




领域情感词汇的自动构建 语义消歧

王燕燕¹ · 富联阴¹ · 刘建波¹ · 马可² 

收到日期: 2019年5月13日/修订日期: 2020年2月14日/接受日期: 2020年5月5日/

在线发布时间: 2020年5月22日

©施普林格科学+商业媒体有限责任公司, 施普林格自然2020的一部分

摘要

情感词汇用于判断词汇的情感, 在情感分析中起着重要的作用。现有的情感词汇忽略了词汇在不同语境下的情感模糊性, 只为情感分配正极性或负极性。本文提出了一种自动构建领域特定情感词汇 (SDS-lex) 的方法, 利用我们改进的TF-IDF算法 (ITF-IDF), 不仅包含了现有词汇, 还包含了语料库中的情感信息。ITF-IDF算法通过考虑单词的重要性和不同情绪倾向标记的语料库中不同词性 (POS) 的分布来计算单词的情绪。在真实数据集上的实验表明, 我们构建的词汇改善了情感歧义, 在执行文本情感分类任务时, 在覆盖范围和准确性方面优于许多现有的词汇。

关键词改进TF-IDF情绪词汇情绪分类Word感觉消除歧义

1介绍

随着社交网络[5, 7, 13, 20, 33, 41]的不断兴起, 情绪分析 (SA) 作为从文本中获取信息的重要方法而受到越来越多的关注。基于词汇的[33]方法和基于机器学习的[12, 28]方法是情绪分析的两种主要方法。特别是基于词汇的方法, 为词汇分配情感值, 仍然是其简单性[34]的主流。

早期对SA的研究主要集中在构建一般情感词汇[6, 18, 23]上, 其中SentiWordNet [2, 8]是最著名的一个。它通过世界知识库WordNet为每个同步集的概念分配积极和消极的情绪值,

✉ Fulian Yin
yinfulian@cuc.edu.cn

¹中国通信大学信息工程研究所, 北京100024

²工业和应用数学实验室, 约克大学, 多伦多, M3J 1P3, 加拿大

每个单词的词都有不同的情感表达。此外，在2005年，Wilson等人。[39]为推特的特定语料库构建了一个主观词汇，称为MPQA主观词汇，它用情感极性、强度和词性来标记词汇。这些情感词汇对许多词汇给出了相应的情感极性或价值，这将作为我们论文的先验情感知识使用。然而，大多数现有的词汇都是一般的[2, 8, 15, 31, 39, 42]，其灵活性有限，不能直接应用于某些特定的领域。因此仍存在一些因素值得考虑在进一步研究这个话题，如单词的情感多样性在不同的上下文中，单词的情感极性的波动在不同领域和多义词的情感极性的识别。

针对特定领域的情感词汇的构建已成为解决上述问题的关键。这项工作通常需要特定领域的语料库的资源，而现有的情感词汇构成了它的基础。2013年，穆罕默德等人。[24]利用SVM分类器将情绪词汇构造问题转化为词分类。在他们的论文中，我们使用大量的推特语料库构建了两个情感词汇，分别命名为NRC话题标签情绪词汇和情感140词汇。后来，Vo和Zhang（2016）[36]使用了一个简单的神经网络架构来重新分配词情，其性能优于NRC的[24]。2014年，唐等人。[32]提出了基于神经网络的词情绪训练方法。他们使用种子集扩展了情感词汇，并构建了一个基于twitter的大型情感词汇TS-Lex。但是他们的方法需要提前手动选择种子集。基里琴科等人。[17]自动从语料库中生成情感词汇，如#快乐、#悲伤和#愤怒，但他们没有仔细过滤训练数据，也忽略了一些讽刺和直率的案例。邓等人。[5]提出了一种层次监督模型，考虑主题和情感词汇，构建主题适应性情感词汇。上述方法只根据不同的语料库来区分不同的词情。

然而，即使在同一语料库下，一个词也可能有多种含义。我们以“病人”一词为例，图。1表明“病人”这个词不仅表达了病人的意思，还表达了用来描述一个人的好角色。为了解决语义消歧的问题，在2019年，Wu等人。[41]考虑到一个情感词在描述不同目标时可能表达不同的情绪取向，构建了一个特定目标的情绪词汇。例如，在句子“屏幕太薄”中。，“屏幕”是目标，“瘦”是描述目标“屏幕”的情感词。该方法在[41]中的适用性受到限制，因为它忽略了一些句子的不完整和语法结构问题。

因此，如何区分词汇的情感表达是构建情感词汇时值得考虑的一个问题。从无花果。1、“病人”一词的词源是一个具有病人意思的名词和一个人的良好性格意思的形容词。因此，单词的POS可以作为区分不同情感表达的一个指标。基于这一事实，我们使用了ITF-IDF算法



Sentences	Sentiment word	Part-of-speech	Sentiment polarity
Jane is patient to children.	patient	adjective	
Now there is a patient in the class.	patient	noun	

图1以“病人”一词为例，对同一个词的不同表示

本文同时研究了不同极性语料库中情感词的区别。

在本文中，我们提出了一种领域特定情感词汇的自动构建方法，它集成了情感信息的两个部分：基于语料库的ITF-IDF算法计算的情感值和从现有情感词汇中获得的先验情感知识。我们的ITF-IDF算法使用不同的极性标签和POS因子，在特定领域获得更精细的单词情感价值，而传统的TF-IDF只关注单词在不同文档中的重要性。然后将先验情感知识与从特定领域语料库中计算出的词汇情感信息相结合，克服词汇情感对语料库的严重依赖，并对训练语料库中未包含的词汇情感进行补充。最后，实验结果表明，本文在情绪分类任务下构建的情绪词汇优于现有的情绪词汇，且该方法更适合于长文本语料库。

本文的其余部分组织如下：我们在第2节中介绍情感词汇的构建方法，并在第3节中描述我们的模型。然后，我们在第4节中介绍结果和性能比较，然后在第5节中提出结论和下一个研究计划。

2相关工程

为了进行情绪分析，研究人员提出了许多使用手动或自动方法的一般或领域特定的情绪词汇。

2.1. 现有情感词汇的资源

现有的情感词汇主要由一般词汇和领域特定词汇组成。一般词汇包括一般询问者（GI）[31]、发送者wordNet[2, 8]、意见词汇[14]等。GI [31]是最早的情感词汇，用极性、强度、词性等来标记每个单词。此后，Hu等人[14]提出的观点词汇由许多情感词组成，在工业上有广泛的应用。两年后，来自世界网的知识库的网络由Esuli和塞巴斯蒂亚尼·[2, 8]提出。在这个词典中，为WordNet的每个同步集分配三个分数来描述积极、消极和客观的程度，并进一步用来确定每个词的精确情绪。然而，大多数一般的词汇都是手工标记的，这不仅很耗时，而且还需要大量的人力。

