Hans Journal of Data Mining 数据挖掘, 2024, 14(4), 207-217
Published Online October 2024 in Hans. https://www.hanspub.org/journal/hjdm
https://doi.org/10.12677/hjdm.2024.144019

Hans汉斯

中国乡村形象在田园生活短视频中的国际构建

张　欣[1]，罗　琪[2]

[1]浙大城市学院外国语学院，浙江　杭州
[2]上海海事大学外国语学院，上海

收稿日期：2024年8月27日；录用日期：2024年9月20日；发布日期：2024年9月30日

摘　要

以李子柒为代表的中国田园生活类自媒体短视频，在国际社交媒体平台上受到海外网友的追捧和喜爱，为新时代我国如何推动中国文化走出去提供了参考路径和示范经验。运用卷积神经网络(CNN)和潜在狄利克雷分布(LDA)主题模型对李子柒短视频下的评论文本进行文本挖掘和情感分类，可快速识别文本中的情感倾向，了解海外受众在观看李子柒的视频后对中国文化产生的文化情感和文化认知，从而为我国主流媒体如何利用好海外社交媒体平台和短视频，向世界讲好中国乡村故事、传播中国乡土文化提供借鉴。

关键词

李子柒，评论文本，文本挖掘，情感分析，卷积神经网络

The International Construction of Chinese Rural Imagery in Pastoral Lifestyle Short Videos

Xin Zhang[1], Qi Luo[2]

[1]School of Foreign Languages, Hangzhou City University, Hangzhou Zhejiang
[2]College of Foreign Languages, Shanghai Maritime University, Shanghai

Received: Aug. 27th, 2024; accepted: Sep. 20th, 2024; published: Sep. 30th, 2024

Abstract

Represented by Li Ziqi, short videos about Chinese pastoral lifestyle have garnered praise and affection from overseas internet users on international social media platforms. They provide a reference and exemplary experience for how China can promote its culture abroad in the new era. By employing Convolutional Neural Networks (CNN) and Latent Dirichlet Allocation (LDA) topic models, text mining and sentiment classification are conducted on comments over Li Ziqi's short videos. This enables rapid identification of emotional tendencies in the text, understanding overseas

文章引用: 张欣, 罗琪. 中国乡村形象在田园生活短视频中的国际构建[J]. 数据挖掘, 2024, 14(4): 207-217.
DOI: 10.12677/hjdm.2024.144019

audiences' cultural emotions and perceptions towards Chinese culture after watching Li Ziqi's videos. Thus, it offers insights for mainstream Chinese media on effectively utilizing overseas social media platforms and short videos to tell compelling Chinese rural stories and promote rural Chinese culture to the world.

Keywords

Li Ziqi, Comment Text, Text Mining, Sentiment Analysis, Convolutional Neural Network (CNN)

Copyright © 2024 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

文化是国家的根本和命脉，一个国家、一个民族的强大往往以文化的繁荣为基础。如何将推动中华优秀文化“双创”与加强国际传播能力建设有机融合，是当代中国面临的重大课题。

当前，充分利用国际化传播平台，开展中华优秀文化的国际传播，是全面提升国际传播效果、更加积极主动地讲好中国故事、传播好中国声音的必由之路。近年来，随着新兴技术的不断涌现，国际化传播平台正在经历深刻变革。各国民众的信息获取习惯日益向新兴媒体、移动传播等方面转变。Twitter、YouTube、Facebook、TikTok 等社交媒体平台在全球范围内迅速崛起与蓬勃发展，短视频正以前所未有的影响力渗透进人们的生活。短视频在社交媒体平台中的比重也日益显著，在全球传播领域具有着越来越重要的地位和作用，为中国文化走向世界、讲好中国故事提供了新的国际化舞台。

因此，在当前的对外传播格局和“讲好中国故事”、“文化强国”等国家战略的背景下，越来越多的中国媒体积极进驻海外社交媒体平台，希望通过与海外受众直接对话，塑造客观真实、良好的国家形象。然而，目前我国在利用国际社交媒体传播中国文化方面仍处于起步阶段，经验不足，存在过度依赖主流媒体、宣传味道浓重、倾向于宏大叙事等问题，传播效果并不尽如人意。

