

doi: 10.3969/j.issn.0490-6756.2019.06.009

基于短语级情感分析的不良信息检测方法

明弋洋, 刘晓洁

(四川大学网络空间安全学院, 成都 610065)

摘要: 针对基于关键词字符匹配和粗粒度情感分析方法的传统不良信息检测方法准确率低的问题, 提出一种基于短语级情感分析的不良信息检测方法. 该方法制定语法规则来提取敏感词所在短语, 结合二次分类的情感词典, 通过分析短语的情感倾向来判断表达者对敏感关键词的情感倾向, 从而判定内容的敏感性. 本文方法克服了字符串匹配方法的缺点, 如: 传统的字符串匹配方法忽视敏感词上下文, 从而导致大量误报的缺点, 及粗粒度情感分析方法由于无法准确定位情感对象, 只能以文本整体情感倾向代表表达者主观情感, 导致不能精准分析与敏感词相关的情感倾向及敏感性. 实验表明, 本文方法与传统方法相比, 准确率有较大的提升.

关键词: 不良信息; 语法规则; 短语提取; 情感词典; 情感分析

中图分类号: TP301 **文献标识码:** A **文章编号:** 0490-6756(2019)06-1042-07

Sensitive information detection based on phrase-level sentiment analysis

MING Yi-Yang, LIU Xiao-Jie

(College of Cybersecurity, Sichuan University, Chengdu 610065, China)

Abstract: String matching method and coarse-grained sentiment analysis method are often used in sensitive information detection, the accuracy, however, is rather low. To mitigate this issue, this paper proposed a method based on phrase-level sentiment analysis. This method takes advantage of a rule set which is used to extract phrases concerning the sensitive word, by analyzing the sentiment orientation of relative phrases, it is possible to determine the sentiment orientation of the expresser towards the sensitive word, thus determine the sensitivity of the information. This method takes the context of the sensitive word into consideration and is able to extract sentiment orientation towards the sensitive word instead of the whole text, which is missing from string matching method and coarse-grained method respectively. Experimental results suggest that the accuracy is considerably increased compared to string matching and coarse-grained sentiment analysis method.

Keywords: Sensitive information; Semantic rules; Phrase extraction; Sentiment dictionary; Sentiment analysis

1 引言

我国网民数量众多, 互联网环境复杂, 存在少

部分人员利用网络传播暴力、反动、色情及邪教等不良信息, 如何检出这些信息是一个重要问题. 当前, 对不良信息的检测研究主要集中在对敏感词的

收稿日期: 2019-04-19

基金项目: 国家重点研发计划(2016YFB0800604, 2016YFB0800605); 国家自然科学基金项目(61572334, U1736212); 四川省重点研发项目(2018GZ0183)

作者简介: 明弋洋(1995-), 男, 四川德阳人, 硕士生, 研究方向为网络与信息安全. E-mail: shurrik11@163.com

通讯作者: 刘晓洁. E-mail: liuxiaojie@scu.edu.cn

检测上,即基于敏感词库,如何高效、准确地检出敏感词本身或其变体形式.目前广泛采用的方法是基于字符串匹配通过建立敏感词库,对待检测文本进行检索,若其中含有敏感词,则标记为不良信息.这种方法实现简单,但仅能识别敏感词库中的词.对此,文献[1]提出一种基于决策树的过滤算法,该算法通过将敏感词按照拼音首字母分类,最终形成一个前缀树结构的决策树,将待检测文本视作信息流,查找决策树,最终输出文本中所含敏感词.文献[2]提出了一种识别敏感词变形体的方法,该方法结合了汉字的发音结构与特征,能够识别出中文敏感词的拼音、简称和拆分三种变形体形式.文献[1-2]的方法拓展了检测的范围,可以识别敏感词的部分变形形式,但由于仍是基于字符串匹配,检出对象都为“词”,缺少上下文信息的考虑,因此准确率较低.

