

文章编号: 1674-6864(2014)04-0016-05

基于语义分析的不良倾向文本的识别算法研究

刘梅彦 张仰森 张 涛

(北京信息科技大学 智能信息处理研究所 北京 100192)

摘 要: 分析了目前常用的不良倾向文本识别方法存在的困难和不足,设计了一种基于语义分析的不良倾向文本识别算法。该算法以语句为基本处理单元,采用依存句法获得句子的语义结构,结合 HowNet 词汇褒贬倾向性判别,识别不良信息。实验结果表明,此算法能够较好地提高不良文本识别效率和准确率。

关 键 词: 不良倾向文本; 语义分析; 依存句法分析

中图分类号: TP 391.1

文献标志码: A

Bad polarity text recognition algorithm based on semantic analysis

LIU Mei-yan ZHANG Yang-sen ZHANG Tao

(Institute of Intelligence Information Processing, Beijing Information Science and Technology University, Beijing 100192, China)

Abstract: The difficulties and problems in commonly used bad polarity text recognition method are analyzed. A bad text information recognition algorithm is designed based on semantic analysis. Using sentences as the processing unit, the dependency parsing is adopted to get semantic structure of sentence. Combined with pattern matching and semantic vocabulary tendentiousness discrimination based on HowNet, this algorithm can identify the harmful information. Experiments show that this algorithm can improve the efficiency and accuracy of bad text recognition.

Key words: bad polarity text; semantic analysis; dependency parsing

0 引 言

随着互联网技术的迅速发展,越来越多地人们在网络上发布信息和获取信息,如何有效地处理和过滤这些信息,已引起人们广泛的关注和研究。在网络信息空间中,隐藏着各种虚假、色情、暴力、反动和邪教等不良信息^[1],这些不良信息严重影响了人们的生活和社会稳定。

互联网上的文本信息可分为主题性信息和具有情感倾向的信息^[2]。前者内容直观清晰,可以通过特征词汇来识别;后者则表述了信息发布者的观点、态度、立场。例如,针对“法轮功”主题,既有对其批判的文章,也有对其进行反动宣传的文章,这就需要在判断其所属主题的基础上,进一步分析其情感倾向性,从而实现对不良信息文本的识别。

文本情感倾向性分析是实现不良倾向文本识别的重要环节。目前的文本情感倾向性分析一般采用基于统计的方法和基于语义分析的方法。基于统计

的方法,把文本看成词袋,忽略了词与词之间的位置关系、修饰关系等语义关联,该方法还需要构建大量基础语料,同时由于不良信息文本的语料不充分,这些因素都影响了倾向性分析结果的准确性。基于语义分析的方法,能弥补统计方法缺乏语义信息的弊端和语料不足的限制,更细粒度地获取文本的语义信息,能够有效地识别网络不良信息。

考虑到只有主观句才会影响文本情感倾向,本文作者提出了基于情感词典和敏感词典识别主观句的方法。另外,文本中还会经常出现对不同观点引用的情况,因此,在计算文本的情感倾向值时,要综合考虑局部情感倾向值与全局情感倾向值。

1 不良倾向文本识别的语义分析

语义分析是自然语言理解的主要技术,它是指根据句子的句法结构和句中实词的词义推导出能够反映这个句子的意义(句义)的某种形式化表示^[3]。然而,由于自然语言的复杂性和灵活性,目前还不能

收稿日期: 2014-05-49

基金项目: 国家自然科学基金资助项目(61070119, 61370139);北京市属高等学校创新团队建设与教师职业发展计划项目(IDHT20130519);北京市教委专项资助(PXM2013_014224_000042, PXM2014_014224_000067)

作者简介: 刘梅彦,女,副教授。

达到对句子的完全理解,在实际应用中,通常采用语义词典和语义框架近似地表达句子的语义。例如,文献[3]就采用语义框架表示句子的施事者、受事者和施加的动作,用语义模式识别句子的语义,在文本倾向性分析中取得了较好的效果。

在对文本进行褒贬倾向性判别时,并不需要精确获取句子的语义,只要知道各语义成分的倾向性即可。因此,本文采用文献[3]提出的语义模式概念作为文本倾向性判别的基础。但在抽取语义模式时,句子的一些细节被忽略,如修饰动词、名词、形容词等的修饰语,这些修饰语也会影响句子的褒贬倾向,因此,综合运用语义模式和词汇倾向性计算,才是实现不良倾向文本识别的有效途径。

