DOI:10.16799/j.cnki.csdqyfh.250354

基于APP在线评论的网约车服务质量评价指标体系构建

罗靖雯,王　富
(武汉工程大学 土木工程与建筑学院,湖北 武汉 430073)

摘　要: 网络预约出租汽车(以下简称网约车)是城市客运交通的重要组成部分,其服务质量直接影响着人们的出行体验和行业的发展,因此亟须构建一套科学、系统的网约车服务质量评价指标体系,为规范和提升我国网约车行业的服务质量提供参考。对滴滴出行APP在线评论进行语义分析,结合3种提取关键词的方法,构建了关键词词典和LDA主题模型运行结果;查阅相关文献、行业规范,借鉴SERVQUAL(服务质量)模型,构建了包含6个维度、29个评价指标的网约车服务质量评价指标体系。使用在线评论作为基础数据构建的网约车服务质量评价指标体系能够很好地满足科学性、实用性和独立性3项评价原则。

关键词: 交通运输;服务质量;SERVQUAL模型;LDA主题模型
中图分类号: U491.2　　**文献标志码:** A　　**文章编号:** 1009-7716(2026)01-0069-06

Construction of Evaluation Index System for Quality of Online Car-hailing Services Based on APP Online Comments

LUO Jingwen, WANG Fu
(School of Civil Engineering and Architecture, Wuhan Institute of Technology, Wuhan 430073, China)

Abstract: Online-hailing taxi (hereinafter referred to as "online car-hailing") is an important part of urban passenger transport. Its service quality directly affects the travel experience of the people and the development of the industry. Therefore, it is urgent to build a scientific and systematic evaluation index system for the quality of online car-hailing services, which provides a reference for regulating and improving the service quality of online car-hailing industry in China. The semantic analysis on the online comments of the Didi Travel APP is conducted. Combining three methods for extracting keywords, a keyword dictionary and the operation results of the LDA topic model are constructed. By consulting the relevant literature and industry norms and drawing on the SERVQUAL (Service Quality) model, a service quality evaluation index system for online car-hailing services is constructed, which includes 6 dimensions and 29 evaluation indicators. The evaluation index system for the quality of online car-hailing services constructed using online reviews as the basic data can well meet the three evaluation principles of scientificity, practicality and independence.

Keywords: transportation; service quality; SERVQUAL model; LDA topic model

0　引　言

2010年,第一个打车软件Uber在美国推出,与传统的巡游出租车服务模式相比,网约车平台具有即时响应性、费用公开透明性以及服务质量的标准化管理架构等优势。这种新型出行方式通过技术创新和运营模式革新,已在全球交通服务领域实现快速市场渗透,开启了网约车服务的新纪元。根据统计数据,2024年全球网约车市场规模约为1 949亿美元,预计到2029年,市场规模将进一步扩大至2 965亿美元,在预测期内(2024—2029年)复合年增长率为8.75%[1]。然而,随着使用群体的持续扩大,用户需求的异质性倾向逐渐显现,平台需要深入理解用户需求,针对用户提供更加精细化的服务。

目前,网约车服务质量评价的研究大多依赖问卷调查获得数据支撑,但问卷调查在数据客观性和有效性方面存在一定的局限性。相比之下,在线用户评论数据更客观,能够很好地解决这一问题。因

收稿日期:2025-04-01
基金项目:武汉工程大学研究生教育创新基金项目(CX2024513)
作者简介:罗靖雯(2001—),女,硕士在读,从事交通运输方向研究工作。
通信作者:王富(1978—),男,博士,教授,从事交通运输规划与管理方面的研究。电子信箱:550616898@qq.com

此，本文旨在使用APP在线用户评论数据来构建网约车服务质量评价指标体系，为网约车平台提供服务质量参考。优质的服务质量不仅是网约车平台公司生存发展的关键，也是促进网约车行业健康、可持续发展的保障。