大多数特定领域的词汇都是使用出现在特定领域的语料库中的单词的共现来构建的，如产品评论[7, 15]、社交平台[11, 24, 32, 43, 44]。威尔逊等人。[39]扩展了由Riloff和Wiebe等人首先收集的词汇MPQA主观性词汇。[29]从多角度的问题回答（MPQA）意见语料库。扩展词汇有助于消除情绪表达的极性的歧义，并以情感强度和词性标记每个单词。虽然这种情感词汇认为单词的域或主题，它忽略了单词的分布在不同的情感极地公司和一些词在同一主题可能有不同的情感极性，如“骄傲”这个词，这可能有两个意思：骄傲和自满在相同的主题下，导致不同的情绪表达。此外，由于不同的语言具有不同的句法结构，因此基于不同的语言构建情感词汇

已成为一个重要的研究趋势,包括越南[34],马来语[43],韩国[26],中国[42, 44],阿拉伯语[1, 25],斯洛文尼亚[3]。需要注意的是,语料库本身可能存在的错误严重影响了所构建的领域特定词汇的准确性,即领域特定词汇通常依赖于所选择的语料库。因此,我们利用现有词汇中的先验信息来解决依赖性问题。

2. 2个情感词汇的构建方法

基于词汇、基于语料库和混合的方法是构建情感词汇[37]的三种主要方法。

第一类是基于一些现有的情感词汇或知识库。WordNet是最常用的词汇之一,它通过各种同义词和反义词[8, 10, 16]来构建情感词汇。以下一些研究基于WordNet构建了情感词汇。坎普斯等人。[16]提出了一种利用WordNet测量词的语义取向的方法,但他们只考虑形容词而不是其他词性词的常用词。Esuli等。[2, 8]在WordNet中基于同义词关系和POS信息计算概念的情绪,然后对不同概念的同一词的情绪值进行整合,得到单词的情绪。他们的方法目前使用较少,因为它需要大量的人力,效率较低。此外,他们没有正确使用特定领域的语料库,导致对特定领域的词汇进行情绪分析失败。例如,“大”这个词在描述一个显示屏时是一个积极的词汇,但在描述衣服时可能会有一种消极的情绪。

第二种方法通过分析领域特定的语料库[4, 32, 36, 38, 42]中单词的关联,自动构建情感词汇。唐等人。[32]首先学习了情绪特异性短语嵌入(SSPE),然后将情绪词汇的构建转化为基于神经网络训练的分类问题。然而,它高度依赖于语料库。杨等人。[42]通过神经语言模型训练了一个大规模的语料库,并构建了一个包含10个情感标签维度的情感词汇。这个情感词汇通过将原来的三个类扩展到当前的十类,对单词的情感进行了更详细的判断,但没有考虑同一个词的消歧义。

最后一种方法是将语料库与现有的情感词汇相结合,共同学习情感词汇[9, 12, 19, 40, 44]。刘等人。[19]通过整合当前的情感词汇,构建了一个基本的微博情感词汇,并基于CHI平方提取了新的词汇,改进了SO-PMI算法。然而,这种方法并不健壮,因为SO-PMI算法需要手动选择种子。吴等人。[40]通过整合从语料库中提取的特定领域词汇知识、从所有语文中提取的情感相似性知识以及从现有情感词汇中提取的先验情感知识,构建了特定领域情感词汇。然而,他们并没有考虑到一个词在不同的语境中可能有不同的极性。最近,Han等人。[12]利用互信息与POS相结合,生成了特定领域的情感词汇,并在情感分析任务中取得了良好的效果。然而,互信息需要手动设置情绪的种子集,这种限制增加了该模型的不稳定性。我们的模型采用了改进的TF-IDF方法,通过考虑同一情感类别中每个词的POS的不同分布来解决这一问题。我们的方法比互信息方法提供了一种更详细、更稳定的词汇情绪分析方法。

有人试图防止情感词汇构建中的情感歧义问题。陆等人。[21]整合了来自不同来源的信息来学习一个基于上下文的情感词汇。这种方法可以使不同的单词在不同的主题下具有不同的情感表征。Saif等人。[30]利用从特定领域提取的上下文信息和语义信息，共同更新单词的情绪趋势，以缓解上下文变化时词汇情绪的差异。邓等人。[5]提出了高级分类任务的主题自适应情绪词汇（TaSL），联合考虑单词的主题和情绪，以捕捉单词在不同主题下的不同情绪表达。然而，由于一个词在不同的上下文和主题下可能有不同的极性和强度表达，因此仅基于上下文或主题进行情绪分析是不够合理的。

3. 构建一个基于特定领域的情感词汇
改进的TF-IDF算法

现有的情感词汇不能表达在不同语境中具有不同含义的词语的精确情感，而没有情感强度的简单极性标签所包含的信息有限。例如，如果单词“旗帜”被用作名词，它的意思是一个横幅（指一个国家或组织及其信仰和价值观），并表达了一种积极的倾向。然而，当它被用作动词时，它的意思是减轻疲劳，表达一种消极的情绪。此外，有些词只是出现在特定的领域，我们在一般的情感词汇中找不到这些词的情感强度。

为了解决这些问题，我们提出了一个自动构建领域特定领域情感词汇的框架，它联合利用了特定领域语料库和现有的情感词汇。图2显示了整体框架，包括三个部分：基于改进的TF-IDF算法的基于语料库的情绪计算，从现有的情感词汇中提取先验的情感知识，以及构建特定领域的情感词汇。对于每个单词，我们使用POS方法对细粒度的词义进行标记，这一步骤可以减少语义消歧。然后，通过ITF-IDF算法计算一个具有不同POS的单词的不同情绪值。由于特定领域语料库的词汇量限制，我们也包括了现有的

