

文献分享

Title: Finding People with Emotional Distress in Online Social Media: A Design Combining Machine Learning and Rule-based Classification

Source: MIS Quarterly Vol. 44 No. 2 pp. 933-955/June 2020

Author: Michael Chau; Tim M. H. Li; Paul W. C. Wong; Paul S. F. Yip (from HONG KONG)
Jennifer J. Xu; Hsinchun Chen (from U.S.A)

Keywords: Social media, emotional distress, suicide research, design science, classification

一、研究背景

1, 情绪痛苦是现代社会普遍存在的一个复杂的社会和公共健康问题, 是自杀行为的一个强大的风险因素, 早期发现高危人群是预防自杀行为的关键(Turecki et al. 2016)。

全世界9.2%的人一生至少有过1次自杀的想法, 2%的人在过去的12个月里有过自杀想法 (Borges et al. 2010)

每年约有80.4万人结束自己的生命 (世界卫生组织 2014年)

2, 人们在网上留言表示情绪痛苦, 甚至是自杀遗书, 是一种正在形成的趋势(Ruder et al. 2011)。互联网上的内容, 特别是在线书写的叙事和日记, 在理解人们的情绪痛苦和自杀行为方面具有很大的潜力(Cheng et al. 2015; Hessler等2003年;Huang et al. 2007)。

香港约30%的自杀学生曾在社交媒体上表达过自杀意图(香港教育局 2016年)

3, 有鉴于此, 一些非政府组织(NGOs)开始积极在社交媒体上搜索这些痛苦和消极的自我表达, 以识别潜在的严重抑郁症患者, 以便提供帮助和后续服务。但依赖于社交媒体搜索引擎的简单关键字搜索是劳动密集型和时间无效的, 很容易混淆。

例如: 报道自杀事件的新闻文章

二、研究目的

研究目的：利用先进的技术来提高识别有情绪困扰的人的时间和成本效率。

KAREN的系统：根据用户输入的搜索关键词，得到多个博客搜索引擎的搜索结果并提取内容，结合机器学习分类(支持向量机和遗传算法)和基于规则的分类(专家提供的规则)，自动对内容结果进行分析和分类，分为有情绪困扰和没有情绪困扰，并聚合成最终的分类决策。