本文基于APP在线用户评论的网约车服务质量评价指标体系构建流程框架见图1。

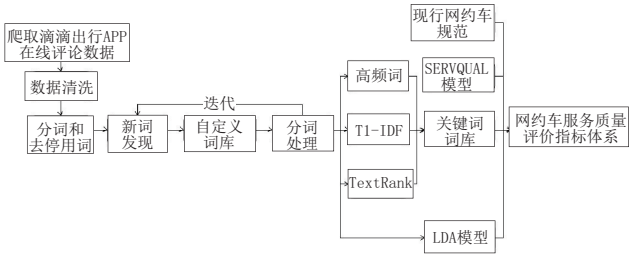

图1　基于APP在线用户评论的网约车服务质量评价指标体系构建流程框架

1　在线数据获取和预处理

1.1　数据的采集

使用网页爬虫的方法对所需数据进行采集。数据来自于第三方移动数据监测机构七麦数据（https://www.qimai.cn/）。七麦数据本身并不产生数据，而是收录iOS和Android上移动APP的实时数据，并构建历史数据库。使用八爪鱼采集器从历史数据库初步获得各应用市场2019年12月1日至2024年12月1日的滴滴出行APP评论数量信息（见表1）。

表1　APP评论数量信息

应用市场	APP Store	华为	小米	vivo	oppo
评论数量/条	122 608	7 953	5 227	6 239	7 019

1.2　数据预处理

1.2.1　数据的清洗

数据清洗主要包括以下3个步骤：

（1）去除重复评论。使用Excel自带的“删除重复项”功能，将评论文本中内容完全相同的评论剔除，最终得到了143 916条唯一项。

（2）去除非中文字符。通过pandas模块中的replace函数和Unicode编码，使用正则表达式，将评论文本内容中的数字、字母和符号等非中文字符替换成空字符串。

（3）去除过短评论[2]。使用Excel中的LEN函数，去除3个字段以内的评论共计395条。

经过筛选后共获得了有效评论数据143 521条。

1.2.2　在线评论分词和去停用词

为了更好地理解评论的语义，并进一步分析文本，采用python中的jieba模块对评论数据进行分词处理。但是，分词中存在如“啊”、“呀”、“我”、“你”等没有实际意义的停用词，为了提高分词效率，采用哈工大停用词库将停用词去除。

1.2.3　新词发现

将分词过程中出现的高频新词加入自定义词库可以显著提高分词效果。线上评论数据为自然语言，随着网络的发展会出现大量新兴的网络热词，分词库中未添加过这些词汇，但它们经常前后一起出现。根据这一特点，利用python算法，通过计算词频和左右熵从在线评论数据中识别和提取出未在词典中出现的新词或短语，部分示例见表2，在里面进行筛选，将准确的词加入自定义词库，再迭代进行分词处理，直至新词发现中不再包含网络热词、领域词、专业名词等。

表2　新词发现部分示例

词语	左右熵值	词频/次
霸王条款	1 473.131 727	72
评差评差	960.999 107 3	219
团队付出	797.371 822 4	47
网信办	763.423 659 7	45
取消费	474.929 404 3	141
割韭菜	446.729 752 8	34
花小猪	438.138 796	95
实名认证	436.285 283 5	67
薅羊毛	402.932 718 2	34
虚假宣传	258.084 121 4	40

针对线上打车领域，另外添加了例如“派单”、“过城费”、“杀熟”等口语化词汇到自定义词库中。经过迭代和词库更新，不断优化分词，得到最终的分词结果，为后续关键词提取和主题分析奠定了基础。

2　网约车服务质量评价指标体系构建

2.1　关键词词典的构建

为了更好地了解用户需求并构建有效的服务质量评价指标体系，基于用户在线评论数据，采用3种不同的关键词提取方法来建立关键词词典，从大量在线用户评论中快速筛选出与服务质量相关的关键信息，去除评论中冗余的噪声信息，调整指标细节。

