

北京大学学报(自然科学版)
Acta Scientiarum Naturalium Universitatis Pekinensis
doi: 10.13209/j.0479-8023.2014.012

基于句法分析的跨语言情感分析

陈强^{1,2} 何炎祥^{1,2,†} 刘续乐¹ 孙松涛^{1,2} 彭敏^{1,2} 李飞^{1,2}

1. 武汉大学计算机学院, 武汉 430072; 2. 武汉大学软件国家重点实验室, 武汉 430072; † 通信作者, E-mail: yxhe@whu.edu.cn

摘要 提出利用句法分析模型, 将语句分成若干组合词, 根据组合词的主谓成分中情感词对于句子情感贡献的不同, 分别赋予不同的权重。统计分析该语句的情感分布特征, 利用得到的特征参数训练分类器, 再将训练好的分类器用于测试语料的情感分类。实验结果表明, 与已有的判别方法相比, 该方法的情感分类判别准确率较理想。此研究方法也可用于语句的比较级判别和否定句的极性判断等领域。

关键词 跨语言; 情感分类; 句法分析; 贝叶斯分类

中图分类号 TP391

Cross-Language Sentiment Analysis Based on Parser

CHEN Qiang^{1,2}, HE Yanxiang^{1,2,†}, LIU Xule¹, SUN Songtao^{1,2}, PENG Min¹, LI Fei^{1,2}

1. School of Computer of Wuhan University, Wuhan, 430072; 2. State Key Laboratory of Software Engineering of Wuhan University, Wuhan University, Wuhan 430072; † Corresponding author, E-mail: yxhe@whu.edu.cn

Abstract Using the syntactic analysis model, the statement is divided into several combinations of words. According to the subject-predicate component of compound words and emotional color difference of emotional words, different weights are given respectively. Statistical analysis of the distribution of the emotional statement, use the characteristic parameter training the classifier, and employ the trained classifier for the test corpus emotion classification. Experiment results show that the emotion classification discriminant accuracy rate and recall rate of this method is more ideal, compared with the existing discrimination methods. This research method can also be used in the statement of comparative discrimination and negative polarity judgment.

Key words cross-language; sentiment analysis; parser; Bayes classification

目前, 互联网(如社交网络、电子商务平台和论坛等)上产生了大量的用户参与的, 对于人物、事件、产品等有价值的评论信息。这些评论信息表达了人们的各种情感色彩和情感倾向, 如喜、怒、哀、乐和批评、赞扬等。基于此, 潜在的用户就可以通过浏览这些主观色彩的评论来了解大众舆论对于某一事件或产品的看法。目前, 随着文化交流的增加, 多语种的评价资源日益普遍, 如何进行跨语言情感分析已成为自然语言处理的热点问题。

目前, 对于语句情感倾向性的识别有多种方法, 句法分析方法的优势在于处理复杂语句时, 能够很好地定位到情感词语以及其修饰副词与评价对象的

关系。王根等^[1]使用 CRF 和冗余标签对句子序列进行情感倾向性标注, 这种方法的缺陷在于需要人工标注的语料进行训练, 需消耗大量人力。另一种是基于句子结构的分析, 薛丽敏等^[2]提出基于句法分析的 5 种结构用于情感倾向性识别。

而关于跨语言情感分类任务方面, Wan^[3]提出利用协同训练的方法, 具体做法是: 在中文情感语料及资源缺乏的情况下, 利用已经标注的英文情感语料和翻译工具, 将未标注的中文评论语料翻译成未标注的英文语料, 再将已标注的英文评论语料翻译形成已标注的中文评论语料, 这样就将中英文评论语料看成两个独立的视图, 然后, 利用协同训

国家自然科学基金(61070083, 61303115)资助

收稿日期: 2013-06-29; 修回日期: 2013-08-29; 网络出版时间: 2013-11-11 10:25

练的方法,抽取情感特征,形成一个 SVM 分类器完成跨语言的情感分类任务。

同样, Meng 等^[4]提出一种跨语言混合模型,利用大量未标注的双语平行语料,选取适合的使得双语平行语料似然最大化的参数用来学习,获得情感词典中未出现的情感词来提高情感词典的词覆盖度,通过学习目标语言句子和源语言句子间的对齐关系,借助跨语言混合模型扩充的情感词典训练出分类器,完成跨语言的情感分类任务。

