能与2006年将博客垃圾意见过滤作为TREC比赛新增的任务有关系^[41]。博客垃圾(blog spam)一词最早出现在文献[49]中,但第一次对博客垃圾意见展开深入研究的工作出现在文献[50]中,作者提出了一种基于语言模型差异的垃圾意见检测方法,该方法通过比较博客日志、回复以及回复中链接的页面的语言模型差异,从而辨别出回复是否是一个垃圾意见,该方法不需要事先训练、也不依赖于规则库和Web链接的知识。文献[51]提出了一种利用博客日志、回复等中频率相同的子串个数(称为词汇量)识别垃圾意见的方法,该方法基于一个事实——正常博客日志中频率相同的子串个数遵循Zipf分布,而垃圾意见则不是。文献[47]基于垃圾博客和正常博客在统计特征上的差异,对多种针对博客分类有效的统计特征进行了分析,提出了基于博客页面统计特征的

过滤方法。文献[52]提出了通过计算评论和文章的相关度来判断是否为垃圾意见的识别方法,该方法不需要训练样本。文献[53]基于一些已有的方法和一些新的无需训练集的方法研究了博客垃圾意见的特征,诸如文章—评论相似度、词重复、锚文本数量、名词集中度、停用词比例、句子个数和垃圾相似度等。文献[45]中通过分析博客垃圾意见的特点,将垃圾意见分为显式垃圾意见和隐式垃圾意见两大类。对前者采用了基于规则的方法进行识别,对后者采用了LDA(latent dirichlet allocation)模型对博文抽取隐含主题信息,然后通过这些主题信息,使用基于主题的特征选取和基于主题的检索模型两种方法发现垃圾意见。文献[54]提出了一种利用网络常用语先对短小评论进行垃圾意见的识别后利用改进的相似度公式对剩余评论进行识别,之后对识别出的垃圾意见进行第二次过滤,提取出其中的合法意见,从而在一定程度上提高了垃圾意见的识别准确度。文献[46]提出了从用户名、发帖时间间隔、博文内容、锚文本和链接地址、分类标签等博客的结构特征出发的特征提取方法,在特征提取的基础上,提出了基于多结构特征的识别方法,并使用支持向量机和朴素贝叶斯模型作为分类器进行垃圾意见识别分类。

2.2 在线购物网站上针对产品的垃圾意见检测方法研究

针对产品的垃圾意见研究是用户生成内容中垃圾意见研究工作的另外一个热点。文献[55]提出“针对产品的意见是否有帮助”问题,并采用支持向量机方法实现了自动评估。文献[56]提出了“预测产品评论意见的实用分数”问题,并采用回归的方法对产品评论的实用性进行评估。文献[17]提出了“评价产品评论意见的质量”问题,并采用支持向量机的方法对评论意见的质量进行评估。以上文献中所提到的问题,是垃圾意见研究工作中的一个部分但非重点,文献[10,20,38]第一次提出了针对购物网站上产品的垃圾意见问题,并将垃圾意见分为三种不同类型,分别是非可信意见、针对品牌的意见、无实质内容的意见。研究的垃圾意见检测方法是:a)查找重复的意见;b)对于第二、三类垃圾意见,通过建立回归模型的方式进行识别;c)对于第一种垃圾意见,通过将重复的意见作为目标样本,而将其他的样本作为非目标样本,以此建立模型进行垃圾意见的识别工作。文献[34]提出了基于行为的垃圾意见作者识别方法,其中重点的行为包括垃圾意见作者会重点关注特定的产品或者产品组,他们对所关注产品给予的评价往往与其他用户不同等。基于此,作者建立了回归模型并用来识别垃圾意见的作者。文献[57]提出了识别垃圾意见作者的典型行为的方法,其中,作者首先将问题定义为“发现不期望的行为模式或模式组”,之后在定义了期望的模式基础上,对不期望的模式进行了定义,代表了不正常的或者说反常的意见作者行为,即垃圾意见行为。文献[16]中基于设计科学的思路,提出了一种针对产品的垃圾意见检测和影响评估架构,在这个体系架构中,作者采用语言模型的方法检测不可信垃圾意见,并通过实验证实了检测方法的有效性;在此基础上,作者又对不可信的负面垃圾意见的经济影响作了实证性的分析。

2.3 其他的垃圾意见检测方法研究

文献[58]中提出了针对社交网站的垃圾意见检测方法,研究中采用后向神经网络、朴素贝叶斯等方法对其中的垃圾意见进行检测。文献[59]中提出了推荐系统中基于内存的推荐算法面对垃圾意见时的脆弱性问题,并提出了基于模型的算法对其进行改进。文献[60]中提出了一种针对视频共享网站的垃圾意见检测方法,其所采用的分类技术是支持向量机。除上

