文章编号：1003-0077(2018)09-0093-10

利用准私密社交网络文本数据检测抑郁用户的可行性分析

刘德喜[1,2]，邱家洪[1,2]，万常选[1,2]，刘喜平[1,2]，钟敏娟[1,2]，郭海峰[3]，邓 松[4]

（1. 江西财经大学 信息管理学院，江西 南昌 330013；
2. 江西财经大学 江西省高校数据与知识工程重点实验室，江西 南昌 330013；
3. 江西财经大学 学生工作处，江西 南昌 330013；
4. 江西财经大学 软件与通信工程学院，江西 南昌 330013）

摘 要：社交媒体的发展为抑郁用户的检测提供了一条新的途径。已有的相关研究通常是利用用户在 Twitter、微博等社交网络平台上的用户行为数据或公开发表的文本内容，较少有利用微信朋友圈、QQ 空间这种相对比较私密的社交网络数据。直观地，这类准私密社交网络数据更能反映用户的心理健康状况。该文主要讨论利用准私密社交网络文本数据检测抑郁用户的可行性，包括训练样本的选择、特征量化方法、检测模型选择和不同文本特征下的模型分类效果等。实验表明，采用平衡高低分组的方法选择样本比非平衡高低分组样本和离散化的高低分组样本训练的分类器要好；利用 Z-score 标准化的特征量化方法比直接使用频次或归一化频率要好；随机梯度下降模型 SGD 较支持向量机 SVM 等其他用于对比的分类模型要好。实验还发现，相对于词袋、词向量等文本特征，主题特征有较好的效果，可以使社交网络用户抑郁检测模型的 F 值达到 0.753，而对抑郁用户的检测精度达到 0.813。

关键词：准私密社交网络文本；抑郁用户检测；可行性分析

中图分类号：TP391 **文献标识码**：A

Feasibility of Detecting Depressive Users Using Quasi-private Social Text

LIU Dexi[1,2], QIU Jiahong[1,2], WAN Changxuan[1,2], LIU Xiping[1,2],
ZHONG Minjuan[1,2], GUO Haifeng[3], DENG Song[4]

(1. School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China;
2. Jiangxi Key Laboratory of Data and Knowledge Engineering,
Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China;
3. Students' Affairs Division, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China;
4. School of Software and Communication Engineering, Jiangxi University of Finance and Economics,
Nanchang, Jiangxi 330013, China)

Abstract: The development of social network has provided an innovative perspective for detecting depressive users. Few works have been done on private data which come from the relatively private social network such as WeChat friends circle or QQ Zone to detect depressive users. This paper discusses the feasibility of detecting depressive users on quasi-private social network data, including training samples, feature extraction, detection model, etc. The experimental results show that, to train an effective model and overcome the challenge of unbalance samples, we should firstly select almost the same amount of positive and negative samples with the highest and the lowest scores of self-report tests, which corresponding to the most depressive users and the most normal users. Secondly, the features should be quantified by Z-score standardized frequency, which is more powerful than the other two quantifying methods such as frequency or normalized frequency. Thirdly, the SGD classifier performs better than the other classifiers such as SVM. The results also show that, compared to other features such as bag-of-words or word-to-vector, topical features performs better with 0.813 detection precision and 0.753 F-measure.

Key words: quasi-private social text; depressive users detecting; feasibility analysis

收稿日期：2017-09-05 **定稿日期**：2017-10-29

基金项目：国家自然科学基金(61762042, 61363039, 61562032, 61662027, 61462037)；江西省科技落地计划(KJLD14035)；江西省自然科学基金(20171BAB202021)

0 引言

世界健康组织(WHO)在2012年的研究表明,全世界约有3.5亿人患有抑郁症,严重的抑郁可以导致自杀[1]。由于缺少心理健康知识以及心理疾病显著区别于身体疾病的无疼痛感,导致许多人身患抑郁而不知或是由于抑郁羞耻感而不敢主动寻找专业人士帮助。心理学上通过抑郁自评量表检测的方法属于侵入型检测方法,在适时性和自评频率方面存在不足,导致不能及时检测出抑郁症患者,延误治疗。随着互联网和信息技术的发展,Twitter、微博、微信等社交媒体已经成为人们互相交流必不可少的工具,形成与物理空间相对等的网络社区,用户网络行为信息也记录在社交网络中,为检测用户抑郁症等心理健康疾病提供了一种新的途径。

目前,已有较多利用用户在社交网络上的行为和发布的文本进行用户心理健康分析的相关研究工作,所选取的社交网络平台大都是Twitter、微博、人人网等公开社交网络。公开社交网络支持单向关注的特点使得用户隐私权无法得到有效的保障。因此,用户在公开社交网络上更倾向于表达话题性观点,大部分用户仅仅是在热点话题上表现活跃。

与公开社交网络相比,QQ、微信等社交网络因为朋友圈的划分和有限的用户访问权限设置等,更能保障用户隐私不被泄露,私密性更强,更加受到用户的青睐。本文称这种信息只在好友圈可见的社交网络为准私密社交网络,准私密社交网络越来越成为人们日常生活不可分割的一部分。直观上,相比公开社交网络数据,准私密社交网络数据能够更有效地反映出用户的生活状态与心理状态,更能反映用户的抑郁等心理健康问题。

