

基于双重注意力模型的微博情感分析方法

张仰森, 郑佳, 黄改娟, 蒋玉茹

(北京信息科技大学 智能信息处理研究所, 北京 100101)

摘要: 微博情感分析是获取微博用户观点的基础。该文针对现有大多数情感分析方法将深度学习模型与情感符号相剥离的现状, 提出了一种基于双重注意力模型的微博情感分析方法。该方法利用现有的情感知识库构建了一个包含情感词、程度副词、否定词、微博表情符号和常用网络用语的微博情感符号库; 采用双向长短记忆网络模型和全连接网络, 分别对微博文本和文本中包含的情感符号进行编码; 采用注意力模型分别构建微博文本和情感符号的语义表示, 并将两者的语义表示进行融合, 以构建微博文本的最终语义表示; 基于所构建的语义表示对情感分类模型进行训练。该方法通过将注意力模型和情感符号相结合, 有效增强了对微博文本情感语义的捕获能力, 提高了微博情感分类的性能。基于自然语言处理与中文计算会议(NLPCC)微博情感测评公共数据集, 对所提出的模型进行评测, 结果表明: 该模型在多个情感分类任务中都取得了最佳效果, 相对于已知最好的模型, 在2013年的数据集上, 宏平均和微平均的 F_1 值分别提升了1.39%和1.26%, 在2014年的数据集上, 宏平均和微平均的 F_1 值分别提升了2.02%和2.21%。

关键词: 情感分析; 双重注意力模型; 微博; 语义表示; 情感符号

中图分类号: TP391.1

文献标志码: A

文章编号: 1000-0054(2018)02-0122-0130

DOI: 10.16511/j.cnki.qhdxxb.2018.22.015

Microblog sentiment analysis method based on a double attention model

ZHANG Yangsen, ZHENG Jia, HUANG Gaijuan, JIANG Yuru

(Institute of Intelligent Information Processing,
Beijing Information Science and Technology University,
Beijing 100101, China)

Abstract: Microblog sentiment analysis is used to get a user's point of view. Most sentiment analysis methods based on deep learning models do not use emotion symbols. This study uses a double attention model for microblog sentiment analysis that first constructs a microblog emotion symbol knowledge base based on existing emotional semantic resources including emotion words, degree adverbs, negative words, microblog emoticons and common Internet

slang. Then, bidirectional long short-term memory and a full connection network are used to encode the microblog text and the emotion symbols in the text. After that, an attention model is used to construct the semantic representations of the microblog text and emotion symbols which are combined to construct the final semantic expression of the microblog text. Finally, the emotion classification model is trained on these semantic representations. The combined attention model and emotion symbols enhance the ability to capture the emotions and improve the microblog sentiment classification. This model gives the best accuracy for many sentiment classification tasks on the Natural Language Processing and Chinese Computing (NLPCC) microblog sentiment analysis task datasets. Tests on the 2013 and 2014 NLPCC datasets give F_1 -scores for the macro and micro averages that are 1.39% and 1.26% higher than the known best model for the 2013 dataset and 2.02% and 2.21% higher for the 2014 dataset.

Key words: sentiment analysis; double attention model; microblog; semantic representation; emotion symbol

以微博为代表的新兴社交媒体, 以其独特的开放性、实时性、互动性和低门槛性为人们进行意见表达和信息交流提供了媒介, 以极快的速度影响着社会的信息传播格局^[1]。对微博文本进行情感分析有助于了解公众的情感状态, 及时获取公众的观点、态度, 对于舆情控制、产品营销、民意调查等具有重要的意义。

现有的情感分析技术主要分为3种: 基于情感知识的方法、基于特征分类的方法、基于深度学习的方法。基于情感知识的方法主要是构建一些情感词典, 通过情感词典对文本中的情感词进行一定的组合计算, 实现对文本情感的分析^[2-3]。人们在进行情感表达时, 往往采用某些特殊的情感符号实现情

收稿日期: 2017-08-22

基金项目: 国家自然科学基金资助项目(61370139, 61772081, 61602044)

作者简介: 张仰森(1962—), 男, 教授。

E-mail: zhangyangsen@163.com

感的呈现,常用的情感符号主要为带有情感色彩的情感词,有时为了对情感进行更加深入的表达,人们会将情感词与程度副词或是否定词配合使用,起到情感加强或情感减弱的效果。微博中的情感表现形式除了基本的情感词外,还包含微博自身的表情符号和一些流行的网络用语。本文将情感词、程度副词、否定词、微博表情符号、网络用语统称为情感符号,将其作为评价微博文本情感倾向的一个重要组成部分。

基于特征分类的方法主要是采用文本分析技术提取文本中蕴含的相关语言特征,采用机器学习的方法将情感分析看作分类问题进行处理^[4-5],但是该方法对于情感特征提取方法的依赖性较大,提取情感特征的准确与否将直接影响最终的分类结果。

