

基于主题模型和情感分析的垃圾评论识别方法研究

金相宏 李琳 钟珞

(武汉理工大学计算机科学与技术学院 武汉 430070)

摘要 随着电子商务的飞速发展,网络购物越来越被消费者认同,而随之产生的产品评论给消费者的购买决策带来了影响。产品评论是指用户在购物站点上对商品的评价信息,而经过分析和研究发现这些评论中充斥着大量的垃圾评论,因此垃圾评论的识别成了电子商务在提高服务质量的过程中需解决的重要问题之一。根据垃圾评论的主要特点提出 LDA-SP(LDA-Sentiment Polarity)垃圾评论识别方法。首先利用 LDA 主题模型过滤出内容型垃圾评论,然后结合情感分析识别出欺骗型垃圾评论。对网络商城的大量评论数据进行准确度分析实验的结果表明,LDA-SP 方法的识别准确度高于传统的 LDA 主题模型和单一的情感极性分析方法,能够有效地检测垃圾评论,从而使产品评论信息更加客观准确,为电子商务用户提供了有效的参考信息。

关键词 产品评论,垃圾评论,主题模型,情感分析

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.10.046

Review Spam Detection Approach Based on Topic Model and Sentiment Analysis

JIN Xiang-hong LI Lin ZHONG Luo

(School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China)

Abstract With the rapid development of e-commerce, consumers have accepted online shopping increasingly, and the product reviews then have an great influence on consumers' purchase decision. Product reviews refer to the evaluation or comment information of items or products written by online shopping users. These comments usually include some review spams that may hurt user shopping experiences. Review spam detection, therefore, becomes one of the important problems to improve service quality. In this paper, a review spam detection approach called LDA-SP(LDA-sentiment polarity) was proposed by carefully analyzing the main characteristic of review spams. First, we used LDA topic model to filter the irrelevant reviews, and then applied sentiment analysis to identify the untruthful reviews. Experiments were conducted on a large number of reviews data on a online shopping mall. Our experimental results show that the detection accuracy of LDA-SP method is higher than that of the traditional LDA topic model and the single sentiment polarity analysis method. It can effectively detect review spams, so that more objective and accurate information about products will be displayed to the users of e-commerce.

Keywords Product reviews, Review spam, Topic model, Sentiment analysis

1 引言

近年来,随着信息技术的不断进步和电子商务的飞速发展,网络购物已经融入社会大众的日常生活中,其不但对消费者的行为模式和消费习惯产生了深刻的影响,也使互联网在该领域的数据分析具有巨大的研究价值。Double Click Inc.^[1]曾在美国进行一项对网络用户的调查发现,大部分消费者在购买这些行业的产品之前,都会在互联网上先搜索相关产品的介绍以及查看其他已购买用户的评论信息,这些信息表达着用户的情感观点。因此,用户评论数据对消费者的购买决策有着重要的影响,挖掘用户的评论信息无论是对商

家还是消费者都具有重要的意义。但是在网络商品交易领域中只是简单地挖掘用户评论数据显然是不够的,互联网中的大量情感观点所隐藏的价值引起了人们的高度关注。评论的褒贬直接影响着产品的销售业绩,因此催生了大量的垃圾评论。这些垃圾评论的发布者以盈利为目的,利用欺骗性的评论企图误导潜在客户,并且他们还学习如何躲避垃圾评论识别技术的追踪^[2]。何海红、孙升芸等人^[3]对垃圾评论识别做了相关研究,其主要关注的是购物网站中的垃圾评论,这种评论被用来误导消费者的购买决策。Gilbert 等人^[4]的研究发现 10%~15%的评论本质上是受早期评论所影响的,并且很可能被垃圾评论所影响,甚至在评论内容上某些垃圾评论和

到稿日期:2016-08-05 返修日期:2016-12-11 本文受国家自然科学基金(15BGL048),国家 863 计划项目(2015AA015403)资助。

