

The Micro-Blog Users' Social Attitude towards Leftover Men and Women Based on the Text Sentiment Analysis

Xue Zhang^{1,2}, Peng Wang^{1*}

¹School of Psychology, Shandong Normal University, Jinan Shandong

²Department of Psychology, Peking University, Beijing

Email: *pengsdnu@163.com

Received: Jun. 11th, 2015; accepted: Jun. 23rd, 2015; published: Jun. 30th, 2015

Copyright © 2015 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This study investigated the Internet users' social attitude towards leftover males and females. Related to the topic of leftover males and females, 1000 original micro-blog messages selected from Sina Weibo were used to identify the emotional tendencies of these messages. The results showed that the micro-blog users held negative attitude towards leftover males and females. In addition, the emotional tendencies of the 900 original micro-blog messages that were recently published from 60 Sina Weibo users who were marked by leftover males or females were analyzed as well. The results showed that the emotional tendencies of these messages were neutral. Besides, there were dramatically differences on the topics focused by leftover males and females. The research not only induces a rational social attitude towards leftover males and females, but also plays a guiding role for them in setting up the correct conception on choosing spouse.

Keywords

Leftover Men and Women, Attitude, Micro-Blog Messages, Chinese word segmentation, Analysis of Emotional Tendency

微博用户对剩男剩女的社会态度研究——基于微博文本情感倾向分析

张雪^{1,2}, 王鹏^{1*}

*通讯作者

¹山东师范大学心理学院, 山东 济南

²北京大学心理系, 北京

Email: pengsdnu@163.com

收稿日期: 2015年6月11日; 录用日期: 2015年6月23日; 发布日期: 2015年6月30日

摘要

本研究探讨了网络用户对剩男剩女的社会态度。采用微博文本情感倾向分析的方法对1000条新浪原创微博进行态度研究。采用微博文本情感倾向分析的方法对1000条新浪原创微博进行态度研究, 其结果显示微博用户对剩男剩女群体持负向态度。另对标签为剩男剩女的60名新浪微博用户近期发表的900条原创微博进行情感倾向性分析, 发现剩男剩女群体的微博呈中性态度, 且剩男和剩女的微博关注话题存在明显性别差异。本研究结果对于社会正确理性看待逐渐增多的剩男剩女具有一定的启示意义, 对剩男剩女在择偶问题上树立正确择偶观具有引导作用。

关键词

剩男剩女, 社会态度, 微博文本, 中文分词, 情感倾向分析

1. 引言

剩男剩女(Leftover men and women)指到达适婚年龄而处于非婚状态的青年, 作为伴随社会发展而出现的一个特殊群体, 其数量正以一定的比例逐年扩张。中国剩男剩女的增多, 将会导致社会成本的增加, 心理压力的累积, 甚至对社会的安定构成潜在的威胁[1]。剩男剩女作为一种社会现象引发了学者的探讨, 目前主要研究领域集中在剩男剩女产生的原因及其折射的社会问题, 有关社会对剩男剩女的态度研究较少。

态度(Attitude)包含认知、情感、行为三种成分, 态度具有内隐性, 需要借助中间变量通过外显的行为或语言进行测量。态度同时具有倾向性, 包括中性态度、积极的正向态度、消极的负向态度, 可反映态度持有者对人或事物的观点、看法及立场。由于态度对行为具有一定的预测作用, 通过探究微博用户对剩男剩女的社会态度, 在客观上可以反映广大微博用户在日常行为或言语中对剩男剩女是否存在偏见, 从而有助于针对社会大众对剩男剩女持有的态度及观点提出合理化建议。本文同时对剩男剩女这一群体的微博进行了态度研究, 以比较大众对该群体的态度与该群体本身的社会态度是否存在差异, 同时以性别为变量分别探讨了剩男关注的微博话题与剩女关注的微博话题是否存在性别差异。

2. 微博文本界定与研究方法

2.1. 微博及微博文本情感倾向分析

2014年1月16日, 中国互联网络信息中心(CNNIC)发布的《第33次中国互联网络发展状况统计报告》[2]表明: 截至2013年12月, 我国网民规模达6.18亿, 我国微博用户规模为2.81亿, 网民中微博使用率为45.5%, 即接近一半的网民在使用微博。由此可以看出微博已经成为数以亿计的中国网民发表、分享、获取信息的主要平台。2009年8月, 国内网站新浪推出“新浪微博”, 成为国内首家提供微博服务的门户网站。新浪微博作为全球使用量最大的两家微博客提供商(美国的Twitter和新浪微博)之一, 自推出以来, 受

