

基于抖音数据的地铁运营舆情主题识别研究

张诗怡

同济大学经济与管理学院, 上海

收稿日期: 2026年2月27日; 录用日期: 2026年3月19日; 发布日期: 2026年4月1日

摘要

近年来, 短视频平台已成为公众表达意见的重要渠道, 聚焦短视频中关于地铁运营管理的讨论, 从中挖掘出网友对地铁运营管理的意见和建议, 对于提升地铁运营管理水平具有重要意义。本研究以抖音平台为数据源, 爬取2018年至2025年间地铁运营相关视频1842条, 采用BERTopic主题聚类模型对视频描述文本进行主题建模, 识别出24个语义区分度较高的初始主题, 并参考国家标准将其归纳为8个大类构成的多层次舆情主题体系。分析表明, 公众讨论呈现出从基础服务评价向社会价值认知延伸的特征, 既关注安全事故、设备故障、基准票价等运营核心议题, 也涉及无障碍服务、智能化运营、人文化服务等人文与社会价值层面, 乘客行为主题的识别进一步揭示了地铁空间中个体行为与制度管理的深层互动。本研究构建的主题体系具有多维度、场景化、体验导向的特点, 能够系统反映公众对地铁服务的关注焦点与价值诉求, 为地铁运营部门开展精细化服务改善、舆情监测预警与公众沟通策略制定提供了清晰的结构化框架与决策依据。

关键词

地铁运营, 舆情分析, 主题建模, BERTopic, 抖音

Research on Topic Identification of Metro Operation Public Opinion Based on Douyin Data

Shiyi Zhang

School of Economics and Management, Tongji University, Shanghai

Received: February 27, 2026; accepted: March 19, 2026; published: April 1, 2026

Abstract

In recent years, short video platforms have become an important channel for public opinion expres-

文章引用: 张诗怡. 基于抖音数据的地铁运营舆情主题识别研究[J]. 统计学与应用, 2026, 15(4): 1-12.

DOI: 10.12677/sa.2026.154066

sion. Focusing on discussions related to subway operations and management within these platforms, and extracting opinions and suggestions from netizens, is of significant importance for improving the level of subway operation and management. This study uses Douyin as the data source, collecting 1,842 subway operation-related videos from 2018 to 2025. The BERTopic topic modeling approach is applied to analyze video description texts, identifying 24 initial topics with high semantic discriminability. These topics are further categorized into a multi-level thematic framework comprising eight major categories, based on national standards. The analysis reveals that public discussions exhibit a shift from basic service evaluation to social value recognition, covering not only core operational issues such as safety incidents, equipment failures, and fare mechanisms, but also humanistic and social dimensions including barrier-free services, intelligent operations, and cultural expression. The identification of passenger behavior topics further uncovers the deep interaction between individual conduct and institutional management within the subway space. The thematic framework constructed in this study is characterized by multidimensionality, contextualization, and an experience-oriented approach. It systematically reflects the public's focal points of concern and value demands regarding subway services, providing a clear structured framework and decision-making basis for subway operators to implement refined service improvements, conduct public opinion monitoring and early warning, and develop communication strategies with the public.

Keywords

Subway Operation, Public Opinion Analysis, Topic Modeling, BERTopic, Douyin

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

由于城市的地域、经济规模与集聚效应之间的相互影响关系，城镇人口逐年向国内各经济发达的大中城市集中，而城市人口的高度聚集与城市规划、基础设施建设水平的相对落后之间存在着种种矛盾，容易引发一系列的社会问题，造成“城市病”，其中交通拥堵问题尤为突出[1]。地铁在运输能力、运输速度和运输效率方面都优于私家交通和传统公交，并且具备节能、准时、舒适和安全的特点，因此日益成为现代城市重要的交通方式之一[2][3]。然而，由于地铁系统的复杂性和外部环境不确定性，我国的地铁运营在安全、效率和乘车环境等方面也存在诸多问题，比如设备故障、管理运维成本高、乘客体验差等[4]，这也引起了公众的反感和激烈讨论。因此，在这种背景下，密切关注公众舆论焦点，对于及时发现和解决目前我国地铁运营存在的问题、满足公众在地铁运营方面的需求以及营造地铁运营的良好舆论氛围具有重要意义。

现有的关于地铁运营公众关注点及态度研究多采用问卷调查、访谈等方式进行[5][6]，然而，问卷调查是一种被动的信息披露过程，受访者在此过程中会考虑隐私问题及其可能带来的负面影响，这会导致问卷数据的真实性出现问题[7]。随着大数据时代的到来和相关技术的普及，在社会科学研究中，文本挖掘与网络爬虫相结合进行非结构化数据的挖掘与分析，比传统方法更加客观有效[8]，目前已被广泛应用于风险识别[9][10]、情感分析[11]、城市治理[12]和安全管理[13]等领域。并且，随着互联网的普及和繁荣，社交网络平台近年来成为发布、传播和交流信息的重要渠道，人们在这些平台上自由表达对一些话题或事件的看法并进行互动。抖音平台作为基层百姓表达声音的重要平台，具有信息丰富、实时性强和互动性高等优点，能够反映公众对于热点事件、社会议题以及日常生活的真实情感与观点，是研究集体

行为、情感演变和信息传播模式的理想数据源。近年来,许多学者基于抖音平台开展了关于舆情方面的研究[14][15]。而地铁作为一个涉及公共安全、运行复杂且与城市居民日常生活密切相关的交通系统,与其他交通工具的运营管理存在较大差异,有必要对其舆情主题进行深入研究,尤其是在抖音平台等新兴社交媒体环境下。

