

文章编号: 1003-0077(2013)05-0093-07

基于序列标注模型的情绪原因识别方法

李逸薇¹, 李寿山^{1,2}, 黄居仁¹, 高伟²

(1. 香港理工大学 中文及双语学系, 香港;

2. 苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

摘要: 情绪原因识别是情绪分析中的一个重要研究任务。该任务旨在自动分析出导致某一情绪发生的原因描述。该文将情绪原因识别任务建模为序列标注模型, 即将情绪词相关的子句当成序列, 进而整体标注出哪些属于原因子句。具体实现中, 我们使用条件随机场(CRF)模型进行求解, 并结合了基本词特征、词性特征、距离特征、上下文特征及语言学特征等多种特征进行原因识别。实验结果表明, 所采用的这些特征对于原因识别都有一定帮助, 特别是上下文特征。此外, 我们发现在使用类似特征集合的情况下, 序列标注模型能够获得比分类模型更好的识别效果。

关键词: 序列标注; 情绪原因识别; 上下文特征; 语言学规则特征

中图分类号: TP391

文献标识码: A

Detecting Emotion Cause with Sequence Labeling Model

LI Sophia Yat Mei¹, LI Shoushan^{1,2}, HUANG Churen¹, GAO Wei²

(1. Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hongkong, China;

2. School of Computer Sciences and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

Abstract: Emotion cause detection is an important task in the research on emotion analysis. This task aims to detect the cause description of a emotion happening. In this study, we model this task as a sequence labeling problem and predict each related sentence to be in a emotion cause or not. Specifically, we apply the conditional random field (CRF) model to solve this problem with various of features, such as basic word features, POS features, context features and linguistic rule features. Empirical studies demonstrate that these features are effective for the task, especially the context features. Moreover, we find that the sequence labeling model is superior to the classification model when similar features are employed.

Key words: sequence labeling; emotion cause detection; context feature; linguistic rule features

1 引言

文本情绪分析是自然语言处理(Natural Language Processing, NLP)研究中的一个重要研究方向。该方向旨在研究如何自动分析文本所表达的情绪及与情绪相关的信息。目前, 主流的情绪分析研究主要集中在情绪分类上面, 该任务旨在对文本表达的情绪(例如, 高兴、伤心、惊讶等)进行自动分

类^[1-3]。然而, 该任务所关注的情绪信息仅仅是情绪类别, 属于比较浅层的情绪信息。为了更好的理解文本表达的情绪信息, 迫切需要挖掘文本表达的关于情绪的更深层次的信息。例如, 情绪的体验者、情绪的原因、情绪的结果等, 进行进一步的探讨和研究。本文关注一种关于情绪的更深层次的信息, 即情绪原因, 研究如何自动识别给定情绪的触发原因。以下面句子为例, 我们的目标是识别“伤心欲绝”情绪的原因“你遗弃我”。(注: [$*01e$]与[$*02e$]之

收稿日期: 2013-06-15 定稿日期: 2013-08-02

基金项目: 香港 GRF 项目(543810), 国家自然科学基金资助项目(61003155, 61273320)

作者简介: 李逸薇, 女, 助理教授, 主要研究方向为语言学; 李寿山(1980—), 男, 副教授, 主要研究方向为自然语言处理, 模式识别; 黄居仁(1958—), 男, 讲座教授, 国际计算语言学委员会(ICCL)会士, 主要研究方向为自然语言处理, 语言学。

间的内容为情绪原因,<emo id=0>与</emo>之间的内容为情绪关键词)

例1 [*01e]你遗弃我[*02e]后,我<emo id=0>伤心欲绝</emo>。

在语言学研究方面,大多数关于情绪语言学的模型都将情绪的原因事件作为一个重要的组成部分^[4-7]。因此,研究情绪原因可以帮助理解情绪的产生和发展的语言学机制。此外,情绪原因识别的研究有助于帮助自动处理和分析同情绪相关的事件,进而应用于突发事件监测、情感摘要等其他任务中。