为了将文本上下文纳入考虑,文献[3]使用情感分析来识别微博中的敏感信息.作者对微博的情感极性进行了度量,通过计算文本情感得分将文本情感倾向分为正向情感和负向情感两类,通过分析发现负向情感的微博内容具有较高的敏感性,由此建立了含有情感极性因素的敏感度模型,通过情感倾向计算敏感度,使用敏感度来判别微博的敏感性.文献[4]提出了一种 Laplace 平滑情感判定算法,通过细化文本语义分析规则计算文本情感值.文献[5]通过结合情感词典和卡方模型,将情感词的情感极性值与句中的否定词纳入考虑,提升了判定的准确率.以上通过计算文本情感得分或极性值来判断文本情感倾向方法的问题在于一段文本通常包含多个情感词语,这些情感词所修饰的对象互不相同,因此整段文本所展现的整体情感不一定代表表达者对敏感词的情感倾向,以文本为粒度进行的情感分析得出的情感倾向十分笼统,因此无法制定通用的告警策略,导致其准确率仍然较低.如:

例句 1) 支持法轮功学员研习法轮功;

例句 2) 谴责公安机关迫害法轮功.

在以上两个例句中,敏感词都为“法轮功”,从“支持”和“迫害”可以看出表达者对其情感倾向是正面的.但如果采用文献[3]和文献[4]的方法,例句 1) 的输出将是正面情感,例句 2) 的输出将是负面情感,然而两句包含的内容显然都应属于不良信息.粗粒度情感分析不能正确地提取表达者对“法轮功”的情感倾向,因此需要细化情感分析的粒度.细粒度情感分析的关键是对评论对象及其对应

情感词进行识别^[6].例如,在例句 2) 中,如果能提取出“迫害法轮功”,就可以仅对这一短语进行情感分析,从而得出表达者对“法轮功”的情感.目前广泛使用的从文本中提取短语的统计方法是 PMI (pointwise mutual information, 点互信息)^[7],计算公式如下.

$$PMI(x, y) = \log \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

其中, $PMI(x, y)$ 表示 x 和 y 的点互信息值, $p(x)$, $p(y)$ 分别表示 x 和 y 在文本中的出现概率, $p(x, y)$ 表示 x 和 y 共现的概率.点互信息描述了两个物体之间的相关性,点互信息值越大说明两个词语越有可能形成一个短语.但在不良信息的检测中,检测对象通常是网页中的文本,其规模远达不到统计规模,PMI 的正确性无法保证^[8].

基于上述问题,本文提出一种基于短语级情感分析的不良信息检测方法,该方法考虑了敏感词上下文信息,在基于词典的情感分析方法的基础上,将情感词典细分为主体正向、主体负向、受体正向、受体负向、褒义词和贬义词 6 类而不仅是正面情感和负面情感两类,同时将敏感词按照其性质归为正向敏感词、负向敏感词和超级敏感词三类,使用一种基于语法规则的短语提取策略,通过提取敏感词所在短语,分析其情感倾向,结合敏感词的归类作出告警决策.该方法将分析的粒度由文本细化到包含敏感词的上下文短语,检出对象由“敏感词”扩展到“不良信息”,有效地提高了准确率.最后通过在真实数据上的实验验证了该方法的正确性.

2 基于短语级情感分析的不良信息检测

在情感分析中,“情感”的概念被认为是表达者的主观观点、态度和情绪等,所以情感分析也被成为观点挖掘、评价抽取和主观分析^[9].按照分析粒度的不同,情感分析可分为粗粒度情感分析和细粒度情感分析,粗粒度情感分析只能判断文本的整体情感倾向,对于某产品来说,产品评论文本的整体情感倾向并不一定能和产品各个属性的情感倾向保持一致^[10].本文使用的短语级情感分析属于细粒度情感分析,目的是分析表达者对某敏感词的主观情感,而不是表达者的整体情感倾向.

由于 PMI 方法不适用于小规模文档,本文提出一种基于语法规则的短语提取策略来提取短语,并与二次分类后的情感词典结合,提出一种短语级

情感分析的方法. 方法总体流程如图 1.

目前广泛使用的情感词典有知网情感词典、台湾大学情感词典等. 这类情感词典仅简单地将情感词分为正面情感、负面情感两类,但在中文中,同一个词表达的情感可能因其使用方式不同而不同. 如“迫害”一词,在情感词典中属于负面情感,但在句子“A 迫害 B”中,表达者对 A 的主观情感是负面的,而对 B 的主观情感是正面的(包含一种同情的感情),因此不能简单地将“迫害”分为负面情感词,而应视实际情况分类. 对此,本文将情感词典进行了二次分类,进一步细化了情感词的情感倾向,然后使用二次分类后的情感词典,制定了语法规则集用于提取短语,同时将敏感词按照其性质归类. 在检测阶段,首先进行文本预处理,将文本按标点符号分句后,对每个短句进行分词、去除停用词处理. 然后对每个词进行标注,标明该词属于情感词,敏感词,还是其他词. 若属于情感词,同时标注其所属类别. 接下来逐一一对分词后的短句进行短语级情感分析,使用语法规则提取敏感词所在短语并输出其情感倾向,最后结合敏感词所归类别进行告警决策.

