

基于Python爬虫关于独生子女对双方父母养老态度的分析

张丰钰, 张申奥, 张诗笛, 李松涛, 孔亮宸

大连外国语学院, 辽宁 大连

收稿日期: 2022年4月23日; 录用日期: 2022年5月24日; 发布日期: 2022年5月31日

摘要

“十四五”规划伊始,我国开启全面建成社会主义现代化国家新征程,随着人口出生率的不断下降,我国人口老龄化程度逐年加剧,老年人口数量逐年攀升,积极应对老龄化已经成为国家重大战略之一。养老逐渐成为每个独生子女家庭的问题,人口老龄化已成为现代社会的常态。在这一大背景下,中国家庭“老龄化”的现象也在不断加剧,主要表现为有老年人的家庭比重上升和家庭中老年人口比重增加。本文针对新时代新养老的热点问题,在我国人口老龄化加重的背景下进行创新实验,实验中首先使用Python网络爬虫技术,针对独生子女长大后面对父母养老问题的态度搜寻并进行数据爬取,之后根据文本情感进行分析并使用作图工具进行可视化展示,最终通过正负情感的对比分析得出结论。

关键词

爬虫, 数据分析, 独生子女, 养老

Python-Based Crawler Analysis of the Only Child's Attitude towards Both Parents' Old-Age Care

Fengyu Zhang, Shenao Zhang, Shidi Zhang, Songtao Li, Liangchen Kong

Dalian University of Foreign Languages, Dalian Liaoning

Received: Apr. 23rd, 2022; accepted: May 24th, 2022; published: May 31st, 2022

Abstract

At the beginning of the “14th Five-Year Plan”, China started a new journey of building a socialist modernized country in an all-round way. With the declining birth rate, the aging degree of China's

文章引用: 张丰钰, 张申奥, 张诗笛, 李松涛, 孔亮宸. 基于 Python 爬虫关于独生子女对双方父母养老态度的分析[J]. 数据挖掘, 2022, 12(3): 203-210. DOI: 10.12677/hjdm.2022.123021

population is increasing year by year, and the number of elderly people is increasing year by year. Actively coping with aging has become one of the major national strategies. Pension has gradually become a problem for every one-child family, and the aging of population has become the normal state of modern society. Under this background, the phenomenon of “aging” of Chinese families is also increasing, which is mainly manifested in the rising proportion of families with elderly people and the increasing proportion of elderly people in families. In this paper, aiming at the hot issue of the new age and the aging of the population in China, an innovative experiment was carried out. In the experiment, Python web crawler technology was first used to search for the attitude of the only child towards the parents’ old-age problem after he grew up, and data was crawled. Then, the text emotion was analyzed and visual display was carried out by using mapping tools. Finally, a conclusion was drawn through the comparative analysis of positive and negative emotions.

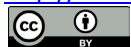
Keywords

Web Spyder, Data Analysis, Only Child, Providing for the Aged

Copyright © 2022 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

随着信息技术的快速发展，互联网对社会的各个领域都产生了巨大的影响，人们已经习惯使用互联网来了解国家发展、关注日常生活、表达利益诉求，这就对传统媒体的发展带来了极大的影响。在互联网时代下，网络舆情是一种“虚拟社会”的产物，网络信息内容具有多样性[1]。现有的网络舆情治理手段主要分为算法设计分析手段和治理对策分析手段，前者主要是基于大数据算法，针对宽泛的网络舆情分析手段进行设计和研究，后者主要基于分析模型进行分析和研究。本文主要针对国家重大养老问题进行探究，针对养老问题进行具体化、精细化的研究和分析，使用具体的主题研究代替宽泛的理论研究。

目前，微博、论坛等网络社交平台为实验数据的首选地，但是针对本文的研究主题，其上搜索到的数据与理想实验数据差距较大，并且一部分评论属于发泄式评论，没有论述的逻辑。因此，知乎平台作为目前许多人寻找问题答案的优先选择平台，其中的许多回答也与理想实验数据特征较好的符合，具有较高关注度和较多的用户评论数量。

