

# 后疫情时代背景下网络用语情感倾向分析 ——基于微博语料库的研究

胡逸清 姜一卉 王雪陈 张云飞 张敬宇  
北京科技大学 100083

**【摘要】**网络“情感词典”是分析网络用语使用特点及情感极向的常规途径之一。然而，目前对网络“情感词典”的研究缺乏基于特定语境、特定人群的分析。本文基于后疫情时代跨文化交流这一特定语境，结合机器深度学习，以微博、推特为语料库，爬取主题词与青年相关的语料。在提升情感识别精确度的基础上，分析跨文化语境下特定词语在后疫情时代的情感极性，与疫情前、疫情时代这两个时代的情感极性进行对比。以“口罩”为例，该词在疫情前期情感权值差异较大，而在疫情期间因政策要求与社会共识而上升。对应的英文词汇“mask”则在疫情期间进一步下降，反应西方社会个人主义思想。通过不同词汇的情感倾向差异进一步反映了深层次跨文化社会文明差异。

**【关键词】**后疫情时代；情感词典；机器学习；网络用语

Analysis of the emotional tendency of network language in the post-epidemic era —— Research based on microblog corpus

A make: Hu Yiqing Two for: Yi-hui jiang corresponding author: Wang Xuechen, Zhang Yunfei, Zhang Jingyu  
University Of Science and Technology Beijing 100083

**【Abstract】**The network "emotional dictionary" is one of the conventional ways to analyze the characteristics of network language use and emotional extreme. However, current research on online "emotion dictionary" lacks context-specific, population-specific analysis. Based on the specific context of cross-cultural communication in the post-epidemic era, combined with machine deep learning, this paper takes microblog and Twitter as the corpus to climb the corpus of subject words and words related to youth. On the basis of improving the accuracy of emotion recognition, the emotional polarity of specific words in the post-epidemic era in the cross-cultural context is analyzed, and compared with the emotional polarity of the two eras before the epidemic and the epidemic era. Take "mask" as an example, the emotional weight varied greatly in the early stage of the epidemic, but increased during the epidemic period due to policy requirements and social consensus. The corresponding English word "mask" declined further during the epidemic, reflecting western social individualism. The differences in emotional tendencies of different words further reflect the deep cross-cultural social civilization differences.

**【Key words】**post-epidemic era emotional dictionary, machine learning, network language

## 1 引言

2019年12月，新冠疫情突然爆发，给社会和生活方式带来极大影响。在这一事件下，“身体流动”被限制，互联网平台成为疫情期间人们宣泄情绪、寻求慰藉的主要窗口。情感分析和社交平台的兴起之间存在密切的关系。社交平台提供了海量的用户内容和交流数据，这些数据为情感分析提供了更多的数据来源和应用场景，如市场营销、娱乐休闲及政治政务等。目前对网络话语分析的研究中，针对公共语言分析较多，而针对某一特定语境、特定人群的个性化语言分析较少。本文以青少年群体为研究对象，基于后疫情时代这一特定语境，以微博、推特为研究语料库，并通过细分情感权值的方法完善 Hownet 知网情感词典，提升情感识别精确度，探讨中美青少年对疫情相关情感变化差异的深层原因。

## 2 中文情感词典

中文情感词典主要基于知网 Hownet 中文情感词典。经整合，词典中共包括 4566 个正向词汇与 4370 个负向词汇。

①褒贬情感词典：包含形容词、动词、名词等，往往带有情感极性，但难以区分其情感程度。因此，将正、负向词汇的情感权值赋为+1 与-1。

②程度副词词典：程度副词与动词和形容词搭配，在知网 Hownet 基础上本文将程度副词按强烈程度细化为 10 类，采用人工标注法。情感修饰作用越弱权值越接近 0，反之则越强。

③否定词词典：否定副词改变动词和形容词的情感倾向。本文在知网 Hownet 的否定词情感权值划分的基础上，将否定词权值细化为 7 类，范围 (-1,1)