3 情感词典的二次分类

本文使用的情感词典衍生自知网情感词典和台湾大学情感词典,将两个情感词典合并、去重后,从结果中人工挑选出具有明显情感倾向的词语,将其二次分类为 6 大类:主体正向、主体负向、受体正向、受体负向、褒义和贬义. 前四类主要包含动词,后两类主要包含形容词,“主体”、“客体”泛指动词的主语和宾语,“正向”、“负向”表示该词的情感倾向,如“主体正向”代表该词对主体的情感是正面的. 其中, (1) 主体正向(负向)类包括动作的主语

一般为正面(反面)形象的词语,或表达者对动作的主语带有明显正面(负面)情感的词语;(2) 受体正向(负向)类包括动作一般表达了对宾语正面(负面)情感的词语,或表达者对动作的宾语带有明显正面(负面)情感的词语;(3) 褒义、贬义分别包含褒义词和贬义词.

分类后得到的情感词典数量如表 1 和表 2 列出了部分例词. 除褒义、贬义外,其余四类所包含的词语允许交叉,如“迫害”、“歼灭”二词都属于两个不同的类别.

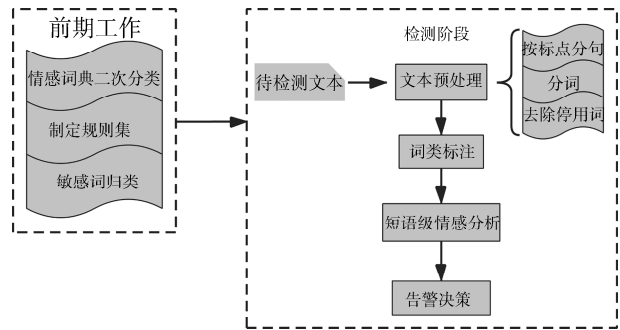


图 1 整体流程

Fig. 1 Processing workflow

表 1 情感词种类及数量

Tab. 1 Types and quantities of sentiment dictionaries

类别	数量/个
主体正向	1 704
主体负向	2 381
受体正向	2 647
受体负向	3 476
褒义词	2 258
贬义词	2 839
总计	15 305

表 2 情感词库部分例词

Tab. 2 Examples from sentiment dictionaries

类别	主体正向	主体负向	受体正向	受体负向	褒义词	贬义词
例词	歼灭、惩治、	侵略、剥削、	赞赏、关爱、	厌恶、憎恨、	美好、伟大、	残忍、傲慢、
	指引、铲除、	污蔑、陷害、	褒奖、感激、	讨厌、抵制、	无畏、严谨、	龌龊、不堪、
	救助、伸张、	推卸、迫害、	迫害、污蔑、	怀疑、打倒、	虔诚、神圣、	蛮横、虚伪、
	牺牲、奉献	勾结、行贿	抹黑、诬陷	歼灭、痛斥	忠实、优秀	邪恶、恶毒

4 短语级情感分析

4.1 基于语法规则的短语提取

在二次分类情感词典的基础上,可以根据其中

的情感词类别组合出人们习惯使用的表达不同情感的词语搭配. 如在例句 2) 中,“迫害”一词属于受体正向类,“法轮功”是其受体且为待检测敏感词对象,则二者可以组成“受体正向+敏感词”短语,该

短语明显地表达了对敏感词的情感倾向。

本文根据中文语法,结合经验制定了一个语法规则集用以提取不同类别的词语组合形成的短语.该规则集的制定主要参照了常见的中文短语构成方式与表达习惯,如主谓短语(中心词与谓语动词构成)、动宾短语(动词与宾语中心词构成)、偏正短语(形容词与中心词构成),与简单的被动句等.规则集中共包含14条语法规则.部分规则如下:

Rule1. $\{pos_{obj}\} + \{nword\} \rightarrow \text{DISCARD}$

Rule2. $\{nword\} + \{neg_{sub}\} \rightarrow \text{DISCARD}$

Rule3. $\{pos_{obj}\} + \{word\} \rightarrow \{pos_{obj}_P\}$

Rule4. $\{word\} + \{neg_{sub}\} \rightarrow \{neg_{sub}_P\}$

Rule5. $\{word\} + \{pos_{adj}\} \rightarrow \{pos_{sub}_P\}$

Rule6.