本文根据 NLP&CC2013 测评任务中的跨语言情感分类的测评任务,利用不同语言成分对于句子情感的贡献不同,提出一种基于句法分析的跨语言情感分析方法。首先,利用翻译平台将中文语料翻译为英文,同时利用句法分析,根据句子主谓两部分对于句子情感贡献的不同,本文将情感词色彩的强弱赋予不同权重,计算语句中积极情感词和消极情感词的比例,并用训练语料不断迭代修改权重,训练分类器,最后用训练好的分类器完成分类工作。本文在上述情感分类的方法上对亚马逊上的评论语料进行实验,实验结果表明此方法可以有效提高跨语言情感分析的准确率,对情感细粒度分析以及电子商务决策提供必要支持。

1 句法分析

句法分析是指根据给定的语法,自动地识别出语句含有的句法单位以及它们之间的关系。句法分析是对自然语言进一步语义分析的基础,已成为大多数自然语言处理应用的关键因素。句法分析大体分为两种途径:基于规则的方法和基于统计的方法。基于规则的方法是以知识为主体的理想主义方法,采用非歧义的规则形式描述或解释歧义行为或特性。基于统计的句法分析^[5-6]必须以某种方式对语言的形式和语法规则进行描述,通过对已知句法分析结果进行训练获得句法分析模型。

本文采用基于树库的统计句法分析^[7-9],这也是现代句法分析的主流技术。利用句法分析得到的语法树形式的结果,提取情感词所在的主谓成分,作为该词的属性记录,为统计整个语句的情感特征作为权重赋值标准之一。同时,可以通过对语法树的分析,抛弃没有情感色彩的连接词,从而提高算法效率。以“My dog also likes eating sausage.”为例进行句法分析,得到语法树如图 1 所示。

通过对语法树的分析,得到如下结果:(My,

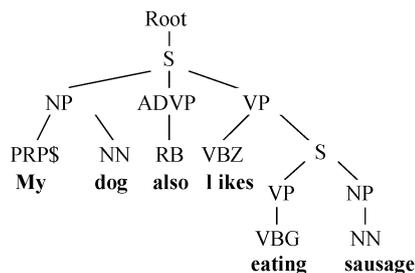


图 1 句法分析树示例
Fig. 1 Example of parse trees

NP), (dog, NP), (likes, VP), (eating, VP)和(sausage, VP), 由于 also 是主谓连接词,不具有感情色彩,因此可以丢弃。

利用句法分析得到的句子主谓部分对于句子的情感贡献大小不同对于情感句的情感分类有着重要影响。例如,对于例句“肮脏的他做出那么一件感人、伟大、美好的事。”如果单纯考虑该句的情感词的个数,“感人”、“伟大”、“美好”这种积极色彩的词语多于“肮脏”这种消极词语的数量,就会将该句子标记为积极句。如果加入主谓情感贡献度,可以看出“肮脏”位于句子的主语部分,而“感人”、“伟大”、“美好”这类积极情感词则位于句子的谓语部分,当主谓赋予不同权重后,该情感句就有可能标记为消极句,说明主谓语部分对于整个情感句的情感贡献值并不相同,因而利用主谓情感权重的不同对于正确判断情感句的极性具有重要的意义。再例如,情感句“看似这么正直的人居然说出如此狠毒的话。”如果单纯考虑这句话的两个情感词“正直”和“狠毒”,它们分别位于主语和谓语部分,而整个情感句表现出来的是消极情感。经过统计和研究我们发现:如果单纯考虑情感句的主谓成分的情感,情感句实际上是重心后移的,即谓语部分对于情感句的情感贡献度明显大于主语部分。

综合上述,本文对于情感句进行句法分析分别抽出主语和谓语两个部分,然后分别对于主谓两部分设置不同的情感贡献权重,对句子的正负情感分别进行累计,作为最终情感句的正负情感度量。