2.1.1　词频统计

词频统计（term frequency，TF）是从文本中提取关键词的一种常用方法。通过计算文本中各个词语出现的频率，可以识别出最重要或最具代表性的词。本文借助jieba库对评论数据进行词频和词性统计。

词性统计如图 2 所示，名词、动词和形容词占比超过 90%。名词，直观反映了用户关注的核心对象或需求点，指导快速识别出用户关心的服务或问题；动词，揭示了事物之间的关系和过程，指导理解用户的行为模式或需求变化；形容词，反映了用户的情感态度，它们通过修饰名词来表达对事物的评价，指导识别出用户的情感倾向。本文只保留这 3 种词性进行后续分析，对其他词性词语进行识别并剔除。

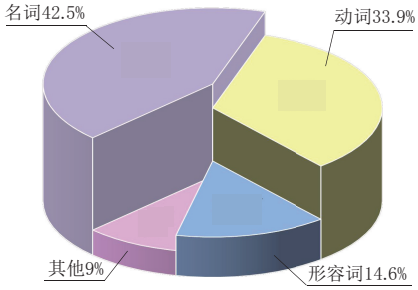

图 2　词性统计图

表 3 为词频统计结果部分示例（词频从高到低）。根据网约车在线评论词频统计结果，可以基本看出用户针对网约车服务质量的关注焦点。比如，在评论中出现频率较高的词汇包括“打车”、“软件”、“服务”、“价格”、“司机”等。此外，评论中出现的“喜欢”、“专业”、“推荐”等主观性词汇频率也很高，表现用户对于网约车服务质量的感知。

表 3　词频统计结果部分示例

序号	词语	词频/次	序号	词语	词频/次
1	软件	99 432	11	垃圾	12 543
2	打车	87 260	12	一个	12 349
3	不错	84 130	13	专业	11 146
4	滴滴	49 313	14	平台	10 708
5	司机	29 811	15	推荐	10 269
6	好用	27 592	16	支持	9 061
7	出行	25 278	17	下载	8 949
8	好好	19 456	18	时间	8 868
9	喜欢	16 583	19	客服	8 676
10	真的	13 739	20	感觉	8 570

2.1.2　TF-IDF 算法

TF-IDF 算法结合词频（term frequency，TF）和逆文档频率（inverse document frequency，IDF）来计算词语在文本数据中的重要性[3]。TF 是一个特定数据在数据集中出现的频率；IDF 是指反向数据的发生频次，其用来度量某一数据集对整体数据库的重要性[4-6]且取的是对数值 log，IDF 越大表示这个词越稀有，因此更加重要。TF-IDF 法实际是 TF×IDF，可以去除常见词而保留重要的词。利用 Python 中 jieba 分词包的 TF-IDF 算法，获取该算法下在线评论关键词按照权重值的排序，排序前 20 的词汇见表 4。

表 4　TF-IDF 算法关键词提取结果

序号	TF-IDF	权重值	序号	TF-IDF	权重值
1	打车	0.583 02	11	应用	0.061 54
2	软件	0.537 34	12	客服	0.055 29
3	不错	0.382 73	13	专业	0.050 74
4	滴滴	0.356 73	14	平台	0.049 15
5	好用	0.247 30	15	下载	0.047 02
6	司机	0.149 18	16	很好	0.046 51
7	出行	0.146 56	17	推荐	0.041 86
8	方便	0.077 26	18	网约	0.041 63
9	垃圾	0.073 67	19	代言	0.041 18
10	喜欢	0.069 52	20	值得	0.035 88

2.1.3　TextRank 算法

TextRank 算法是一种无监督的图排序算法，该算法基于图的结构，通过词语或句子之间的相似性来进行排序，从而选出文本中最重要的部分。TextRank 的核心思想借鉴了 Google 的 PageRank 算法，用于计算文本中词语或句子的“重要性”[7]，它能够利用文本数据中语义共现信息进行关键词、关键短语以及关键句抽取。本文主要将文本中的词语作为节点，通过计算词之间的共现频率建立图结构，然后通过 PageRank 算法计算每个节点的权重，最终抽取重要关键词。利用 Python 中 jieba 分词包的 TextRank 算法，获取该算法下在线评论关键词按照分数从大到小的排序。表 5 为排序前 20 个关键词提取结果。