相比之下，以李子柒为代表的中国田园生活类自媒体短视频、中国美食生活类短视频则异军突起，在国际社交媒体平台上受到海外网友的追捧和喜爱。自 2015 年开始，李子柒以家乡四川乡村为背景，创作了一系列展现中国文化的美食生活类短视频。虽然李子柒自 2021 年 7 月已经停止更新视频，但她在 YouTube 的粉丝量不降反增，过往的视频被持续播放，截至 2023 年 10 月，“李子柒 Liziqi”账号在 YouTube 平台上的订阅量达 17,800,000，仅排名前十的视频累计播放量就已超过 6.4 亿次。通过制作短视频，如为奶奶做千层底棉鞋、笔墨纸砚、活字印刷、竹编、蚕丝被等，李子柒娴熟地传播着中国乡土文化。

李子柒短视频的成功“出海”，为中国乡村形象的国际建构及中国乡土文化国际传播提供了参考路径和示范经验。

2. “李子柒”短视频的跨文化传播研究

李子柒的中国传统田园生活视频赢得全球粉丝的持续关注，成为具有巨大跨文化影响力的传播案例，引起了学界研究热潮。

李子柒的短视频在内容层面引发了学术界的深入研究。曾一果等从情感视角对李子柒所制作的系列短视频展开媒介文本解读，深入探究社会加速时代“李子柒现象”背后的情感肌理、文化机制和消费逻辑，发现李子柒短视频从视觉、心理和审美等不同层次同时“按摩”异域文化中焦虑的观众，唤醒了海外受众对逐渐消逝的乡村田园生活的、世界性的共通情感[1]。刘志颖从审美角度研究并提出李子柒短视

频通过内容和形式的双重特色完成了“民族”与“世界”之间的审美转换，充分体现了受众期待视野的公约性心理，从而得到不同文化背景受众的普遍欢迎[2]。田玉霞从李子柒的短视频总结出了当前中国文化海外传播的启示，如视频要扎根生活、加强创新转型、注重专业化生产等[3]。宋玉将其划归为美食类短视频进行个案研究，探索其传播传统文化的成功要素[4]。

部分研究聚焦李子柒短视频在社交媒体领域的重要性以及其国际传播策略的影响，张红芸以李子柒短视频社交自媒体角色和对外传播策略为研究对象，探究如何在中国文化国际传播的过程中“讲好中国故事”[5]。路鹃等从中华田园风视角进行研究，他们提出以诗意叙事，通过“诗意符号”搭建异质文化的沟通桥梁，以跨文化的“视觉共同体”充分展示人类命运共同体的精神内核，是民间话语软渗透的有效途径[6]。

关于海外受众对李子柒短视频的评价和接受度，辛静和叶倩倩采用词频计量和文本分析结合的传统文本方法分析海外受众对李子柒视频的评论，认为引发海外受众热议的认知符号是传统的中国文化符号，积极的情感表达占据绝对主导，提出了优化中国文化跨文化传播的路径[7]。

传统的评论文本分析方法受制于特征工程的复杂性和规模效应，难以满足大规模数据的处理要求。社交网络文本篇幅短，针对性强，往往包含更丰富的情感信息。近年来，随着深度学习技术的兴起，通过构建复杂的神经网络结构，深度学习能够自动地从原始文本数据中提取高层次的语义特征，从而实现高效准确的情感分类[8]。因此，本研究采用深度学习技术对李子柒短视频下的评论文本进行情感分类和文本挖掘，旨在快速识别文本中的情感倾向，以探究海外受众在观看李子柒的视频后对中国文化产生了怎样的文化认知和文化情感。

3. 文本挖掘和情感分析在评论文本研究中的应用

文本挖掘作为一种高效的信息检索和挖掘技术，又称为文本数据挖掘或文本知识发现，是指在大规模的文本中发现隐含的、以前未知的、潜在有用的模式的过程[9]。传统研究往往针对单个文本进行结构，而文本挖掘可以针对海量的文本进行整体趋势挖掘。初级文本挖掘主要是指传统的词频统计、共现分析、聚类分析、社会网络分析等,高级文本挖掘包括篇章分析、情感分析、本体构建、人物关系、可视化网络等[10]。