$\{word\} + \{passive\} + \{neg_{obj}\} \rightarrow \{neg_{obj}_P\}$

这部分规则包含了规则集中出现的所有标签,各标签含义如下: pos 表示正向情感; neg 表示负向情感; sub 表示主体; obj 表示客体; adj 表示形容词; $passive$ 表示被动词; $word$ 表示敏感词; $nword$ 表示其他词.因此, pos_{obj} 表示二次分类后的情感词典中对客体正向类别; neg_{adj} 表示贬义词,类似地可得其他类别的含义; pos_{obj}_P 表示由一个对客体正向的情感词与一个敏感词形成的短语,类似地可得其他短语的含义; DISCARD 表示丢弃.每条规则中词语的顺序不可调换.

由于中文表达方式的多样性,本文所使用的规则集并不能完全覆盖所有的表达方式与句型,但覆盖了常见的较为正常、简单的表达方式,如rule3, rule4分别表示了简单句中的动宾、主谓结构, rule6表示了一个简单的被动句.这些句型的特点很容易被总结为一个通用的模式,因此也适合被归纳为语法规则.复杂句型则很难具备这一特点,如句子“A对B进行了迫害”,其表达的含义与“A迫害B”、“B被A迫害”相同,却远没有后两者易于归纳,因此没有纳入规则集中.除规则集中的14条规则外,另增一条出口规则,如遇规则集无法解析的复杂句型,使用出口规则,将内容交由专家审核.

在本文研究背景下,不存在文献[11]指出的隐藏属性问题,可以认为短语的评论对象就是敏感词.制定规则后,逐一调用规则解析待检测文本.

定义1 预处理后的文本 $t = \{t_0, t_1, \dots, t_n\}$, $t_i (0 \leq i \leq n)$ 表示组成文本的词语;规则集 $ruleset = \{r_0, r_1, \dots, r_m\}$, $r_j (0 \leq j \leq m)$ 表示每条规则, $m, n \in \mathbb{N}$;类别标注集 $tag = \{t_0: attr_0, t_1: attr_1, \dots, t_n: attr_n\}$, n 为 t 中元素个数, $t_i (0 \leq i \leq n)$ 对应 t 中的元素, $attr_i (0 \leq i \leq n)$ 是集合 $\{word, nword, pos_{sub}, pos_{obj}, neg_{sub}, neg_{obj}\}$ 的一个子集,代表 t_i 所属类别.为了说明,将所有规则表示为

$pattern \rightarrow result$

短语提取算法步骤如算法1.

算法1 短语提取算法

(1) 对于 $\exists r \in ruleset$, 若 $\exists t_i, t_j \in t, tag[t_i] + tag[t_j] = pattern$, 则将 t_i, t_j 替换为 $result$; 否则,若当前为第一次匹配,进入出口规则,结束解析;

(2) 若有规则匹配,重复步骤(1),否则进入步骤(3);

(3) 对于 $\forall t_i \in t$, 如果 $t_i \in \{pos_{obj}_P, neg_{sub}_P, pos_{sub}_P, neg_{obj}_P\}$, 则将 t_i 加入到新集合 res 中;

(4) 返回集合 res , 算法结束.

算法伪代码描述如下.

Parse function

1) function $P_{ARSE}(text, ruleset)$

2) $res \leftarrow \text{none}$

3) if no rule matches at beginning then

4) enter exit rule

5) else if parse done then

6) for each in text do

7) $res \leftarrow IsSubsetOf(\text{each}, \text{phrases})$

8) end for

9) return res

10) else

11) $text \leftarrow RuleParse(text, ruleset)$

12) $P_{ARSE}(text, ruleset)$

13) end if

14) end function

其中, $RuleParse$ 函数将逐条调用规则,完成算法1中步骤(1)描述的匹配与替换过程; $IsSubsetOf$ 函数完成步骤(3)描述的过程,将解析结果中的短语作为最终结果返回.递归退出的条件为当前解析不为第一轮解析,且已无规则可以继续匹配.