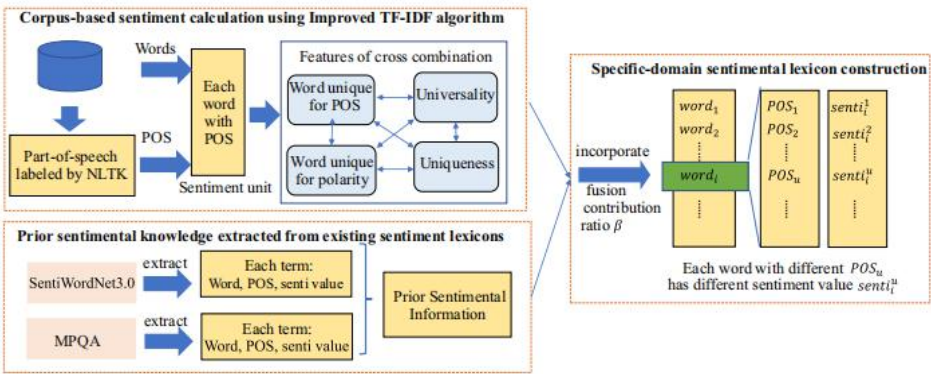


图2. 我们的模型的总体框架

一般情感词汇构建一个特定领域的情感词汇，旨在增加词汇的覆盖范围，微调词汇的情感价值。

3.1 基于语料库的情绪计算

构造情感词汇最常用的方法之一是点态互信息 (PMI) [35]。尽管被广泛使用，但这种方法的明显缺点是，一些单词需要手动选择作为正或负的种子集。在本文中，我们提出了改进的TF-IDF (ITF-IDF) 算法，并将POS信息与词在不同情感语料库中的分布相结合，以计算特征词的重要性，减少训练过程中的不确定性。

假设这个词 w_i 及其相应的POS p_u 构建一个新的情感单元 $w_i^{p_u}$ ， x_j 是对应的情感极性，并有两种状态：正极性 x_p 负极性 x_n 。当计算 w 的情绪时 i^{p_u} ，我们设置了四个假设：

-H1: Word对POS的独特性: w 的情感强度 i^{p_u} 与 w 的频率的比值呈正相关吗 i^{p_u} 以及 x_j 语料库中 w_i 的频率。

-H2: Word独特的极性: w 的情感强度 i^{p_u} 是否与频率与 W 的比值呈正相关 i^{p_u} 和数量 x_j 语料库在每个极性语料库中。

独特性: W 的情感强度 i^{p_u} 与正语料库和负语料库中单词的频率差的绝对值呈正相关。

普遍性: W 的情感的强度 i^{p_u} 是否与总频率的 W 呈负相关 i^{p_u} 在正和消极语料库中。

这些属性是决定词语的情感强度的关键。H1在数学上可以转化为：

$$F_1(w_i^{p_u}, x_j) = N\left(\frac{c(w_i^{p_u}, x_j)}{\sum_u c(w_i^{p_u}, x_j)}\right), \quad (1)$$

其中 $c(w_i^{p_u}, x_j)$ 是 w 的频率 i^{p_u} 在标记的语料库中 x_j ，分母为的总频率 w_i 在不同的POS x_j 语料库 $N(\cdot)$ 表示所得到的结果

是标准化的，这是用来稳定的值吗F10到1之间。公式(1)测量的是一个特定的POS内的情感强度。

H2形式化为：

$$F_2(w_i^{p_u}, x_j) = \frac{c(w_i^{p_u}, x_j)}{c(x_j)} = N(\cdot), \quad (2)$$

其中， $c(x_j)$ 为被标记的语料库的数量 x_j ， $N(\cdot)$ 表示结果被归一化，用于稳定的值F20到1之间。式(2)表示 w 的分布 i^{p_u} 在 x_j 语料库，用来衡量这个词在某一极性类别中的重要性。

H3表示为：

$$F_3(w_i^{p_u}) = N\left(\frac{|c(w_i^{p_u}, x_p) - c(w_i^{p_u}, x_n)|}{\sum_j c(w_i^{p_u}, x_j)}\right), \quad (3)$$

其中 $c(w_i^{p_u}, x_p)$ ， $c(w_i^{p_u}, x_n)$ and对 j $c(w_i^{p_u}, x_j)$ 分别为 w 的频率 i^{p_u} 在正、负和两个语料库中，和 $N(\cdot)$ 表示结果是归一化的，用于稳定的值F30到1之间。频率差的绝对值有助于删除一些经常出现的单词

并为每个极性类别选择一些最相关的词。公式 (3) 保证了语料库中特定词分布的唯一性。H4 定义为:

$$F_4(w_i^{pu}, x_j) = N(\log(\frac{\sum_i c(w_i^{pu}, x_j)}{c(w_i^{pu}, x_p) + c(w_i^{pu}, x_n) + \varepsilon})), \quad (4)$$

其中, ε 是保证分母不为零的常数; $N(\cdot)$ 表示结果被归一化, 用于稳定的值 F_4 到 1 之间。公式 (4) 显示了情感词在语料库中特定词性内分布的普遍性。任何一个 w 的情感价值 w_i^{pu} 在 x_j 语料库中, 计算方法为:

$$S_I(w_i^{pu}, x_j) = \prod_{n=1}^4 F_n, \quad (5)$$

在哪里平头的表示上述四个假设中的第 n 个结果, $SI(w_i^{pu}, w)$ 表示 w_i^{pu} 关于 x_j 倾向和我们采用乘法来表示各因素对情感结果的影响。每一个情感价值都通过减去 w 的情感价值来得到 w_i^{pu} 在正极性和负极性类别中:

$$SI(w_i^{pu}) = SI(w_i^{pu}, x_p) - SI(w_i^{pu}, x_n), \quad (6)$$

在哪里斯洛文尼亚 (w_i^{pu} 是积极类别中的情感价值, 斯洛文尼亚 (w_i^{pu} 是消极范畴中的情感价值斯洛文尼亚 (w_i^{pu} 是 w 的情感价值 w_i^{pu}).

3. 2 基于多个 POS 构建领域特定的情感词汇

我们引入了 ITF-IDF 算法来解决词义消歧和发现新词, 但在使用小的训练语料库时, 它对一般情感词的覆盖范围较差。为了解决这一问题, 我们将先验情感知识与基于语料库的情感信息相结合, 提出了一种特定领域的情感词汇。

我们首先收集现有的词汇 MPQA 和 SentiWordNet 3.0 来获取—精神知识。前者包含了单词的极性和强度标签, 但没有特定的情感价值。对于任何情感词 w 和 POS p , 先前的情感价值 $SM(w_i^{pu})$ 的定义类似于在 MPQA 中定义的

(7)

$$SM(w_i^{pu}) = \begin{cases} \alpha(x) & , \text{ if } w_i^{pu} \text{ is positive in MPQA} \\ -\alpha(x) & , \text{ if } w_i^{pu} \text{ is negative in MPQA} \end{cases}, \quad (8)$$