已有的研究工作大部分是基于公开社交网络的,鲜有文献分析准私密社交网络数据是否可用于分析用户的抑郁倾向,以及如何利用这些数据分析用户的抑郁倾向。本文从训练样本选择、特征量化方法、分类模型的选择、文本内容特征四个角度考察利用准私密社交网络文本检测抑郁用户的可行性,并与基于公开社交网络数据进行抑郁检测的相关文献进行比较。论文的结构安排如下:首先介绍研究背景,然后介绍利用社交网络数据分析用户心理健康的研究现状,接下来介绍数据采集与预处理、候选特征抽取与量化、训练样本选择、相关性分析、检测模型选择,并通过实验考察样本选择、特征量化方法对抑郁用户检测模型的影响,分析不同的文本特征在检测模型上的表现,最后对全文进行总结。

1 相关研究

利用社交网络数据分析用户心理健康状态具有实时性、高效性、无侵入性等特点,对心理健康状况欠佳人员的及时检测、辅导和诊疗具有重要意义,得到心理学领域和计算机科学领域研究者的关注。已有的研究工作通常把利用社交网络数据分析用户心理健康状态视为一个分类问题,通过样本训练分类模型,将社交网络用户的自杀倾向、抑郁等心理健康问题分类为"有""无"两大类。下面主要从社交网络数据与抑郁的相关性分析、数据源选择、特征选择和量化、训练样本选择、分类模型五个方面对研究现状进行描述。

大量研究发现可以通过社交网络活动记录对用户的抑郁状态进行检测[1-19],严重的抑郁症患者在社交网络上的行为与正常人存在显著的差异[1]。Choudhury等[3]通过研究Amazon用户的语言风格和网络行为,发现抑郁用户社会活动少,消极情感更为严重,对人际关系和药物的使用更为担心,同时更注重宗教思想的表达。Park等[5]发现抑郁用户使用消极情感词和愤怒词明显较正常用户多,用户在社交网络上不仅表达抑郁情感也会发布一些非常隐私的信息。

数据源方面,大多数相关研究使用了当前比较流行的社交网络平台,如Twitter[2-5,13-15]、Facebook[9-11,19]、论坛[16-17]、新浪微博[1,6-7,18]、人人网[20]等。也有利用用户的其他上网痕迹,如网关记录的网页浏览、搜索行为等[21]。而Hiraga[22]使用了来自Yahoo Japan、Livedoor等多个blog平台的数据。

特征选择方面,被采用较多的特征主要包括语言特征、行为特征、属性特征、社交关系特征等。语言特征是指用户的社交网络用语表现出来的特征,主要有情感词、人称代词、表情符号的使用等[1,5-7,10,16-18]。行为特征主要有点赞数、转发数、原创帖子数等[8],研究者认为不同心理健康状态的用户网络行为表现不同。属性特征是指社交网络用户的属性,主要包括年龄、性别、职业等[2,8,11-12]。社交关系特征是社交网络中错综复杂的社交关系的表现,主要有好友个数、互动频数、亲密度等[7]。由于

LIWC 词典(Linguistic Inquiry and Word Count)[①]是从心理学的角度描绘用户的用词特点，因此经常被用作心理健康分析的语言特征[5,16-18,23]。除以上几类特征外，也有文献直接利用文本中的 n-gram、词性(POS)等信息[22]。

相对其他用户，社交网络上的抑郁用户数量非常少，因此采集的样本通常是极度不平衡的，大量的研究工作采用高底分组的方法构建平衡训练样本。文献[6]采用随机抽取的方式，而文献[24]则采用高低分组的方式，分别抽取了自杀风险最高的和最低的 80%的用户构成自杀用户数据集。为了在训练样本中反映抑郁用户和正常用户的真实分布，文献[21]采用非平衡采样的方式，其中 449 个抑郁用户、279 个正常用户。

分类模型的选择方面，线性回归[2,22]、多任务线性回归[18]、SVM[4,22]、朴素贝叶斯[21]、贝叶斯网络[6,22]、神经网络[18,21]、决策树[6,21]、规则决策表[6]等常用的分类模型大都被使用或比较过。

2　分析方法

利用准私密社交网络文本数据进行抑郁用户检测的可行性分析，主要包含六个阶段：数据采集与预处理，候选特征抽取与量化，训练样本选择，相关性分析，检测模型选择，检测效果评估与分析。本节仅介绍前五个阶段，最后一阶段在下一节介绍。

2.1　数据采集与预处理

2.1.1　数据采集

通过用户填写抑郁自评问卷得到用户抑郁状况，即标签；同时，收集用户的 QQ 和微信账号并获取数据使用授权，采集得到用户准私密社交网络数据。本研究邀请了江西财经大学 6 378 位大一新生于 2016 年 10 月参与研究，所有参与者完成抑郁测评问卷，同时签署数据保密协议，获取参与者 QQ 空间和微信朋友圈数据(问卷截止日期前一年内的数据)。为了保证数据质量，采取了一系列措施，包括：采用 CES-D[②] 和 BDI[③] 双量表形式设计问卷，舍弃两个量表分值相差过大的用户；去除问卷得分为零分或满分的特殊用户以及问卷完成时间少于 4min 的用户；去除无法采集到 QQ 空间及微信朋友圈数据的用户。

CES-D 和 BDI 是心理学上常用于测量抑郁症的抑郁量表，从多个维度综合考查了用户的抑郁状态，同时也是典型的 4 点李克特度量量表(每道题有四个选项，得分为 0～3，对应抑郁程度由无到严重)。CES-D 量表有 20 道题，得分区间在[0，60]，分值分布区间为："＜＝10 分"为无抑郁，"11～20 分"为可能有抑郁，"21～60 分"为肯定有抑郁[25]；BDI 量表有 21 道题，得分区间在[0，63]，"＜＝15 分"为无抑郁，"16～35 分"为轻度抑郁或中度抑郁，"36～63 分"为重度抑郁[26]。合并两个量表的分值分布区间得到问卷分值分布区间[0，123]，本文设置正常用户得分区间为[1，25]，轻度抑郁用户得分区间为[26，55]，重度抑郁用户得分区间为[56，123]。