金相宏(1991-),男,硕士生,主要研究方向为自然语言处理,E-mail:339449572@qq.com;李琳(1977-),女,博士,教授,主要研究方向为自然语言处理、信息检索、推荐系统,E-mail:cathylilin@whut.edu.cn(通信作者);钟珞(1957-),男,博士,教授,主要研究方向为智能方法。

非垃圾评论并无区别。

Jindal^[5]把垃圾评论分为 3 类:1) 欺骗性评论^[6],指发布虚假评论来蓄意诋毁或者吹捧产品以误导消费者和观点挖掘系统,因此也可称欺骗性评论为虚假评论;2) 不相关评论,指某些评论仅仅只评论了产品的品牌、销售商或者厂家等信息,尽管从某些方面看,这些评论仍有作用,但并非是针对某一特定产品,并且带有歧视;3) 非评论信息,指某些评论不是评论信息,主要是一些链接、随机文字和广告等无用的非观点信息。本文主要根据 Liu 等人^[7]对垃圾评论的阐述将其分为两类:欺骗型垃圾评论和内容型垃圾评论,欺骗型垃圾评论指上述第一类评论,内容型垃圾评论指上述第二类和第三类垃圾评论,这类垃圾评论的识别难度较大,其主要目的是在消费者心中获取不可信评分,以影响消费者的选择^[8]。

本文对大量的产品评论进行分析,结合 LDA 主题模型和情感极性,设计出 LDA-SP(LDA-Sentiment Polarity)垃圾评论识别方法。近年来,LDA 模型、LDA 的扩展模型以及它们在自然语言和智能信息处理中的应用得到了充分的重视和深入的研究,但它们较少用于垃圾评论的识别,Diao 等人^[9]利用 LDA 模型来识别垃圾评论并取得了不错的效果。由于 LDA 可以挖掘隐含主题,可以很好地识别内容型垃圾评论,但对欺骗型评论的识别能力较差。因此,本文引入情感分析的处理方法,分析每条评论的情感值,检测有正负异常情感值的评论并判断是否为垃圾评论。实验结果表明,所提方法对垃圾评论的识别率有一定的提高。

2 相关工作

相较于垃圾网页识别和垃圾邮件识别,垃圾评论识别的研究相对较晚,当前垃圾评论识别的主要技术是分类算法。Diao 等人^[9]提出了一种基于 LDA 的博客垃圾评论识别模型,该模型使用 LDA 获取评论内容主题特征,采用支持向量机分类模型判断垃圾评论;Lai 等人^[10]使用语言模型对评论进行建模,最后通过 Kullback-leibler Divergence 和语言模型对垃圾评论进行建模;而 Jin 等人^[11]则使用半监督的 Co-training 对产品的评论质量进行分析检测,在评论质量的分析中除二元分类模型以外,也可以对评论内容进行评分和排序,将排序靠后的评论视为低质量评论。

目前垃圾评论识别的主流方法是分析评论内容特征和评论发布者的行为特征,然后通过分类模型实现对垃圾评论的检测。本文主要研究评论内容特征,挖掘评论内容中的主题信息和情感信息来建立一套识别方法。本文将收集到的产品的大量非垃圾评论看作一个长文本,依词频抽取主题词,在除去无用的停用词以及无关该产品的情感词后构建语料库,并与评论的主题相匹配。另外,本文使用的是中科院的分词工具 ICTCLAS^[12]以及大连理工大学信息检索实验室的情感词汇本体^[13]。由于某些字符较少的短文本(比如“真的很不错”、“还可以吧”)没有隐含主题,在收集的所有产品评论中出现的次数较少,但很多类似句子并非垃圾评论,因此利用情感词汇本体来进行预处理,保留这些包含情感词的短文本,在做情感分析之前视其为非垃圾评论。

3 LDA-SP 方法

本文主要考虑用户对产品进行评论的文本信息,对于用户的其他行为以及图像识别将不予考虑。

3.1 方法概述

本文提出的 LDA-SP 方法的流程如下:

- (1) 从网络中收集需要的语料库数据集和测试数据集,语料库的数据集不含垃圾评论,用于构建主题模型。
- (2) 对收集到的所有评论去除标点和分词。
- (3) 对收集到的语料库数据集去除停用词并构建相关产品的语料库。
- (4) 对测试数据集中的垃圾评论进行人工鉴定和标记,以满足后期实验的需要。
- (5) 将评论内容长度小于 7 个字符的评论与情感词汇本体匹配,包含情感词的评论保留到步骤(7),其余的则进行下一步。
- (6) 利用 LDA 主题模型对所有产品评论抽取主题,并与语料库进行匹配对比从而判断垃圾评论。
- (7) 对步骤(6)处理后的判断为非垃圾的评论和步骤(5)保留的评论进行情感分析,判断正负异常情感,将正负异常情感评价的评论视为垃圾评论。
- (8) 评估实验结果。

LDA-SP 方法的具体流程如图 1 所示。

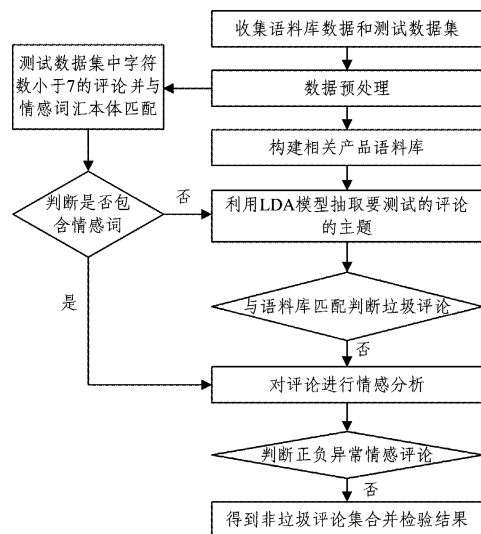


图 1 LDA-SP 方法的流程

3.2 基于 LDA 的内容型垃圾评论识别

LDA(Latent Dirichlet Allocation)^[14]主题模型是一个多层的产生式全概率生成模型,它是对文本数据的主题信息进行建模的一种方法。使用此模型来处理本文问题的主要原因有:1) LDA 作为全概率生成模型有着清晰的内在结构,还可以利用高效的概率推理算法进行计算;2) 自然语言处理的领域依赖性很强。因此,若采用有监督的方法对文本进行标注分析,则要求针对不同领域的文本提供不同人工标注的样例来供计算机学习。而经典的 LDA 模型是无监督的,不需要训练样本,适合处理大规模文本语料。

前文已指出内容型垃圾评论的特点,无论是广告还是随

机文字等无关信息,都与需要的产品主题无关。本文采用概率检索模型来识别这些垃圾评论。将评论识别视为检索问题,设评论 C 为查询串,语料库 L 为该产品相关内容的文档集合,在还未引入主题信息之前,计算 C 属于 L 的概率,建立简单的概率检索模型,计算公式如下:

$$P(C|L) = \prod_{w \in C} P(w|L) \tag{1}$$

其中, C 为一条评论, L 为整个语料库, w 为 C 中的一个词,设 C 中的词相互之间是独立的, $P(C|L)$ 为 L 产生 C 的概率, $P(w|L)$ 为词 w 出现在语料库 L 中的概率。

在上述的检索模型中,加入引入 LDA 模型后发现的隐含主题集合,对 $P(w|L)$ 进行平滑处理。与其他聚类模型不同, LDA 模型不会把评论当作单一主题集合,而是将其看作多个主题的集合,将评论划分为多个主题的集合并以不同比例进行区分,大大地增强了灵活性。本文利用这些隐含的主题信息来共同建立主题检索模型,如式(2)、式(3)所示:

$$P(C|L) = \alpha \prod_{w \in C} P(w|L) + (1-\alpha) \prod_{w \in C} P_1 \tag{2}$$

$$P_1 = \sum_{t \in t_L} \prod_{w \in t} P(w|t) \cdot P(t|L) \tag{3}$$

其中, α 为参数, α 值越小,评论主题的相关权重越大,而评论的全局权重越小。很显然,主题的相关性更能代表评论内容的相关性,但由于网络商城的评论内容通常较短,而 LDA 模型对短文本的处理效果不佳,因此改进策略,引入评论的全局权重,本文 α 取值为 0.3。 t_L 为收集到的语料库 L 的主题集合, t 是 L 中的一个主题。 $P(t|L)$ 表示主题出现在语料库中的概率, $P(w|t)$ 表示词 w 在主题 t 中出现的概率。

