

文章编号: 1004-4353(2019)02-0153-08

# 基于情感词典的课程评论情感分析

胡荣, 崔荣一, 赵亚慧\*

( 延边大学 工学院, 吉林 延吉 133002 )

**摘要:** 采用极性计算方法, 对 MOOCs 上的课程评论进行情感分析. 首先, 从 MOOCs 上搜集课程评论, 并对所有评论按学习者、授课方式、课件、平台和视频这 5 个属性进行分类. 其次, 基于情感词典对各属性评论进行极性计算, 得到各属性的正向评论、负向评论、中性评论和无效评论, 将正向评论和负向评论作用于课程评分. 最后, 分别计算课程评论与 5 个属性在网页中的共现频率, 并将单个共现频率与总共现频率之比作为评分的权重, 从教育者角度、学习者角度和平台角度对课程进行评分. 将本文方法应用于某高校的课程分析中, 结果表明本文方法所得的结果客观、合理.

**关键词:** 情感分析; 课程评论; 属性分类; 极性计算

**中图分类号:** TP391

**文献标识码:** A

## Sentiment analysis of curriculum review based on sentiment dictionary

HU Rong, CUI Rongyi, ZHAO Yahui\*

( College of Engineering, Yanbian University, Yanji 133002, China )

**Abstract:** A method of polarity calculation was used to analyze the sentiment of curriculum reviews on MOOCs. First, course reviews were collected from MOOCs, and all reviews were classified according to five attributes: learners, teaching methods, courseware, platform and video. Then, based on the sentiment dictionary, we calculated the polarity of each attribute comment, and got the positive, negative, neutral and invalid comment of each attribute. The positive and negative comment acted on the course score. Finally, the co-occurrence frequencies of curriculum reviews and five attributes in web pages were calculated, and the ratio of single co-occurrence frequency to total co-occurrence frequency was taken as the weight of the scoring. The curriculum was scored from the perspectives of educators, learners and platforms. After applying this method to the educational administration department of a university, the analysis shows that the results obtained by this method are objective and reasonable.

**Keywords:** sentiment analysis; curriculum review; attribute classification; polarity calculation

## 0 引言

情感分析是指通过机器学习或者自然语言处理技术, 从文本中分析出人们对实体或属性所表达的观点、情感、评价、态度和情绪<sup>[1]</sup>. 目前, 情感分析被广泛应用于消费产品分析、社会舆情监控、股票市场预测和客户反馈跟踪等领域. 其主要的分析方法为基于语义的方法和基于机器学习的方法. 基于语义的方法主要是通过情感词典计算文本的情感值, 以此为依据确定文本的情感倾向<sup>[2]</sup>. 基于机器学习的方法主要是通过提取文本中的特征, 利用机器学习中的分类算法预测文本的情感倾向<sup>[3]</sup>. 近年来, 有关情感

分析的相关研究取得了一些成果.例如:张磊<sup>[4]</sup>提出了一种基于主题的评价信息分类方法,在评价信息明确的情况下该方法可以取得较好的分类效果,但存在少数评价信息和主题不相关的情况.杨伊态<sup>[5]</sup>提出了一种基于半监督学习的细粒度属性获取模型对评价信息进行属性分类的方法,该分类算法获取细粒度属性集和分类的效率虽然较好,但过于依赖数据的质量.刘磊等<sup>[6]</sup>提出了一种基于集成学习的情感模糊计算分类方法,该方法的分类效果优于单分类器,但需要构造极性量化情感字典和量化情感词典.

截至 2016 年,大规模在线课程(Massive Open Online Course,简称 MOOCs)平台上已汇集了 700 多所学校的 6850 多门课程,约有 58 万学习者利用该平台进行了学习<sup>[7]</sup>.对 MOOCs 评论进行情感分析不仅有助于教育者了解学生的需要,学习者选择适合自己学习的课程,还有助于管理者进行平台的建设和改善.目前,针对 MOOCs 评论进行文本情感分析的研究较少,例如冯君<sup>[8]</sup>提出了一种基于规则的情感单元提取算法,该方法虽然能够正确提取在线文本的情感单元,但不能有效提取不在规则内的内容.基于此,本文提出一种基于属性权重对 MOOCs 评论进行情感分析的方法,并通过某高校教务部门的应用,验证本文方法的合理性和准确性.

