

HITIRSYS:COAE2012 情感分析系统¹

唐都钰, 石秋慧, 王沛, 赵妍妍, 秦兵, 刘挺

哈工大社会计算与信息检索研究中心, 哈尔滨, 150001

E-mail: dytang@ir.hit.edu.cn

摘要: 情感分析是自然语言处理领域的热点问题, 旨在通过评论文本分析用户的主观倾向性。本文主要研究第四届中文倾向性分析评测 (COAE 2012) 中的句子级倾向性分析任务 (任务 1) 和篇章级倾向性打分任务 (任务 3)。在句子级倾向性分析任务中, 本文结合子句极性识别和句子主干分析提出了基于主干分析的句子极性识别框架; 在篇章级倾向性打分任务中, 本文分别使用文本分类的方法和基于情感相似度的方法计算篇章的倾向性打分, 最后将二者的结果融合作为最终的篇章倾向性打分。本系统在两个评测任务中均取得了较好的成绩。

关键词: 情感分析; 极性识别; 倾向性打分; 主干分析

HITIRSYS: Sentiment Analysis System in COAE2012

Duyu Tang, Qiuhui Shi, Pei Wang, Yanyan Zhao, Bing Qin, Ting Liu

Research Center for Social Computing and Information Retrieval, Harbin Institute of Technology, Harbin 150001

E-mail: dytang@ir.hit.edu.cn

Abstract: Recently, sentiment analysis has become a hot research field in natural language processing. It aims to analyze people's subjective opinions through review. This paper focuses on the two sub-tasks in fourth COAE evaluation, namely sentence-level sentiment analysis task and document-level rating prediction task. In the sentence-level sentiment analysis task, we proposed a trunk based framework to identify the polarity of sentence, in which combines sub-sent level sentiment classification and trunk analysis. In rating predicting task, we both use text classification method and sentiment similarity based method to predict the rating. Then, the combined answer between these two results is used as the final rating. Our system achieves good performance across the two tasks.

Keywords: sentiment analysis; polarity classification; rating predicting; trunk analysis

1. 引言

情感分析又称意见挖掘, 旨在研究人们针对实体、人物、事件、主题及其属性的主观意见和情感[1, 2, 3]。按照处理文本粒度的不同可以将情感分析分为篇章级[4, 5, 6]、句子级[7, 10]、短语级[8]和词语级[9, 11]。前三届中文倾向性分析评测 (COAE) 在词语级、句子级、篇章级进行了中文倾向性分析评测, 任务涉及主客观分析、情感极性分析、评价对象抽取以及搭配抽取[14]等方面, 为中文倾向性分析提供了一个很好的平台。为了探索中文倾向性分析的新技术、新方法, 推动中文倾向性分析理论和技术的研究与应用, 本次评测共包含 3 个任务:

- 任务一 句子级倾向性分析
- 任务二 比较句的识别与要素抽取
- 任务三 篇章级倾向性打分

哈工大社会计算与信息检索研究中心开发了 HITIRSYS 情感分析系统, 参与 COAE2012 的任务一和任务三。本文内容安排如下: 第二部分介绍 HITIRSYS 情感分析系统在

¹ 本工作受国家自然科学基金面上项目资助, 基金号 60975055; 本工作受国家自然科学基金重点项目资助, 基金号 61133012。

COAE2012 的两个任务中所使用的词典资源; 第三部分和第四部分分别阐述句子级倾向性分析(任务一)和篇章级倾向性打分(任务三)的主要方法; 第五部分给出评测结果及结果分析; 第六部分对本文进行总结。

2. 词典资源构建

在介绍句子极性识别框架之前, 简要介绍本文使用的词典资源。词典资源是情感分析任务的基础, 在各个情感分析任务中起着至关重要的作用。本文使用了哈工大社会计算与信息检索研究中心的情感词典、修饰词典和连词词典资源。

(1) 情感词典共有 22419 个词语, 其中褒义和贬义情感词是针对不同评价对象情感极性一致的情感词语; 歧义情感词是针对不同评价对象情感极性不一致的情感词语。

