



申请代码	F020516
接收部门	
收件日期	
接收编号	6190070309



国家自然科学基金 申 请 书

(2019 版)

资助类别：	青年科学基金项目		
亚类说明：			
附注说明：			
项目名称：	基于大数据的青少年群体心理压力动态模型与预警系统建构		
申 请 人：	李琦	电 话：	010-58806836
依托单位：	北京师范大学		
通讯地址：	北京市海淀区新街口外大街19号后主楼1424A		
邮政编码：	100875	单位电话：	58807807
电子邮箱：	qili0308@163.com		
申报日期：	2019年01月24日		

国家自然科学基金委员会



基本信息

申请人信息	姓名	李琦	性别	女	出生年月	1991年03月	民族	汉族
	学位	博士	职称	工程师	每年工作时间（月）	10		
	是否在站博士后	否	电子邮箱	qili0308@163.com				
	电话	010-58806836	国别或地区	中国				
	个人通讯地址	北京市海淀区新街口外大街19号后主楼1424A						
	工作单位	北京师范大学/心理学部						
	主要研究领域	计算心理学，心理大数据						
依托单位信息	名称	北京师范大学						
	联系人	林小鹏	电子邮箱	linxiaojuan@bnu.edu.cn				
	电话	58807807	网站地址	http://keyanyuan.bnu.edu.cn/				
合作研究单位信息	单位名称							
项目基本信息	项目名称	基于大数据的青少年群体心理压力动态模型与预警系统建构						
	英文名称	Dynamic Model of Group-Level Stress and Construction of Early Warning System for Adolescents Based on Big Data						
	资助类别	青年科学基金项目				亚类说明		
	附注说明							
	申请代码	F020516. 计算医疗与健康技术				F020505. 社交网络与社会计算		
	基地类别							
	研究期限	2020年01月01日 -- 2022年12月31日				研究方向：其他研究方向		
	申请直接费用	21.3000万元						
中文关键词		心理大数据；压力源事件；群体动态；青少年；应激过程模型						
英文关键词		psychological big data; stressor events; group dynamic; adolescents; process-based model of stress						



中文摘要	<p>多项研究表明,青少年心理健康问题日趋严重。本项目基于群体心理压力动态的分析视角,深度融合心理学传统方法与大数据分析方法,旨在为青少年心理压力预测预警提供新方案,具有重要的理论与实践意义。针对研究问题的复杂性,拟采用多质多法、交叉验证的研究策略,深度交叉融合心理学领域专家知识、问卷调查、实验研究与社会媒体大数据分析方法,通过在试点中学开展群体施测,构建涵盖青少年个体因素、微博数据以及学校日程事件的综合数据集。在此基础上,构建群体心理压力动态模型,预测群体内心理压力的发展规律,并设计实现一个“群体心理应激过程分析、预测平台”,实时可视化地展示:①相似心理压力群体的聚簇、动态应激变化过程,②群体压力源事件与应激过程分析,③选定节点的实时压力预测结果,为学校、团体、社会提供群体层面的心理压力辅助分析预测工具。</p>
英文摘要	<p>A number of studies have shown that adolescent mental health problems are becoming more serious. Based on the analysis of group dynamics stress, this project deeply integrates traditional methods of psychology and big data analysis methods. It aims to provide new programs for predicting early warning of psychological stress in adolescents, which has important theoretical and practical significance. In view of the complexity of research questions, it is proposed to adopt multi-quality and multi-method, cross-validation research strategies, and deeply integrate expert knowledge in the field of psychology, questionnaires, experimental research and social media big data analysis methods. A comprehensive data set covering adolescent individual factors, microblog data, and school schedule events is to develop by conducting group testing in pilot secondary schools. On this basis, the dynamic model of group psychological stress is to construct to predict the development law of psychological stress within the group, and a “group psychological stress process analysis and prediction platform” is to design and implement. The platform will display in real time visually: 1) clustering of similar psychological stress groups, dynamic stress change process, 2) group stress source events and stress process analysis, and 2) real-time pressure prediction results of selected nodes, to provide group-level psychological stress analysis and forecasting tools for schools, groups, and society.</p>



国家自然科学基金项目资金预算表（定额补助）

项目编号：6190070309

项目负责人：李琦

金额单位：万元

序号	科目名称	金额
	(1)	(2)
1	项目直接费用合计	21.3000
2	1、设备费	0.0000
3	(1)设备购置费	0.00
4	(2)设备试制费	0.00
5	(3)设备升级改造与租赁费	0.00
6	2、材料费	3.00
7	3、测试化验加工费	3.00
8	4、燃料动力费	0.00
9	5、差旅/会议/国际合作与交流费	6.30
10	6、出版/文献/信息传播/知识产权事务费	4.00
11	7、劳务费	4.20
12	8、专家咨询费	0.80
13	9、其他支出	0.00



预算说明书（定额补助）

（请按照《国家自然科学基金项目预算表编制说明》的有关要求，对各项支出的主要用途和测算理由，以及合作研究外拨资金、单价 ≥ 10 万元的设备费等内容进行必要说明。）

项目直接费用合计：21.30 万元

1. **设备费：**0 万元
2. **材料费：**计划投入3.00 万元，用于购买存储微博大数据的大容量存储介质（刻录光盘、U盘、移动硬盘等），日常科研办公用耗材等。
3. **测试化验加工费：**计划投入3.00 万元，具体测算如下：
青少年人格量表测验被试费：200人*10元/次*1次 = 0.20万元
青少年PSS压力量表测验被试费：200人*5元/次*4次 = 0.40万元
青少年微博数据标注：0.1元/条*10万条 = 1.00万元
青少年生活事件量表测验被试费：200人*10元/次*4次 = 0.80万元
项目所开发“群体心理应激与预测平台”测试被试费：100人*2元/天*30天 = 0.60万元
4. **燃料动力费：**0 万元
5. **差旅/会议/国际合作与交流费：**计划投入6.30万元，具体测算如下：
 - （1）差旅费：计划投入3.05万元，测算如下：
 - a) 前往外地进行课题相关的科学考察、业务调研以及国内学术会议活动，平均每年1人次，出差3天，每人每次平均需5680元（=机票或火车票 2000元+3*（补助180元+住宿380元）+会议注册费2000元），3年共需约1.70万元。
 - b) 市内交通费，用于参加学术交流活动、文献检索、调研学习。每年0.45万元，3年共计1.35万元。
 - （2）国际合作与交流费：计划投入约3.25万元。
计划参加EDBT/UbiComp等心理学/计算机国际会议一次，1人次，预计6天，每人每次出国参加会议平均费用约3.25万元（其中会议注册费0.4万元，机票和城间交通1.8万元，住宿费6.7汇率*160美元/天*6天=0.6432万元，城间交通/公杂费6.7汇率*100美元/天*6天=0.402万元）。
6. **出版/文献/信息传播/知识产权事务费：**计划投入4.00万元。具体预算如下：
 - （1）论文版面费：预计发表论文3篇，版面费平均每篇0.60万元，总计1.80万元；
 - （2）文献检索查新费、打印费、资料费、图书费等预计每年0.50万元，3年总计1.50万元；
 - （3）专利申请费：预计申请专利1项，投入0.50万元。
 - （4）软件著作权申请费：预计申请软件著作权1项：投入0.20万元
7. **劳务费：**计划投入4.20万元。具体如下：
用于支付参加项目博士生和硕士生的劳务费。博士生500元/月，硕士生200元/月，平均每年两名硕士和本科生参与项目，工作时长10个月，共计（0.05万元/人/月*2人*10月+0.02万元/人/月*2人*10月）*3年=4.20万元。
8. **专家咨询费：**计划投入0.80万元。具体如下：
拟聘请高级职称专家5人次，每次预计1600元，总计0.16万元/次*5次=0.80万元
9. **其他支出：**0 万元



报告正文

(一) 立项依据与研究内容 (建议 8000 字以内):

1. 项目的立项依据 (研究意义、国内外研究现状及发展动态分析, 需结合科学研究发展趋势来论述科学意义; 或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景。附主要参考文献目录);

1.1 研究背景与意义

青少年心理健康问题日趋严重, 及时预测其心理压力的发展趋势, 对尚未发生的心理压力危机进行预警, 对于保障青少年健康成长具有重要意义。

心理压力, 也称为“应激反应”, 是由应激源引起的生理反应、心理反应以及行为改变。长期反复经历心理压力将引发病毒感染、头痛、睡眠障碍、消化问题、生育问题、泌尿问题、免疫系统弱化[NCI12], 以及严重的心理健康问题如抑郁症、焦虑症甚至自杀倾向[JER13]。处于特殊发展时期的青少年承受着来自心理压力的巨大冲击。我国卫生部门对中小学生的心理健康调查显示, 1.5 亿青少年中受到心理压力困扰的高达 3000 万人[LGJ14]。2017 年, 上海市教委对 2500 名中小学生抽查显示, 有 24.39% 的学生曾考虑结束自己的生命, 更有 5.85% 的孩子曾有过自杀计划。种种数据表明, 由心理压力引发的青少年心理健康问题危害巨大, 在心理危机发生之前进行及时预警十分重要。

心理大数据旨在融合心理学坚实的理论基础、实证方法与信息时代大数据的科学方法, 将是解决青少年心理健康问题的全新探索。

一方面, 传统心理学关于心理压力的研究至今已有 70 多年的历史。学者通过对大规模特定人群的调研, 构建了世界范围内通用的普适性量表 (如 PSS 压力量表[COH83], 生活琐事量表[KAN81], 青少年琐事量表[BO86][WD11]等), 积累了丰富的实验探索成果, 为心理大数据的开展提供了坚实的理论支撑。基于量表的心理压力研究具有理论坚实、测评结果可靠的优势, 但受到样本规模、时效性、人力需求等较大限制, 利弊并存。

另一方面, 近几年研究者基于社交网络对个体心理压力感知、个体心理压力预测等课题进行了系列探索[XUE14, LIN14, LY15], 使得大规模、实时、无入侵的心理压力研究成为可能。目前, 相关研究多从计算方法的角度展开, 偏重用户微博的表层特征分析, 心理学领域专家知识有待提升, 且尚未触及到压力源触发个体心理压力发生、发展过程的心理学本质所在。