经过以上筛选，获取了 1 522 个有 QQ 空间数据的有效用户，710 个有微信朋友圈数据的有效用户，这些用户心理健康状况分布如表 1 所示。本文获取的准私密社交网络数据与文献[21]有较大的不同，数据不平衡问题更严重。在 QQ 空间用户中，抑郁自评量表反映出正常用户占 60.5%，轻度抑郁用户占 36.7%，重度抑郁用户占 2.8%；而微信用户中，正常用户占 36.8%，轻度抑郁用户占 61.3%，重度抑郁用户占 1.9%。导致这种分布差异的可能原因有两个：一是不同抑郁状态的用户在 QQ 空间和微信朋友圈的使用上有差异；二是由于部分用户(特别是有重度抑郁倾向的用户)的 QQ 空间设置了密码无法抓取，导致样本分布的改变。

表 1　准私密社交网络上有抑郁倾向的用户分布

	正常用户	轻度抑郁用户	重度抑郁用户
QQ 空间用户	921(60.5%)	559(36.7%)	42(2.8%)
微信朋友圈用户	261(36.8%)	435(61.3%)	14(1.9%)

对微信朋友圈和 QQ 空间中用户发布的帖子数的统计显示，大部分用户发布帖子数量都在 50 条以下(截止填写自评量表前一年内)。

2.1.2　数据预处理

数据预处理主要包括去除和转换两个操作。去除内容包括：(1)转义字符，例如，以"\t"和"\n"的形式出现的空格和换行符；(2)偏僻字符，例如，"卐、※、ぷ"等；(3)英文文本，本研究只针对中文文本。

转换操作：(1)将表情符转换为＜emoticon＞：社交文本中存在大量的表情符，表情符是用户表达情感的最直接方式，本文将其视为一类特殊符号(由

① http://liwc.wpengine.com/.
② 流行学研究中心抑郁量表的缩写.
③ 贝克抑郁量表的缩写.

于本文并没有分析文本的情感类别，因此，不区分表情符的极性）；（2）将 url 链接转换为＜url＞：用户通过转载图片、视频、网页等形式表述情感或观点，而这些图片、视频、网页通常以 url 链接的形式出现在原始数据中（本文不考察 url 链接的对象类型及其内容）；（3）将@及其对象转换为@符号：@符号是社交网络平台中用户互动的常用方式，能在一定程度上表明用户之间的交流。

数据预处理还包括分词，本文选用的分词工具是 NLPIR 汉语分词系统①，它针对微博等数据有优化、有新词识别能力，比较适合微博、微信、QQ 空间上的文本。

2.2 候选特征抽取与量化

当前研究对特征的选取主要有两种方法：一是借助心理学家对抑郁用户社交文本、网络行为、用户属性的统计和分析，归纳出抑郁用户的特征[6,10]；二是通过统计用词或行为的频率，根据相关性分析，得出抑郁用户与正常用户在用词或行为上的不同[2]。本文使用了如下候选特征。

（1）行为特征。行为特征是用户在社交网络上所表现出的行为，包括用户发布帖子、用户之间的互动等，本文考察的网络行为特征主要有：转载帖子数、原创帖子数、点赞数、用户在凌晨 0 点到 6 点之间发布的帖子数、@符号数、帖子评论数等。

（2）语言特征。本文考察的语言特征主要来自 LIWC，包括表情符号、第一人称单复数等 71 个词类。LIWC 中的每个词类被视为一个特征，特征值为样本中包含该类词的帖子数量。

（3）文本内容特征：利用用户在准私密社交网络上发布的文本来检测其是否有抑郁倾向的问题，可以视为文本分类的问题，因此用于文本分类的特征可以被借鉴。本文在实验中考察了以下特征：

Bag of Words（BOW，或 1-gram）：以用户发布的文本中全部的词为特征，以词的 TFIDF 值为权重。

主题（Topics）：对数据集进行主题分析，以用户发布的文本的主题分布为特征。本文利用 Gensim 工具②中的 LDA 模型进行主题分析。

词向量（Word2vect）：将用户发布的文本中的词转换为词向量，并将文本中全部词的词向量平均值作为特征。本文利用 Gensim 工具，在维基百科数据上进行训练，词向量的维度设置为 400。

对行为特征和语言特征采用了三种量化方式，以探讨不同的量化方式对检测效果的影响。根据相关工作中的研究结果，抑郁用户和正常用户在社交网络上的行为和词汇的使用上是有区别的，这种区别可以通过行为或词汇的使用频次、频率的差异来度量。

频次（TF，Term Frequency）。对语言特征，统计某用户发布的全部帖子中包含第 j 类特征词的帖子总条数。例如，对于第一人称单数，统计包含第一人称单数的帖子总条数。对行为特征，统计用户帖子中包含或具有该行为特征的帖子总条数，例如，统计点赞数不为 0（被点赞过）的帖子的总条数，如式（1）所示。

$$\mathrm{TF}_j = \sum_{i=1}^{n} \mathrm{include}(d_i, w_j)$$

$$\mathrm{include}(d_i, w_j) = \begin{cases} 1, \text{如果帖子 } d_i \text{ 包含特征 } w_j \\ 0, \text{其他} \end{cases} \tag{1}$$