(2) 修饰词典共有 450 个词语, 其中包含否定词和程度副词, 顾名思义二者分别表示可以对情感词语的极性进行翻转、加强和减弱的修饰性词语。

(3) 连词词典共有 33 个词语, 其中包含句间的总结性连词和转折、让步关系连词。总结性连词是可以对文章的语义总结和概括的连词; 转折关系连词是可以使相邻文本的语义发生翻转的连词。

其中情感词典的构建共包含两部分工作:

(1) 基于同义词词林的情感词典构建。我们首先以知网情感词典²作为种子, 使用哈工大《同义词词林(扩展版)》³对种子情感词语进行同义词扩充, 同时建立基于同义词关系的图模型, 最后使用主题相关的链接分析算法[11]建立情感词典, 记生成的词典为 SentiDict1。

(2) 基于微博的情感词典构建。鉴于微博上涌现出来的新词和流行语, 我们使用大规模的微博数据自动挖掘情感词语。首先我们将微博数据中的一元、二元和三元语法作为候选情感词语, 继而对候选词语过滤并利用候选词语的上下文信息构建大规模的图模型, 最后使用图传播算法[15]生成情感词典, 记生成的词典为 SentiDict2。

最后, 我们将情感词典 SentiDict1 和 SentiDict2 合并后进行人工过滤, 得到最终的情感词典; 修饰词典和连词词典均由人工构建。各词典的数据规模及样例如表 1 所示。

表 1 词典资源

词典资源类型	词典名称	数据规模	样例
情感词典	褒义情感词	10442	漂亮、美丽
	贬义情感词	11814	恶劣、差
	歧义情感词	163	高、低
修饰词典	否定词	108	不、绝非
	程度副词	342	很、非常
连词词典	总结性连词	10	总之、因此
	转折关系连词	9	但、但是
	让步关系连词	14	虽然、尽管

3. 句子级倾向性分析

3.1. 任务描述与分析

本任务的目的是检测系统在面对包含否定句、转折句时的倾向性判别能力, 句子的倾向

² <http://www.keenage.com/>

³ http://ir.hit.edu.cn/phpwebsite/index.php?module=pagemaster&PAGE_user_op=view_page&PAGE_id=162

性包含褒义（1）、贬义（-1）和中性（0）三种。通过分析语料，我们认为虽然很多句子包含否定、转折成分，但是本任务的最终目标是判断句子的整体倾向性，而不是判断否定、转折成分所在片段的倾向性。同时，如果句中包含多种倾向性同时具有主干成分，我们认为主干成分的极性可以代表整体句子的观点极性。

在本文中，句子的主干成分是指句子中可以强调、概括和汇总整个句子语义的成分。以数码领域的训练数据 Doc130 为例，句子内容如表 2 所示，（1）和（2）分别为该句的两个子句。其中子句（2）可以代表整个句子的总体语义，因此我们认为子句（2）是整个句子的主干成分。

表 2 主干成分样例

句子来源	句子内容
数码领域 Doc130	（1）虽然淘宝有些衣服不错 （2）但是衣服每次买回来都和照片上不一样

从表 2 我们可以看出子句（1）的极性是褒义，子句（2）的极性是贬义。但是由于子句（2）是整个句子的主干成分，并且该主干成分的极性代表了整个句子的极性，因此该句的整体情感极性为贬义（-1）。

3.2. 基于主干分析的句子极性识别

通过对任务一的训练语料进行统计分析，我们发现汽车领域评论中有 20.2% 的评论包含明显的主干成分，即句中包含总结性连词和转折关系连词，其中主干成分极性与句子整体极性一致的比例为 90.5%；类似地，数码领域的评论中有 25.8% 的评论包含主干句，其中主干极性与句子整体极性一致的比例为 85.6%。基于上述的统计分析，我们认为句子的极性识别应分为两个步骤：即（1）首先判断每个子句的情感极性，（2）随后通过主干分析识别句子的主干成分，并使用主干成分的极性代表整个句子的总体极性。鉴于此，本文提出了基于主干分析的句子极性识别方法，其中主要包含两个模块：子句极性识别和句子主干分析。具体流程如表 3 所示。