综合上述两点，本项目基于群体心理压力动态的分析视角，通过深度融合心理学实证分析方法与大数据分析方法，旨在发现具有心理学意义的青少年群体，预测群体内心理压力的发展规律，并构建“群体心理应激过程分析、预测平台”。其中“群体”和“心理压力”是心理学概念，“预测”则是大数据分析的方法。该体系使得群体层面的青少年心理压力得到多学科透视，也决定了本项目在研究思路上的多角度交叉：从心理学角度提出问题，从计算大数据角度解决问题，再回归到心理学角度深化问题。本项目多质多法、交叉验证的研究策略符合科学研究的一般方法论，使得群体压力的心理学本质在大数据中得以充分显现。

申请人于清华大学计算机系取得博士学位，目前就职于北京师范大学心理学部，所在单位具备国内顶尖的心理学科研究基础，可为当前项目提供心理学领域专家知识、问卷调查、实验研究基础，为心理大数据分析的结果可信性与可解释性提供保障，这也是以往相关研究中的一大短板与难题。

1.2 国内外研究现状及动态

本节将从以下两方面阐述与本项目紧密相关的国内外研究现状及发展动态：

1) 传统心理学领域的群体心理压力研究方法，2) 基于微博的用户心理压力分析，以及 3) 社交网络中的生活事件检测。

1.2.1 传统心理学领域的群体心理压力研究方法

心理学研究认为，群体是个体心理压力产生、发展的重要媒介。一方面，群体共同经历的压力源事件是青少年产生心理压力的重要来源，如学校生活中的压力源事件，同伴关系产生的压力源事件[KAN81]。另一方面，情绪感染理论认为，人类倾向于模仿周围人的面部表情、语言表达、动作及行为，从而影响自身的情绪体验[HOF01]。传统心理学领域主要通过访谈法、量表法以及生理信号检测法对群体心理压力状态进行感知。

- 访谈法指的是由访员和受访人进行面对面交谈，从而对受访人的行为及心理状态进行评估的心理领域的基本研究方法。心理咨询师通过对个体在生活事件中扮演的角色、相关背景信息、个体的脆弱性等信息对其心理压力状态进行判断和评估。访谈法是公认解决心理压力问题最为专业的途径。

- 量表法以生活时间为指标，收集被试自我报告的主观压力感受。以 Holmes 提出的社会再适应量表为例[HOL67]，该量表包含 43 项生活事件，被试报告过去一年内是否发生各项事件，并累计相应事件的“生活变化得分”，分数大于 300 表示存在精神疾病风险，分数在 150-299 之间表示中度的压力，分数低于 150 表示轻微压力。量表法具有相对高的信效度，被广泛应用于心理学领域的研究。

- 传感器检测方法基于生理信号（如心率变化 HRV [BO13]，心电图 ECG



[SUN10], 皮电反应 GSR [SIE11], 血压和肌电图[RIG12]) 和身体信号(如手势、语音、眼睛注视、面部表情、瞳孔扩张和眨眼率等[BA12]), 从个体的行为、情绪和身体特征中感知其心理压力水平, 在实验室环境具有检测精准度高的优势。

目前, 针对青少年的群体心理压力监测主要基于访谈和量表的方法, 存在一些难以避免的问题。首先, 访谈法对时间和人力要求较高, 且面临着心理咨询师资源不足的现状[MZL17]; 其次, 量表法在个性化衡量方面存在不足, 难以做到实时检测和对未来心理压力的预测。

鉴于传统心理学研究方法具有高信效度, 但难以捕捉青少年心理压力变化的过程和深层机制。本项目拟将传统心理学领域的研究方法与大数据分析方法相结合, 基于心理学理论刻画群体压力变量, 基于大数据方法构建群体应激的动态模型, 搭建群体层面的压力预测体系, 从而解决传统方法存在的瓶颈问题。

1.2.2 基于微博的用户心理压力分析

I. 微博用户心理压力检测

目前基于微博数据进行用户心理压力分析的相关工作主要采用情感分析技术, 基于微博内容特征展开。压力级别检测的本质是一个分类问题, 通过抽取单条微博的内容特征(文本特征、图片情感特征、轨迹特征、好友互动特征等), 训练机器学习模型, 对压力级别进行分类。目前, 对于单条微博的心理压力检测工作已经比较成熟, 为后续开展更为深入的微博用户心理压力分析奠定了基础。

- **基于微博文本、图片等内容特征:** [XUE14]针对青少年四种类型的心理压力构建相应的压力主题词典, 基于单条微博的内容信息归纳了压力相关的 16 维特征, 应用高斯过程分类器, 对每条微博的压力级别进行 0-5 级别的划分, 取得了较好的效果。[XUE16]基于上述算法进一步对青少年微博慢性心理压力的变化模式进行了分析和归纳。[LIN14, LII14]通过训练深度稀疏神经网络模型, 加入微博所发布图片的情感属性, 基于跨媒体的特征空间对单条微博是否存在压力进行二分类检测。

- **融合用户轨迹特征:** [JIN16]提出融合用户的轨迹异常特征与微博特征进行心理压力检测。该研究通过收集用户的轨迹异常信息和微博数据, 构建了协同训练的半监督学习模型, 对用户的心理压力级别进行分类。

- **融合好友互动特征:** [ZH16]对用户微博下的评论特征进行了进一步挖掘和特征补充; [LIN17]发现用户的心理压力状态与其好友在社交媒体上的状态存在关联, 提出心理压力与社交状态的相关性, 并基于文本、图像和互动特征, 构建了融合卷积神经网络的因子图模型, 从而基于微博内容和社交状态进行用户心理压力检测。



II. 微博用户心理压力趋势预测

以单条微博心理压力检测技术为基础,对时间序列上用户未来某一时刻或一段时间的心理压力水平进行预测也被证明是可行的。

[LF15]考虑到个体心理压力变化趋势与股票价格走势之间存在的相似性,定义了压力级别烛台图以及压力变化趋势的反转信号,并基于微博心理压力序列,应用烛台图预测当前青少年在未来时刻的压力变化趋势。[LY15]进一步将模糊理论与烛台图表示方法相融合,提出了一种基于模糊烛台图的模型,对青少年压力水平达到峰值/谷值的时刻进行预测。

[LH15]应用线性季节自回归移动平均(SVARIMA)的时间序列预测方法,并考虑将青少年在几个预设的历史事件区间的平均压力值叠加在压力预测结果上。该方法叠加特定事件的影响,但忽略了来自微博的语义信息,以及事件与压力变化趋势之间的潜在关联。

不同于个体心理压力预测问题,群体层面的压力预测则是一个全新的问题。如何刻画群体级别的关键心理变量,捕捉群体内部个体之间的压力动态影响,预测群体的压力变化规律,并将理论模型应用于系统实践,均有待深入探索。

1.2.3 社交网络中的生活事件检测

社交网络个人事件检测是一个相对新兴的研究方向,主要应用自然语言处理技术对微博文本进行语义分析,关注于重大生活事件的定性抽取,较少关注伴随用户生活事件的情绪和行为特征,对事件引发个体心理压力变化的关联关系尚处于探索阶段。

[LJ13]使用无监督方法将微博中的个人重要生活事件生成了一个按时间顺序排列的时间表,其中列出了关于个人事件和特定时间事件的推文。该研究介绍了非参数多级别的狄利克雷过程模型,并进一步基于这些推文提取个人生活事件并生成事件时间线。

[LJ14]通过检测好友回复中是否出现了“祝贺!”(congratulations!)或“节哀”(condolences)等关键短语,从而确定该条推文与重要事件相关,并在此基础上应用条件随机场(CRF)模型提取事件属性。

[LIN16]提出基于社交网络的海量数据源,定性地检测压力源事件。通过应用混合多任务模型(hybrid multi-task model),从微博中识别有压力的生活事件主题和类别,并将检测到的事件关键词映射到SRRS量表[HOL67]的事件类别中,对照量表标记所检测到事件的心理压力级别。

上述基于单条微博的生活事件检测存在两个问题:

- **漏检误警问题。**压力源事件的发生及其影响是一个持续的过程,而上述基



于单条微博的事件检测工作没有考虑微博与微博之间的潜在关联，因此对于间隔发博的行为可能会出现漏检的情况；对于连续微博序列中的异常点，则可能出现误警的问题；

- **语料质量问题**。由于微博中修辞及表达方式的非正式性质、微博的字数限制等因素，基于单条微博识别具体的压力源事件实例和完整的发生时间、角色等信息存在较大困难。因此，如何跳出单点检测的惯有思路，对压力源事件进行更为精准的识别，当前项目将给出新的解决思路和实施方案。

1.3 学术价值与社会意义

本项目提出的群体心理应激过程分析、预测与实证体系，将深度融合传统心理学研究方法与大数据的分析技术。首先，这一创新理念回应了以往传统心理学研究难以触及的大规模、及时性、对群体应激过程实现动态监测的学术难题；第二，通过构建融合领域专家知识的结构化神经网络预测模型，给出学科交叉共融的解决方案；第三，将理论应用于实践，拟构建“群体心理应激过程分析、预测平台”，从而为学校、社会提供群体层面的青少年心理压力辅助检测、预警工具，对于探索信息网络时代心理学研究的新模式，推进面向公众心理健康的大数据服务技术，辅助青少年心理健康发展具有重要意义。

主要参考文献

- [AF15] Afshar H., Roohafza H. R., Keshteli A. H., et al. The association of personality traits and coping styles according to stress level[J]. Journal of Research in Medical Sciences the Official Journal of Isfahan University of Medical Sciences, 2015, 20(4):353.
- [BA12] Bauer G., Lukowicz P. Can smartphones detect stress-related changes in the behaviour of individuals? PerCom Workshops. 2012:423-426.
- [BO13] Bousefsaf F., Maaoui C., Pruski A. Remote assessment of the heart rate variability to detect mental stress[C]. International Conference on Pervasive Computing Technologies for Healthcare. ICST, 2013.
- [BO86] Bobo J. K., Gilchrist L. D., Elmer J. F., et al. Hassles, Role Strain, and Peer Relations in Young Adolescents[J]. The Journal of Early Adolescence, 1986, 6(4):339-352.
- [CH10] Che W., Li Z., Liu T. LTP: A Chinese Language Technology Platform[C]. The 23rd International Conference on Computational Linguistics, Demonstrations Volume, 23-27 August 2010, Beijing, China. 2010.
- [COH83] Cohen S. A global measure of perceived stress[J]. J Health Soc Behav, 1983.
- [DI08] Diaconescu E. The use of NARX neural networks to predict chaotic time series[J]. WSEAS Transactions on Computer Research, 2008, 3(3).