情感分析，作为文本挖掘的重要分支，是指从文字中提取出人们的情感倾向或特征，目前情感分析的主要方法有三种。一是基于情感词典的方法。通过建立情感词典，将文本中出现的情绪词汇与情感词典进行情感词语的匹配，并根据匹配特征给与该词语相应的情感极性分数，从而计算其情感倾向。二是基于统计的机器学习方法。该方法通过标注的语料进行统计学习，把分类模型的选择和分类特征提取作为关注重点。常见方法有支持向量机(SVM)、朴素贝叶斯和逻辑回归等。三是基于深度学习的方法。该方法利用神经网络模型自我学习处理情感分析任务，能够进一步理解和分析文本，并根据语境和上下文的联系来提取特征，因此判断文本倾向的准确率更高。常用的神经网络模型包括卷积神经网络(CNN)、循环神经网络(RNN)、长短期记忆模型(LSTM)、门控循环单元(GRU)、注意力机智(Attention)等[11]。相关研究被应用在多重领域，如舆情监督、产品调研、文学作品海外接受度等。

目前，基于深度学习的情感分析方法被证明效果最佳，广泛应用于在线评论文本的特征提取和情感分类。例如，李杰等[12]基于卷积神经网络(CNN)有效实现了短文本评论产品特征提取及高准确率情感分类，证明了卷积神经网络(CNN)是在线评论分析的有效模型。

文本主题挖掘通常用到主题概率模型，目前常见的主题概率模型是潜在狄利克雷分布(LDA)主题模型。LDA 模型是 Blei 等于 2003 年提出的一种主题模型，作为一种无监督的机器学习方法，可计算给定的每一篇文档主题的概率分布，可以此为依据进行文本分类和主题聚类等任务[13]。目前，LDA 在文本挖掘和情感分析等领域得到了广泛应用。例如，张敏[14]以某电商平台上的购买用户评论数据为基础，构建了 LDA 主题模型，实现了对文本主题的提取和隐藏信息的挖掘。

在社交媒体短视频评论文本挖掘领域，网络评论文本数量庞大，存在句式口语化严重、网络流行词和特殊符号泛滥等特点，单纯采用文本挖掘、情感分析或者文本主题挖掘的任意一种，并无法彻底地挖掘出蕴含在文本内部的深层语义信息，因此，本研究在文本挖掘的基础上，选取卷积神经网络(CNN)进行情感分析，依据情感分类结果分别对正向评论和负向评论进行 LDA 主题模型的构建以获取特征主题词及对应的概率分布，最后系统化地梳理和概括主题聚类的结果，并基于研究发现，为我国主流媒体如何利用好海外社交媒体平台和短视频，向世界讲好中国故事，特别是中国乡村故事，传播中国乡土文化，建构可信、可爱、可敬的中国乡村形象，提供具有可行性的建议。

4. “李子柒”短视频评论文本挖掘

(一) 研究语料

本研究爬取了“李子柒 Liziqi”账号在 YouTube 平台播放量排名前十的视频的评论数据(见表 1，统计时间为 2023 年 10 月 8 日)。为了研究的可行性和获得更具代表性评论样本的需要，本研究选取全部热门评论，最后共获取 40,000 条评论。

Table 1. The top ten videos of Li Ziqi by popularity on YouTube
表 1. 李子柒在 YouTube 平台热度排行前十的视频

	视频标题和上传时间	观看量	评论量	点赞量
1	《花生瓜子糖葫芦，肉干果脯 雪花酥——年货小零食》 Peanut and melon seeds, dried meat, dried fruit, snowflake cake—snacks for Spring Festival (2019/1/31)	126,763,880	53,578	1,500,000
2	《柳州螺蛳粉》 Liuzhou “Luosifen”: Slurpy, Spicy, and Absolutely Satisfying (2019/8/11)	84,077,105	35,597	817,000
3	《瓜间一壶酒，西瓜和葡萄的一生》 A pot of wine among watermelon fields. So…The life of watermelons and grapes? (2020/9/14)	81,965,904	44,493	1,000,000
4	《柿子》 It’s a red mountain, and in the fall, it’s natural to make some sweet persimmons (2018/10/9)	72,218,228	19,301	549,000
5	《竹沙发》 Using bamboo to make some sophisticated old furniture (2018/9/25)	63,930,775	57,782	1,100,000
6	《紫米南瓜的一生……，还有花生》 The life of purple rice, pumpkins, and…peanuts (2020/10/11)	59,066,963	36,524	771,000
7	《棉花的一生》 The Life of Cotton, I promise next time I will come up with a new name of series (2020/1/20)	55,390,687	37,156	836,000
8	《番茄的一生》 Before the tomatoes fall, pick them home and get something delicious (2019/7/30)	52,079,338	21,854	557,000
9	《黄瓜的一生》 The life of cucumbers. It tastes like Li Ziqi’s summer. How about yours? (2020/7/20)	51,888,878	32,344	669,000
10	《菜籽油的一生》 The life of rapeseed oil (2020/6/19)	48,698,198	29,310	616,000