2. 研究综述

2.1. 地铁运营相关研究

地铁作为城市公共交通的重要组成部分,近年来受到广泛关注,相关研究涵盖了多个方面,包括地铁建设、运营效率、社会经济影响及运营安全风险等。随着城市化进程的加快,地铁系统的建设与运营在缓解交通拥堵、促进经济发展和提升城市形象等方面发挥了关键作用。学者们从不同视角对地铁运营进行深入探讨,揭示了其在城市可持续发展中的重要性。

首先,地铁的建设与运营效率是研究的核心内容之一。研究表明,地铁系统的高效运营能够显著提高城市的交通效率,降低出行时间。许多研究集中于地铁网络的规划、站点布局和运行调度等技术性问题,探讨如何在资源有限的情况下优化地铁的运营效率。同时,地铁的运行效率还与其经济可持续性密切相关。一些研究指出,高债务风险和低客流量可能导致地铁运营面临盈利困境[16][17],因此,分析地铁的经济效益和社会回报成为了重要的研究方向。

其次,地铁运营对社会经济发展的影响也是相关研究的重要议题。研究表明,地铁建设往往能够提升沿线区域的地价和房价,促进商业发展[18]-[21]。此外,地铁的便捷性和高效性还被认为有助于提高居民的幸福感和生活质量[22]。并且,地铁运营的社会影响并不仅限于经济层面,还涉及环境保护和城市规划等方面。地铁作为低碳交通工具,能够有效减少城市的汽车排放污染,改善空气质量[23][24]。

近年来,地铁运营安全风险问题逐渐引起学术界和社会的广泛关注。研究表明,地铁运营中的安全风险主要源于设备故障、人员失误以及外部环境因素,其中,设备故障通常是引发事故的主要原因[25]。并且,由于致因因素与事故之间存在复杂的、动态的相互关系,有研究通过构建地铁运营事故成因网络来捕捉地铁运营中复杂的事故安全风险与成因关系,并为优化安全决策、降低成因和控制事故提供有效建议。

综上所述,地铁运营相关研究在技术、经济和社会等多个维度上都展现出广泛的学术价值。这些研究不仅为地铁系统的安全管理和效率提升提供了理论基础,也为政策制定者和城市管理者在规划和实施地铁项目时提供了实用的参考依据。随着城市化进程的加快和公共交通需求的增加,未来的研究应更加关注地铁运营中的动态因素与复杂关系,利用新兴技术推动更为科学的决策方法和管理实践。

2.2. 基于社交媒体数据的主题建模相关研究

在过去的研究中,大多数学者主要使用新闻报道等数据进行自然语言处理[26]。然而,随着互联网技术的发展,公众愈加倾向于在社交媒体上表达自己对于事物的看法和意见,比如推特、微博和抖音等。社交媒体数据的时效性和多样性使得研究者可以捕捉到公众情感和态度的瞬时变化,从而提供更为全面和动态的视角,帮助政策制定者和相关利益方更好地理解 and 回应公众需求,因此,利用社交媒体数据可以进行更为细致的研究。

主题建模技术在社交媒体数据研究中具有重要地位,它能够从大量文本数据中自动提取潜在的主题,帮助研究者理解公众对某一事件或项目的关注点。Tang 等人[27]利用 Twitter 数据对美国建筑行业的四个主要群体(建筑工人、建筑公司、建筑工会和建筑媒体)进行了主题建模和情感分析,揭示了不同群体对行业的看法及其日常关注的焦点。随着时间的推移,公众关注的焦点会逐渐发生变化。为了捕捉这些变化,

Hong 等人[28]结合时空信息进行主题建模,揭示了自然灾害期间公民与地方政府在推特交流中的动态主题变化及其与灾害严重程度的关系,研究表明,主题建模有效捕捉了公众关注点的时空变化,为地方政府的信息传递策略调整提供了重要依据。Xue 等人[29]基于 LDA 开发了动态主题模型,揭示了公众关注点随时间变化的趋势以及各个关注点与利益相关者之间的关联性。在之前研究的基础上,Tang 等人[30]进一步爬取中国建筑工人、建筑公司、建筑工会和建筑媒体的微博数据,比较了中美两国公众对建筑行业看法的异同。翟羽佳等人[31]分析了新冠疫苗信息在抖音和微博两个社交平台中的传播特征与互动关系,采用 Top2Vec 主题建模和格兰杰时间序列检验方法,揭示了不同主题的演化规律和平台间的联动效应,为提高健康信息传播率提出了对策建议。於世为等人[32]以 2021 年“拉闸限电”事件为研究对象,基于群体智慧理论提出大数据驱动的研究框架,用 BERTopic 方法深入分析公众对事件的评论数据,从而揭示事件原因和影响,该研究为突发事件舆论引导策略提供了参考。

上述研究展示了主题建模方法在不同领域中的广泛应用,尤其是在理解公众对重大项目和技术的关注焦点和态度演变方面。这些研究成果为政策制定者和研究人员提供了重要的参考依据,有助于制定更加有效的沟通策略和公众参与机制。此外,这些研究为未来的公众舆论研究奠定了坚实的理论基础,鼓励后续学者在更广泛的情境中探索相关议题。

3. 数据收集与预处理

3.1. 数据检索与采集

抖音作为当前国内极具代表性的短视频社交平台,拥有庞大的用户基数与高频的交互特征,其用户生成内容(UGC)能够及时、真实地反映公众对特定社会事件的情感诉求。本研究利用 Python 网络爬虫技术获取“地铁运营”相关的抖音视频内容。