- [GR08] Gretton A., Borgwardt K., Rasch M. J, et al. A Kernel Method for the Two-Sample Problem[J]. 2008.
- [HOF01] Hoffman M. L. How automatic and representational is empathy, and why[J]. Behavioural & Brain Sciences, 2001, 25(1):38-39.
- [HOL67] Holmes, T. H. and Rahe, R. H. The Social Readjustment Rating Scale[J]. Journal of Psychosomatic Research, 1967, 11(2):213-218.
- [JER13] Jeronimus, B. F. and Ormel, J. and Aleman, A. and Penninx, B. W. J. H. and Riese, H. Negative and positive life events are associated with small but lasting change in neuroticism. Psychological Medicine. 43(11), 2403-2415, 2013.
- [JIN16] Jin L., Xue Y., Li Q., et al. Integrating Human Mobility and Social Media for Adolescent Psychological Stress Detection[C]. 21st International Conference on Database Systems for Advanced Applications. 367-382. 2016.
- [KAN81] Kanner A. D., Coyne J. C., Schaefer C., et al. Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events[J]. J Behav Med, 1981, 4(1):1-39.
- [KU99] Kumpfer K L. Factors and processes contributing to resilience: The resilience framework[J]. 1999.
- [LF15] Li Y , Feng Z , Feng L . Using Candlestick Charts to Predict Adolescent Stress Trend on Micro-blog[J]. Procedia Computer Science, 2015, 63:221-228.
- [LGJ14] 李国军. 浅谈心理咨询师发展现状与对策[J]. 人才资源开发, 2014(17):41-42.
- [LH15] Li Y., Huang J., Wang H, et al. Predicting Teenager's Future Stress Level from Micro-Blog[C]. IEEE International Symposium on Computer-based Medical Systems. 208-213. 2015.
- [LI97] 刘贤臣, 刘连启, 杨杰,等. 青少年生活事件量表的信度效度检验[J]. 中国临床心理学杂志, 1997(1):34-36.
- [LI14] Lin H., Jia J., Guo Q., et al. User-level psychological stress detection from social media using deep neural network[C]. Proceedings of the {ACM} International Conference on Multimedia. 507-516. 2014.
- [LIN14] Lin H , Jia J , Guo Q , et al. Psychological stress detection from cross-media microblog data using Deep Sparse Neural Network [C]. IEEE International Conference on Multimedia and Expo (ICME). 2014.
- [LIN16] Lin H., Jia J., Nie L., et al. What does social media say about your stress? International Joint Conference on Artificial Intelligence. AAAI Press, 2016.
- [LIN17] Lin H , Jia J , Qiu J , et al. Detecting Stress Based on Social Interactions in Social Networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2017:1-1.



- [LIN96] Lin T., Horne B. G., Tino P., et al. Learning long-term dependencies in NARX recurrent neural networks[J]. IEEE Transactions on Neural Networks, 1996, 7(6):1329-1338.
- [LJ13] Li J., Cardie C. Timeline Generation: Tracking individuals on Twitter[J], 643-652. 2013.
- [LJ14] Li J., Ritter A., Cardie C., et al. Major Life Event Extraction from Twitter based on Congratulations/Condolences Speech Acts[C]. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.
- [LY15] Li Y., Feng Z., Feng L. When a Teen's Stress Level Comes to the Top/Bottom: A Fuzzy Candlestick Line Based Approach on Micro-Blog[J]. Smart Health. 2015.
- [MZL17] 马兹林. 我国心理咨询发展的现状及对策[J]. 科学与财富, 2017(6).
- [NCI12] National Cancer Institute. Psychological Stress and Cancer. <https://www.cancer.gov/about-cancer/coping/feelings/stress-fact-sheet>. 2012.
- [RIG12] Rigas G., Goletsis Y., Fotiadis D. I. Real-Time Driver's Stress Event Detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(1):221-234.
- [SC86] Schilling M. Multivariate Two-Sample Tests Based on Nearest Neighbours [J]. Publications of the American Statistical Association, 1986, 81(395):8.
- [SIE11] Sierra A. D. S., Avila C. S., Casanova J. G., et al. A Stress-Detection System Based on Physiological Signals and Fuzzy Logic[J]. IEEE Transactions on Industrial Electronics, 2011, 58(10):4857-4865.
- [SUN10] Sun F. T., Kuo C., Cheng H. T., et al. Activity-Aware Mental Stress Detection Using Physiological Sensors[J]. 2010.
- [TA10] Tausczik Y. R., Pennebaker J. W. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods[J]. Journal of Language and Social Psychology, 2010, 29(1):24-54.
- [TA99] Taylor S E . Health psychology (4th ed.)[J]. Health Psychology, 1999.
- [VA88] Vaux, A. (1988). Social Support: Theory, Research and Interventions. My, NY: Praeger.
- [WD06] 王登峰, 崔红, 胡军生, et al. 中国青少年人格量表 (QZPS—Q) 的编制[J]. 心理发展与教育, 2006(3):110-115.
- [WD11] Wang D. Y. The Development of College Students' Stressors Questionnaire[J]. 2011.
- [XUE14] Xue Y., Li Q., Jin L., et al. Detecting Adolescent Psychological Pressures from Micro-Blog [C]. HIS, 83-94. 2014.
- [XUE16] Xue Y , Li Q , Zhao L , et al. Analysis of Teens' Chronic Stress on Micro-blog[C]. International Conference on Web Information Systems Engineering. Springer International Publishing, 121-136. 2016.
- [ZH16] Zhao L., Li Q., Xue Y., et al. A systematic exploration of the micro-blog feature space for teens stress detection [J]. Health Information Science and Systems, 2016, 4(1):3.



2. 项目的研究内容、研究目标，以及拟解决的关键科学问题（此部分为重点阐述内容）；

2.1 研究内容

本项目旨在通过分析青少年微博大数据，探索青少年心理应激行为的群体性规律，并最终达到预测青少年未来时刻心理压力的目的。面临三项主要研究内容：1) 如何发现具有心理学意义的相似压力群体？2) 如何构建群体应激过程的关键心理变量？3) 融合群体应激状态，如何预测个体在未来时刻的心理压力？

研究内容一：首先，本项目拟在试点学校开展群体施测，构建“微博-实证”数据集。在此基础上融合心理学应激过程理论，发现具有相似应激模式的群体

如何定义心理应激模式的相似性？基于 Shelley 提出的应激过程理论(图 1)，个体的应激过程可分为五个环节：压力源事件的刺激，对压力源事件的评估和解释，个体的应对反应和策略，应对以及应对的结果[TA99]。其中，“个性因素”（如人格、性别、年龄等）和“应对方式”是个体调节情绪的关键中介变量[AF15]，本项目将以此为核心构建青少年心理应激模式的相似性模型。

本项目依托北师大心理学部青少年健康成长计划¹的试点学校，拟在北京市三帆中学、北师大第二附属中学开展群体施测与实证分析，构建“微博-实证”数据集。①通过量表自测的方式（中国青少年人格量表[WD06]、青少年生活事件量表[LI97]，压力知觉量表 PSS-14），构建青少年个性因素数据集；②经过用户授权和匿名处理，构建青少年微博数据集；③基于学校日程规划，构建学校生活相关的压力源事件数据集。

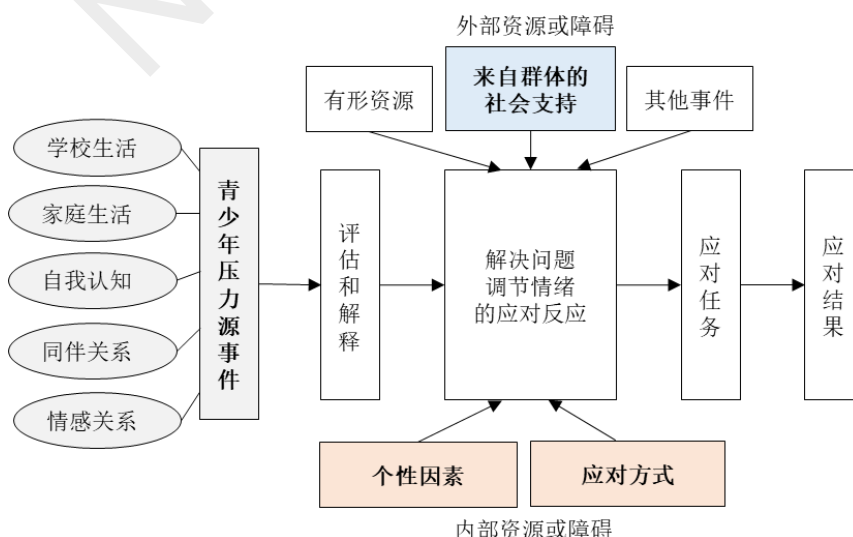


图 1 Shelley 提出的应激过程理论模型

¹ <http://psy-abc.bnu.edu.cn/>



研究内容二：其次，基于相似应激群体，量化与个体压力变化密切相关的群体心理变量（群体社会支持的缓冲因子，群体心理弹性因子，群体压力源影响因子），对于预测心理压力的后续发展十分重要。

群体社会支持的缓冲因子：研究表明，群体内的社会支持行为可以减少压力源对个体带来的伤害[VA88]。本项目拟对微博中的“情绪焦点型缓冲”和“问题焦点型缓冲”两种群体内缓冲因子进行分析（图2）。其中情绪焦点型是指针对压力源的负性应激反应而采取相应方法，以缓解、减轻这种反应；问题焦点型是指通过排除应激源来避免和克服这种反应。

群体心理弹性因子：心理弹性是在压力源事件中使个体得以积极应对的品质。基于 Kumpfer 提出的心理弹性理论[KU99]，青少年所在的群体作为外部保护性因素，有助于青少年发展起心理弹性方面的特质，减轻压力源事件对青少年的冲击，从而促进个体的健康发展。