本文主要以独生子女养老态度为主题，详细分析了独生子女目前养老的现状、存在问题以及内心的养老观念，最终提出疫情常态化下，针对养老问题独生子女应该有的新态度和新观念。

2. 实验数据

数据来源于知乎平台“独生子女以后怎么照顾好双方父母”话题。该话题在知乎平台中属于关注度较高的一个话题，一共有 1.8 万关注量，530 条回答以及 317 万的浏览量。按照该话题下方回答的评论数量高低，分别选取前 3 条评论数量高于 100 条的回答，对每一条回答中的用户评论进行数据爬取。其中发表以上 3 条回答的用户的用户名分别为“若非鱼”、“穆谢特”和“唐旭”。

3. 实验流程与方法

3.1. 实验流程

实验选择的实现方法主要为数据爬取，本次实验利用 Python 语言对数据进行爬取。其工作原理是：

向客户端发送 HTTP 请求, 将 URL 提交给服务器端, 服务器端根据 URL 的信息进行逻辑处理, 并将需要的数据返回给客户端[2]。首先确定数据来源的 URL 地址, 利用 Python 中 requests 库进入目标网页。对目标网页源代码进行观察和分析, 找到需要数据在源代码中的存在位置。确定了数据所在位置以及网页源代码结构后, 先将网页源代码全部爬取下来, 再对源代码进行信息提取。实验所选择的数据分析方法为产生图云、统计词频和进行评论情感分析。对用户评论进行分词、去停用词操作, 将得到的结果通过 Word Cloud 库生成词云图。同时对分词后的结果进行词频统计, 结果以图表形式展现, 最后将用户评论数据进行情感分析。

3.2. 具体实现

数据爬取过程中, 首先分析网页源代码。本次实验选择对“独生子女以后怎么照顾好双方父母”话题下的 3 条回答的用户评论进行数据爬取。将数据所在网址作为 URL 地址, 通过 request 函数获取到网页源代码。上述流程所得到的数据格式为 JSON 格式, 某些数据可能被转换为十六进制数, 所以必须对数据进行类型转换, 将数据中的十六进制数转换为字符串, 进而得到正确的用户评论数据。

得到用户评论数据后, 将数据集中保存在一个 txt 文件中, 以便后续操作。然后是分词和去停用词操作。在作生成词云技术的时候用到 jieba 和 wordcloud 这两个第三方库, 他们两个分别是用来进行分词操作, 以及生成词云的第三方库。

分词完成后, 在网上搜索到了对于中文来说比较常用的停用词, 之后用来分隔评论中的重要部分, 之后将分割出的关键词填写到 txt 文本中, 将遍历过程中得到的词语与停用词表中的词进行比较。若该词语存在于停用词表中, 则不选择此词语进行保存。将上述流程得到的结果保存到另一个 txt 文件中并将该文件作为参数, 并使用 wordcloud 库生成词云图, 文件中某一词语出现次数越多, 在词云图中该词语格式越大。其中我们规定生成的图片的长和宽分别为 1000 和 700, 背景颜色为白色, 这样更利于我们最终结果的展示。最后, 将生成的图片命名并进行存储。

情感分析过程实现方法为利用百度智能 API 接口对用户评论数据进行分析。在使用接口之前, 先在情感定向分析的创建模型一栏中, 通过传入正负项预料来自训练和生成模型, 再通过 API 调用, 使得得到的数据准确性高。将分析完成后的结果分为两部分进行保存, 分别为积极和消极。

词频统计过程所使用的是上述流程得到的消极情感数据。对该部分用户评论数据进行分词和产生词云图操作, 从词云图中选择若干个与实验主题具有联系的词语进行词频统计。将得到的词频通过 matplotlib 库进行表格形式展示。

