基于心愿墙的大学生情绪结构文本挖掘研究

越 缙 吴慧琴

安徽文达信息工程学院计算机工程学院 安徽 合肥 231201

【摘 要】：本文对当代大学生的情绪状态及其关联进行了研究，利用学校里较为常见的心愿墙这种匿名社交媒介收集研究的数据，再对收集的数据运用扎根理论进行初步处理，处理后的数据运用关联规则算法进行文本的挖掘。主要研究结果有：通过扎根理论归纳出影响大学生情绪的 8 个主要类型范畴，进一步归纳为 4 个主要类属；通过关联规则挖掘得出大学生情绪影响因素的联系机制，大学生负面情绪的出现受到学生所处环境、社群、个体因素共同作用。在此基础上提出了应对建议和进一步的研究方向。

【关键词】：情绪结构；文本挖掘；关联规则

Research on the Text Mining of College Students' Emotional Structure Based on Wish Wall

Jin Yue Huiqin Wu

School of Computer Engineering, Anhui Wenda Information Engineering College Hefei Anhui 231201

Abstract:This paper studies the emotional state and its association of contemporary college students, using the anonymous social media of the more common wish wall in schools to collect research data, and then preliminarily processes the collected data using the root theory, and the processed data uses the association rule algorithm to excavate the text. The main research results are: through the grounded theory to summarize the 8 main types of categories affecting the emotions of college students, and further summarized into 4 main categories; through the correlation rules to explore the linkage mechanism of the factors affecting the emotions of college students, the emergence of negative emotions of college students is affected by the environment, community and individual factors of students. On this basis, suggestions for response and further research directions are proposed.

Keywords:Emotional structure;text mining;Correlation rules.]"

最早的许愿墙（也叫心愿墙）用来承载人们的愿望，一般是建筑或者树木，人们在上面写字涂画或者粘贴纸片，表达自己的愿望、寄语、祝福等。现在，随着计算机网络的发展，人们可以在互联网上建立虚拟的许愿墙，功能与实物的一样，一般是一个网站或者一个独立的网页，学校、企业单位中常有应用，网络许愿墙给人提供一个抒发感情、宣泄心事以及能够互帮互助的平台。

大学生是社会上年轻且极为活跃的群体，对网络许愿墙的接受程度相对较高，许愿墙发表言论一般是匿名的，是一种匿名社交。[1]参考 Mob 研究院发布的《2019 年陌生人社交行业洞察报告》，在匿名社交应用中，年龄在 24 岁以下用户占总数的比例为 95%，年龄在 45 岁及以上的用户群体占比大概只有 1%；从学历层次分布的来看，本科以上学历占比不大于 20 %。[2]从 Mob 研究院调查的用户学历和年龄特征来看，匿名社交媒介中的主要参与者以年轻用户为主，不同于实名社交的微信、微博，用户群体特征较为相似，且年轻用户普遍缺少社会经验。受人的心理因素影响，心愿墙显然是年轻大学生们吐槽与获得心理慰藉的“安全场所”。话题也较多以愿望需求与情绪释放为主。对大学校园内的心愿墙网站数据进行研究，有助于了解当代大学生在移动互联网下的社会生活形态、[3]思想观念、心理状态、情绪影响因素，以及各个因素之间的关联，对于解决大学教育的相关问题具有一定的指导意义。[4]本文以大学校园心愿墙的文本词条为研究样本，运用数据挖掘方法探索当代大学生的情绪特点及其关联因素。

1 原始数据及其预处理

1.1 数据来源

本研究的文本样本来自于某高校的校园许愿墙网站，许愿墙发布的留言允许匿名。我们对发布许愿墙网站上的留言词条进行收集整理，共采集到 5500 余条语句。

1.2 数据的预处理

原始的留言词条难以直接进行挖掘，需要进行预处理。本文采用扎根理论法对收集的语句进行整理归类，同时扎根理论也可以得出相应的分析结论，可以与关联规则分析的结

论相互验证。应用扎根理论研究的主要流程如图 1 所示。

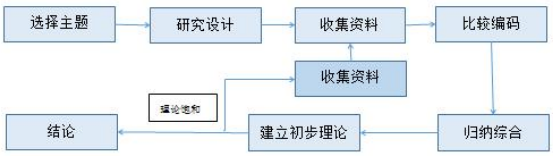

图 1 扎根理论研究流程

在收集的 5500 余条语句中随机选取 1200 条，运用 ROST Content Mining System 挖掘软件的词语标注工具对 1200 条语句进行标注，部分没有确切意义的语句直接进行删除，最后剩下 1102 条可标注，形成标注词表。对标注词表的词语进行编码（见表 1）并将词表内容归到尽可能多的概念类属下面以后，将编码过的词语在同样和不同的概念类属中进行对比，为每一个概念类属找到属性。这里每一个概念类我们称之为主范畴，再将主范畴与它们的属性进行整合，对这些主范畴进行比较，考虑它们之间存在的关系，将这些关系用某种方式联系起来。形成主要类型，确定该理论的内涵和外延，将初步理论返回到原始资料中进行验证，对剩余的 4300 条语句进行编码，并归为相应的概念属性中。同时不断地优化现有理论，使之变得更加精细。 对理论进行陈述，将所掌握的资料、主范畴、主范畴的属性以及主范畴类属之间的关系一层层地描述出来。

