# 研究生就读期间状态的聚类分析

**摘要：**为深入了解新时代下研究生就读期间的心里状态和压力，促进研究生德智体美劳全面发展，本文改变以往传统使用问卷或者面对面访谈的形式采集数据，运用爬虫软件，python，R语言等工具，直接爬取微博等平台上关于研究生压力以及一些当下研究生话题热点等情况的评论，对收集到的文本数据进行分词，按照数据分析进行权值叠加，进行词频统计和聚类分析。研究生状态可分为学业状态（包括毕业，就业，实验等）和生活状态（包括学习，情感等），以及心情期望（研究生们通过各种方式释放压力，调节心情，对明天充满期望）。

**关键词：**压力，心理，聚类分析，文本分词

## 引言

进入新的世纪，高等教育范围不断普及［1］，研究生成为我国科教兴国的骨干力量，在高等教育要求不断提升的情况下，在校研究生的工作、生活、学习、科研等都发生了天翻地覆的变化，各种压力不断增加。随着研究生扩招的深入推进，研究生的精英角色受到了冲击，这对于研究生的毕业和就业问题产生了巨大的影响，由此也产生了更多的压力，压力的增加使得当代研究生的心理健康问题日渐突出。

不久前南京大学女博士失足坠楼，大连理工大学研三学生在实验室自缢身亡，留下一封遗书离开了这个美丽的世界，说自己下辈子只想做一只猫，慵懒悠闲，再次把研究生的心理健康问题推向了舆论热点，引发大众的思索。研究生作为社会的高素质人才，对当今社会的发展起着不可或缺的作用，如何充分了解研究生的吐槽点，进行合理解释并提出相对应的解决方向，成了当下研究的一个热点。2010 年，教育部《关于进一步加强和改进研究生思想政治教育的若干意见》明确提出要“加强研究生心理健康教育和咨询工作” ［2］

压力对人的身心都会产生巨大的影响，不仅仅是正面的积极的影响，更为人知的是对身心的负面影响。一个人如果长期生活在压力下，会引起很多负面的情绪反应，例如恐惧、焦虑、抑郁、冷漠等，它是形成心理适应困难甚至心理疾病的重要原因之一。辛玲玲与张平指出研究生主要压力可分为学业压力、就业压力、人际关系、经济压力和其他压力，且研究生的身心健康状况普遍差于本科生［3］。蔡爽指出研究生主要压力为科研压力、经济压力、感情压力、就业压力［4］。

本研究采用质性研究方法，获取微博平台上一些话题例如研究生辛苦、研究生难等，当下一些热点事件，例如大连理工大学化学系研三学生自缢等。爬取相关论题的评论、点赞、转发等数据作为本文的数据源。通过聚类分析深入挖掘文本数据背后的信息［5］，研究新时代研究生对自身处境的看法、建议与展望，让研究生能够充分在象牙塔尖绽放自己的光彩。

## 工具与方法分析

本研究采用的工具包括R语言，python，SPSS等软件，采取不同方法尽可能的爬取相关网站上与课题相关的信息。网络爬虫数据来源可分为三大类：第一类是研究生相关话题下所有的评论、点赞、转发信息；第二类是研究生压力相关的文章下面的评论；第三类是当下热点事件例如“大连理工大学研究生自缢”，“南京大学女博士失足坠楼”等的相关评论。平台上爬取的数据采用excel的格式进行保存，如表1和表2所示，采集的时间段位2020年10月15号-10月31，共计5272条数据。

表1：热点事件评论摘要



表2：相关话题下评论摘要



整理爬虫搜集的文本数据作为基础数据，利用Python 工具中的分词数据分析包将文本进行分词处理。工具包中的分词功能可以将句子较为精确地切开，把句子中所有可以成词的词语都扫描出来，方便我们进行文本分析，再进行数理统计分析得到词频统计数据。根据词频数据生成词频矩阵，利用 SPSS 软件系统谱系图，以便直观地分析评论的类型。在聚类分析基础上，统计每个类别中分词出现的频率，按照频率高低进行词语分类。筛选出一部分高频代表词语生成词频矩阵［6］，进行编码，再进行聚类得到最佳分类。

## 数据分析

### （一）数据统计

共搜集到关于研究生现状相关的爬虫数据5272条，其中突发事件共1331条，话题热点等共3941条，包括用户ID、用户头像、用户名、字段长度、发布时间、微博链接、关键词、话题、文字链接、手机型号、评论、转发、点赞等等一系列共计47个字段，存在大量重复分词字段，大量空白、缺失、重复等情况。