(二) 研究方法

1) 数据预处理

本研究对原始数据的预处理主要包括三个步骤。首先，筛选掉除英语之外语言的评论以及纯表情评论，得到纯英文评论文本 23,990 条。接着，对英文评论文本进行字母大小写与单词时态转换，并替换同义词；最后，过滤停用词，过滤掉符号、介词、冠词等无意义词汇，实现文本去噪。

2) 高频词分析

评论中的高频词代表着用户的关注和讨论重点，以此可以推断出海外受众在观看李子柒短视频时的主要关注点。本研究首先对 YouTube 上李子柒短视频英文评论的词频进行统计分析，得到前 100 的高频词表(见表 2)。

Table 2. The top 100 most frequent words list
表 2. 前 100 高频词表

排名/R.	词频/word	频率/freq	排名/R.	词频/word	频率/freq
1	video	3421	26	grandma	595
2	like	2534	27	girl	571
3	Li Ziqi	2259	28	channel	568
4	life	2002	29	world	539
5	love	1904	30	watch	537
6	people	1412	31	need	521
7	make	1326	32	feel	515
8	beautiful	1191	33	always	459
9	China	1021	34	well	458
10	know	964	35	living	456
11	Chinese	937	36	hope	454
12	work	902	37	could	451
13	amazing	825	38	never	450
14	good	804	39	new	446
15	live	779	40	every	422
16	want	767	41	wish	419
17	time	764	42	please	405
18	thing	758	43	right	404
19	think	754	44	nice	394
20	miss	691	45	best	386
21	everything	684	46	Youtube	376
22	food	666	47	real	373
23	woman	660	48	eat	368
24	hard	659	49	place	366
25	thank	602	50	nature	366

续表

排名/R.	词频/word	频率/freq	排名/R.	词频/word	频率/freq
51	garden	363	76	money	262
52	great	361	77	peaceful	257
53	god	356	78	long	256
54	cooking	356	79	skill	254
55	bamboo	344	80	show	253
56	say	342	81	talented	250
57	comment	340	82	little	249
58	ziqi	325	83	agree	249
59	true	325	84	keep	247
60	care	323	85	sure	246
61	happy	322	86	village	240
62	kind	320	87	lady	237
63	dog	318	88	everyone	236
64	country	316	89	give	227
65	actually	306	90	different	226
66	family	303	91	maybe	224
67	person	303	92	cook	221
68	grandmother	292	93	wonderful	219
69	working	285	94	cute	214
70	better	284	95	home	213
71	city	276	96	ever	213
72	man	269	97	learn	213
73	wow	267	98	anything	212
74	thanks	265	99	dislike	212
75	help	263	100	going	208

从高频词表中可以清晰地看出，排名前 100 的高频词，以 Li Ziqi (李子柒)、video (视频)、life (生活)、love (喜爱)、like (喜欢)、beautiful (漂亮的)为显著高频词，出现频次达到 1000 次以上。这些高频词反映了海外受众在观看李子柒短视频时的主要关注点，同时也揭示了他们对中国文化的初步感知。大部分评论侧重于李子柒的短视频本身，涉及到诸如 Li Ziqi (李子柒)、video (视频)、watch (观看)、channel (频道)等词汇。此外，评论中表达了积极的情感态度，如“love”(喜爱)、“like”(喜欢)、“thank”(谢谢)、“wow”(哇)、“good”(好的)、“amazing”(惊人的)、“talented”(有才能的)等词汇。还有关注中国社会生活文化的词汇，如“life”(生活)、“grandma”(奶奶)、“food”(食物)、“cook”(烧饭)、“bamboo”(竹子)、“music”(音乐)等。此外，评论还涉及到中国地域和地理文化，如“city”(城市)、“garden”(庭院)、“live”(生活)等词汇。综合来看，海外受众的关注点涵盖了中国传统美食、风土、生活方式、音乐等多个方面。这表明他们对李子柒个人、中国美食文化以及田园生活有着浓厚的兴趣。这样的关注点也为深入了解海