群体压力源影响因子：压力源事件的影响因子可体现在微博的语义信息、压力曲线波动、行为规律等多方面，体现了当前群体对不同压力源事件的敏感度。

在实际情况中，上述线索往往不能同时出现，因此如何为上述线索分配权重，做出取舍，是量化群体压力源事件影响的关键问题。

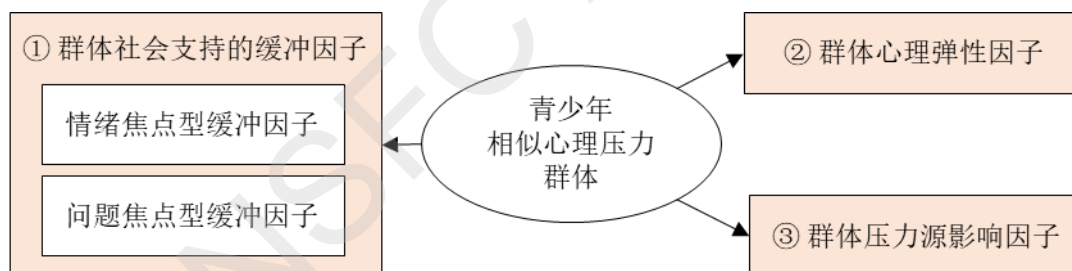


图2 青少年相似压力群体应激过程的关键心理变量

研究内容三：融合上述群体应激过程中的关键心理变量，预测青少年个体的未来时刻压力，主要解决两个难题：1）不活跃的青少年个体的数据稀疏问题，以及2）如何融合压力源事件对未来压力发展的驱动效果。

一方面，针对不活跃的微博用户数据存在稀疏这一问题，基于研究二的基础，考虑融合所在群体的心理应激变量，推断群体动态心理压力迁移路径，对稀疏的青少年个体压力序列进行补充；另一方面，由于外界对于青少年心理压力的发展甚至极端情况的发生往往是后知后觉的，而压力源事件的存在，是未来心理压力变化的驱动根源。因此，基于研究二的基础，如何将群体应激变量与青少年个体的历史心理压力相融合，并基于多重压力源事件的发生预测个体心理压力的动态发展，将是一个全新的挑战。



2.2 研究目标

通过对上述研究内容的探讨，本项目拟达到如下研究目标：

- 1) 本项目旨在融合青少年应激行为的群体性规律，预测其未来时刻的心理压力。围绕群体心理应激过程，勾勒出其中蕴含的一系列关键科学问题（诸如如何基于微博刻画应激理论体系中的关键中介变量，进而发现具有相似应激模式的群体？如何结合微博的语义表现、时序变化、行为特点等方面，量化与个体压力变化密切相关的群体心理变量？如何基于群体心理压力动态，融合压力源事件的影响，达到预测不活跃的青少年未来时刻心理压力的目的？），并提交可以付诸实施的解决方案与系统，破解传统心理应激研究的局限性问题的。
- 2) 基于心理学理论体系，挖掘个体显著的心理压力“应对方式”，并融合个体的“个性因素”，构建相似群体划分模型。
- 3) 刻画群体应激的关键心理变量（“群体社会支持的缓冲因子”、“群体心理弹性因子”、“群体压力源影响因子”）。在相似压力群体中结合微博的语义描述、压力时序变化、行为特点等表征，给出上述心理变量的量化和实证策略。
- 4) 构建基于神经网络的群体心理压力预测模型。基于群体应激的特征指标，构建群体内的心理压力动态迁移途径。
- 5) 设计并实现一个青少年“群体心理应激过程分析、预测平台”，实时可视化地展示：①相似青少年心理压力群体的聚簇、动态应激变化过程，②群体压力源事件与应激过程分析，③选定节点的心理压力预测结果，为中学提供群体层面的心理压力辅助分析预测工具，验证项目所提出的理论与方法。

2.3 拟解决的关键科学问题

围绕 2.1 节所拟定的三项研究内容，本项目将具体解决如下三个科学问题。

问题一：如何定义青少年心理应激模式的相似性，并构建相似压力群体？

节点属性的相似，群体结构的紧密互联，是构成具有心理学意义群体的必要因素。首先，由于个体因素（人格特质、年龄、性别等）和应对方式的差异，青少年个体心理压力水平在各类型压力源事件的牵引下不断变化，呈现出个性化的表现；其次，青少年在微博平台的互动行为，既体现了压力源事件的群体特性（转发行为），又体现了来自群体的社会支持（评论-回复/点赞）。

这里，如何统一考虑青少年应激属性的相似性与互动行为的关联程度，构建具有相似应激模式的群体？如何获取可靠的用户数据？如何解决群体特征缺失的问题？如何合理设置与动态学习各因素的权重？等均为本项目面临的挑战。

问题二：如何量化与个体压力变化密切相关的群体级别心理变量？

为了探讨群体级别心理变量（群体社会支持的缓冲因子，群体心理弹性因子，群体压力源影响因子）对个体心理压力变化的影响，本项目拟从微博的内容信息（包括文本、表情符、图片、分享音乐等）、时序变化（压力水平曲线波动、趋



势、极值点等)、行为特点(发布原创/有压力微博频率、好友互动、分享、点赞等)多个维度展开研究,在群体压力源事件影响区间内对于上述因子进行量化分析,构建各因子影响下的个体压力变化概率模型。

问题三:如何融合群体应激过程的关键心理变量,预测个体的心理压力?

本项目考虑根据群体的心理弹性因子,推断个体压力持续的概率;融合来自群体社会支持的缓冲因子,推断个体压力缓解的概率;围绕群体压力源事件的影响因子,推断个体压力持续增加的概率。如何将上述三种群体心理变量的影响概率与个体历史压力状态有效融合,是预测未来心理压力的核心问题。

3. 拟采取的研究方案及可行性分析(包括研究方法、技术路线、实验手段、关键技术等说明);

为了解决以上关键科学问题,本节讨论拟采取的研究方案及其可行性,包括1)群体施测与数据收集的设计方案,2)青少年相似压力群体的划分策略,3)群体心理变量的微博特征刻画与概率模型,以及4)融合群体应激过程的神经网络预测方案及平台架构。

3.1 群体施测与数据收集的设计方案

本项目拟在北京市三帆中学、北师大第二附属中学开展用户调研,针对青少年个性因素、学校日程事件和青少年微博数据,采用在线群体施测、学校官方网站数据收集、微博数据爬取等方法构建综合数据集。

(1) 在线群体施测

本项目拟基于在线调查平台 Qualtrics (<https://login.qualtrics.com/>)编写、发放、收集量表,并以班级为单位,邀请用户填写调查问卷。每次问卷结束之后,对有效问卷进行评定,并发放被试费(1元/分钟)。群体在线施测拟进行5周,流程设计如表1所示。

- 青少年人格量表[WD06]是基于大五人格量表编制的适合中国青少年应用的量表,符合中国人格结构模型,由青少年人格的7个维度(外向性维度、才干维度、善良维度、人际关系维度、处事态度维度、情绪性维度和行事风格维度)、19个二级因素及三个综合性维度构成,共111个项目。

- 青少年生活事件量表[LI97]包含27件青少年常见的生活压力源事件,本项目评定期限设置为最近1个月,统计指标为各事件发生频度和刺激量,设定为无影响、轻度、中度、重度和极重,并累积各事件刺激量作为总的刺激量。

- 压力知觉量表[COH83]目标是检测生活中普遍存在的压力,共包含14个项目,每个项目得分为1-5共五个级别,根据累积得分区间对被试的整体压力情况



进行判断，其中 1-13 为低压力，14-26 为中等压力，27-40 为较高压力。

表 1 本项目拟开展的用户调研流程

用户调研项目	拟开展时间段	目标
被试基本情况调研	2020/3/2 - 2020/3/8(1 次)	年级、年龄等信息（匿名）
中国青少年人格量表	2020/3/2 - 2020/3/8(1 次)	人格的 7 个人维度
青少年生活事件量表	2020/3/30- 2020/4/5(1 次)	27 种青少年生活事件
压力知觉量表 PSS-14	2020/3/2 - 2020/4/5(5 次)	心理压力级别
学校日程事件	2017/8/1 至 2020/7/1	在校期间的学校事件
微博	2017/8/1 至 2020/7/1	过去三年的微博数据

(2) 学校日程事件

为了涵盖所有调研用户在校期间的学校事件，拟收集 2017/8/1 至 2020/7/1 期间，学校网站发布的每周日程规划表²，并整理事件属性（压力源事件/积极事件）、事件类别、相关角色（家长、老师、同学）、事件起止日期、参与事件年级等。

(3) 微博

基于群体施测中的微博 ID 信息，拟收集 2017/8/1 至 2020/7/1 期间的用户微博数据，包括个人简介（昵称、签名、头像图片），发布时间、微博内容（文本、图片、音乐分享）、互动信息（评论、回复、点赞、转发）等。对于调研期间发布的微博（2020/3/2-2020/4/5），邀请被试对每条微博进行相关事件的标注。

上述所有数据均需被试签署知情同意书，并在匿名情况下开展后续实验分析。

3.2 青少年相似压力群体的划分策略

接下来，围绕个体应激模式，对用户间的心理应激差异进行量化。采用基于属性子空间的目标群体划分策略，发现具有相似心理压力模式的群体。

(1) 个体心理应激模式概率模型

经统计发现，青少年在压力源事件发生时发布有压力微博更频繁。若将青少年个体在不重叠的时间区间内发布的有压力微博数看作独立随机的泊松变量，则时间区间 T 内发布 n 条有压力微博的概率为：

$$\Pr[N=n | \lambda] = \frac{e^{-\lambda T} (\lambda T)^n}{n!}$$

因此，若将个体在压力源区间与非压力源区间的发博行为看作两个独立齐次泊松过程，为其分别定义发布有压力微博的频率 λ_1 和 λ_0 ，则该场景下 $\lambda_1 > \lambda_0$ 。为了衡量以上推断的置信度，可利用个体历史发布微博的行为对 λ_1 和 λ_0 进行估计，并根据后验概率分布计算推断的置信度，作为个体心理压力模式的判定指标。

² <http://www.sanfan.cn/article-21-2035.aspx>
<http://www.bjxcsy.net/cms/19-cl-7-268.htm>