突发事件和话题热点爬取的数据储存格式、字段、类别不一，需要进行数据清洗整理。爬虫时往往将话题、热点、事件等作为一个字段爬取下来，因此存在一部分词语例如#研究生累#、#大连理工大学研究生自杀#等需要进行标注，方便在分词的时候进行剔除。整理统计“研究生”，“数字”，“看到”等作为停词、重复词、无效词，否则将影响主题词的重复计数和后期关键词的统计。

存在一部分噪声干扰数据，与我们的研究方向完全不相关，例如广告，水军等情况需进行识别剔除；存在一部分数据含有英文词语和表情包等特殊文本，无法进行分词分析，需要进行删除整理。

### （二）数据清洗

将不同excel文件进行数据整合，利用excel软件进行数据清洗，只保留用户名、发布时间、文本、评论、点赞、转发这6个字段 ，删除评论空白值后共计5173条评论数据。

利用过滤词txt文件进行噪声数据筛选，剔除无用嘈杂的干扰评论479条后剩下4694条评论数据。对文本数据进行去重操作，剔除反复出现的文本数据保留点赞最多或者最近时间发布的一条，删除重复评论之后剩下3274条真实有效数据。

按照转发、评论、点赞数，可以得出该条评论或者微博在人们心目中的赞同程度，因为转发和评论数据缺失严重，故本实验按照点赞数进行文本信息的权值附加，对于高点赞数的文本按照权值进行翻倍，以生成5000条有效评论，使样本数达到一定期望。

### （三）分词选择

所有评论数据最终只剩下用户名、发布时间、文本、点赞四个特征，其中点赞存在一小部分空白值，使用0值进行填充，确保没有缺失值。

调用python中的分词包将所有评论进行分词，生成词包，统计所有分词出现的频率，按照从大到小进行排列，生成分词频率表，选出出现频率最高的50个词语，剔除停词、重复词、无效词例如大学生，研究生等。见表3

表3：分词频次统计表



根据分词频次统计表生成词云，更为清楚的看到当下人们更为关注的一些词语。见图1

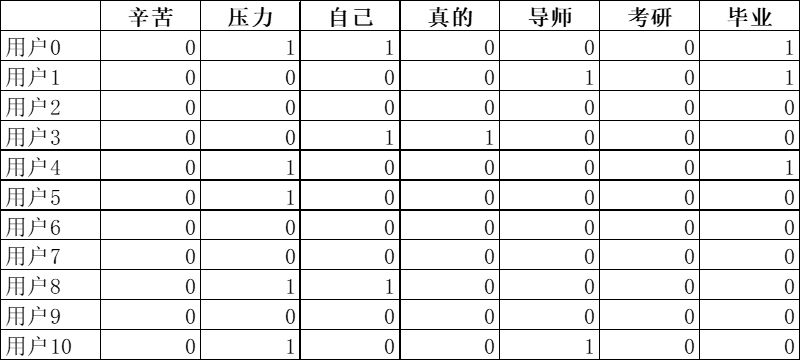


图1：词云

### （四）聚类分析

将这些词语代入原始数据。利用one-hot编码判断每一个用户是否出现该词，若出现则为1，反之则为0，生成用户词频分析表。见表4

表4：用户词频分析表



将所有数据放进SPSS统计分析软件，采用极差标准化方法对分类变量进行标准化处理，采用平方Euclidean距离作为区间变量测度，以相似性最大化为原则，利用离差平方和法进行聚类分析，得到聚类树状图[7]，如图2，将用户分成3小类，用户数分别为948，844，1482。

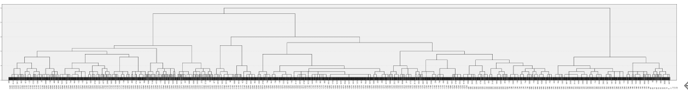


图2：聚类树状图

统计三个分类中各个分词出现的频率生成分类词频表，选择每个类中分词出现最多的前十个词语作为该类的代表情况，若有词语交叉出现不同分类之中，则以频率作为分类主导，将三类高频关联词集合分别进行命名为学业状态、生活状态、心情期待，学业状态和生活状态可以归为压力来源。详情见表5