虽然情绪分析研究已经开展多年,但是大部分的研究主要集中在情绪识别和分类方面。相对而言,对于情绪原因分析的研究还非常缺乏。Lee等^[8]和Chen等^[9]是仅有的针对情绪原因识别进行的研究工作,分别采用了规则的方法和分类模型方法进行情绪原因识别,取得了一定的效果。然而,这些方法对于情绪原因识别任务存在一些不足。具体来讲,规则的方法需要领域专家定义大量规则,代价较高。分类模型的构建相对容易,并且取得更佳的识别效果,但是分类模型是将文本中每个分句单独对待,并不能捕捉文本中分句与分句之间的关系。事实上,分句与分句之间的关系往往对情绪原因识别起到关键作用。以下面句子为例,由于“因为”这种明显的功能词,很容易识别出来“感情失意”是情绪“伤心”的原因,在已经识别出原因句后,其前面的句子及其他句子是原因的可能性就很小。这种利用样本标签之间关系的特征是分类模型没有办法捕捉到的。

例2 江祖平不只在「九指新娘」每集都有哭戏,在「后山日先照」也不少;林依晨在「十八岁的约定」饰演痴情种子,因为[*01e]感情失意[*02e],所以有不少<emo id=0>伤心</emo>流泪的镜头;许玮伦在「十」剧饰演感情错综复杂的凉子,为情落泪、痛哭流涕的镜头更是让人印象深刻。

本文对情绪原因识别重新建模,将之视为一个序列标注问题,进而采用条件随机场(Conditional Random Field,CRF)进行求解。新模型的优势在于不仅仅很好的融合了词法、距离、语法规则等多种特征,还能够充分考虑到样本上下文之间的特征信息。实验结果表明,本文提出的序列标注模型相对于分类模型能够获得更佳的识别性能。

本文其他部分组织如下:第2节中介绍自然语言处理中情绪分析方向的相关研究工作;第3节给

出情绪原因识别任务描述及相关语料介绍。第4节提出基于序列标注的情感原因识别方法及相关特征介绍;第5节给出了实验结果和分析;第6节是本文的内容总结及工作展望。

2 相关工作

文本情绪分析已经成为自然语言处理领域中的一个研究热点。按照情绪类别的粒度来分,基本上可以将相关工作分为粗粒度和细粒度情绪分析任务。粗粒度情绪分析具体是指对文本表达的情感极性进行分析,涉及的情感类别一般包括两类,即正类和负类^[10]。自从Pang等^[10]关于情感分类研究开始,关于粗粒度情绪分析的研究受到了NLP领域广大研究者的关注。粗粒度情绪分析研究中主要集中在两个任务的研究:情感分类和情感信息抽取。其中,情感分类研究旨在对文本所表达的情感极性进行分类。目前主流的情感分类方法是基于机器学习的分类方法,具体包括基于监督学习的情感分类方法^[11],基于半监督学习的情感分类方法和基于非监督学习的情感分类方法^[12]。情感信息抽取是抽取出自情感表达相关的元素,例如,评价对象和观点持有者。近几年来,相继出现了大量的关于情感信息抽取的研究^[13]。

细粒度情绪分析具体是指对文本表达的情绪进行分析。情绪类别包括的种类远远多于两类,例如,高兴、伤心、惊讶等。关于细粒度情绪分析的研究主要包括下面几个内容:情绪词典构建^[14-15]、句子级和篇章级的情绪资源建设^[16]、基于监督学习的情绪分类方法研究^[1,9,17]。

本文研究的内容是属于细粒度情绪分析。但是,不同于大多数已有研究,我们研究的关注点是情绪的原因识别。据我们所知,关于情绪原因识别的研究才刚刚起步。密切相关的研究主要有两篇文章:文献^[8]和^[9]。这两篇文章中的情绪原因识别方法分别是基于规则的和基于分类模型的识别方法。与他们不同的是,本文的情绪原因识别方法是基于序列标注模型,能够更好的对该任务建模,从而获得更好的识别效果。