其中, $\alpha(x)$ 用由强度标签决定 $\alpha(x) = \begin{cases} 1 & , \text{ if } w_i^{pu} \text{ is strongsubj in MPQA} \\ 0.5 & , \text{ if } w_i^{pu} \text{ is weaksubj in MPQA} \end{cases}$ 标签。它的值 " 表示单词有较弱的情绪。

sentiwordnet3 是一个基于概念的情感词汇, 每个概念包括 POS、情感值、类似的词集等。但它忽略了这样一个事实, 即一些词虽然有相同的词性, 但也可能有多种情感价值。为了解决这个问题, 我们计算了多重情绪的平均值, 并得到了先验情绪值:

$$SW(w_i^{pu}) = \frac{\sum_{d \in D} s_d(w_i^{pu}, v_d)}{|D|}, \quad (9)$$

where $S_W(W_i^{pu})$ 是任何W的先验情绪值吗 i^{pu} 使用SentiWordNet3.0, D表示W的次数 i^{pu} 出现在SentiWordNet3.0, $s(W_i^{pu}, \text{任何})$ 的任何价值 i^{pu} 包含在类似的概念之词中。对于每个 w_i^{pu} , 先前的情感价值SP (w_i^{pu}) 计算方法为:

$$sp(w_i^{pu} \frac{S_M(W_i^{pu}) + S_W(W_i^{pu})}{2}) = , \quad (10)$$

最后, 将先验的情感知识与基于语料库的ITF-IDF情感信息相结合, 构建了我们的特定领域的情感词汇:

$$SU(w_i^{pu}) = (1 - \beta) \times SI(w_i^{pu}) + \beta \times SP(w_i^{pu}). \quad (11)$$

where the $S_U(W_i^{pu})$ 表示通过结合先验知识和基于语料库的ITF-IDF情感信息而获得的情感价值, β 是融合对比-情感比, 在0到1之间变化, 用于调整先验知识与基于语料库的情感知识之间的比例。

. 33评价方法

我们选择了定义如下的不同测量标准, 包括精度(P)、查全率(R)、F1、准确性(Acc)和覆盖范围(Cov), 以测试我们提出的情绪词汇在执行文本情绪分类任务中的效率。

对于一个文本序列 $x = w_1, \dots, w_k, \dots, w_K$, 我们首先标记相应的POS使用NLTK python工具, 并获得新的序列表达式 $x(w, p) = w, w, w_1^{p1}, \dots, w_k^{pk}, \dots, w_K^{pK}$, 其中 w_k^{pk} 表示序列s中的第k个字 p_k , K是序列s的单词总数。然后得到相应的情感值 $Senx = (SU(w_1^{p1}), \dots, SU(w_k^{pk}), \dots, SU(w_K^{pK}))$ 为每个 w_k^{pk} 在x中来合成最后一个句子的情感值包:

$$S_x = \sum_k^K S_U(w_k^{pk}), \quad (12)$$

在哪里 $S_U(w_k^{pk})$ 为序列x中带有POS标签的第k个单词的情感值 p_k 由我们的方法计算, 包是序列x的最终情感值。我们把 S_x 转换为极性值税以下是为了方便地判断我们的情绪分类任务的表现。

$$T_x = \begin{cases} 1, & \text{if } S_x > 0 \\ 0, & \text{if } S_x = 0 \\ -1, & \text{if } S_x < 0 \end{cases}, \quad (13)$$

我们进一步计算一下税每种情况下, 分别用NP、NN和ZN表示。最后采用了精度(P)、查全率(R)、F1、覆盖范围(Cov)和精度(Acc)来评价我们提出的方法。为简单起见, 我们只提供详细的定义测量P(pos)、R(pos)和F1(pos)的积极趋势。

(14)

(15)

(16)

表1实验数据集

数据集	积极的	负	平均长度 每句话	水平
拉姆德	5000	5000	231	长文本
摩斯	5000	5000	21	短文

其中TP是使用我们提出的词汇的正确正句子的数量，N_{pos}是评价数据集中的真正句。特别是，F1是精确度和召回率之间的权衡。负趋势的测量P（阴性）、R（阴性）和F1（阴性）可以用类似的方法定义。Acc和Cov是我们在实验中定义的情绪分类任务的准确性和覆盖度的度量如下

$$Acc = \frac{TP + TN}{N}, \tag{17}$$

$$Cov = \frac{NP + NN}{N}. \tag{18}$$

其中TN是使用我们提出的词汇的正确负句子的数量，N是评价数据集的总数。Acc反映了我们的模型在情绪分类上的性能，Cov表示了我们构建的情绪词汇的单词覆盖范围。

4实验

在本节中，我们将进行比较实验来测试我们所构建的情感词汇的效率。

4.1实验数据

我们的实验数据集来自大型电影评论数据集（LMRD）[22]和电影评论数据集（MRD）[27]。大型的电影评论数据集¹由Maas等人提供。[21]是从IMDB中提取的。它包含50000个标记样本（阳性和阴性样本，每个25000例）和50000个未标记样本。电影评论数据²由Pang和Lee [26]提供，摘自烂番茄网站，包括5331个正短文语料库和负短文语料库。

为了提高我们的结果的鲁棒性，我们在我们的实验中使用了十倍交叉验证的方法。特别是，为了保持训练语料库和测试语料库之间的平衡分布，我们分别从每个数据集中随机选择相同数量的正面（每个5000条）。因此，我们的训练数据集有4500个正和负句子，测试数据集在每个实验中包含500个正和负句子。表1报告了数据集的一些详细特性。

¹<http://ai.斯坦福.edu/~amaas//data/sentiment/>
²<http://www.cs.cornell.edu/people/pabo/movie-review-data>

4. 2. 对我们的SDS-lex的构建进行了实验

在本节中，我们将分析构建特定领域情感词汇过程中的重要步骤。其中一步是确定参数 β ，它加权基于语料库的情绪值和先验基于知识的情绪值。公式(1)–(4)中定义的对词语的情绪计算有不同的影响。公式(5)将这些因素相乘，以衡量总的影 响。另一个步骤是探讨这些因素如何影响情感词汇的构建。如表2所示，通过分别考虑每个假设因素或结合其中的部分因素，共得到15个特征。

4. 2. 1调整因子 β

本文构建的情感词汇同时使用了现有词汇的先验知识和基于语料库计算的情感值，其中 β 和 $1-\beta$ 是对这两个值的权重。因此， β 的测定在应用中很重要。

我们采用不同的值对 β 在不同的特征下进行情绪分类任务，以确定是否存在一个适用于每个或大多数特征的最优 β 。在我们的实验中，我们使用MRD语料库，使用十倍交叉验证的方法来训练我们的模型。图3绘制了我们在不同 β 值的不同特征下的实验精度。我们发现，当 β 在0.2左右时，精度达到了最大值，这一结论对于除两个特征外的大多数不同特征的情况都是有效的F1和F1* F4。因此，在接下来的实验中，我们选择0.2作为 β 的值。