为确保研究样本的全面性与代表性,本研究构建了三维关键词检索策略。首先,在实体维度上,以“地铁”及“1 号线~20 号线”为关键词,覆盖国内主流地铁线路名称;其次,在空间维度上,采用“城市/省份名称 + 地铁”组合词,涵盖截至 2025 年底已开通地铁的所有行政区划;最后,在主题维度上,围绕运营核心领域,以“地铁运营 + 安全/管理/线路/过程/系统/风险/乘客/服务/票价”为关键词进行检索。

通过上述策略,初步获取视频数据 26756 条。经过去重处理及人工逻辑校验(剔除与地铁运营主题无关的视频),最终筛选出核心视频数据 1842 条。视频发布时间范围在 2018 年 8 月 18 日和 2025 年 12 月 31 日之间。采集字段包括视频作者 ID、发布时间、视频 ID、视频链接、视频时长,以及点赞量、评论量、收藏量等交互特征指标。

3.2. 文本处理

为进行深入的文本挖掘,需对获取的多源异构数据进行标准化清洗与整合。考虑到视频标题往往难以全面涵盖其承载的语义信息,本研究利用科大讯飞开放平台的语音识别技术,将获取的 1842 条视频音频转换为文本,并将抖音视频的标题与语音转换的文本组合为视频描述文本。

3.3. 描述性统计

为更清晰地呈现样本特征,本文对最终纳入分析的 1842 条抖音视频进行了描述性统计。如图 1 所示,样本在 2018~2025 年间呈现明显的阶段性增长趋势。2018~2020 年相关视频数量总体较少,分别为 4 条、10 条和 15 条;自 2021 年起样本数量开始上升,2021 年和 2022 年分别达到 77 条和 98 条;2023 年后增长更为显著,2023 年为 289 条,2024 年增至 650 条,2025 年进一步上升至 699 条。整体来看,地铁

运营相关讨论在抖音平台上的可见度与活跃度呈持续增强趋势，说明短视频平台日益成为公众表达地铁运营认知与体验的重要场域。

从时间结构上看，样本主要集中于 2023~2025 年，这一方面与近年来短视频平台用户规模扩大、内容生产活跃度提升有关；另一方面也可能与地铁建设扩张、运营事件曝光增加以及平台算法对公共议题内容的传播放大效应有关。因此，本文所识别出的主题体系在较大程度上反映的是近几年地铁运营舆情的关注结构，而对于早期年份的舆情特征代表性相对有限。

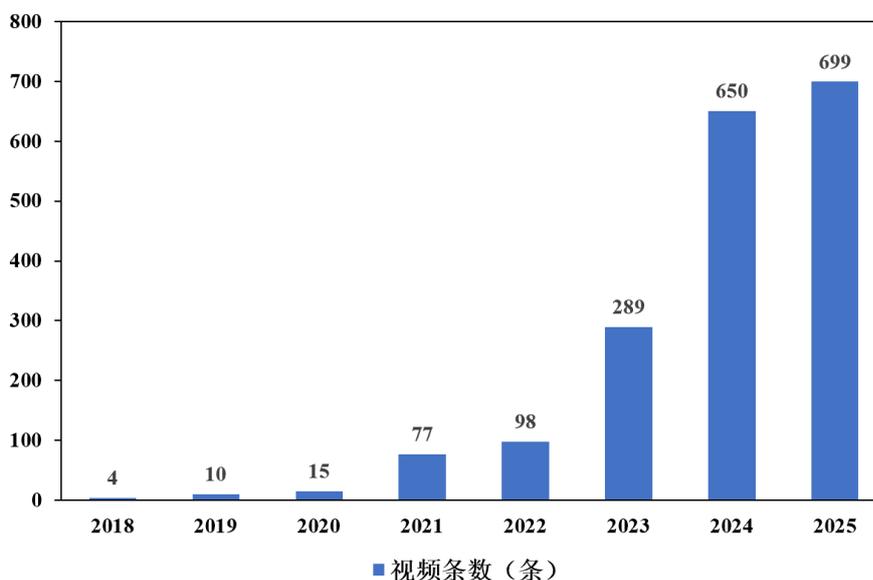


Figure 1. Annual distribution of sample videos (2018~2025)
图 1. 样本视频年度分布(2018~2025)

4. 地铁运营舆情主题识别分析：基于 BERTopic 模型

4.1. BERTopic 主题聚类模型介绍

主题建模是一种旨在从大规模文档集合中自动识别潜在主题结构的无监督机器学习方法，能够在缺乏人工标注的情况下挖掘文本数据中隐含的语义信息，已被广泛应用于舆情分析、公共管理与社会计算等研究领域。相较于以词袋假设为基础的传统主题模型(如 LDA)，近年来基于深度学习架构的主题建模方法在语义表示能力方面表现出明显优势，尤其适用于短文本、碎片化文本及语境依赖性较强的数据场景。

BERTopic 模型是一种融合了预训练语言模型与传统聚类思想的主题建模方法，其核心思想是通过上下文语义嵌入对文本进行高维表示，并在此基础上完成主题聚类与语义表征。该模型以 Transformer 架构为基础，能够充分利用上下文信息捕捉文本的深层语义关系，从而在语义一致性与主题区分度方面优于传统概率主题模型。对于抖音平台中以视频标题与描述为代表的地铁运营舆情文本而言，其表达方式高度口语化且语境依赖性强，BERTopic 模型在此类场景下具有较好的适配性。

从算法流程上看，BERTopic 模型主要由文本向量化、特征降维、文档聚类和主题表征四个核心环节构成，其整体流程如图 2 所示。首先，在文本向量化阶段，模型利用预训练语言模型将原始文本映射至高维语义空间，以获得能够反映上下文语义信息的文档嵌入向量；其次，在特征降维阶段，通过降维算法对高维嵌入向量进行压缩，以缓解“维度灾难”并提高聚类效率；再次，在文档聚类阶段，依据文本在