3 任务描述及语料分析

情绪原因识别任务是指识别文本中针对某一情绪体验的触发原因。该任务是一种特殊的信息抽取

问题。同一般信息抽取问题不一样的是,所要抽取的内容同一个情绪关键词关联。由于同某一情绪关键词关联的情绪原因一般离情绪词的距离较近,Chen 等^[9]指出多于 80% 的情绪原因事件或者实体会出现在情绪关键词所在的子句的前后两个子句之间。因此,我们的情绪原因识别任务是指对情绪词周边的子句进行标注。如图 1 所示,在例 3 中,我们对情绪关键词“伤心”周边的分句(位置分别为 Left2, Left1, Left0, Right0 和 Right1)给出其是否属于原因分句的标签。如果原因在该分句中,则标注这个分句为“1”,否则为“0”。关于例 3 中的子句标注结果如表 1 所示。

表 1 例 3 中情绪原因识别中的标签设置

分句位置	样本	标签
Left2	陆陆续续的	0
Left1	我又听到几位同事也因这位主管的口不择言	1
Left0	感到受挫	0
Right0	甚至消极好一阵子	0
Right1	这位主管做事果断有决心	0

例 3 <PrefixSentence>只为了澄清谣言中不实的部分,而非强辩他对我批评,于是私底下与他沟通,最后他向我道歉了事。</PrefixSentence><FocusSentence>陆陆续续的,我又听到几位同事也因[*01e]这位主管的口不择言[*02e],感到受挫、<emo id=0>伤心</emo>,甚至消极好一阵子。</FocusSentence><SuffixSentence>这位主管做事果断有决心,但是喜欢批评别人,而且经常断章取义的搬弄是非,弄得同事们心里不快。</SuffixSentence>

为了对该任务进行分析并构建相应的机器学习识别系统, Lee 等^[8]从 Sinica Corpus 平衡语料^①中收集并标注了 5 964 篇情绪实例。语料中,情绪原因分为两个类别:事件型情绪原因和名词型情绪原因。具体来讲,事件型情绪原因是指情绪发生的原因是一个事件所导致,在语法表现方面,这些原因里面一般有动词。名词型情绪原因是指情绪发生的原因是某一具体事物,在语法表现方面,这些原因一般由名词短语组成。例 3 是语料中情绪标注的一个实例,其中,原因事件标记[*01e]和[*02e]之间的内容,具体的,0 是与它相关的情绪关键词的标号,1 表

示原因事件的开始,2 表示原因事件的结束,“e”表示该原因属于事件型情绪原因,名词型情绪原因由“n”表示。该语料的相关统计信息及示例如表 2 所示。

表 2 Lee 等^[8]情绪原因语料库相关统计信息及示例

	数目	示 例
句子群	5 964	例 1、例 2、例 3
带有情绪的句子群	4 260	<emo id=0>伤心</emo>
带有情绪原因的句子群	3 460	[*01e]这位主管的口不择言[*02e],
事件型情绪原因的句子群	2 941	[*01e]这位主管的口不择言[*02e]感到受挫、<emo id=0>伤心</emo>
名词型情绪原因的句子群	519	提到[*01n]球场[*02n],就很<emo id=0>伤心</emo>

从表 2 可以看出,情绪原因在情绪相关文本中出现的比率非常高,81% 的情绪关键词都有对应的情绪原因内容。在两种情绪原因类型中,事件型情绪原因的数目远远多于名词型情绪原因。

4 基于序列标注模型的情绪原因识别方法

4.1 情绪原因识别的序列标注建模

已有的研究将情绪原因识别作为分类问题来看,如引言里面所述,分类问题并不能捕捉子句和子句之间的类别关系。为了更好的利用子句之间类别的关系,本文对情绪原因识别重新建模,将之作为一个序列标注问题。情绪原因识别任务的序列标注模型如图 1 所示。该模型同现有的情绪原因识别模型^[9]最大的区别在于,可以充分考虑利用子句和子句的之间的关系,提高识别性能。除了子句和子句之间的关系特征,该模型也可以方便地将子句内部的特征,例如,词特征和词性特征,融入到模型中进行学习和识别。