1 评论的属性分类

为了对课程评论进行情感分析,首先需要根据情感分析的目的对评论文本进行属性分类.本文经过调研分析,将课件、语速、上课方式、时间安排、学习感受和接受度作为课程指标.综合 MOOCs 评论的显式评价对象抽取结果与调研课程指标,制定出 5 个适用于 MOOCs 课程评论的属性,分别为视频(V)、课件(C)、学习者(L)、平台(P)和授课方式(M).通过分类器训练,将每个课程评论划分为概率最大的所属类别.另外,为了保证属性分类的正确性,引入信息熵,即通过计算文本的信息熵,判断某一评论属于某一类别是否具有合理性,以及是否需要实行软分类.

1.1 文本表示

Word2vec 模型是 Mikolov 提出的一种基于大规模语料库计算词向量的工具<sup>[9]</sup>,它包括 CBOW 和 Skip-gram 2 个训练模型.这 2 个模型结构相似,都包括输入层、隐含层和输出层<sup>[10]</sup>,如图 1 所示.因 CBOW 模型能够精确地度量词与词之间的关系,挖掘词与词之间的联系,因此本文选用该模型作为文本的表示模型.CBOW 模型的目标是最大化平均对数似然函数  $L$ :

$$L = \frac{1}{T} \sum_{t=k}^{T-K} \log p(w_t \mid w_{t-k}, \cdots, w_{t+k}).$$

(1)

公式(1)表示在上下文为  $w_{t-k}, \cdots, w_{t+k}$  时,模型对当前词  $w_t$  进行预测的可能性.

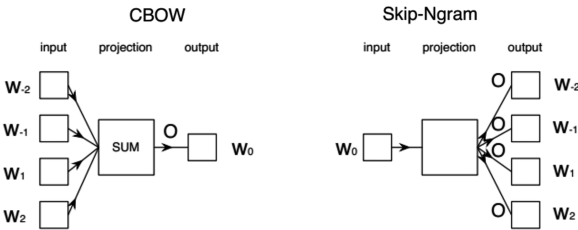


图 1 CBOW 和 Skip-gram 模型结构

1.2 文本分类

应用 TextCNN 作为分类器对课程评论进行分类,以此得到评价句所对应的属性类别. Textcnw 模型主要包括 4 个层面:数据预处理层、卷积层、池化层和 softmax 分类层,层与层之间的关系是底层为上层提供服务<sup>[11]</sup>.因 TextCNN 结构中的卷积层对信息敏感,连续的卷积池化层可以获取局部的敏感信息,加之其网络结构高效、简单且降维速度快,因此本文选取卷积神经网络进行文本分类<sup>[12]</sup>.使用卷积神经网络模型进行分类的步骤如下:

1) 将评论语句向量化.利用词嵌入的方法得到每个词  $x_i$  的向量表示,向量的维度为  $K$ .对于长度为

$n$  的句子,其向量表示为

$$\mathbf{x}_{1:n} = \mathbf{x}_1 \oplus \mathbf{x}_2 \oplus \cdots \oplus \mathbf{x}_n.$$

(2)

用矩阵  $\mathbf{T}(n \times k)$  表示评论短文本,其中  $n$  是文本的长度,  $k$  是词嵌入的维度.

2) 将带有多个卷积核的卷积层进行特征提取及特征映射. 利用卷积核操作进行特征提取,映射得到特征输出,为:

$$\mathbf{c}_i = f(\mathbf{w} \cdot \mathbf{x}_{i:i+h-1} + b).$$

(3)

其中  $\mathbf{w}$  表示卷积核,对于包含  $h$  个单词的滑动窗口可表示为  $\mathbf{x}_{i:i+h-1}$ . 每次通过滑动  $h$  个单词,得到特征  $\mathbf{c}_i$ , 每个卷积核对句子进行  $h$  个单词的窗口滑动,分别得到对应的特征映射,为:

$$\mathbf{c} = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_{n-h+1}].$$

(4)

3) 选择文本特征. 采用最大池化处理进行特征选择,即在每个卷积核得到的特征映射向量中分别挑选出最大值,其目的是提取出每个特征映射中最重要的信息. 最大池化处理公式为

$$\mathbf{C}^* = \max_i \{\mathbf{c}_i\}.$$

(5)

4) 利用全连接层获得类别概率. 将池化层得到的特征通过全连接 softmax 层后,可得到每个类别的概率  $p$ . 全连接的输入是特征向量,输出的是判断的类别概率,其计算公式为

$$p = \text{softmax}(\mathbf{w}_c \mathbf{v} + b_c).$$

(6)

1.3 信息熵

熵值  $H$  的计算公式为

$$H = - \sum_i^n p_i \log p_i.$$

(7)

熵值越小,确定性越好,实例的分类可信度越高<sup>[13]</sup>.

