

新浪微博用户中自杀死亡和无自杀意念者特征差异的研究

管理 郝碧波 刘天俐 程绮瑾 叶兆辉 朱廷劭

【摘要】目的 探讨新浪微博用户中自杀死亡与无自杀意念者行为和语言特征差异。**方法** 经新浪微博认证用户提供的信息,收集31名网络识别自杀死亡用户(自杀死亡组);邀请微博用户填写自杀意念相关筛查量表,收集30名无自杀意念用户(对照组)。比较两组用户在10种微博行为和88种语言特征方面的差异。**结果** (1)行为特征中,自杀死亡组的微博链接率(链接微博数与公开微博总数的比值)和微博互动率(平均每篇微博@其他用户的次数)均低于对照组[0.04(0.04) vs. 0.06(0.04), $P=0.029$; 0.60(0.27) vs. 0.69(0.18), $P=0.028$],自我关注程度(平均每篇公开微博使用的第一人称单数次数)高于对照组[0.47(0.25) vs. 0.30(0.10), $P=0.010$];(2)语言特征中,自杀死亡组在数量单位词、工作词、省略号使用率低于对照组(均 $P<0.05$),在代名词、特定人称代名词、第三人称单数、非特定人称代名词、社会历程词、焦虑词、排除词、性词、宗教词、第二人称单数、人类词、消极情绪词、愤怒词、悲伤词和死亡词的使用率均高于对照组(均 $P<0.05$)。**结论** 自杀死亡用户的微博互动更少,更加关注自我,更频繁地使用表达排除意义的词语,从情感层面上有更多负性表达,使用更多与死亡、宗教相关而更少与工作相关的表达。这些特点对开展针对网络用户的自杀学研究有重要启示。

【关键词】 自杀; 微博; 网络行为; 文本分析; 社交媒体

A pilot study of differences in behavioral and linguistic characteristics between Sina suicide microblog users and Sina microblog users without suicide idea Guan Li^{1,2}, Hao Bibo², Liu Tianli³, Cheng Qijin⁴, Yip Paul Siu Fai⁴, Zhu Tingshao¹. 1 Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China; 2 University of Chinese Academy of Sciences; 3 Institute of Population Research, Peking University; 4 Hong Kong Jockey Club Center for Suicide Research and Prevention, the University of Hong Kong

Corresponding author: Zhu Tingshao, Email: tszhu@psych.ac.cn

【Abstract】Objective To investigate how suicide microblog users in China “act” and “speak” differently from other microblog users without suicide idea. **Methods** The suicide group consisted of 31 Chinese microblog users identified as suicide via online information provided by a Sina microblog user, and the control group consisted of 30 active microblog users without suicide idea screened by using suicide-related psychological scales. The differences in 10 microblog use behavioral characteristics and 88 linguistic characteristics between the suicide group and the control group were compared with normality test and rank sum test respectively. **Results** In the behavioral characteristics, the suicide group used hyperlinks and “@” less frequently than the control group [0.04 (0.04) vs. 0.06 (0.04), $P=0.029$; 0.60 (0.27) vs. 0.69 (0.18), $P=0.028$], and was more self-focused [0.47 (0.25) vs. 0.30 (0.10), $P=0.010$]. In the linguistic characteristics, the suicide group showed less frequency in using measure word, work related word and apostrophe than the control group ($P<0.05$), and showed more frequency in using pronoun, personal pronoun, third person singular, non-specific pronoun, word expressing social experience, word expressing anxiety, word expressing exclusion, sexual word, religious word, second person singular, human being related word, negative emotion related word, anger related word, sadness or death related word ($P<0.05$). **Conclusion** Suicides seemed to interact less with others, showed more self-concern and more negative expressions, use more cognitively exclusive, death-related, religion-related words, and use less work-related words. The results of this study might be helpful for the research on suicide among netizen.

