基于文本信息的自杀倾向检测

王宗杰[1,2],彭艳兵[2],姚方来[2]
(1.武汉邮电科学研究院 湖北 武汉 430000;2.南京烽火天地通信科技有限公司 江苏 南京 210000)

摘要: 近年来互联网的快速发展使我国网民人数急剧增加,部分网民通过微博、论坛、贴吧、bbs等应用程序使用文字、音频、视频、图片等发布自杀倾向信息。本文提出一种基于网民发言信息的自杀倾向行为检测技术,利用自杀倾向用户在网络上发布的文本信息,使用TF-IDF算法分析出自杀倾向的关键词,通过这些关键词在网络上爬取命中关键词的文本,利用机器学习算法将这些文本信息中具有自杀倾向的用户检测出来。分别对比SVM、KNN、NB、LR、DT、RF模型效果,实验结果表明RF模型效果最好精确率达85.1%。研究发现具有自杀倾向的用户敏感发言时间等信息对于检测自杀倾向也有很好的效果,通过加入这些指标,再次进行实验对比发现,自杀倾向检测的精确率达到了86.8%。

关键词: 文本信息; 自杀倾向; 机器学习; TF-IDF; RF
中图分类号: TP391 **文献标识码:** A **文章编号:** 1674-6236(2020)18-0030-04
DOI: 10.14022/j.issn1674-6236.2020.18.007

Suicidal tendency detection based on text information

WANG Zong-jie[1,2], PENG Yan-bing[2], YAO Fang-lai[2]
(1. *Wuhan Institute of Posts and Telecommunications*, *Wuhan* 430000, *China*; 2. *Nanjing Fenghuo World Communication Technology Co., Ltd.*, *Nanjing* 210000, *China*)

Abstract: with the rapid development of the Internet in recent years, the number of netizens in China has increased dramatically. Some netizens use text, audio, video and pictures to send messages of suicide intention through microblog, forum, post bar, BBS and other applications. This paper proposes a suicidal behavior detection based on netizens speech information technology, the use of suicide user text information released on the Internet, use the TF-IDF algorithm analysis suicidal keywords, through these keywords crawl in the network of the key words of the text, using machine learning algorithms to the text information is suicidal detected by the user. Comparing the model effects of SVM, KNN, NB, LR, DT and RF, the experimental results show that the RF model has the best accuracy rate of 85.1%. The study found that users with suicidal tendencies, such as sensitive speech time, also had a good effect on detecting suicidal tendencies. By adding these indicators and comparing experiments again, it was found that the accurate rate of detecting suicidal tendencies reached 86.8%.

Key words: text information; suicidal tendency; machine learning; TF-IDF; RF

随着互联网的快速发展,越来越多的人通过互联网来获取信息。部分互联网用户在社交网络上发布自杀倾向信息[1]并存在邀约自杀和直播自杀等行为[2-4]。社交网络上信息的传播很快很广对用户产生很大影响[5-6],自杀倾向与其它心理问题之间存在联系[7],文献[8]使用机器学习的方法检测了抑郁,文献[9]使用青少年的电子健康记录数据来预测是否有自杀企图,文献[10]使用文本信息检测了抑郁倾向。

从国内外研究来看,文献[11-13]使用了心理词汇和语言特征研究了微博用户的自杀倾向,但是没有结合敏感发言时间。文献[14-15]使用了利用机器

收稿日期: 2019-12-10 **稿件编号:** 201912091
作者简介: 王宗杰(1993—),男,安徽滁州人,硕士研究生。研究方向:机器学习。

学习算法进行检测自杀,但数据来源是问卷调查没有对发言信息进行检测。文献[16]使用私信的方式识别有自杀倾向的用户,识别用户数量有限不能大量检测。本文在前人研究的基础上,利用随机森林(RandonForst,RF)模型对社交网络上的自杀倾向用户进行检测,实验结果表明该方法能很好的检测出具有自杀倾向的用户。

1　分析过程

1.1　关键词提取

本次研究采用常用的基于离散词袋的词频-逆文本频率(Term Frequency-Inverse Document Frequency,TF-IDF)算法,根据已发现自杀倾向用户发布的文本信息,从这些文本信息中提取关键词。

基于TF-IDF算法计算关键词权重的公式如公式(1)所示。

$$W_i = tf_i \times \log\left(\frac{N}{df_i + 1}\right) \tag{1}$$

其中tf_i是表示分词后关键词i在文本中出现的频率,N是表示语料库中的文档总数,df_i是表示总的文档中包含该关键词i的文档个数。

结合人工核查最终筛选了24个有典型特征的关键词。具体关键词如表1所示。

表1　自杀倾向关键词表

关键词
好想死、安乐死、不想活、没人理解、烧炭
痛苦、活的好累、好绝望、抑郁、我完蛋了、彼岸花
接受不了、上吊、生不如死、约走、zs、聊4
自杀、约死、活不下去、欺凌、s友、我完了、厌世