2 基于情感词典的极性计算

为了确定每个评论文本的情感倾向性,需要对每个文本进行基于情感词典的极性计算. 首先通过极性计算可以统计正向文本、负向文本、中评文本和无效文本在某一课程评论中(按授课方式、视频、平台、课件和学习者 5 个属性)所占的比例,然后从教育者角度、学习者角度和平台 3 个方面对该课程进行评分.

2.1 课程评论情感词典的构建

目前我国还没有通用的情感词典,相关研究大多是在某一情感词典(如 HowNet 中英文情感词典、大连理工大学的中文情感词汇本体库、台湾大学的 NTUSD 和知网的情感分析用词语集)的基础上进行融合和扩展<sup>[14]</sup>. 由于这些词典都是根据中文的正式用语制定的,难以完全适用于网络文本的情感分析,因此本文基于知网的情感分析用词语集<sup>[15]</sup>,从课程评论语料中添加部分领域的情感词,以此构建 MOOCs 课程评论的情感词典. 部分课程评论领域情感倾向的词语示例如表 1 所示.

表 1 课程评论基础情感词表

	词语的情感极性	
	褒义词	贬义词
数量	5 309	4 848
部分示例	效果、干货、质优、安全、 非常好、释然、切切实实、入木三分	差劲、漏洞、灌输式、希望、 填鸭式、稀碎、矛盾、凌乱

2.2 评论文本的极性计算

本文依据否定词词典、副词词典和课程评论基础情感词典中的情感词的极性来判断文本的极性. 将倾向积极、肯定和赞扬等正面的文本看作正例,将倾向消极、否定和批评等负面的文本看作负例<sup>[16]</sup>. 如果一个评论文本中不含情感倾向的词语,则视为中性文本,不作为情感倾向分析的对象.

评论文本的极性计算过程如下:

Input 分词后的测试文档、情感词汇表、程度副词表、否定词表

output 结果\_正向文档、结果\_负向文档、结果\_中性文档、结果\_无法判断

file=open("test.txt")

for each sentence in file do

  for each word in current sentence do

    if word exists in emtion\_pos\_dict then

      score=正向情感词典中对应的情感强度

      pos\_word\_num+=1

    else if word exists in emtion\_neg\_dict then

      score=负向情感词典中对应的情感强度

      neg\_word\_num+=1

    else then

      continue

    endif

  if word 前 3 个词存在否定词(可叠加) then

    score \*= -1

  endif

  if word 前 3 个词存在程度副词(可叠加) then

    score \*= 程度副词表中对应的强度值

  endif

  if word exists in emtion\_pos\_dict then

    result\_pos+=score

  else if word exists in emtion\_neg\_dict then

    result\_neg+=score

  endif

endfor

if pos\_num==0 and neg\_num==0 then

  将当前 sentence 输出到“结果\_无法判断.txt”中

else then

  result\_total=(result\_pos+result\_neg)/(pos\_num+neg\_num)

  if pos\_num>neg\_num then

    result=(result\_pos/pos\_num+result\_total)/2

  else if pos\_num<neg\_num then

    result=(result\_neg/neg\_num+result\_total)/2

  else if pos\_num==neg\_num then

    result=result\_total

  endif

  if result>0 then

    将当前 sentence 输出到“结果\_正向.txt”中

  else if result<0 then

```
    将当前 sentence 输出到“结果_负向.txt”中
else if result==0 then
    将当前 sentence 输出到“结果_中性.txt”中
endif
endif
endfor
```

2.3 基于属性权重的情感分析

利用属性权重进行情感分析时,因研究角度不同,同一属性所占的权重会有所不同.因此,在对某一课程进行评价时,需要从教育者角度、学习者角度和平台角度,依据不同属性文本所占的权重进行评分.

首先,根据不同情感倾向文本的比例和极性计算的正确率计算出所有属性的得分.单个课程评论属性的计算公式为

$$Score = positive * acc\_pos + (-negative * acc\_neg).$$
 (8)

其中 *positive* 表示某一属性的正向评论文本比例, *negative* 表示一个属性的负向评论文本比例, *acc\_pos* 表示计算 1 000 条正向评论极性的准确率, *acc\_neg* 表示计算 1 000 条负向评论极性的准确率.