【Key words】 Suicide; Microblog; Microblog use behavior; Text analysis; Social media

DOI: 10.3760/cma.j.issn.0254-6450.2015.05.003

作者单位: 100101 北京, 中国科学院心理研究所社会与工程研究室(管理、朱廷劭); 中国科学院大学(管理、郝碧波); 北京大学人口研究所(刘天俐); 香港大学香港赛马会防止自杀研究中心(程绮瑾、叶兆辉)

通信作者: 朱廷劭, Email: tszhu@psych.ac.cn

自杀是一种严重的心理健康问题。2014 年我国城市居民自杀年死亡率为 4.82/10 万,农村居民则高达 8.58/10 万^[1]。传统自杀风险评估研究主要采用心理测验、访谈和问卷等^[2],其成本耗费较大。研究表明,尽管自杀死亡的个体生前很少主动寻求帮助,但在日常言语表达中往往有所流露自杀意念^[3-4]。近年来微博成为社交网络用户具有时效性自我表达的途径之一,因其具有开放可见性,部分学者已关注通过该平台开展自杀相关研究^[5-7]。不论是线上或者线下,语言能够显露出个体身份和社会关系等重要信息^[8-9]。近年来国内外学者在心理健康评估中使用网络社交媒体的文本分析方法已取得一些成果^[10-14]。但由于网络的匿名特性和研究自杀本身的困难^[2],总体而言该方法对自杀风险评估和预警仍处在初级阶段,目前尚无有效在线评估用户个体自杀意念和自杀未遂的研究报道。为此本研究在新浪微博平台上探讨自杀死亡用户的微博行为和语言模式,比较网络识别自杀死亡用户和无自杀意念用户的行为和文本特征差异,为在社交媒体平台开展自杀干预提供思路和数据支持。

对象与方法

1. 研究对象:自杀死亡用户样本来自社交网络。通过媒体报道,了解并联络新浪微博认证用户“逝者如斯夫 dead”,在其协助下获得被描述为自杀死亡的微博用户账号。在每个疑似自杀死亡账号的微博主页中,从至少 3 名被该账号关注的微博用户留言中再次确认该账号所有者确系死于自杀。2014 年 2—5 月共收集网络识别自杀死亡新浪微博用户账号 35 个,从中剔除外国籍 1 人, ID 由经纪人代理 1 人,认证名 1 人,以及原创微博数过少(<20 条)1 人,最终用于分析的网络识别自杀死亡(自杀死亡组)31 人(男性 12 人,女性 19 人)。

采用 3 个量表筛选无自杀意念微博用户(对照组)。①贝克自杀意念量表(Scale for Suicide Ideation, SSI)^[15]共 19 个项目,如果用户做到项目 5 则可通过,且总分越低自杀意念越低;②自杀可能性量表中文版(Suicide Possibility Scale, SPS)^[16]共 36 个项目,分数越高其自杀风险越大,本研究按刘国华和张桂英^[17]的筛选标准(<50 分);③抑郁自评量表(Self-rating Depression Scale, SDS)^[18]共 20 个项

目,分数越高抑郁倾向越大,如<53 分则通过。在知情同意下,于 2014 年 5 月共向 1 500 名新浪微博用户发放问卷,回收 1 385 份,筛选出无自杀意念用户 91 人,其中保留发布公开微博数>100 条者 52 人,为避免“试验者效应”,从中剔除 22 名与作者有社交关系的用户,最终确认符合对照组标准的新浪微博活跃用户 30 人(男 15 人,女 15 人)。

本研究通过新浪微博数据开放平台,应用程序编程接口经伦理许可的方法,完成两组用户自注册起至 2014 年 10 月 8 日所有公开可见的微博数据,并获得中国科学院心理研究所伦理委员会的审查批准。

2. 研究方法:

(1)行为特征提取:微博用户行为特征主要体现在自我表达的特点、与其他用户的互动,以及活跃周期等,共提取 10 类用户行为数据(表 1)。由于自杀死亡组不再有更新,而对照组一直处于活跃状态,为了排除时间因素对用户微博特征值的影响,行为特征指标大多使用比率数据,或平均每篇微博所含特征计数值。行为特征值的提取和计算由计算机程序完成。