2 基于句法分析的跨语言情感分析模型

2.1 跨语言情感分析

情感分类系统依赖于高质量的情感资源,然而,现实生活中,不同语种中的情感资源分布极为不平

衡。因此当某种语言的情感资源匮乏时，利用其他语言中的情感资源来实现该语言的情感分类被称为跨语言情感分析，跨语言分析逐渐成为情感分析领域的热门研究课题。此次，NLP&&CC2013 跨语言评测使用亚马逊评论的中文语料，并只利用英文情感词典进行中文语料的情感分类。

面对跨语言分析问题，首先都应应将中文语料翻译为英文进行情感识别，接下来提出两种方案：1) 直接利用分词系统，将翻译后的英文语句分为单个单词，并添加该词语的词性作为属性，并逐个单词匹配情感词典，得出情感词的感情色彩及其强弱程度，并根据情感词处于“summary”或“text”中的不同分别赋予不同权重，统计语句的情感特征；2) 对翻译后的英文语句使用句法分析，记录词在句子成分(主要区别为主语部分和谓语部分，以下提到情感词所在句子成分，均为此意)中的位置，同样根据每个单词及其词性匹配情感词典，对每个情感词赋予不同权重，在考虑情感词情感强弱和位置的同时，还需考虑该词所处句子的主语或者谓语部分。通过两个方案的比较，方案 2 显然考虑了情感词的更多特性，更为全面地分析了情感语句中不同情感词的作用。因此本文采用方案 2 建立跨语言分析模型，实现跨语言情感分类任务。

2.2 基于语法分析的跨语言情感分析模型

2.2.1 模型训练

1) 将评论切分成句子，并对各个句子进行句法分析。

2) 根据 NLP&&CC2013 测评发布的英文情感词典匹配句子中的情感词，并且记录情感词在评论中的“summary”或者“text”部分，在情感词在对应句子成分标签：“sub”(主语部分)或者“pred”(谓语部分)，情感词典中标注的情感词的情感强度标签“strongsubj”或者“weaksubj”以及其情感极性“positive”或者“negative”。

3) 根据步骤 2 中的对于识别出的情感词的记录项，做以下权重设定：

① 根据情感词所在句子成分不同分别赋予情感权重 α_1 (情感词位于句中主语部分)和 α_2 (情感词位于句子谓语部分)；

② 根据 NLP&&CC2013 测评发布的英文情感词典匹中所标识的情感强弱分别赋予 β_1 和 β_2 ；

③ 根据句子所处评论的“summary”和“text”部分设定权重 λ_1 和 λ_2 。

4) 统计一个词的权重大小为累积权重，积极、消极情感词的情感值分别可由式(1)和(2)计算：

$$v_{\text{-pos}}^{(s)} = \alpha_i^{(s)} \times \beta_j^{(s)} \times \lambda_k^{(s)}, \quad (1)$$

$$v_{\text{-neg}}^{(t)} = \alpha_i^{(t)} \times \beta_j^{(t)} \times \lambda_k^{(t)}, \quad (2)$$

$\alpha_i^{(s)}$, $\beta_j^{(s)}$ 和 $\lambda_k^{(s)}$ 表示评论中第 s 个积极情感词根据步骤 3 的记录选择的对应权重，同样 $\alpha_i^{(t)}$, $\beta_j^{(t)}$ 和 $\lambda_k^{(t)}$ 表示评论中第 t 个消极情感词的对应权重。

5) 整条评论的积极情感值和消极情感值分别由式(3)和(4)给出：

$$W_{\text{-positive}} = \sum_{s=1}^m v_{\text{-pos}}^{(s)} = \sum_{s=1}^m \alpha_i^{(s)} \times \beta_j^{(s)} \times \lambda_k^{(s)}, \quad (3)$$

$$i, j, k \in \{1, 2\},$$

$$W_{\text{-negative}} = \sum_{t=1}^n v_{\text{-neg}}^{(t)} = \sum_{t=1}^n \alpha_i^{(t)} \times \beta_j^{(t)} \times \lambda_k^{(t)}, \quad (4)$$