其次,确定各属性所占的权重.本文基于微软新推出的搜索引擎 Bing,通过限定领域构造术语的联合检索式,检索课程评论分别与视频、授课方式、学习者、平台和课件在 MOOCs 领域的检索的命中数<sup>[17]</sup>.属性权重的确定是基于各属性与课程评论在 Bing 上关于 MOOCs 领域所检索出来的命中数与与其相关的属性检索出的总命中数之比.各属性的检索式为(“课程评论”NEAR“属性”)AND (MOOCs).

从教育者角度,课程评分基于视频、授课方式、学习者、课件 4 个属性进行计算,计算公式为

$$Score\_edu = a_1 * S\_vedio + a_2 * S\_method + a_3 * S\_learner + a_4 * S\_courseware.$$
 (9)

其中 *a*<sub>1</sub> 表示视频在教育者角度所占的权重, *a*<sub>2</sub> 表示授课方式在教育者角度所占的权重, *a*<sub>3</sub> 表示学习者在教育者角度所占的权重, *a*<sub>4</sub> 表示课件在教育者角度所占的权重.

从学习者角度,课程评分基于视频、授课方式、课件 3 个属性进行计算,计算公式为

$$Score\_learner = b_1 * S\_vedio + b_2 * S\_method + b_3 * S\_courseware.$$
 (10)

其中 *b*<sub>1</sub> 表示视频在学习者角度所占的权重, *b*<sub>2</sub> 表示授课方式在学习者角度所占的权重, *b*<sub>3</sub> 表示课件在学习者角度所占的权重.

从平台角度,课程评分只需要考虑平台这一个属性即可,其计算公式为

$$Score\_plat = positive * acc\_pos + (-negative * acc\_neg).$$
 (11)

3 实验及分析

3.1 实验数据的采集

实验数据为 2017 年和 2018 年已经结课的所有 MOOCs 课程的学生评论(人工收集).在收集的数据中去除重复发表的样本和其他无效的数据样本,并将处理过的课程评论数据形成一个总的数据集 Data.将评论数大于 2 000 的课程单独存放.部分原始课程评论数据如表 2 所示.

表 2 部分原始文本数据

序号	评 论
1	作业的完成度比较好,考试前能复习的很好,上课不会跟不上,学习感受还不错,学得很快乐
2	课件制作很用心,老师课前也会用心备课
3	视频画质良好,视频添加字幕就更好了
4	老师授课方式多样,能够带动课堂气氛,让每个小组合作学习完成课程任务

3.2 实验过程

1)语料预处理. 首先利用结巴分词对数据集(18 000 条评论)进行分词,去停用词;然后对数据集进行 word2vec 词向量的训练,获得词向量模型;最后使用 Stanford Corenlp 对 5 500 条课程评论进行分词和去停用词处理,并将预处理后的评论作为实验的训练数据和测试数据.

对评论句属性类别进行人工划分时,当属性类别判定出现不一致时,以少数服从多数确定评论所属类别. 实验数据如表 3 所示.

表 3 属性类别数据

数据量	属性类别				
	授课方式	平台	课件	视频	学习者
总评论数量	1 100	1 100	1 100	1 100	1 100
训练数据量	700	700	700	700	700
测试数据量	400	400	400	400	400

2)文本表示. 预处理完成实验语料后,将每个词语作为一个独立的单元用以 word2vec 输入. 本文经过多次实验得出 word2vec 的最优训练参数:词向量维度为 50,最小窗口为 1,迭代次数为 15, sample 值为 0.001.

3)分类器训练. 由于 TextCNN 的输入是矩阵多维向量,因此需要对训练评论文本进行建模. 定义每个评论文本为  $T(n \times k) = w_1 \otimes w_2 \otimes w_3 \otimes \cdots \otimes w_n$ , 其中  $T(n \times k)$  是评论文本的矩阵表示,  $n$  是文本的长度,  $k$  是词嵌入的维度. 实验中每个属性各选取 700 条评论,共计 3 500 条评论. 分类器训练时,  $k$  取 50,  $n$  取 3 500.