式（1）中，d_i 是用户发布的第 i 条帖子，w_j 是第 j 类特征词，n 是该用户发布的帖子总数量。

归一化频率（NTF，Normalized TF）：把某用户第 j 类特征发生的频次转换为频率，即映射到[0，1]之间，如式（2）所示。

$$\mathrm{NTF}_j = \frac{\mathrm{TF}_j}{n} \tag{2}$$

式（2）中，TF_j 是某用户发布的包含第 j 类特征的帖子数量（频次），n 是该用户发布的贴子总数。

Z-Score 标准化频率（ZTF，Z-Score normalized TF）：对全部用户某一特征的归一化频率进行 Z 分值标准化，Z 分值标准化如式（3）所示。

$$\mathrm{ZTF}_j = \frac{\mathrm{NTF}_j - \mu}{\sigma} \tag{3}$$

式（3）中，NTF_j 是式（2）所计算的归一化频率，μ 和 σ 是 NTF_j 在全部用户上的平均值和标准差。

2.3 训练样本选择

相对正常用户，社交网络上的抑郁用户数量非常少，因此采集的样本通常是极度不平衡的，如表 1 所示。大量的研究工作采用随机选择或利用高底分组的方法构建平衡训练样本。本文对是否需要构建以及如何构建平衡样本进行探讨。

在数据采集阶段，用户被分成了三组：正常组，

① http://ictclas.nlpir.org/.

② http://radimrehurek.com/gensim/.

轻度抑郁组，重度抑郁组。实验阶段采用三种不同的样本选择方式来构建数据集。

(1) 非平衡高低分组样本(UHLSG，unbalance high/low scores grouping)：选取表 1 中所有的正常用户组和所有的重度抑郁用户组的数据，构成数据集。

(2) 平衡高低分组样本(BHLSG，balance high/low scores grouping)：由于重度抑郁用户数量与正常用户数量差异巨大，因此，为构建平衡样本，根据抑郁问卷得分由低到高(分值越高，抑郁越严重)选取表 1 中与重度抑郁用户组人数相同的正常用户，与重度抑郁用户组一起构成数据集。

(3)离散化高低分组样本(DHLSG，discretized high/low scores grouping)：参照文献[20]对用户抑郁问卷得分由低到高进行排序，通过式(4)对用户进行离散化：

$$\begin{aligned}\alpha &= E(x) - \sigma(x)\\ \beta &= E(x) + \sigma(x)\end{aligned} \tag{4}$$

其中，$E(x)$代表所有用户抑郁问卷自评得分的平均值，$\sigma(x)$代表所有用户问卷得分的标准差。将抑郁问卷分值的区间[1，123]划分为三段，分值介于[1，α]的用户为低分组用户，分值介于[β，123]的用户为高分组用户，数据集由低分组用户与高分组用户构成。式(4)的实质是找分值有显著差异的样本。

本文对 QQ 空间数据集(简称 QD)和微信朋友圈数据集(简称 WD)都进行了如上三种样本选择，得到的样本数量如表 2 所示。其中，低分组或正常组用户被贴上 normal 或“+”标签，高分组或严重抑郁组用户被贴上 depressed 或“−”标签。微信数据集因重度抑郁人数只有 14 人，样本太少，实验中放弃使用相应的平衡高低分组的样本采样方法。

表 2　QQ 空间和微信朋友圈数据集样本选择结果

数据集	数据集__采用方法	normal	depressed	样本总量
QD	QD_UHLSG	921	42	963
	QD_BHLSG	42	42	84
	QD_DHLSG	213	213	426
WD	WD_UHLSG	184	105	289
	WD_BHLSG	—	—	—
	WD_DHLSG	105	105	210

2.4　相关性分析

由于文本内容特征中的主题特征 Topics 和词向量特征 Word2Vect 是基于数据集分析的结果，不依赖于某个具体的词或词类，因此，相关性分析只在行为特征和语言特征两类上开展。在 QQ 空间和微信朋友圈数据集上各得到 78 个语言和行为特征，但是并不是所有的特征都是与抑郁相关的。因此，本文通过分析各特征值与抑郁自评量表得分之间的相关性，选择相关性较高且显著的特征用于分类模型中。本文假设所有特征的取值服从正态分布，采用皮尔逊相关系数分析特征值与用户抑郁自评量表得分之间的相关性。

2.2 节中介绍了对 QQ 空间数据集和微信朋友圈数据集上的候选特征的三种特征量化方法，本文在三种不同的候选特征量化方法上分别进行相关性分析和显著性分析。相关性分析结果显示，选择频次 TF 量化方法时，两个数据集上的各候选特征与抑郁自评量表得分的相关性都小于 0.1，且相关性不显著(显著水平均远大于 0.05)。因此，本文后续实验只考虑除频次 TF 量化方法以外的其他两种候选特征量化方法。本文选取显著水平小于 0.05 的特征，即该特征有 95%以上的可能性与用户的抑郁自评量表得分是相关的。由于筛选后的特征主要为来自 LIWC 的语言特征，因此统称它们为 LIWC 特征。

表 3 是在 QQ 空间数据集 QD_BHLSG 上通过相关性分析筛选得到的特征，特征量化方法为 Z-Score 方法。包括微信朋友圈数据集在内的不同数据集、不同特征量化方法上的特征选择过程类似，选择结果不再赘述。

表 3　QD_BHLSG 数据集上特征选择结果(Z-Score 量化)

特征	意义	Pearson 系数	显著水平
mention	@符号	−0.23	0.035
shehe	第三人称单数代名词	−0.238	0.029
verb	动词	−0.228	0.037
preps	介词	−0.257	0.019
interjunction	语气助词	−0.228	0.037
multiFun	多用途词	−0.223	0.041
family	家族词	−0.219	0.045

续表

特征	意义	Pearson系数	显著水平
discrep	差距词	−0.24	0.028
excl	排除词	−0.257	0.018
percept	感知历程词	−0.215	0.049