$$i, j, k \in \{1, 2\}.$$

6) 计算得到积极词语和消极词语的比值 P 作为该评价最终的情感极性的判别特征，记为式(5)：

$$P = \frac{W_{\text{-positive}}}{W_{\text{-negative}}}, \quad (5)$$

当 $W_{\text{-positive}}$ 和 $W_{\text{-negative}}$ 中的某一个值为零时，不计算 P 值。

7) 再利用评论中积极情感值 $W_{\text{-positive}}$ 和消极情感值 $W_{\text{-negative}}$ 作为分类特征，利用已有的标注的评论语料作为训练语料(去除只有单个情感值的评论)，计算出所有英文评论的积极消极情感比值 P ，训练得出贝叶斯分类器中 P 的阈值 Threshold ，具体训练步骤参见 2.3 节贝叶斯情感分类方法。

8) 对已标注的中文语料(规模较小)进行翻译，经过步骤 1~7 算出 P 值，利用贝叶斯情感分类方法继续训练，在已有阈值 Threshold 基础上进行微调得到最终的贝叶斯分类器的阈值 Threshold 。其中对于中文语料翻译部分做如下说明。

跨语言情感分析必须经过翻译，由于不同语种的语法习惯，语句结构的不相同，翻译后的语料必定与原有语料的语义有少许差错，但是语句表达的感情色彩仍得到保留。本文利用 Google 翻译平台完成中英文语料的翻译，通过模拟访问 Google Translate 在线获得翻译结果，这是由于 google 翻译采用了全球范围内样本的“统计机器翻译”技术，整体机器翻译水平处于全球领先，得到的结果可以保留原有语句中较多的情感词。本文在实验过程中同时对比金山在线翻译和微软 Translator API，最终采用 Google Translate 在线翻译获得翻译对应英文表

述。

2.2.2 测试

1) 利用 Google Translate 在线, 将测试语料翻译成英文;

2) 进行模型训练中步骤 1~6, 统计出 $W_{positive}$ 和 $W_{negative}$, 并计算出 P 值, 将这 3 类特征作为分类依据, 结合 P 值阈值 $Threshold'$ 比较, 分类公式如下:

$$f(W_{positive}, W_{negative}, P) = \begin{cases} -1, & \text{当 } W_{positive} = 0 \&\& W_{negative} \neq 0, \\ & \text{或 } W_{positive} \neq 0 \&\& W_{negative} \neq 0 \&\& P \leq Threshold', \\ 1, & \text{当 } W_{positive} \neq 0 \&\& W_{negative} = 0, \\ & \text{或 } W_{positive} \neq 0 \&\& W_{negative} \neq 0 \&\& P \leq Threshold', \end{cases} \quad (6)$$

其中, -1 表示评论表现消极情感, 1 表示评论表现为积极情感。

2.2.3 跨语言情感分类流程

跨语言情感分类模型训练流程如图 2 所示。

2.3 贝叶斯情感分类方法

本文将每条评论得到的积极情感值与消极情感值作为情感分类特征, 比值 P 作为分类依据。因为贝叶斯分类算法的准确度和较小的计算量等特点, 本文采用贝叶斯分类模型进行跨语言情感分类任务。

2.3.1 贝叶斯分类器

Bayes 分类算法是统计学分类方法, 它利用概率统计知识进行分类的算法。贝叶斯分类算法方法简单、速度快。情感分类的过程分为 3 个阶段。

1) 准备工作: 主要为训练分类器做准备, 该阶段依据英文评论语料, 根据情感词的情感强弱、所处句子成分以及所处评论中的位置(“summary”和“text”两个位置)计算得到出 3 个分类特征 $W_{positive}$, $W_{negative}$ 和 P 值, 对英文训练语料进行初步处理, 去

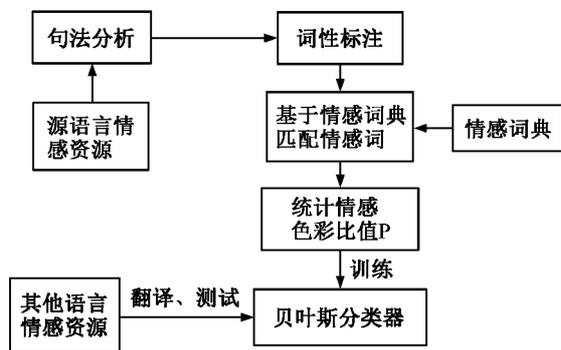


图 2 跨语言情感分类系统模型

Fig. 2 Cross-language sentiment classification system model

除只包含单一情感的英文评论得到训练语料 C_{train} 。

2) 分类器训练阶段: 训练出高质量分类器。该阶段的输入是在准备工作阶段人工选定的特征属性 $W_{positive}$, $W_{negative}$, P 值和训练语料 C_{train} , 输出训练得到的分类器, 主要是得到分类阈值 $Threshold'$ 。