4. 2. 2探讨不同特性的重要性

我们的SDS-lex的构造取决于所选择的特性。我们探讨了表2中报告的15个不同特征是如何影响情绪分类任务的。

基于短文本和长文本语料库，表3和表4分别报告了我们的方法在不同特征下的性能，其中前一节定义的Cov、Acc、P（pos）、R（pos）、F1（pos）、P（阴性）、R（阴性）、F1（阴性）分别表示单词覆盖率、分类准确性和对正文本和负文本的影响，粗体条目表示不同特征之间达到最优值的测量。

表2 15个由假设因素及其组合得到的特征

不	特点	不	特点
1	F1	9	F2*F4
2	F2	10	F3*F4
3	F3	11	F1*F2*F3
4	F4	12	F1*F2*F4
5	F1*F2	13	F1*F3*F4
6	F1*F3	14	F2*F3*F4
7	F1*F4	15	F1*F2*F3*F4
8	F2*F3		

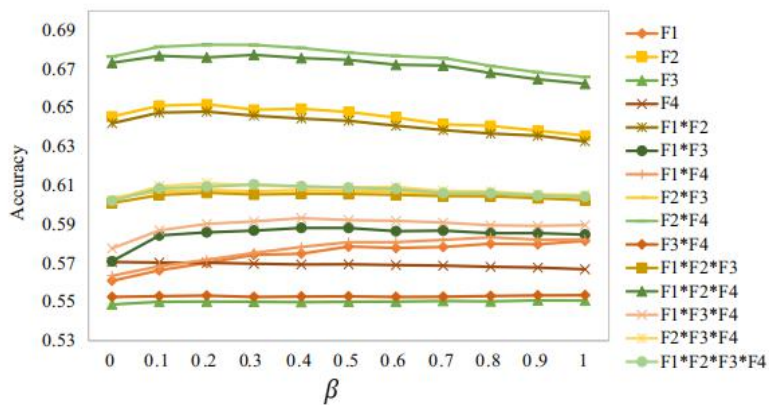


图3使用基于电影评论数据（MRD）语料库的不同β值对不同特征下的情绪分类任务的准确性

从表3可以看出，大多数特征下基于短文本语料库的Cov都大于98%，P和R的正负文本值不同，以及特征F2*F4在f1和Acc值方面达到最佳性能。从表4中可以看出，基于长文本的情绪分类任务在所有措施方面都比基于短文本的情绪分类任务具有明显的优势。

因此，情绪分类任务的表现明显依赖于所选择的特征，即不同特征的重要性有很大的不同。为了进一步探索每个特征的性能，我们使用MRD数据集和LMRD数据集在不同特征下进行情绪分类任务，并进行了展示

表3使用MRD数据集测量F1（pos）、F1（阴性）和Acc在不同特征下的情绪分类任务的表现

特点	科夫	P（pos）	R（pos）	F1（pos）	P（neg）	R（neg）	F1（内）	Acc
F1	0.9902	0.5683	0.6568	0.6093	0.5864	0.4834	0.5297	0.5701
F2	0.9961	0.6422	0.6996	0.6695	0.6696	0.6042	0.6350	0.6519
F3	0.9157	0.5819	0.6866	0.6299	0.6353	0.4138	0.5010	0.5502
F4	0.9157	0.6006	0.6940	0.6439	0.6611	0.4464	0.5327	0.5702
F1*F2	0.9963	0.6379	0.6982	0.6665	0.6665	0.5980	0.6300	0.6481
F1*F3	0.9867	0.5769	0.7154	0.6386	0.6227	0.4564	0.5265	0.5859
F1*F4	0.9875	0.5699	0.6574	0.6104	0.5918	0.4860	0.5336	0.5717
F2*F3	0.9959	0.5842	0.7766	0.6664	0.6634	0.4388	0.5269	0.6077
F2*F4	0.9963	0.6966	0.6568	0.6759	0.6752	0.7084	0.6912	0.6826
F3*F4	0.9157	0.5849	0.6852	0.6310	0.6392	0.4214	0.5077	0.5533
F1*F2*F3	0.9959	0.5831	0.7746	0.6650	0.6611	0.4380	0.5257	0.6063
F1*F2*F4	0.9963	0.6883	0.6540	0.6706	0.6700	0.6982	0.6837	0.6761
F1*F3*F4	0.9867	0.5809	0.7148	0.6409	0.6271	0.4656	0.5342	0.5902
F2*F3*F4	0.9959	0.5826	0.8154	0.6790	0.6900	0.4070	0.5096	0.6112
F1*F2*F3*F4	0.9959	0.5813	0.8120	0.6770	0.6863	0.4068	0.5087	0.6094

表4使用LMRD数据集测量的F1（pos）、F1（neg）和Acc在不同特征下的情绪分类任务的表现

特点	科夫	P（pos）	R（pos）	F1（pos ）	P（neg）	R（neg）	F1（内 ）	Acc
F1	1.0000	0.5745	0.7700	0.6579	0.6527	0.4294	0.5175	0.5997
F2	1.0000	0.7305	0.7908	0.7592	0.7722	0.7070	0.7378	0.7489
F3	1.0000	0.5924	0.8380	0.6940	0.7241	0.4220	0.5329	0.6300
F4	1.0000	0.6056	0.8404	0.7038	0.7403	0.4512	0.5602	0.6458
F1*F2	1.0000	0.7303	0.7746	0.7515	0.7601	0.7126	0.7353	0.7436
F1*F3	1.0000	0.5804	0.7606	0.6583	0.6536	0.4498	0.5326	0.6052
F1*F4	1.0000	0.5738	0.7610	0.6541	0.6465	0.4344	0.5192	0.5977
F2*F3	1.0000	0.5808	0.9536	0.7216	0.8690	0.3092	0.4536	0.6314
F2*F4	1.0000	0.7323	0.8884	0.8015	0.8588	0.6690	0.7488	0.7787
F3*F4	1.0000	0.5928	0.8384	0.6944	0.7250	0.4226	0.5336	0.6305
F1*F2*F3	1.0000	0.5837	0.9508	0.7231	0.8654	0.3194	0.4642	0.6351
F1*F2*F4	1.0000	0.7297	0.8738	0.7940	0.8437	0.6706	0.7441	0.7722
F1*F3*F4	1.0000	0.5827	0.7702	0.6634	0.6619	0.4480	0.5341	0.6091
F2*F3*F4	1.0000	0.5973	0.9588	0.7356	0.8951	0.3494	0.4974	0.6541
F1*F2*F3*F4	1.0000	0.5995	0.9548	0.7360	0.8892	0.3578	0.5048	0.6563