4)分类器测试. 每个属性选取 400 条评论,共计 2 000 条课程评论. 测试时先对评论进行预处理(用词嵌入表示评论),然后再输入 TextCNN 模型进行测试,以此得出评论的分类结果以及每一个测试文本所对应的信息熵. 对部分分类结果以及对应的信息熵进行人工检测,结果表明以信息熵作为分类的参考依据是可靠的.

5)极性计算. 根据上述提出的极性计算算法对每一类属性评论文本进行情感极性计算,得到每一类属性文本的正向文本、负向文本、中性文本和无效文本的比例.

6)计算课程得分. 根据各属性和课程评论在 Bing 上共同检索出的命中数与全部属性检索的总命中数之比,确定各属性在 3 个研究角度中所占的权重. 依据公式(8)计算出课程各属性的得分,依据公式(9)计算课程在教育者角度的得分,依据公式(10)计算课程在学习者角度的得分,依据公式(11)计算课程在平台角度的得分. 最后将不同角度(教育者角度、学习者角度和平台角度)计算出的分数转化为 5 分制的分值.

3.3 实验结果及分析

选取 3 500 条课程评论作为训练数据,2 000 条课程评论作为测试语料,并依据课程属性(授课方式、课件、平台、学习者、视频)对评论进行属性分类. 属性分类结果显示,分类的正确率为 89.9%. 为了进一步证明本文分类方法的合理性和可靠性,本文对测试语料(2 000 条)的信息熵分布情况进行实验验证,所得信息熵分布情况见表 4. 由表 4 可知,96.25% 的文本信息熵小于 1. 由公式(7)可知,当某个文本属于某一个类的概率大于 0.5 时,其信息熵小于 1. 由此说明,本文的属性分类方法是合理的,可作为后续课程评论评分(3 个研究角度)的依据之一.

表 4 属性分类的熵分布情况

熵范围	文本数	所占比例/%
$H \geq 2.0$	1	0.05
$2.0 > H \geq 1.58$	16	0.80
$1.58 > H \geq 1.0$	58	2.90
$1.0 > H \geq 0$	1 925	96.25



在收集的评论中,人工挑选出 1 000 条正向评论和 1 000 条负向评论,然后基于情感词典计算课程评论极性的正确率,结果见表 5。由表 5 结果可以看出,本文方法对课程评论极性判断准确、有效,因此本文方法可作为判断未知评论极性的方法。

表 5 评论的极性计算结果						%
文本类型	正确率	错误率	文本倾向所占比例			
			正向文本	负向文本	中性文本	无效文本
正向评论	89.1	10.9	89.1	4.2	2.1	4.6
负向评论	88.4	11.6	5.7	88.4	4.8	1.1

选取 python 课程的 2 000 条课程评论进行情感分析。情感分析时首先进行属性分类,然后计算各文本极性。属性分类及极性计算结果如表 6 所示。

表 6 python 课程的属性分类及其极性计算结果					
文本属性	评论数量/条	文本倾向所占比例/%			
		正向文本	负向文本	中性文本	无效文本
视频	388	87.37	9.54	2.32	0.77
授课方式	418	86.39	10.39	3.22	0.00
学习者	432	71.76	21.99	4.86	1.39
课件	391	87.72	9.21	1.79	1.28
平台	371	73.05	17.52	7.55	1.88

分别从教育者、学习者和平台 3 个研究角度,对 python 课程进行课程评分。以(“课程评论”NEAR “视频”)AND(MOOCs)、(“课程评论”NEAR “授课方式”)AND(MOOCs)、(“课程评论”NEAR “学习者”)AND(MOOCs)、(“课程评论”NEAR “课件”)AND(MOOCs)、(“课程评论”NEAR “平台”)AND (MOOCs)分别在 Bing 上检索,得到的检索命中数分别为 5 460、4 610、3 290、13 700、5 270。

教育者角度的各属性所占的权重如表 7 所示。表 7 中的各属性权重值表示在 Bing 上检索视频和课程评论共同出现的命中数与视频、授课方式、学习者和课件分别和课程评论检索出的命中数的和之比,例如视频的权重计算为: $5\,460/(5\,460+4\,610+3\,290+13\,700)=0.20$ 。根据公式(8)和公式(9),计算出教育者角度的百分制得分为 69.45,5 分制得分为 3。

