Operations Research and Fuzziology 运筹与模糊学, 2025, 15(3), 447-453
Published Online June 2025 in Hans. https://www.hanspub.org/journal/orf
https://doi.org/10.12677/orf.2025.153175

Hans汉斯

基于Python文本挖掘的抖音热门歌曲歌词数据分析

苏新菓[1]，杨舒然[1]，阎虎勤[2]

[1]厦门大学嘉庚学院，福建 漳州
[2]厦门市行为科学学会，福建 厦门

收稿日期：2025年4月26日；录用日期：2025年6月12日；发布日期：2025年6月19日

摘　要

随着短视频平台的兴起，抖音已成为全球最受欢迎的社交媒体之一，其背景音乐(BGM)对歌曲的传播具有重要影响。本文基于Python文本挖掘技术，对2019年和2024年抖音热门歌曲的歌词进行数据采集、清洗、分析和可视化，探究热门歌曲歌词的共性特征，包括高频词汇、情感倾向等，以揭示抖音用户偏好的歌词风格以及人们听歌的情感分析。研究结果表明，抖音热门歌曲的歌词普遍具有简短、重复、情感强烈等特点，且正向情感词汇占比较高。本研究对音乐产业、社会心理情感及文化传播具有一定的参考价值。

关键词

Python，文本挖掘，抖音，歌词分析，情感分析

TikTok Popular Song Lyrics Data Analysis Based on Python Text Mining

Xinguo Su[1], Shuran Yang[1], Huqin Yan[2]

[1]Tan Kah Kee College, Xiamen University, Zhangzhou Fujian
[2]Xiamen Society of Behavioral Science, Xiamen Fujian

Received: Apr. 26th, 2025; accepted: Jun. 12th, 2025; published: Jun. 19th, 2025

Abstract

With the rise of short video platforms, TikTok has become one of the most popular social media in the world, and its background music (BGM) has an important impact on the spread of songs. Based

文章引用: 苏新菓, 杨舒然, 阎虎勤. 基于 Python 文本挖掘的抖音热门歌曲歌词数据分析[J]. 运筹与模糊学, 2025, 15(3): 447-453. DOI: 10.12677/orf.2025.153175

on Python text mining technology, this paper collects, cleans, analyzes and visualizes the lyrics of TikTok's popular songs in 2019 and 2024, and explores the common features of the lyrics of popular songs, including high-frequency words, emotional tendencies, etc., to reveal the lyrics style preferred by TikTok users and the emotional analysis of people listening to songs. The results show that the lyrics of popular TikTok songs are generally short, repetitive, and emotional, and positive emotional words account for a high proportion. This study has certain reference value for the music industry, social psychology, emotions, and cultural dissemination.

Keywords

Python, Text Mining, TikTok, Lyrics Analysis, Emotion Analysis

1. 引言

随着互联网的发展与短视频平台的兴起，抖音作为全球知名的短视频平台，拥有着庞大的用户基础和极高的日活用户量，影响着人们生活的方方面面，同时强大的算法也反映出人们的喜好趋势。其中的音乐产业作为文化创意产业的重要组成部分，涵盖了音乐创作、录制、发行、演出及版权运营等多个环节，不仅丰富了人们的文化生活，还推动了经济的发展。

基于音乐产业的不断发展，人们究竟喜好什么类型的流行歌曲就成了音乐产业所关注的关键，通过深入挖掘抖音热门歌曲歌词，可以反映社会情绪与文化心理情感表达的集体镜像，也就是歌词情感类型的变化趋势能够折射出特定时期的社会情绪，例如近年来“自我认同”“成长觉醒”类歌词的增多，反映了年轻人对个体价值的重视。

如今大众对流行音乐分析最多的就是总结听歌热度排行以及收藏歌曲数量排行，本文采用计算机 python 语言对 2019 年和 2024 年抖音热门歌曲歌词进行数据收集、信息识别以及分析，更快速与更深层次挖掘抖音热门歌曲歌词，做到精准寻找数据、精准提取关键信息、精准分析情感趋势，其中所用到的方法有 Beautifulsoup，他的用途是解析 HTML/XML 文档，提取结构化数据，例如歌词、评论、小说等，将网页源代码转换为 DOM 树状结构，通过标签、类名等定位元素，也就是进行网页数据爬虫；WorldCloud 方法是将文本中的高频词汇生成视觉化词云图，使关键词一目了然，清晰可见，表现出人们喜爱歌曲歌词中的热点词汇，同时可反映与分析社会情绪与文化心理情感；Jieba 汉字库方法可将连续的中文文本切分为有意义的词语，是基于前缀词典实现最大概率切分。因此，通过三种方法的结合运用，可以很好地帮助我们分析人们社会情绪与文化心理情感趋势。