述针对特定领域的垃圾意见检测方法研究工作外,文献[61~63]提出了用于专门检测垃圾机器人的基于浏览行为的检测方法,方法中基于垃圾机器人和正常用户之间浏览行为的不同,采用机器学习的方法识别垃圾机器人所产生的垃圾意见,这种方法的应用可不限于特定的领域(如博客垃圾意见识别等),是通用的。

为理解方便和进一步条理化,对以上这些研究工作,按照其在垃圾意见检测工作中所针对的主要对象不同,分为以下的类别:基于意见内容进行垃圾意见识别的方法(如文献[20, 52]中的部分研究工作)、基于意见作者的垃圾意见识别方法(如文献[46, 61]中的部分研究工作)、基于特定对象(如关键词)的垃圾意见识别方法(如文献[45]中的部分研究工作)等,如图2所示。

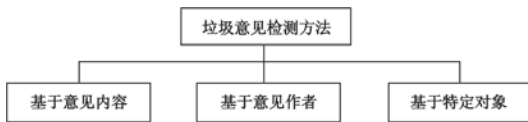


图2 垃圾意见检测方法分类

3 结束语

总体上,用户生成内容中的垃圾意见研究工作是一个比较新的研究领域,也是一个很重要的、非常有应用前景的研究领域。国际上目前的研究工作主要集中在针对博客、产品的垃圾意见研究,尚未发现对论坛、微博、维基网站中垃圾意见的研究工作;国内目前对用户生成内容中的垃圾意见研究工作刚刚开始,且主要集中在博客领域。从本文所提到的这些研究工作可以看出,此领域中很多工作尚未有效开展,对垃圾意见的特征理解尚未到位,目前的研究工作主要是针对一些侧面的尝试,垃圾意见的检测效果也有待改进。比如一些文献中认为非常有效的方法在其他的文献中却被认为效果相对不好,这也从一个角度表明目前所研究的垃圾意见检测方法的普适性尚存在不足。未来需要在进一步研究用户生成内容中垃圾意见的特征、深入分析其产生机理的基础上,研究更有效的检测方法和算法。

参考文献:

- [1] 中国互联网络信息中心. 第27次中国互联网络发展状况统计报告[R/OL]. 2011-01-18. http://www.cnnic.cn/research/bgxz/tjbg/201101/t20110120_20302.html.
- [2] JIN Wei, HO H H. A novel lexicalized HMM-based learning framework for Web opinion mining[C]//Proc of the 26th International Conference on Machine Learning. New York: ACM Press, 2009:465-472.
- [3] WAN Xiao-jun. Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis[C]//Proc of Conference on Empirical Methods in Natural Language Processing. 2008:553-561.
- [4] PANG Bo, LEE L. Opinion mining and sentiment analysis[J]. Foundations and Trends in Information Retrieval, 2008, 2(1-2):1-135.
- [5] SU Qi, XU Xin-ying, GUO Hong-lei, et al. Hidden sentiment association in Chinese Web opinion mining[C]//Proc of the 17th International Conference on World Wide Web. New York: ACM Press, 2008:959-968.
- [6] TITOV I, McDONALD R. Modeling online reviews with multi-grain topic models[C]//Proc of the 17th International Conference on World Wide Web. New York: ACM Press, 2008:111-120.
- [7] CHOI Y, CARDIE C. Learning with compositional semantics as structural inference for subsentential sentiment analysis[C]//Proc of Conference on Empirical Methods in Natural Language Processing. 2008:793-801.
- [8] ZHAO Jun, LIU Kang, WANG Gen. Adding redundant features for CRFs-based sentence sentiment classification[C]//Proc of Conference on Empirical Methods in Natural Language Processing. 2008:117-126.
- [9] ZHANG Min, YE Xin-yao. A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval[C]//Proc of the 31st International Conference on Research and Development in Information Retrieval. 2008:411-418.
- [10] LIU Bing. Web data mining: exploring hyperlinks, contents and usage data[M]. New York: Springer, 2007:441-448.
- [11] HU Min-qing, LIU Bing. Mining and summarizing customer reviews[C]//Proc of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2004:168-177.
- [12] DAVE K, LAWRENCE S, PENNOCK D. Mining the peanut gallery: opinion extraction and semantic classification of product reviews[C]//Proc of the 12th International Conference on World Wide Web. New York: ACM Press, 2003:519-528.
- [13] PANG Bo, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques[C]//Proc of Conference on Empirical Methods in Natural Language Processing. 2002:79-86.
- [14] TURNEY P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews[C]//Proc of the 40th Annual Meeting on Association for Computational Linguistics. 2002:417-424.
- [15] 赵妍妍, 秦兵, 刘挺. 文本情感分析综述[J]. 软件学报, 2010, 21(8):1834-1848.
- [16] LAU R Y K, LIAO S S Y, XU Kai-quan. An empirical study of on-line consumer review spam: a design science approach[C]//Proc of the 31st International Conference on Information Systems. 2010:103-123.
- [17] LIU Jing-jing, CAO Yun-bo, LIN C Y, et al. Low-quality product review detection in opinion summarization[C]//Proc of Joint Conference on Empirical Methods in Natural Language and Computational Natural Language Learning. 2007:334-342.
- [18] HAYATI P, POTDAR V, TALEVSKI A, et al. Definition of spam 2.0: new spamming boom[C]//Digital Ecosystems and Technologies. 2010:580-584.
- [19] HAYATI P, POTDAR V. Toward spam 2.0: an evaluation of Web 2.0 anti-spam methods[C]//Proc of the 7th International Conference on Industrial Information. 2009:875-880.
- [20] JINDAL N, LIU Bing. Opinion spam and analysis[C]//Proc of International Conference on Web Search and Web Data Mining. 2008:219-229.
- [21] BECCHETTI L, CASTILLO C, DONATO D, et al. Link analysis for Web spam detection[J]. ACM Trans on the Web, 2008, 2(1):1-42.
- [22] WANG Yi-ming, MA Ming, NIU Yuan, et al. Spam double-funnel: connecting Web spammers with advertisers[C]//Proc of the 16th International Conference on World Wide Web. New York: ACM Press, 2007:291-300.
- [23] FETTE I, SADEH-KONIECPOL N, TOMASIC A. Learning to detect phishing e-mails[C]//Proc of the 16th International Conference on World Wide Web. New York: ACM Press, 2007:649-656.
- [24] WU Bao-ning, GOEL V, DAVISON B D. Topical TrustRank: using topicality to combat Web spam[C]//Proc of the 15th International Conference on World Wide Web. New York: ACM Press, 2006:63-72.
- [25] CASTILLO C, DONATO D, BECCHETTI L, et al. A reference col-