对于所有的产品评论 C ,都需要计算 $P(C|L)$ 来建立检索模型,若此值小于某一特定阈值,则视其为垃圾评论,否则不是垃圾评论。

3.3 基于情感分析的欺骗型垃圾评论识别

欺骗型垃圾评论又被称为虚假评论,隐蔽性极高,很难对其进行精准判断,在逻辑上不可能通过简单的阅读来进行辨别。某些虚假评论表面上看几乎不虚假,某人可以基于他真实购买了一个坏手机而做出假正面评价,比如“这手机用起来真心不错!”的简单情感句,根本不可能单从文本信息来判断是否虚假。但是欺骗型垃圾评论也有其特点,即使无法识别出所有隐藏的虚假评论,但也可以部分识别,从而使最后的结果更加精准。

欺骗型垃圾评论主要包括两方面^[15]: 1) 用于推广目标对象的一些不切实际的正面评论; 2) 为诋毁目标对象声誉的一些不公平或者恶意的负面评论。这些虚假评论通常采用感情比较强烈的情感词来影响其他用户,因此情感倾向表现得比较夸张的评论更可能是垃圾评论。基于该特点,对用户评论的情感分析便有了意义,可以从评论内容中找出有别于其他大多数评论的情感,检测出垃圾信息。本文采用 python 的一个类库 SnowNLP,它是中文文本的情感分析工具,训练的数据主要是用户购物后的评价,它可以通过输入的文本得到一个介于 0 到 1 之间的情感值来表示该短文本表达的情感强烈程度,小于 0.5 是负面评价,大于 0.5 是正面评价。情感值越接近 0 说明负面情感越强烈,情感值越接近 1 说明正面情感越

强烈。因此我们需要计算每条评论的情感值,进而判断出正负异常情感的评价。首先需要计算评论的情感平均值,为了减小垃圾评论对该值的影响,先将所有评论按情感值降序排列,然后去除情感值极大和极小的评论使之不参与计算,主要公式如下:

$$p = \frac{\sum_{i \in V} u_i}{(1-\lambda) \cdot n} \tag{4}$$

其中, p 为所求的情感平均值, n 为所有评论的总数, λ 为不参与计算的评论的比例, V 为去除情感极性评论的集合, u_i 为第 i 个评论的情感值。在不同环境下,欺骗型垃圾评论的占比不同,因此没有确切的值。通常为了避免计算时丢失过多的真实评论,应使 λ 的值尽可能小,本文中 λ 取 0.1。

将得到的 p 值与 u_i 进行比较,得到 s_i , s_i 表示第 i 条评论的情感值与大部分非虚假评论的平均情感值的距离,以 s_i 来判断对应评论是否为正负异常评价的评论,显然, s_i 越大,说明该评论的情感值极性越高,越有可能是垃圾评论。 s_i 的计算公式如下:

$$s_i = \frac{|u_i - p|}{p} \tag{5}$$

所有评论计算完毕之后, s_i 的值会落在一个区间内,对其进行降序排列,排名越靠前的评论表示其与真实相关评论的相似度越低,因此设置阈值 θ ,将排名靠前的 $n \cdot \theta$ 条评论视为垃圾评论。通过实验测试,在保证准确率的前提下,阈值越小,检测的准确度越高,但召回率越低,因此以 F 值为参考标准,得出阈值取 0.15。

4 实验

4.1 实验数据

本文所有实验数据都来自淘宝网络商城(<http://www.taobao.com>)的用户评论,产品对象为手机、相机以及电脑笔记本,分别从 25209, 24535 和 23100 条面向多个品牌产品的非垃圾评论中抽取主题词构建相关的语料库。另外,再从网络商城中分别各自收集 3518 条相关产品的评论并进行标注作为测试集来检验方法的有效性,其中每个产品对象都包括 2518 条非垃圾评论以及 1000 条垃圾评论。