判断一个中文句子是正常文本还是不良文本,仅仅知道句子的倾向性是不够的,还需要进一步知道观点持有者和评价对象的身份倾向(如代表正义一方或反动一方),因此,在句中识别观点持有者和评价对象的极性,也是不良信息文本判别的重要任务。

基于以上分析,本文采用简化的语义分析方法,使用哈工大开发的依存句法分析器获取句子的语法结构及各成分之间的关系,并将句子划分为主谓宾模式、主谓模式、动宾模式和独立成分4种语义模式,句子的情感倾向值根据语义模式和词汇的情感倾向值得。

下面给出使用依存句法提取文本中每个句子语义结构的算法。

输入:待判断文档 d 的句法分析结果XML文件

输出:文档 d 中每个句子的主干成分和语义模式,用九元组 $\langle s_id, schema, subject, predicate, object, ModifiedPolarity(subject), ModifiedPolarity(predicate), ModifiedPolarity(object), sentiment \rangle$ 表示,其中 s_id 表示句子编号, $schema$ 表示句子所属的语义模式, $subject$ 、 $predicate$ 、 $object$ 分别表示句子的主语、谓语和宾语, $ModifiedPolarity(subject)$ 、 $ModifiedPolarity(predicate)$ 、 $ModifiedPolarity(object)$ 分别表示主语、谓语和宾语的上下文极性(也称修饰极性^[4]), $sentiment$ 表示句子的极性。

对于文档中的每个句子:

步骤1 首先提取包含核心动词的依存关系对SBV、VOB关系对,经过处理后得到句子的主语($subject$)、谓语($predicate$)、宾语($object$)等,从而得到句子的语义模式 $schema$ 。

步骤2 利用ATT、ADV依存关系对,计算出句子的主语、谓语和宾语在上下文中的修饰极性(详见2.2.1节)。

步骤3 生成用于表示该句语义模式的九元组,其中的 $sentiment$ 在2.2.2节给出的算法中求得。

2 基于语义分析的文本倾向性判别

文本的倾向性判别包括2个阶段:第1个阶段是识别主观句;第2个阶段是在判断主观句的情感倾向的基础上,计算整个文本的情感倾向值,情感倾向包括正向、负向和中性。

2.1 词汇的情感倾向计算与主观句的识别

表达对事物的观点、态度、立场的句子称为主观句。对文本中所有句子进行分词和词性标注后,通过判断句中是否存在敏感词和情感词来断定句子是否属于主观句。

网络不良文本中包含了大量不良敏感词汇,而且,为躲避内容监控和信息过滤,这些敏感词汇经常变形,如“法轮功”变形为“法*轮*功”或“珺轮功”。因此,为识别不良信息需要构建一个敏感词典。但是,由于网络不良信息制造者不断寻找新的规避手段,因此在识别不良词汇时还需要实时检测,得到其想表达的敏感词汇,同时把检测到的敏感词汇添加到敏感词典中。本文采用正则表达式识别敏感词汇的变形体。

句子中情感词的词性一般为动词、名词、形容词、副词,可依据情感词典,得到这些词的情感倾向值。情感词典是主观句识别的基础,本文构建的情感词典,是以HowNet情感词典为基础,融合了大连理工大学整理的情感本体,情感词典的每个词条的格式为:序号、词汇内容、情感倾向值。

对于文本中的词汇,如果情感词典中包含该词,则直接得到该词汇的情感值;如果情感词典中不包含该词,则通过计算其与HowNet中的基准词近似度而得到该词的情感值,基准词是从HowNet中抽取的已标注“良”(褒义)和“莠”(贬义)的词汇。

下面给出基于HowNet的语义相似度计算公式计算词汇倾向极性值的方法^[5]。

对于两个汉语词汇 w_1 和 w_2 ,如果 w_1 有 n 个义项(概念): $c_{11}, c_{12}, \dots, c_{1n}$, w_2 有 m 个义项(概念): $c_{21}, c_{22}, \dots, c_{2m}$, w_1 和 w_2 的相似度取各个概念的相似度之最大值,即:

$$\text{Sim}(w_1, w_2) = \max_{i=1, \dots, n, j=1, \dots, m} \text{Sim}(c_{1i}, c_{2j}) \quad (1)$$

