

# 从Facebook“孔子学院”关键词看中国文化的海外传播

- 基于python 的数据采集与情感分析

刘金芳\* · 杨蓉蓉\*\*

## 目 录

1. 引言
  - 1) 研究背景与研究问题
  - 2) 研究综述
2. 研究方法与设计
  - 1) 数据来源
  - 2) 数据处理
3. 研究结果
  - 1) 话题量变化明显, 2019年激增
  - 2) 整体情感态度趋势呈正面、积极的态度
  - 3) 词频和词云分析
4. 结语与反思

## 1. 引言

### 1) 研究背景与研究问题

随着中国国际影响力的提高和话语权的增强, 世界各地人民对于汉语学习的需求也在急速增长, 汉语文化的传播越发显示出重要性。近年来, 关于中国文

---

\* 复旦大学 国际文化交流学院 汉语国际教育在读硕士

\*\* 复旦大学 国际文化交流学院 副教授

化的海外传播、“文化输出”等话题也一直是国内汉语国际教育领域讨论的一大热点。在中国文化的海外传播中，孔子学院是一个非常重要的载体，孔子学院在国外汉语教学和文化传播中发挥了很大的作用。

截至目前，全球已有162国家（地区）设立了541所孔子学院和1170个孔子课堂<sup>1)</sup>。但是近年来也有越来越多反对孔子学院的声音，孔子学院也面临很多质疑和争议。关于孔子学院的文化传播，要有效开展跨文化交流，就需要更好了解中华文化在目的国的接受程度和影响程度。本文以孔子学院在海外传播现状研究为出发点，依托Facebook上2007年至2019年间共计5022条帖子，以python语言的情感分析为工具，对Facebook上“孔子学院”关键词展开分析，力图展示孔子学院在海外的传播现状。有助于我们通过“孔子学院”这一关键词了解更多中国文化传播现状，为进一步的开展国际间的交流，发展友好对外关系，帮助世界人民了解中国和中国文化打下基础。

## 2) 研究综述

笔者主要对“孔子学院”相关话题的研究现状进行介绍，关于孔子学院话题，主要有三个大方向的研究：孔子学院文化传播；孔院汉语教学；孔子学院发展和自身建设。

笔者搜录了孔子学院相关的文化传播研究，可以细分为两个大的方向，一是孔子学院在文化传播中的作用和存在问题研究：瓮俊燕[1]以“5w”模式为视角，讨论孔子学院在中华文化国际传播的过程中的成效与不足；毛向樱[2]（2019）重点分析了孔子学院在文化传播过程中遇到的困境，主要有“西方国家文化威胁论、孔子学院在全球范围的布局存在盲点、师资力量缺乏、跨文化传播形式单一等”。马梦真[3]（2018）主要从传播内容和传播方式总结了孔子学院在文化传播方面存在的问题：内容趋于表面化，方式趋于程式化。并且提出借鉴塞万

1) 数据参考孔子学院官网：[http://www.hanban.org/confuciusinstitutes/node\\_10961.htm](http://www.hanban.org/confuciusinstitutes/node_10961.htm)

提斯和歌德学院的经验。张周洲, 陈越[4] (2019) 分析了中华文化国际传播存在的短板: 海外传播理念、海外传播策略、文化产品供给质量, 传播主体等; 传播渠道仍然是政府为主, 民间和个人力量没有得到充分发挥, 应该坚持“以人为本”的理念。

关于孔子学院文化传播方面的研究, 第二个主要方向是针对数据分析外媒对中国形象和孔子学院的情感态度: 焦俊峰, 刘美兰[5] (2019) 对《卫报》和《纽约时报》相关报道展开分析, 研究这些报道关于孔子学院、中国文化(刻板)形象的构建及其意识形态产生的根源。葛宇飞, 徐欣[6]以英美主流媒体对“孔子学院”的报道为例, 探讨西方主流媒体对“孔子学院”接受程度和情感态度。房乐宪[7] (2019) 分析了三大群体: 西方某些政客、研究机构、新闻媒体对“中国渗透论”的渲染。对中国影响力的提高产生的焦虑和不安, 中国面临的外部舆论环境更加复杂。