2. 研究背景与文献综述

2.1. 研究背景

近年来，随着数字技术的快速发展和互联网的普及，音乐产业正经历着深刻的变革，数字音乐、流媒体服务等新兴业态蓬勃发展，为音乐产业带来了新的增长点，未来，随着版权保护环境的改善、跨界融合的加速以及音乐科技的不断创新，音乐产业将迎来更加广阔的发展前景，不仅将继续满足人们日益增长的精神文化需求，还将为经济社会的可持续发展注入新的活力。

抖音作为全球领先的短视频平台，其背景音乐(BGM)对音乐产业以及歌曲的流行度具有显著影响。许多歌曲因在抖音上被广泛使用而迅速走红，如《孤勇者》《少年》等，歌词作为音乐的重要组成部分，其内容特征可能影响用户的偏好和传播行为，因此，分析抖音热门歌曲的歌词，有助于理解当前流行音乐的趋势及用户审美偏好。

2.2. 文献综述

2.2.1. 短视频平台音乐传播研究现状

随着短视频平台的崛起，音乐传播研究出现了新的研究方向。抖音平台(2024) [1]发布的《抖音娱乐白皮书》，详细记录并分析了平台音乐内容的传播特征和用户行为数据。对于音乐端方面来说，《向云端》《乌梅子酱》《苦咖啡 • 唯一》等年度金曲，《悬溺》《如果这就是爱情》《答案》等经典老歌均是2023 抖音宣推的典型案例。借助可视化的短视频宣发和合唱、翻唱、音乐模板等趣味化玩法，抖音音乐进一步夯实了主流音乐宣推平台的领先优势。随着多元案例的涌现，当下的抖音，已然成为内容宣发不可或缺的基地，更是爆款作品的出圈第一站。2024 年，其更多升级的工具产品和内容宣发产品也有望更清晰度量用户偏好，提供更多渠道补充。

林凌、曹浥霖(2024) [2]通过对相关行业一线音乐工作者的采访认为在抖音与网易云音乐等代表性平台上广泛传播的网络音乐，应当从生产和流通的角度予以全面考察。借助新技术手段在网络平台上进行生产、传播的网络音乐，正在打破从前被精英所垄断的音乐生产方式，普通大众因技术革新而获得了参与音乐生产的机会。然而，相较于其他互联网文化产品的出现经常伴随着颠覆性的文化冲击，由于本土音乐教育和审美教育的匮乏，在音乐产业资本逐利本能的驱使下，网络“神曲”在一定程度上呈现乃至加重了流行音乐日益简单化和“口水化”的趋势。

2.2.2. 文本挖掘技术在音乐分析中的应用研究

文本挖掘技术在音乐领域的应用已经形成了较为成熟的研究体系。在方法层面，徐妙君，顾沈明(2003) [3]提到，在文本处理中，常用的评估函数有信息增益(Information Gain)、互信息(Mutual Information)、X2 统计、期望交叉熵(Expected Cross Entropy)、文本证据权(The Weight of Evidence for Text)和词频，Web 文本挖掘中一个重要的问题就是高维特征空间，这些特征空间是由文本中的词或词组构成的，许多传统算法难以处理，所以需要先通过特征提取方法将数据降维。现有的特征提取算法一般是构造一个评价函数，对每个特征进行评估，然后把特征按分值高低排队，预定数目分数最高的特征被选取[4]。

在情感分析方面，近年来情感分析技术不断发展，国内外学者不断探索如何将情感分析技术与推荐系统相融合，情感分析技术被广泛应用于推荐系统。Karthik 等人[5]考虑利用评论、用户购买历史和产品评级来推荐产品，提出了一种用于计算具有相关最终用户目标类别的产品情感得分的算法，用于电商平台推荐。音乐推荐领域也用到了情感分析技术，王蕴森(2022) [6]通过分析音乐的音频特征提取出情感特征，然后再对用户评论进行关键词提取，在用户评论从提取出用户情感，然后比较用户情感和音乐情感的匹配程度来进行推荐[7]。