低维空间中的语义距离对文档进行聚类，从而形成若干语义相近的主题簇；最后，在主题表征阶段，模型基于每个主题簇内文本的词汇分布特征，提取具有代表性的关键词，实现主题的语义表达与命名。

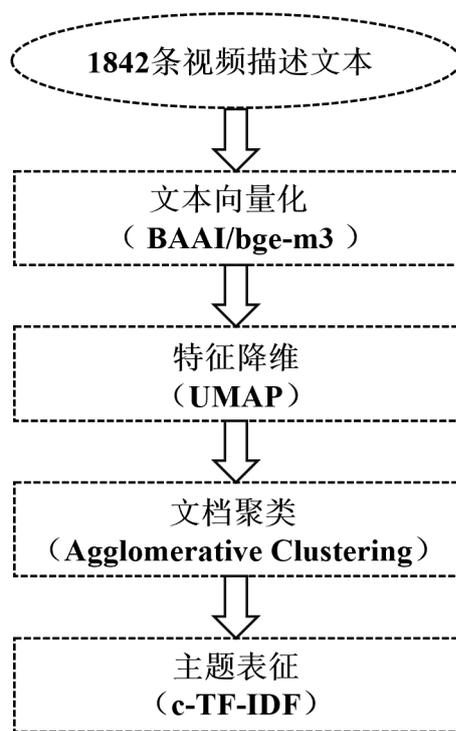


Figure 2. BERTopic model pipeline
图 2. BERTopic 模型流程

4.2. BERTopic 主题聚类模型构建

在具体实现过程中，本文基于前文获取的 1842 条地铁运营相关视频文本，构建 BERTopic 主题聚类模型，对视频舆情文本进行主题识别分析。模型构建过程包括文本向量化、特征降维、文档聚类和主题表征四个步骤。

在文本向量化阶段，本文以视频描述的原始文本作为输入，选取 BAAI/bge-m3 作为底层语义嵌入模型，对每条文本生成高维文档嵌入向量，以充分保留上下文语义信息并避免分词处理可能带来的语义损失。在获得文档嵌入向量后，采用 UMAP (Uniform Manifold Approximation and Projection) 算法对高维向量进行降维映射，以提高聚类效率并尽可能保持文本之间的语义邻近关系。在文档聚类阶段，本文基于降维后的文本向量表示，采用层次聚类算法 (Agglomerative Clustering) 对视频描述文本进行聚类分析，从而将语义相近的文本划分至同一主题簇。相较于基于密度的聚类方法，层次聚类在主题数量可控性与聚类结构解释方面具有一定优势，更有利于后续主题体系的结构化分析。在主题表征阶段，本文对视频描述文本进行分词与清洗处理，以构建主题关键词统计所需的词汇集合。具体而言，采用 Jieba 分词工具对文本进行中文分词，然后结合哈尔滨工业大学停用词表与人工增补的停用词，对分词结果进行过滤，以剔除无实际语义或对建模产生干扰的词汇。在获得文本的主题归属关系后，基于 c-TF-IDF 方法分别计算各主题簇内部词汇的重要性权重，从而提取能够代表主题语义特征的关键词，并据此完成主题命名。

在完成上述建模流程后，BERTopic 模型对 1842 条地铁运营相关视频描述文本进行了主题识别分析，最终生成 24 个初始主题。上述主题构成了后续地铁运营舆情主题结构分析与解释的基础。

4.3. 地铁运营舆情主题识别结果分析

4.3.1. 初始主题识别结果

基于前文构建的 BERTopic 主题聚类模型，本文对 1842 条地铁运营相关视频描述文本进行了主题识别分析。模型运行结果显示(表 1)，共识别出 24 个语义区分度较高的初始主题，各主题在关键词构成、语义指向方面均表现出较为清晰的差异性，能够较好地反映抖音平台中地铁运营舆情的主要讨论内容。

从主题内容分布来看，识别出的初始主题覆盖了地铁运营舆情的多个方面。一方面，部分主题直接指向运营系统的基础属性，包括安全管理、设备运行与票价机制等；另一方面，围绕乘客出行过程与服务体验形成了多项细分主题，涉及站点可达性、进出站与换乘便捷性、信息引导与播报服务、无障碍设施配置等具体环节。此外，模型还识别出若干与乘车环境、技术应用、外部影响以及乘客行为相关的主题，反映出抖音平台地铁运营舆情在内容上的多元性与场景化特征。

从关键词构成情况看，各主题内部关键词在语义指向上具有较高一致性，不同主题之间的关键词组合差异较为明显，表明模型能够在当前数据规模与文本特征条件下实现较为有效的主题区分。

尽管初始主题在语义层面具有较好的可区分性，但由于主题数量相对较多，且部分主题在现实语义上存在内在关联，若直接基于初始主题开展后续分析，难以从整体层面把握地铁运营舆情的结构特征。因此，有必要在保持初始主题语义完整性的基础上，对其进行进一步的结构化归纳与整合。