2.5 检测模型选择

在检测模型上，选择了相关工作中分类效果较好的模型，同时也对比了其他具有代表性的分类模型，包括 Naïve Bayes、LibSVM、SMO、Voted Perceptron、SGD(Stochastic Gradient Descent)，其中 Naïve Bayes、LibSVM、SMO、Voted Perceptron 模型来自 Weka，SGD (Stochastic Gradient Descent) 模型来自 Python scikit-learn，模型参数基于网格搜索法进行设置。

3 实验分析

在 QQ 空间数据和微信朋友圈数据上均进行了同样的实验，限于篇幅，重点对 QQ 空间数据集上的实验结果进行分析，同时也对微信数据集上的一些有趣的结果进行说明。

选用的评价指标有精确率 P、召回率 R、F_1 值，评测得分为十折交叉验证的结果。实验结果中，P_-、R_- 和 F_- 分别表示对抑郁用户分类的精确率、召回率和 F_1 值；P_+、R_+ 和 F_+ 分别表示对正常用户分类的精确率、召回率和 F_1 值。$P_\pm$、$R_\pm$ 和 $F_\pm$ 表示相应指标在两类用户上的加权平均，如式(5)所示。

$$X_\pm = X_+ \cdot \mathrm{Per}_+ + X_- \cdot \mathrm{Per}_- \quad (5)$$

式(5)中，X 表示 P、R 或 F，Per_+ 和 Per_- 表示正常用户和抑郁用户的比例。

3.1 样本选择对抑郁用户检测的影响

Z-Score 标准化是文献中通常采用的一种特征量化方法[21]，也是在本文的实验中表现较好的特征量化方法，因此，在考察样本选择对抑郁用户检测的影响时，采用 Z-Score 标准化方法(ZTF)对特征进行量化，分类器用到的特征为 LIWC 特征。表 4 是不同的样本选择方法在分类器为 LibSVM、Voted-Perceptron、NaiveBayes、SMO 和 SGD 上的表现。

表 4 中的实验结果显示：总体上，非平衡高低分组样本 QD_UHLSG 效果最差，平衡高低分组样本 QD_BHLSG 比离散化高低分组样本 QD_DHLSG 效果要好。在非平衡高低分组样本 QD_UHLSG 上，$P_\pm$、$R_\pm$、$F_\pm$ 均达到了 0.9 以上，然而 P_-、R_-、F_- 却非常小，表明在 QD_UHLSG 数据集上构建的模型无法识别抑郁用户，将几乎全部的抑郁用户都识别成了正常用户，原因是 QD_UHLSG 是一个极度不均衡的数据集，正常用户 921 个，抑郁用户 42 个，而本文所选择的模型没有处理样本的不均衡问题。

表 4 样本选择对抑郁用户检测的影响
(特征：LIWC；特征量化方法：ZTF)

样本选择 \ 分类模型		LibSVM	Voted Perceptron	Naive Bayes	SMO	SGD
QD_UHLSG	$P_\pm$	0.915	0.915	0.922	0.922	0.915
	$R_\pm$	0.956	0.955	0.936	0.936	0.956
	$F_\pm$	0.935	0.935	0.928	0.928	0.935
	P_-	0	0	0.115	0.150	0
	R_-	0	0	0.071	0.071	0
	F_-	0	0	0.088	0.088	0
QD_BHLSG	$P_\pm$	0.626	0.637	0.618	0.624	0.651
	$R_\pm$	0.619	0.631	0.607	0.619	0.643
	$F_\pm$	**0.614**	**0.627**	**0.598**	**0.616**	**0.638**
	P_-	0.596	0.608	0.582	0.600	0.615
	R_-	0.738	0.738	0.762	0.714	0.762
	F_-	0.660	0.667	0.660	0.652	0.681
QD_DHLSG	$P_\pm$	0.604	0.556	0.500	0.567	0.581
	$R_\pm$	0.601	0.556	0.500	0.566	0.580
	$F_\pm$	**0.598**	0.556	0.461	0.564	0.579
	P_-	0.586	0.554	0.500	0.559	0.573
	R_-	0.685	0.582	0.770	0.624	0.629
	F_-	0.632	0.568	0.606	0.590	0.600

在平衡高低分组数据集 QD_BHLSG 上，大部分模型的评测分值均大于其在离散化高低分组数据集 QD_DHLSG 上的分值(Naive Bayes 分类器上的 R_- 稍小)，表明平衡高低分组样本选择效果比离散化高低分组样本选择效果要好。对比 QD_BHLSG 数据集和 DQ_DHLSG 数据集上的实验结果，如果仅从高低分组数据集的角度考虑，使用抑郁自评得分越极端的用户，所训练出的模型评测结果越优良。原因是，相对 QD_DHLSG 数据集(正负样本各 213

个），QD_BHLSG 数据集上样本更少，正负样本各 42 个，分值分布更极端，用户更集中，抑郁特征更突出、更显著，而离散化高低分组样本的高分组中同时包含严重抑郁用户和轻度抑郁用户。

使用归一化的特征量化方法 NTF 时，得到的实验结论与 ZTF 上的结论是一致的。

3.2　特征量化方法对抑郁用户检测的影响

表 5 显示了在 QD_BHLSG 数据集上，选择不同的特征量化方法对抑郁用户检测的影响。可以看出，使用 Z-Score 标准化频率 ZTF 对特征进行量化比使用归一化频率 NTF 效果好。使用 ZTF 特征量化方法时，$P_{\pm}$、$R_{\pm}$、R_{-}、$F_{\pm}$ 和 F_{-} 在所有分类模型上均大于或等于 NTF 方法，特别是 LibSVM 和 VotedPerceptron 两个分类模型在 R_{-} 上表现明显。一个可能的原因是，由于 QD_BHLSG 数据集样本数量有限，该数据集上的特征值波动较大，且特征值的分布与其实际分布有较大差异，ZTF 量化方法降低了这种波动，而 NTF 却没有。