由于所有的概念都最终用义原来表示,所以义原的相似度计算是概念相似度计算的基础。根据2个义原在树状义原层次体系中路径距离 d ,2个义原 p_1 和 p_2 之间的语义距离采用式(2)计算,其中 α 是一个可调节的参数。

$$\text{Sim}(p_1, p_2) = \frac{\alpha}{d + \alpha} \quad (2)$$

取 2 个义项的义原之间相似度最大值作为 $\text{Sim}(c_{1i}, c_{2j})$ 。

最后,将待定词汇 W 与基准词集中褒义词 WP_i 的近似度减去待定词汇 W 与基准词集中贬义词 WN_j 的近似度,得到词汇 W 的语义倾向性 polarity , 计算公式^[6]如下:

$$\text{polarity}(W) = \sum_{i=1}^{k_p} \text{Sim}(WP_i) - \sum_{j=1}^{k_n} \text{Sim}(WN_j) \quad (3)$$

式中 k_p 为褒义词可能的近义词词数; k_n 为贬义词可能的近义词词数。

2.2 主观句的情感倾向判别

主观句中短语的情感极性判断是句子倾向分析中一个重要元素。通常的方法^[7]是:将词组作为特征项,按一定的抽取规则抽取词组,将词组的情感值累加求和,得出整个文本的情感倾向值。此类方法的缺点是没有考虑词组与句子其他成分之间的语义关系,不能准确获取句子的语义倾向。由于句子中情感短语的极性值,由情感词的原始极性与修饰词或关联词的极性联合决定,因此,在计算句子的情感极性时,为了考虑词组与句子其他成分之间的语义关系,本文将情感短语的极性用该短语中被修饰词(简称中心词)的上下文极性(简称修饰极性)来表示,例如褒义词“支持”添加否定词前缀后修饰极性为正。

2.2.1 词汇在句子上下文中情感倾向值计算

构成情感短语的词组由动词、名词、形容词和副词构成,短语中的各词之间构成修饰关系或关联关系。使用哈工大开发的依存句法分析器,可以识别出情感短语中各词汇之间存在的 ATT(定中)关系、ADV(状中)关系、SBV(主谓)关系和 VOB(动宾)关系,前 2 种为修饰关系,后 2 种为关联关系。另外,此处为了避开句子的主语、谓语、宾语等主干成分,规定情感短语的依存关系对 SBV 和 VOB 中不能出现整个句子的核心动词。

如果某依存关系对的中心词在另一个依存关系中作修饰语,这就构成了关系链。依次处理关系链中每个关系对的中心词的修饰极性(按照表 1 给出的规则),最后则得到关系链末尾的中心词的修饰极性。

令 m 表示关系对中的修饰词或关联词, $\text{polarity}(m)$ 表示 m 的极性;令 c 表示关系对中的中心词, $\text{polarity}(c)$ 表示 c 的原始极性, $\text{ModifiedPolarity}(c)$ 表示 c 的修饰极性。当 m 是普通情感词时,按照表 1

给出的规则计算 c 的修饰极性。

表 1 普通情感词作修饰词或关联词时中心词的修饰极性

修饰词或关联词的极性 $\text{polarity}(m)$	中心词的原始极性 $\text{polarity}(c)$	中心词的修饰极性 $\text{ModifiedPolarity}(c)$
褒义	褒义	褒义 $\text{ModifiedPolarity}(c) = (\text{polarity}(m) + \text{polarity}(c)) / 2$
褒义	贬义	贬义 $\text{ModifiedPolarity}(c) = \text{polarity}(c)$
贬义	褒义	贬义 $\text{ModifiedPolarity}(c) = \text{polarity}(m)$
贬义	贬义	褒义 $\text{ModifiedPolarity}(c) = -(\text{polarity}(m) + \text{polarity}(c)) / 2$
中性	褒义/贬义	褒义/贬义 $\text{ModifiedPolarity}(c) = \text{polarity}(c)$
褒义/贬义	中性	褒义/贬义 $\text{ModifiedPolarity}(c) = \text{polarity}(m)$

否定副词和程度副词一般在 ADV 关系对中修饰动词或形容词,通过匹配否定词表和程度副词表,可以识别 ADV 关系对中的否定副词和程度副词。

否定副词将被修饰词的情感极性变向,例如“支持”是正向情感词,“不支持”则变成了负向情感词。本文采用文献[7]给出的否定词表,该否定词表包含了 22 个否定副词,即“并非、不、不对、不再、不着、不至于、从不、毫不、毫无、决非、绝非、没、没有、尚未、未、未必、未尝、未曾、永不、不大、不太、不很”。当修饰词为否定副词时,情感词的修饰极性值计算公式如下:

$$\text{ModifiedPolarity}(c) = -\text{polarity}(c) \quad (4)$$

程度副词可以增强、减弱或改变情感词的极性,不同的程度副词对情感词的影响强度不同。本文采用了文献[7]中给出的程度副词词典,如表 2 所示。

表 2 程度副词词典

序号	程度副词	强度系数
1	最、最为、极、极为、极其、极度、极端	1.5
2	大、至、至为、顶、过、过于、过分、分外、万分、何等	1.4
3	很、挺、怪、老、非常、特别、相当、十分、甚、甚为、异常 深为、蛮、满、够、多、多么、殊、何其、尤其、无比、尤为	1.3
4	不堪、不胜、好、好不、颇、颇为、大、大为	1.2
5	稍稍、稍微、稍许、略微、略为、多少	1.1
6	较、比较、较为、还	0.9
7	有点、有些	0.8