(2)语言特征提取:本研究使用中文版 LIWC (Linguistic Inquiry and Word Count)词库^[19-21],涉及词类包括总体描述词、语言过程词、心理过程词、个人关切词等共计 88 类(图 1),并由中国科学院心理研究所计算机网络心理实验室研发的“文心中文心理分析系统”完成语言特征的提取,即由计算机程序从每名用户的微博文本中自动输出 88 类语言特征,计算每一种类别的词语使用频率百分数值。

3. 统计学分析:使用 SPSS 17.0 软件。首先对特征值分布进行 Kolmogorov-Smirnov (K-S)正态性检验,对于符合正态分布的行为和语言特征,两组之间的差异使用独立样本 *t* 检验;对于不符合正态分布的

表 1 本研究新浪微博用户 10 类行为特征及其操作定义

行为特征	操作定义
自我描述长度	显示在用户主页顶端简短自我描述的字节
微博原创率	用户原创微博数与公开微博总数的比值
微博链接率	用户含有链接的微博数与公开微博总数的比值
微博互动率	用户平均每篇公开微博@其他人的次数
集体关注度	用户平均每篇公开微博使用的第一人称复数次数
自我关注度	用户平均每篇公开微博使用的第一人称单数次数
夜间活跃度	用户在 22:00 至次日 06:00 公开发布微博数与公开微博总数比值
积极表情图片使用率	用户平均每篇公开微博使用的积极表情次数
消极表情图片使用率	用户平均每篇公开微博使用的消极表情次数
社交活跃度	用户与其他人互相关注次数与被关注次数的比值