4. 研究成果

4.1. 可视化词云

根据图 1 可以看出, “孩子”、“独生子女”、“计划生育”、“钱”、“生活”、“工作”等词语所占的比例比较大。其中, 孩子问题和养老问题是对立的根本, 由于父母的年代存在计划生育的政策使得他们只能生一到两个孩子。父母步入老年往往伴随着子女进入中年, 此时独生子女既要承担自己的生活压力又要承担父母的养老压力。独生子女结婚后, 在没有孩子的前提下, 独生子女双方最多需要赡养 4 位老人。独生子女生育后, 将在孩子的抚养上花费大量的时间与精力, 这也极大压缩了对父母的照顾。其次, 经济问题也是独生子女面临的问题之一, 新冠疫情肆虐下经济不景气, 独生子女面对着工作上与生活上双重的经济压力, 如住房压力、工作压力等, 没有足够的资金支撑父母的养老。工作繁忙也是独生子女面对的养老问题之一, 如今大部分人都过着严重加班的生活, 无法支配自己的时间, 每天都被工作占满, 没有时间照顾自己的父母, 这表明传统的养老方式无法匹配当前的社会现状。

通过文本分析,数据表明少部分积极态度的人群是对未来有着期望而不是抱怨生活的糟糕,但大部分积极态度的人群可能经济条件较为优越,自己本身生存压力较小,双方父母身体健康。因此这部分人群希望通过二胎和三期来让自己的生活更加充实,同时老人带孩子既解决养老问题也解决教育孩子的问题。同时,也有很多人认为在新的生产力条件下不能被旧观念束缚,养老院如果能让老人生活的更舒适、更开心、更长寿,子女实在没有办法照顾老人的情况下可以养老院。

反观消极情感,评论消极的原因主要是生存的压力,主要包括经济上的压力、精神上的压力和工作上的压力,这些压力无法让子女本身的生活得到保障,因此他们更没有多余的精力投入在照顾双方父母身上。父母和子女的关系是相对应而存在的,如果子女较多必然会增大早年间抚养孩子的成本和压力,但同时也会减轻养老上子女的压力。但是现实问题是一个独生子女承受了父母全部的爱,同时在结婚后也要承担赡养四个父母的重担。其中,经济压力是生存压力的主要表现,近些年,国家经济迅速发展,但居民平均收入并没有完全等价的提高,大部分人的收入仅仅能维持自己的生活,而无法留下更多的给父母。

综上所述,情感分析和数据解读表明大众在不同情景下和不同压力下所表现的想法是不同的。面对金钱的压力很多人都抱有焦虑的心理,而面对生活上的压力,大家更多的是出于积极态度,希望通过二胎三胎的增加改变养老问题。

4.3. 消极情感的原因及表现

由图3和图4可以看出消极情感的关键词主要是“住院”、“经济”、“孩子”、“身体”、“生育”、“生活”、“孤独”、“工作”。其中这8个关键词中出现频率从高到低的是“孩子”、“生活”、“生育”、“工作”、“孤独”、“经济”、“身体”、“住院”。