4.2 CRF 学习算法及相关特征

针对序列标注模型的求解,表现较好的是条件随机场方法,即 CRF 方法。条件随机场模型是 Lafferty 等^[18]在最大熵模型和隐马尔可夫模型的基础

① <http://www.sinica.edu.tw/ftms-bin/kiwi1/mkiwi.sh>

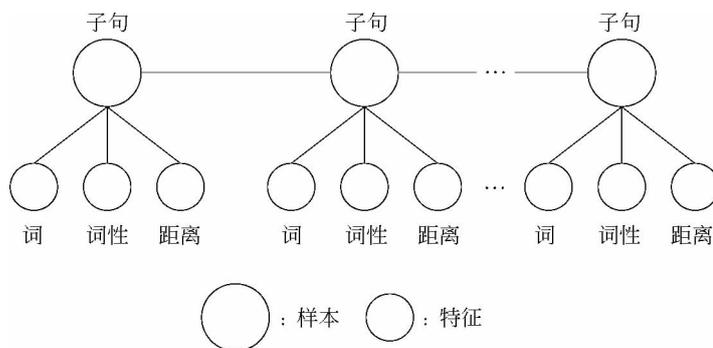


图1 情绪原因识别任务的序列标注模型示意图

上提出的一种无向图学习模型,是一种用于标注和切分有序数据的条件概率模型。CRF 方法是对序列的整体优化。具体求解公式如式(1)所示:

$$p_{\lambda}(Y | W) = \frac{1}{Z(W)} \exp\left(\sum_{t \in T} \sum_k \lambda_k f_k(y_{t-1}, W, t)\right) \quad (1)$$

其中: $Y = \{y_t\}$ 表示对应的输出标记序列, $y_t \in \{1, 0\}$ 代表对应的子句是否是情绪原因或者情绪原因里面的一个子句。 W 表示待测子句序列。 $Z(W)$ 是归一化因子, f_k 为特征函数, t 为相应特征的下标。

特征集包括描述当前对象基本特征和描述上下文信息的两类特征,分别由表 3 和表 4 进行描述。其中描述当前对象自身信息的特征主要为当前子句包括的名词、动词及其数目。只挑选名词和动词的原因是考虑到情绪原因主要包括两个类型,名词表示的名词型原因和动词表示的事件型原因。距离特征是另外一个重要的位置特征。一般来说,情绪的原因会出现在情绪关键词的左边,距离特征能够很好地捕捉这个特性。上下文特征主要包括上一句和下一句中的词法特征。此外,如引言所述,上一句的类别标签在序列标注问题中能够提供充分的信息。

表 3 基本特征: 当前子句中的词特征和距离特征

特征	说明
名词	当前句子中的名词,如果当前句子中没有名词,对应位置填上“NULL”
动词	当前句子中的动词,如果当前句子中没有动词,对应位置填上“NULL”
名词个数	当前句子中所包含的名词个数
动词个数	当前句子中所包含的动词个数
距离	当前子句与情绪关键词所在子句的距离, left2, left1, left0, right0, right1 的距离值分别为 -2, -1, 0, 0, 1

表 4 上下文特征

特征	说明
上一句中的动词	上一子句中包括的动词,如果没有,对应位置填上“NULL”
上一句中的名词	上一子句中包括的名词,如果没有,对应位置填上“NULL”
上一句的标签	上一子句的类别标签,如果没有,对应位置填上“NULL”
下一句中的动词	下一子句中包括的动词,如果没有,对应位置填上“NULL”
下一句中的名词	下一子句中包括的名词

因此,我们也加入了上一句的标签特征。

除上述特征外,我们还加入了语言学规则特征,这些特征集合来自 Chen 等^[9]的论文。具体的规则集合如表 5 所示。表 5 中的规则用来帮助识别情绪原因子句,其中 I/II/III/IV/V/VI 为关键词集,具体定义如下。

表 5 规则特征集^[9]