表 5　TextRank 算法关键词提取结果

序号	Text Rank	TextRank 分数	序号	Text Rank	TextRank 分数
1	软件	1.000 00	11	客服	0.149 11
2	打车	0.909 27	12	方便	0.148 98
3	滴滴	0.671 60	13	喜欢	0.123 76
4	不错	0.608 70	14	专业	0.117 74
5	司机	0.439 66	15	时间	0.109 90
6	好用	0.273 61	16	应用	0.109 53
7	出行	0.264 18	17	支持	0.089 45
8	垃圾	0.225 26	18	推荐	0.086 16
9	平台	0.169 95	19	乘客	0.084 92
10	没有	0.163 29	20	下载	0.079 19

2.1.4　关键词词典

为确保提取的关键词能够准确反映用户评价的原始语义，本研究结合词频统计、TF-IDF 算法和 TextRank 算法的结果，综合筛选高频且具有区分度的关键词，仅保留 3 种方法均提到的前 100 个词语，

避免单一方法的偏差，构建基于在线评论的网约车服务质量关键词词典，对关键词词典进行人工校验，将关键词与实际评论内容做对比，验证关键词是否保持和继承原有评价语义，避免算法和人工误判。词云图是数据可视化的一种常见形式，特别适用于文本数据的处理和分析[8]。图3为对关键词词典中关键词的词频进行统计后得出的词云图。由此词云图可以直观地看到，网约车用户最关注软件、平台、服务、价格、时间，其次是环境、功能等要素。

图3　词云图

2.2　LDA主题模型

LDA（Latent Dirichlet Allocation，潜在狄利克雷分配）是一种无监督学习方法[9]，不需要标注数据就能进行训练和主题发现，LDA能够帮助从大量无结构的文本中提取潜在的语义信息。本文利用LDA主题模型对分词后的在线评论进行主题分布分析，发现潜在的用户需求，指导建立指标体系。

2.2.1　确定主题个数

主题数量K值是LDA模型中一个关键性的超参数，随着K值的变大，主题会更具有特点和精细化，但是当K值过大时会引起过拟合，主题之间会出现重叠，并且由于LDA模拟结果需要人工识别出主题，过多的主题可能导致每个主题包含的信息变得稀薄，从而影响可解释性。因此，需要通过一致性（Coherence）和困惑性（Perplexity）确定主题的数量K。一致性，衡量每个主题下的词语能否共同表达一个具有一致语义的概念，通过计算主题中词汇之间的共现关系来评估。困惑性，衡量模型对新文档集合的预测能力，通过计算文档的似然度来评估。这两个指标用于评估主题模型的质量，帮助理解模型生成的主题是否具有实际意义。通常期望在提高一致性的同时，能够保持较低的困惑性。通过网格搜索（K=2~15）结合一致性和困惑性综合确定LDA模型的主题数量（K值）。利用python代码，分别计算出当主题数设置为2~15时的一致性和困惑度输出结果（见图4）。当K=5时，一致性得分达到峰值（表明主题内词语语义高度相关），同时困惑度处于较低水平（表明模型对新数据的泛化能力较强）。根据结果显示最佳主题数K=5。