3) 应用阶段: 使用分类器对待分类项进行分类。输入是分类器和待分项, 输出是分类结果 -1 和 1, 分别代表评论表现为消极和积极情感。

2.3.2 情感分类

本文采用 Bayes 分类算法, 用两次贝叶斯分别得出各商品英文评论的最佳阈值和修正得到对应商品中文评论最佳情感阈值。

1) 第一次贝叶斯分类: 根据每条评论中统计得出的积极情感值与消极情感值作为分类特征, 比值 P 作为分类依据。对测评任务提供的英文语料各商品类型分别进行训练。根据经验判断, 当 $P=1$ 时, 消极与积极情感相抵消, 评论应该表现为中性; 当 $P>1$ 时, 积极情感大于消极情感, 评论表现为积极情感; 当 $P<1$ 时, 积极情感小于消极情感, 评论表现为消极情感。但是, 由于实际计算统计与理论之间的误差以及对于不同商品的情感分布的差异, 可能会发生阈值情感中性值偏移的情况, 因此本文将判别阈值 $Threshold$ 的范围定位为以 1 为中心的一个区域 $[1-\epsilon, 1+\epsilon]$, 且以 Δx 为步长进行训练。

先验工作: 根据训练语料计算得到的 P 值, 统计得到的 $p(P > Threshold | positive)$, $p(positive)$ 和 $p(P > Threshold)$ 。

具体训练: 当 $P > Threshold$ 时, 计算语料的贝叶斯概率(计算如式(7)), 根据 $Threshold$ 的范围 $[0.8, 1.2]$, 取步长为 0.05, 对不同阈值 $Threshold$ 计算 8 次贝叶斯概率, 取贝叶斯概率最高的 $Threshold_{best}$ 作为一次贝叶斯分类器。

$$p(positive | P > Threshold) = \frac{p(P > Threshold | positive) * p(positive)}{p(P > Threshold)} \quad (7)$$

2) 第二次贝叶斯分类: 与第一次贝叶斯分类过程相似, 主要区别在于, 这次的初始阈值设定为 $Threshold_{best}$, 阈值区域为 $[Threshold_{best} - 0.1, Threshold_{best} + 0.1]$, 步长为 0.01, 分别对不同阈值 $Threshold'$ 计算 20 次贝叶斯概率, 选取贝叶斯概率最高的 $Threshold'_{best}$ 作为二次贝叶斯分类器, 即最终分类器。

3 实验结果与分析

3.1 实验数据

本文的实验是在斯坦福句法分析器 Stanford Parser 下, 利用 2013 年 NLP&&CC 情感评测大会提供语料进行跨语言情感分类任务。包括: 1) 标注的英文评论训练语料; 2) 英文情感词典; 3) 标注的中文评论语料; 4) 中文未标注测试语料。数据均采用 XML 格式, UTF-8 编码存储。中英文评论语料按商品品类不同分类, 且出自亚马逊的用户评论, 包含 3 类商品的评论: DVD, Book 和 Music。每个领域内有 4000 条积极评价和 4000 条消极评价。MPQA 情感词典 (Multi-Perspective Question Answering Subjectivity Lexicon)中共包含 2789 个积极情感词和 6079 个消极情感词, 同时 MPQA 情感词典还给出每个词的情感强度、词性等信息。

3.2 实验结果及分析

本次测评任务中, 本文具体模型中涉及的权重设定如下: 1) 对于情感词位于句子成分的不同, 设定的权重为经验值为 $\alpha_1 : \alpha_2 = 0.75 : 1$; 2) 对于情感词情感强弱不同, 设定的权重为经验值 $\beta_1 : \beta_2 = 1 : 0.5$; 3) 对于所处评论中的位置不同, 设定的权重为经验值 $\lambda_1 : \lambda_2 = 1 : 0.5$ 。