外受众对中国文化的认知和喜好提供了重要线索。

3) 共现词词对分析

通过筛选具有代表性的高频词相关的共现词词对，可进一步呈现出海外受众的热点关注内容(见表 3)：

Table 3. High-frequency co-occurring word pairs
表 3. 高频共现词词对

1	“video”, “amazing”	14	“ancient”, “China”
2	“video”, “liziqi”	15	“Chinese”, “people”
3	“like”, “liziqi”	16	“Chinese”, “countryside”
4	“love”, “liziqi”	17	“hard”, “work
5	“liziqi”, “life”	18	“traditional”, “food”
6	“good”, “life”	19	“ideal”, “woman”
7	“healthy”, “lifestyle”	20	“wonder”, “woman”
8	“village”, “life”	21	“virtuous”, “woman”
9	“watching”, “liziqi”	22	“feel”, “peaceful”
10	“amazing”, “food”	23	“dream”, “place”
11	“amazing”, “liziqi”	24	“nice”, “girl”
12	“beautiful”, “world”	25	“talented”, “girl”
13	“beautiful”, “country”		

通过对高频共现词和词对进行比较分析，可以发现评论主要集中在对李子柒个人和视频的正面评价上。例如，表达了对李子柒的喜爱(love)和对视频的惊叹(amazing)，以及对其美丽(beautiful)、有才华(talented)、努力(hardworking)等方面的肯定。同时，评论还表达了对中国文化的积极情感，展现了对中国田园生活的向往和对中国人的认知。例如，使用了与生活相关的词汇，如 life(生活)、lifestyle(生活方式)，以及描述中国人的特质，如 simple (简单)、healthy (健康)、happy (快乐)、beautiful (美丽)、hardworking (努力)、creative (有创造力)等。

此外，评论还关注了视频中的地点和地方文化空间。例如，提到了庭院(garden)、乡村(village)、梦想之地(dream place)等词汇，表达了对视频中环境和地域文化的关注和向往。

综上所述，海外受众的评论主要围绕着对李子柒个人、视频内容以及中国文化的积极评价和认知展开，体现了他们对中国田园生活和乡土文化的浓厚兴趣和向往。

5. “李子柒”短视频评论情感分析

基于高频词和共现词次对分析海外受众对李子柒的短视频产生的中国文化认知仍然相对表面，为深入了解他们对李子柒的短视频的情感或态度倾向，本研究利用 Kaggle 公开数据集 Twitter 和 Reddit 情感分析数据集作为训练数据[15]，并构建了构建卷积神经网络 CNN 模型。该模型的设计包括一个嵌入层、三个卷积层、三个最大池化层以及一个全局最大池化层，并最终连接了一个具有 softmax 激活函数的全连接层。输出层的神经元数量设置为 3，分别对应着积极、中立和消极三种情感类别。在模型训练之后，整体准确率(accuracy)达到了 0.91(见表 4、图 1、图 2)。

Table 4. Model training results
表 4. 模型训练结果

类别	精确率(P)	召回率(R)	F1 值	真实样本数量
类别 0 (中立情感)	0.91	0.91	0.91	2602
类别 1 (积极情感)	0.91	0.88	0.90	3137
类别 2 (消极情感)	0.91	0.94	0.93	3347
准确率			0.91	9086