从以上研究可以看出, 针对“孔子学院”文化传播的研究还有很大的研究空间, 主要集中在孔子学院文化传播内部分析。针对外部传播效果及情感分析较少, 数据来源也大多以国外主流媒体和官方聚到为主, 而民间对孔子学院的接受程度和情感态度研究很少, 因而笔者从广大网民使用最广泛的脸书出发, 分析“孔子学院”在外国网民中的接受程度及他们对孔子学院的情感态度。

## 2. 研究方法与设计

### 1) 数据来源

Facebook是美国的社交网络服务网站, 是在中国以外的其他地方使用最广泛的社交网络工具之一, 在全球拥有数量广大的用户。Facebook也成为国外人们获取信息的重要来源之一, 该平台思想与观点表达较为自由丰富, 数据开放程

度也相对较高。因此本研究利用Facebook平台，使用python数据抓取工具，对“Confucius institute”这一关键词的Facebook帖子进行抓取并研究。本文主要抓取了2007年至2019年间，不同发帖人讨论、发布的有关于“孔子学院”话题的帖子，选取了帖子的发布时间、发布地区、发布内容、评论数量、点赞数量这五个维度。帖子数量共计5022条。

## 2) 数据处理

首先，笔者从抓取的5022条帖子中，随机抽取50条，基于帖子的文本内容，人工判断该帖子持有的情感态度“积极态度”、“消极态度”、“中立”倾向[9]。为了提高判断的准确性，笔者会根据文本语境、说话语气等综合判断。例如“China's Plan A : Gain access to the wealth of oceania countries, by: -- Confucius Institute - influencing the current and next generations through re-education of the communist way in our universities...”（中国的A计划：通过孔子学院获得大洋洲国家的财富——在我们的大学里通过共产主义方式的再教育影响后世后代……）全帖没有一个消极情感的词汇，但是通过文本的讽刺语气，可以判断并不是真的讲述中国的成功，实则是“共产主义”“间谍组织”“中国威胁论”言论下对孔子学院教育的抵制和讽刺。

第二步：对数据进行人工处理，对抓取的数据进行清理调整，去掉与话题无关的帖子、删减重复内容、去掉一些无法翻译出的小语种帖子，剩余有效数据4733条。

第三步：1、使用python工具，对数据进行预处理，使用python Jieba 分词[10]将帖子文本切分为词语列表，进行关键词提取并过滤停用词，删掉帖子文本包含的网址等无用信息。2、使用R语言sentimentr，计算出每条帖子的情感分析数值，根据时间序列，将文本的情感态度均值变化进行可视化处理，更清晰的反映广大网民对孔子学院的态度变化趋势。3、根据切分处理的词语，利用python工具计算出不同情感态度下出现的词频比例，并作出词云图。对高频出

现的关键词进行排名，更直观有效的反映出人们最关心的是“孔子学院”哪些相关议题。

### 3. 研究结果

#### 1) 话题量变化明显，2019年激增

如图1表1所示，2019年该话题出现次数激增，2019年以前话题次数分布均匀且都很少不超过100次。思考原因可能是由于时间过早的帖子已经被删除或者爬虫技术无法获取，因而话题量较少。本次数据相关内容无法探究出话题次数变化的背后原因。

年份	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	总计
次数	1	18	14	17	20	20	33	70	52	56	51	58	4612	5022