序号	规则
1	i) C(B/F)+III(F)+K(F) ii) C=the nearest N/V before III in F/B
2	i) IV/V/I/II(B/F)+C(B/F)+K(F) ii) C=the nearest N/V before K in F
3	i) I/II/IV/V (B)+C(B)+K(F) ii) C=the nearest N/V after I/II/IV/V in B
4	i) K(F)+V/VI(F)+C(F/A) ii) C=the nearest N/V after V/VI in F/A
5	i) K(F)+II(A)+C(A) ii) C=the nearest N/V after II in A
6	i) III(F)+K(F)+C(F/A) ii) C=the nearest N/V after K in F or A

续表

序号	规 则
7	i) yue4 C yue4 K “the more C the more K” (F) ii) C=the V in between the two yue4’s in F
8	i) K(F)+C(F) ii) C=the nearest N/V after K in F
9	i) IV(B)+C(B)+K(F) ii) C=the nearest N/V after IV in B
10	i) C(B)+K(F) ii) C=the nearest N/V before K in B

- I={“为”,“为了”,“对”,“对于”,“以”}
- II={“因”,“因为”,“由于”,“于是”,“所以”,“因而”,“可是”}
- III={“让”,“令”,“使”}
- IV={“想到”,“想起”,“想来”,“说道”,“说起”,“讲到”等}
- V={“听”,“听到”,“听说”,“看”,“看到”,“看见”,“见到”等}
- VI={“的是”,“的说”,“于”,“能”}

C、K、B、F、A 缩写的含义分别为: C 代表情绪原因;K 代表情绪关键词;F 代表包含情绪关键词的子句;B 代表 F 之前的子句(位于 F 左边);A 代表 F 之后的子句(位于 F 右边)。

5 实验

5.1 实验设置

本文采用 Chen 等^[9]的实验语料,语料的相关介绍可以参考第 3 节。我们选取 80%的语料作为训练集,20%的语料作为测试集。文本处理中所用的词性信息是由斯坦福 Parsing 工具产生^①。

评价标准为精确度(Precision),召回率(Recall)和 F 值(F-score)。由于本任务中牵涉到两个类别,即属于情绪原因和不属于情绪原因,对应的评价结果也有两组打分。例如,Precision-P 表示分类结果中针对子句属于情绪原因的精确度;Precision-N 表示分类结果中针对子句不属于情绪原因的精确度。

本实验中,序列标注模型的实现借助于条件随机场模型 CRF++^②工具,采用默认参数。特征模板中加入了 4.2 节中所介绍的所有特征,具体使用 unigram 当前特征和 Bigram 的上下文特征等。

5.2 实验结果及分析

表 6 给出了基于序列标注模型的原因识别方法在依次加入不同类型特征后的识别结果,其中词特征选取了子句中的名词,动词,名词个数和动词个数作为子句的特征,距离特征选取了子句与包含情绪关键词子句的距离作为特征,语言学规则特征为表 3 中的各项规则对应的特征。由于情绪原因标注任务的特殊性,非情绪原因的子句(N)要远远多于情绪原因子句(P),这种不平衡性导致 Fscore-N 的值要远远高于 Fscore-P 的值。从表中的数据可以看出以下几点。

1) 仅仅使用最基本的词特征,虽然 Fscore-N 已经达到了很高的值,但是 Fscore-P 仅有 0.294。该结果表明词特征远远不能满足我们的需求。

2) 在词特征的基础上增加上下文特征以后,Fscore-P 有大幅度的提升。该结果表明子句的上下文对于识别情绪原因有着很重要的启示作用。

3) 在上面的特征基础上加入距离特征后,Fscore-P 有小幅度的提升。该结果表明子句与情绪关键词所在子句的距离对于情绪原因句的识别有一定帮助。

4) 增加了语言学规则特征以后,Fscore-P 有了进一步提升。该结果表明语言学规则对分析情绪原因句的有效性。然而,Fscore-N 变化不大。这可能是因为 Fscore-N 已经达到峰值,很难再有大幅度的提升。