1.2　人工指标构建

定义用户在夜晚11点到次日凌晨5点之间发言中包含自杀倾向关键词的时间称为敏感发言时间。研究发现自杀倾向用户与非自杀倾向用户之间的敏感发言时间、发送的平均文本长度和敏感发言条数等信息都有不同,通过前期分析总结的自杀倾向用户特征,初步构建以下5个指标具体信息如表2所示。

表2　人工构建指标表

指标名	说明
Diff_Key	命中的不同自杀关键词个数
Sum_Key	命中自杀关键词的总个数
Avg_Content_Length	平均发言长度:发言文本长度之和/总发言次数
Sen_Time	敏感发言时间:命中自杀关键词的发言时间
Day_count	发言天数

1.3　文本的向量化表示

对于每一个用户的文本数据需要转换成向量,本文使用的文本表示模型是向量空间模型(Vector Space Model,VSM)。把对文本内容的处理简化为向量空间中的向量运算,并且它以空间上的相似度表达语义的相似度,直观易懂。当文档被表示为文档空间的向量,就可以通过计算向量之间的相似性来度量文档间的相似性。文本处理中最常用的相似性度量方式是余弦距离。对于向量空间模型需要文档、特征词和权重。

1)文档:通常是文章中具有一定规模的字符串。文档通常我们也叫文本;

2)特征项:是VSM中最小的不可分的语言单元,可以是字、词、词组、短语等。一个文档内容可以被看成是它含有的特征项的集合。表示为一个向量:$D(t1,t2,...,tn)$,其中tk是特征项;

3)特征项权重:对于含有n个特征项的文档$D(t1,t2,..,tn)$,每一个特征项tk都依据一定的原则被赋予了一个权重wk,表示该特征相在文档中的重要程度。这样一个文档D可用它含有的特征项及其特征项所对应的权重所表示:$D(t1=w1,t2=w2,...,tn=wn)$,简记为$D(w1,w2,...,wn)$,其中wk就是特征项tk的权重。

在本文研究中,文档是某一用户在一段时间内命中关键词的发言信息,特征项是自杀倾向关键词和人工构建的指标,特征项权重是关键词出现的次数,对于关键词来说就是词频 对于人工构建的指标是计算的结果。具体实现框图如图1所示。

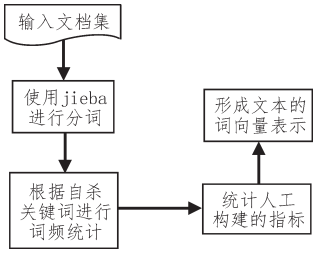

图1　VSM实现框图

1.4　算法的原理

本文使用的是基于分类和回归树(Classification and Regression Tree)算法构建的RF,采用随机抽样的方式生成随机样本,从而产生多个训练集,并在每个训练集上构建分类器,最终的结果由多个分类器结果投票产生。

对于数据集D,假设有K个分类,样本属于K类

的概率为P_K，则此时概率分布的基尼指数如式(2)所示。

$$Gini(p)=\sum_{K=1}^{K}P_K(1-P_K) \tag{2}$$

根据特征A将数据集划分成子数据集D_1，D_2，基尼指数如式(3)所示。

$$Gini(D,A)=\frac{D_1}{D}Gini(D_1)+\frac{D_2}{D}Gini(D_2) \tag{3}$$

*Gini*指数越小，对应的特征即为当前的最优特征，就是决策树的根节点，后面的叶节点以此类推。

1.5 分析流程图

根据前期获取的自杀倾向样本提取关键词，使用关键词构成关键词向量结合人工构建的量化指标组成整体的量化指标。利用关键词爬取互联网上数据，对训练集进行人工打标。根据对比各种算法模型效果选择最优模型。然后对爬取的数据进行预测，推荐出自杀倾向重点用户。具体流程图如图2所示。

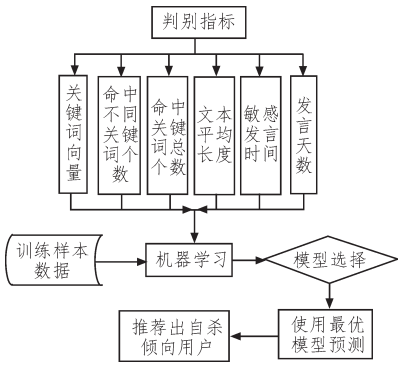

图2 自杀倾向分析流程图

2 实验结果及分析

2.1 实验数据

本次实验数据是用微博、贴吧、论坛和bbs上利用自杀倾向关键词爬取的数据，并对数据进行人工打标。用于模型训练的数据集是432个，用于实验预测的数据集是118个。

2.2 实验设置

为了使对比效果更好，使用相同的数据集进行训练和测试。本文采用十折交叉验证的方法(将训练集分成10份，轮流将其中的9份作为训练数据，1份作为测试数据)，目的是为了避免测试一次带来的较大误差。对训练样本进行10次训练测试，分别对比RF算法、K邻近(K-Nearest Neighbor，KNN)算法、逻辑回归(Logistic Regression，LR)算法、朴素贝叶斯(Naive Bayes，NB)算法、支持向量机(Support Vector Machine，SVM)算法和决策树(Decision Tree，DT)算法的模型效果。参数设置都是模型默认的参数，首先使用的指标都是自杀倾向关键词指标，然后随机森林算法使用不同指标进行对比，评价指标采用的都是模型的精确度，最后用训练好的RF模型进行预测。