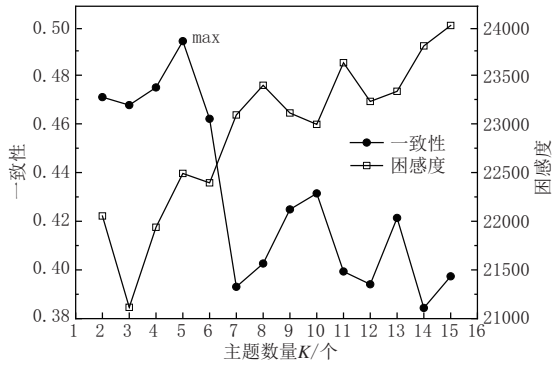

图4　一致性与困惑度变化曲线

2.2.2　LDA主题模型

先验分布是统计推断中基于已有认知设定的初始概率分布，用于量化对模型参数的初始认知，并约束分析结果的合理性。根据文献中常见的参数设定，将Dirichlet先验分布的超参数α和β分别设置为：α=0.1（控制文档-主题分布的稀疏性，避免主题分布过于集中），β=0.01（控制主题-词语分布的稀疏性，确保每个主题由少量核心词汇表征）[10]。根据上文确定的文本主题个数5，设定代码参数num_topics=5，便可以得到5个主题下各个词语及其强度，强度越大表征该词越贴近主题，提取出每个主题中有实际意义且按照权重大小排序的词语部分列出（见表6）。

利用pyLDAvis模块和gensim库生成LDAvis图对模型结果进行可视化，如图5所示。这是一个交互式的图像，左边的气泡代表不同主题，气泡越大表示属于这个主题的词越多，各个气泡间的距离代表主题间的相似程度，气泡越接近，代表主题间相似度越高。通过图5观察主题间距离与词语分布，各主题均语义独立且无重叠。右边是一个个主题词，图例中浅色调条代表词的频数，当把鼠标放在左边的气泡上时，气泡会变成深色，相应的右边条形图也会变为所选气泡代表的主题词，深色调条代表这个词在这个主题内的频数。

根据LDA主题模型运行出来的结果以及可视化图像可以识别出：Topic 1为好用、专业、喜欢、下载，与用户对整体服务的感知有关；Topic 2为客服、广告、功能、拼车，与平台功能有关；Topic 3为师傅、手机、服务态度、良心，与用户对司机师傅的评价有关；Topic 4为不错、推荐、有意思、垃圾，与用户对软件的评价有关；Topic 5为希望、取消、师傅、时间、体验，表

表 6　LDA 主题模型主题词及其概率分布

Topic 1		Topic 2		Topic 3		Topic 4		Topic 5	
词语	强度	词语	强度	词语	强度	词语	强度	词语	强度
出行	0.144 73	客服	0.028 15	滴滴	0.091 34	不错	0.397 33	滴滴	0.057 32
软件	0.143 30	广告	0.024 73	师傅	0.072 76	垃圾	0.052 64	希望	0.030 35
好用	0.112 63	拼车	0.022 71	乘客	0.012 69	有意思	0.016 46	师傅	0.025 18
喜欢	0.032 37	更新	0.019 60	良心	0.008 50	软件	0.016 39	取消	0.015 02
专业	0.024 65	功能	0.013 80	用户	0.008 23	推荐	0.014 88	打车	0.012 82
下载	0.014 87	打车	0.012 59	时间	0.006 85	超级	0.013 75	时间	0.012 60
真心	0.013 95	界面	0.010 76	投诉	0.006 84	产品	0.008 85	专车	0.012 54
推荐	0.013 55	优惠券	0.008 73	服务	0.006 80	感觉	0.007 01	服务	0.010 66
支持	0.012 02	地图	0.008 03	手机	0.006 19	很棒	0.007 00	快车	0.010 16
好评	0.009 61	支付	0.007 49	客服	0.005 26	良心	0.005 97	体验	0.008 89
实用	0.008 81	运行	0.007 23	发现	0.004 84	朋友	0.005 93	定位	0.008 83
值得	0.008 48	简洁	0.007 08	客户	0.004 74	真不错	0.005 60	价格	0.006 32