表5：分类词频表

|  |  |  |
| --- | --- | --- |
| 类别 | 高度关联频率词 | 类型定义 |
| 第一类 | 自己，压力，一个人，问题，毕业，我们，导师，大学，生活，很多，因为，论文，实验，科研 | 学业状态 |
| 第二类 | 辛苦，工作，现在，努力，考研，考上，今天，一定，看到，读研，学习，每天，本科，一直 | 生活状态 |
| 第二类 | 希望，世界，快乐，知道，活着，生命，可能，顺利，好好，开心，人生，朋友，同学 | 心情期待 |

根据之前所有的数据分析，再将加权处理后的5000条微博数据进行分类识别，将聚类得到的三类状态结果做成柱状图（图3），可以从中看出出现学业状态的评论存在3169条，占比为63.38%，可以说绝大部分的人都谈到了自己的学业状态问题，讨论频次甚至超过生活状态和心情期待的总和，出现生活状态的评论较少，比学业状态的评论一半略多，存在1847条，占比36.94%，出现心情期待的评论最少，大概是生活状态的一半，存在842条，占比16.84%。

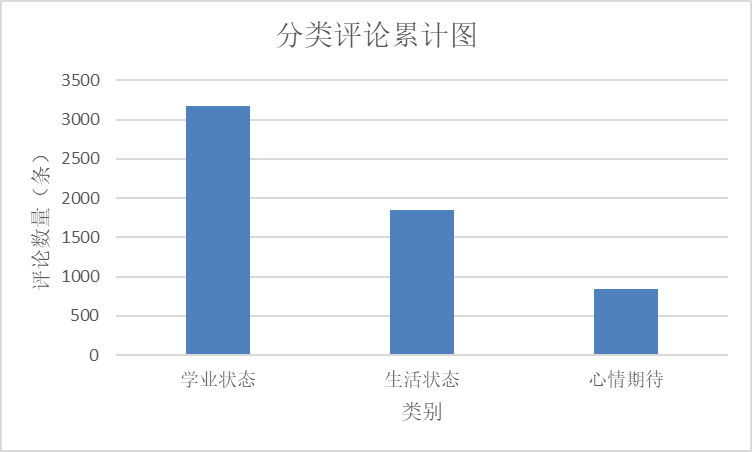


图3：分类评论累计图

## 情况分析

一般的问卷或者访谈有所遮掩或者不方便回答，而微博上的数据由于匿名性完全是内心的真实体会，不需要去掩饰什么，我们可以通过数据分析精准的进行探讨。本文通过微博数据进行分析将压力分解为学业状态和生活状态，将面对压力时的心理状态单独划分为心情期待。其中学业状态包括论文、实验、就业等方面，生活状态分为经济、学习，心情期待包括人生、希望等。通过数据分析得到的结果和当下研究生心理压力的分析大致相同。

压力与心理健康有着十分紧密的联系，而巨大的压力和不健康的心理正是研究生们选择不理智做法的根本原因[8]。关于研究生就读期间状态和心理压力将具体从这三个方面出发来进行分析。

### 学业状态

学业状态是所有研究生都比较注重的一个问题，出现频率也是最高，几乎参与评论的人都提到了，其中包含着毕业，论文，试验，导师，实验室等等一系列问题，困扰着每一个研究生，如何进行针对性的分析这些问题出现的原因，合理引导研究生的压力，减少负担，成为当下研究的一个重点，这里我们从高频词汇出发研究其起源和可能解决方向。

自己、我们词语出现在前列，结合感同身受这个词，可以看出当下研究生们面对的状态、压力等大都类似，每每遇到话题总能够引起大家的共鸣，可以进行整体化分析。压力，众所周知，随着研究生体制的不断变革，研究生作为高等教育精英需要面对的压力也在不断增加[9]，生活、科研、毕业、情感等多方面的压力都在升级，故需要在保证研究生质量的前提下合理的减压，结合分词一个人可以得出研究生的心理问题来自于压力过大。

由于大连理工大学自缢研究生提到了导师的不作为的问题，许多人也开始纷纷吐槽自己的导师，当然更多人开始庆幸自己有一个有担当有作为的导师。研究指出有些研究生非常希望被导师认可和关心[3]，但是缺乏与老师友好相处的能力，在遇到困难时或者被询问时惴惴不安，出现逃避和拖延行为，既无法获得导师的指导，也无法解决问题。其实关键就在于如何和导师沟通，导师更像是家长、朋友，完全可以和导师就各方面进行心贴心的交流，不管是学业还是生活，导师都可以给你提出建议和方法，然而很多的问题都是沟通不够、不充分引起的，反倒还会加剧压力[10]。