他们的表现由F1（pos）、F1（neg）和Acc三个测量方法报告。分别是4和5。

从图中可以看到。4和5表明，(1)使用两个不同的数据集执行的任务具有相似的性能；(2)在该特征下获得了最佳的性能F2在四个独立特征中，词汇在不同情绪极性类别中的分布对情绪词汇的构建具有重要意义；(3)的特征F2*F4始终保证所有特征中F1（fos）、F1（neg）和Acc的最优值；(4)F2*F4和的结果F1* F2 * F4在正语料库和负语料库中都具有平衡的性能，而其他特性通常对单一面向方向的语料库有用。

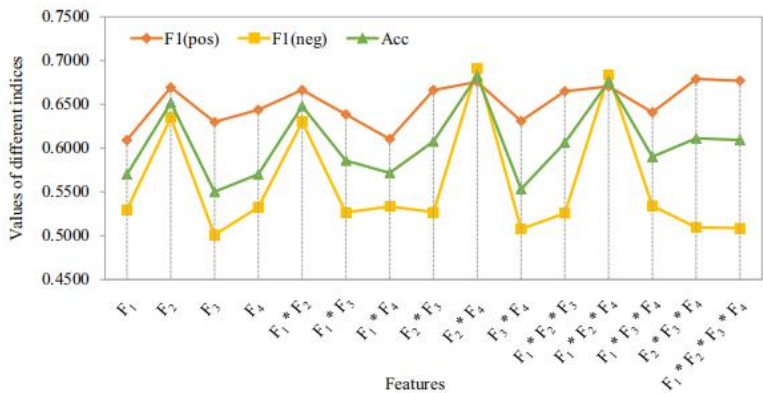


图4这三个准确性指标，F1（pos）和F1（pos）对电影评论数据（MRD）的影响

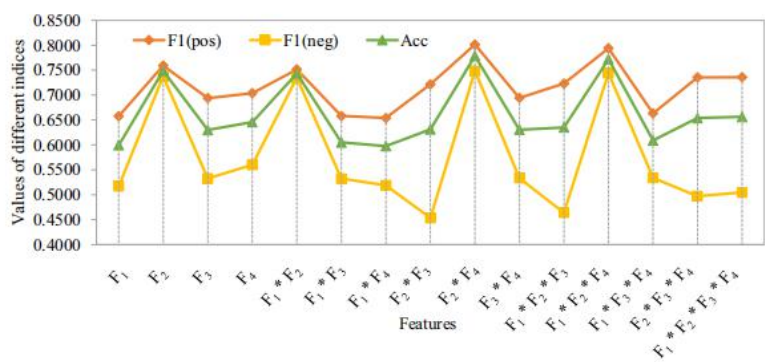


图5这三个准确性指标，F1（pos）和F1（pos）对大型影评数据集（LMRD）的影响

因此，我们使用F2*F4计算基于语料库的情感值，并将其与之前的情感词汇相结合，构建我们的SDS-lex。

.34比较实验

我们进行了情绪分类任务，并将我们提出的方法与一些现有的情绪词汇进行了比较，包括GI [31]，MPQA [39]，SW [2]、NRC [24]、S 140[24]，ETSL[17]，HIT[32]、NN [36]和HSSWE [38]。

表5和表6分别给出了基于MRD中的短文本语料库和LMRD中的长文本语料库使用不同词汇进行的文本情绪分类任务的性能。很明显，我们的方法的Acc是最高的，这提高了

表5基于长文本语料库MRD的不同特征的文本情绪分类任务的效果

方法	Cov P (pos)	R (pos)	F1 (pos)	P (neg)	R (neg)	F1 (内)	Acc
通用词典							
gi	0.7127 0.6055	0.5858	0.5952	0.6806	0.3118	0.4275	0.4488
有部分发言权的墨西哥人							
MPQA	0.7202 0.6486	0.5104	0.5711	0.6455	0.4214	0.5097	0.4659
西南部	0.9485 0.5787	0.6430	0.6090	0.6096	0.4788	0.5361	0.5609
域特异性词汇							
NRC	0.9969 0.6730	0.4902	0.5671	0.5990	0.7580	0.6691	0.6241
S140	0.9958 0.5696	0.8324	0.6764	0.6885	0.3652	0.4771	0.5988
ETSL	0.7747 0.5349	0.6364	0.5812	0.6160	0.2216	0.3259	0.4290
击球	0.9799 0.6268	0.6376	0.6321	0.6329	0.5966	0.6142	0.6171
神经	0.9984 0.6569	0.5192	0.5799	0.6025	0.7268	0.6588	0.6230

HSSWE	0.9955	0.5804	0.7550	0.6562	0.6491	0.4478	0.5297	0.6014
我们的领域特异性词汇								
我们的SDS-lex	0.9963	0.6966	0.6568	0.6759	0.6752	0.7084	0.6912	0.6826

比短文本和长文本语料库中现有情感词汇的最优结果分别高出6%和9%。

对于短文语料库，表5显示了我们的方法与其他方法具有相似的覆盖范围。然而，我们的方法比N网络、NRC和S140。从表5中得到的另一个观察结果是，情感词汇S140和NN仅对特定的情绪语料库有效，而我们的SDS-lex、MPQA和SW的情绪词汇对正语料库和消极语料库都有潜力。这是因为后三个词汇包含了词性信息。特别是，我们构建的词汇在所有指标方面都优于MPQA和SW。

对于长文本语料库，表6显示，基于长文本语料库，所有情感词汇的Cov均大于90%。我们再次发现，POS信息的使用提高了情感词汇的适用性。

综上所述，该方法构建的情感词汇在大规模语料库数据下不仅与现有的情感词汇具有相同的覆盖范围，而且对正语料库和负语料库都具有较高的准确性和良好的稳定性。

. 44语料库数据集大小对SDS-lex构建的影响

我们选择不同的数据集大小来测试它们对SDS-lex构建的影响。选择的正和负语料库数据集大小分别为1000、3000、5000、7000，LMRD的数据分别为9000、11000和12500。

图6绘制了详细的结果，并显示了数据化对所构建的情感词汇的覆盖率没有影响。我们发现，指标P和R是波动的，指标f1和Acc随着数据集大小的增加而增加。由于F1是P和R之间的权衡，而Acc是情感词汇在执行分类任务方面的准确性，我们得出结论，我们构建的SDS-lex的性能可以通过增加数据集的大小来提高。