## 3 语料库及研究过程

### 3.1 数据爬取

本文采用 scrapy 爬虫框架对微博中的情感语料进行爬

取,使用 Python 3.7.0 编译器版本号,PyCharm 2020.1.5 集成开发环境。

根据疫情特殊时期,本文设置数据爬取三个时间段:疫情前(2018.12.01--2019.11.30)、疫情期间(2019.12.01--2020.03.31)、后疫情时代(2020.04.01--2022.1.07)。爬取的数据包括博文内容与发布时间。中文方面选取五个在疫情期间较具代表性的词语作为爬取对象,分别是【居家、快递小哥、戴口罩、网课、返乡】。英文方面寻找对应的关键词【stay at home, online class/course, mask, homecoming, deliver man】。在数据爬取后,中文语料通过叠加“青年”“大学生”等关键词限制语料范围并进行第一次过滤;英文语料由于词汇的时态、语态变化,通过“词根”确定关键词。在语料主题方面,一个中文词语可能对应多个英文表达,因此要收集以“青年”为主题的语料,与直接关键词对应的表达之外,还需加上“学校”、“课程”等不同表达形式以确保数据量。

### 3.2 数据清洗

#### 1. 导入数据和初步文本处理

#### 2. 建立数据集和计算 TF-IDF 值

将分词后的文本数据转换为数据集的有效格式 TF-IDF (词频-逆文档频率)特征矩阵,并计算每个词语在文档中的权重,反映其在文本中的重要性。

#### 3. 构建聚类模型和确定聚类 K 值

本文中 K-means 聚类采用 Python 的 Scikit-Learn 库实现,并用肘部图法确定聚类的 K 值。通过计算不同 K 均值下的误差平方和,绘制肘部图,选取图中的“肘部”K 值,表示增加更多簇不再显著降低误差,为本实验中的最佳 K 值。

#### 4. 聚类和人工判断

用 K-means 算法进行聚类,把文本数据分为 K 个簇。对每个簇进行人工分析和判断,剔除可能的无效文本或异常簇,实现数据清洗。

### 3.3 中文情感权值计算

对爬取语料进行清洗后的数据为有效语料,进行情感权值计算。本文情感权值计算基于已有的知网 HowNet 情感词典,并对其中的副词程度词典及否定词词典的情感权值进一步细分。此处以“今天,我非常非常不开心”为例,通过分词得到[“今天”,“非常”,“非常”,“不”,“开心”],与情感词典依次比对,直到找到第一个情感词“开心”,其情感分值为 1.5。随后查找开心前的程度副词,发现“不”为第一个程度词,分值为-1.0,再发现两个程度词“非常”,分值为 2.0。继续向前寻找,直到遇到非程度词“今天”则结束此语句的情感值判断。因此,“开心”的情感分值为  $1.5 * (-1.0) * 2 * 2 = -6.0$ 。接下来向后寻找下一个情感词,得到的所有分值即为语句的情感分值。

## 4 讨论

在散点图中,平均值表示整体趋势,对异常值敏感,可能偏离实际分布。中值对异常值不敏感,只受数据集大小和

排序影响。散点分正值表示积极情感,负值反之,接近零表示中性情感。词语的情感变化趋势反映人们对问题的态度。英文情感词的权值由于通过 NLTK 库直接进算,而 NLTK 库预先将情感权值的取值范围设定在  $(-1,1)$  之间,所以不存在极值的影响。因此,通过平均值的变化反应即可反应英文关键词的情感倾向在各个阶段的变化。

### 4.1 “居家”

疫情之前,“居家”情感中值相对较高(4.00),可能表示青年对家庭的温馨和舒适的感受,以及对回家的渴望。疫情期间,“居家”的情感中值进一步下降(3.40),可能反映了青年群体被迫待在家中,面对疫情带来的压力和焦虑。然而,居家政策也带来了更多与家人相处的机会,并有利于防疫,因此整体情感倾向仍保持在正值水平。后疫情时代,“居家”的情感中值下降至较低水平(2.60),可能表明青少年对居家生活的热情减退,开始渴望恢复正常的社交和外出活动。“stay at home”的平均值大体可以反应美国社会对该词语的情感倾向变化。该词语的整体取值较小,可以反映出人们对这个词语持较为消极的态度。在疫情前期,美国社会对居家的态度本身较为消极。是因为美国社会文化强调自由与独立,人们可能更倾向于追求社交互动、外出活动和经济发展,因而居家在疫情前的分值较低。这与中国文化中重视亲情、家庭关系的传统文化有很大差别,因此两国民众在疫情前对这一词语的情感态度也有很大差别。而在疫情时期,数值进一步下降到 0.037,是由于美国在疫情期间部分地区实行了居家隔离政策,这一政策被许多美国人视为侵犯自身基本的权利,因而受到了强烈反抗。