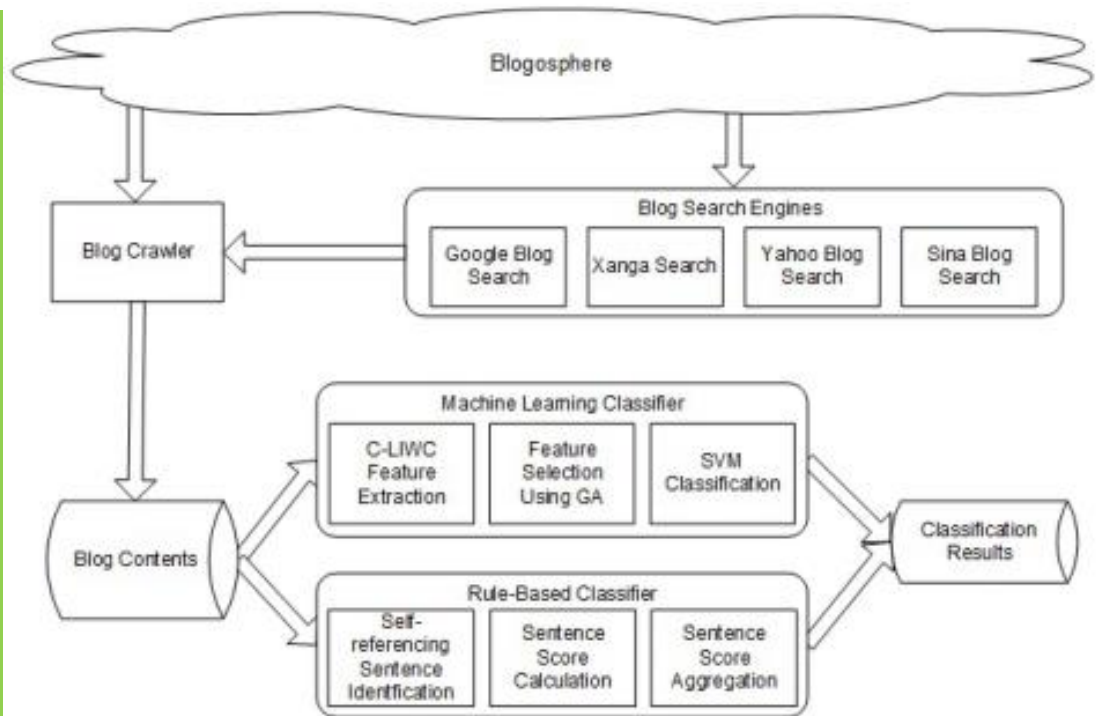


Figure 1. System Architecture

问题思考

为什么采用机器学习和规则分类相结合？

只有在分析整个文档的上下文时，才能识别一些表现出情绪困扰的表达，这对于SVM来说是不可能的，因为它不考虑文档中的单词顺序。

而规则分类可以将人类的判断纳入分类过程，执行句子级和段落级的分析，在其规则中合并上下文特定的启发式。

因此作者认为两者结合可以相互补充，获得更好的性能。

尽管基于规则的方法已被用于情绪分析和情绪检测研究（例如，Hutto和Gilbert，2014；Neviarouskaya等人，2010年，2011年；Wu等人，2006年），但它们尚未被用于情绪困扰的分类。

四、机器学习分类

1, 文本分词。使用中国科学院开发的**ICTCLAS**中文分割工具。



这是一种流行的工具，在许多之前的研究中都使用过(Zeng et al 2011;Zhang 等, 2003)。

2, 特征提取。采用了LIWC的中文版本C-LIWC。类别单词列表中所有单词频率计数的总数用于计算71个类别中每个类别的特征值 (f_i)，并将文档d长度作为第72个特征。从而为每个文档d创建了一个包含72个值的向量。

C-LIWC(Huang et al 2012)提供了积极或消极情绪、自我参照、因果词等71个词类，用于情感和认知词汇分析。许多研究表明，通过观察单词所属的LIWC类别，可以预测人们在写作中使用词汇时的心理健康状况(Pennebaker 2003)。

$$f_i = \sum_{\text{all words } w \text{ in category } i} \frac{\text{frequency}_w}{\text{total number of words in document } d}$$



四、机器学习分类

3, 特征选择。由于对所有可能的特征子集进行穷举搜索是不可行的，因此可以在特征选择中使用随机、基于人群的启发式搜索技术，如遗传算法（GA）（Fang等人，2007；Oreski和Oreski 2014；Yang和Honavar 1998）。



4, SVM分类。

为什么使用支持向量机分类器？

（1）它在各种文本分类任务中取得了最佳性能(Abbasi et al. 2008;Yang and Liu 1999)。

（2）显示情绪痛苦的博客比例比常规博客小得多，SVM是一种在积极训练实例数量较少时表现更好的分类器（Yang 和 Liu 1999）。

先前的研究表明，特征选择可以显著提高机器学习文本分类器的性能（Saad 2014）。

基于遗传算法的特征子集选择方法基于达尔文的自然选择理论，根据“适者生存”的原则搜索最优子集。该算法从随机选择一定数量的特征子集开始，这些特征子集代表一组潜在的解。每个子集都用适应度函数进行评估。然后通过选择具有较高平均适应度得分的子集来形成新的群体。新群体的一些子集经历了转换，如交叉和突变。经过多次迭代后，GA从所有总体中选择最佳特征子集。

五、规则分类

1, 创建情绪痛苦词典。由熟悉网络话语术语的专业人士对博客内容进行手动检查而构建。词典中的单词被分为十组：自我参照、积极情绪、消极情绪、危险因素、自杀、时间、否定、休闲、参照和感激表达。

以前的研究中也使用了类似的词汇创造方法，并显示了令人鼓舞的结果（Abbasi 和 Chen 2007；Subasic和Huettnner 2000）。

Table 1. Examples and Number of Words in the Ten Lexical Groups.