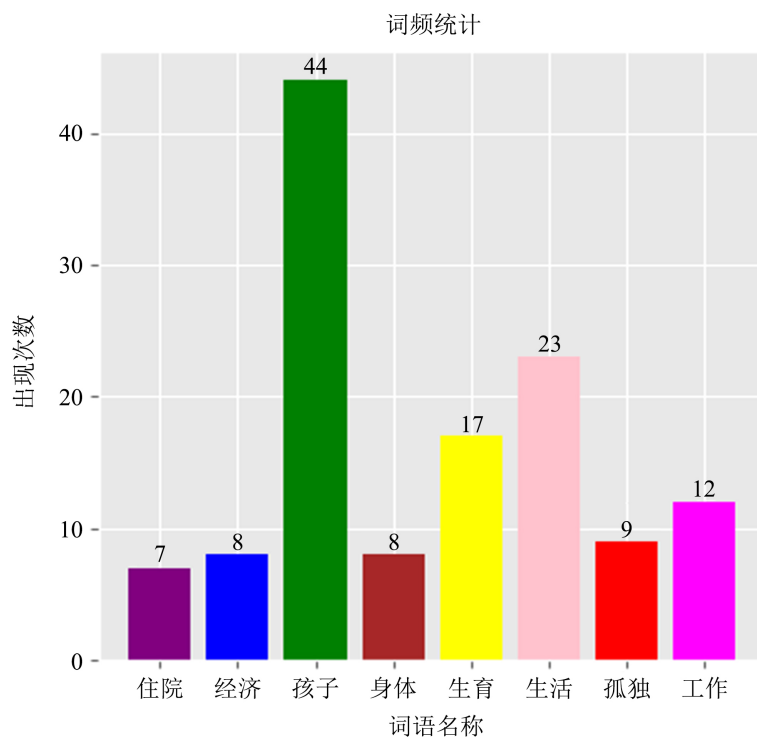


Figure 3. Word frequency statistical histogram

图3. 词频统计柱状图



Figure 4. Cloud image of negative comments
图 4. 消极评论词云图

由以上的关系可以很明显的看出孩子是独生子女赡养老人的最大的问题，由于年轻人存在孩子要抚养，他们可能没有太多的时间和精力来抚养他们的父母。大部分的独生子女步入中年之后都结婚生子，进入了“上有老、下有小”的生活阶段。许多年轻人选择延缓生育来保障自己的生活，近些年年轻人的生育率逐渐降低，疫情下 2020 年全国人口出生率首次跌下 1%，刚刚过去的 2021 年，我国人口出生率仅为 8.52‰，是中华人民共和国成立以来的最低出生率，中国人口基本上进入了零增长的阶段。与此相对的是中国的人口老龄化规模不断加剧，其表现为规模大、速度快、养老体系不完善、养老负担加大等等。据统计，1962~1976 年婴儿潮人口在未来 5~10 年进入老龄化。因此，预计 2033 年左右进入占比超过 20% 的超级老龄化社会，之后持续快速升至 2060 年的 35%。

在健康与生活方面，生活中面临的一些问题都极大的消耗了当代年轻人的精力和时间。生活与工作的压力必然会导致疾病的年轻化，例如高血压、冠心病、肥胖、糖尿病、恶性肿瘤等慢性疾病的发病率越来越年轻化。这导致现在的年轻人不得已将花费大量的金钱来照顾自己的身体，相反更没有时间、金钱和精力陪伴照顾父母。

在生育方面，有部分独生子女选择了生二胎、三胎，近几年国家大力建议开放二胎政策，让年轻人加大生育欲望，这也无疑导致了年轻人的抚养孩子的压力极度增加。众所周知，养育一个孩子的成本是极高的，这也间接的导致了年轻人对自己的父母关爱程度较低。对于现在的年轻人来说，他们在工作中遇到的压力是很大的，这使得他们的身体以及心理上都受到了严重的压力，现在年轻群体的生活节奏是很快的，他们大部分都会面临加班的问题，将所有的精力和时间都花费在工作上，这也间接的导致了现在年轻群体很少照顾他们的父母。对于经济这一关键条件，是年轻群体面临的十分重要的问题，由于在生活、住房、抚养孩子上的花销是极其严重的，所以，年轻人们又很少的一部分钱来帮助他们的父母。

综上所述，消极情感主要表现在孩子问题、工作生活问题、身体健康问题，国家如今更加注重养老保障体系的建设，未来尽力用政策带动养老而不是让每个独生子女家庭承担着巨额的养老负担与养老压力。人口既是经济增长的红利，也是社会共担的负债。养老保障体系尤其是养老金体系至关重要，微观