学习者角度的各属性所占的权重如表 8 所示。表 8 中的各属性权重值表示在 Bing 上检索视频和课程评论共同出现的命中数与视频、授课方式和课件分别与课程评论检索出的命中数的和之比,例如视频的权重计算为: $5\,460/(5\,460+4\,610+13\,700)=0.23$ 。根据公式(8)和公式(10),计算出学习者角度的百分制得分为 69.45,5 分制得分为 3。

表 7 教育者角度的各属性所占权重

属性	视频	授课方式	学习者	课件
权重值	0.20	0.17	0.12	0.51

表 8 学习者角度的各属性所占权重

属性	视频	授课方式	课件
权重值	0.23	0.19	0.58

平台角度的评分主要取决于平台。根据表 6 和公式(11),计算得出平台角度的百分制得分为 46.93,5 分制得分为 2。

各角度(教育者、学习者、平台)的评分值与 python 课程(在 MOOCs 平台上)的评价结果的对应关系如表 9 所示。根据上述不同角度计算出的 python 课程得分,

表 9 评分值与评价结果的对应关系

评分值	评价结果
$\geq 4$	优秀
3	良好
2	中等
$\leq 1$	较差

可得到以下评价结论:从教育者角度(3分)和学习者角看(3分)来看,该课程评价良好;从平台角度来看(2分),课程平台建设有待进一步提高。

## 4 结论

本文以授课方式、视频、学习者、平台和课件 5 个属性对 MOOCs 课程评论进行情感分析。在分析过程中,利用卷积神经网络进行属性分类,并引入信息熵判定分类属性。经实验测定,本文属性分类的正确率是 89.9%。在情感极性计算过程中,正向评论的正确率达到 89.1%,负向评论的正确率达到 88.4%,验证了课程评论情感词典的有效性和实用性。在课程总体评价方面,利用属性在 Web 网页中的共现频率确定了 5 个属性的权重,并从教育者、学习者和平台 3 个角度进行了评分计算。将本文提出的情感分析方法应用于某高校 MOOCs 课程评论,所得结果客观、合理。在今后的研究中,我们拟将机器学习方法与情感词典相结合进行极性计算,以进一步提高极性计算的正确率和课程评价的准确性。

## 参考文献:

- [1] 刘兵. 情感分析:挖掘观点、情感和情绪[M]. 刘康,赵军,译. 北京:机械工业出版社,2017:1-2.
- [2] 林斌. 基于语义技术的中文信息情感分析方法研究[D]. 哈尔滨:哈尔滨工业大学,2006.
- [3] PEI J, ZHANG C, HUANG D, et al. Combining word embedding and semantic lexicon for Chinese word similarity computation[C]//Natural Language Understanding and Intelligence Applications. Springer International Publishing, 2016:766-777.
- [4] 张磊. 细粒度意见挖掘研究[D]. 南京:南京师范大学,2016.
- [5] 杨伊态. 在线评论细粒度属性情感分类模型研究[D]. 武汉:武汉纺织大学,2018.
- [6] 刘磊,李学勇,黄永峰. 基于集成学习的情感模糊计算分类方法[J]. 计算机工程与设计,2018;39(7):1872-1876.
- [7] 李志民. 信息技术发展与教育变革[J]. 中国教育,2015(9):39-43.
- [8] 冯君. 基于条件随机场的情感分析模型在 MOOCs 评论文本分析中的应用研究[D]. 武汉:华中师范大学,2017.
- [9] 陈雷. 面向股票评论的情感分析系统研究与实现[D]. 杭州:浙江工商大学,2017.
- [10] 周练. Word2vec 的工作原理及应用探究[J]. 科技情报开发与经济,2015(2):145-148.
- [11] 刘娇. 基于深度学习的多语种短文本分类方法的研究[D]. 延吉:延边大学,2018.
- [12] 王儒. 基于卷积神经网络的短文本表示与分类研究[D]. 济南:山东师范大学,2018.
- [13] 施寒潇. 细粒度情感分析研究[D]. 苏州:苏州大学,2013.
- [14] 朱军,刘嘉勇,张腾飞,等. 基于情感词典和集成学习的情感极性分类方法[J]. 计算机应用,2018(38):95-98.
- [15] 王科,夏睿. 情感词典自动构建方法综述[J]. 自动化学报,2016;42(4):495-511.
- [16] 吴金源,冀俊忠,赵学武,等. 基于特征选择技术的情感词权重计算[J]. 北京工业大学学报,2016(42):142-151.
- [17] 徐健. 术语相似度计算方法研究[M]. 广州:中山大学出版社,2012:83-97.