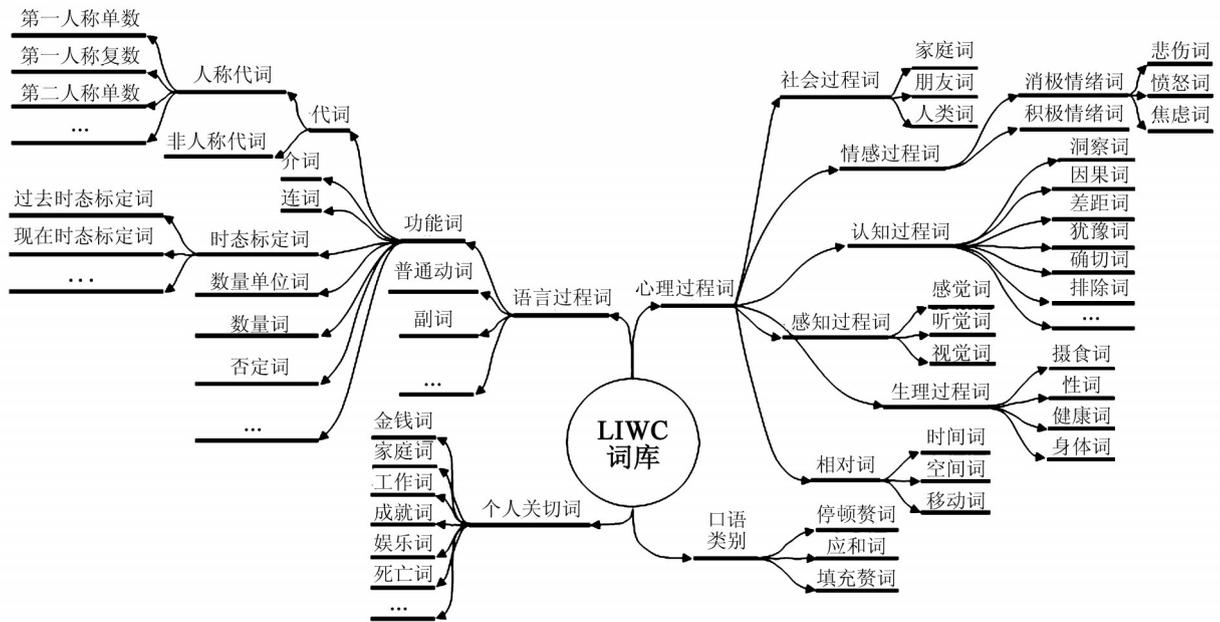


图1 简体中文LIWC词典词类特征

行为和语言特征，两组之间的差异使用 Mann-Whitney *U* 秩和检验。

结果

1. 一般特征：自杀死亡组生前公开发表微博数 $M=316.0$ [四分位间距 (Q_R) = 1 216.0]，平均活跃天数为 (198.0 ± 162.2) d；对照组公开发表微博数 $M=932.0$ ($Q_R=915.8$)，平均活跃天数为 (471.5 ± 197.3) d。表明两组在生前或现阶段均较深程度使用微博，并有一定的网络社交水平。

2. 组间差异分析：

(1) 微博行为特征：K-S 正态性检验结果表明，10 种行为特征中有“微博原创率”、“集体关注度”、“夜间活跃度”和“社交活跃度”4 个特征服从正态分布 ($P < 0.05$)。对这些特征采用独立样本 *t* 检验显示，自杀死亡组和对照组在“社交活跃度”特征上的差异有统计学意义 (表 2)。其他行为特征进行

Mann-Whitney *U* 秩和检验，结果表明，自杀死亡组的“微博链接率”、“微博互动率”和“自我关注度”与对照组的差异有统计学意义 (表 3)。

表2 两组人群符合正态分布的4项行为特征对比 ($\bar{x} \pm s$)

行为特征	自杀死亡组 ($n=31$)	对照组 ($n=30$)	<i>t</i> 值	<i>P</i> 值
微博原创率	0.47 ± 0.26	0.55 ± 0.20	-1.28	0.205
集体关注度	0.07 ± 0.05	0.06 ± 0.02	1.10	0.275
夜间活跃度	0.17 ± 0.10	0.19 ± 0.08	-0.68	0.499
社交活跃度	0.14 ± 0.13	0.42 ± 0.16	-7.52	<0.001

注：“集体关注度”单位是个/篇，其余特征值为比率

(2) 微博语言特征：采用 K-S 正态性检验，88 种语言特征中有 61 类特征值服从正态分布 ($P > 0.05$)。独立样本 *t* 检验显示，自杀死亡组和对照组在“代名词” (如他、你们、在下)、“特定人称代名词” (如你、在下、他们)、“第三人称单数” (如他、她)、“非特定人称代名词” (如一切、这些、其他)、“数量单位

表3 两组人群不符合正态分布的6项行为特征对比

行为特征	自杀死亡组 ($n=31$)		对照组 ($n=30$)		<i>U</i> 值	<i>W</i> 值	<i>Z</i> 值	<i>P</i> 值
	<i>M</i>	<i>Q_R</i>	<i>M</i>	<i>Q_R</i>				
自我描述长度	13.00	15.00	9.00	7.20	374.0	839.0	-1.32	0.188
微博链接率	0.04	0.08	0.06	0.08	314.0	810.0	-2.18	0.029
微博互动率	0.60	0.54	0.69	0.36	313.0	809.0	-2.19	0.028
自我关注度	0.47	0.50	0.30	0.20	286.0	751.0	-2.58	0.010
积极表情使用率	0.04	0.08	0.04	0.06	445.5	941.5	-0.28	0.778
消极表情使用率	0.07	0.20	0.09	0.14	384.5	880.5	-1.16	0.245

注：*U* 值为 Mann-Whitney *U* 秩和检验；*W* 值为 Wilcoxon rank sum test 检验的秩和值；“自我描述长度”单位是字节数，“自我关注度”、“微博互动率”和“积极/消极表情使用率”单位是个/篇，其余特征值为比率数据