Table 1. Initial topic identification results

表 1. 初始主题识别结果

序号	主题	关键词
1	安全事故	事故、安全、追尾、受伤
2	安检制度	安检、安全、检查、配合
3	设备故障	故障、延误、屏蔽门、对齐
4	优惠政策	免费、优待证、退役军人、优惠
5	基准票价	票价、便宜、涨价、成本
6	站点可达性	开通、方便、线路、背篓专线
7	进出站便捷性	闸机常开门、刷卡、进出、效率
8	乘车便捷性	大站快车、行车间隔、跨线运营、大小交路
9	换乘便捷性	换乘、电梯、通道、距离
10	导视服务水平	导视、标识、指引、系统
11	播报服务水平	播报、报站、声音、语言
12	无障碍服务便捷性	无障碍、残疾人、轮椅、导盲犬
13	其他服务便捷性	人性化、厕所、充电、设置
14	人流密度	客流量、拥挤、早高峰、晚高峰
15	热环境舒适性	温度、空调、强弱、冷暖
16	声环境舒适性	噪音、刺耳、车厢、内部
17	座椅舒适性	座椅、通勤、板凳族、舒服
18	智能化运营系统	无人驾驶、科技、全自动、首条
19	人文化服务	地铁站、文化、主题、设计

续表

20	周边噪音影响	噪音、隔音、扰民、入睡
21	周边房价影响	房价、通车、利好、规划
22	购票及逃票行为	逃票、冒用、查验、处罚
23	文明礼让行为	让座、先下后上、爱心、素质
24	不安全行为	冲门、屏蔽门、夹住、危险

4.3.2. 主题体系的结构化归纳

为提升主题识别结果的整体性与解释性，本文在初始主题识别的基础上，对 24 个主题进行结构化归纳，构建多层次的地铁运营舆情主题体系。

在具体实施过程中，本文参考国家标准《城市公共交通乘客满意度评价方法》(GB/T 36962.3-2018)所提出的评价维度体系，对初始主题进行归类整合。该标准从服务安全性、可靠性、经济性、便捷性、舒适性等多个方面系统刻画了城市公共交通服务质量，为理解公众对地铁运营的评价逻辑提供了较为成熟的理论参照。基于上述标准，本文将语义关联度较高、现实指向相近的初始主题进行归并，最终形成由 8 个主题大类和 24 个具体主题构成的地铁运营舆情主题体系，其结构如表 2 所示。

总体来看，该主题体系在保留初始主题细分特征的基础上，实现了从碎片化舆情议题向结构化分析维度的整合提升，有助于从整体层面梳理公众对地铁运营的主要关注方向及其内在逻辑。通过引入层级的主题结构，地铁运营舆情的内容构成与主题边界得以更加清晰地呈现，为后续围绕地铁运营舆情主题体系展开系统分析提供了结构基础。

Table 2. Topic system of metro operation public opinion

表 2. 地铁运营舆情主题体系

主题大类	主题	关键词	视频数量
服务安全性	安全事故	事故、安全、追尾、受伤	48
	安检制度	安检、安全、检查、配合	43
服务可靠性	设备故障	故障、延误、屏蔽门、对齐	72
服务经济性	优惠政策	免费、优待证、退役军人、优惠	76
	基准票价	票价、便宜、涨价、成本	57
服务便捷性	站点可达性	开通、方便、线路、背篓专线	229
	进出站便捷性	闸机常开门、刷卡、进出、效率	70
	乘车便捷性	大站快车、行车间隔、跨线运营、大小交路	137
	换乘便捷性	换乘、电梯、通道、距离	39
	导视服务水平	导视、标识、指引、系统	60
	播报服务水平	播报、报站、声音、语言	93
	无障碍服务便捷性	无障碍、残疾人、轮椅、导盲犬	126
服务舒适性	其他服务便捷性	人性化、厕所、充电、设置	64
	人流密度	客流量、拥挤、早高峰、晚高峰	127
	热环境舒适性	温度、空调、强弱、冷暖	57
	声环境舒适性	噪音、刺耳、车厢、内部	61

续表

	座椅舒适性	座椅、通勤、板凳族、舒服	49
服务创新性	智能化运营系统	无人驾驶、科技、全自动、首条	66
	人文化服务	地铁站、文化、主题、设计	115
外部影响	周边噪音影响	噪音、隔音、扰民、入睡	40
	周边房价影响	房价、通车、利好、规划	57
乘客行为	购票及逃票行为	逃票、冒用、查验、处罚	53
	文明礼让行为	让座、先下后上、爱心、素质	65
	不安全行为	冲门、屏蔽门、夹住、危险	38

4.3.3. 主题体系分析

从语义结构与舆情内容来看，地铁运营舆情主题体系整体呈现出以服务评价为核心、向社会影响与价值认知延伸的特征。关注点不仅包含运营效率与服务质量本身，更进一步触及地铁系统与城市发展之间的深层互动关系。

首先，在服务基础属性维度，“服务安全性”、“服务可靠性”、“服务经济性”、“服务便捷性”与“服务舒适性”构成了舆情讨论的底层框架。在“服务安全性”维度中，“安全事故”与“安检制度”主题表现出较高的舆论敏感性，关键词如“事故”、“追尾”、“受伤”等集中反映了公众对运营安全底线的关切。“服务可靠性”维度中，“设备故障”主题直接映射了公众对于地铁运营系统稳定性的关注。“服务经济性”方面，公众围绕“优惠政策”与“基准票价”两个主题展开讨论，“免费”、“优待证”、“退役军人”等关键词反映了地铁作为公共交通工具，其公益属性与定价机制在社会舆论中具有极高的讨论权重。“服务便捷性”维度的讨论覆盖乘客出行的全流程，既包括进出站、乘车、换乘等环节，也涉及导视、播报等信息系统；此外，“轮椅”、“无障碍”、“导盲犬”等关键词也表明公众对地铁服务的评价逻辑已由普适化的效率需求转向差异化的人文关怀。“服务舒适性”方面，舆情聚焦于物理环境体验，“客流量”、“拥挤”、“早晚高峰”指向运能压力与拥挤痛点，而“空调”、“冷暖”、“噪音”则体现了公众对车厢内部环境精细化管理的期待。