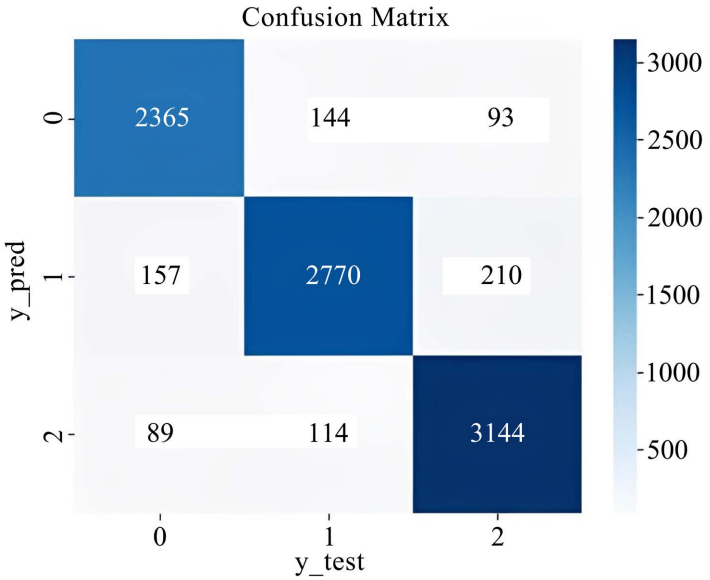

Figure 1. Confusion matrix
图 1. 混淆矩阵

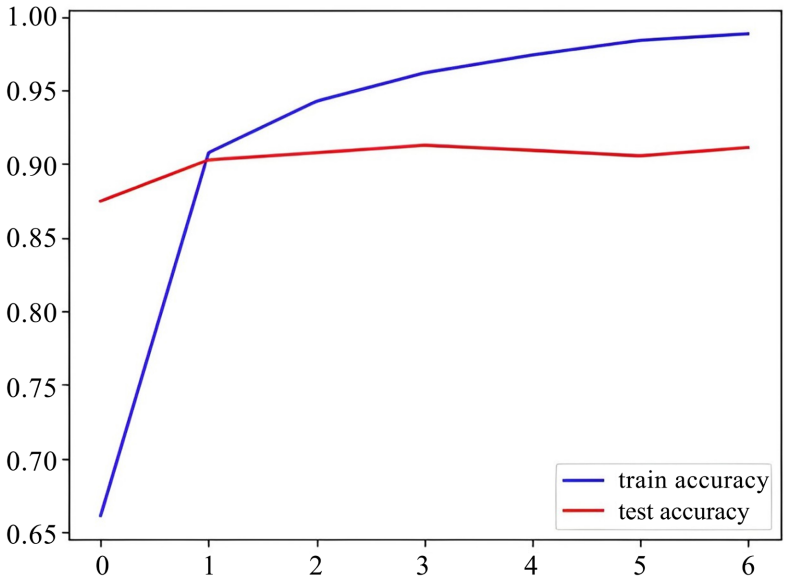

Figure 2. Training and testing accuracy curves
图 2. 训练和测试准确率曲线图

综上所述，模型在实验中取得了较好的性能表现，各类别的精确度、召回率和 F1 值也都表现出良好的性能，通过运用经过训练的模型对本研究所选取的李子柒视频下的 23,990 条英文评论文本进行情感分类预测，得到了相对可靠的结果。最终情感分类结果为：积极评论的数量 11,633 条，占比 48.5%；中立评论的数量 9126 条，占比 38%；消极评论的数量 3231 条，占比 13.5% (详见图 3)。

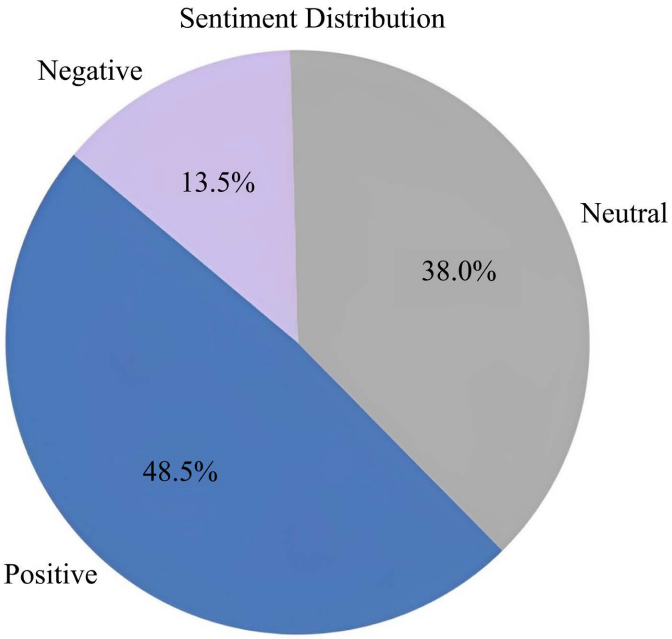

Figure 3. Sentiment classification results
图 3. 情感分类结果

6. “李子柒”短视频评论 LDA 主题建模

通过对情感分类结果为正向和负向的评论文本进行 LDA 文本聚类，LDA 主题分析关键词及其权重分布情况如下(表 5、表 6)：

Table 5. Keywords and weights from LDA topic analysis for positive comments
表 5. 正向评论 LDA 主题分析关键词及其权重