表 3 句子极性识别方法

基于主干分析的句子极性识别方法
输入：句子 S，子句极性识别模块 f()，主干分析模块 trunk() 输出：句子 S 的极性 $pol(S) \in \{-1,0,1\}$
1. 分句。对句子 S 按照 {, : . ? ! } 分句，分句结果记为 SubS[1~N]，N 为子句的总个数。 2. 预处理。对每个子句 SubS[i] 使用哈工大语言技术平台 LTP ⁴ [12] 分词和词性标注。 3. 子句极性识别。识别每个子句 SubS[i] 的极性，结果 f(subS[i]) 记为 $polSubS[i] \in \{-1,1\}$ 。 4. 强情感规则。对每个子句的情感线性叠加： 若 $\sum_i polSubS[i] > \varepsilon_p$ $pol(S) = 1$ ，返回。 若 $\sum_i polSubS[i] < \varepsilon_n$ $pol(S) = -1$ ，返回。 5. 主干分析。识别句子 S 的主干成分，若整个句子 S 包含多个主干候选，则把位置靠后的候选作为句子 S 的主干成分，记为 trunk(S)。 6. 极性判别。若句子 S 包含主干，则 $pol(S) = polSubS[trunk(S)] \in \{-1,0,1\}$ ，返回。 若句子 S 不包含主干，则 $pol(S) = f(S) \in \{-1,0,1\}$ ，返回。

从表 3 中我们可以看出，对于一条待测试的语句，首先对其分句（第一步）和预处理（第二步）；随后使用 3.2.1 中的子句极性识别模块判断每个子句的极性（第三步）；如果各子句

⁴ <http://ir.hit.edu.cn/demo/ltp/>

极性的线性求和满足预先定义的强情感规则，直接返回结果（第四步）；否则在分句的基础上使用 3.2.2 中的主干分析模块（第五步）；如果句子包含主干成分，把主干成分的极性作为整个句子的极性（第六步-1）；如果句子不包含主干成分，则把句子作为整体使用子句极性识别模块进行极性识别（第六步-2）。

3.2.1. 子句极性识别

文本的极性识别可以看作文本分类问题，已有众多学者使用文本分类算法识别文本的极性，其中以有指导的方法最具代表性[4]。有指导学习方法两个重要因素是标准的训练语料和特征提取方案。鉴于上述分析，本文使用机器学习的方法自动判断文本的情感极性，主要内容包括训练语料的自动获取和特征提取。

(1) 训练语料自动获取

互联网上存在海量的用户评论信息，部分数据包含了用户明确的情感倾向。以表 4 中的京东评论数据为例，评论者认为该产品的优点是“品质不错”，缺点是“送货太慢”，总体的评价是“品质不错，有游戏卡送”，并且打了“5.0 分”⁵以表示评论者对该产品的满意。通过观察数据，本文决定从京东评论数据中自动获取训练语料。

表 4 评论数据样例

打分	优点	缺点	总体评价
5.0	品质不错	就是送货太慢了，用了十几天才收到	品质不错，有游戏卡送

具体地，本文使用的京东评论数据来自数据堂⁶。为了保证获得优质的训练语料，我们只抽取 4 分和 5 分评论中的“优点”作为褒义的训练语料，1 分和 2 分评论中的“缺点”作为贬义的训练语料。经过去重和褒贬语料平衡之后，共剩余 50000 条褒义句和 50000 条贬义句，即为本小节的训练语料。

(2) 特征提取

给定一段文本，本小节旨在利用已有的词典资源提取特征，以供后续的机器学习方法训练模型。Pang[4]针对电影评论的倾向性分析实验表明：使用一元文法特征可以获得最佳的分类效果。然而在本文自动获取的训练语料上经过特征选择，我们获得了效果最佳的 3 类特征，具体特征描述及样例如表 5 所示。

表 5 句子特征抽取模板

特征类别	文本样例	特征样例
情感词+否定词	不怎么喜欢这个手机	neg_喜欢
歧义词+上下文	油价很高	油价_高
情感统计函数	不怎么喜欢这个手机	fP=0 fN=1,fP_N=-1