### 4.2 戴口罩:

“戴口罩”一词在疫情前、疫情时期和后疫情时代的情感中值均为 0.00。这可能是因为口罩在疫情前只在特定职业或特殊场合中使用,被视为常见且普遍接受的行为。而在疫情时期和后疫情时代,大多数人普遍认同戴口罩作为一种防疫措施,人们普遍理解戴口罩对社会公共卫生的重要性,因此对这个词语的情感倾向相对中立<sup>[1]</sup>。图中平均值在疫情前是-0.65,在疫情时期是 1.37,在后疫情时代为 1.31。一些人对青少年戴口罩持消极情绪可能是因为担心其身体发育和社交发展。然而,随着疫情的发展,公众对戴口罩的认知发生了变化,情感倾向转向积极。

反观西方对青年对戴口罩的情感倾向不仅保持在较低水平,而且持续下降。这与中美两国文化差异有着较大的联系。疫情前,美国社会将口罩视为是医生和患者之间隔离和交流障碍的象征。<sup>[2]</sup>在疫情时期,大部分西方人仍然坚持认为戴口罩没有用,并以一种自以为是的态度认为自己是上帝的选民,相信自己会平安无事。这与美国文化深受《圣经》思想的影响有很大联系。<sup>[3]</sup>而美国对个人主义近乎狂热的追求,使美国民众对任何限制其自身自由的行为极度排斥,因此口罩成为了“约束”的代名词。而中国社会奉行“集体主义”,主张“依顺”与“服从”<sup>[4]</sup>因此中国青年更倾向于接受自觉佩戴口罩的政策要求,以维护社会和谐。

### 4.3 快递小哥:

“快递小哥”情感变化在疫情前 0.80,疫情时期 0.00,

后疫情时代 1.00。疫情前，人们对“快递小哥”持积极态度，可能是对其工作和服务的认可。随着疫情爆发，人们对外部人员接触持谨慎态度，情感倾向降至中性。后疫情时代，“快递小哥”情感倾向升至积极水平，可能是对快递员在经济复苏和生活便利方面重要性的认知。相比中值，平均值较高，分别为疫情前 4.11，疫情时期 3.73，后疫情时代 2.78。然而，随着居家隔离政策取消，人们恢复出行，购物和出行需求增加，对快递小哥的重要性相对减弱，持强烈积极情感的人数减少。

这一词语在国外的的情感倾向变化较小。与中国外卖平台高度普及以及价格相对便宜不同，美国外卖平台的价格相对高昂，根据《华盛顿邮报》调查，由于外卖平台的收入信息并非一直透明，加上西方传统支付小费的习惯，因此顾客的满意度并不高。这也是为何疫情前对外卖小哥的情感倾向相对较低<sup>[5]</sup>。而据美联社报道，在疫情期间美国的外卖行业得到了较快发展，人们对外卖小哥这一职业的情感倾向也有所增长。中西方这种差异不仅反映了中美两国在外卖平台发展方面的不同特点。

#### 4.4 网课：

“网课”一词的情感中值在疫情前是 2.00，在疫情时期是 1.00，在后疫情时代为 1.76。

疫情前，人们对网课持中立或稍积极的态度。网课被认为是方便灵活的学习选择，能够提供更多学习机会和资源。然而，疫情期间，学校关闭迫使学生转为线上学习。此外，技术设备等问题也可能影响了人们对网课的感受，因此人们对网课的态度变得更加负面。后疫情时期，尽管学校逐渐恢复线下教育，但仍保留线上教育。从平均值的下降趋势可以反映出，对网课持消极情感的人数仍在增加，可能是因为学生对线下教育的期待，希望能够回归传统面对面教学的互动和体验。