以例句2)为例,解析过程如图2.图2仅展示了部分规则的匹配结果(如“公安机关”与“迫害”也将与一条规则匹配),如果“公安机关”不属于敏感词,则其形成的短语都将被丢弃.最终解析的结果为对敏感词“法轮功”,形成一个 pos_{obj}_P 类型的短语并返回.

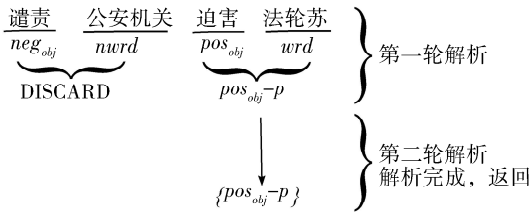


图 2 对例句 2 的解析
Fig. 2 Parse of example text 2

4.2 情感分类与告警决策

首先,为了使用基于规则的短语提取的结果,需要指定每一类短语对应的情感倾向,如表 3.

表 3 短语的情感倾向

Tab. 3 Sentimental orientation of phrases

短语	情感倾向
pos_{obj_P}	POS
neg_{obj_P}	NEG
pos_{sub_P}	POS
neg_{sub_P}	NEG

将表 2 与语法规则相结合,即可得到短语与其表达情感的对应关系,如:

Rule3. $\{pos_{obj}\} + \{word\} \rightarrow POS$

Rule4. $\{word\} + \{neg_{sub}\} \rightarrow NEG$

其中,POS 表示正面情感;NEG 表示负面情感.

其次,为了配合使用每种短语表示的情感倾向,要将敏感词库归类.本文收集了涉政、色情、暴恐、违禁品及邪教共五种敏感词,每种数量如表 4.

表 4 敏感词种类及数量

Tab. 4 Classes and quantities of sensitive words

种类	数量/个
涉政	750
色情	1 100
暴恐	17
违禁品	2 172
邪教	274
总计	4 313

本文方法的特点在于,由于检出对象由“敏感词”扩展到了“不良信息”,只有与敏感词相关的不良信息会产生告警,因此可以将原本不属于敏感词的词语添加进敏感词库而不会产生大量误报.如在涉政一类中就包含了国家机构和国家领导人.根据五种词的特性,将五种词库归为三类:正向敏感词、

负向敏感词和超级敏感词,如表 5.

表 5 敏感词归类

Tab. 5 Categories of sensitive words

种类 class	归类 category
涉政	负向
色情	超级
暴恐	正向
违禁品	超级
邪教	正向

将某种敏感词库 $class$ 归为某类 $category$ 的含义为“若表达者对 $class$ 种类的敏感词的主观情感为 $category$ 类,则属于不良信息”,例如涉政类属于负向敏感词的含义是“如果表达者对涉政类敏感词的主观情感是负面的,则属于不良信息”.超级敏感词库中的词本身就属于不良内容,因此不经过情感分析,仅做简单字符串匹配.将归类后的敏感词与情感分析产生的结果结合,情感分析与告警决策算法如算法 2(除超级敏感词外).

算法 2 情感分析与告警决策算法

(1) 对于短语提取的结果集合 res , $\forall P_n \in res$,根据表 2 中的映射关系,将 P_n 对应的情感倾向加入到集合 so_set 中;

(2) 处理 so_set 中的冲突,得到最终情感倾向 so ;

(3) 查询敏感词 $word$ 对应类别 cat ;

(4) 若 $cat = so$,告警;

(5) 算法结束.

“冲突”是指对同一个敏感词,不同规则匹配结果的情感倾向不同.当出现冲突时,根据文献[12]描述的观点词邻近原则,即表达观点的词语通常出现在被评论的主体附近,考察出现冲突的短语中的情感词与敏感词的距离,最终取与敏感词距离小的情感词所在短语表达的情感倾向.二者距离计算公式如下:

$$\text{dist}(x, y) = |\text{index}(x) - \text{index}(y)| \quad (2)$$

其中, $\text{dist}(x, y)$ 表示词 x, y 的距离; $\text{index}(x)$ 和 $\text{index}(y)$ 表示二者在分词后的短句中所处位置的下标,下标由 1 开始递增.如在例句 2) 中,“谴责”所处位置下标为 1,“公安机关”所处位置下标为 2,则二者距离为 1.