Group	Number of Words	Examples
Self-Reference	9	自己 (self), 在下 (I), 小弟 (I), 本人 (myself), 我 (I)
Positive Emotion	34	窩心 (heartwarming), 雀躍 (joyful), 暢快 (carefree), 驚喜 (pleasant surprise), 酷愛 (love)
Negative Emotion	56	心痛 (sad), 失望 (disappointed), 失落 (down), 沮喪 (frustrated), 焦慮 (anxious)
Risk Factors	15	分手 (separation), 離婚 (divorce), 疾病 (illness), 貧窮 (poverty), 比人厄 (cheated)
Suicide Words	18	界手 (self-laceration), 跳樓 (jumping from a building), 燒炭 (charcoal burning), 食安眠藥 (taking sleeping pills), 永別 (part forever)
Time	16	今朝 (this morning), 每天 (every day), 昨晚 (last night), 聽日 (tomorrow), 宜家 (now)
Negation	39	唔 (not), 不 (no), 別 (don't), 否 (negative), 沒 (without), 非 (not)
Leisure	184	義工 (volunteer), 健身 (fitness), 煲劇 (watching TV drama), 動漫 (animation and comics), 旅行團 (tour)
References	108	本報訊 (news), 專訊 (special news), 參考資料 (references), 摘錄 (excerpts), 撰稿 (written)
Gratitude Expressions	11	共勉之 (encourage each other), 加油 (make effort), 鼓勵 (encourage), 感恩 (appreciate), 謝天謝地 (thank god)

五、规则分类

情绪痛苦包括不同情境下的多重影响和生活压力源。例如，与丧亲相关的情绪困扰会产生悲伤和紧张等影响(Chen et al 1999)，而与糖尿病相关的情绪困扰会产生恐惧和担忧等影响(Snoek et al 2000)。

此外，人们会谈论日常生活中发生的事件，而不是使用许多负面情绪的词汇，这些事件可能是他们情绪困扰的原因。

因此，除了分析消极和积极的情绪词汇外，研究人员还研究了其他与情绪困扰相关的词汇，如各种风险因素和自杀词汇，以及表示积极的幸福和态度的词汇。

与C-LIWC相比，该词典：

(1) 更精确，更适合于心理健康领域。因为C-LIWC的覆盖面广，包含与应用领域不太相关的类别和单词。

(2) 包含了博客作者在网上情感表达中实际使用的词汇，其中包括C-LIWC中没有的口语词汇。

五、规则分类

2, 自我参照句识别。

带有自我参照词的句子可以提供更多的线索来识别脱离行为，从而识别情绪痛苦。

心理语言学的研究表明，目前有抑郁症或自杀意念的人具有独特的语言风格，在写作中倾向于使用更多的自我参照词（例如，我、本人、我自己），这意味着强烈的自我导向（Li等人2014; Ramirez Esparza等人，2006年; Rude等人2004），甚至退出社会关系（Stirman 和 Pennebaker 2001）。

```
1. Inputs:
2.   s, a sentence
3.   lexicon, a lexicon of words divided into 10 groups
4. Output:
5.   score, the emotional distress score for sentence s
6. Procedure:
7.   score = 0
8.   if s contains (Self-reference)
9.     if s contains (Negative Emotion and not Negation) 焦虑
10.      or s contains (Positive Emotion and Negation) 不开心
11.      score = 1
12.      for each (Risk Factors or Suicide) in s
13.        if s contains (Time)
14.          score = score + 2
15.        else
16.          score = score + 1
17.      else if s contains (Positive Emotion and not Negation) 开心
18.        or s contains (Negative Emotion and Negation) 不焦虑
19.        score = -1
20.        for each (Leisure) in s
21.          score = score - 1
22.      else
23.        for each (References or Gratitude expressions) in s
24.          score = score - 1
25.      return score
```

1, 情绪

2, 事件

五、规则分类

3, 计算句子情绪痛苦得分。分数为正表示该句子表现出了情绪困扰, 而为零或负值则表示相反。

(1) 如果同时发现了积极和消极情绪的单词, 则该句子被归类为表达消极情绪。这是为了避免忽略可能的情绪痛苦。

(2) 即使句子分别包含多个消极或积极情绪词, 我们也只给出1或-1分。这是为了更多地关注与情绪痛苦和精神健康领域相关的词汇, 从而与标准的情绪分析方法区分开来。

```
1. Inputs:
2.   s, a sentence
3.   lexicon, a lexicon of words divided into 10 groups
4. Output:
5.   score, the emotional distress score for sentence s
6. Procedure:
7.   score = 0
8.   if s contains (Self-reference)
9.     if s contains (Negative Emotion and not Negation) 焦虑
10.      or s contains (Positive Emotion and Negation) 不开心
11.      score = 1
12.      for each (Risk Factors or Suicide) in s
13.        if s contains (Time)
14.          score = score + 2
15.        else
16.          score = score + 1
17.      else if s contains (Positive Emotion and not Negation) 开心
18.        or s contains (Negative Emotion and Negation) 不焦虑
19.        score = -1
20.        for each (Leisure) in s
21.          score = score - 1
22.      else
23.        for each (References or Gratitude expressions) in s
24.          score = score - 1
25.      return score
```