表 7 给出了基于序列标注模型和分类模型比较结果,其中分类模型采用的是最大熵分类器,使用了词特征、距离特征及语言学规则特征。从表 6 中可以看出,序列标注模型明显优于分类模型。其中,Fscore-P 高出约 5%,Fscore-N 高出约 4.1%,序列标注模型与分类模型的主要差别在于该模型充分考虑了上下文信息,序列标注模型体现出的优势也充分说明了上下文信息在情绪原因识别任务中的重要性。值得一提的是,Chen 等^[9]中汇报的结果同我们的结果有些差异,这个主要是因为 Chen 等^[9]中所采用规则的词性信息等是人工标注结果(标准答案),而我们的结果是在词性标注工具识别的结果上面进行的。因此,我们使用的规则在准确率方面受到一定影响,并不能给系统带来大幅度的提高。

① <http://crfpp.sourceforge.net/>

② <http://nlp.stanford.edu/software/lex-parser.shtml>

表6 序列标注模型中加入不同特征的比较结果

特征 \ 评价标准	Precision-P	Recall-P	Fscore-P	Precision-N	Recall-N	Fscore-N
词特征	0.506	0.207	0.294	0.833	0.951	0.888
+上下文	0.519	0.344	0.414	0.854	0.923	0.887
+距离特征	0.523	0.364	0.429	0.858	0.920	0.888
+语言学规则特征	0.520	0.377	0.437	0.860	0.917	0.888

表7 分类模型与序列标注模型比较结果

模型 \ 评价标准	Precision-P	Recall-P	Fscore-P	Precision-N	Recall-N	Fscore-N
分类模型	0.312	0.500	0.384	0.900	0.803	0.849
序列标注模型	0.537	0.364	0.434	0.858	0.924	0.890

5.3 错误分析

基于序列标注的模型已经取得一定效果,但是对于一些特殊情况的情绪原因句识别仍然不理想,主要分析结果如下。

(1) 当情绪原因句与情绪词距离过远的时候(如例4所示),往往容易将近距离的子句识别为情绪原因句。CRF模型错误地将“老公公一点也不觉得痛”识别为情绪原因句。

例4 小精灵的领袖说完,[*01e]红色的精灵马上伸手摸摸老公公脸颊上的大瘤,那个瘤立刻被摘下来[*02e],可是,老公公一点也不觉得痛。老公公<emo id=0>高兴</emo>极了,用手摸摸突然变轻的脸颊,笑嘻嘻的走下山,回到家里。好心的老公公所住的村里,还住着一位坏心的老公公。

(2) 当句子中并没有情绪原因的时候(如例5所示),CRF模型有时候会错误识别出情绪原因句。例5中只有情绪,而没有与情绪直接相关的情绪原因句,这种情况占少数,多数情况下情绪原因句位于情绪词的周边,CRF模型将“就载着满船的鹅”识别为其情绪原因句,导致错误的识别。

例5 王羲之大喜过望,道士也赶紧把早就准备好的笔墨跟绢拿出来请他写。王羲之写完了道德经,就载着满船的鹅,<emo id=0>高高兴兴</emo>的回去了。他又听说,有一个老太太养了一只鹅,不但长得好,而且叫声洪亮。

(3) 当原因包括多个子句时候,CRF模型仅仅识别出其中一个原因子句(如例6所示)。例6中情绪原因子句为“他们只知道学成后,我依然可以当老师”,两个子句,而CRF模型只识别出“我依然可以

当老师”这一句为情绪原因句,忽略了前一句。

例6 老画家范洪甲回忆当初自台南师范学校毕业后,要继续到东京美术学校深造时父母并没有反对。「[*01e]他们只知道学成后,我依然可以当老师[*02e],就十分<emo id=0>高兴</emo>,很少出门的祖父还到港口来送我,」范洪甲回忆。帝展的滋味然而为了跨入帝展、台展等所谓官办的沙龙展门槛,不可免的,画家就必须进入比赛约定俗成的风格中。