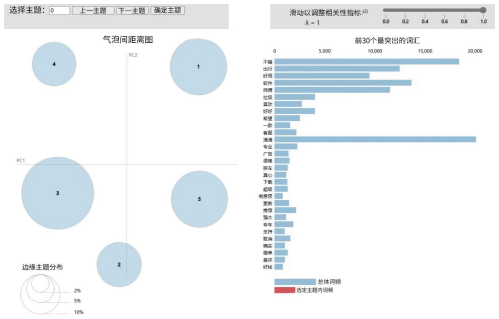

图 5　LDA 主题模型结果可视化

达了用户对网约车各方面服务的期待，如司机师傅的服务态度、取消打车的取消费用责任认定、打车时间的期望、专车服务的体验感、定位的精确度、价格的透明度等。根据 5 个主题及其下主题词对指标体系进行指导构建。

2.3　指标体系的构建

2.3.1　指标选取原则

为确保评价结果科学、合理，在选取评价指标时应遵循以下原则[11]：一是科学性，选取的指标应具备理论依据和严谨性，遵循一定的科学原则和标准；二是实用性，选取的指标应当易于采集，且应避免选取网约车平台难以采取措施加以改进的指标；三是独立性，选取的指标之间概念界限应清晰，易于区分，并且相互独立，无包含关系。

2.3.2　确定维度

本研究评价维度以 SERVQUAL 模型有形性、可靠性、响应性、保证性和移情性这五大维度为理论基础[11]。然而，LDA 主题分析显示，用户评论中隐含的语义焦点与理论框架存在差异。LDA 主题 4（软件评价）中高频词“价格”“优惠券”凸显了用户对服务成本的敏感，而主题 5（服务期望）中“取消费”“定位”等词进一步揭示了用户对价格透明度的关注，因此在 SERVQUAL 量表的五大维度中增加了一个“经济性”维度。同时，LDA 主题 3（司机评价）中“安全”“合规”等词，以及主题 5 中“定位”“取消”等词，均指向用户对驾驶安全与合规性的强烈关注。对于作为出行领域的网约车来说，值得充分关注的头等问题依旧是乘车安全问题，因此将“保证性”维度改成“安全性”维度。此外，LDA 主题的语义分布表明，用户对服务质量的评价呈现交叉性特征。例如，主题 1（服务感知）中“专业”“推荐”等词补充了“移情性”的内涵（如服务专业性），而主题 2（平台功能）中“客服”“功能”等词同时支撑了“可靠性”与“响应性”维度。根据前面的调整，本文将网约车服务质量测量的维度分为六个维度，即有形性、可靠性、安全性、响应性、移情性、经济性（见表 7）。这一过程既尊重了 SERVQUAL 模型的系统性，又通过 LDA 主题分析捕捉了用户评论中的实际诉求，避免了理论与数据的割裂。需要说明的是，LDA 主题作为无监督的语义聚类结果，其与维度的映射并非严格一一对应（如主题 5 同时涉及安全性与经济性），而维度设计需要考虑到服务全流程，因此允许语义交叉。这种非严格映射恰恰体现了数据与理论的互补性：LDA 模型修正了理论模型，而理论框架确保了维度的完整性与可解释性。

2.3.3　评价指标体系

根据关键词词典以及 LDA 主题模型运行结果，对评价指标进行适应性调整与改进，将关键词词典以及主题词中未提及的指标进行删改，得到最终网

表7　网约车服务质量评价维度及主题支持

评价维度	维度释义	LDA主题支持
有形性	服务过程中的可见因素	Topic 1、Topic 4
可靠性	服务提供者能否按时、按要求提供服务	Topic 2、Topic 5
安全性	用户能免于担心危险、风险式疑惑等状况	Topic 3、Topic 5
响应性	服务提供者对用户需求、问题的响应速度和态度	Topic 2
移情性	服务提供者能够理解并关心用户的具体需求	Topic 1、Topic 3
经济性	用户对服务成本的感知	Topic 4、Topic 5