词”(如条、头、只)、“社会历程词”(如家人、接纳、打招呼)、“焦虑词”(如不安、挣扎、紧绷)、“排除词”(如取消、但是、除外)、“性词”(如上床、性欲、裸体)和“宗教词”(如上帝、慈悲、信仰)特征值的差异有统计学意义(表4)。其他语言特征进行Mann-Whitney U秩和检验。结果表明,自杀死亡组在“第二人称单数”(如你、您)、“人类词”(如人民、成员、群众)、“消极情绪词”(如担忧、猜疑、嫉妒)、“愤怒词”(如可恶、抱怨、破坏)、“悲伤词”(如心痛、沮丧、无力)、“工作词”(如工厂、面试、薪水)、“死亡词”(如亡故、自杀、遗嘱)和“省略号”(…)的使用频率上与对照组的差异有统计学意义(表5)。

表4 两组人群中符合正态分布语言特征的对比($\bar{x} \pm s$)

语言特征	自杀死亡组 (n=31)	对照组 (n=30)	t值	P值
代名词	6.20±2.96	4.22±1.09	3.44	0.001
特定人称代名词	4.46±2.35	2.76±0.97	3.67	0.001
第三人称单数	0.30±0.21	0.20±0.11	2.30	0.025
非特定人称代名词	0.96±0.49	0.70±0.15	2.80	0.007
数量单位词	1.50±0.59	1.85±0.67	-2.20	0.032
社会历程词	4.23±2.27	3.12±1.40	2.29	0.025
焦虑词	0.20±0.15	0.11±0.07	2.75	0.008
排除词	2.25±1.12	1.77±0.36	2.22	0.030
性词	0.15±0.11	0.09±0.04	3.03	0.004
宗教词	0.16±0.11	0.10±0.04	2.62	0.011

注:仅报告差异有统计学意义的结果;特征值均为词语使用频率(%)

表5 两组人群中不符合正态分布的语言特征对比

语言特征	自杀死亡组 (n=31)		对照组 (n=30)		U值	W值	Z值	P值
	M	Q _R	M	Q _R				
第二人称单数	0.99	0.68	0.59	0.20	196.0	661.0	-3.88	<0.001
人类词	0.90	0.46	0.65	0.18	255.0	720.0	-3.03	0.002
消极情绪词	0.59	0.44	0.38	0.16	230.0	695.0	-3.39	0.001
愤怒词	0.12	0.22	0.06	0.04	281.0	746.0	-2.65	0.008
悲伤词	0.15	0.16	0.07	0.06	162.0	627.0	-4.37	<0.001
工作词	0.33	0.26	0.45	0.42	292.0	788.0	-2.50	0.013
死亡词	0.16	0.14	0.11	0.08	269.0	734.0	-2.83	0.005
省略号	0.02	0.08	0.08	0.20	302.5	798.5	-2.35	0.019

注:同表4

讨 论

本研究显示,自杀死亡组比对照组自我关注程度更强,链接到他人微博的频率更低,微博中和其他用户的互动也更少。结果表明,有自杀意念的个体可能更关注于自身^[22],这与以往大量研究结果一致^[14,23-24]。研究中还发现自杀死亡组与对照组相比有更多的负面表达。如自杀死亡组在情感方面使用

了更多的消极情绪词以及更为具体的类别,包括焦虑、愤怒、悲伤;在功能词方面使用了更多表达排除的词语;在个人关切方面使用更多表达死亡的词语,包括与自杀直接相关的词汇。已有的研究更多关注自杀死亡者或高自杀风险者表达的消极情感^[23,25-26],有自杀意念的个体在文本中更多使用死亡意象^[23]。

本研究显示在宗教词的使用频率上自杀死亡组高于对照组。以往研究表明皈依宗教信仰是应对个体内心极端冲突的一种方式^[27],自杀死亡的个体有可能采取该方式。研究还显示,工作词的使用频率自杀死亡组低于对照组。以往有研究认为存在心理健康问题(如抑郁症)的个体,其工作积极性及其业务绩效均可受负面影响^[28]。此外,自杀死亡组在多种人称和非人称代词、社会历程词、性词、人类词的使用频率高于对照组,数量词、省略号的使用频率低于对照组,但此结论目前尚无理论依据支持,这些词语类别特征上的差异可能与我国文化语言表达特点相关,值得进一步探讨。