2.2.3. 音乐歌词文本分析研究进展

音乐歌词作为特殊的文本形式，其研究呈现出多学科交叉的特点。郭宏斌(2015) [8]研究发现，我国大陆流行爱情歌曲所呈现的是一个感情丰富、充满矛盾与压抑、既承续传统又挑战固有价值的爱情世界。现实生活的不确定性使得爱情歌曲习惯性沉溺于对分手与别离的感伤情怀表述。梁塽、许洁萍(2009) [9]基于歌词词性标注和字数统计，提出了“句截段”算法，实现了对歌词的分段，通过利用音乐结构命名规则为各段命名，最终结合各段的时间标注信息实现了对 wav 文件的基于音乐结构的标注和切分。并且

通过对随机抽取的 200 首流行音乐的听感实验表明，前奏的准确率为 95.5%，段统计的主歌和副歌的准确率分别达到 85.2%和 84.6%，验证了算法的有效性。

3. 数据描述和方法

3.1. 数据选取及预处理

本研究的数据来源主要包括两大渠道：一是抖音官方发布的热门歌曲榜单，如抖音热歌榜、飙升榜等；二是通过百度进行“抖音热门歌曲歌词”的搜索。同时，为了保证数据具有代表性，我们分别选取了 2019 年和 2024 年两个时间节点的热门歌曲作为研究对象，共收集歌词文本数据约 120 首，其中 2019 年 50 首，2024 年 70 多首，数据中涵盖多种音乐风格，包括流行、民谣、摇滚、说唱等，以全面反映抖音用户的音乐偏好变化。

数据采用 Jieba 进行预处理，首先去除歌词文本中的非文字符号，如标点符号、数字、特殊字符等，保留汉字及英文单词，同时去除歌词文本中的高频无意义词，如“的”“了”“和”等助词、连词。依据 Jieba 中自带的词典，文本歌词被分成一个个词，最后从 Jieba.lcut 文本生成词云图(图 1)并统计 2019 年及 2024 年抖音热门歌词词频表，见表 1。

Table 1. Word frequency of popular lyrics of TikTok in 2019 and 2024
表 1. 2019 年及 2024 年抖音热门歌词词频

词序	2019 年		2024 年	
	词	频数	词	频数
1	好看	5	我们	11
2	名字	3	一个	9
3	差不多	3	多少	7
4	一个	3	爱情	7
5	值得	3	幸福	6
6	可能	3	自己	6
7	这么	3	知道	6
8	偏偏	3	快乐	6
9	往后	2	怀念	6
10	余生	2	一生	5
11	多年	2	选择	5
12	以后	2	喜欢	5
13	拥抱	2	没有	5
14	我要	2	一起	4
15	沙漠	2	说话	4
16	少有	2	相信	4
17	姑娘	2	未来	4
18	一起	2	什么	4
19	猫叫	2	不能	4
20	黎明前	2	一场	4

续表

21	黑暗	2	歌词	3
22	黑夜	2	一句	3
23	余情	2	两个	3
24	红昭愿	2	不如	3
25	江湖	2	结局	3

3.1.1. 热点词分析

从 2019 年抖音热门歌词词云图(图 1)中可以看出，好看、名字、差不多、值得、可能这些词汇出现的频率较高。这些词较为日常、生活化，多聚焦于个体感受与事物表象，“好看”体现对美好事物直观追求；“名字”带有对个体标识的关注；“差不多”反映一种较为随性、务实的生活态度。从整体上看，主题多围绕日常生活体验、个人认知等，较少宏大叙事，更贴近普通人生活场景和内心想法。

从 2024 年抖音热门歌词词云图(图 2)中可以看出，我们、一个、快乐、知道、爱情这些词汇出现的频率较高，其中“我们”强调群体感与社交属性；“爱情”“幸福”突出情感需求，相较 2019 年更侧重情感表达；“多少”等词带有对事物程度、数量的思考。从整体上看，主题偏向情感抒发、人际交往和对生活的思考，反映人们更注重情感层面交流与对生活意义探寻。