表6基于长文本语料库LMRD的不同特征的文本情绪分类任务的效果

方法	Cov	P (pos)	R (pos)	F1 (pos)	P (neg)	R (neg)	F1 (内)	Acc
通用词典								
gi	0.9258		0.8390	0.6829	0.7460	0.2940	0.4214	0.5665
	0.5759							
有部分发言权的墨西哥人								
MPQA	0.9358	0.6608	0.7120	0.6852	0.7074	0.5610	0.6253	0.6365
西南部	0.9991		0.8086	0.6892	0.7084	0.4616	0.5587	0.6351
	0.6007							
域特异性词汇								
NRC	1.0000		0.2482	0.3876	0.5633	0.9688	0.7123	0.6085
	0.8891							
S140	1.0000		0.9024	0.7126	0.7915	0.3690	0.5027	0.6357
	0.5889							
ETSL	0.9990	0.5209	0.9736	0.6787	0.8025	0.1034	0.1829	0.5385
击球	0.9994		0.7692	0.7126	0.7262	0.6098	0.6627	0.6895
	0.6640							
神经	1.0000		0.4522	0.5794	0.6196	0.8912	0.7309	0.6717
	0.8068							
HSSWE	1.0000	0.6228	0.8374	0.7143	0.7525	0.4924	0.5951	0.6649
我们的领域特异性词汇								

我们的SDS-lex	1.0000	0.7323	0.8884	0.8015	0.8588	0.669	0.7488	0.7787
------------	--------	--------	--------	--------	--------	-------	--------	--------

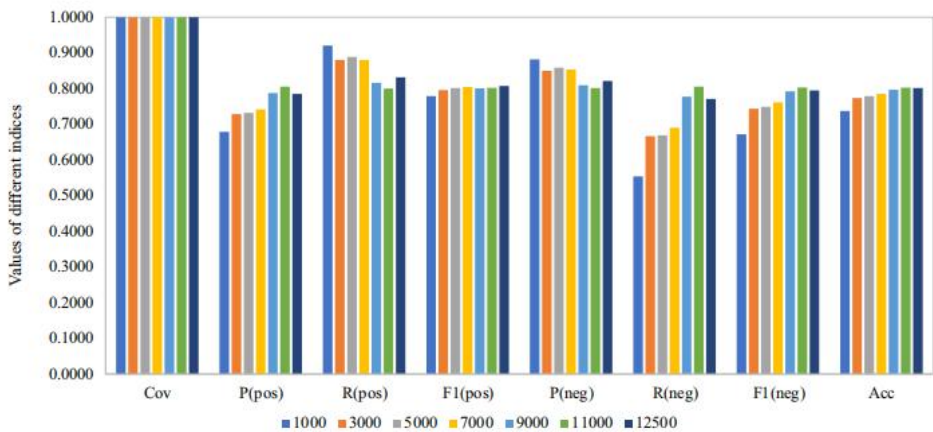


图6这三个准确性指标（F1（pos）和F1（pos））对大型影评数据集（LMRD）的影响

.54案例分析

我们构建的SDS-lex的一个优点是可以识别同一个词的不同情绪。我们以不同句子中包含的“旗帜”和“口音”，以分析这一优势。表7显示了我们构建的情感词汇提供的两个词的情绪极性和POS。单词“flag”（“重音”）既是名词，也是动词，它在不同的句子中具有不同的情感值。我们从表7中发现，我们构建的情感词汇有可能识别这一事实。

请注意，由于同一个词通常有不同的形式，如果不同的形式用不同的POS标记，则会大大增加计算负担。我们构建的情感词汇避免了这个问题，并以相同的方式用相同的形式标记相同的单词。包含不同形式的动词、副词、副词和名词分别标记为“v”、“a”、“r”和“n”，其他罕见词的POS保持不变。

表7以这两个词为例，分析了单词语义消歧的性能

单词	情绪	语言部分	句子
旗	积极乐观的	n	在兰德尔·华莱士导演的挥舞旗帜的战争电影中，梅尔·吉布森在越南展开了精彩的战斗。
	消极的	v	这部电影有时会旗帜鲜明...
口音	积极乐观的	v	霍夫曼突出了痛苦的细微差别，但他那聪明、尖锐的声音和摇摇摆摆的侧面轮廓突出了威尔逊困境中的幽默。
	消极的	n	轻率的，随意的，肤浅的幽默和很多非常糟糕的恶心口音

5结论

本文利用现有词汇的先验信息和特定领域的语料库信息，构建了一个情感词汇。我们的ITF-IDF算法使用不同的极性标签和POS因子，在特定领域获得更精细的词汇情感价值，防止了许多现有词汇中出现的情感歧义问题。然后将先验的情感知识与特定领域语料库计算出的情感值相结合，克服了词汇情感对语料库的严重依赖问题。该方法对训练语料库中未包含的词汇的情感也有较好的补充效果。最后，我们在真实数据集上进行了文本情绪分类任务，以测试我们构建的情感词汇的效率，发现该方法与一些常用方法相比具有明显的优势。

在今后的工作中，我们希望进一步研究时态等因素，构建一个更细粒度的情感词汇，并多样化其评价方法。

国家自然科学基金资助项目。61801440)，北京（中国通信大学互联网信息专业高校优质前沿学科建设项目）和中央高校基本科研业务费基金。

参考文献

1. Assiri A, Emam A, Al-Dossari H (2018), 旨在加强基于词汇的沙特方言情绪分析方法。J Inf Sci 44(2):184–202
2. 3. 0: 用于情感分析和意见挖掘的增强词汇资源。在: 语言资源和评价国际会议, 瓦莱塔
3. (2018) 斯洛文尼亚注释新闻语料库和情绪分析词汇。LangResourEval52(3): 895–919
4. 5: 通过上下文嵌入发现情感分析的概念原语。32th Int. 会议使发生联系Adv. Artif. Intell, 第1795–1802页
5. 邓, 静, 余, 孙. L., Ng MK (2019) 情绪词汇构建与分层监督主题模型。IEEE-ACM跨音频语音处理过程27(4): 704–718
6. Denecke K (2008) 使用感知字网进行多语言情感分析。见: 2008 IEEE第24届数据工程国际研讨会, 第507–512页
7. (2018) 句子n克: 一个用于情绪分析的n克词汇。专家系统系统App1103: 92–105
8. (2006) 网络: 一个公开的意见挖掘词汇资源。见: 语言资源与评价国际会议 (LREC-2006), 第417–422页
9. 冯杰, 龚C, 刘RYK (2018) 移动购物评论情感词汇的自动生成方法。无线通信和移动计算
10. 情感词汇: 为情感分析提供一个高精度和高覆盖率的词汇。IEEE反式影响Comput7(4): 409–421
11. Go A, Bhayani R, 黄L (2009) 推特情绪分类使用远程监督。见: 斯坦福大学自然语言处理小组2008/2009年春季CS224N的最终项目
12. 韩慧, 张跃, 杨强, 沈年, 张年 (2018) 生成了特定领域的情绪词汇, 用于回顾情绪分析。多定时工具App177(16): 21265–21280
13. 使用混沌salp群算法进行特征分类。阿拉伯文学科学论文集44(4): 3801–3816
14. Hu MQ, Liu B (2004) 挖掘和总结客户评论。见: ACM SIGKDD, 第168177页
15. Khoo CSG, Johnkhan SB (2018) 基于词汇的情绪分析: 六种情绪词汇的比较评价。J Inf Sci 44(4):491–511