令压力事件区间时间跨度为 T_1 ，发布有压力微博条数为 N_1 ；相应地，定义非压力区间长度及有压力微博数量分别为 T_0 和 N_0 。在估计理论中， N_1 和 N_0 分别是估计发博率 λ_1 和 λ_0 的最小充分统计量。此处对 λ_1 和 λ_0 应用 Jeffreys 无信息先验，即对于 $\lambda_1, \lambda_0 > 0$ ，先验密度函数和相应的边界先验为

$$p(\lambda_1, \lambda_0) \propto \frac{1}{\sqrt{\lambda_1 \lambda_0}}, \quad p(\lambda_1) \propto \frac{1}{\sqrt{\lambda_1}}, \quad p(\lambda_0) \propto \frac{1}{\sqrt{\lambda_0}}$$

根据贝叶斯法则， λ_i ($i=0, 1$) 的后验概率分布为：

$$p(\lambda_i | N_i) \propto \Pr[N_i | \lambda_i] p(\lambda_i) = \left(\frac{1}{\sqrt{\lambda_1 \lambda_0}} \times \frac{e^{-\lambda_i T_i} (\lambda_i T_i)^{N_i}}{N_i!} \times \frac{1}{\sqrt{\lambda_i}} \right) \propto \lambda_i^{N_i - 0.5} e^{-T_i \lambda_i}$$

由此，当前青少年具有显著心理压力模式的概率为：

$$p(\lambda_1 > \lambda_0 | N_1, N_0) = \iint_{x > y} p(\lambda_1 = x | N_1) p(\lambda_0 = y | N_0) dx dy$$

- 对于用户 u_i ，当 $pattern = p_i(\lambda_1 > \lambda_0 | N_1, N_0)$ 高于一个阈值 $\tau \in [0, 1)$ ，则当前青少年具有显著的心理压力模式。
- 对于任意两位用户 u_i 和 u_j ，其心理压力模式差异函数为：

$$Sim(pattern_i, pattern_j) = \frac{|pattern_i - pattern_j|}{\max(pattern_i, pattern_j)}$$

(2) 基于心理应激模式的相似压力群体发现策略

本项目拟采用面向压力源事件的目标群体发现方法，仅挖掘共享压力源事件的群体。该方法可以避免划分整个网络，从而提高算法运行效率，且群体划分结果更具有心理学意义。

设 $S = (\mathcal{V}, \varepsilon, \mathcal{F}, \mathcal{P})$ 表示当前网络，其中， \mathcal{V} 代表所有青少年用户节点的集合， ε 为用户间所有连边的集合， \mathcal{F} 表示用户的所有属性集， \mathcal{P} 为社交网络平台，则目标群体定义为 $G = (V, E, F)$ ，其中 $V \subseteq \mathcal{V}$ ， $E \subseteq \varepsilon$ ， $F \subseteq \mathcal{F}$ ，且 V 在 F 上满足属性相似性条件 $\Phi(V, F)$ ， V, E 满足结构内聚性条件 $\Psi(E)$ 。

令 F_i 表示用户 u_i 的属性集合，包括个体心理压力模式 $pattern$ 和个体基础特征 $profile = \langle personality, age, gender \rangle$ （个体基础特征基于问卷实证与微博特征分析）。令 $Sim(.)$ 为相似度计算函数，基于属性子集设置属性相似度为 $\Phi(F_i, F_j) = \max(Sim(pattern_i, pattern_j), Sim(profile_i, profile_j))$ 。

令 $E_{i,j}$ 表示用户 u_i 和 u_j 间的连接特征，为了表示其交流密切程度，令 $E_{i,j} = \langle act_{i,j}, stress_act_{i,j} \rangle$ ，其中， $act_{i,j}$ 表示互动频率（转发、评论、点赞），为其设置弱关联参数 α ； $stress_act_{i,j}$ 表示压力相关的互动频率，即在有压力的微博

下进行转发、评论或点赞，为其设置强关联参数 β ，则有 $\Psi(E_{i,j}) = \alpha \times \text{Sim}(\text{act}_{i,j}) + \beta \times \text{Sim}(\text{stress_act}_{i,j})$ 。

接下来，基于属性相似度 $\Phi(F_i, F_j)$ 和结构相似度 $\Psi(E_{i,j})$ ，构建相似群体发现模型。这里考虑两种情况，情况一，两位属性相似度很高的用户（如 c 和 b）可能不存在微博上的互动，甚至互不相识，例如两位同时备战高考且具有突出外向人格的青少年，此时，结构内聚条件 $\Psi(E_{i,j})$ 不满足但属性相似条件 $\Phi(F_i, F_j)$ 很强；情况二，两位处于同一班级的好友，存在互动行为，共享压力源事件的概率增加，此时 $\Psi(E_{i,j})$ 很强但 $\Phi(F_i, F_j)$ 可能并不一致（用户 a 和 d）。因此，本项目将综合考虑属性相似度 $\Phi(F_i, F_j)$ 和结构相似度 $\Psi(E_{i,j})$ ，挖掘相似青少年群体。

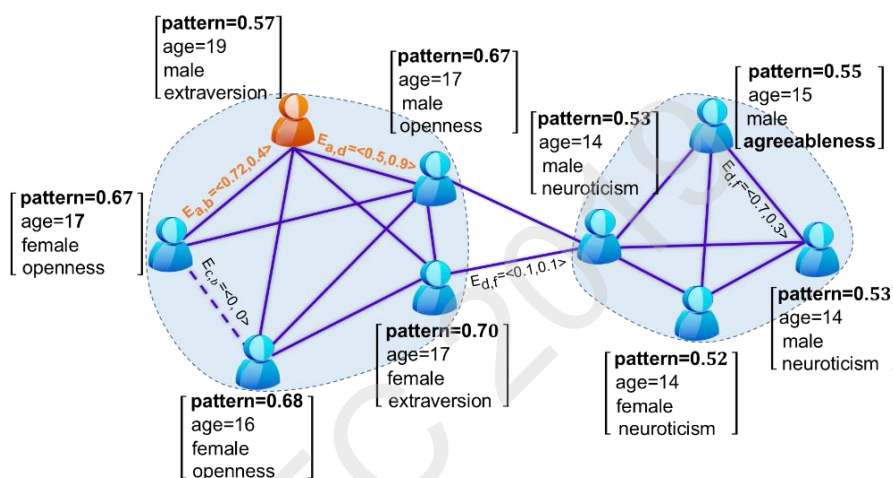


图 3 围绕压力属性相似度和互动结构相似度的相似群体建构

3.3 群体心理变量的微博特征刻画与概率模型

基于相似压力群体，接下来，从压力源事件与压力序列之间关联关系的角度出发，对群体内的关键心理变量与个体压力状态变化的关联关系进行分析。

(1) 群体心理弹性因子与群体缓冲因子的微博特征

首先，当前群体的心理弹性因子与压力时序变化和语义分布相关。

– 时序变化描述群体心理弹性对时间序列上压力变化模式的影响。可以形式化地将压力源事件 e 的影响区间 $I(e)$ 内的压力变化模式描述为一个矩阵 $P = [P^L \cup P^V \cup P^R] \in \mathbb{R}^{N \times 9}$ ，其中 $P^L = [l_1, l_2, \dots, l_n]^T \in \mathbb{R}^{N \times 3}$ ， $P^V = [v_1, v_2, \dots, v_n]^T \in \mathbb{R}^{N \times 3}$ ， $P^R = [r_1, r_2, \dots, r_n]^T \in \mathbb{R}^{N \times 3}$ ，分别表示区间 $I(e)$ 内的压力级别序列矩阵，压力偏差序列矩阵，以及压力趋势序列矩阵。

– 语义分布表示在压力源事件 e 影响下的区间 $I(e)$ 内，群体话题的集中程度。为衡量“青少年的压力是由当前压力源事件引起的”这一假设的置信度，从自然语言理解的视角出发， $I(e)$ 内微博的压力语义分布可表示为压力源事件主题词频率向量 d^{event} 和角色词频率向量 d^{role} 的集合，即 $d = [d^{\text{event}} \cup d^{\text{role}}]$ ， $d \in \mathbb{R}^{N \times 14}$ 。



其次,当前群体对个体 u 的缓冲因子拟从以下两方面进行考虑。

– 基于群体支持行为的角度,群体内与个体 u 的交流密度、交流类型、交流模式与 u 的压力缓解有关。其中交流密度指在时间序列上所在群体与 u 互动的频次分布;交流类型指具有缓解指向的交流类型所占比例;交流模式刻画个体 u 对群体支持的回应比例。

– 基于群体支持内容的角度:一是群体层面的语义分布情况,以刻画当前个体 u 所处的情绪环境;二是群体与个体 u 互动过程中的语义分布特点,以刻画来自群体的支持语义,以及个体在互动中的压力缓解模式。

此外,为解决不同压力源事件导致群体划分不一致的问题,对于用户 u_i 和 u_j ,在时间区间 $[t_s \rightarrow t_e]$ 内,压力源 e ,压力源序列 E ,对类型为 x 的群体压力源事件序列进行筛选,可表示为:

$$\begin{aligned}\Gamma_{i,j}(x) &= \Gamma(E_{t_s \rightarrow t_e}^x(u_i), E_{t_s \rightarrow t_e}^x(u_j)) \\ &= \{e_{t_k} : e_{t_k} \in E^x(u_i) \wedge e_{t_k} \in E^x(u_j) \\ &\quad \wedge \text{Duration}(e_{t_k}^i) \cap \text{Duration}(e_{t_k}^j) \neq \emptyset\}\end{aligned}$$

(2) 融合时序关联和语义关联的群体心理变量概率模型

为了克服时间序列分析中的异常值干扰问题,本项目拟将群体心理变量归纳为一个多变量的双样本假设检验问题。基于当前群体微博体现出的时序变化和语义分布,推断两个样本集合是否来自不同的基本分布,从而推导出当前群体心理变量是否具有显著特征。

这里,以群体心理弹性因子为例。设当前压力源事件 e 平均区间长度为 k ,随机选取 n 个受其影响的压力区间,形成集合 S^{event} ;为了表示时间序列上当前用户的一般压力水平,随机截取长度为 k 的子序列,经过 n 次随机截取,构成一个子序列的集合 $S^{\text{random}} = \{L_j^{\text{random}} | 1 \leq j \leq n\}$ 。