经过贝叶斯分类器训练得到的对于 Book, DVD 和 Music 的领域中设定的阈值 $\text{Threshold}'_{\text{best}}$ 分别为 1.15, 1.0 和 1.0。根据测评返回结果, 本次跨语言情感分类任务中我们在各领域的识别准确率如表 1 所示。其中, DVD 和 Music 商品评价情感极性的分类准确率在所有提交结果中为第一名, Book 商品评价情感极性的分类准确率在所有提交结果中为第二名。测试结果中整体的准确率为 77.09%, 在所有测评结果位居第二。

由于后期情感评测组织方给出了评测语料的标注结果。本文针对本次评测任务, 对于情感词位于句子成分的不同, 设定的权重调整为为 $\alpha_1 : \alpha_2 = 1 : 0.75$, 并比对标注结果, 得到各商品类别评论的识别准确率如表 2 所示。

表 1 基于句法分析的跨语言下 DVD, Book 和 Music 领域情感分类准确率

商品类别	训练样本数	测试样本数	准确率/%
Book	4000	4000	*77.00
DVD	4000	4000	*78.33
Music	4000	4000	*75.95

表 2 基于句法分析的跨语言下 DVD, Book 和 Music 商品类情感分类准确率

商品类别	训练样本数	测试样本数	准确率/%
Book	4000	4000	66.67
DVD	4000	4000	67.33
Music	4000	4000	63.45

通过本文第 2 节对于情感句中主谓成分对于情感贡献度的分析, 本文假定情感句的情感重心具有“情感后移”的特性。从表 1 和 2 可以看出, 如果设定情感句中的主语成分中的情感以更大的情感权重, 则情感极性的分类准确率会有很大程度的下降, 这也表明本文统计分析得出的情感句“情感重心后移”特性的假设的正确性。这对于分析情感句极性具有重要的意义。

在利用基于句法分析进行跨语言情感分析评测任务后, 对在本文方法的框架结构下去掉句法分析模块后的跨语言情感分析评测数据也进行实验, 实验识别准确率如表 3 所示。

表 3 未加入句法分析信息下 DVD, Book 和 Music 商品类情感分类准确率

商品类别	训练样本数	测试样本数	准确率/%
Book	4000	4000	68.25
DVD	4000	4000	69.33
Music	4000	4000	66.75

从表 1 和 3 可以发现, 与未加入句法分析的传统情感因素打分方法相比, 利用本文提出的基于句法分析的跨语言情感分类方法在 3 种商品的情感分类任务取得的准确率有较大程度的提升, 说明加入句法分析信息对于提升情感分类的准确率具有重要的作用。

从表 2 和 3 可以发现, 在设定不恰当的主谓情感贡献权重时, 即打破“情感后移”假设的条件下, 加入句法分析后的情感判定准确率会出现逆增长情况, 进一步印证本文假设的合理性。

4 总结

基于上述研究, 本文提出了一种基于句法分析的跨语言情感分析方法。根据句法分析器的工作原理以及句法分析树的句子主谓部分对于句子情感的贡献度, 建立情感分析统计模型。结合句法分析的依存关系解析, 将句子划分成主谓两大成分, 并且

记录情感词所在句子成分后赋予相应不同权重。情感词加入句子成分的句法信息后的量化的情感,大大增加了有效情感词的比重和分析组合情感词的情感特性的准确率。因为贝叶斯分类算法的简单性和高效性,并结合本文分类研究工作,最终选取贝叶斯分类算法作为研究分类任务的工具,阐述如何结合分类器进行测试语料的分析,最后通过实验,进一步验证了该方案的有效性,为今后跨语言情感分类问题提供了支持。

本次测评中 Book 商品的情感预测准确率没有达到理想状况的原因在于 Book 评论中中文训练样本数目较少,导致利用贝叶斯分类训练得到的阈值 1.15 较大。在接下来的实验中,可以收集更多 Book 评论并人工标注后加入到原中文训练预料中继续训练,再次实验得到预测准确率进行分析。