基于深度学习的方法主要是采用词向量表示技术对文本中的词语进行表示,进而构建句子和篇章的语义表示。在句子和篇章的语义表示的基础上,采用深度学习模型对文本中蕴含的情感进行学习,从而实现对文本情感的分析,常用于情感分析的深度学习模型包括:卷积神经网络(convolutional neural network, CNN)^[6]、循环神经网络(recurrent neural network, RNN)^[7]、长短记忆网络(long short-term memory, LSTM)^[8]等。

现有的情感分析方法大多是基于深度学习的方法,将文本看作一个整体进行编码,对于情感符号的作用体现不够,而传统的基于情感词典的方法则过分依赖于情感词典的作用,并没有考虑文本整体的语义关系。为了解决以上问题,本文将在机器翻译任务中表现出优秀性能的注意力模型^[9-10]引入到情感分析任务中,利用注意力模型对微博文本和情感符号进行双重编码,提出了一种基于双重注意力模型(double attention model, DAM)的微博情感分析方法:

1) 通过利用现有的情感语义资源,对中文微博中经常使用的情感符号进行分类和统计,构建了一个微博情感符号库。

2) 提出了一种基于双重注意力模型的微博情感分类方法,将注意力模型引入微博文本的情感分析之中,通过设计相关实验验证了双重注意力模型在情感二分类和情感多分类任务中的有效性。

3) 通过设计相关实验,验证了在深度学习模型中引入注意力机制和情感符号,对情感分析性能的提升作用。实验表明,本文模型的性能超越了一

般的深度学习模型。

1 相关工作

1.1 注意力模型

最早将注意力模型引入自然语言处理领域的是Bahdanau等^[9],他们认为传统的Encoder-Decoder模型中Encoder生成的固定长度的中间语义向量是提升神经机器翻译模型性能的瓶颈。为此,Bahdanau等提出了一种计算前一时刻输出状态与每一时刻输入状态对齐概率的对齐模型,将其应用到英语与法语的翻译系统中,取得了不错的效果,该对齐模型是现在大多数注意力模型的基础。随后,基于注意力机制的Encoder-Decoder模型在自然语言处理的很多领域得到了应用,如机器人的短文本对话、文本摘要生成、文本分类、句法分析等。

随着注意力模型在自然语言处理领域的广泛应用,各种改进模型也应运而生,如Luong等^[10]针对注意力的全局对齐的模式,提出了一种局部注意力模型,首先确定对齐的大致位置,然后向左向右各拓展一定的窗口长度,在指定的窗口范围内计算每个单词的对齐概率;Hermann等^[11]根据人们在做阅读理解题目时,往往是阅读完整个问题后再去文档中寻找答案而不是每读完一个单词就去文档中寻找答案的实际情况,将逐个对齐的方式修改为将整个问题与文档中的单词进行对齐的模式;Raffel等^[12]为了解决长距离的语义依赖问题,将注意力模型中的对齐模型进行改进,通过对当前输入的自适应加权,将原来前一时刻输出的隐含状态与当前时刻输入的隐含状态进行对齐的模式,修改为仅依赖当前时刻输入的隐含状态的对齐模式。

1.2 微博文本情感分析

自从微博出现以来,国内外对于微博文本的情感分析研究就从未间断过。多数的研究工作仍然是承袭了传统的基于情感知识的方法、基于特征分类的方法和基于深度学习的相关模型。例如:栗雨晴等^[13]在大规模微博文本的基础上,利用相似度计算的方法对现有的情感词汇知识库进行扩展,构建了一个双语情感词典,采用半监督Gauss混合模型和对称相对熵的K近邻算法实现对微博文本的情感分类计算;Barbosa等^[14]在普通文本特征的基础上,提取了微博文本特有的一些特征,包括转发、回复、hashtag、URL、标点符号、表情符号以及以大写字母开头的单词数目等,采用有监督的方法实现Twitter文本的情感分类。由于微博文本中的表

情符号具有较强的情感表现力,越来越多的研究将其引入到情感分析中,如 Jiang 等^[15]利用表情符号的词向量构建了表情符号向量空间模型(emoticon space model, ESM),并将微博文本中所有的词映射到向量空间中,采用支持向量机(support vector machine, SVM)模型实现对微博文本的情感分类;何炎祥等^[16]为常见的表情符号构建情感空间的特征表示矩阵,通过对文本词向量矩阵与表情符号的特征表示矩阵进行乘积运算,实现词义到情感空间的映射,从而构建微博文本的情感表示矩阵,然后采用多通道的卷积神经网络对微博文本情感表示矩阵进行建模,实现对微博文本的情感分类。