令 1_x 表示上述两个集合 S^{event} 和 S^{random} 中的某一个数据点,即 $1_x \in S^{\text{event}} \cup S^{\text{random}}$,则 1_x 由两部分组成,分别是 $P = [P^L \cup P^V \cup P^R] \in \mathbb{R}^{N \times 9}$ 和 $d = [d^{\text{event}} \cup d^{\text{role}}]$ 。假设 S^{event} 和 S^{random} 是分别基于分布 $F(\cdot)$ (概率密度为 $f(\cdot)$) 和分布 $G(\cdot)$ (概率密度为 $g(\cdot)$) 随机产生的,则该问题的双样本检验的假设表示为:

$$\begin{cases} H_0: F = G \\ H_1: F \neq G \end{cases}$$

上述假设检验可基于多种核方法[GR08]和基于距离的方法[SC86]求解,不同方法在当前问题上的性能表现需要在后续实验中进一步探讨。

3.4 融合群体应激过程的青少年心理压力预测方案及平台架构

本项目的心理压力预测模型拟考虑: 1) 融合群体心理变量的驱动效应(符

合应激过程理论），2）融合群体压力背景（解决个体数据稀疏问题），3）基于个体历史压力曲线（符合个体压力变化趋势）等主要原则。为此，项目拟采用基于神经网络的压力预测方案，该方案的优点在于它完全由数据驱动，无需规则或模板指导。但是，此方案离不开合理的特征选择以及异构特征的融合策略。

（1）基于神经网络的心理压力预测方案

本项目将从个体级别、群体级别两个层面，构造心理压力预测的特征空间。

– **个体级别心理压力特征**：基于[XUE14][LIN14]的工作基础，拟采用的青少年个体心理压力特征包括压力源事件特征（角色关键词分布（8 维）、压力源类型关键词分布（5 维）、情绪关键词频率（1 维）、压力源事件阶段（2 维）、个体压力趋势特征（压力趋势（3 维）、压力偏差（3 维）、发博行为特征（发布微博频率（3 维）、发布微博类型（3 维）、发布微博时间分布（24 维）、压力源事件影响特征（各类型压力源事件的时序关联（5 维）、各类型压力源事件的语义关联（5 维））等四个特征子空间。

– **相似群体压力及心理变量特征**：在时间单位 t 上的群体压力特征定义为群体平均压力（压力水平、均方压力值、压力趋势所占比例等）、各类型压力源事件出现比例及其影响因子、 t 时刻的群体心理弹性因子、 t 时刻的群体缓冲因子。

在此基础上，对于用户 u ，分别对其在下一单位时间 $t+1$ 以及未来时间段 $[t+1, t+k]$ 内的压力水平进行预测。拟构建一个嵌入双层前馈网络的非线性递归神经网络模型 (NARX-RNN)，将青少年心理压力预测这一问题看做一个动态的多步超前的非线性时间序列预测问题。文献表明 NARX 神经网络框架适合解决事件影响下的时间序列预测任务，其在处理输入变量及输出变量之间隐藏关系的复杂问题时被证明优于传统的线性预测方法[LIN96]。

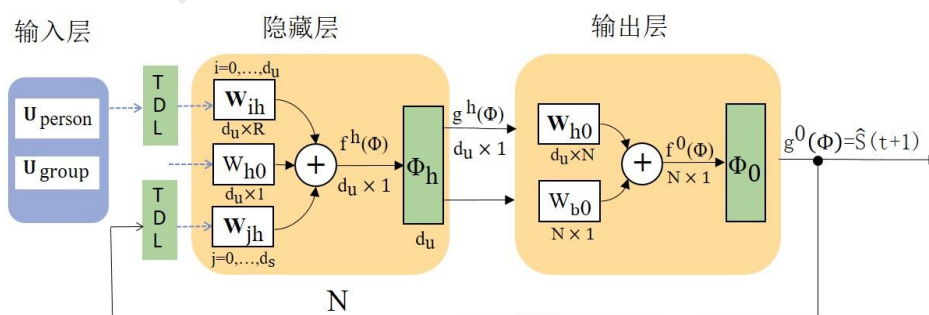


图 4 青少年心理压力预测问题的 NARX 回归神经网络拓扑结构

根据研究方案 3.2 中的相似群组发现策略，对于用户 u ，截止到 t 时刻，其历史上 d_y 个单位时间区间内压力水平序列（按单位时间区间聚合）表示为 $S^t = \{y(t), y(t-1), \dots, y(t-d_y+1)\}$ ，其个体压力特征序列为 $U^t = \{u(t), u(t-1), \dots, u(t-d_u+1)\}$ ，历史 d_c 个单位时间内的群体压力及群体心理变量特征序列为 $C^t = \{c(t), c(t-1), \dots, c(t-d_c+1)\}$ ，则当前压力预测问题建模为：



$$y(t+1)=\Psi(y(t), \cdots, y(t-d_y+1), \\ u(t), \cdots, u(t-d_u+1), c(t), \cdots, c(t-d_c+1))$$

该模型在各层神经元之间提供加权反馈连接，并在训练过程中为整个网络基于时间分配权重。为了提高模型的学习能力和更快的收敛能力，拟嵌入梯度下降方法（gradient-descending）。模型第一层网络的激活函数设置为 sigmoid 函数 Φ ，第二层网络的激活函数设置为线性函数（图 4）。

为了克服模型训练过程中的梯度消失问题，提高模型的泛化性能，考虑到在基于梯度的训练过程中对函数 Φ 的估计策略，拟在当前前馈网络中采用嵌入记忆的策略，并将其表示为图 4 中的时间延迟线，通过额外的记忆时间变化模式，可以动态地降低模型的灵敏度[DI08]。

（2）群体心理应激过程分析预测平台设计方案

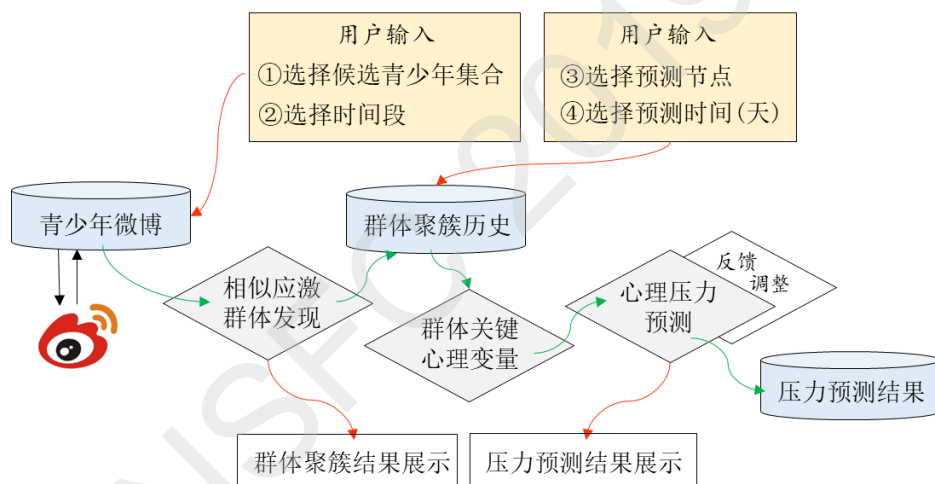


图 5 群体心理应激过程分析预测平台架构图

本项目设想一个旨在发现相似青少年心理压力群体，并对群体心理应激的动态过程进行分析，在此基础上提供心理压力预测功能的平台架构(如图 5 所示)。该平台包含相似应激群体发现、群体关键心理变量分析、未来心理压力预测等核心部件。其中，1) **相似应激群体发现模块**负责接收用户的输入（①候选用户集合、②目标时间段），从青少年微博数据库中提取相应微博、个体因素数据，经过模型分析，向用户展示目标时间段内的可视化群体聚簇结果，并将分析结果进行存储。2) **群体关键心理变量模块**负责根据群体发现模块的结果，对各群体内的系列关键心理变量进行建模运算，并将量化计算结果输出给心理压力预测模块。该模块作为平台的中间模块，不与用户直接交互。3) **心理压力预测模块**负责接受用户的输入（③选择预测节点，④选择预测时间），在用户根据相似群体划分结果，选择其中的单一/部分节点作为目标节点之后，根据用户需求对给定