表 5　特征量化方法对抑郁用户检测的影响
（特征：LIWC；数据集：QD_BHLSG）

量化方法 \ 分类模型		LibSVM	Voted Perceptron	Naive Bayes	SMO	SGD
归一化频率 NTF	$P_{\pm}$	0.612	0.596	0.589	0.624	0.641
	$R_{\pm}$	0.607	0.583	0.583	0.619	0.631
	$F_{\pm}$	0.603	0.570	0.576	0.616	0.625
	P_{-}	0.636	0.630	0.566	0.600	0.604
	R_{-}	**0.500**	**0.405**	0.714	0.714	0.762
	F_{-}	0.560	0.493	0.632	0.652	0.674
Z-Score 标准化频率 ZTF	$P_{\pm}$	0.626	0.637	0.618	0.624	0.651
	$R_{\pm}$	0.619	0.631	0.607	0.619	0.643
	$F_{\pm}$	0.614	0.627	0.598	0.616	**0.638**
	P_{-}	0.596	0.608	0.582	0.600	0.615
	R_{-}	**0.738**	**0.738**	0.762	0.714	0.762
	F_{-}	0.660	0.667	0.660	0.652	**0.681**

3.3　不同分类模型在抑郁用户检测上的效果

表 4 和表 5 列出了五种分类模型在不同数据集和不同特征量化时的表现。总体上看，SGD 分类器的性能表现更突出，其在 QD_BHLSG 数据集和 ZTF 特征量化时表现达到最佳，$F_{\pm}$ 和 F_{-} 的值分别为 0.638 和 0.681。但表 4 和表 5 同时也显示，在不同的数据集上、采用不同的特征量化方法时，不同的分类模型的表现并不完全一致，例如，在 QD_DHLSG 数据集上使用 ZTF 特征量化时，LibSVM 较其他模型要好（表 4）。

3.4　文本内容特征对抑郁用户检测的影响

以上实验所使用的特征主要是语言学特征，即 LIWC 特征。本节讨论其他文本特征，包括 BOW、Topics、Word2Vect。根据 3.3 节可知，在 QQ 空间数据集上，使用平衡高低分组的样本选择方法、Z-score 标准化的特征量化方法，以及 SGD 分类模型，得到的检测效果较好，因此本节的实验延用这些方法。Z-score 标准化还可以应对不同类型特征取值范围的差异给检测模型带来的挑战。

表 6 是在 QD_BHLSG 数据集上，SGD 分类器在 LIWC、BOW、Topics、Word2Vect 上的检测效果。其中 Topics 特征上，主题数设置为 25，主题数对检测效果的影响如图 1 所示。

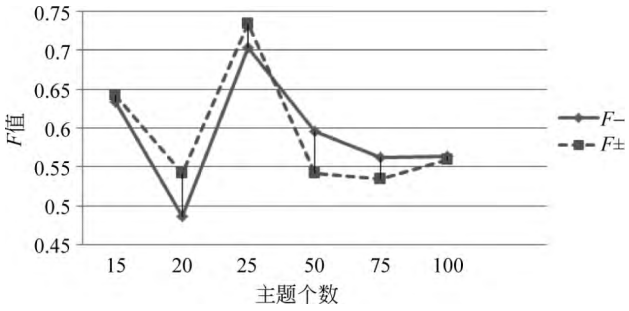

图 1　主题数对检测模型的影响

实验结果表明，相对于 BOW 和 Word2Vect，LIWC 特征的效果较好。原因有两个方面，一是 LIWC词典本身是从心理学的角度对文本内容进行统计分析，二是 2.4 节中通过相关性分析保留了与抑郁自评结果相关性较高的词类，减少了潜在的噪声干扰。在 LIWC、BOW 和 Word2Vect 这三类特征中，词袋特征 BOW 表现最差，这与其在其他文本分类问题中的表现类似。

相对于 LIWC、BOW 和 Word2Vect，主题特征 Topics 的表现更佳，其 F 值达到 0.753，而对抑郁用户的检测精确率 P_{-} 达到 0.813。主题特征考虑了上下文之间的语义关联，从更深层次挖掘出了文本之间的语义关联性，取得较好的效果。然而图 1 也让我们也看到，主题个数的选择对于检测模型有较大的影响。

比较意外的是，当在主题特征 Topics 的基础上增加其他特征时，检测的效果反而下降。但从另一个角度看，在 LIWC、BOW 和 Word2Vect 这三类特征的基础上，增加主题特征 Topics 后，检测效果都有显著提高，这也证明了主题特征在抑郁用户检测中的重要作用。

表 6 SGD 分类器在文本内容特征上的效果
（数据集：QD_BHLSG；特征量化方法：ZTF）

	P_-	R_-	F_-	$P_\pm$	$R_\pm$	$F_\pm$
LIWC	0.615	0.762	0.681	0.651	0.643	0.638
Word2Vect	0.583	0.500	0.538	0.573	0.572	0.569
BOW	0.571	0.476	0.519	0.561	0.560	0.556
Topics(k=25)	0.813	0.619	0.703	0.753	0.738	0.734
Topics+LIWC	0.635	0.786	0.702	0.677	0.667	0.662
Topics+Word2Vect	0.591	0.619	0.605	0.596	0.595	0.595
Topics+BOW	0.632	0.571	0.600	0.621	0.619	0.619