限于篇幅，表 1 中列出了其中具有代表性的部分语句，共计 16 个范畴。

表 1 心愿墙语句情绪范畴编码统计表

序号	初范畴	标注（条）	百分比（%）	代表性原始词条语句
1	人际关系	55	4.97	今天认识了新的朋友，挺谈得来的
2	孤独情绪	53	4.81	孤单其实并非真的孤单，至少有寂寞可以相伴
3	融入障碍	52	4.72	无法融入这个集体，感觉就是融入不了
4	相处体验	50	4.54	知人知面不知心，有的人真是难以相处
5	不良作息	47	4.26	决定了！今晚还去网吧包夜
6	手机强迫	42	3.81	老是不自觉地碰手机，尝试不打电话不碰手机
7	就业压力	41	3.72	班里不少人都确定工作了！工作真难找啊
8	适应环境	40	3.63	迎来向往的大学生活，不知道是否都习惯呢
9	厌学情绪	39	3.54	这段时间心情烦躁，无心学习
10	情绪调节	39	3.54	在这里吐槽一下情绪，无力在现实中诉说
11	物化焦虑	33	2.99	上课不给带手机，真是煎熬
12	情绪扩散	32	2.90	大家不要怪我发火，我也是在外受了气
13	考试紧张	31	2.81	明天考四级，最后一次机会，睡不着，希望能过
14	认知迷茫	30	2.72	感觉学的课程没什么用，自己未来不知道能做什么
15	沉迷游戏	30	2.72	沉迷游戏有不好也有好处吧
16	抑郁倾向	26	2.36	我讨厌现在的自己，一无所有，满身疲惫，就连“未来”两个字，都不敢轻易提出口

对多个范畴进行联系归类，最终归纳出了 8 个主范畴：集体融入、环境感知、情绪障碍、情绪感染、课程考试压力、毕业就业压力、手机依赖、网络成瘾。通过进一步的分析，这 8 个主范畴又可分为 4 个主要类属：群体认同、情绪动力、压力应对、网络依赖。

2 关联规则的挖掘

在对数据进行基础的清理转换后，我们使用 Apriori 算法挖掘关联规则，得出相关因素的支持度和置信度。在运用算法之前，首先需要详细了解关于传统关联数据分析的一些重要基本概念。关联数据分析过程是一种在一个大规模海量数据集中不断寻找相互关系的复杂过程，这些相互关系大致可以认为有三种表现形式：

①项集：事物（物品）的集合。

②频繁项集：经常出现在一起的事物（物品）的集合。

③关联规则：暗示两种不同事物（物品）之间可能存在很强的关系。

而数据挖掘关联规则算法的作用就是寻找这些数据当中的那些具有强烈关联性的规则。强关联规则指的是满足最低支持度和最低置信度的规则，它们是关联规则挖掘算法在应用过程中最重要的概念。[5]

支持度：记作 Support(X=>Y)，指的是某一项集的频繁程度，是关联规则重要性的衡量准则，用于表示该项集的重要性。假设 count(X U Y)为同时包含项 X、Y 的项集数量，它与项集总数 countall 的比值，反映了 X、Y 项集同时出现的频率。

公式如下：

$$\text{Support}(X \Rightarrow Y) = \frac{count(X \cup Y)}{countall} \quad （公式 1）$$

置信度：记作 Confidence(X=>Y)，用来确定同时包含项 X、Y 的事务在包含 X 的事务中出现的频繁程度，即 Y 在 X 条件下的条件概率，是对关联规则准确度的衡量准则，表示规则的可靠程度。假设同时包含项 X、Y 的项集数量为 Support（X=>Y），它与含 X 事务项数量的比值为置信度。公式如下：

$$\text{Confidence}(X \Rightarrow Y) = \frac{Support(X \Rightarrow Y)}{X} \quad （公式 2）$$

一般来说，从 Apriori 的分析结果来看，关联规则的强度可以从它的支持度和置信度度量上展示出来。[6]

本研究分析行动者多次发布的词条内容，每个词条归为一个或者多个范畴，行动者和所发词条的范畴形成一个事务，采集样本中的所有事事务形成事务集合，利用 SPSS modeler 中的 Apriori 算法对事务集合进行关联规则挖掘。