文^[15-16]虽然考虑了微博文本中表情符号对于情感表现的重要性,但是都是采用全局的表情符号集实现对微博文本的情感加强,并未考虑微博文本自身情感符号的意义以及文本自身的语义特征,同时也未考虑微博文本中常见的情感词以及常用的网络用语对于情感的表现能力。受文^[15-16]的启发,本文将考察常见的情感符号,包括情感词、程度副词、否定词、微博表情符号、网络用语,对情感表现的作用,采用注意力模型对微博文本和文本中包含的情感符号分别进行建模,构建了一种基于双重注意力模型的微博文本情感分析方法,并采用自然语言处理与中文计算会议(NLPCC)2013年和2014年的微博情感测评数据集构建了多个微博文本情感分类任务,对本文提出的模型进行验证。

2 双重注意力模型

2.1 情感符号库的构建

本文通过对现有情感知识库的整理,构建了微博情感符号库,如表1所示。

表1 微博情感符号库

情感符号	数目	举例
情感词	34 606	高兴、绝望、恶心
表情符号	1 866	[哈哈]、[伤心]、[失望]
程度副词	219	很、相当、着实
否定词	62	不、从未、绝非
网络用语	675	伤不起、羡慕嫉妒恨、你想多了

表1中,情感词主要基于大连理工大学的《情感本体库》和董振东构建的《HowNet》中的情感词典进行构建;表情符号采用新浪应用程序接口(API)从新浪微博开放平台上获得;程度副词以《HowNet》中的程度副词列表为基础进行构建;否

定词和网络用语以 Song 等^[17]构建的否定词表(negation words)和网络用语词表(NetLex)为基础进行构建。

2.2 情感符号的提取

通过构建的情感符号库,提取微博文本中包含的情感符号,为每一条微博文本构建情感符号集合。对于情感符号的提取,本文制定了如下提取规则:

规则1 若当前词语为情感词、表情符号和网络用语,直接将当前词语加入到情感符号集合中。

规则2 若当前词语为程度副词,且当前词语的下一个词语为情感词,则将当前程度副词和情感词作为一个整体加入到情感符号集合中,若情感词存在于情感符号集合中,则将其删除。

规则3 若当前词语为否定词,且当前词语的下一个词语为情感词,则将当前否定词和情感词作为一个整体加入到情感符号集合中,若情感词存在于情感符号集合中,则将其删除;或者当前词语的下一个词语为程度副词,并且程度副词的下一个词语为情感词,则将当前否定词和程度副词以及情感词作为一个整体加入到情感符号集合中,若程度副词与情感词的整体存在于情感符号集合中,则将其删除。

2.3 双重注意力模型

本文在双向长短记忆(bidirectional long short-term memory, BLSTM)网络深度学习模型中引入注意力机制和情感符号,构建了双重注意力模型,模型的整体架构如图1所示。

2.3.1 词向量表示层

词向量表示层为整个模型的输入,主要包括两部分:每一条微博文本词语的词向量矩阵 \mathbf{R}_T 和情感符号集合的词向量矩阵 \mathbf{R}_E 。词语及情感符号词向量的获取可以看作是一个查词典的过程。词典 $\mathbb{R}^{d \times N}$ 通过大规模语料采用词向量训练模型学习得到。其中: d 表示词向量的维数, N 表示词典中词语的个数。对于一个文本序列 $\mathbf{T} = \{w_1, w_2, \dots, w_n\}$,将文本中词语的词向量拼接起来,就可以得到整个文本序列的词向量表示,拼接方式如式(1)所示:

$$\mathbf{R}_T = \mathbf{v}_1 \oplus \mathbf{v}_2 \oplus \dots \oplus \mathbf{v}_n. \quad (1)$$

其中: $\mathbf{v}_i \in \mathbb{R}^{d \times N}$ 表示 w_i 对应于词典中的元素, \oplus 表示行向量拼接操作。对情感符号集合中词语的词向量同样采用式(1)的方式进行拼接, \mathbf{R}_T 和 \mathbf{R}_E 的维数即为微博文本中词语的数目和情感符号的数目。