其次，“服务创新性”与“外部影响”维度体现出短视频舆情所特有的信息广度与公众参与特征。在“服务创新性”方面，“智能化运营系统”和“人文化服务”的主题讨论，反映出前沿技术应用与文化表达对地铁品牌与城市形象的塑造作用。

“外部影响”维度主要涵盖“周边噪音影响”和“周边房价影响”两个主题。“房价”、“噪音”、“扰民”等关键词表明，公众对地铁运营的讨论并不局限于乘车过程本身，而是进一步延伸至地铁系统对周边社区、居住环境与城市空间价值的外溢影响。不过，需要指出的是，这类主题在抖音平台上的出现频率，并不能被简单等同于现实社会中相关问题的重要性排序。尤其是“周边房价影响”主题，可能在一定程度上受到平台用户结构与传播机制的共同塑造：一方面，抖音用户中较为活跃的年轻群体、城市发展关注者及投资导向型内容消费者，更倾向于围绕“区位利好”和“房价变化”等议题进行表达；另一方面，涉及房价预期与城市发展红利的话题本身具有较强的话题性与传播性，更容易获得平台推荐与互动反馈。因此，该主题既反映了地铁建设外部效应进入公众讨论视野，也可能带有明显的平台可见性偏差。

“乘客行为”维度聚焦于个体在轨道交通公共空间内的动态表现，涵盖了“购票及逃票行为”、“文明礼让行为”及“不安全行为”三个核心主题，表明乘客的个体行为已演变为影响运营秩序与社会感知的关键因素。这不仅揭示了地铁密闭空间内个体行为与制度管理间的深度博弈，更折射出新媒体舆情环

境下，城市公共交通在社会治理与文明引导方面面临的复杂挑战。

综上，基于抖音数据构建的地铁运营舆情主题体系具有多维度、场景化、体验导向的特点。公众讨论既覆盖运营全环节，也深入人文关怀与社会价值层面。此地铁运营舆情主题体系为地铁运营部门开展精细化服务改善、舆情响应与公众沟通提供了清晰的决策依据与切入点。

4.3.4. 与外部资料的交叉比对

为增强研究结论的可信度，本文进一步将抖音平台识别出的地铁运营舆情主题，与官方发布的轨道交通服务质量评价结果进行交叉比对。以广州市交通运输局发布的《2025 年度广州城市轨道交通服务质量评价结果》为例，该评价基于第三方评估展开，覆盖乘客满意度、服务保障能力和运营服务关键指标等维度。评价结果显示，广州城市轨道交通整体服务水平较高，但在“通风与温度”、“列车运行平稳性及噪声”、“候车乘车秩序”等方面仍存在优化空间；同时，官方还提出应进一步优化标志标识设置，提升设施设备精细化运维水平。

这一结果与本文基于抖音视频文本识别出的主题具有较强一致性。首先，官方指出“列车运行平稳性及噪声”仍有改进空间，这与本文主题体系中的“声环境舒适性”高度呼应，说明乘客在短视频平台中表达的不适体验，并非纯粹由个别情绪性内容构成，而是与正式服务评价中呈现的问题相互印证。其次，官方提到“候车乘车秩序”仍需优化，也与本文识别出的“人流密度”相关讨论形成对应，反映出高峰拥挤、候车体验及秩序管理确为公众持续关注的运营议题。再次，官方强调“设施设备精细化运维”仍需加强，这与抖音舆情中“设备故障”等主题具有一致性，说明设施可靠性问题同样在民间表达与官方评价中被反复提及。

不过，二者之间也存在一定差异。官方评价整体上更强调服务水平持续提升、关键指标保持高位以及运营管理制度的完善，其叙述方式更偏向制度化、综合性的“管理话语”；而抖音平台中的相关讨论则更多体现为具体场景中的个体体验和即时情绪，例如拥挤不适、噪声干扰、设备失灵带来的不便等。官方文本倾向于在总体肯定中指出“优化空间”，而平台舆情则更容易聚焦负面体验和具体痛点，这说明二者在问题呈现方式上存在明显差别。前者反映的是运营治理视角下的综合评估，后者则更接近乘客日常出行中的感受表达和体验叙事。

这种一致性与差异性表明，抖音平台舆情具有一定现实基础，能够在相当程度上捕捉地铁运营中的真实服务短板；同时，它又受到平台传播机制与用户表达偏好的影响，更容易放大具象化、情绪化和冲突性较强的问题。因此，将短视频平台舆情与官方服务评价资料结合起来，有助于更全面地理解地铁运营中的问题结构：前者揭示“乘客如何感知问题”，后者体现“管理部门如何界定问题”，两者之间的对照也反映出官方话语与民间话语之间既相互印证又存在张力的关系。

5. 结论

本研究以抖音平台为数据源，利用 Python 爬虫技术获取 2018 年至 2025 年间地铁运营相关的 1842 条视频。在数据预处理基础上，采用基于预训练语言模型的 BERTopic 主题聚类方法，对视频描述文本进行主题建模，识别出 24 个具有明确语义指向的初始主题。进一步参考国家标准《城市公共交通乘客满意度评价方法》(GB/T 36962.3-2018)，对初始主题进行结构化归纳，构建了由 8 个主题大类和 24 个具体主题构成的地铁运营舆情主题体系，并对各主题的语义内涵与舆情特征进行了系统分析。研究主要得出以下结论：

第一，地铁运营舆情在抖音平台上呈现出多元化、场景化与体验导向的特征。识别出的主题覆盖了从运营系统基础属性(安全性、可靠性、经济性)到乘客出行全过程(便捷性、舒适性)，再到外部影响与价值认知(创新性、外部影响、乘客行为)等多个层面，反映出公众对地铁运营的关注已超越单纯的出行功能，