表1 - 话题次数统计表

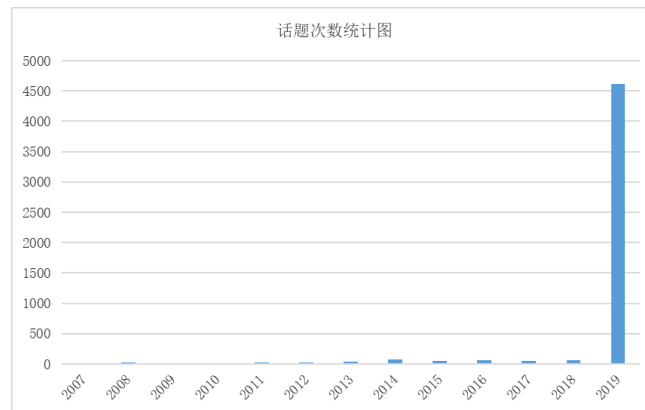


图1 - 话题次数统计图

## 2) 整体情感态度趋势呈正面 积极的态度

R语言的情感分析结果取值分为三大类型：大于0，等于0，小于0。大于0则表示情感态度是积极的可能性；当数值等于0时，情感偏向中立；当数值小于0时，情感态度趋向消极负面态度。

笔者先对所统计出来的情感分析结果数值进行分类统计，饼图是三种情感类型所占比例。如图2显示，整体情感态度呈正面、积极的态度，积极占比55.1%，消极占比26.9%，中立占比17.9%。由此可见，虽然社交媒体上对孔子学院的态度包含很多敌意和歪曲言论，但是民间态度更多的还是积极支持的态度。

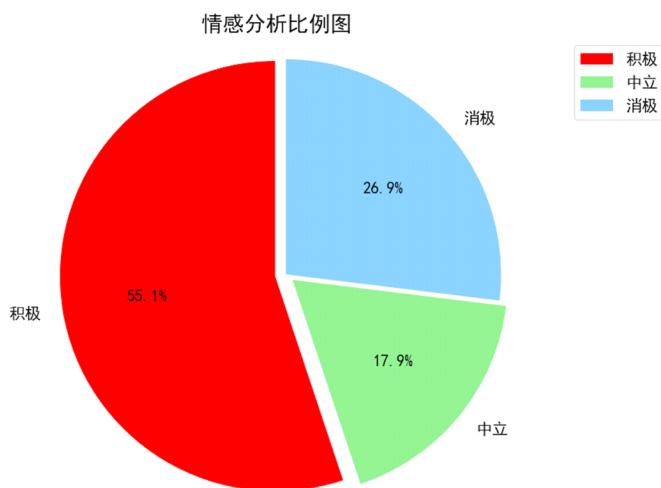


图2 - 情感态度比例图

笔者再对所统计出来的情感分析结果数值进行平均值处理。作出了可视化的均值折线图[11]。根据下图3数据显示：2007至2019年，情感均值都是大于0，平均值呈积极态度；但是从时间序列来看，情感值变化呈波浪式的变化趋势。因而网民对“孔子学院”相关话题的言论情感态度是整体上呈现波浪式趋于正面

的态度。

但是在2012和2019年这两年情感数值较低，查阅历史事件资料，2012年中国十八大完成换届，并提出新的发展目标，同时美国的军事战略重心转向亚太地区。由于中国局势和世界格局力量的变化，中美竞争加剧，因而外媒显示出对中国的言论和形象态度呈现更多负面态度。

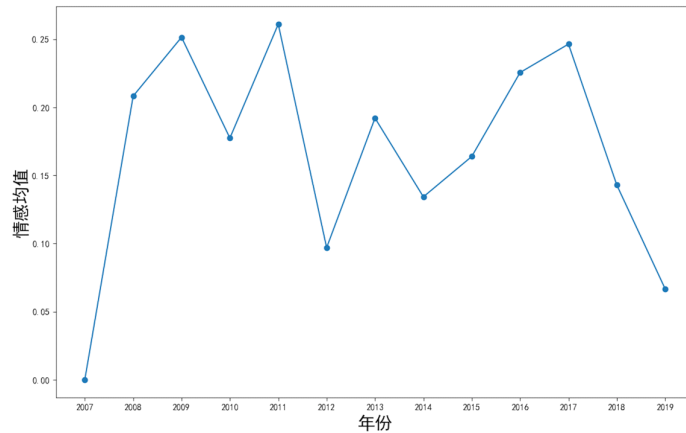


图3 - 情感均值折线图

### 3) 词频和词云分析

为了进一步了解网络上关于“孔子学院”话题所讨论的主要议题，笔者在对帖子文本进行分词处理之后，作了关键词词频统计[12]。保留了排名前30的高频关键词。如下表2所示：