1, 情绪

2, 事件

五、规则分类

(3) 对于无消极情绪的句子，情绪痛苦评分会随着休闲词的出现而降低。在积极心理学中，休闲是整体幸福感和唤起幸福感的核心成分(Newman2014;Zawadzki et al. 2015)。

(4) 在脱离理论下，参考其他来源向他人提供意见或传递信息的人患抑郁症的风险较低(Stirman&Pennebaker 2001)。给予他人感谢和鼓励的话，这被证明可以改善人们的幸福和缓解抑郁，也证明了作者的积极态度(Boiler2013;Lyubomirsky&Layous2013)。

```
1. Inputs:
2.   s, a sentence
3.   lexicon, a lexicon of words divided into 10 groups
4. Output:
5.   score, the emotional distress score for sentence s
6. Procedure:
7.   score = 0
8.   if s contains (Self-reference)
9.     if s contains (Negative Emotion and not Negation)
10.      or s contains (Positive Emotion and Negation)
11.       score = 1
12.       for each (Risk Factors or Suicide) in s
13.         if s contains (Time)
14.           score = score + 2
15.         else
16.           score = score + 1
17.       else if s contains (Positive Emotion and not Negation)
18.         or s contains (Negative Emotion and Negation)
19.           score = -1
20.           for each (Leisure) in s
21.             score = score - 1
22.     else
23.       for each (References or Gratitude expressions) in s
24.         score = score - 1
25.   return score
```

焦虑
不开心

1, 情绪

2, 事件

开心

不焦虑

五、规则分类

4, 计算文档得分（句子分数聚合）。

由于整个文档的情绪波动可能很复杂，文档中间的一些分数可能没有意义，甚至可能令人困惑。因此，聚合集中在文档开头和结尾的主题得分上。

已有几种分割方法用于在全文文档中查找子主题，(例如，将句子分组在块中，并将内容划分为连贯的单元；赫斯特1997年)。

使用文档中自引用语句的第一个和最后一个块来获得最终的预测分数，其中一个块被定义为具有相同极性(+/-)的连续语句集。

由于文档中大量的极性变化代表作者情绪的不一致(即波动和不稳定)，也使用这个数字来计算最终的分数。

Final score(d) = 第一个区块的得分 + 最后一个区块的得分 + 文档中极性转换的次数

如果最终得分为正，则被归为有情绪痛苦。

六、结果聚合

为了不遗漏任何情绪困扰案例，如果一个博客被**任何一个**分类器检测到表现出情绪痛苦，它将被归类为**有情绪痛苦**。

七、性能评估

Table 2. Sources of Blog Data Used in the Experiment.

Data Source	Source	Search Tool Used	No. of Blogs
#1	Hong Kong Federation of Youth Groups	Yahoo blog search	180
#2	Samaritan Befrienders Hong Kong	Various	239
#3	Blog site browsing	Nil; browsing only	150
#4	Google blog searching	Google blog search	235
Total number of blogs in the sample (<i>n</i>)			804