未来时间段内目标节点的心理压力做出预测，进行可视化展示，并将预测结果进行存储。该模块根据青少年后续发布微博情况，对前期预测结果进行反馈调整。

3.5 技术路线

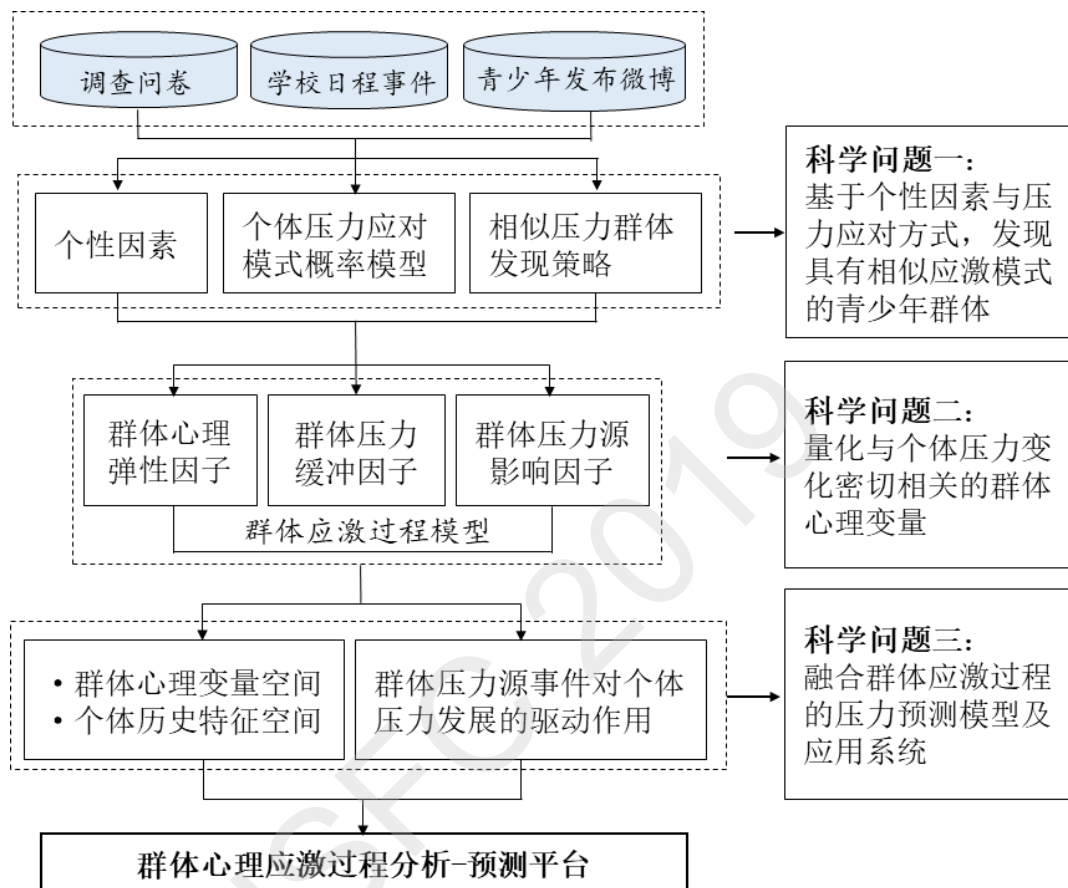


图6 基于大数据的青少年群体心理压力动态模型与预警系统建构

4. 本项目的特色与创新之处；

本项目的特色与创新主要体现在如下三个方面：

- 1) 首次提出将心理学应激过程理论应用于大数据分析的理论-应用体系，回应了传统心理学研究难以触及的大规模、及时性、动态过程、可预测的学术难题。
- 2) 深度交叉融合心理学领域专家知识、问卷调查、实验研究与社会媒体大数据分析方法，提出学科交叉共融的解决方案。根据青少年心理压力的复杂性，采用多质多法、交叉验证的研究策略，为大数据分析的数据可靠性与结果可解释性提供实证基础。
- 3) 构建一个“群体心理应激过程分析、预测平台”。通过在群体层面对青少年应激过程展开结构化分析、预测，为学校、社会提供青少年心理压力辅助预测、预警的有效工具，具有重要的应用价值。



5. 年度研究计划及预期研究结果(包括拟组织的重要学术交流活动、国际合作与交流计划等)。

5.1 年度研究计划

时间段		工作重点
2020 年	1-6 月	➤ 在试点中学开展群体施测, 分批次收集问卷与标注数据; ➤ 在试点中学网站收集学校日程事件列表;
	7-12 月	➤ 设计爬虫, 爬取目标青少年集合的微博数据; ➤ 完成三个数据集的整合、清洗与数据库的搭建;
2021 年	1-6 月	➤ 定义、设计、实现、初步验证青少年相似心理应激模式; ➤ 完成青少年相似压力群体模型建构与实验验证;
	7-12 月	➤ 抽取、度量三个群体关键心理变量; ➤ 设计、构建、优化群体心理变量影响的关联模型;
2022 年	1-6 月	➤ 构建融合群体心理变量的压力预测特征空间; ➤ 训练基于神经网络的青少年群体应激动态预测模型;
	7-12 月	➤ 构建群体心理应激过程分析、预警平台 ➤ 对平台进行测试、反馈、调整

5.2 预期研究成果

本项目拟提交的研究成果包括:

- 1) 在国内外高水平学术会议与期刊上发表关于青少年群体应激动态预测相关问题的原创性论文 3 篇, 其中 SCI 期刊论文至少 2 篇;
- 2) 完成群体应激过程分析、预测的平台系统, 提供群体聚簇、群体压力源事件与应激过程分析、个体心理压力预测的实时分析、可视化展示的功能。
- 3) 提交一份涵盖青少年群体压力源事件、个体因素、微博语料的用户标注大规模数据集, 用于青少年社交网络相关任务领域的研究工作。

(二) 研究基础与工作条件

1. 研究基础(与本项目相关研究工作积累和已取得的研究工作成绩);

申请人在清华大学计算机系攻读博士学位期间, 2013 年~2017 年作为骨干深度参与了国家自然科学基金面上项目《感知与疏导青少年心理压力的微博平台》, 通过开放的微博平台, 开展了实时感知、预测与疏导青少年心理压力的系列研究工作, 包括 1) 感知与疏导青少年微博心理压力的原型系统; 2) 分析青少年个



体心理压力区间与压力源事件；3) 预测青少年个体心理压力趋势。

1.1 基于微博分析青少年个体心理压力区间与压力源事件

提出一种基于微博序列感知青少年心理压力区间及压力源事件的检测方法。该方法利用微博数据蕴含的丰富语义信息，从自然语言处理的角度出发，对压力源事件的结构和细节进行深入挖掘，提取压力源事件的具体信息，并进一步对积极事件的缓解作用进行了探索，为进一步对青少年进行心理压力干预提供可能。

- **Qi Li, Yuanyuan Xue, Liang Zhao, Jia, Ling Feng. Analyzing and Identifying Teens Stressful Periods and Stressor Events from a Microblog. IEEE J. Biomedical and Health Informatics, 2017, 21(5):1434-1448. (JCR 一区期刊, 影响因子 3.48)**

1.2 感知与疏导青少年微博心理压力的方法

1) 提出基于微博的青少年用户心理压力检测的思路与方法。以微博平台数据为基础，以单条微博为单位，构建心理压力词典，以及一个包含发文内容、发文行为、交互内容、户交行为的四维微博特征空间，通过自然语言处理与机器学习，从单条微博中检测出用户压力类别与压力级别。

2) 实现疏导青少年心理压力的平台原型系统。基于微博上的心理压力检测结果，该系统提供了多种青少年心理压力的疏导途径，从文字，图像，视频等素材出发，结合日志模块、家长知会模块、轻松运动、虚拟聊天等方法提供对青少年心理压力疏导的接口。

- **Qi Li, Yuanyuan Xue, Jia and Ling Feng. Helping Teenagers Relieve Psychological Pressures: A Micro-blog Based System. Proceedings of the 17th International Conference on Extending Database Technology (EDBT), 660-663. 2014. (CCF 列表 B 类会议, EI 收录, 引用 9 次)**
- **Yuanyuan Xue, Qi Li, et al. Detecting Adolescent Psychological Pressures from Micro-Blog. Health Information Science, 2014. (Best student paper, 引用 21 次)**

1.3 基于微博预测青少年个体未来心理压力趋势

运用时序分析方法，基于单条微博的压力检测结果，通过对不同粒度（天或周）的压力时间序列，预测青少年个体下一时间区间（如下一天或下一周）的压力级别（平均值、压力最小值和压力最大值）、压力变化趋势（即升高、降低或不变）、压力变化幅度（即压力程度何时达到峰值、谷底），并结合事件对个体心理压力进行校正。



- **Qi Li, Liang Zhao, Yuanyuan Xue, Li Jin, Mostafa Alli and Ling Feng.** Correlating Stressor Events for Social Network Based Adolescent Stress Prediction. Proceedings of the 22nd International Conference Database Systems for Advanced Applications (DASFAA), 642-658. 2017 (CCF 列表 B 类会议, EI 收录)
- **Qi Li, Liang Zhao, Yuanyuan Xue, Li Jin and Ling Feng.** Exploring the Impact of Co-Experiencing Stressor Events for Teens Stress Forecasting. Proceedings of the 18th International Conference on Web Information Systems Engineering (WISE), 313-328. 2017. (CCF 列表 C 类会议, EI 收录)

申请人在青少年微博心理压力感知方面进行了一些探索,取得的研究成果为本项目的顺利实施奠定了一定基础。本项目将从全新的群体应激理论视角,交叉应用心理学领域专家知识、问卷调查、实验研究基础的方法,提出基于大数据的青少年群体心理压力动态理论体系与模型,并建构预警系统。

2. 工作条件 (包括已具备的实验条件, 尚缺少的实验条件和拟解决的途径, 包括利用国家实验室、国家重点实验室和部门重点实验室等研究基地的计划与落实情况);

申请人所在北京师范大学心理学部具有开展“心理学—社交网络大数据”交叉研究的坚实理论支撑和学科优势。北师大心理学部是中国心理学的发源地,在教育部历次学科排名中连续排名全国第一,是中国最重要的心理学教学与科研基地,并获批“国家级实验教学示范中心”和“国家级虚拟仿真实验教学中心”。

目前所在单位已初步完成心理学大数据实验室的硬件建设,正全力开展基于社交网络大数据的心理学探索。实验室配置有惠普 Z8 G4 高性能服务器一台,配有英特尔至强 (Intel Xeon 8160) 双颗 48 核心处理器,英伟达 (NVIDIA Quadro P5000) 16G 独立显卡,以及 64T 容量的 NAS 网络附加存储,为社交网络大数据的收集、存储,以及机器学习模型的训练速度准备了必需的硬件设施。

3. 正在承担的与本项目相关的科研项目情况 (申请人正在承担的与本项目相关的科研项目情况,包括国家自然科学基金的项目和国家其他科技计划项目,要注明项目的名称和编号、经费来源、起止年月、与本项目的关系及负责的内容等);

无。



4. 完成国家自然科学基金项目情况（对申请人负责的前一个已结题科学基金项目（项目名称及批准号）完成情况、后续研究进展及与本申请项目的关系加以详细说明。另附该已结题项目研究工作总结摘要（限 500 字）和相关成果的详细目录）。

无。

（三）其他需要说明的问题

1. 申请人同年申请不同类型的国家自然科学基金项目情况（列明同年申请的其他项目的项目类型、项目名称信息，并说明与本项目之间的区别与联系）。

无。

2. 具有高级专业技术职务（职称）的申请人是否存在同年申请或者参与申请国家自然科学基金项目的单位不一致的情况；如存在上述情况，列明所涉及人员的姓名，申请或参与申请的其他项目的项目类型、项目名称、单位名称、上述人员在该项目中是申请人还是参与者，并说明单位不一致原因。

无。

3. 具有高级专业技术职务（职称）的申请人是否存在与正在承担的国家自然科学基金项目的单位不一致的情况；如存在上述情况，列明所涉及人员的姓名，正在承担项目的批准号、项目类型、项目名称、单位名称、起止年月，并说明单位不一致原因。