上影响每个人一生的财富分配，宏观上关系到经济可持续发展的动力、社会稳定的根基。

5. 结论

5.1. 疫情下人口老龄化带来的影响

根据以上的实验结果，不难看出人口老龄化已成为现代社会的常态，中国的人口老龄化程度在全世界处于中上水平。在人口老龄化背景下，养老产业的发展对缓解供需矛盾，保障人民生活，实现全面建成小康社会的历史目标战略意义重大[3]。在这一大背景下，中国家庭“老龄化”的现象也在不断加剧，主要表现为有老年人的家庭比重上升和家庭中老年人口比重增加。中国的人口老龄化重点体现在未富先老和养老负担加重两大特殊趋势。人口老龄化会对我国经济发展产生一定程度的不利影响，老年人的需求是普遍低于年轻人的，需求的降低会遏制消费，经济的增长率会逐步降低。老年人口过多导致劳动力短缺，进而导致我国劳动力成本不断增加。但随之而来的也是养老产业逐步成为“朝阳行业”，扩大老年人的消费市场，进而一定程度上拉动经济增长。

世界正处于新型冠状病毒肺炎疫情的大流行阶段，新冠病毒的肆虐对老年人冲击巨大，人口老龄化逐渐增加潜移默化地提升了防疫的难度，老年人相较于年轻人身体较差，基础疾病较多，并且有相当大部分的老年人没有接种新冠疫苗，不利于抵抗病毒的入侵，大部分时间居家生活，或者两点一线从家到超市。这种疫情下常态化的生活模式，很容易让老年人的精神世界产生孤独感和空虚感。落实到每个家庭，老年人的物质需求和精神需求得不到充分的满足，国家的养老体系、养老产业和养老能力无法适应如今的老齡化社会。

一旦父母居住地的城市爆发疫情，子女必然无法第一时间到家里陪伴父母。如果子女不能基于父母及时的照顾和关心，常态化疫情防控模式下很容易造成父母不明白防疫要求，如不会使用手机打开健康码。这导致老年人基本的生活质量都无法保障。此外，老年人的娱乐方式以线下聚集为主，如打牌、打麻将、广场舞等，疫情的反复必然导致线下聚集活动不定时的停摆。养老院也发生过因防疫不当造成全院传播的风险，因此传统的养老产业不能满足疫情下养老模式的需求。绝大多数的老年人不会使用电子产品进行接打电话、收发微信以外的任何操作，这大大增加了疫情下购买物资的难度，如果老年人的生活基本得不到保障，那么更大的养老消费市场也吸引不到消费。

5.2. 疫情下新型养老观念

“十四五”规划下，老龄化问题已经上升到国家战略问题的高度，但社会上对养老问题的重视程度明显不够。本文对实验结果的分析充分表明了老龄化社会下独生子女的压力在逐渐增加，正是因为社会上普遍的养老观念还没有跟上社会与时代的发展。常态化疫情防控的背景下，独生子女应该更加注重精神上的养老而并非物质上的给予，子女应该主动帮助父母熟悉手机在疫情下的用法，如下载注册防疫 APP、展示健康码、预约核酸检测等等。如今，父母需要的已经不是简单的金钱、陪伴，他们更需要的是疫情来临时子女对他们在配合防疫上的帮助，是疫情下能保障基本生活的物资，是自己的生命健康安全。此外，老人的娱乐方式可以逐步的由线下向线上过渡，养老产业可以开发老年人专用的线上养老平台供老年人交友娱乐，以满足老年人足不出户的进行娱乐，既减少了感染的风险，保证了健康又丰富了生活，降低了空虚感和孤独感。

基金项目

大连外国语大学 2022 年大学生创新创业训练计划，项目编号：202210172A195；“颐馨家园”——人口老龄化背景下新型养老平台研究。

参考文献

- [1] 张明杰, 朱焯行. 网络舆情监测系统的设计与实现[J]. 电脑编程技巧与维护, 2017(22): 18-19+33.
<https://doi.org/10.16184/j.cnki.comprg.2017.22.006>
- [2] 田雪丽, 郭志斌, 刘梦贤. 基于 Python 的网页数据爬取与可视化分析[J]. 电脑知识与技术, 2022, 18(6): 24-26.
<https://doi.org/10.14004/j.cnki.ckt.2022.0312>
- [3] 高凯, 胡秋明. 可持续养老保障运营模式研究[J/OL]. 西南金融: 1-11, 2022-05-16.