在构建情感词+否定词特征（表 5 第一行）时，需要判断每个情感词周围是否包含否定词。通过尝试不同的窗口设置，窗口最终设为[-2,0)，即考虑从当前情感词向左 2 个窗口内的词语是否包含否定词；类似地，对文本中的每个歧义词（表 5 第二行），我们在[-2,0)窗口内匹配与歧义词距离最近的名词，共同组成歧义词+上下文特征；情感词统计函数（表 5 第三行）共包含 3 种，分别是褒义词出现频率、贬义词出现频率以及褒贬频率的差值。

特征提取完成后，本文使用最大熵模型训练褒贬二元分类器。表 3 中第三步只需输出褒贬 2 类情感结果，本文使用最大熵的直接判别结果。而第五步和第六步需要判断褒贬中三类，本文的策略是对最大熵的输出概率设置阈值。通过在开发数据集上调参，我们最终选定：若最大熵分类器输出褒义的概率大于 0.8，句子为褒义；若贬义概率的概率大于 0.8，句子为贬

⁵ 京东的评论数据打分范围是 1.0~5.0，分值以 1.0 递增。1.0 分表示非常不满意，5.0 分表示非常满意。

⁶ <http://www.datatang.com/data/15516>，<http://www.datatang.com/data/43444>

义；若褒贬概率均小于 0.8，句子为中性。

3.2.2. 句子主干分析

主干分析旨在抽取句子中可以强调、概括和汇总整个句子语义的成分，即主干成分。通过语料分析，我们发现总结性连词和转折关系连词可以很好地指示整句话的主干信息。如在表 2 中，子句 (2) 包含了转折关系连词“但是”，子句 (2) 强调了整句话的语义；类似地，总结性连词如“总之”、“因此”可以很好地概括和汇总一句话的语义。

鉴于以上分析，本文使用总结性连词和强转折连词抽取句子的主干成分。通过分析数据，我们发现大部分总结性连词所在的子句即为主干成分；而包含转折关系连词的主干成分则分散在 (a) 转折关系连词所在子句，(b) 转折关系连词所在的子句的最后一邻接子句，(c) 整句话的最后一个子句。所以，我们使用了下述两条规则对句子进行主干分析。

(1) 若句子的某一子句包含总结性连词，则该子句为整句话的主干成分。

(2) 若句子的某一子句包含转折关系连词，记该子句的 id 为 i 。我们将从第 i 个子句开始到最后一个子句的子句群整体作为该句的主干成分。若有多个子句包含转折关系连词，则以位置最靠后的转折关系连词作为起始位置抽取主干成分。在识别此类主干成分的极性时，我们将该主干成分作为一个整体进行特征提取，继而判断其情感极性。

4. 篇章级倾向性打分

4.1. 任务描述与分析

本任务的目的是对给定的篇章评论进行打分，分值范围是 (1~5) 分，5 分表示强烈褒义，1 分表示强烈贬义。如评论“服务态度很差，晚上 7:00 到的呷哺，见有 2 对在等位，于是向服务员要个等位号，发号的服务员直接冲我吼道：没有号了，发完了。去过那么多家的呷哺店，这种情况我还是第一次遇到，态度太恶劣了，以后不会再去这家店”，标准的输出结果为 1 分。

通过任务分析，我们认为篇章的倾向性打分受两方面因素的影响：

第一个因素是篇章文本内出现的词语或短语。如果评论文本包含“非常好”、“非常满意”，则整个评论更可能具有较高的倾向性打分（如 4 分、5 分）；相反，如果评论文本包含“太差了”、“很不满意”，则整个评论更可能具有较低的倾向性打分（如 1 分、2 分）。

第二个因素是篇章的褒义句比例。如果一篇评论中有超过 90% 的句子是褒义句，整个评论更可能具有较高的倾向性打分；相反，如果褒义句的比例小于 10%，则整个评论更可能具有较低的倾向性打分。

4.2. 文本分类和情感相似度融合的篇章倾向性打分

鉴于以上分析，本文使用了两个模块判断篇章的倾向性打分。针对第一个影响因素，本文把每个分值作为一个类别，使用文本分类的方法识别篇章的倾向性打分；针对第二个影响因素，本文首先计算评论文本的褒义句比例，继而通过计算待测文本与训练数据的情感相似度判断篇章的倾向性打分。最后，我们将上述两个结果的平均值作为最终的倾向性打分。