主题一		主题二		主题三		主题四		主题五	
特征词	权重	特征词	权重	特征词	权重	特征词	权重	特征词	权重
video	0.0771	love	0.0645	make	0.0328	China	0.0255	life	0.0389
love	0.0377	beautiful	0.0445	food	0.0175	people	0.0251	like	0.0309
watch	0.0312	amazing	0.0312	like	0.0168	Chinese	0.0175	music	0.0286
thank	0.0287	good	0.0186	eat	0.0155	live	0.0169	just	0.0180
new	0.0187	woman	0.0169	use	0.0121	life	0.0119	video	0.0165
good	0.0186	video	0.0167	just	0.0118	like	0.0101	make	0.015
Liziqi	0.0175	girl	0.0167	love	0.0113	good	0.0098	Beautiful	0.0145
Culture	0.0163	work	0.0163	grow	0.0107	just	0.0097	Work	0.0138

Table 6. Keywords and weights from LDA Topic analysis for negative comments
表 6. 负向评论 LDA 主题分析关键词及其权重

主题一		主题二		主题三		主题四		主题五	
特征词	权重	特征词	权重	特征词	权重	特征词	权重	特征词	权重
video	0.0259	eat	0.032	like	0.0448	video	0.0721	make	0.0234
think	0.0247	dog	0.0276	life	0.0379	miss	0.0584	ill	0.0213
Chinese	0.0235	chicken	0.0245	hard	0.0303	watch	0.0285	hate	0.017
fake	0.0167	food	0.0241	work	0.0299	liziqi	0.0263	country	0.017
people	0.0141	China	0.0185	make	0.0259	channel	0.0225	covid	0.016
China	0.0109	little	0.0152	video	0.0230	come	0.0213	watch	0.014
just	0.0107	Chinese	0.0152	feel	0.0223	new	0.0198	lazy	0.0138
propaganda	0.0097	think	0.0136	live	0.0213	really	0.0165	sorry	0.0127

主题挖掘结果显示，在文化情感层面，大多数海外受众对李子柒短视频所展现的中国田园生活场景持有积极的态度，尤其是对视频中呈现的中国传统非遗文化、田园乡土文化、饮食文化、以及中国风的视听元素以及李子柒个人形象表现出极大的兴趣和喜爱。负面评论主要集中在对视频真实性的质疑，这些消极评论中词频较高的词汇包括 fake (虚假)、propaganda (宣传)等，这部分观众认为视频内容过于精心制作，以至于显得不真实。

在文化认知层面，海外受众对于李子柒短视频所呈现的中国文化有着相对明确的认知。他们特别关注中国社会生活文化、饮食文化，以及中国的乡土文化，表现出对中国文化的明确且积极的认同感。这进一步反映了李子柒短视频在国际传播中取得的文化认知成果，为增进国际社会对中国文化的理解提供了有益的信息。

7. “李子柒”短视频评成功“出海”的启示

本研究基于李子柒的短视频成功“出海”的案例，对新时代背景下，我国主流媒体如何利用好短视频这一新形式，更好地向世界讲好中国乡村故事、传播好中国乡土文化提出以下建议：

一是用“土味”乡村唤醒不同文化背景受众对中国乡土文化的情感共鸣。李子柒的“一生”系列视频结合二十四节气，对周期较长的农耕过程采取季节跨度性拍摄，自然万物埋种、破土、开花的生长过程，在镜头下得以完美呈现，体现了人与自然的和谐共生，彰显了中华传统思想的道法自然。李子柒一日三餐及水果蔬菜等都取之于大自然，与奶奶过着朴素质朴的生活，视频中“一箪食一瓢饮”的舒缓乡村日常，成功勾勒了中国乡村生活图景，让身处现代都市喧嚣的受众对中国田园乡村生活无限向往，产生了对中国乡土文化的情感共鸣。

二是进行平民化叙事，用乡村小人物故事展现大时代精神。李子柒的个人形象得到了海外观众的积极评价，譬如，李子柒在爆米花即将出炉的那一刻，帮奶奶捂住耳朵，这一细节温暖感动了无数观众，她展现的善良、孝顺、勤劳、自强的品质正是当代中国人的集体精神的体现，也是对中华优秀传统文化的弘扬这种品质能够激发海外观众对中华优秀传统文化的认同和兴趣。