2名心理学专家评估 274个包含了情绪困扰的内容, 530个包含了不包含困扰的内容

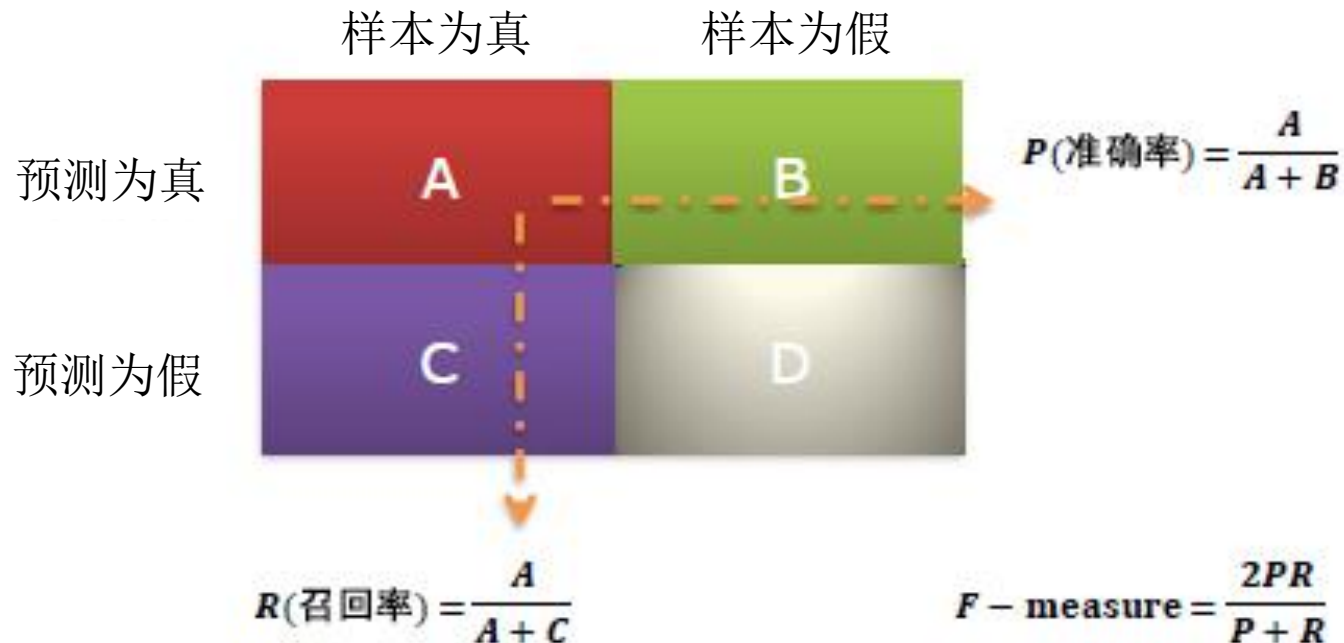
Table 3. Classification Performance

Model	No. of Features	Precision	Recall	F-measure
Model 1: SVM(n-grams)	17,560	0.5732	0.6715	0.6185
Model 2: SVM(C-LIWC)	72	0.7404	0.6350	0.6837
Model 3: SVM(C-LIWC + GA)	38	0.7542	0.6606	0.7043
Model 4: Rule-Based	10	0.4705	0.8723	0.6113
Model 5: SVM(C-LIWC) + Rule-Based	72	0.6171	0.8175	0.7033
Model 6: SVM(C-LIWC + CB) + Rule-Based	17	0.6123	0.8358	0.7068
Model 7: SVM(C-LIWC + RF) + Rule-Based	20	0.6202	0.8285	0.7094
Proposed Model 8: SVM(C-LIWC + GA) + Rule-Based	38	0.6287	0.8467	0.7216

Note: SVM: support vector machine; n-grams: n-grams-based features; C-LIWC: C-LIWC-based category features; CB: correlation-based feature selection; RF: recursive feature elimination; GA: genetic algorithm.

延伸

- 1, **召回率R**是针对原来的样本而言的, 表示的是样本中的真例有多少被预测正确。
- 2, **精确率P**是针对预测结果而言的, 表示的是预测为真的样本中有多少是真正的真样本。
- 3, 但P和R指标有时候会出现矛盾的情况。
- 4, **F值**是对以上两个指标的综合, 可作为折中的评价指标。



八、研究不足及未来工作

研究不足：

没有调查博主的人口统计学特征，如性别。研究发现，在计算机介导的交流中，男性和女性使用不同的情感表达(Thelwall et al 2010)。本研究在分类过程中没有考虑性别差异。

未来工作：

1，将分析博主，而不是单篇文章。例如，对博主过去三个月的帖子、帖子的评论和互动进行分析，来预测他们的情绪波动，以更好地理解博主的想法(Campbell and Pennebaker 2003; Chau和Xu 2012)。

2，可以将表情符号、插入式表达等常用的表达思想和情感的符号纳入分类方法。