- lection for Web spam[J]. *ACM SIGIR Forum*, 2006, 40(2): 11-24.
- [26] LI Wen-bin, ZHONG Ning, LIU Chun-nian. Combining multiple email filters based on multivariate statistical analysis[C]//Proc of the 16th International Symposium on Foundations of Intelligent System. 2006:729-738.
- [27] WU Bao-ning, DAVISON B D. Identifying link farm spam pages [C]//Proc of the 14th International Conference on World Wide Web. New York: ACM Press, 2005:820-829.
- [28] BAEZA-YATES R, CASTILLO C, LOPEZ V. PageRank increase under different collusion topologies[C]//Proc of the 1st International Workshop on Adversarial Information Retrieval on the Web. 2005.
- [29] FETTERLY D, MANASSE M, NAJORK M. Detecting phrase level duplication on the World Wide Web[C]//Proc of the 28th International Conference on Research and Development in Information Retrieval. 2005:170-177.
- [30] CHIRITA P A, DIEDERICH J, NEJDL W. MailRank: using ranking for spam detection[C]//Proc of the 14th ACM International Conference on Informational and Knowledge Management. New York: ACM Press, 2005:373-380.
- [31] COURNANE A, HUNT R. An analysis of the tools used for the generation and prevention of spam[J]. *Computers & Security*, 2004, 23(2):154-166.
- [32] GYONGYI Z, GARCIA-MOLINA H. Web spam taxonomy [R]. Stanford: Stanford University, 2004.
- [33] SAHAMI M, DUMAIS S, HECKERMAN D, *et al.* A Bayesian approach to filtering junk e-mail [EB/OL]. 1998-07. <http://www.scotnpati.com/UNL/Csce810/sahami98bayesian.pdf>.
- [34] LIM E P, NGUYEN V A, JINDAL N, *et al.* Detecting product review spammers using rating behaviors[C]//Proc of the 19th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2010:939-948.
- [35] LIU Wei, TAN Song-bo, XU Hong-bo, *et al.* Splog filtering based on writing consistency[C]//Proc of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Washington DC: IEEE Computer Society, 2008:227-233.
- [36] TAKEDA T, TAKASU A. A splog filtering method based on string copy detection[C]//Proc of IEEE International Conference on Applications of Digital Information and Web Technologies. 2008:543-548.
- [37] SCULLEY D, WACHMAN G M. Relaxed online SVMs for spam filtering[C]//Proc of the 30th International Conference on Research and Development in Information Retrieval. 2007:415-422.
- [38] JINDAL N, LIU Bing. Analyzing and detecting review spam[C]//Proc of the 17th International Conference on Data Mining. 2007:547-552.
- [39] HAN S, AHN Y, MOON S, *et al.* Collaborative blog spam filtering using adaptive percolation search[C]//Proc of the 15th International Conference on World Wide Web. New York: ACM Press, 2006.
- [40] KOLARI P, FININ T, JOSHI A. SVMs for the blogosphere: blog identification and splog detection[C]//Proc of AAAI Spring Symposium on Computational Approaches to Analysing Weblogs. 2006.
- [41] KOLARI P, JAVA A, FININ T, *et al.* Blog track open task: spam blog classification[C]//TREC Blog Track Notebook. 2006.
- [42] KOLARI P, JAVA A, FININ T, *et al.* Detecting spam blogs: a machine learning approach[C]//Proc of the 21st National Conference on Artificial Intelligence. 2006:1351-1356.
- [43] <http://blog.akismet.com/2010/11/23/20-billion-served/> [EB/OL].