三是巧妙运用中国元素表达人类共同审美情趣。海外观众对李子柒视频的古风意境赞叹不已，李子柒身着中国传统丝绸服饰，配以箫笛、琵琶等中国传统乐器所奏曲目作为背景音乐，极尽展现了中国风的唯美，满足了全球观众的审美期待，成功实现了跨文化传播。

总体而言，这项研究提供了一种有效的文本挖掘方法来探索海外受众对李子柒短视频的文化情感和文化认知。相较于基于语料库的词频计量、词语搭配等传统的文本分析方法以及基于情感词典的情感分析方法，通过运用卷积神经网络(CNN)和潜在狄利克雷分布(LDA)主题模型[1]对李子柒短视频下的评论文本进行文本挖掘和情感分类，情感分类整体准确率达到了 91%，能够更精准地理解观众对视频内容的情感倾向，为进一步推动中国文化的国际传播提供了有益的参考。此外，通过构建 CNN 模型，本研究证明了 CNN 模型在处理大规模评论文本情感分类任务中的有效性和可靠性，为类似情感分析任务提供了可复用的技术框架。本研究选取的评论文本为“李子柒 Liziqi”账号在 YouTube 平台播放量排名前十的视频的评论数据，囊括了李子柒短视频中的美食、非遗文化等不同系列，未来可就李子柒短视频的不同系列的评论文本展开针对性的分析和探究，从而为主流媒体如何利用好短视频这一新形式，更好地向世界讲好中国乡村故事、传播好中国乡土文化和非遗文化，提供更为具象的经验借鉴。

基金项目

本文系浙大城市学院妇联女性专项资助项目(项目编号：FZ-202201)的阶段性成果之一。

参考文献

[1] 曾一果, 时静. 从 “情感按摩”到“情感结构”: 现代性焦虑下的田园想象——以“李子柒短视频”为例[J]. 福建师范大学学报(哲学社会科学版), 2020(2): 122-130, 170-171.

[2] 刘志颖. 跨文化传播视域中的李子柒短视频审美解读[J]. 当代电视, 2020(11): 84-87.

[3] 田玉霞. 中华文化走出去的路径选择——以“李子柒现象”为例[J]. 传媒, 2020(10): 81-83.

[4] 宋玉. “李子柒”短视频的传统文化传播[J]. 青年记者, 2020(17): 10-11.

[5] 张红芸. 中国文化对外传播的实践经验和可行路径——以 YouTube 李子柒短视频为例[J]. 出版广角, 2020(12): 77-79.

[6] 路鹃, 付砾乐, 张君昌. 中华田园风短视频跨文化传播价值分析[J]. 青年记者, 2021(14): 63-65.

[7] 辛静, 叶倩倩. 国际社交媒体平台中国文化跨文化传播的分析与反思——以 YouTube 李子柒的视频评论为例[J]. 新闻与写作, 2020(3): 17-23.

[8] 邓钰. 面向短文本的情感分析关键技术研究[D]: [博士学位论文]. 成都: 电子科技大学, 2022.

[9] 谌志群, 张国煊. 文本挖掘与中文文本挖掘模型研究[J]. 情报科学, 2007, 25(7): 1046-1051.

[10] 陆宇杰, 许鑫, 郭金龙. 文本挖掘在人文社会科学研究中的典型应用述评[J]. 图书情报工作, 2012, 56(8): 18-25.

[11] 王婷, 杨文忠. 文本情感分析方法研究综述[J]. 计算机工程与应用, 2021, 57(12): 11-24.

[12] 李杰, 李欢. 基于深度学习的短文本评论产品特征提取及情感分类研究[J]. 情报理论与实践, 2018, 41(2): 143-148.

[13] Blei, D.M., Ng, A.Y. and Jordan, M.L. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*, **3**, 993-1022.

[14] 张敏. 基于文本挖掘的电商评论情感分析[J]. 产业与科技论坛, 2020, 19(2): 63-64.

[15] Gowda, C., Anirudh, Pai, A. and Chaithanya Kumar, A. (2019) Twitter and Reddit Sentimental Analysis Dataset, Kaggle. https://doi.org/10.34740/KAGGLE/DS/429085