通过数据可以看出，环境感知→群体融入（这里的“→”表示一组关联关系，同时也表示一定的因果联系，下同）、情绪障碍→群体融入、群体融入→环境感知、情绪障碍→环境感知四对关联规则支持度相对较大，表示相关的前后项同时出现的概率较大，同时他们对应的置信度数值也较高，表示在总体的事件中出现的概率也较多。单独从置信度数值来分析，情绪障碍→学业压力、网络成瘾→手机依赖、情绪障碍→毕业压力这三组数值较大，反映出学生学业压力大、毕业压力大的时候容易引发学生的情绪障碍，另外对手机依赖的学生，一般也容易网络成瘾。

利用 SPSS Modeler 可以对关联规则挖掘结果进行可视化，运用网络节点工具可以直观的表示各个范畴之间关联的强弱程度，体现他们之间的共现频数。如图 2 所示。

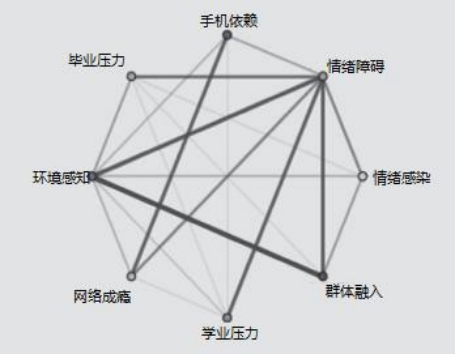

图 2 类别共现网络

图中的点表示各个影响大学生情绪的范畴因素，点之间的线表示关联关系，并用粗细表示关系的强弱。越粗的连线表示越强的关联关系。图中较粗的线有四条，对应前面表 3 中四条关联支持度相对较大的四对前后项。共现网络图中呈现的关联强弱程度基本与表 3 中的数据吻合。

结合所收集数中大学生的年级进行分析，可以进一步得出这样的规律，大学先生在刚进入大学阶段时，周围生活学习环境和集体的融入对大学生的情绪影响较为强烈，新生对新的环境和群体较为陌生，内心容易产生困惑。学校提供一个好的环境感知，教育管理者给予新生较多的帮助和心理疏导，引导他们对新环境和专业有个正确认识，帮助学生尽快地融入集体和新环境，有助于减少不良情绪的产生。[7]

等到学生逐渐适应大学生活后，课程学习和考试等学业问题成为大学生关注的中心问题，如果这时候学生不关心学业问题，那么较为空虚的生活通常会使得学生迷恋手机和网络，这些新的关注点在一定程度上容易使得学生产生负面情绪。这时应该教育学生认识到大学学习与以往的差异性，及时调整学习方法和习惯，建立多样的兴趣爱好，缓解学业压力，培养良好的兴趣。

大学生活的后期，学生面临着毕业、就业问题逐渐凸显，同样会引发学生心理情绪的变化，这时家人和学校老师应该帮助学生认识自己，做好未来规划，客观的衡量自己，确定发展方向，调整好心态。

3 总结

大学生是个特别的群体，智力发育基本成熟，但是人生经历相对较少，心理情绪容易受到内外界因素的影响。本研究针对大学生群体较为常见的负面情绪进行了分析，探查其中的关系，最终目的是希望给学校和家庭教育提供一定参考和帮助。研究的内容还存在一定的局限性，由于分析的数据主要来源于心愿墙这样的匿名社交媒介，这样的媒介所收集的信息具有一定的偏向性，以负面情绪较多。因此，研究结果的适用范围有限制。另外，数据获取的样本数量可能较为有限，从而可能造成结果的片面性，未来的研究中，希望可以采集更多的数据，获得更加科学可信的研究结果。

参考文献：

[1] 林志业，王永菁.匿名社交网络生态现状及存在问题研究——以匿名社交应用 Soul 为探讨对象[J].新闻研究导刊，2021，

12(3):59-60.

[2] 屠嘉俊，万娟，熊红星.父母支持对大学生人际适应性的影响：情绪智力的中介作用[J].心理科学，2016 (04):965-968.

[3] 赵莹.“情绪与面具”：大学生情绪管理创新实践研究[J].云南民族大学学报（哲学社会科学版），2013(02）：95-100.

[4] 王炎冰.融合用户标签和微博内容的用户兴趣社区发现[D].昆明：昆明理工大学,2014.

[5] 钱程.Apriori 算法在大学生情感素质研究中的改进与应用[D].上海：上海师范大学.2019：5-14.

[6] 王平水.关联规则挖掘算法研究[J].计算机工程与应用，2010，46（30)：115-116.

[7] Nakashima, K., Isobe, K., Ura, M.. How does higher in-group social value lead to positivemental health？An integrated model of in-group identification and support[J]. Asian Journal of Social Psychology，2013（16）：269-276.

作者简介：越缙（1983—），男，安徽合肥人，安徽文达信息工程学院计算机工程学院讲师；

吴慧琴（1999—），女，安徽芜湖人，安徽文达信息工程学院计算机工程学院学生。

依托项目：2021 年国家级大学生创新创业训练计划项目，项目号 202112810020；安徽文达信息工程学院 2021 年校级科研，项目号 XZR2021A15。