到广大网友的热捧。新浪微博的主要功能有：发布功能、转发功能、关注功能、评论功能、搜索功能、私信功能，其主要特点有使用门槛低、随时随地、快速传播、实时搜索等。

微博作为人们抒发情感、发表评论的平台，含有海量的富含感情信息的文本。文本情感分析(Text sentiment analysis)就是对大规模文本进行态度分析而发展起来的方法。文本情感分析，在商业领域可以为企业进行市场分析、市场调查、顾客反馈提供更多有价值的信息，在管理领域能够帮助领导者更快地了解群众对各类政策措施的反馈意见，使得官方和民间智慧进行更加良性地互动[3]。所谓情感分析，就是确定说话人或作者对某个特定主题的态度。其中态度可以是他们的判断或评估，他们(演讲、写作时)的情绪状态，或者有意(向受众)传递的情感信息[4]。情感分析的重要问题就是对情感倾向性(Sentiment orientation)的判断，即判断作者的情感观点是积极正面的，还是消极负面的，它是表达自身观点、态度、立场等语言的一种量度，这类问题也被称为情感分类(Sentiment classification)或意见挖掘(Opinion mining)。情感倾向通常由两个标准来衡量：一个是指偏离的方向，即该情感词所要表达的意思是正面的还是负面的；另一个是指偏离的程度，即情感词在表达正面或负面状态下的强弱程度。

2.2. 目标微博的选取与数据收集

本文将进行两个研究，研究一为探讨微博用户对剩男剩女的社会态度，研究二为探讨剩男剩女本身的社会态度并对剩男和剩女关注微博话题进行性别差异研究。

研究一选取和剩男剩女相关话题的新浪微博 1000 条，实际操作时通过搜索新浪微博关键词的方式依次选取微博内容里含有“剩男剩女”或“大龄未婚”或“大龄单身”的原创微博，为保证微博内容的多样性，对某些出现频率较高的相似微博进行人工筛选过滤，最终得到 1000 条微博。

研究二选取标签为“剩男剩女”的新浪微博用户 60 名，其中标签为“剩男剩女”的女性微博用户 30 名，标签为“剩男剩女”的男性微博用户 30 名，实际操作时通过搜索新浪微博找人的方式选取标签为剩男剩女的微博用户 60 名，其中男女各占 30 名，且这些用户为认证用户，并排除非个人用户即机构用户，选取每名用户最近发表的 15 篇原创微博，最终得到 900 条微博。

2.3. 数据分析方法

首先，对获取的文本数据进行文本预处理。文本预处理技术是文本分析的一项基础工作，本研究对文本进行了中文分词、去除停用词两项相关的文本预处理工作。分词技术是语义理解的首要环节，是将语句中的词语正确切分开的一种技术。由于中文本身的复杂性及书写习惯，使得中文词语之间没有明显的区分标记，因此汉语分词技术便成为中文信息处理的关键[5]。本研究采用张华平开发研制的NLPIR2014分词系统[6]进行中文分词处理，该分词系统主要功能包括：中文分词、英文分词、词性标注、命名实体识别、新词识别、关键词提取、支持用户专业词典与微博分析。为了将经过文本预处理后的分词结果进行形象直观的呈现，本文采用了词云(Word cloud)技术对分词结果进行呈现。“词云”[7]又名文字云，是由词汇组成的类似云的彩色图形，“词云”就是对网络文本中出现频率较高的“关键词”予以视觉上的突出，形成“关键词云层”或“关键词渲染”，从而过滤掉大量的文本信息，使浏览者只要一眼扫过词云图片就可以领略文本的主旨。停用词也被称为功能词，是指与其他词相比没有实际含义的词，本研究中的停用词指在微博文本内容中出现频率极高或极低的介词、虚词、代词以及一些与情感无关的字符(如@、V、#、http://等)，通常文本的停用词处理中可采用基于词频的方法将其除去。

其次，借鉴情感词典对经过文本预处理后的分词结果进行情感倾向分析。情感词典是情感词的集合，微博文本的情感倾向大多通过情感词语来体现，因此情感词典的构建是情感倾向分析的基础。目前文本情感分析领域还没有一部完整且通用的情感词典，国内主要应用的资源有董振东开发的知网(HowNet)[8]，