16. (2004) 使用wordNet来测量形容词的语义取向。在: Proc. 4th Int. 会议长的Resur. Eval, 第4卷, 第1115–1118页
17. 李志志 (2014) 非正式文本的情绪分析。J Artif Intell的意思是50: 723–762
18. (2020) 对aadhaar推特数据的情绪分析—一种混合分类方法。国际计算科学与应用会议的进程: ICCSA 2019. 施普林格自然出版社, 第309–318页
19. 刘强 (2016) 基于中国微博的情感词汇构建研究。见: 智能人机系统与控制论国际会议, 杭州, 第56–59页
20. Liu J, Fu X, Liu J等 (2018) 对JD的分析和评价。com.Intell自动软输出24(1): 73–79
21. 翟C (2011) 情境感知情感词汇的自动构建: 一种优化方法。在: Proc. 20th Int. 会议万维网(WWW), 第347–356页
22. Maas AL, Daly RE, Pham PT等人 (2011) 学习情绪分析的词向量。参见: 计算语言学协会会议: 人类语言技术, 波特兰, 第142–150页
23. Mandal S, Singh GK, Pal A (2020) 文本摘要技术的情绪分析和杜鹃搜索算法。工程和技术中的计算。新加坡, 施普林格, 第357–366页
24. (2013) nrc-加拿大: 建立最先进的推文情绪分析技术。见: 第二届词汇和计算语义联席会议, (SEM)。第七届语义评价国际讲习班论文集 (SemEval 2013), 第2卷。亚特兰大, 第321–327页
25. Abdullah S (2018) 阿拉伯语句子-词汇: 为阿拉伯语情感分析构建可用的公共语言资源。J Inf Sci 44(3):345–362
26. 崔M (2018) KNU韩国情绪词汇: 基于双LSTM的方法构建一个韩国情绪词汇。J Intell Inf系统 24(4): 219–240
27. 庞B, 李L (2005) 看明星: 利用等级关系对评级量表进行情感分类。在: 关于计算语言学协会的会议。安娜堡出版社, 第115–124页
28. Rani S, Kumar P (2019) 基于卷积神经网络的深度学习的情绪分析。阿拉伯期刊科学论文集 44(4): 3305–3314
29. Riloff E, Wiebe J (2003) 学习主观表达的提取模式。见: EMNLP, 第105–112页
30. 赛义夫 (2017) 情感词汇与语境和语义的适应。SemantWeb8(5): 643–665
31. (1966) 一般询问者: 内容分析的计算机方法。Inf存储回收装置4(4): 375–376
32. 唐D、魏F、秦B等人 (2014) 构建大规模的twitter特定情感词汇: 一种表示学习方法。参见: 卷, 第172–182页
33. 陶、刘特、于华等 (2018) 在意见挖掘中构建不同情绪情境和多语言环境的本体。自动软输出 24(1): 65–71
34. Tran TK, Phan TT (2018) 一种构建越南情绪词典的混合方法。J Intell模糊系统第35(1)节: 967–978
35. Turney PD (2002) 竖起大拇指还是竖起大拇指? 语义取向应用于无监督分类。见: 计算语言学协会年会会议记录, 第417–424页
36. Vo DT, 张勇 (2016) 不数, 预测! 一种自动学习短篇文本的情感词汇的方法。在: Proc54年度。会议协会的作品。语言学家, 第219–224页
37. 王K, 夏R (2016) 关于情感词汇自动构建方法的调查。ActaAutomSin42(4): 495–511
38. 王Y, Zhang Y, Liu B (2017) 基于神经PU学习、双字典查找和极性关联的情感词汇扩展。在: Proc. 会议《经验方法自然朗法》, 第553–563页
39. 霍夫曼P (2005) 在短语水平的情绪分析中识别语境极性。见: HLT ‘05自然语言处理中的人类语言技术和经验方法会议论文集, 温哥华, 不列颠哥伦比亚省, 第3–354页
40. 吴五、黄勇、宋勇等人 (2016) 面向构建一个高质量的微博特定的中国情感词汇。Decis支持 Syst87(C): 39–49
41. (2019) 目标特定情感词汇的自动构建。专家会议附录116: 285–298

42. Yang XP, Zhang ZX, Wang L等人 (2017) 基于Word2Vec的情感词汇的自动构建与优化。Comput Sci 74(1):42–47
43. Zabha NI, Ayop Z, Anawar S, 哈米德E, Abidin ZZ (2019) 使用基于词汇的方法开发马来语推特数据的跨语言情绪分析。Int J Adv Comput Sci Appl 10(1):346–351
44. 赵德 (2019) 利用社会和地方情境传播诱导中国微博特定情感词汇。Comput演讲Lang55: 57–81

出版商的说明施普林格自然保持中立的管辖权主张在已出版的地图和机构附属机构。



王燕燕于2016年获得中国传播大学工程学学士学位，并决定在攻读硕士学位期间学习文本情感分析。她目前正在攻读博士学位。D. 中国通信大学信息与通信工程学院，北京。她目前的研究兴趣包括表征学习和情感分析。



尹获得学士、硕士、博士学位。2005年、2007年和2010年获哈尔滨工程大学学士学位。现任中国通信大学信息与通信工程学院副教授、硕士导师。她目前的研究工作涉及自然语言处理、数据分析和数据挖掘。



刘建波于1985年获得中国北京清华大学无线电电子系学士学位，1988年在中国北京通信大学获得硕士学位。他是中国通信大学信息与通信工程学院的教授。他的研究兴趣包括有线电视和宽带网络技术。



马尔科·托萨托完成了他的纯数学（代数领域）的理学学士学位，然后决定在他的硕士学位期间转向应用（生物）数学。学士学位和硕士学位都是在意大利特伦托大学完成的，并且是作为荷兰乌得勒支大学的交换生完成的。他目前是加拿大多伦多约克大学吴建红教授的指导下的一名博士生。他的研究兴趣包括神经网络、时滞微分方程。