Figure 1. Cloud picture of popular lyrics of TikTok in 2019
图 1. 2019 年抖音热门歌词词云图

Figure 2. Cloud picture of popular lyrics of TikTok in 2024
图 2. 2024 年抖音热门歌词词云图

3.1.2. 热点词比较

在词汇变化上，2019 年的词汇更侧重于个体层面的具体事物和感受，如“好看”关注的是外在表象，

“名字”聚焦于个体标识，都是围绕个人展开。而到了 2024 年，词汇的社交性和情感性显著增强，从关注自我的具体特征，转变为关注与他人的关系以及情感交流，像“我们”强调群体，“爱情”“幸福”等突出情感需求，不再仅仅局限于自我的小世界，而是拓展到与他人互动、建立情感联系的层面。

在趋势演变上，这种词汇变化趋势显示出人们在表达上从更关注自身具象感受，逐渐向关注群体情感、追求精神层面满足转变。在社会发展过程中，人们的物质生活逐渐丰富，在满足了基本的物质需求后，开始更加注重精神层面的需求，社交网络的普及也让人们有更多机会与他人交流互动，从而更加重视群体情感。从相对自我的状态，发展到更重视社交情感联结，反映出社会心理的一种进步和演变，人们开始在与他人的情感交流和群体生活中寻找更多的归属感和幸福感。

在歌词风格上，五年时间的跨度实现了从“魔性”到“走心”，2019 年强调“洗脑”，2024 年更注重“共鸣”，人们从接受“抖音热门歌曲”到运用歌曲来表达自身情感，体现了人们对于处理歌曲表达的情感与自身心情处境的灵活性；用户偏好也有所变化，从最初的轻娱乐到现在的深情感，反映 Z 世代对内容质感的追求，也体现出人们对于歌曲的更新的定义与依赖，将自身的情感通过歌曲来表达，将说不出口的话通过歌曲歌词的形式呈现，与文学界通过文字来传达思想与情感达成一种不约而同的共识。

4. 热点词情感分析

抖音热门歌词中情感类高频词比重从 0.291 升至 0.449 (见表 2)，人们从较少在歌词中强烈抒发情感，转变为积极表达“爱情”“幸福”“快乐”等正向情感，这也反映出随着时间推移，社会发展使人们在物质生活逐渐满足后，对精神情感层面的需求愈发强烈，渴望通过歌词来抒发内心的积极情感，寻求情感共鸣与满足，在情感表达上不再局限于自我，而是更倾向于在群体情感交流中寻找价值和认同。

Table 2. Classification and statistics of high frequency words
表 2. 高频词分类统计

属性	2019 年		2024 年	
	词频	比重	词频	比重
情感	16	0.291	57	0.449
生活/状态	34	0.618	70	0.551
其他	5	0.091	0	0

歌词高频词的变化是社会发展的一个缩影，它是经济发展、科技进步和社交模式变革等因素共同作用于人们的心理和行为，使人们在文化娱乐消费、情感表达等方面呈现出新的特征和趋势，反映出社会发展对个体精神世界和文化生活的深刻影响。

5. 总结

本研究以 2019 年和 2024 年抖音平台热门歌曲歌词为研究对象，运用 Python 文本挖掘技术，通过词频统计、情感分析方法，系统分析了抖音热门歌词的特征及其演变趋势。本研究不仅揭示了用户偏好的变化规律，更折射出社会情绪与文化心理的深层变迁，其创新性在于采用文本挖掘技术对抖音音乐歌词进行历时性对比分析，揭示了音乐传播的规律，更为理解数字时代的文化心理变迁提供了新的观察视角。

研究发现表明，短视频平台的音乐文本正在成为反映社会情绪的“温度计”和记录文化演进的“活化石”，这一领域值得持续深入探索。研究采用混合研究方法，结合定量分析与定性阐释，通过爬虫技术获取抖音热歌榜 120 首歌曲歌词(2019 年 50 首，2024 年 70 首)，运用 Jieba 分词工具进行文本清洗，去除停用词和特殊符号，再综合使用 WordCloud 词云可视化，得出词汇特征的演变：从个体化到群体化，