本研究存在局限性。首先获得自杀死亡组的样本信息极为困难。收集网络自杀死亡用户账号主要依赖社交网络媒体报导,与传统研究相比其样本量偏小,且由于网络匿名的特征,现阶段难以大规模开展收集和确认工作,故本研究抽样时未对自杀死亡组和对照组的多种人口统计学信息进行控制存在偏倚。此外国内外目前尚无网络自杀特征的评估体系,尽管本研究在预实验中采取了一些避免结果偏倚的措施,如筛选自杀死亡组和对照组要求具有一定数目的公开发表微博,但仍可能难以严格意义上区分微博用户的“行为特征”和“语言特征”;且目前使用的计算机语言特征提取工具还缺乏对每个特征类别明确的定义。

总之,本研究提示在网络微博上自杀死亡用户确实存在特定行为表现以及语言表达,并与无自杀意念的一般用户存在差异。

参 考 文 献

[1] Center of Health Statistics and Information, Ministry of Health of the People's Republic of China. China Public Health Statistical Yearbook 2013 [R/OL]. (2014-04-26) [2014-07-21]. <http://www.nhfpc.gov.cn/htmlfiles/zwgkzt/ptjnj/year2013/index2013.html>. (in Chinese)
 中华人民共和国卫生部统计信息中心. 2013 中国卫生统计年鉴 [R/OL]. (2014-04-26) [2014-07-21]. <http://www.nhfpc.gov.cn/htmlfiles/zwgkzt/ptjnj/year2013/index2013.html>.
 [2] Li GY. Suicide and self-harm[M]. Beijing: People's Medical Publishing House, 2009: 17-22. (in Chinese)