同时,对于本次评测任务中对于影响进一步提高情感分类准确率的一些问题进行总结如下:1)利用翻译平台,会使翻译后的评论资源与原始语言的评论资源的情感色彩部分削弱,中文中的部分修辞如反语、反问等,翻译英文后,与原有情感句的情感色彩有偏差,从而减少了句子中有效的情感词;2)由于实验中只是将否定词后固定距离的情感词极性取反,可能对某些较长的否定句的情感判断产生了影响;3)由于很多语句存在转折词,如 however,这种类型的句子转折词之后才述真正情感,这样如果前半部分表达相反感情的句子较长,同样会对情感句的识别造成干扰;4)有些书名或者专辑的名字含有情感词,如《下一站,幸福》,这样对于情感词较少的评价语句,会对结果的识别造成障碍。

5 工作展望

根据本文所提出的方法,仍有地方需要加强和改进。

1)在进行谷歌翻译时,某一词语或单词存在多种目标语言的词语相对应,可以通过人工干预或机器学习的方法,选择在对应语句情境下最为合适的翻译,从而保留更多的感情词,使翻译后的语句与原有语句的感情色彩相差较小。

2)在句法分析中,可以引入句法层次和依存关系的概念,在句法分析树中,情感词所处的层次越高,对情感句的贡献度就应该越大,因此在进行情感统计时,层次越高的情感词赋予的权重就应该较大。同时利用依存关系可以解决否定词对情感语句的影响,因为依存关系可以直接找出否定词后直接

修饰的情感词,更准确地判断出现否定词后,哪些情感词情感色彩取反,哪些保持不变。

3)总结并扩充转折词词库,从而在后续工作中加入对转折词的判断,利用分词得到的结果判断语句是否存在“however”等一类的转折词。此类语句表达的重点在转折词之后的语句,因此我们可以抛弃转折词之前语句中存在的情感词,使跨语言分类模型更为有效。

参考文献

- [1] 王根, 赵军. 基于多重冗余标记 CRF 的句子情感分析研究 // 内容计算的研究与应用前沿——第九届全国计算语言学学术会议论文集. 北京: 清华大学出版社, 2007
- [2] 薛丽敏, 李殿伟, 肖斌. 中文文本情感倾向性五元模型研究. 通信技术, 2011, 44(7): 136 - 138
- [3] Wan Xiaojun. Co-training for cross-lingual sentiment classification // The 47th Annual Meeting of the Association for Computational Linguistic (ACL 2009). Singapore, 2009: 235-243
- [4] Meng Xinfan, Wei Furu, Liu Xiaohua, et al. Cross-lingual mixture model for sentiment classification // The 50th Annual Meeting of the Association for Computational Linguistic(ACL 2012). Jeju Island, Korea, 2012: 572-581
- [5] Manning C D, Schutze H. Foundations of statistical natural language processing. London: The MIT Press, 1999
- [6] Chelba C, Jelinek F. Structured language modeling. Computer Speech and Language, 2000, 14(4): 283-332
- [7] Xue Nianwen, Xia Fei, Chiou Fudong, et al. The Penn Chinese treebank: phrase structure annotation of a large corpus. Natural Language Engineering, 2005, 11(2): 207-208
- [8] Fung P, Ngai G, Yang Yongsheng, et al. A maximum-entropy Chinese parser augmented by transformation-based learning. ACM Trans on Asian Language Processing, 2004, 3(2):159-168
- [9] Goodman J T. A bit of progress in language modeling. Computer Speech and Language, 2001, 10: 403-434
- [10] Wang G J, Wang Z L, Li L X. An affective model of artificial psychology // IEEE International Conference on Communications, Circuits and System. Hong Kong, 2005: 1061-1064
- [11] Zhang Yu, Li Zhuoming, Ren Fuji, et al. Semi-automatic emotion recognition from textual input based on the constructed emotion thesaurus // Proceedings of 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering. Wuhan, 2005: 571-576
- [12] 徐琳宏, 林鸿飞. 基于语义特征和本体的语篇情感计算. 计算机研究与发展, 2007, 44(S2): 356-360