论文、实验、科研可以归结为学术能力，存在的一系列问题是实验室设备老旧，严重影响科研进度，再影响学生毕业论文的撰写，由此引发延毕、留级等问题，大连理工大学研三学生的自缢莫过于太过担忧延毕，心里想不开才做出令人痛心的行为。近年来国家对研究生培养质量的重视程度不断提升[4]，对研究生毕业的论文的质量要求也在不断提升，一些学校甚至规定必须在SCI收录的杂志上发表论文才允许毕业，这无形中对研究生的实验和科研能力提出了高要求。如何打破唯论文标准、授予导师的一票毕业权等关键问题需要一个个学校的不断试点，学生和老师的充分沟通，学校和教育部的充分沟通，虽推动缓慢，但一直正在努力中。至于设备、环境这些硬件，可以向导师和学校沟通反应，看能否获得更多的支持如设备的优先使用，少部分设备的持续更新、申请网上计算资源等等，也可以和师兄弟们协力合作突破课题。

毕业可以理解为就业压力，研究生教育规模的不断扩大导致原有的教育配套资源跟不上，会导致小部分研究生的素质不足，造成就业压力。还有因为专业、薪水、工作地点等问题都会在就业上被提到，就业压力是当代大学生必须面对的头等大事！南京师大教科院的一项研究覆盖全国17个省市50所高校，发现超过72.3%的研究生都会担忧就业问题。2007 年南京师范大学全国研究生问卷调查，结果“就业压力占 61% [11]。2010年郭敏的研究结果表明，研究生心理压力主要来自就业、学业和经济等方面，其中就业的压力最大[12]。由此可见就业几乎是所有研究生的最大心病，如何使得人尽其用，充分就业是最为重要的一个话题。可从三个方面进行改善：1.学校深化课程改革，拓宽就业途径，跟上社会对专业的各种需求；2.树立第一责任人意识，提升导师的就业指导能力；3.研究生自己正确规划，合理定位，提升就业竞争力[13]

自己、压力、一个人三个分词频率很高，可以看出孤独是研究生就读期间的一个常态，面对着新时代研究生的高标准、高要求、高期待所衍生的各种学业压力、问题都需要自己一个人去承受，去解决，要锻炼自己坚强的意志去面对压力。

### （二）生活状态

超过70%的研究生第一学历大都是普通本科而非985、211高校，本科生活较为闲适，在迈入研究生生涯的时候，对于学校的学业安排以及教学要求等会出现不同程度的不适应[1]。再加上有不少同学考研时候选择了跨专业，面对新学校新环境新专业，面对的压力会更大。

所以出现频率最高的是考研、辛苦、学习、每天、本科等词语，点赞人数最多的一个评论是：考研的时候以为现在已经是最辛苦的时候，结果没想到研究生生活比考研还要辛苦，还要累，还要忙碌，需要承受的东西更多。每一天都在辛苦学习，一直在看论文，一直在做试验，相对应的就是看不完的论文，做不完的实验，写不完的总结，干不完的工作，熬夜、缺觉、困乏成为研究生就读期间的一种常态。如此繁忙的节奏，忙碌的生活安排，影响研究生们的人际交往和情感，给研究生带来相当严重的心理问题，许多研究生没有较好的情绪调节能力，所以会出现焦虑、抑郁等情况。

合理安排生活作息刻不容缓，并不是长时间高负荷的工作就可以带来足够的产出，相反合理的休息、娱乐放松可以带来更为高产的效率。各高校可以开放研究生参与本科社团活动，学校学院多组织研究生活动，增设放松休息区，以此加强研究生的休闲娱乐生活，学校还可以开设心理健康课程和心理辅导站。研究生自己也可以组织活动、组织联谊，认识新朋友，不仅可以交流一下学术，也可以交流一下生活情况，

一个良好的社交环境可以给人带来更好的精神状态，好的精神状态是做好工作的一个基本条件，合理安排生活和工作，适当娱乐，将自己的研究生生活做好规划，既繁忙又充实，生活不仅仅是学习和科研，更是德智体美劳多样化发展。

### （三） 心情期待

虽然都会承受学业、就业、情感等不同的压力和期望，在面对其他人的不良状态或者不佳行为的时候，很多人自发进行了安慰和心理指导。在沮丧、沉闷话题下许多人发了有关于祝愿、期待、希望的评论，展现了压力下研究生们的美好心态。