算法伪代码描述如下.

Alert Function

1) function $A_{\text{LERT}}(word, res)$

```

2)  cat ← CCATEGORY(word)
3)  so ← AANALYSE(res)
4)  if so = cat then
5)      return true
6)  else
7)      return false
8)  end if
9)  end function

```

其中, *Analyse* 函数完成算法 2 中步骤(1)和步骤(2)描述的过程, 得到短语的情感倾向 *so*; *Category* 函数用于获取敏感词 *word* 归属的类别 *cat*.

以图 2 的解析结果为例, 敏感词“法轮功”形成的短语 *pos_{obj}_P* 对应的情感倾向为正向, 而“法轮功”属于归类为正向的敏感词, 因而产生告警。

5 实验结果及分析

采用 Python 2.7 程序语言实现本文算法, 数据集使用搜狗实验室提供的全网新闻数据集, 从中选择了 128 篇新闻文档, 这些新闻文档内容正常, 但都包含敏感词; 另外从网络上下载了 42 篇包含反动、迷信和邪教等不良信息的文档. 在所有文档中, 共出现敏感词 5 820 处, 其中新闻文档中出现的 3 775 处敏感词属于正常信息, 不良文档中出现的 2 045 处敏感词属于不良信息.

使用关键字(字符串)匹配方法和文献[3]使用的粗粒度(文本级)情感分析方法作为对比, 定义准确率、召回率和精确率

$$P_{\text{precision}} = \frac{TP}{TP + FP} \quad (3)$$

$$P_{\text{recall}} = \frac{TP}{TP + FN} \quad (4)$$

$$P_{\text{accuracy}} = \frac{TP + TN}{FP + TP + FN + TN} \quad (5)$$

其中, *TP* 表示正确告警条数; *FP* 表示误报条数; *FN* 表示漏报条数; *TN* 表示正确未告警条数. 实验结果如下.

在关键字匹配方法中, 只要使用的字符串匹配算法正确, 即可识别出文本中出现的所有敏感词, 故其召回率为 100%. 与关键字匹配方法相比, 本文方法在损失少量召回率的情况下, 较大地提高了检测的准确率; 在准确率、召回率上较粗粒度方法也有较大提升. 另外, 虽然本文制定的规则集数量较少, 但在所选数据集上已经取得了较高的召回率.

表 6 实验结果

Tab. 6 Experimental results

方法	准确率/%	召回率/%	精确率/%
本文方法	76.4	91.8	64.6
关键字匹配	26.0	100	41.5
粗粒度方法	44.5	67.2	67.7

为了进一步验证所用规则集的正确性, 从规则集中随机选择 5、8、11、14 条规则进行实验. 由于准确率与召回率能够更加直观地反映规则提取短语的正确性, 因此分别测试不同数量规则在相同数据集下的准确率与召回率. 实验结果如下表 7 所示.

表 7 不同数量规则的准确率与召回率

Tab. 7 Precision and recall rate of rule subsets

规则数量/条	准确率/%	召回率/%
5	85.7	22.3
8	79.3	59.1
11	77.6	77.4
14	76.4	91.8

首先, 检测的准确率在各种情况下相差不大, 这说明不论规则数量的多少, 本文所使用的规则在提取短语时都有一定的正确性; 其次, 随规则数量的增多, 覆盖的短语结构和句型增多, 因此召回率随之升高, 在使用所有规则(14 条)时, 达到 91.8% 的召回率, 说明本文使用的规则集对常见短语结构的覆盖比较完整.

分析本文方法出现误报、漏报的主要原因如下.

(1) 情感词库的正确性与完整性问题导致的误报与漏报. 一方面, 本文方法在判定内容敏感性时依赖于情感词的情感倾向, 即该情感词所属情感词典的类别. 错误分类的情感词将影响情感分析的结果, 从而可能影响最终告警决策的输出, 产生误报. 另一方面, 情感词典的完整性将直接影响短语提取的正确性, 某一情感词在情感词典中的缺失可能导致提取出错误的短语, 或忽略本应被提取出的短语, 从而可能导致误报与漏报. 如正常文本“共产党人从不屑于隐瞒自己的观点”中, “不屑”一词在情感词库中的缺失导致了“共产党”与“隐瞒”被提取成为短语, 导致了误报. 除已有词典类别数量上的不足外, 情感词库的缺乏还表现在其他方面, 本文所用的词库仅包含动词、形容词, 但还有很多名词也表现了强烈的情感倾向. 如不良文本“这仅仅

是共产党的罪行之—”,“罪行”一词不在情感词库中,导致了漏报.