2.3 结果分析

不同模型的精度的对比如表1所示，RF模型不同指标的精确率对照表如表2所示。

表1 不同模型的精确度的对比

算法名称	精确度
RF	0.851
KNN	0.793
LR	0.828
NB	0.773
SVM	0.642
DT	0.828

表2 RF模型不同指标的精确率对照表

指标	精确率
关键词指标	0.851
关键词指标+人工构建指标	0.868

用训练好的RF模型对118个已经打完标签的数据集进行预测，实验结果表明：能正确划分的用户为110个，预测结果和训练结果基本吻合。预测实验结果数据如表3所示。

表3 预测结果表

	真实自杀倾向	真实非自杀倾向
预测自杀倾向	27	5
预测非自杀倾向	3	83

2.4 评价指标

对于给定的数据集，假设*TP*是正确分类到正类的数量，*FP*是错误分类到正类的数量，精确率(Precision)的计算公式如式(4)所示。

$$P=\frac{TP}{TP+FP} \tag{4}$$

其中P表示精确率

对于给定的数据集，假设*FN*是错误分到负类的数量，召回率(Recall)计算公式如式(5)所示。

$$R=\frac{TP}{TP+FN} \tag{5}$$

其中R表示召回率

由精确率和召回率可知，F-measure的计算公式如式(6)所示。

$$F1=\frac{2*P*R}{P+R} \quad (6)$$

1)从训练测试实验结果可知在相同数据集下RF模型的精确率最高,在加入人工构建的指标后RF模型训练测试的精确度有了提高。

2)从预测结果可知该模型预测的精确率是90%,召回率是84.4%,$f1$值是87.1%。精确率与模型训练测试结果相当,召回率与$f1$值较高说明模型对于具有自杀倾向用户有很好检测的效果。

3 结　论

本文提出的基于文本信息的自杀倾向检测的技术,完成了对社交网络上具有自杀倾向行为用户的检测。实验结果表明使用RF模型检测自杀倾向行为精确率较高且预测结果较好。可以及时快速检测出在互联网上发布自杀倾向言论的用户,对于这些自杀倾向用户进行及时的预警和干预阻止其自杀。但是本文仍然存在不足,对于社交网络上关于自杀倾向用户发布的图片、声音和视频等信息没有涉及需要后期深入研究。

参考文献:

[1] 高一虹, 孟玲,高一虹,等. 自杀倾向的话语表述——大学生“走饭”微博分析[J]. 外语与外语教学, 2019, 304(1):47-59.

[2] 许宏, 李洁. 网络相约自杀的现状研究[J]. 精神医学杂志, 2015(2):79-81.

[3] 罗晓东, 邹桥. 网络“约死”:青年群体的自杀戏仿化与主体危机[J]. 天府新论, 2018,203(5):73-83.

[4] 李昂, 黄潇潇, 郝碧波, 等. 公众对社交媒体直播自杀企图现象的态度:基于新浪微博大数据的研究[C].第二十届全国心理学学术会议--心理学与国民心理健康,2017.

[5] 于婉琳. 社交网络对大学生自杀行为的影响及对策研究[D].北京:北京邮电大学,2017.

[6] 黄强,罗继锋,吴志艳. 熟人社交网络平台上信息传播影响指标及其效度研究[J].上海管理科学, 2019(3):35-47.

[7] 段彩彬, 周会, 张冰. 大学生自杀态度与焦虑、抑郁的关系研究[J]. 吉林省教育学院学报, 2016, 423(3):182-184.

[8] 邱家洪. 基于文本挖掘的社交网络抑郁用户检测[D]. 南昌:江西财经大学,2018.

[9] Bhat H S, Goldman-Mellor S J. Predicting Adolescent Suicide Attempts with Neural Networks [J]. 2017(2):3-10.

[10]Trotzek M , Koitka S , Friedrich C M . Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences[J]. IEEE Transactions on Knowledge and Data Engineering, 2018(1):1-8.

[11]Huang X , Lei Z , Liu T , et al. Detecting suicidal ideation in chinese microblogs with psychological lexicons[C]. 2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops. IEEE, 2014.

[12]Zhang L , Huang X , Liu T , et al. Using linguistic features to estimate suicide probability of chinese microblog users[C]. International Conference on Human Centered Computing. Springer International Publishing, 2014.

[13]许立鹏,宋文爱. 基于中文微博语言特征的自杀意念检测[J]. 中北大学学报,2019(4):13-18.

[14]丁楠. 基于机器学习的大学生自杀风险预测与分析[J]. 现代电子技术,2017(21):99-101.

[15]周家智,攸佳宁,李校安,等. 基于机器学习模型预测青少年群体的潜在自杀风险[C]. 中国心理学会会议论文集,2018.

[16]刘兴云, 孙炳丽, 王雪菲,等. 基于微博大数据的自杀宣言主动识别与及时干预[C]. 第二十届全国心理学学术会议--心理学与国民心理健康, 2017.