3.5 与相关文献的对比分析

为进一步分析在准私密社交网络数据上进行抑郁用户检测的可行性，本节介绍相关文献中利用 Twitter、微博、Blog、网关日志等数据检测抑郁用户的效果。

文献[3]以 476 个用户的 Twitter 数据作为数据集，其中抑郁用户 171 个，正常用户 305 个，定义了六种抑郁行为衡量方法，包括 engagement、ego-network、emotion、linguistic style、depression language、demographics，通过相关性分析筛选得到与抑郁最相关的特征，选择 SVM 为检测模型，得到的最好结果中，精确率和召回率分别为 0.742 和 0.629，显著低于本文的 0.753 和 0.738。

文献[6]以中文新浪微博数据为数据源，在行为特征、交互特征和语言特征的基础上，引入微博的情感特征，并借助心理学家对数据的观察分析结果，利用 Bayes、Trees、Rules 等几类模型进行抑郁用户检测，在抑郁和正常用户各 90 个的数据集上，F 值的最好效果为 0.85。文献[7]是在文献[6]的基础上，考虑社会关系（链接）特征后，检测正确率达到 0.95。进一步分析发现，文献[6]和文献[7]取得较好效果主要有以下两个原因。首先，在数据集的采集上，除采用用户自评量表外，还配合访谈的方式进一步确认用户的抑郁倾向，较本文只采用自评量表的方式，采集的数据集质量更高，抑郁用户和正常用户之间的划分更清晰，从而使得特征对数据的区分更强。例如，文献[6]中微博数量特征和情感符数量特征与抑郁自评分值的相关显著水平达 0.002 和 0.003，远低于本文表 3 中的最低值 0.018。其次，使用了情感、社会关系等更丰富的特征，并且通过心理学家辅助特征的筛选。

文献[21]以用户的网关日志为数据源，把 728 个用户分为 449 个抑郁用户和 279 个正常用户，组成训练集，通过聚类和离散傅里叶变换分别得到了聚类特征和频率特征，对抑郁用户检测的精确率和召回率最高分别为 0.756 和 0.623，相应的 F 值为 0.683，低于本文的 F 最高值 0.703。

文献[22]针对包括 49 个抑郁用户和 59 个正常用户的日语博客数据，利用 character n-grams、token n-grams、lemmas（词的原形）、词性等特征，通过特征筛选后，用 Naïve Bayes、SVM、Logistic 回归等模型分类，得到最优结果的准确率达 0.95，而最优结果所采用的特征仅为来自动词和副词两种词性且词干化后的 2007 个词，分类模型为 Naïve Bayes。文献[22]分类效果较好，也与其数据集构建有密切关系，其中抑郁用户和正常用户的识别主要依据用户在博客中是否用了“depression（抑郁”一词并透露了他们是抑郁患者。尽管与“depression”主题相关的博客都在后来的实验中被弃用，但与之相关的词汇仍然会给分类器提供较好的指示。该数据集的不足在于它没有包含那些没在博客中用“depression”一词透露其是抑郁患者的用户，而这部分用户相对更难识别，并且检测出那些潜在的、未被确诊的抑郁患者较确诊的抑郁患者有更重要的意义。

与上述文献相比，本文的优势在于：(1)对抑郁用户和非抑郁用户检测的平均 F 值达到 0.734，而对抑郁用户的检测精确率 P_- 达到 0.813，优于文献[3]和文献[21]；(2)不需要心理学家参与构建数据集和特征选择，仅使用社会网络用户的自评量表，对数据质量的要求较文献[6-7]和文献[22]更低；(3)数据含盖未确诊的潜在抑郁用户，较文献[22]更接近真实数据。

4 总结

从特征量化、训练样本选择、模型选择、文本内容特征四个角度考察了利用 QQ 空间这种准私密社交网络数据进行抑郁用户检测的可行性。对比了常用的特征量化方式：频次、归一化频率、Z-Score 标

准化；对比了常用的训练样本选择方式：平衡高低分组方法、非平衡高低分组方法、离散化高低分组方法；对比了 LibSVM、Voted Perceptron、Naïve Bayes、SGD 等分类模型。实验发现：Z-Score 标准化比其他两种特征量化方法要好；平衡高低分组方法较其他样本选择方法要好；检测模型则比较依赖于数据集、样本选择、特征及其量化方法。

实验还分析了在平衡样本上，不同的文本内容特征对抑郁用户检测的影响。结果发现，主题特征对抑郁用户的检测效果最好，其他特征如语言特征 LIWC、词袋 BOW、词向量 Word2Vect 等，在加上主题特征后对检测效果有明显改善。最后还对比分析了相关文献中基于 Twitter、微博、Blog、网关日志等数据检测抑郁用户的效果，明确了本文的优势，进一步说明了使用准私密社交网络数据检测抑郁用户是可行的。

从实验以及与相关工作的对比可以看出，数据集、特征和检测模型都是基于社会网络数据的抑郁用户检测的关键，不同文献在这几方面的差异也较大，可比性不强。另外，已有工作中各种高达 0.8 以上的准确率都是在平衡样本上得到的，与抑郁用户的实际分布差异较大，也意味着在实际应用中还会面临诸多挑战。最后，采用自评量表的方式获取的样本受用户填写量表时的心情影响较大，而确诊抑郁等心理问题需要更长期、更专业的观察，因此，样本采集需要结合心理医生的诊断才更为准确。

参考文献

[1] Xiaohui Liang, Siqi Su, Jiayuan Deng, et al. Investigation of college students' mental health status via semantic analysis of Sina microblog[J]. Wuhan University Journal of Natural Sciences, 2015, 20(2): 159-164.

[2] Tsugawa Shuo, Mogi Yukiko, Kikuchi Yusuke, et al. On estimating depressive tendencies of Twitter users utilizing their tweet data[C]//Proceedings of the Virtual Reality Conference, 2013: 1-4.