序号	词语	频次 (次)	百分比
1	confucius	4164	10.55%
2	institute	2326	5.89%
3	chinese	683	1.73%
4	university	676	1.71%

5	china	550	1.39%
6	quote	256	0.65%
7	close	239	0.61%
8	life	225	0.57%
9	learn	197	0.50%
10	day	172	0.44%
11	school	160	0.41%
12	Confucius institute	147	0.37%
13	time	147	0.37%
14	student	143	0.36%
15	teach	136	0.34%
16	propaganda	135	0.34%
17	love	133	0.34%
18	matter	123	0.31%
19	fund	122	0.31%
20	fall	119	0.30%
21	beijing	115	0.29%
22	language	113	0.29%
23	superior	113	0.29%
24	campus	110	0.28%
25	spy	108	0.27%
26	action	107	0.27%
27	slowly	97	0.25%
28	people	96	0.24%
29	beauty	95	0.24%
30	tie	92	0.23%

表2 - 高频关键词表

出现频率最高的是“Confucius”（孔子）一词，共出现4164次，此外相关的还有“institute”（机构）2326次、“Chinese”683次，“university”676次，“school”160次。这些日常词语都是和“孔子学院”话题紧密度很高的常用词。

在前30词频中，中性态度的词占比最高，从人工分析角度看，中性词有17个，此外消极情感态度的词有6个，分别是close、propaganda、fall、spy、slowly、tie；而积极情感态度的词是7个，分别为quote、life、learn、teach、

love、superior、beauty。总的结果和研究预想一致，消极和积极呈现均衡态势。

笔者再针对不同情感类别下（呈积极情感态度、消极情感态度、中性态度）的词频作了词云统计图[13]，词云是通过字体大小展示的，在词云中字体越大表示频率越高，字体越小频率越低。由图4-积极词云图也可以看出，积极态度下，出现频率较高的词有：quote、love、learn、life、superior、job、beauty等。可见外国网民对孔子学院的教育和文化传播工作较为认可，很多人也是出于热爱和工作等各种动机，积极学习中国文化。



图4 - 积极词云图

在积极词频中，笔者发现quote一词可以深究，quote一词出现的频率很高，总共出现256次，占总词频的第六位。quote的中文含义是指引用、语录等，笔者追溯到quote相关的句子，发现这一词大多和孔子的经典语录相关，帖子通常直接引用孔子《论语》中的某一言论表达自己的观点。

孔子是中华文化最具代表的象征性人物，孔子在国人中的地位毋庸置疑，梁漱溟先生(1999:150)曾指出：“孔子以前的中国文化，差不多都收在孔子手里；孔子以后的中国文化，又差不多都从孔子那里出来。”

孔子的经典语录《论语》不仅仅在中国影响深远，《论语》中的经典句子对外国学习者也是影响颇深。很多网友经常引用《论语》中关于人生哲理和处世哲学相关的句子，一部分是中文原文，一部分是英文大意。用来描述自己心情或者启迪思维。笔者阅读帖子文本，帖子中出现次数最多的句子分别是：

"It does not matter how slowly you go as long as you do not stop."-Confucius

"Wherever you go, go with all your heart!" -Confucius

"Everything has beauty, but not everyone sees it." - Confucius

可见孔子语录在国外也具有很高的文化价值和教育价值。孔子语录的广受欢迎可以作为孔子学院文化传播的视角之一。

从消极情感态度下的词云分布图，如下图5，可以看出，消极词语主要有如propaganda（政治宣传）、close、fall、espionage（间谍活动）、spy（间谍）、mistake等，但是从文本分析中可以看出，对孔子学院消极言论也几乎只有这一种声音，没有其他关于孔子学院运营、教学、活动、教师形象等相关的消极言论。





