

# 基于网络信息抽取的快递服务质量研究

李洪仪

兰州财经大学经济学院

**[摘要]** 当今每天都有海量数据存储于互联网,如此庞大的数据为各类服务行业了解用户对服务质量的评价提供了前提条件。随着我国快递行业飞速的发展,提高快递服务质量将是各快递企业必然的选择。利用网络评价信息抽取科学合理的关键词对快递服务质量进行评价,为快递企业的服务提供有效的改善意见就成为一个新的研究方向。本文从国内外现有的快递服务质量评价体系出发,指出了目前的评价方法大多是基于传统调查问卷的方式获取顾客对快递服务的评价内容对不同快递企业的服务质量评价。这种方式就目前看来,调查周期长、消耗人力物力、费用高。因此,本文提出了一种基于网络信息的抽取方法,得出反映快递服务质量权重和得分及相关结论。

**[关键词]** 网络爬虫; 快递服务质量; 自然语言技术

**[DOI]** 10.12252/j.issn.2096-627X.2021.09.888

## 一、概述

### (一) 研究背景及意义

结合2018年11月至2019年2月的报告不难发现,快递行业正逐步向着一些头部企业聚合,导致很多中小快递企业不得不在夹缝中求得一线生机。当前,在我国的快递行业市场主要有三个比较大的竞争主体,它们分别是国有快递企业,比如中国邮政;国外快递企业,比如联邦快递;民营快递企业,如申通、圆通、韵达等。其中国有快递企业市场份额正在快速下降,但在政策方面其具备较大优势。外资快递在我国的快递市场发展缓慢,民营快递企业近年来发展势头非常迅猛,它们之间的竞争变得愈发激烈。

随着网络发展,网购成为人们日常生活的重要构成,这就导致快递行业发展飞速,行业间的竞争趋于白热化。在此背景下,一些快递企业的服务并没有因时发展,逐渐难以满足消费者需求,甚至在服务中会出现物品丢失、损坏、速度慢、态度差等问题,这对消费者的合法权益造成了不小的侵害。<sup>[1]</sup>

对于提出的这些问题,选取顾客对快递服务质量评价这一方面进行研究。研究目的是通过一系列分析,获取快递企业的服务质量总体得分的排序以及通过对某家快递的评价指标排序得出亟待改善的服务衡量指标。互联网资源的不断增加的同时,也给广大用户带来了困扰<sup>[2]</sup>。本文具有以下现实意义:在网络评价信息中,评论是对快递服务进行好坏判断的重要标准,因而,快速获取大量快递服务评价并对其进行分析是有意义的;传统的调查问卷形式存在很大弊端,比如发放困难、信息处理效率低等,这就很难满足相应的服务调查需求。另外,若想提升快递服务质量,需要重视对服务信息的检索、分类以及抽取,只有对相应的服务信息准确、快速展开分析,方可更好地了解用户的需求,进而提升之后的快递服务水平。

### (二) 研究综述

近年来,基于信息抽取技术的探究通常以信息抽取、多语言文本处理、篇章分析与推理以及自然语言处理技术等层面为主。实际上,对于英文文本信息的抽取技术至今已经获得较高水平发展,查全率、查准率和指数等指标都有了较大程度的提高。

现阶段,评价方式多是以既往的调查问卷形式展开,以此取得顾客对相应快递企业服务的意见,而后以此为依据展开快递服务质量的相应评价工作。这样的形式在以往互联网不健全时尚可获得不错效果,具备一定优势。但是,如今面对海量顾客发布于网络的评价内容时,调查问卷依旧基于人工收集分析,没有采用新的技术,就显得束手无策。这就更加证明了本文研究基于网络信息抽取的快递服务质量评价,达到有针对性的提升快递服务质量是有意义的。对快递服务质量有针对性的提升,才能够还在各个快递企业的竞争中站稳脚跟。

对百度口碑中文文本内容进行深入分析理解,结合我国现

有快递服务质量评价指标。从自然语言处理技术着手,处理评价内容,通过对快递服务质量词频高的关键词提取,并匹配各项评价指标对应的评价情感得分,把不少的中文评价内容简化为评估标准所占比重,以方便快递服务企业直观的认识服务过程中的优势和劣势。