英文 online course 一词的情感变化趋势大致与中文相似，这侧面表现出在网络技术方面中美青少年的观念差异较小。尽管美国作为科技大国，在疫情期间，相当多的学校都

采用了 Zoom 平台进行授课，可以实现中文网课平台的基本授课功能，还能完成收交作业，虚拟现实场景体验等更加完备的技能。但据美国国家高级信息与数字技术研究中心调查，对网课的不满程度，从疫情前的 9% 上升到 27%<sup>[6]</sup>，主要是由于美国网课系统也无法完全排除技术困扰。

#### 4.5 返乡：

在疫情前，返乡的情感中值为 2.40，因为返乡通常代表着回到家乡、与家人团聚的意义，可能引起人们对温馨、亲情和归属感的积极情感。疫情期间，返乡的情感中值上升至 3.80，可能是因为疫情期间，很多人被困在外地，渴望重返家乡与家人团聚，返乡愿望更加强烈。后疫情时代情感中值回落至 3.00，仍然较为积极。这反映出随着疫情逐渐受控，限制逐渐放松，人们返乡更加容易，因此返乡意愿可能没有疫情时期那么强烈，但仍较为积极。这一情感中值变化的规律提醒我们家庭和家乡在人们的情感中具有重要的地位。

西方 homecoming 一词的情感倾向则一直处于下降状态，这与美国文化高度重视自由主义和个人主义有关。<sup>[7]</sup> 欧美国家的家庭伦理重视个人主义与个人利益，因此在自身面临疫情风险时，家庭观念则进一步削弱。因此对返乡的意愿也进一步下降。而在中国文化中，家庭尊重和孝顺是核心价值观。而在后疫情时代，一些由于求学、工作常年漂泊在外的青年人，在疫情期间逐渐意识到家庭亲情的重要性，更愿意返乡，把更多的时间用于与家人相处。

#### 5 结论

本文通过对比研究中外社交媒体上针对疫情相关词语的情感趋势，分析后疫情背景下中美青少年群体的情感表达和心理状态。本文总结了中西方情感分析的相关技术和研究进展，阐述了本文细化情感词典的构建和情感语料的爬取、过滤，情感权值的计算以及实验验证。实验结果表明，中美青少年在疫情不同时期对待相同事物的情感倾向既有差异也有相似点。这不仅反映了青少年群体共有的心理、生理特征，也反映了中美两国社会背景差异导致的文化差异。

#### 参考文献

- [1]刘玲玲:《疫情之下,要不要戴口罩?看看东西方纷争背后的故事》, [https://view.inews.qq.com/k/20200327AOSJRS00?web\\_channel=wap&openApp=false](https://view.inews.qq.com/k/20200327AOSJRS00?web_channel=wap&openApp=false)
- [2]Gorovitz, Samuel, Alasdair MacIntyre, "Toward a Theory of Medical Fallibility", *Hastings Center Report*, Vol.5, No.6 (1975), pp.13-23.
- [3][英] 安东尼·吉登斯:《社会学》, 赵旭东、齐心等译, 北京: 北京大学出版社, 2007年, 第505页。
- [4][韩] 韩炳哲:《我们不能给病毒找到理由》,《世界报》2020年3月23日。
- [5]何贝贝,安峥. 在美国点外卖:令人头疼的共赢难题[N]. 解放日报, 2023-06-11(006). DOI:10.28410/n.cnki.njfb.2023.003178.
- [6]梁霞.中文网课教学带来的挑战与应对策略及教学设计实例——以美国高等院校中文网课教学为例[J].汉语国际教育学报,2022(01):39-56.
- [7]何玲,汪凤慧,牛莉莉等.西方国家的家庭观研究[J].中华家教,2023(01):21-29.

第一作者:胡逸清;第二作者:姜一卉;通讯作者:王雪陈 张云飞 张敬宇。