4.2.1. 基于文本分类的篇章倾向性打分

受 Pang[6] 的启发，本小节使用有指导的方法判断篇章的倾向性打分。有指导的学习方法包含两个必备的因素，即标准的训练语料和特征提取方案。因此，本小节分别从训练语料的自动获取和特征提取两方面展开。

(1) 训练语料自动获取

通过对 COAE2012 任务三的训练语料进行统计分析,我们发现各个打分的篇章个数存在很大的不平衡性,具体统计信息如表 6 所示。其中数码领域的 3000 篇训练数据中,4 分的评论有 867 篇,5 分的评论有 1231 篇,4 分+5 分共占了所有训练数据的 69.9%;类似地,在汽车领域中,3 分的评论共有 2155 条,占了全部 3000 篇评论的 71.8%。通过实验我们发现,如此严重的数据不平衡性会严重影响倾向性打分的识别效果。因此,为了避免数据不平衡带来的问题,本文决定从互联网上自动挖掘大规模的平衡训练语料。

表 6 COAE2012 篇章级训练语料统计信息

领域	评论个数					
	1 分	2 分	3 分	4 分	5 分	总数
数码领域	311	153	438	867	1231	3000
汽车领域	41	71	2155	440	293	3000

互联网上存在大量的用户评论数据。以表 3 中的京东评论数据为例,其包含用户针对产品的打分以及总体评价;类似地,新浪汽车⁷的“口碑”板块包含了大量网民对汽车的评论,其数据格式和京东评论数据类似,同样包含用户提供的打分信息和评论信息。因此,针对 COAE2012 的数码领域和汽车领域,我们分别从京东评论数据和新浪口碑中自动挖掘训练语料。经过长度过滤、去重和语料平衡之后,两个领域的训练语料情况如表 7 所示,其中第三列表示语料规模。以数码领域为例,我们一共获取了京东评论数据 17435 条,其中 1 分~5 分范围内每个打分的评论数据都包含 3487 条。

表 7 篇章级自动获取语料的统计信息

领域	数据来源	数据规模
数码领域	京东	3487 * 5 = 17435
汽车领域	新浪汽车	3175 * 5 = 15875

(2) 特征提取

针对篇章级倾向性打分任务,大多数学者(如 Pang[6])使用一元语法特征作为篇章的特征表示。但是在本文中,通过在开发集上使用特征选择,我们最终使用了 3 类特征,具体的特征表示和样例如表 8 所示:

表 8 篇章级特征抽取模板

特征类别	文本样例	特征样例
unigram+否定词+程度副词	非常不喜欢这个手机	非常_negation_喜欢
转折、让步关系连词	虽然很喜欢这个手机	Null
情感 bigram	有点喜欢这个手机	有点_喜欢

表 8 第一行以一元语法特征为主,同时判断每个词语在窗口[-2,0)内是否包含否定词或程度副词,如果包含则将该词语和否定词、程度副词结合起来共同组成特征。表 8 第二行使用了转折关系和让步关系特征。具体地,如果某子句以让步关系连词开始(前两个单词内包含让步关系连词,转折关系连词类似),则此子句特征为空;如果某子句以转折关系连词开始,则相邻前一子句的特征为空。表 8 第三行使用了 bigram 特征,与传统的 bigram 不同,本文仅适用包含情感词语的 bigram 作为特征。

特征提取完成后,我们分别尝试了 LibSVM、SVM Regression 和最大熵分类器,通过分析开发集上的实验结果,我们最终使用了最大熵分类器,并用其训练五元分类器判断文本的倾向性打分。

⁷ <http://data.auto.sina.com.cn/>

4.2.2. 基于情感相似度的篇章倾向性打分

受 Pang[6]的启发, 本小节使用情感相似度计算篇章的倾向性打分。主要内容包括 (1) 基于褒义句比例的篇章情感信息 (2) 基于相似度计算的篇章倾向性打分。

(1) 基于褒义句比例的篇章情感表示

借鉴 Pang[6], 我们认为褒义句的比例可以很好地描述篇章的情感信息。如果一篇评论中有超过 90% 的句子都是褒义句, 整个评论更可能具有较高的倾向性打分; 相反, 如果褒义句比例小于 10%, 则整个评论更可能具有较低的倾向性打分。