Table 1. Word segmentation results
表 1. 分词结果

1000 条微博分词结果及权重	30 名标签“剩男剩女”女性用户 微博分词结果及权重	30 名标签“剩男剩女”男性用户 微博分词结果及权重	60 名标签“剩男剩女”用户微博 分词结果及权重
微博 13.75	非诚勿扰 14.12	安卓 1201.6	安卓 1201.66
剩男剩女 10.07	微博 10.48	微博 11.65	非诚勿扰 16.09
吐槽 9.72	失联 7.06	二手房 8.26	微博 14.00
爸妈 9.70	可以 6.27	可以 6.87	失联 8.87
离婚率 9.28	单身 6.20	快餐街 6.49	二手房 8.26
特么 8.70	男人 5.97	二手房出售 6.25	七夕 7.73
恐婚 8.53	女人 5.52	年终总结 5.82	可以 7.72
差评 8.32	杂粮米 5.28	工作 5.80	没有 7.64
大龄单身 8.08	盆友 5.27	没有 5.72	单身 6.61
海归 8.01	喜欢 5.17	时代天街 5.64	快餐街 6.49
悲催 7.91	需要 5.14	时间 5.56	二手房出售 6.25
大龄未婚 7.88	现在 5.11	牛逼 5.48	男人 6.24
苦逼 7.73	逆天 5.04	知道 5.29	还是 6.12
韩剧 7.73	飞机失联 4.99	执行力 5.28	时间 6.04
可以 7.52	坠楼 4.99	商住房 5.28	爱情 5.98
单身 7.51	爱情 4.92	一双 5.28	女人 5.92
相亲节目 7.39	朋友 4.79	喜欢 5.03	工作 5.90
适婚 7.26	时间 4.59	生活 4.92	喜欢 5.88
相亲 7.22	觉得 4.47	发现 4.92	年终总结 5.82
开始 7.22	起来 4.47	朋友 4.79	朋友 5.77
剩斗士 7.11	生活 4.42	公司 4.76	知道 5.67
一堆 7.06	开始 4.40	天街 4.70	生活 5.66
生活 7.04	微信 4.30	还是 4.70	时代天街 5.64
结婚 7.04	就是 4.19	爱情 4.67	起来 5.58
娱乐圈 6.78	上班 4.16	看到 4.65	开始 5.51
觉得 6.74	知道 4.05	分享 4.56	需要 5.51
社会 6.68	世界 4.00	单身 4.54	牛逼 5.48
职场 6.66	说话 3.99	开始 4.35	看到 5.36
妈蛋 6.65	心情 3.93	结婚 4.33	觉得 5.36
恋爱 6.60	看到 3.88	投资 4.17	商住房 5.28
爱情 6.55	人生 3.87	中国足球 4.17	执行力 5.28
神马 6.55	可能 3.75	微彩票 4.16	杂粮米 5.27
男女 6.50	离开 3.75	生命 4.16	盆友 5.27
工作 6.50	变成 3.75	价值 4.14	软饭 5.27
朋友 6.49	流量 3.73	改变 4.12	妹纸 5.27
知道 6.47	准备 3.72	手机 4.12	帅哥 5.27
成为 6.40	感情 3.69	个人 4.09	包邮 5.27
男票 6.40	游戏 3.69	男人 4.07	发现 5.20
盆友 6.39	北京 3.68	觉得 4.03	说话 5.06
暗恋 6.39	现实 3.64	人生 4.00	逆天 5.04
年龄 6.37	支持 3.59	市场 3.99	分享 4.99
孩子 6.37	恋爱 3.57	能力 3.97	坠楼 4.99
看到 6.35	过去 3.57	解决 3.90	结婚 4.95
屌丝 6.32	回家 3.50	相信 3.90	可能 4.92
介绍 6.32	回来 3.47	起来 3.89	手机 4.92
女人 6.31	奋斗 3.47	机会 3.88	公司 4.85
时间 6.12	卡片 3.47	客户 3.87	世界 4.85
需要 6.06	生日 3.43	希望 3.87	人生 4.80
还有 6.04	孩子 3.40	时候 3.86	离开 4.70
喜欢 6.02	还有 3.35	感觉 3.85	天街 4.70



Figure 4. Word cloud of 30 leftover men's micro-blog
图 4. 30 名“剩男”微博分词词云

根据对剩男和剩女的微博分词结果进行分析可得，男女关注的话题有较大的差异：女性关注最多的是“非诚勿扰”，男性关注最多的是“安卓”，即女性更关注相亲类节目，男性更关注电子产品。另外，女性关注较多的话题有“男人”、“女人”、“杂粮米”、“爱情”、“盆友”、“朋友”、“孩子”等，男性关注较多的话题有“二手房”、“年终总结”、“工作”、“执行力”、“公司”、“投资”、“中国足球”、“微彩票”、“市场”、“客户”等，即女性倾向于讨论与人际相关的话题，男性倾向于讨论与权力相关的话题。这一结论符合社会对男女所持有的性别角色期待，也从侧面反映了剩男剩女的社会适应性良好。