## 4. 结语与反思

本文运用Python工具对Facebook中的孔子学院话题进行文本挖掘，并利用R语言sentimentr情感分析，结果初步揭示了外国网民对孔子学院的情感态度和变化趋势，总结了不同情感态度下关注的话题，有助于我们更好的了解中国文化的海外传播现状及存在的国际舆论环境，对于推动中华文化海外传播，增强中华民族文化自信有一定的实践意义。

研究结论主要有以下几个方面：首先，外国网民们对“孔子学院”的态度整体上是波浪式趋于积极的态度，说明孔子学院组织在海外能够受到认可和支持，虽然“中国威胁论”“间谍机构”等网络言论看似汹涌，但实际上外国网民还是非常认可孔子学院的教学和文化传播作用，也非常喜爱《论语》等中华文化经典及其语录。这对于孔子学院的文化传播方向、定位及后续发展都有积极导向作用。值得注意的是，国外对中国影响力的提高充满了焦虑和恐慌，采用歪曲孔子学院性质的方法散播不实言论，企图关闭孔子学院。我们也应该关注到这些不和谐声音，做好应对策略，促进孔子学院更好更快发展。

由于数据样本的局限性，本文在处理数据和分析数据背后原因方面还存在很多不足，这也为后续研究提供了更多方向。

### 參考文獻

- 瓮俊燕,「从“5w”模式探析中华文化国际传播-以孔子学院为例」,『传播力研究』.
- 毛向櫻,「孔子学院跨文化传播效果及提升路径:困境与策略」,『绥化学院学报』,2019.
- 马梦真,「孔子学院文化传播过程中的几个问题及解决方法」,『文学教育』,2018.张周洲·陈越,「中华文化国际传播多元化路径探析」,『人民论坛』,2019.
- 焦俊峰·刘美兰,「英美媒体对孔子学院文化形象构建研究-基于《卫报》与《纽约时报》相关报道的批评话语分析」,『天津外国语大学学报』,2019.
- 葛宇飞·徐欣,「基于语料库和批评话语分析的海外孔子文化传播研究-以英美主流媒体-对“孔子学院”的报道为例」,『海外英语』.
- 房乐宪,「谁在鼓吹“中国渗透论”」,『人民论坛』,2019.
- 马凌远,「孔子学院“走出去”的影响因素研究:理论框架与实证分析」,『云南师范大学学报』,2019.
- 李丹妮,梁嘉,「新浪微博中的“上海垃圾分类”议题文本挖掘研究-基于Python Snow NLP的舆情分析」,『东南传播』,2019.
- 张娜·柳运昌·王若男,「基于文本情感分析的社交媒体数据挖掘」,『河南城建学院学报』,2019.
- 高雅·苏艳·席方园,「基于Python的新浪微博用户数据采集与分析」,2019.
- 骆明·王含含·盛博·廖安灵,「基于文本挖掘技术的网络舆情特征研究」,2019.
- 洪居兴·王璐·王虹·祁俊·陈秀明·王世颖,「基于Python的情感数据分析」,『青海大学学报』,2019.

## Abstract

On the overseas spread of Chinese culture from the key words of "Confucius Institute" on Facebook  
-Data collection and emotion analysis based on Python

Liu Jinfang · Yang Rongrong

In the overseas communication of Chinese culture, Confucius Institute is a very important carrier. But in recent years, there are more and more voices against Confucius Institute, which also faces many questions and disputes. This study analyzes about 5022 keywords of "Confucius Institute" on Facebook with the emotion analysis of Python language, trying to show the current situation of Confucius Institute's overseas communication. In this paper, we first use Python web crawler technology to crawl the "Confucius Institute" keyword on Facebook, then preprocess all the numbers, filter out the useless information, use R language emotion analysis algorithm to calculate the score of each post, and draw a visual emotion score change chart according to the score. It is found that netizens have a positive attitude towards the discussion of "Confucius Institute", but there are still many negative comments and distortions of the image and nature of Confucius Institute.

**Key words :** Facebook, Confucius Institute, python, Data collection ,emotion analysis, overseas communication.

투 고 일 : 2020. 4. 10. / 심 사 일 : 2020. 4. 15. ~ 2020. 5. 15. / 게재확정일 : 2020. 5. 20.