(2) 语法规则完整性问题导致的误报. 在解析复杂句型时,针对复杂句型规则的缺失将导致两种结果:一是没有任何一条规则匹配,直接进入出口规则;二是有至少一条其他规则匹配,继续解析,情况二将可能导致误报. 如正常文本“其成员对党进行了抹黑”,在匹配过程中“党”会与“抹黑”结合形成短语,此时“党”被错误地看作了“抹黑”的主语,“抹黑”一词属于主体负向和受体正向类,因此“党”一词与其关系直接影响了情感分析的结果,导致误报;语法规则不全一般不会导致漏报,因为在一个含敏感词的短句中,只要同时包含了存在于情感词典中的情感词,就一定会计与至少一条规则匹配.

6 结 论

针对使用字符串匹配方法和粗粒度情感分析方法来检测不良信息时出现的准确率低的问题,本文综合考虑了敏感词与其所在上下文表现的情感极性,提出了一种基于语法规则的短语级细粒度情感分析方法,结合二次分类的情感词典判定短语情感倾向进行告警决策. 实验表明,本文方法在准确率上较传统方法有较大的提升.

从实验结果来看,该方法的召回率和准确率依赖于情感词典与语法规则的完整性. 本文使用的情感词典数量偏少,还应进行扩充;目前的规则集由经验产生,由于中文语法的多样性,该规则集不能完全覆盖所有句型,在以后的工作中将持续更新规则集提高其覆盖率.

参考文献:

- [1] 邓一贵, 伍玉英. 基于文本内容的敏感词决策树信息过滤算法[J]. 计算机工程, 2014, 40: 300.
- [2] 付聪, 余敦辉, 张灵莉. 面向中文敏感词变形体的识别方法研究[J]. 计算机应用研究, 2018, 36: 988.
- [3] 李杨, 潘泉, 杨涛. 基于短文本情感分析的敏感信息识别[J]. 西安交通大学学报, 2016, 50: 80.
- [4] 姚艳秋, 郑雅雯, 吕妍欣. 基于 LS-SO 算法的情感文本分类方法[J]. 吉林大学学报: 理学版, 2019, 57: 375.
- [5] 胡思才, 孙界平, 琚生根, 等. 基于扩展的情感词典和卡方模型的中文情感特征选择方法[J]. 四川大学学报: 自然科学版, 2019, 56: 37.
- [6] Schouten K, Fransincar F. Survey on aspect-level sentiment analysis[J]. IEEE T Knowl Data En, 2016, 28: 813.
- [7] Om P D. Improving pointwise mutual information (PMI) by incorporating significant co-occurrence [C]//Conference on Computational Natural Language Learning. [S. l.]: Computer Science, 2013.
- [8] 王建新. 计算机语料库的建设与应用[M]. 北京: 清华大学出版社, 2005.
- [9] Liu B. Sentiment analysis and opinion mining [M]. California: Morgan & Claypool Publishers, 2012.
- [10] 黄胜. Web 评论文本的细粒度意见挖掘技术研究[D]. 北京: 北京理工大学, 2014.
- [11] Kauter M V, Breesch D, Hoste V. Fine-grained analysis of explicit and implicit sentiment in financial news articles [J]. Expert Syst Appl, 2015, 42: 4999.
- [12] Hu M Q, Liu B. Mining opinion features in customer reviews [C]//19th National Conference on Artificial Intelligence. [S. l.]: [s. n.], 2004.

引用本文格式:

中文: 明弋洋, 刘晓洁. 基于短语级情感分析的不良信息检测方法[J]. 四川大学学报: 自然科学版, 2019, 56: 1042.

英文: Ming Y Y, Liu X J. Sensitive information detection based on phrase-level sentiment analysis [J]. J Sichuan Univ: Nat Sci Ed, 2019, 56: 1042.