在心情期待这一类词语中出现了希望、活着、开心、快乐等词语，可以归结为大家对于美好生活的期待，对美好明天的期盼，实验有结果，科研有进步，导师有鼓励等等。朋友、同学、家人等词语说明了大家在调整心情时的可求助对象，虽然压力、问题和事情只能你自己去抗，但是你拥有很多支持你的人，他们会听你倾诉，跟你一起奋斗和前进，前进的时候路阻且艰，总有人陪在你身边，做你的灯塔，做你的指路牌，做你的拐杖，支持你，和你一起前进。

乐观一词说明拥有一份乐观的心态是人生中不可或缺的，从大连理工大学化学系研三学生留下一份遗书，以自缢来结束自己尚未绽放的生命，到南京大学博士失足坠楼，无不在说明调整乐观、积极面对生活的重要性。

人生、世界、希望、顺利等词语，表达了新时代研究生们对于美好人生的追求，世界那么大那么精彩，值得你好好的去追求，去奋斗。人生是那么绚丽多彩，还有那么多的东西没有体验，那么多的风景没有欣赏，你怎么就可以屈服在眼前，被这么一点压力就击垮，阳光总在风雨后，研究生仅仅是真正精彩人生的开端。

所以，请对未来充满希望，好好的努力、奋斗，一定会一切顺利的。

## 五、结语

根据这些搜集的研究生评论的分析，可得出研究生就读期间学业压力繁重，休闲娱乐项目较少，心情十分压抑、沮丧。经过二十多年的研究积累，对于研究生心理方面的研究已经取得了相当丰富的成果，但是一些新出现问题和具体实施还需要大家深入探讨。高校应当重视新时代下研究生的状态和心理压力，并想办法去缓解这些情况，本研究综合一些学者的论文研究，提供了一些大概的解决方向。

新中国迈入新时期，在全面建成小康社会、决战脱贫攻坚的基础上迈向社会主义现代化国家中新征程中，迫切需要大批高素质、高水平的德才兼备的优秀研究生。需要深化研究生培养模式改革，完善研究生教育体系，推动研究生教育高质量发展。

参考文献：

1. 陶林.近20年来国内关于研究生压力的研究综述[N].北京科技大学学报.2020(4)：36-42
2. 周莉,李安祺.研究生心理压力成因和自我调节策略探析[J].北京教育.2020（9）：72-74
3. 辛玲玲,张平．理工科院校研究生压力源探析［J］.思想教育研究.2008（6)：52-54
4. 蔡爽.当代研究生心理健康教育探析[J].当代教育实践与教学研究.2019(24)：235-236
5. CABEDO R, EDMUNDO T C, CASTRO M. (2016). A Benchmarking Study of Clustering Techniques Applied to a Set of Characteristics of MOOC Participants. 2016 ASEE Annual Conference & Exposition，New Orleans, Louisiana.
6. 贾文军.大学生在线学习体验的聚类分析研究[J].中国高教研究.2020(4)：23-27
7. 刘博文.基于Ward系统聚类的吉林省农用地整治分区及策略研究[N] 江西农业学报.2019(3)：119-124
8. 张清扬.当代研究生心理压力来源与疏导[J]科技视界.2019(11)：245-246
9. 禹玉兰,郑希付.扩招时期研究生心理压力现状及其相关影响因素[J].心理发展与教育.2005(3)：99-103
10. 张新颖.95后研究生心理压力状况及对策建议[J].管理观察. 2019(36)：118-120
11. 钱洁,吴海娜,王磊.研究生就业: 站在理想与现实的十字路口［J］.教育与职业， 2007（19）：70-72
12. 郭敏.武汉大学研究生心理健康现状调研报[J].学校党建与思想教育.2010(28)：60-52
13. 谭晓雪,胡光丽.基于就业压力的医学研究生就业现状调查[J]. 中国高等医学教育.2019(2)：32-33

Clustering Analysis of Postgraduate mental state

FU Hangtian

Qiqihar university

**Abstract:** In order to deeply understand the psychological state and pressure of graduate students in the new era, and promote the comprehensive development of graduate students' moral, intellectual, physical, and artistic, this article changes the traditional use of questionnaires or face-to-face interviews to collect data, using crawler software, python, R language and other tools. Directly crawl the comments on postgraduate pressure and some current postgraduate topic hotspots on platforms such as Weibo, segment the collected text data, superimpose weights according to data analysis, and perform word frequency statistics and cluster analysis. The status of graduate students can be divided into academic status (including graduation, employment, experiments, etc.) and life status (including learning, emotions, etc.), as well as mood expectations. Graduate students use various methods to release stress, adjust their mood, and look forward to tomorrow.

**Key words:** pressure, psychology, cluster analysis, text segmentation