无。

4. 其他。

无。



李琦 简历

北京师范大学，心理学部，工程师

教育经历（从大学本科开始，按时间倒序排序；请列出攻读研究生学位阶段导师姓名）：

(1) 2013.9 - 2018.7, 清华大学, 计算机科学与技术, 博士, 导师：冯铃

(2) 2009.9 - 2013.7, 大连理工大学, 软件工程, 学士, 导师：无

科研与学术工作经历（按时间倒序排序；如为在站博士后研究人员或曾有博士后研究经历，请列出合作导师姓名）：

(1) 2018.9-至今, 北京师范大学, 心理学部, 工程师

曾使用其他证件信息（申请人应使用唯一身份证件申请项目，曾经使用其他身份证件作为申请人或主要参与者获得过项目资助的，应当在此列明）：

主持或参加科研项目（课题）情况（按时间倒序排序）：

1) 北京师范大学青年教师基金项目，基于社交网络大数据的用户心理压力分析，2018.10 - 2020.09, 10万元, 在研, 主持

2) 重点研发专项，面向高端制造领域的大数据管理系统，2016.07 - 2019.06, 8718万元, 在研, 参加

3) 国家自然科学基金重点项目，20151311404, 碎片化知识聚合方法研究，2016.01 - 2020.12, 85万元, 在研, 参加

4) 国家自然科学基金面上项目，20131351107, 感知与疏导青少年心理压力的微博平台，2014.01 - 2017.12, 78万元, 已结题, 参加

代表性研究成果和学术奖励情况

（请注意：①投稿阶段的论文不要列出；②对期刊论文：应按照论文发表时作者顺序列出全部作者姓名、论文题目、期刊名称、发表年代、卷（期）及起止页码（摘要论文请加说明）；③对会议论文：应按照论文发表时作者顺序列出全部作者姓名、论文题目、会议名称（或会议论文集名称及起止页码）、会议地址、会议时间；④应在论文作者姓名后注明第一/通讯作者情况：所有共同第一作者均加注上标“#”字样，通讯作者及共同通讯作者均加注上标“*”字样，唯一第一作者且非通讯作者无需加注；⑤所有代表性研究成果和学术奖励中本人姓名加粗显示。）

按照以下顺序列出：①代表性论著（包括论文与专著，合计5项以内）；②论著之外的代表性研究成果和学术奖励（合计10项以内）。

一、代表性论著



(1) Li Qi^(#); Xue Yuanyuan; Zhao Liang; Jia Jia; Feng Ling^(*), [Analyzing and Identifying Teens' Stressful Periods and Stressor Events From a Microblog](#)[✓], IE EE Journal of Biomedical and Health Informatics, 2017.09, 21(5): 1434~1448

(期刊论文)

(2) Qi Li^(#); Yuanyuan Xue; Jia Jia; Ling Feng^(*), Helping Teenagers Relieve Psychological Pressures: A Micro-blog Based System, Proceedings of the 17th International Conference on Extending Database Technology (EDBT), 2014.3.24-2014.3.28 (会议论文)

(3) Li Qi^(#); Zhao Liang; Xue Yuanyuan; Jin Li; Alli Mostafa; Feng Ling^(*), [Correlating Stressor Events for Social Network Based Adolescent Stress Prediction](#)[✓], 22nd International Conference on Database Systems for Advanced Applications (DASFAA), 2017.03.27-2017.03.30 (会议论文)

(4) Qi Li^(#); Liang Zhao; Yuanyuan Xue; Li Jin; Ling Feng^(*), Exploring the Impact of Co-Experiencing Stressor Events for Teens Stress Forecasting, Web Information Systems Engineering (WISE) 2017 - 18th International Conference, 深圳, 2017.10.7-2017.10.11 (会议论文)

(5) Yuanyuan Xue^(#); Qi Li; Ling Feng^(*); David A. Clifton; David A. Clifton, Detecting Adolescent Psychological Pressures from Micro-Blog, Health Information Science - Third International Conference (HIS 2014), 2014.4.22-2014.4.23 (会议论文)

二、论著之外的代表性研究成果和学术奖励

(1) 冯铃; 李琦; 薛媛媛, 一种基于微博的青少年心理压力区间及其压力源事件检测方法, 2016.12, 中国, 201610576199.1 (专利)

(2) 冯铃; 李琦; 薛媛媛, 青少年心理压力感知与疏导平台软件, 2017SR463762, 原始取得, 全部权利, 2016 (软件著作权)

(3) 冯铃; 李琦; 薛媛媛, 基于微博的青少年压力区间及压力源事件检测软件, 2018SR156409, 原始取得, 全部权利, 2017.05 (软件著作权)

(4) 冯铃; 李琦; 薛媛媛, 青少年心理压力疏导虚拟聊天软件, 2017SR463650, 原始取得, 全部权利, 2016 (软件著作权)

(5) 冯铃; 薛媛媛; 李琦, 基于微博和模糊认知图的自杀风险检测方法及系统, 2017.9.20, 中国, 201710237046.9 (专利)



附件信息

序号	附件名称	备注	附件类型
1	1_Analyzing and Identifying Teens' Stressful Perio	IEEE Journal of Biomedical and Health Informatics. SCI一区期刊. 2017	代表性论著
2	2_Helping Teenagers Relieve Psychological Pressure	EDBT. CCF-B类国际会议. 2014	代表性论著
3	3_Correlating Stressor Events for Social Network B	DASFAA. CCF-B类国际会议. 2017	代表性论著
4	4_Exploring the Impact of Co-Experiencing Stressor	WISE. CCF-C类国际会议. 2017	代表性论著
5	5_Detecting Adolescent Psychological Pressures fro	HIS. 2014. Best student paper	代表性论著
6	基于微博的青少年压力区间及压力源事件检测软件证书.pdf	软件著作权1	其他
7	青少年心理压力感知与疏导平台软件证书.pdf	软件著作权2	其他
8	青少年心理压力疏导虚拟聊天软件证书.pdf	软件著作权3	其他



项目名称： 基于大数据的青少年群体心理压力动态模型与预警系统建构

资助类型： 青年科学基金项目

申请代码： F020516. 计算医疗与健康技术

国家自然科学基金项目申请人和参与者公正性承诺书

本人**在此郑重承诺**：严格遵守中共中央办公厅、国务院办公厅《关于进一步加强科研诚信建设的若干意见》规定，所申报材料和相关内容真实有效，不存在违背科研诚信要求的行为；在国家自然科学基金项目申请、评审和执行全过程中，恪守职业规范和科学道德，遵守评审规则和工作纪律，杜绝以下行为：

- （一）抄袭、剽窃他人科研成果或者伪造、篡改研究数据、研究结论；
- （二）购买、代写、代投论文，虚构同行评议专家及评议意见；
- （三）违反论文署名规范，擅自标注或虚假标注获得科技计划等资助；
- （四）购买、代写申请书；弄虚作假，骗取科技计划项目、科研经费以及奖励、荣誉等；
- （五）在项目申请书中以高指标通过评审，在项目计划书中故意篡改降低相应指标；
- （六）以任何形式打听尚未公布的评审专家名单及其他评审过程中的保密信息；

（七）本人或委托他人通过各种方式及各种途径联系有关专家进行请托、游说，违规到评审会议驻地游说评审专家和工作人员、询问评审或尚未正式向社会公布的信息等干扰评审或可能影响评审公正性的活动；

（八）向评审工作人员、评审专家等提供任何形式的礼品、礼金、有价证券、支付凭证、商业预付卡、电子红包，或提供宴请、旅游、娱乐健身等任何可能影响评审公正性的活动；

（九）其他违反财经纪律和相关管理规定的行为。

如违背上述承诺，本人愿接受国家自然科学基金委员会和相关部门做出的各项处理决定，包括但不限于撤销科学基金资助项目，追回项目资助经费，向社会通报违规情况，取消一定期限国家自然科学基金项目申请资格，记入科研诚信严重失信行为数据库以及接受相应的党纪政纪处理等。

编号	姓名 / 工作单位名称（应与加盖公章一致） / 证件号码 / 每年工作时间（月）	签字
1	李琦 / 北京师范大学 / 1*****2 / 10	
2		
3		
4		
5		
6		
7		
8		
9		
10		



项目名称： 基于大数据的青少年群体心理压力动态模型与预警系统建构
资助类型： 青年科学基金项目
申请代码： F020516. 计算医疗与健康技术

国家自然科学基金项目申请单位公正性承诺书

本单位依据国家自然科学基金项目指南的要求，严格履行法人负责制，**在此郑重承诺：**本单位已就所申请材料内容的真实性和完整性进行审核，不存在违背中共中央办公厅、国务院办公厅《关于进一步加强科研诚信建设的若干意见》规定和其他科研诚信要求的行为，申请材料符合《中华人民共和国保守国家秘密法》和《科学技术保密规定》等相关法律法规，在项目申请和评审活动全过程中，遵守有关评审规则和工作纪律，杜绝以下行为：

（一）采取贿赂或变相贿赂、造假、剽窃、故意重复申报等不正当手段获取国家自然科学基金项目申请资格；

（二）以任何形式探听未公开的项目评审信息、评审专家信息及其他评审过程中的保密信息，干扰评审专家的评审工作；

（三）组织或协助项目团队向评审工作人员、评审专家等提供任何形式的礼品、礼金、有价证券、支付凭证、商业预付卡、电子红包等；宴请评审组织者、评审专家，或向评审组织者、评审专家提供旅游、娱乐健身等任何可能影响科学基金评审公正性的活动；

（四）包庇、纵容项目团队虚假申报项目，甚至骗取国家自然科学基金项目；

（五）包庇、纵容项目团队，甚至帮助项目团队采取“打招呼”等方式，影响科学基金项目评审的公正性；

（六）在申请书中以高指标通过评审，在计划书中故意篡改降低相应指标；

（七）其他违反财经纪律和相关管理规定的行为。

如违背上述承诺，本单位愿接受国家自然科学基金委员会和相关部门做出的各项处理决定，包括但不限于停拨或核减经费，追回项目经费，取消一定期限国家自然科学基金项目申请资格，记入科研诚信严重失信行为数据库以及主要责任人接受相应党纪政纪处理等。

依托单位公章：

日期： 年 月 日

合作研究单位公章：

日期： 年 月 日

合作研究单位公章：

日期： 年 月 日