4.2 实验结果与分析

在根据词频构建相关语料库时,抽取收集的相关产品的评论数据集中的主题词,表 1 列出了手机的主题词。

表 1 主题词列表

主题词	词频
手机	28172
屏幕	18167
电池	15011
系统	10150
软件	9239
游戏	8090
外观	7665
...	...
闪光灯	1018

本文使用的停用词表为百度停用词表、哈工大停用词表、四川大学机器智能实验室停用词库。

在将评论的主题与语料库进行匹配时,经实验分析,构建语料库时选取的主题词数目对结果存在一定的影响,如图 2 所示。

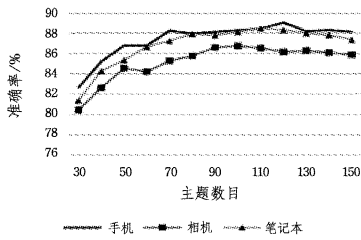


图 2 选取的主题数与准确率的关系

从图 2 可以看出,随着选取的主题数目的增加,LDA 模型检测的准确率有小幅增加进而波动,但波动不大,面向手机、相机、笔记本 3 个产品,本文最终选取词频排名前 120, 100 以及 110 的主题词来构建相关产品的语料库,此时建立的 LDA 模型的匹配性能最佳。同时根据产品类别也可以适当地导入主题词,比如手机产品就可以导入手机品牌的名词“黑莓”、“波导”等,因为它们属于手机的小众品牌,不能以高词频出现,但与主题十分相关。

使用 LDA-SP 方法处理所有手机评论的对比实验结果如图 3 所示。

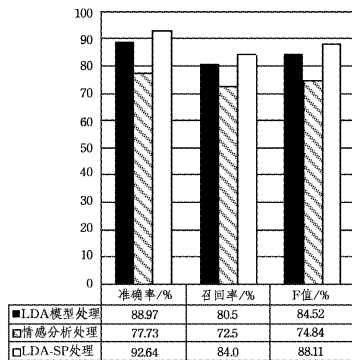


图 3 实验结果 1

对手机、相机、电脑笔记本 3 种产品采用不同垃圾评论检测方法时 F 值关系的实验结果如图 4 所示。

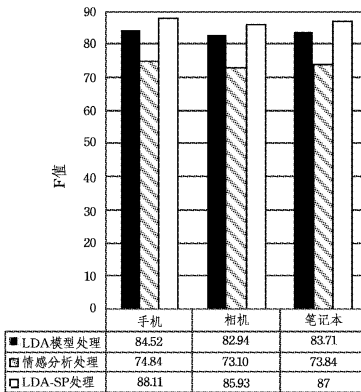


图 4 实验结果 2

本文主要考虑信息检索的 3 个重要指标:准确率(Precision)、召回率(Recall)、F 值。从图 3 中可以清晰地看出,单使用 LDA 模型可以识别出大部分垃圾评论,且情感分析方法显然不如 LDA 的识别效果好,因为欺骗型评论比内容型评论更

难以识别,它的隐蔽性更高。而由图 4 可以看出,即使面对不同的数据集,根据 F 值进行判断,LDA-SP 的检测效果依然优于其他两者。

实验中使用 LDA-SP 处理后的各项指标均高于其他两者,主要原因在于 LDA 模型对欺骗型垃圾评论的识别效果不佳。实验中某条评论内容为“屏幕效果不错,但待机时间是个问题”,经过与语料库的匹配发现,主题信息与我们所需要的一致且大量出现,是一条非垃圾评论;另一条评论为“今天天气不错,适合到处逛逛。”,该评论的主题明显与语料库的主题不一致且出现频率极低,则视其为垃圾评论。再比如“华为手机真是太烂了,大家不要上当,根本不能跟三星比,三星惊艳的屏幕和华丽的外观,才是安卓机最强品牌。”这条评论用夸张的词语明显地诋毁某品牌,推广了另一个品牌,内容过于主观,缺少客观性,对用户而言,虽然评论内容与主题相关,但毫无参考作用。情感分析处理方法作为主题模型处理的补充,处理对象为主观性文本,找出情感倾向特别夸张且有别于其他评论的正负异常情感评论。