具体地, 对待测的评论文本 T , 经过分句、句子极性识别后共包含褒义句 pc 个, 贬义句 nc 个, 中性句 oc 个, 褒义句的比例 PSP 如公式 (1) 所示。值得强调的是, 本文假设中性句并不影响篇章的总体情感倾向, 因此在计算褒义句比例时并没有考虑中性句的个数 oc 。

$$PSP(T) = \frac{pc}{pc + nc} \quad (1)$$

在公式 (1) 的基础上, 我们构建情感向量 $A = (PSP(T), 1 - PSP(T))$ 表示每个文本, 并将该向量提供给后续的相似度计算模块, 用来计算篇章的倾向性打分。

(2) 基于相似度计算的篇章倾向性打分

我们把 1 分~5 分看作 5 个不同的倾向性类别, 本小节的方法基于以下假设: 两个评论文本的相似度越大, 他们具有相同倾向性打分的概率就越大。鉴于此, 我们从训练数据中挑选出与待测文本相似度最高的 K 个文本, 并从这 K 个文本中通过投票的方式选出数目最多的倾向性类别, 作为待测文本的倾向性打分。

通过在开发集上尝试不同的 K 值, 我们最终选择 $K=15$ 。

由于篇章的情感向量只有 2 维, 并且每维向量的值域都在 $[0,1]$, 因此我们选择余弦相似度计算情感向量之间的相似度, 余弦相似度的计算方案如公式 (2) 所示。

$$\text{sim}(A_1, A_2) = \frac{\sum_{i=1}^2 A_{1i} \times A_{2i}}{\sqrt{\sum_{i=1}^2 (A_{1i})^2} \times \sqrt{\sum_{i=1}^2 (A_{2i})^2}} \quad (2)$$

5. 评测结果及分析

5.1. 句子级倾向性分析评测结果

句子级倾向性分析的评价指标包括准确率(P)、召回率(R)、F 值(F1)、精确率(Accuracy)、宏平均值和微平均值[13]。本次评测由 COAE 提供测试数据的标准答案, 在此基础上给出评测指标。汽车领域和数码领域的正例 F 值、负例 F 值、中性 F 值和精确率如表 9 所示; 总体的宏平均和微平均评测指标如表 10 所示。其中第一行 HITIRSYS 是本文所提交的系统, 第二行 Average 是所有提交结果的平均值, 第三行 Median 代表所有提交结果中处于中间排名的队伍, 第四行 Best 是所有队伍的最好结果。

表 9 句子级倾向性分析评测结果 (各领域)

任务一	汽车领域				数码领域			
	正-F1	中-F1	负-F1	Accuracy	正-F1	中-F1	负-F1	Accuracy
HITIRSYS	0.617	0.464	0.811	0.698	0.784	0.418	0.851	0.772
Average	0.405	0.383	0.660	0.562	0.579	0.271	0.685	0.597
median	0.383	0.407	0.702	0.553	0.548	0.322	0.701	0.581
Best	0.693	0.644	0.834	0.754	0.807	0.473	0.863	0.787

表 10 句子级倾向性分析评测结果（总体）

任务一	宏平均				微平均			
	正-F1	中-F1	负-F1	Accuracy	正-F1	中-F1	负-F1	Accuracy
HITIRSYS	0.706	0.441	0.831	0.735	0.728	0.447	0.830	0.735
Average	0.500	0.333	0.674	0.580	0.521	0.348	0.673	0.580
median	0.469	0.356	0.707	0.579	0.490	0.397	0.707	0.579
Best	0.730	0.557	0.848	0.749	0.739	0.573	0.848	0.749

从表 9 中可以看出，本系统在汽车和数码领域的各项指标都大幅超过 Average 水平，数码领域的正-F1、负-F1 和 Accuracy 均接近 Best 水平；从表 10 中可以看出，本系统的微平均、宏平均指标都接近 Best 水平，证明了本方法的有效性。同时我们发现本文的数码领域的实验结果要优于汽车领域，我们分析可能是因为两个领域中我们都使用京东的评论数据作为训练语料，而该语料和数码领域的语言表达更相似，所以在数码领域上取得了更好的实验效果。此外，表 9 中正例和负例的识别效果要优于中性的识别效果，这说明了本文仅使用分类器的输出阈值判断句子是否中性尚有提升的空间。