约车服务质量评价指标体系（见表8）。

表8　网约车服务质量评价指标

维度	评价指标	关键词
有形性	车内外卫生情况	干净，车型
	安全设施配置	专业，安全
	车辆性能良好	专车，晕车
	司机形象	师傅，良心
	APP操作界面友好度	界面，弹窗
可靠性	打车成功率	接单，拒载
	实际里程	定位，路径
	等待时间	等待，迟到
	平台杀熟情况	杀熟，选择
	驾驶员绕路情况	绕路，定位
	驾驶员中途甩客或中途接客情况	甩客，目的地
	定位准确度	定位，位置
安全性	驾驶员安全驾驶情况	安全，变道
	车辆、驾驶员合规性[11]	专车，合法化
	支付安全边界，提供发票	发票，下载
	隐私信息安全性	信息，账号
	驾驶员是否有使用手机、抽烟等不良行为	手机，抽烟
响应性	接单响应性	接单，打不到
	客服响应性	反馈，回复
	投诉响应性	投诉，客服
	失物招领[12]	活动，体验
移情性	个性化服务	丢失，找回
	驾驶员态度	态度，素质
	代言	代言人，广告
经济性	价格透明度	价格，便宜
	高峰时段加价情况	早高峰，加价
	计价标准清晰度	取消费，显示
	平台优惠券	优惠券，优惠
	无线下收费情况	收费，高速费

3　结　论

本文旨在基于滴滴出行APP在线用户评论数据集构建一套系统的网约车服务质量评价指标体系。

（1）利用滴滴出行APP的在线评论数据，通过语义分析和3种关键词提取算法，构建了关键词词库。通过LDA主题模型分析，识别出了用户对网约车服务质量的主要关注点，涵盖了服务、平台功能、司机评价、软件体验及服务期望等方面。

（2）结合关键词词库、LDA运行结果、SERVQUAL模型以及现行的《网络预约出租汽车运营服务规范》（JT/T 1068—2016），共同指导构建了包含6个维度和29个评价指标的网约车服务质量评价指标体系，为规范和提升我国网约车行业的服务质量提供了参考。

（3）使用在线评论构建出来的网约车服务质量评价指标体系能较好地捕捉到实际乘车过程中用户的感知质量，但本文所构建的指标体系仍然在指标适用范围的拓展和数据处理方法的完善方面存在一定的局限性，需要通过实践探索进行不断完善，期望在未来网约车行业发展过程中，能够进一步丰富该领域的研究成果。

参考文献：

[1] 韩淼.平台经济C2C模式税收征管问题研究[D].昆明：云南财经大学，2024.
[2] 徐文平，张媛.基于ISM-MICMAC的生鲜物流服务质量影响因素[J].物流技术，2022，41（10）：32-36.
[3] 徐晓彤.人工智能技术在个性化推荐系统中的应用与效果评估[J].信息记录材料，2025，26（1）：94-96.
[4] 李辉，郝翠萍，常欢欢.基于朴素贝叶斯的多模态异构大数据分类优化方法[J].电子设计工程，2025，33（2）：158-162.
[5] 孙二华，胡云冰.基于鲸鱼优化和深度学习的不平衡大数据分类算法[J].西南师范大学学报（自然科学版），2021，46（5）：127-133.
[6] 高艳.基于分簇算法的传感器网络大数据传输优化方法[J].传感技术学报，2022，35（9）：1262-1267.
[7] 向姝璇.基于专利文本特征的核心技术信息抽取方法研究[D].成都：四川大学，2022.
[8] 陈杰，王艳丽，周峰.基于电商大数据的物流信息系统课程实验设计[J].实验科学与技术，2025，23（4）：118-125.
[9] 刘泽琨，江采欣.中国新能源汽车产业政策变迁及阶段特征研究[J].科研管理，2025，46（6）：146-156.
[10] 宋俊毅，赖昀耀，胡飞菊，等.基于LDA的民宿评论情感分析研究[J].科技传播，2024，16（23）：111-115.
[11] 李晋，江勇东，顾天奇，等.基于乘客视角的网约车服务质量评价[J].城市交通，2021，19（2）：103-111.
[12] 杨亚琪.基于SERVQUAL和IPA模型的网约车服务质量评价研究[D].成都：成都理工大学，2018.