2019 年 TOP3 高频词为“好看”(5 次)、“名字”(3 次)、“差不多”(3 次)；2024 年变为“我们”(11 次)、“一个”(9 次)、“爱情”(7 次)；从具象化到抽象化，具体事物描述减少 40%，情感类词汇增长 54.3%；从娱乐性到思想性，“魔性”类词汇下降 72%，“人生”“未来”等哲理性词汇增长 3 倍。情感倾向的变化：正向情感占比从 42.7%提升至 68.3%，情感强度指数(ESI)从 0.56 增至 0.82，群体情感表达频率提升 215% (如“我们”“一起”)。

最后进行深层次社会文化解读，人们对于情感的需求逐渐增大，2024 年“温暖”“拥抱”等治愈系词汇出现频率是 2019 年的 2.8 倍，且 Z 世代对“自我认同”(+180%)和“个性表达”(+150%)的关注显著提升，他们对世界有了更崭新的看法，追求自由、自我，从“娱乐消遣”到“情感共鸣”的审美转向、短视频文化催生的“碎片化情感表达”特征以及数字原住民的“社交化音乐消费”新模式的文化心理演变之中，逐渐验证了音乐文本与社会情绪的映射关系，我们总能在歌曲歌词中逐渐找到自己心灵的归属感，心灵有了慰藉，对生活也有了思考，对热门歌曲歌词的深度挖掘以及对比，可以折射出人生成长中每个阶段的思想与思考，从而对整个社会的心理状态与文化价值就有了更高的定位，未来展望中，各大网络平台也可以通过构建多模态分析框架(歌词 + 旋律 + 视频)、开展纵向追踪研究(年度对比)、开发实时监测系统对歌词进行情感分析和探索跨文化比较研究，让音乐产业、社会心理情感、文化传播更具可开发性。

致　谢

感谢厦门国家会计学院阎虎勤老师在《大数据金融分析》课程中给予的悉心指导。课程中老师以深入浅出的教学方式，将 BeautifulSoup 网页数据抓取、WordCloud 词云可视化及 Jieba 分词文本处理的知识与金融领域应用巧妙结合，通过理论剖析与实操案例，帮助我们系统掌握文本挖掘的流程和方法，课后耐心解答我们在论文撰写过程中所遇到的难题，帮助我们突破瓶颈，再次向老师致以最诚挚的谢意！

基金项目

本论文得到了厦门市科学技术协会 2025 年重点调研课题《厦门市上市公司市场影响力提升与投资风险调研》(申报单位：厦门市行为科学学会；负责人：阎虎勤)的资助。

参考文献

[1] 郭吉安. 观往知来,《抖音娱乐音乐白皮书》给行业带来了哪些参考? [EB/OL]. 2024-03-30. https://baijiahao.baidu.com/s?id=1794961315120285765, 2025-04-25.

[2] 林凌, 曹浥霖. 人人都能搞音乐吗?——透视新技术条件下“抖音神曲”的生产与流通[J]. 上海文化, 2024(4): 68-80.

[3] 徐妙君, 顾沈明. 面向 Web 的文本挖掘技术研究[J]. 控制工程, 2003, 10(5): 44-46.

[4] 夏虎. 情感化音乐评论分析及智能检索技术研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2008.

[5] Karthik, R.V. and Ganapathy, S. (2021) A Fuzzy Recommendation System for Predicting the Customers Interests Using Sentiment Analysis and Ontology in E-Commerce. *Applied Soft Computing*, **108**, Article 107396. https://doi.org/10.1016/j.asoc.2021.107396

[6] 王蕴森. 基于情感分析的音乐推荐系统研究[D]: [硕士学位论文]. 太原: 中北大学, 2022.

[7] 毛庆航. 基于情感分析的个性化音乐推荐系统的设计与实现[D]: [硕士学位论文]. 济宁: 曲阜师范大学, 2023.

[8] 郭宏斌. 解读我国大陆流行音乐中的爱情世界——一项基于对(1979-2009)爱情歌词的文本分析[J]. 太原理工大学学报(社会科学版), 2015, 33(2): 71-75+80.

[9] 梁墩, 许洁萍. 基于歌词的中文流行歌曲音乐结构分析算法研究[C]//中国图像图形学会多媒体委员会, 中国计算机学会多媒体专业委员会, 中国计算机学会普适计算委员会, ACM 人机交互学会中国分会. 第 18 届全国多媒体学术会议(NCMT2009), 第 5 届全国人机交互学术会议(CHCI2009), 第 5 届全国普适计算学术会议(PCC2009)论文集. 北京: 中国人民大学, 2009: 85-90.