结束语 本文主要分析两类垃圾评论:内容型垃圾评论和欺骗型垃圾评论的特点,提出 LDA-SP 方法,利用主题词的匹配来识别内容型垃圾评论,利用情感分析方法识别欺骗型垃圾评论。由两个实验结果的各项指标可知,该方法能够有效地识别垃圾评论。由于网络商场独特的开放性,用户可以自由地发表任何言论,其评论内容十分丰富并且隐含大量的垃圾评论,某些隐蔽性极高的垃圾评论几乎不能被识别,仍可能误导消费者做出错误的判断。本文的处理方法虽从结果上分析行之有效,但在判断正负异常情感评论时,处理量的多少仍很难精确把握。针对不同的网络商城和产品,此类垃圾评论在整个评论中的占比不同,而且某些真实评论的情感表达也可能因夸张而导致误判。自然语言处理工具在计算情感值时会存在一定偏差,因此在取阈值时会不可避免地丢失非垃圾评论。本文并未考虑全面欺骗型垃圾评论的特点,所选取的特征并不能识别所有欺骗型垃圾评论。此外,本文的某些语料由笔者手工收集,语料中内容的丰富度和校验工作还需进一步进行。以上情况都有待做进一步细致的研究。

参考文献

[1] DoubleClick Search before the purchase Understanding Buyer Search Activity as it Builds to Online purchase[OL]. [http://www. Doubleclick. com /insight/pdfs/searchpurchase _ 0502. pdf.](http://www.Doubleclick.com/insight/pdfs/searchpurchase_0502.pdf)

[2] HEYDARI A, ALITAVAKOLI M, SALIM N, et al. Detection of review spam; A survey[J]. Expert Systems with Applications, 2015, 42(7):3634-3642.

[3] SUN S Y, TIAN X. Product review comment spam detection research[J]. Computer Science, 2011, 38(10): 198-201. (in Chinese)

孙升芸,田萱. 产品垃圾评论检测研究综述[J]. 计算机科学, 2011, 38(10):198-201.

[4] GILBERT E, KARAHALIOS K. Understanding Deja Reviewers [C]//Proc. of ACM Conference on Computer Supported Cooperative Work. New York, USA, 2010:225-228.

- [5] JINDAL N, LIU B. Opinion spam and analysis[C]//International Conference on Web Search and Data Mining. ACM, 2008; 219-230.
- [6] OTT M, CHOI Y J, CARDIE C, et al. Finding Deceptive Opinion Spam by Any Stretch of the Imagination[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011; 309-319.
- [7] LIU B. Sentiment Analysis and Opinion Mining[M]. Chicago: Morgan & Claypool, 2012; 113-115.
- [8] MA Y, LI F. Detecting review spam: Challenges and opportunities[C]//2012 8th International Conference on Collaborative Computing, Networking, Applications and Worksharing (CollaborateCom). IEEE, 2012; 651-654.
- [9] DIAO Y F, LIN H F. LDA-based Opinion Spam Discovering [J]. Journal of Chinese Information Processing, 2011, 25(1): 41-47. (in Chinese)
刁宇峰, 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.
- [10] LAI C L, XU K Q, LAU R Y K, et al. Toward a language modeling approach for consumer review spam detection[C]//2010 IEEE 7th International Conference on e-Business Engineering (ICEBE). IEEE, 2010; 1-8.
- [11] JIN J, JI P. Co-training Algorithm for Quality Analysis of Online Customer Reviews[J]. Journal of Shanghai University (Natural Science Edition), 2014, 20(3): 289-295. (in Chinese)
靳健, 季平. 用于在线产品评论质量分析的 Co-training 算法[J]. 上海大学学报(自然科学版), 2014, 20(3): 289-295.
- [12] 中科院分词系统[DB/OL]. <http://ictclas.org>.
- [13] XU L H, LIN H F, PAN Y, et al. The structure of the emotional vocabulary ontology[J]. Journal of Emotion, 2008, 27(2): 180-185. (in Chinese)
徐琳宏, 林鸿飞, 潘宇, 等. 情感词汇本体的构造[J]. 情感学报, 2008, 27(2): 180-185.
- [14] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003(3): 993-1022.
- [15] QIU Y F, WANG J K, SHAO L S. Research on Product Review Spammer Detection Based on Users' Behavior [J]. Computer Engineering, 2012, 38(11): 254-257. (in Chinese)
邱云飞, 王建坤, 邵良杉. 基于用户行为的产品垃圾评论者检测研究[J]. 计算机工程, 2012, 38(11): 254-257.