4. 讨论

4.1. 大众对剩男剩女存在偏见的心理机制探讨

根据研究一的结果分析，微博用户对剩男剩女的社会态度倾向明显偏于负面，而研究二结果显示剩男剩女自身的社会态度是中性的，该结果在一定程度上反映了当前社会对剩男剩女群体的一种偏见，较多的微博内容显示剩男剩女已成为社会问题，需要受到大家的关注，如：“这些年只涨不跌的八个数据：房价、物价、车牌价、剩男剩女数量、离婚率、污染指数、癌症发病率、对外移民率”，且较多微博认为剩男剩女产生的原因是由于男女择偶观的问题。这表明随着剩男剩女数量的增多，剩男剩女正逐渐成为一个社会热点话题，但是以微博用户为代表的网络大众更多的将该问题做个人归因，即认为是个人择偶观有偏差，较少地将该问题做社会归因，如性别比例失调等。这一现象符合社会心理学归因理论中的观察者效应(Observer effect)，即观察者对行为者的活动倾向于做内归因，认为该活动是由行动者稳定的内部特质(如个人特质)引起的而非情境因素导致的。将观察者效应引入剩男剩女这一话题，可以帮助网络大众了解自己对该群体进行讨论时在社会态度和归因上产生的偏差，从而促进社会公众更加理性客观地看待和对待这一群体，进而减少剩男剩女自身所感受到的社会偏见。

4.2. 剩男剩女关注话题

在研究二中可发现，剩男剩女除关注飞机失联、富士康坠楼等社会热点问题外，其讨论最多的话题是婚姻，相关词汇包括“非诚勿扰”、“七夕”、“单身”、“爱情”、“喜欢”、“结婚”。一方面，剩男剩女也渴望“爱情”和婚姻，希望脱离“单身”状态，和自己“喜欢”的人“结婚”；另一方面，

由于他们的生活中缺少两性情感的部分，因此有些剩男剩女将很多时间投入到独自活动中，比如宅在家里上网看新闻，看“韩剧”等。对研究中中和剩男剩女相关话题的1000条微博分词结果进行分类，将“剩男剩女”、“大龄单身”、“大龄未婚”、“剩斗士”归为标签类，将“娱乐圈”、“海归”归为领域类，将“相亲节目”、“相亲”、“适婚”、“结婚”、“恋爱”、“爱情”、“暗恋”、“介绍”归为相亲类，将“爸妈”、“离婚率”、“韩剧”、“单身”、“生活”、“社会”、“职场”、“男女”、“工作”、“朋友”、“男票”、“盆友”、“年龄”、“孩子”、“女人”、“时间”、“需要”归为生活类。领域类为微博用户讨论剩男剩女较多的领域，包括娱乐圈和海归，这两类领域的人士在中国人传统观念里是较为高端的明星人物，一般拥有好职业、高收入、高学历，但如今却有越来越多的剩男剩女集中在这两个领域，推断其产生原因是由于个人择偶观的偏差。相亲类词汇在分词结果中占较大比例，表明剩男剩女话题与相亲类话题的相关程度较高。这也从侧面表明，“适婚”的剩男剩女们内心也憧憬“爱情”，也渴望自己能够尽早地找到自己的人生伴侣，和有缘人“恋爱”和“结婚”，而有些男女则由于“暗恋”心仪的对象没有表白而错过了姻缘。剩男剩女身边的亲朋好友、同事同学们则往往热心给他们“介绍”对象，于是“相亲”活动成为剩男剩女脱单的重要途径，而具有综艺性和娱乐性的“相亲节目”也正符合了这个时期剩男剩女的心理需求。生活类词汇涵盖了剩男剩女的生活内容。其中，爸妈的权重最高，可推测剩男剩女感受到最大的压力来自于父母。这与中国儒家传统文化特性有密切关系，“不孝有三，无后为大”的思想观念深入人心，父母都希望自己的孩子都能够早日组建幸福的家庭，对于剩男剩女的父母而言，他们的心情则更为迫切，于是父母的“催婚”成了很多剩男剩女的紧箍咒。另外，“工作”和“朋友”相关词汇均出现两次，表明在剩男剩女的单身生活中，工作和朋友占据了比较重要的位置；“年龄”和“时间”两个词语则体现了剩男剩女对年龄增长的无奈和焦虑。