- 李功迎. 自杀与自伤[M]. 北京: 人民卫生出版社, 2009: 17-22.
- [3] Jiang GR, Xia M. Psychological help-seeking: current research and the phases-decision-making model [J]. *Adv Psychol Sci*, 2006, 14(6): 888-894. (in Chinese)
江光荣, 夏勉. 心理求助行为: 研究现状及阶段-决策模型[J]. *心理科学进展*, 2006, 14(6): 888-894.
- [4] Ray LY, Johnson N. Adolescent suicide [J]. *Person Guid J*, 1983, 62(3): 131-135.
- [5] Silenzio VMB, Duberstein PR, Tang W, et al. Connecting the invisible dots: Reaching lesbian, gay, and bisexual adolescents and young adults at risk for suicide through online social networks [J]. *Soc Sci Med*, 2009, 69(3): 469-474.
- [6] Fu KW, Cheng QJ, Wong PWC, et al. Responses to a self-presented suicide attempt in social media: A social network analysis [J]. *Crisis*, 2013, 34(6): 406-412.
- [7] Ruder TD, Hatch GM, Ampanozi G, et al. Suicide announcement on Facebook [J]. *Crisis*, 2011, 32(5): 280-282.
- [8] Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods [J]. *J Lang Soc Psychol*, 2010, 29(1): 24-54.
- [9] Barak A, Miron O. Writing characteristics of suicidal people on the internet: a psychological investigation of emerging social environments [J]. *Suicide Life-Threat Behav*, 2005, 35(5): 507-524.
- [10] Pestian J, Nasrallah H, Matykiewicz P, et al. Suicide note classification using natural language processing: a content analysis [J]. *Biomed Inform Insights*, 2010, 2010(3): 19-28.
- [11] Jashinsky J, Burton SH, Hanson CL, et al. Tracking suicide risk factors through Twitter in the US [J]. *Crisis*, 2014, 35(1): 51-59.
- [12] Zhang JW. Research on emotion recognition technology of microblog based on psychological prewarning model [D]. Hefei: Hefei University of Technology, 2013. (in Chinese)
张金伟. 微博情感分析的心理预警模型与识别研究 [D]. 合肥: 合肥工业大学, 2013.
- [13] Li L. Internet suicide notes recognition research [D]. Wuhan: Central China Normal University, 2012. (in Chinese)
李隆. 网络文本自杀倾向性识别方法研究 [D]. 武汉: 华中师范大学, 2012.
- [14] Wang XY, Zhang CH, Ji Y, et al. A depression detection model based on sentiment analysis in micro-blog social network [M]. *Trends and Applications in Knowledge Discovery and Data Mining*. Berlin Heidelberg: Springer, 2013: 201-213.
- [15] Li XY, Phillips MR, Tong YS, et al. Reliability and validity of the Chinese version of Beck Suicide Ideation Scale (BSI-CV) in adult community residents [J]. *Chin Mental Health J*, 2010, 24(4): 250-255. (in Chinese)
李献云, 费立鹏, 童永胜, 等. Beck 自杀意念量表中文版在社区成年人中应用的信效度 [J]. *中国心理卫生杂志*, 2010, 24(4): 250-255.
- [16] Liang YN, Yang LZ. Study on reliability and validity of the Suicide Probability Scale [J]. *Chin J Health Psychol*, 2010, 18(2): 225-227. (in Chinese)
梁瑛楠, 杨丽珠. 自杀可能性量表的信效度研究 [J]. *中国健康心理学杂志*, 2010, 18(2): 225-227.
- [17] Liu GH, Zhang GY. Gender differences of the undergraduates' suicide probability [J]. *Chin J Sch Health*, 2012, 33(10): 1198-1200. (in Chinese)
刘国华, 张桂英. 大学生自杀可能性的性别比较 [J]. *中国学校卫生*, 2012, 33(10): 1198-1200.
- [18] Shu L. Self-rating depression scale [J]. *Chin Mental Health J*, 1999, Suppl: S194-196. (in Chinese)
舒良. 抑郁自评量表 (SDS) [J]. *中国心理卫生杂志*, 1999, 增刊: 194-196.
- [19] Pennebaker JW, Francis ME, Booth RJ. Linguistic inquiry and word count: LIWC 2001 [J]. Mahway: Lawrence Erlbaum Associates, 2001, 71: 2001.
- [20] Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: Our words, our selves [J]. *Annu Rev Psychol*, 2003, 54(1): 547-577.
- [21] Gao R, Hao BB, Li H, et al. Developing simplified Chinese psychological linguistic analysis dictionary for microblog [M]. Berlin Heidelberg: Springer, 2013: 359-368.
- [22] Psych MWD, Sedway J, Bulik CM, et al. Linguistic analyses of natural written language: Unobtrusive assessment of cognitive style in eating disorders [J]. *Int J Eat Disord*, 2007, 40(8): 711-717.
- [23] Li TM, Chau M, Yip PSF, et al. Temporal and computerized psycholinguistic analysis of the blog of a Chinese adolescent suicide [J]. *Crisis*, 2014, 35(3): 168-175.
- [24] Stirman SW, Pennebaker JW. Word use in the poetry of suicidal and nonsuicidal poets [J]. *Psychosom Med*, 2001, 63(4): 517-522.
- [25] Li TM, Chau M, Wong PW, et al. A hybrid system for online detection of emotional distress [M]. Berlin Heidelberg: Springer, 2012: 73-80.
- [26] Pestian JP, Matykiewicz P, Linn-Gust M, et al. Sentiment analysis of suicide notes: A shared task [J]. *Biomed Inform Insights*, 2012, 5 Suppl 1: 3-16.
- [27] Pienaar J, Rothmann S, van De Vijver FJR. Occupational stress, personality traits, coping strategies, and suicide ideation in the South African Police Service [J]. *CJB*, 2007, 34(2): 246-258.
- [28] Lerner DMS, Henke RM. What does research tell us about depression, job performance, and work productivity? [J]. *J Occup Environ Med*, 2008, 50(4): 401-410.

(收稿日期: 2014-11-14)

(本文编辑: 张林东)