- [44] RIDZUAN F, POTDAR V, TALEVSKI A. Key parameters in identifying cost of Spam 2.0[C]//Proc of the 24th IEEE International Conference on Advanced Information Networking and Applications. Washington DC: IEEE Computer Society, 2010:789-796.
- [45] 刁宇峰, 杨亮, 林鸿飞. 基于 LDA 模型的博客垃圾评论发现[J]. *中文信息学报*, 2011, 25(1): 41-47.
- [46] 何苑, 谭红叶. 基于多结构特征的垃圾博客识别研究[J]. *计算机工程与设计*, 2010, 31(22): 4932-4935.
- [47] 刘伟, 廖祥文, 许洪波. 基于统计特征的垃圾博客过滤[J]. *中文信息学报*, 2008, 22(6): 86-91.
- [48] 孙峰. 基于评论分析的 Blog 观点提取技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2008.
- [49] McFEDRIES P. Technically speaking: slicing the ham from the spam [J]. *IEEE Spectrum*, 2004, 41(4): 72-72.
- [50] MISHNE G, CARMEL D, LEMPEL R. Blocking blog spam with language model disagreement [C]//Proc of the 1st International Workshop on Adversarial Information Retrieval on the Web. 2005.
- [51] NARISAWA K, YAMADA Y, IKEDA D, *et al.* Detecting blog spams using the vocabulary size of all substrings in their copies[C]//Proc of the 3rd Annual Workshop on Weblogging Ecosystem. 2006.
- [52] 何海江, 凌云. 由向量空间相关模型识别博客文章的垃圾评论[J]. *长沙大学学报*, 2008, 22(2): 63-66.
- [53] BHATTARAI A, RUS V, DASGUPTA D. Characterizing comment spam in the blogosphere through content analysis [EB/OL]. 2009-05-14. https://www.c3e.info/uploaded_docs/blogspam_ieeesci09.pdf.
- [54] 邓冰娜, 王煜, 刘宇. 一种应用于博客的垃圾评论识别方法[J]. *郑州大学学报: 理学版*, 2011, 43(1): 65-74.
- [55] KIM S M, PANTEL P, CHKLOVSKI T, *et al.* Automatically assessing review helpfulness[C]//Proc of Conference on Empirical Methods in Natural Language Processing. 2006:423-430.
- [56] ZHANG Zhu, VARADARAJAN B. Utility scoring of product reviews [C]//Proc of the 15th ACM International Conference on Informational and Knowledge Management. New York: ACM Press, 2006:51-57.
- [57] JINDAL N, LIU Bing, LIM E P. Finding unusual review patterns using unexpected rules[C]//Proc of the 19th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2010:1549-1552.
- [58] ZINMAN A, DONATH J. Is britney spears spam[C]//Proc of the 4th Conference on Email and Anti-Spam Mountain View. 2007.
- [59] MOBASHER B, BURKE R, SANDVIG J J. Model-based collaborative filtering as a defense against profile injection attacks[C]//Proc of the 21st National Conference on Artificial Intelligence. 2006:1388-1393.
- [60] BENEVENUTO F, RODRIGUES T, ALMEIDA V, *et al.* Identifying video spammers in online social networks[C]//Proc of the 4th International Conference on Adversarial Information Retrieval on the Web. New York: ACM Press, 2008:45-53.
- [61] HAYATI P, CHAI K, POTDAR V, *et al.* Behaviour-based Web spambot detection by utilising action time and action frequency[C]//Proc of International Conference on Computational Science and Applications. 2010:351-360.
- [62] HAYATI P, POTDAR V, CHAI K, *et al.* Web spambot detection based on Web navigation behaviour[C]//Proc of the 24th IEEE International Conference on Advanced Information Networking and Applications. Washington DC: IEEE Computer Society, 2010:797-803.
- [63] HAYATI P, POTDAR V, TALEVSKI A, *et al.* Rule-based on-the-fly Web spambot detection using action strings[C]//Proc of the 7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference. 2010:3.
- [64] POPESCU A M, ETZIONI O. Extracting product features and opinions from reviews[C]//Proc of Conference on Empirical Methods in Natural Language Processing. 2005:339-346.