5.2. 篇章级倾向性打分评测结果

篇章级倾向性打分任务的评价指标是均方最小错误率（mean square error, mse）和精确率（Accuracy），均方最小错误率的计算公式如公式（4）所示。

$$mse = \frac{1}{n} \sum_{i=1}^n (gold(i) - predict(i))^2 \quad (4)$$

评测结果如表 11 所示，各行的含义与表 9、表 10 相同，在此不再赘述。

通过表 11 可以看出，本系统在汽车领域和数码领域均取得了较好的效果，证明了本方法的有效性。此外，汽车领域的实验结果要优于数码领域，我们分析可能是因为数码领域的训练语料来自于京东评论数据，其包含用户对多种商品的评论文本；而汽车领域的训练语料来自新浪汽车的评论数据。和数码领域相比，汽车领域的训练语料和测试数据具有更相似的语言现象，因此汽车领域取得了更好的实验结果。

表 11 篇章级倾向性打分评测结果

任务三	汽车领域		数码领域	
	MSE	Accuracy	MSE	Accuracy
HITIRSYS	0.467	0.750	1.133	0.479
Average	1.229	0.489	2.101	0.368
median	0.928	0.463	1.946	0.345
Best	0.301	0.832	0.759	0.691

6. 小结

哈工大社会计算与信息检索研究中心 HITIRSYS 情感分析系统参与了 COAE2012 的句子级倾向性分析任务（任务 1）和篇章级倾向性打分任务（任务 3）。在句子级倾向性分析任务中，本系统结合子句极性识别和句子主干分析识别句子的倾向性，实验结果表明本系统的性能接近所有评测单位最好水平，证明了本方法的有效性；在篇章级倾向性打分任务中，本系统分别使用基于文本分类的方法和基于情感相似度的方法计算篇章的倾向性打分，最后将二者的结果融合作为最终的倾向性打分，本系统在数码和汽车两个领域均中取得了较好的成绩。

- [1] 赵妍妍, 秦兵, 刘挺, 文本情感分析. 软件学报, 2010. 21(8): p. 1834-1848.
- [2] Pang, B. and L. Lee, Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2008.2(1-2): p. 1-135.
- [3] Liu, B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 2012.5(1):1–167.
- [4] Pang, B., L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. 2002: Association for Computational Linguistics.
- [5] Turney, P.D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. 2002: Association for Computational Linguistics.
- [6] Pang, B. and L. Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. 2005. Association for Computational Linguistics.
- [7] Yu, H. and V. Hatzivassiloglou. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. 2003: ACL.
- [8] Wilson, T., J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. 2005: Association for Computational Linguistics.
- [9] Hatzivassiloglou, V. and K.R. McKeown. Predicting the semantic orientation of adjectives. 1997: Association for Computational Linguistics.
- [10] Zhou, L., et al. Unsupervised discovery of discourse relations for eliminating intra-sentence polarity ambiguities. 2011: Association for Computational Linguistics.
- [11] Esuli, A. and Sebastiani, F. (2007). Pageranking wordnet synsets: An application to opinion mining. In *Annual meeting-association for computational linguistics*, volume 45, page 424.
- [12] Che, W., Li, Z., and Liu, T. (2010). Ltp: A Chinese language technology platform. In *Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations*, pages 13–16. Association for Computational Linguistics.
- [13] Manning, C., Raghavan, P., and Schütze, H. (2008). *Introduction to information retrieval*, volume 1. Cambridge University Press Cambridge.
- [14] Zhao, Y., Qin, B., and Liu, T. (2012). Collocation polarity disambiguation using web-based pseudo contexts. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 160–170.
- [15] Velikovich, L, Sasha B, Kerry H, and Ryan M. The viability of web-derived polarity lexicons. In *Proceedings of Annual Conference of the North American chapter of the Association for Computational Linguistics*. 2010.