逐步延伸至人文关怀、技术创新与社会治理等更深层次的价值维度。

第二，从主题结构来看，公众讨论的重心集中于服务便捷性维度。其中，“站点可达性”、“乘车便捷性”等子主题具有较高的讨论热度，表明出行效率与乘坐体验是公众最敏感的关切点。同时，“安全事故”、“基准票价”、“不安全行为”等主题的情感倾向整体偏低，构成了地铁运营舆情的核心痛点与潜在风险点，需在运营管理中予以优先关注与系统应对。

第三，乘客行为维度的识别揭示了社交媒体舆情环境下城市公共交通面临的社会治理挑战。“购票及逃票行为”、“文明礼让行为”、“不安全行为”等主题的涌现，表明乘客个体行为已从单纯的个体选择演化为影响运营秩序与社会评价的关键变量。公众在讨论中同时扮演着行为观察者、规范呼吁者与治理参与者的多重角色，为运营方完善乘客引导、开展公众教育与构建共建共治的运营环境提供了舆情依据。

第四，本研究构建的地铁运营舆情主题体系具有较强的结构性与解释力。该体系将碎片化的舆情议题整合为层次分明、逻辑清晰的分析框架，能够系统揭示公众对地铁运营的关注焦点、体验诉求与价值诉求，为运营管理部门开展精细化服务改善、舆情监测预警与公众沟通策略制定提供了明确的决策依据与切入点。

总体而言，本研究通过引入基于深度学习的大数据文本分析方法，拓展了地铁运营舆情研究的视野与方法路径，揭示了短视频平台中公众对地铁运营的认知图景与情感结构。未来研究可在以下方面进一步深化：一是融合视频画面、音频等多模态数据，开展更为全面的舆情分析；二是进一步引入更细粒度的时间序列分析方法，追踪不同主题在重大运营事件、节假日出行高峰及政策调整节点下的动态演化规律；三是拓展多平台对比分析，比较不同社交媒体平台中地铁运营舆情的异同特征，以构建更具普适性的地铁运营舆情认知框架。

参考文献

- [1] Guo, Y., Tang, Z. and Guo, J. (2020) Could a Smart City Ameliorate Urban Traffic Congestion? A Quasi-Natural Experiment Based on a Smart City Pilot Program in China. *Sustainability*, **12**, Article 2291. <https://doi.org/10.3390/su12062291>
- [2] Yang, J., Chen, S., Qin, P., Lu, F. and Liu, A.A. (2018) The Effect of Subway Expansions on Vehicle Congestion: Evidence from Beijing. *Journal of Environmental Economics and Management*, **88**, 114-133. <https://doi.org/10.1016/j.jeem.2017.09.007>
- [3] Zhang, J., Xu, X., Hong, L., Wang, S. and Fei, Q. (2011) Networked Analysis of the Shanghai Subway Network, in China. *Physica A: Statistical Mechanics and Its Applications*, **390**, 4562-4570. <https://doi.org/10.1016/j.physa.2011.06.022>
- [4] Deng, Y., Zhang, Y., Yuan, Z., Li, R.Y.M. and Gu, T. (2023) Analyzing Subway Operation Accidents Causations: Apriori Algorithm and Network Approaches. *International Journal of Environmental Research and Public Health*, **20**, Article 3386. <https://doi.org/10.3390/ijerph20043386>
- [5] 谭启生. 公共交通购票支付的便利化研究——以广州地铁为例[D]: [硕士学位论文]. 广州: 广东财经大学, 2022.
- [6] 赵晓利. 基于用户空间体验的地铁导向标识系统设计研究[D]: [硕士学位论文]. 济南: 山东大学, 2018.
- [7] Wu, M., Long, R., Chen, F., Chen, H., Bai, Y., Cheng, K., et al. (2023) Spatio-Temporal Difference Analysis in Climate Change Topics and Sentiment Orientation: Based on LDA and BiLSTM Model. *Resources, Conservation and Recycling*, **188**, Article 106697. <https://doi.org/10.1016/j.resconrec.2022.106697>
- [8] Hu, H., Ge, Y. and Hou, D. (2014) Using Web Crawler Technology for Geo-Events Analysis: A Case Study of the Huangyan Island Incident. *Sustainability*, **6**, 1896-1912. <https://doi.org/10.3390/su6041896>
- [9] Gao, N., Touran, A., Wang, Q. and Beauchamp, N. (2024) Construction Risk Identification Using a Multi-Sentence Context-Aware Method. *Automation in Construction*, **164**, Article 105466. <https://doi.org/10.1016/j.autcon.2024.105466>
- [10] Li, S., You, M., Li, D. and Liu, J. (2022) Identifying Coal Mine Safety Production Risk Factors by Employing Text Mining and Bayesian Network Techniques. *Process Safety and Environmental Protection*, **162**, 1067-1081. <https://doi.org/10.1016/j.psep.2022.04.054>