(上接第 236 页)

当搜索引擎的数目较多时, 如何高效地选出最适合的搜索引擎依然是一个挑战性的问题。目前人工神经网络已经被应用于引擎选择中, 并已被证明具有有效性^[15]。更进一步, 我们将考虑不同的启发式方法和神经网络在引擎选择中的结合使用。

参 考 文 献

- [1] MENG W Y, YU C, LIU K L. Building efficient and effective meta-search engines [J]. ACM Computing Surveys (CSUR), 2002, 34(1): 48-89.
- [2] XUE Y, SHEN X P, CHEN J B. Research on an Algorithm of Metasearch Engine Based on Personalized Demand of Users [C]//2010 International Forum on Information Technology and Applications (IFITA). IEEE, Kunming, China, 2010; 240-243.
- [3] SUN Y C, LI Q S. The research situation and prospect analysis of meta-search engines[C]//2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering (URKE). IEEE, Bali, Indonesia, 2012; 224-229.
- [4] CALLAN J P, LU Z H, CROFT W B. Searching distributed collections with inference networks[C]//Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, Massachusetts, USA, 1995; 21-28.
- [5] SI L, CALLAN J. Relevant document distribution estimation method for resource selection[C]//Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, Toronto, Canada, 2003; 298-305.
- [6] CETINTAS S, SI L, HAO Y. Learning from past queries for resource selection[C]//Proceedings of the 18th ACM Conference on Information and Knowledge Management. ACM, Hong Kong, China, 2009; 1867-1870.
- [7] RAVI S, GANESAN N, RAJU V. Search Engines Using Evolutionary Algorithms[J]. International Journal of Communication Network Security, 2012, 4(1): 39-44.
- [8] KUMAR R, SINGH S K, KUMAR V. A heuristic approach for search engine selection in meta-search engine[C]//2015 International Conference on Computing, Communication & Automation (ICCCA). IEEE, Uttar Pradesh, India, 2015; 865-869.
- [9] SHEO D, KULDEEP S R. Search Engine Selection Approach in Metasearch Using Past Queries[J]. Oriental Journal of Computer Science & Technology, 2014, 3(23): 177-183.
- [10] SI L, CALLAN J. A semisupervised learning method to merge search engine results [J]. ACM Transactions on Information Systems (TOIS), 2003, 21(4): 457-491.
- [11] KUMAR R, GIRI A K. Learning based approach for search engine selection in meta-search engine[J]. International Journal of Engineering and Management Research, 2013, 10(3): 82-88.
- [12] XU Z Y, WANG X X. A predictive modified round robin scheduling algorithm for web server clusters[C]//2015 34th Chinese Control Conference (CCC). IEEE, Hangzhou, China, 2015; 5804-5808.
- [13] DEB K, GOYAL M. A combined genetic adaptive search (GenAS) for engineering design[J]. Computer Science and Informatics, 1996, 26(16): 30-45.
- [14] GORDON M. Probabilistic and genetic algorithms in document retrieval[J]. Communications of the ACM, 1988, 31(10): 1208-1218.
- [15] GOEL P, JAIN T, BHATIA M P S. Learning from training query in Meta search using Artificial neural network[C]//2015 Annual IEEE India Conference (INDICON). IEEE, Delhi, India, 2015; 1-6.