6 结语

本文旨在研究情绪原因识别任务,将该任务建模为序列标注问题,具体使用条件随机场模型进行求解。学习过程中,结合了基本词特征、词性特征、距离特征、上下文特征及语言学特征等多种特征进行原因识别。实验结果表明,所采用的这些特征对于识别原因都有一定帮助,特别是上下文特征。此外,在使用类似的特征集合的情况下,序列标注模型能够获得比分类模型更好的识别效果。

从实验结果可以看出,情绪原因识别任务仍然具有很大的挑战。目前的方法取得的效果还非常有限,子句属于情绪原因的F值还仅在45%左右。下一步工作中,我们将重点解决错误分析中的无情绪原因情况和多个子句为情绪原因的情况。分别针对这两种情况进行研究和改进模型,进一步提升情绪原因识别的性能。

参考文献

- [1] Alm C, D Roth, R Sproat. Emotions from Text; Ma-

- chine Learning for Text-based Emotion Prediction [C]//Proceedings of EMNLP-05, 2005: 579-586.
- [2] Mihalcea R, H Liu. A Corpus-based Approach to Finding Happiness [C]//Proceedings of the AAAI Spring Symposium on Computational Approaches to Weblogs. 2006.
- [3] Tokuhisa R, K Inui Y Matsumoto. Emotion Classification Using Massive Examples Extracted from the Web[C]//Proceedings of COLING. 2008: 881-888.
- [4] Descartes R. 1649. The Passions of the Soul[M]. J. Cottingham et al. (Eds), The Philosophical Writings of Descartes. 2008, Vol(1): 325-404.
- [5] James W. 1884. What is an Emotion? Mind, 9(34): 188-205.
- [6] Plutchik R. Emotions: A Psychoevolutionary Synthesis[M]. New York: Harper & Row. 1980.
- [7] Wierzbicka A. Emotions across Languages and Cultures: Diversity and Universals [M]. Cambridge: Cambridge University Press. 1999.
- [8] Lee S, Chen Y, Huang C et al. Detecting Emotion Causes with a Linguistic Rule-based Approach [J]. Computational Intelligence. 2012.
- [9] Chen Y, S Lee, S Li, et al. Emotion Cause Detection with Linguistic Constructions [C]//Proceeding of COLING-10. 2010: 179-187.
- [10] Pang B, L Lee. Opinion Mining and Sentiment Analysis: Foundations and Trends [J]. Information Retrieval, 2008, vol. 2(12): 1-135.
- [11] Cui H, V Mittal, M Datar. Comparative Experiments on Sentiment Classification for Online Product Comments [C]//Proceedings of AAAI-06. 2006, 1265-1270.
- [12] Li S, C Huang, G Zhou, et al. Employing Personal/Impersonal Views in Supervised and Semi-supervised Sentiment Classification [C]//Proceedings of ACL-10. 2010, 414-423.
- [13] Kim S, E Hovy. Identifying Opinion Holders for Question Answering in Opinion Texts [C]//Proceedings of the Workshop on Question Answering in Restricted Domain at AAAI-05. 2005.
- [14] Xu G, X Meng, H Wang. Build Chinese Emotion Lexicons Using A Graph-based Algorithm and Multiple Resources [C]//Proceeding of COLING-10. 2010: 1209-1217.
- [15] Volkova S, W Dolan, T Wilson. CLex: A Lexicon for Exploring Color, Concept and Emotion Associations in Language [C]//Proceedings of EACL-12. 2012: 306-314.
- [16] Quan C, F Ren. Construction of a Blog Emotion Corpus for Chinese Emotional Expression Analysis [C]// Proceedings of EMNLP-09. 2009: 1446-1454.
- [17] Purver M, S Battersby. Experimenting with Distant Supervision for Emotion Classification [C]//Proceeding of EACL-12. 2012: 482-491.
- [18] Lafferty J, McCallum A, Pereira F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data [C]//Proceedings of ICML-2001. 2001: 282-289.
- [19] 宗成庆. 统计自然语言处理 [M]. 清华大学出版社: 北京, 2008.