- [11] Wang, Y., Li, H. and Wu, Z. (2019) Attitude of the Chinese Public toward Off-Site Construction: A Text Mining Study. *Journal of Cleaner Production*, **238**, Article 117926. <https://doi.org/10.1016/j.jclepro.2019.117926>
- [12] Estévez-Ortiz, F.J., García-Jiménez, A. and Glösekötter, P. (2016) An Application of People's Sentiment from Social Media to Smart Cities. *El Profesional de la Información*, **25**, 851-858. <https://doi.org/10.3145/epi.2016.nov.02>
- [13] Ghosh, S., Chen, Y. and Dou, W. (2024) Railroad Safety: A Systematic Analysis of Twitter Data. *Case Studies on Transport Policy*, **15**, Article 101154. <https://doi.org/10.1016/j.cstp.2024.101154>
- [14] 刘洋, 柳卓心, 金昊, 等. 基于 BERTopic 模型的用户层次化需求及动机分析——以抖音平台为例[J]. 情报杂志, 2023, 42(12): 159-167.
- [15] 沈洪洲, 朱佳, 黄仕靖, 等. 应急知识短视频传播效果研究: 基于不同发布端类型的分析[J]. 情报理论与实践, 2024, 47(11): 101-110.
- [16] Anderson, M.L. (2014) Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion. *American Economic Review*, **104**, 2763-2796. <https://doi.org/10.1257/aer.104.9.2763>
- [17] Reddy, A.V., Lu, A. and Wang, T. (2010) Subway Productivity, Profitability, and Performance: A Tale of Five Cities. *Transportation Research Record: Journal of the Transportation Research Board*, **2143**, 48-58. <https://doi.org/10.3141/2143-07>
- [18] Gallo, M. (2018) The Impact of Urban Transit Systems on Property Values: A Model and Some Evidences from the City of Naples. *Journal of Advanced Transportation*, **2018**, 1-22. <https://doi.org/10.1155/2018/1767149>
- [19] He, D., Sun, G., Li, L. and Webster, C. (2024) New Metro and Housing Price and Rent Premiums: A Natural Experiment in China. *Urban Studies*, **61**, 1371-1392. <https://doi.org/10.1177/00420980231208560>
- [20] Li, S., Chen, L. and Zhao, P. (2019) The Impact of Metro Services on Housing Prices: A Case Study from Beijing. *Transportation*, **46**, 1291-1317. <https://doi.org/10.1007/s11116-017-9834-7>
- [21] Xu, T., Zhang, M. and Aditjandra, P.T. (2016) The Impact of Urban Rail Transit on Commercial Property Value: New Evidence from Wuhan, China. *Transportation Research Part A: Policy and Practice*, **91**, 223-235. <https://doi.org/10.1016/j.tra.2016.06.026>
- [22] Li, W., Sun, B., Yin, C., Zhang, T. and Liu, Q. (2018) Does Metro Proximity Promote Happiness? Evidence from Shanghai. *Journal of Transport and Land Use*, **11**, 1271-1285. <https://doi.org/10.5198/jtlu.2018.1286>
- [23] Li, K., Yuan, W. and Li, J. (2022) Causal Association between Metro Transits and Air Quality: China's Evidence. *Environmental Science and Pollution Research*, **29**, 70435-70447. <https://doi.org/10.1007/s11356-022-20724-x>
- [24] Sun, C., Zhang, W., Luo, Y. and Xu, Y. (2019) The Improvement and Substitution Effect of Transportation Infrastructure on Air Quality: An Empirical Evidence from China's Rail Transit Construction. *Energy Policy*, **129**, 949-957. <https://doi.org/10.1016/j.enpol.2019.03.005>
- [25] Li, Q., Song, L., List, G.F., Deng, Y., Zhou, Z. and Liu, P. (2017) A New Approach to Understand Metro Operation Safety by Exploring Metro Operation Hazard Network (MOHN). *Safety Science*, **93**, 50-61. <https://doi.org/10.1016/j.ssci.2016.10.010>
- [26] Li, Z., Zhang, S., Meng, Q. and Hu, X. (2021) Barriers to the Development of Prefabricated Buildings in China: A News Coverage Analysis. *Engineering, Construction and Architectural Management*, **28**, 2884-2903. <https://doi.org/10.1108/ecam-03-2020-0195>
- [27] Tang, L., Zhang, Y., Dai, F., Yoon, Y., Song, Y. and Sharma, R.S. (2017) Social Media Data Analytics for the U.S. Construction Industry: Preliminary Study on Twitter. *Journal of Management in Engineering*, **33**, Article 04017038. [https://doi.org/10.1061/\(asce\)me.1943-5479.0000554](https://doi.org/10.1061/(asce)me.1943-5479.0000554)
- [28] Hong, L., Fu, C., Wu, J. and Frias-Martinez, V. (2018) Information Needs and Communication Gaps between Citizens and Local Governments Online during Natural Disasters. *Information Systems Frontiers*, **20**, 1027-1039. <https://doi.org/10.1007/s10796-018-9832-0>
- [29] Xue, J., Shen, G.Q., Li, Y., Wang, J. and Zafar, I. (2020) Dynamic Stakeholder-Associated Topic Modeling on Public Concerns in Megainfrastructure Projects: Case of Hong Kong-Zhuhai-Macao Bridge. *Journal of Management in Engineering*, **36**, Article 04020078. [https://doi.org/10.1061/\(asce\)me.1943-5479.0000845](https://doi.org/10.1061/(asce)me.1943-5479.0000845)
- [30] Tang, L., Griffith, L., Stevens, M. and Hardie, M. (2020) Social Media Analytics in the Construction Industry Comparison Study between China and the United States. *Engineering, Construction and Architectural Management*, **27**, 1877-1889. <https://doi.org/10.1108/ecam-12-2019-0717>
- [31] 翟羽佳, 马柏玉, 陈晓灿. 新冠疫苗信息扩散的共振与联动效应——基于抖音与微博的对比研究[J]. 情报科学, 2023, 41(12): 127-135.
- [32] 於世为, 郭迎迎, 陆永香, 等. “拉闸限电”事件原因和影响解析: 公众评论大数据驱动视角[J]. 工程管理科技前沿, 2024, 43(1): 30-38.