互联网资源的不断增加在满足人们多样需求的同时,也给广大用户带来了更大的困扰。随着近年来网络的高速发展,人类的信息出现雪崩式的增长。从快递服务质量评价这一方面入手,本文具有以下现实意义:1.在网络评价信息中,评论是对快递服务进行好坏判断的重要标准,因而,快速获取大量快递服务评价并对其进行分析是有意义的。2.以传统的问卷调查发放回收、人工处理信息的方法已经远远无法适应人类对于信息的需求。信息抽取、信息分类、信息检索问题也越来越受到人们的重视。3.从这信息的海洋中快速、准确地为用户提供其所感兴趣的信息,目前已经成为亟待解决的问题。

## 二、信息抽取及数据分析

面对大量冗余的信息,对于顾客来说处理不了,但是对于网络爬虫来说又是一大优势。现在是大数据时代,数据分析是处理相应问题不可或缺的关键依据。数据分析结果的准确性有很大一部分取决于数据量是否足够大。所以我们需要的数据量通常比较大,一般不可能借助人力实现数据采集的相应工作。这时候大量的数据对于网络爬虫信息抽取来说,有助于增加数据分析结果的准确性。

### (一) 数据的内容分析

本文借鉴了SERVPERF模型的5个维度22个指标<sup>[13]</sup>,结合本文抽取的关键词,指出影响跨地服务质量评价的13个评价标准以及较为重要的7个维度。

$weight(focus_i)$  是评价指标  $focus_i$  的权重;

$n(focus_i)$  是评价指标  $focus_i$  在语料中出现的频数;

$n(count\_list)$  是语料中词频的总和;

$sentence_{ni}$  是根据评价指标  $focus_i$  分组的  $n$  条语句;

$sensibility(sentence_{ni}, focus_i)$  是对根据  $focus_i$  分组后的每一条语句计算情感得分  $x_i$  是评价指标  $focus_i$  的情感得分均值;

权重表示了各个评价标准受消费者的重视程度以及对快递服务质量的影响力度。本文利用评价指标的频度计算权重。频度权重的思路是评价指标在文本中显示的频次越多,则权重越大<sup>[14]</sup>。权重的计算公式为

$$weight(focus_i) = \frac{n(focus_i)}{n(count\_list)} \quad (1)$$

实施情感判断主要是采用对信息语义的类似程度展开判别,其中李钝<sup>[15]</sup>的情感值计算方法,使用的是褒贬义词为基准对相应的评价标准展开计算,以此获取情感值。最后,由计算获得相应的情感值平均数。情感得分均值的计算式为

$$x_i = \frac{\sum sensibility(sentence_m, focus_i)}{n} \quad (2)$$

计算权重和情感得分均值是为了得出快递企业总体得分，能够更为直接地体现出各快递企业的排名。总体得分计算式为

$$Total = \sum [x_i \times weight(focus_i)] \quad (3)$$

(二) 快递服务质量的评分结果

根据式(3)，计算快递企业的服务质量的总体得分如表1总体得分所示

表1 总体得分

	$focus_i$	Total	$T_i(sf)$	$T_i(tt)$
1	速度	0.026657541	0.035401404	0.019863392
2	服务态度	0.010082469	0.028705286	0.004636712
3	规范性	0.004035727	0.002346078	0.00796746
4	货物完好程度	0.011055769	0.021724992	0.006661199
5	丢件少件	0.037965927	0.072864194	0.016656507
6	客服态度	0.075088524	0.262760215	0.00116303
7	客服可得性	0.056637186	0.067651058	0.040588132
8	投诉处理有效性	0.013620464	0.033146814	0.000449424
9	智能终端设备	0.015368266	0.040663192	0.002259577
10	反馈交互性	0.016496962	0.022466067	0.046775989
11	价格水平	0.062366411	0.120912428	0.035391043
12	派送范围	0.010656413	0.013023011	0.000170727
13	国民形象	0.048964038	0.140671171	0.005274641
	总体得分:	0.388995696	0.862335908	0.187857834