当修饰词为程度副词时,中心词的修饰极性不变,极性值计算公式如下:

$$\text{ModifiedPolarity}(c) = \text{polarity}(c) \times \text{value}_{\text{adv}} \quad (5)$$

式中 $\text{value}_{\text{adv}}$ 为程度副词的强度系数。

2.2.2 主观句的情感倾向值计算

经过语义分析后,每个主观句已经表示为一个九元组 $\langle s_id, schema, subject, predicate, object, Mod-$

ifiedPolarity(subject) , ModifiedPolarity(predicate) , ModifiedPolarity(object) , sentiment > , schema 中保存了句子的语义模式类型 , sentiment 的值待求。若 sentiment < 0 , 表明此句为不良句子 , 若 sentiment > 0 , 表明此句为正常句子 , 若 sentiment = 0 , 表明此句为中性句子。本文在判断句子的情感倾向时 , 综合考虑句子的各语义成分及相互语义关系 , 尤其是将句子的主语和宾语名词所代表的观点持有者和评价对象的潜在倾向加以综合考虑 , 全面考虑了句子的语义。例如, “张三是大家学习的榜样。”这句话中若主语“张三”是负面敏感词汇, 则此句为不良句子。

下面给出基于语义模式的句子倾向极性值 sentiment 的计算方法。

1) “主谓模式”的极性计算规则

若 ModifiedPolarity(subject) > 0 , 则 sentiment = ModifiedPolarity(predicate) ;

若 ModifiedPolarity(subject) < 0 , 则 ,

若 ModifiedPolarity(predicate) < 0 ,

sentiment = - (ModifiedPolarity(subject) + ModifiedPolarity(predicate)) / 2 ,

否则 sentiment = ModifiedPolarity(subject) ;

若 ModifiedPolarity(subject) = 0 , 则 sentiment = 0。

2) “动宾模式”的极性计算规则

若 ModifiedPolarity(object) = 0 , 则 sentiment = 0;

若 ModifiedPolarity(predicate) > 0 , 则 sentiment = ModifiedPolarity(object) ;

若 ModifiedPolarity(predicate) < 0 , 则 sentiment = - ModifiedPolarity(object) 。

3) “主谓宾模式”的极性计算规则

若 ModifiedPolarity(subject) > 0 , 则 sentiment 的值与“动宾模式”取值规则一致;

若 ModifiedPolarity(subject) < 0 ,

则当 ModifiedPolarity(predicate) ≤ 0 且 ModifiedPolarity(object) ≤ 0 时 ,

sentiment = - (ModifiedPolarity(subject) + ModifiedPolarity(predicate) + ModifiedPolarity(object)) / 3;

其他情况下 sentiment = - (| ModifiedPolarity(subject) | + | ModifiedPolarity(predicate) | + | ModifiedPolarity(object) |) / 3;

若 ModifiedPolarity(subject) 是中性 ,

则当 ModifiedPolarity(predicate) ≥ 0 时 , sentiment = ModifiedPolarity(object) ;

当 ModifiedPolarity(predicate) < 0 且 ModifiedPolarity(object) ≠ 0 时 , sentiment = - ModifiedPolarity(object) , 否则 , sentiment = 0。

4) “独立成分”的极性计算规则

sentiment = ModifiedPolarity(predicate) 。

2.3 文本的情感倾向值计算

计算出每个主观句的情感值后 , 通过综合统计来判断整个文本的情感倾向性 , 即对全部主观句的情感值累加 , 再除以主观句的个数 (考虑了文本的长度) , 计算公式参见公式 (6) 。设待识别的文本用 d 表示 , 句子的褒贬倾向性用 $S(d)$ 表示 , 每个句子的情感值为 sentiment_i , n 为主观句的总数。

$$S(d) = \frac{1}{n} \times \sum_{i=1}^n \text{sentiment}_i \quad (6)$$

若 $S(d) > 0$, 表明该文本是正向文本; 若 $S(d) < 0$, 表明该文本为不良文本 , 否则该文本为中性文本。

3 实验及结果分析

为了测试本文设计的不良信息文本识别算法 , 设计如下实验步骤:

步骤1 对文本进行预处理 , 包括分词、词性标注和去除停用词;

步骤2 根据敏感词典和情感词典 , 识别并抽取文本中主观句;

步骤3 采用语义结构抽取算法 , 获得每个主观句的语义结构及语义模式;

步骤4 计算每个主观句的各主干成分的原始极性和修饰极性 , 并按照基于语义模式的句子倾向性算法 , 求出句子的褒贬倾向性;

步骤5 根据式 (6) 计算文本的褒贬倾向性;

步骤6 抽取文本的情感词汇 , 采用基于词汇的方法计算文本的褒贬倾向性;

步骤7 抽取文本的情感词汇作为特征项 , 采用基于支持向量机 (SVM) 方法计算文本的褒贬倾向性。

步骤8 评估各方法的准确率和召回率。

为了测试本文设计的不良信息文本识别算法 , 作者收集了与法轮功主题有关正向、负向和无关文档 280 篇 , 其中正向文本 50 篇 , 负向文本 30 篇 , 无关文本 200 篇。无关文本中 50 篇是与佛教主题相关且无倾向性 , 另外 150 篇是其他主题的文本。

分别采用基于词汇的褒贬倾向性方法、基于 SVM 方法与本文提出的基于语义模式的方法进行了文本的褒贬倾向性实验比较。实验结果如表 3 所示。

表 3 实验结果

类别	数量/篇	方法 1		方法 2		方法 3	
		准确率/%	召回率/%	准确率/%	召回率/%	准确率/%	召回率/%
A 正向文本	50	65	40	66	45	76	58
B 负向文本	30	63	37	63	47	78	64
C 无关文本	50	68	43	72	53	84	87

从实验结果看,与不考虑句法结构的基于词汇的倾向性计算方法和基于统计的 SVM 方法相比,本文采用的基于语义模式的褒贬倾向性分析方法的准确率和召回率稍好。这是因为本文的方法不仅考虑了词汇的情感极性,还考虑了词汇在句子上下文中的位置信息、修饰关系以及否定词和程度副词。然而,由于本文的方法对依存句法分析结果的高度依赖、句子本身的不规范以及句法分析存在的误差,影响了此方法的准确率和召回率。后续的工作将考虑如何改进算法,以修正句法分析不准确带来的影响。

4 结束语

本文介绍了一种基于语义分析的不良文本识别方法。通过对相关主题语料的分析研究,给出了利用敏感词典情感词典和句法分析中的依存关系,识别待处理的句子、计算句子情感极性和文本情感极性的解决方案。这种不良文本识别方法在句子规范条件下,能够通过识别依存关系对获取句子的语义模式、句子各词语之间的修饰关系,从而有效计算句子的情感倾向性。然而,在句子不规范时,句法分析的结果与实际相差较大,严重影响此方法的效果。另外,基于语义的褒贬倾向性分析算法自身的主观性和经验性,也影响了文本的褒贬倾向性计算的准确率,在后续的工作中应继续完善此算法。本文下一步的工作中还要增加对指代消解和复合句的处理,提高对真实文本识别的准确率和效率。

参考文献:

- [1] 12321 网络不良与垃圾信息举报受理中心. 网络不良与垃圾信息举报受理情况月报 [EB/OL]. [2014-05-10]. <http://www.12321.cn/>
- [2] 刘永丹,曾海泉,李荣陆,等. 基于语义分析的倾向性文本过滤 [J]. 通讯学报, 2004, 25(7): 78-85
- [3] 金峰,刘永丹,江宝林,等. TTFS: 一个倾向性文本过滤系统的设计与实现 [J]. 计算机工程与应用, 2003, 30: 137-140
- [4] 姚天昉,娄德成. 汉语语句主题语义倾向分析方法的研究 [J]. 中文信息学报, 2007, 21(5): 73-79
- [5] 朱嫣岚. 基于 HowNet 的词汇语义倾向计算 [J]. 中文信息学报, 2006, 20(1): 14-20
- [6] 徐琳宏,林鸿飞,杨志豪. 基于语义理解的文本倾向性识别机制 [J]. 中文信息学报, 2007, 21(1): 96-100
- [7] WEN Bin, HE Tingting, LUO Le, et al. Text sentiment classification research based on semantic [J]. Computer Science, 2010, 20(1): 261-264 (in Chinese)
- [8] 闻彬,何婷,罗乐,等. 基于语义理解的文本情感分类方法研究 [J]. 计算机科学, 2010, 20(1): 261-264