[3] Munmun De Choudhury, Michael Gamon, Scott Counts, et al. Predicting depression via social media [C]//Proceedings of the Association for the Advancement of Artificial Intelligence, 2013: 1-10.

[4] Munmun De Choudhury, Scott Counts, Eric Horvitz. Social media as a measurement tool of depression in populations[C]//Proceedings of the ACM Web Science Conference, 2013: 47-56.

[5] Minsu Park, Chiyoung Cha, Meeyoung Cha. Depressive moods of users portrayed in Twitter[C]//Proceedings of the HI-KDD'12, 2012: 978-985.

[6] Xinyu Wang, Chunhong Zhang, Yang Ji, et al. A depression detection model based on sentiment analysis in Micro-blog social network[C]//Proceedings of the Revised Selected Papers of PAKDD 2013 International Workshops on Trends and Applications in Knowledge Discovery and Data Mining, 2013: 201-213.

[7] Xinyu Wang, Chunhong Zhang, Li Sun. An improved model for depression detection in Micro-blog social network[C]//Proceedings of the IEEE 13th International. Conference on Data Mining Workshops, 2013: 80-87.

[8] 李昂，郝碧波，白朔天，等. 基于网络数据分析的心理计算：针对心理健康状态与主观幸福感[J]. 科学通报, 2015(11): 994-1001.

[9] Youn Soo Jeong, Trinh Nhi-Ha, Shyu Irene, et al. Using online social media, Facebook, in screening for major depressive disorder among college students[J]. International Journal of Clinical and Health Psychology, 2013, 13(1): 74-80.

[10] Megan A Moreno, Lauren A Jelenchick, Katie G Egan, et al. Feeling bad on Facebook: Depression disclosures by college students on a social networking site[J]. Depression and Anxiety, 2011, 28(6): 447.

[11] Megan A Moreno, Lauren A Jelenchick, Rajitha Kota. Exploring depression symptom references on Facebook among college freshmen: A mixed methods approach[J]. Open Journal of Depression, 2013, 2(3): 35-41.

[12] Wei Tong Mok, Rachael Sing, Xiuting Jiang, et al. Investigation of social media on depression[C]//Proceedings of the International Symposium on Chinese Spoken Language Processing, 2014: 488-491.

[13] Minsu Park, David W Mcdonald, Meeyoung Cha. Perception differences between the depressed and non-depressed users in Twitter[C]//Proceeding of the Seventh International AAAI Conference on Weblogs and Social media, 2013: 476-485.

[14] Munnmun De Choudhury, Scott Counts, Eric Horvitz. Predicting postpartum changes in emotion and behavior via social media [C]//Proceeding of the Sigchi Conference on Human Factors in Computing Systems. ACM, 2013: 3267-3276.

[15] Munnmun De Choudhury, Scott Counts, Michael Gamon. Not all moods are created equal ! Exploring human emotional states in social media[C]//Proceedings of the 6th International AAAI Conference on Weblogs and Social Media, 2012: 66-73.

[16] Thin Nguyen, Dinh Phung, Bo Dao, et al. Affective and content analysis of online depression communities

[J]. IEEE Transactions on Affective Computing, 2014,5(3): 217-226.

[17] Bo Dao, Thin Nguyen, Dinh Phung, et al. Effect of mood, social connectivity and age in online depression community via topic and linguistic Analysis[C]//Proceeding of the 15th Web Information Systems Engineering, 2014: 398-407.

[18] Shuotian Bai, Bibo Hao, Ang Li, et al. Depression and anxiety prediction on Microblogs[J]. Molecular Microbiology, 2014, 5(8): 814-820.

[19] Sungkyu Park, Sang won Lee, Jinah Kwak, et al. Activities on Facebook reveal the depressive state of users[J]. Journal of Medical Internet Research, 2013, 15(10): 217.

[20] Shuotian Bai, Tingshao Zhu, Cheng Li. Big-five personality prediction based on user behaviors at social network sites[J]. Computer Science, 2012, 8(2): 2682-2682.

[21] Changye Zhu, Baobin Li, Aang Li, et al. Predicting depression from Internet behaviors by time-frequency features[C]//Proceedings of the Ieee/wic/acm International Conference on Web Intelligence, 2017: 383-390.

[22] Misato Hiraga. Predicting depression for Japanese blog Text[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Student Research Workshop, 2017: 107-113.

[23] **钟毓,费定舟. 基于稀疏主成分分析的非正式语词的心理-人格特征研究**[J]. **中文信息学报**, 2017, 31(1): 192-204.

[24] Guan Li, Bibo Hao, Qijin Cheng, et al. Identifying Chinese microblog users with high suicide probability using internet-based profile and linguistic features: classification model[J]. Jmir Mental Health, 2015, 2(2): 17.

[25] Lenore Sawyer Radloff. The CES-D Scale: A self-report depression scale for research in the general population[J]. Applied Psychological Measurement, 1977, 1(3): 385-401.

[26] Aaron T Beck, Robert A Steer, Margery G Carbin. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation[J]. Clinical Psychology Review, 1988, 8(1): 77-100.

刘德喜(1975—),**博士,教授,主要研究领域为社会媒体处理、自然语言处理、信息检索。**
E-mail: dexi.liu@163.com

万常选(1962—),**博士,教授,主要研究领域为知识工程与数据挖掘、情感计算、XML数据库。**
E-mail: wanchangxuan@263.net

邱家洪(1991—),**硕士研究生,主要研究领域为数据挖掘、信息检索和自然语言处理。**
E-mail: jiahong3837@foxmail.com