利用本文的算法最终得到的总体得分和百度口碑好评率百分比对照差异如图 1 总体评分与百度口碑差距所示

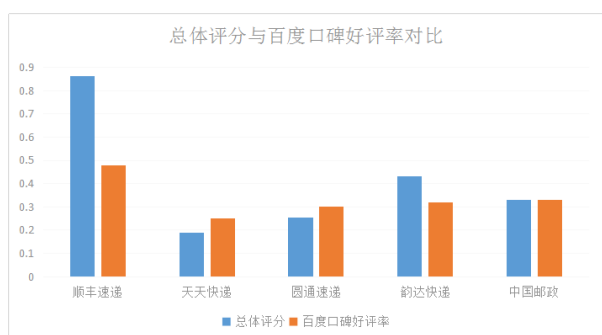


图 1 总体评分与百度口碑好评率对比

以上的计算结果表明:本文提出的基于网络信息抽取情感分析的快递企业服务质量评价与百度口碑的星级评分，除顺丰速递偏差过大之外，趋势大体一致。

### 三、总结与展望

不管是采用传统方式调查还是网络信息抽取，得出的结论在准确度上取决于样本的容量，对于本文来讲语料的取样空间不仅限于百度口碑上评价信息，而是整个互联网上的信息。随着现阶段我国经济不断稳步发展，人们的收入及生活水平大幅提升，供给侧的结构也获得了持续优化，在“稳出口+扩内需”的政策支持下，我国电商经济获得了更为长效的发展，快递产业也迎来了新的风口。我国快递行业在经过了十几年的爆炸式发展后，已逐渐停止了野蛮生长，正面临着一个新的洗牌

期。中小体量的快递企业的生存空间被不断蚕食，行业巨头在市场上的发言权变得更大。在快递行业快速发展的新时代，对行业巨头和中小快递企业都是一个挑战。想要在竞争空前激烈的时代下站稳脚跟，快递行业都要积极改善，甩掉人们对其“傻大黑粗”的印象，变得越来越“细”。

基于此结果，本文还研读了快递企业物流服务质量评价研究中使用传统调查方式的过程。传统调查方式存在的优势是：1. 问题设置明确，便于最终结论的判断；2. 获取准确的被调查者信息，包括年龄、性别、使用快递的频率等；存在的问题有：1. 调查周期长，此篇论文调查周期前后共计为4个月；2. 有效问卷数量低，此篇论文总共发放390份调查问卷，收回有效问卷263份。

对于传统调查方式的优势，通过构建针对快递评价的分词数据字典，可以获得接近于调查问卷中明确的题设以及对对应情感词和程度副词；对于被调查者的更多信息，可以爬取网页上每一个顾客id的更多属性或参考更精确的大数据研究方法。

不论本文的基于网络信息抽取技术还是传统调查方式，得出的结论在准确度上很大一部分取决于样本的容量，对于本文来讲语料的取样空间不仅限于百度口碑上的24, 237, 225条评价信息，而是整个互联网上的有效信息，这点传统调查方式的样本容量是达不到的。

### 参考文献:

[1] 李仲勇. 对我国物流业“灰犀牛”现象加强科学化管理的思考[J]. 物流科技, 2017, (12): 47-49.  
 [2] 孙骏雄. 基于网络爬虫的网站信息采集技术研究[D]. 大连海事大学, 2014.  
 [3] Aggarwal C, Yu P S, Watson I T J, et al. Yu P: Intelligent Crawling on the World Wide Web with Arbitrary Predicates[C]. International Conference on World Wide Web, 2001.  
 [4] Hatzivassiloglou V, Mckeown K. Predicting the Semantic Orientation of Adjectives[J], 1997: 174--181.  
 [5] 王雪雅. 面向微博的企业竞争情报获取研究[D]. 安徽大学, 2015.  
 [6] 杨亮. 面向社交媒体的文本情感分析关键技术研究[D]. 大连理工大学, 2016.  
 [7] Feng, Yi-Xiong, Zheng, et al. Exploratory study of logistics service quality scale based on online shopping malls[J], 2007, 8 (6): 926-931.  
 [8] Mentzer J T, Gomes R, Krapfel R E. Physical distribution service: A fundamental marketing concept? [J], 2009, 17 (1): 53-62.  
 [9] 闫景民. 面向顾客的快递业服务质量评价指标体系研究[D]. 2012.  
 [10] 赵雪娟. B2C电子商务物流服务质量评价研究及应用[D]. 中北大学, 2017.  
 [11] 任敬波. 网购快递服务质量对顾客满意影响研究[D]. 2017.  
 [12] 曹霞, 李玲. 基于SERQUAL模型的网购快递服务质量评价体系研究[J], 2015, (8): 93-98.  
 [13] 曾凡虎. 基于在线学习的微博过滤技术研究[D]. 哈尔滨工程大学, 2013.  
 [14] 李钝, 乔保军, 曹元大, et al. 基于语义分析的词汇倾向识别研究[J], 2008, 21 (4): 000482-487.  
 [15] 周姣. 快递企业物流服务质量评价研究[D]. 安徽理工大学, 2018.