4.3. 微博话题折射男女性别差异

通过对剩男和剩女的微博分别进行分析发现了明显的性别差异，即女性倾向于讨论与人际相关的话题，男性倾向于讨论与权力相关的话题，这与进化心理学中男女的社会角色所导致的心理差异相符合。根据性别角色理论，男性或女性会依据社会或文化的期待来表现适当的行为模式[11]。本研究的结果发现，剩男剩女分别关注较多的话题与其性别角色是相吻合的，剩女更多关注情感、生活、人际等方面的话题，体现出更多的是女性感性思维的风格；剩男则更多的关注科技、工作、时事等方面的话题，体现出更多的是男性理性思维的风格。这可能表明，剩男剩女对自身的性别角色感知比较符合性别角色的社会规范，其社会适应性水平整体处于良好的状态。

5. 建议

5.1. 大众应理性对待剩男剩女群体

剩男剩女由于逐年的增多成为一种社会现象而引发大众的关注，而大众对该现象多持消极态度，认为剩男剩女的增多会导致社会问题。本文通过研究在一定程度上证实了大众对剩男剩女持负向态度，且多数人认为剩男剩女增多的原因在于青年本身，但通过探讨剩男剩女本身的社会态度发现其社会态度呈中性，即剩男剩女能以客观理性的态度看待社会，且剩男和剩女符合社会对他们的性别角色期待，侧面反映其社会适应性良好。有学者指出剩男剩女不是当代青年和青年群体出了问题，而是当代社会的体制性问题[12]。因此社会大众应以理性客观的态度对待剩男剩女，不让剩男和剩女感到周围人对他们的压力及偏见。剩男剩女感受到最大的压力来自于父母，这就要求身为父母应当从孩子的角度考虑婚姻问题，理解孩子的选择，成为剩男剩女坚实的依靠。

5.2. 剩男剩女应树立正确的择偶观

从个人角度讲,有些剩男剩女之所以被剩下是由于存在不恰当的择偶观。某些适婚青年由于受经济社会的影响,过多的将外在条件纳入到择偶选择中,导致择偶困难。婚姻和爱情本质上是男女双方建立在情投意合基础上的一种情感契约,如果将情感之外的经济条件、社会地位、外在评价等因素掺入进来,必将会侵蚀本该单纯如一的情感本身,也会形成扭曲的择偶观,进而将自己一步步推向剩男剩女的境地。有些青年过于挑剔对方或其家庭的经济条件而难以找“理想”的伴侣,殊不知“成家立业”一针见血地指出“成家”不但可以在“立业”的前面,而且“成家”对更好的“立业”还有积极作用。由于一份真挚的情感而选择和所爱的人相知相伴,不必要也不应该附加情感之外的各种条件。

基金项目

山东省应用基础型人才培养特色名校建设(应用心理学专业);2014年度山东省统计科研重点课题(No.2014重点29)。

参考文献 (References)

- [1] 雪杉 (2010) 剩男剩女现象对社会的影响不容忽视. *东北之窗*, 20-21.
- [2] 网易科技 (2014) 第33次中国互联网络发展状况统计报告. <http://www.199it.com/archives/187745.html>
- [3] 杜伟夫 (2010) 文本倾向性分析中的情感词典构建技术研究. 博士学位论文, 哈尔滨工业大学, 哈尔滨, 3-4.
- [4] 王洪伟, 刘颢, 尹裴, 廖雅国 (2010) Web 文本情感分类研究综述. *情报学报*, 5, 931-932.
- [5] 孙铁利, 刘延吉 (2009) 中文分词技术的研究现状与困难. *信息技术*, 7, 187-189.
- [6] 张华平 (2014) NLPIR 汉语分词系统. NLPIR 简介. <http://ictclas.nlpir.org/>
- [7] Tagxedo (2014) 词云技术. Tagxedo-Creator. <http://www.tagxedo.com>
- [8] CNKI (2007) 知网(HowNet). HowNet's Home Page. <http://www.keenage.com>
- [9] 陈晓东 (2012) 基于情感词典的中文微博情感倾向分析研究. 硕士学位论文, 华中科技大学, 武汉, 21-23.
- [10] 余伟成 (2013) 基于微博的热点发现与情感倾向分析. 硕士学位论文, 云南大学, 昆明, 40-41.
- [11] Scanzoni, J. and Fox, G.L. (1980) Sex roles, family and society: The seventies and beyond. *Journal of Marriage and the Family*, 42, 743-756.
- [12] 王彬 (2012) 从经济学视角分析“剩女”现象. *中国青年研究*, 6, 81.