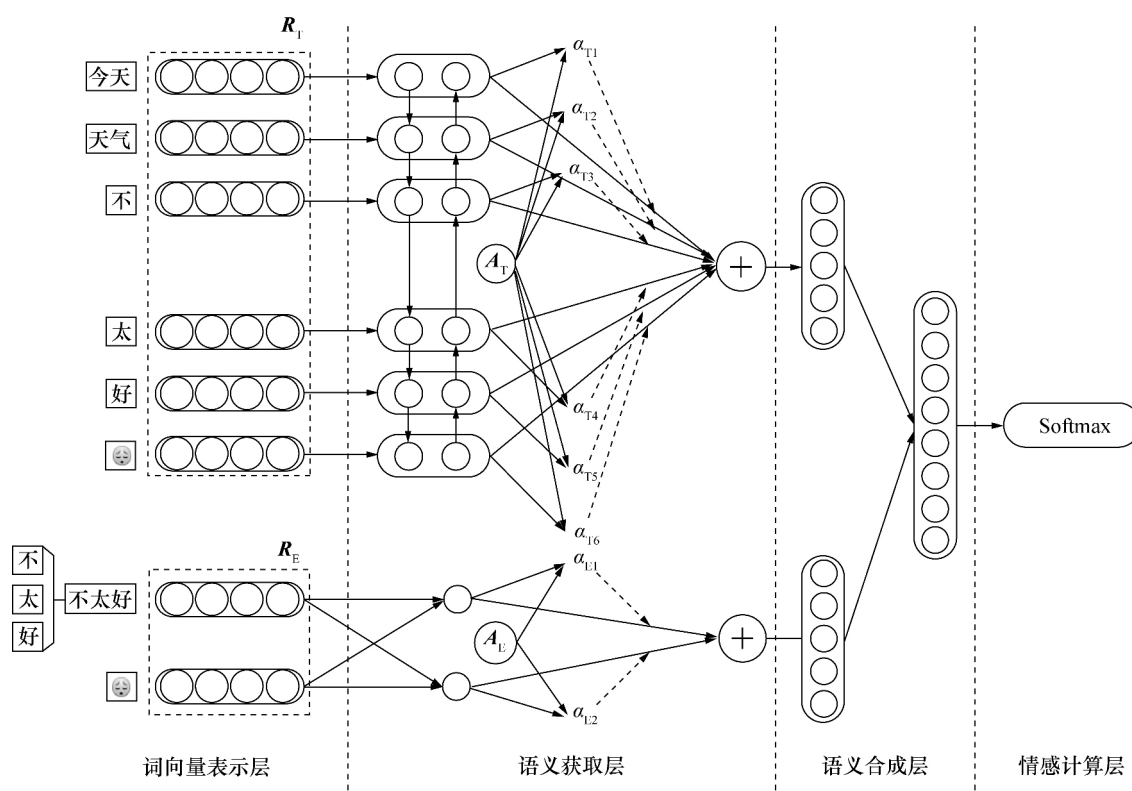


图1 双重注意力模型架构

2.3.2 语义获取层

对于微博文本,除了需要考虑词语之间的语义依赖关系,还要体现出哪些词语对于情感的表现更加重要,对于情感表现重要的词语需要赋予更高的权重。因此,本文采用 BLSTM 和注意力模型对微博文本的语义信息进行编码。对于情感符号集合,为了找出其中最重要的情感信号,采用全连接网络与注意力模型相结合的方式对情感符号进行编码。

由于 LSTM^[18]具有较强的长距离语义捕获能力,因此在文本处理中被广泛使用。但是,普通的 LSTM 模型仅能捕获文本中的正向语义信息,缺乏对逆向语义信息的捕获,因此本文在对文本的语义信息进行编码时采用 BLSTM 模型^[19]。BLSTM 模型是正向的 LSTM 模型和逆向的 LSTM 模型的结合,能够同时捕获正向的语义信息和逆向的语义信息。LSTM 通过一个记忆单元 c 、3 个门结构(输入门 x 、遗忘门 f 和输出门 o)来控制当前网络中信息的流量。具体来说,当前时刻记忆单元的信息含量 c_t 和输出的信息量 o_t 是由当前时刻的输入 x_t 、当前时刻遗忘门的状态 f_t 、前一时刻记忆单元的信息含量 c_{t-1} 以及前一时刻的输出 o_{t-1} 共同决定的。在

BLSTM 中,各个结构之间的关系可以简单地表示为:

$$\vec{c}_t, \vec{o}_t = G(\vec{c}_{t-1}, \vec{o}_{t-1}, \vec{f}_t, x_t), \quad (2)$$

$$\bar{c}_t, \bar{o}_t = G(\bar{c}_{t-1}, \bar{o}_{t-1}, \bar{f}_t, x_t). \quad (3)$$

其中 G 表示 LSTM 模型中各个结构之间的关系。在 BLSTM 中,某一时刻 t 的输出状态由当前时刻的正向 LSTM 和逆向 LSTM 的输出相连接组成,可以表示为 $o_t = [\vec{o}_t, \bar{o}_t]$ 。正向的语义信息和逆向的语义信息以相同的地位被考虑。

由于每条微博文本中的每个情感符号之间只是具有情感加强和减弱的功能,如相同极性的情感符号同时出现或是出现增强性的程度副词,情感将会得到加强,相反极性的情感符号同时出现或是出现减弱性的程度副词或者否定词,情感将会得到减弱甚至反转,情感符号相互之间并不存在较强的语义依赖关系,因此采用全连接网络对情感符号集合进行非线性变换,进行最初的语义编码。

为了捕获更直接的语义依赖关系,本文将微博文本 BLSTM 模型的输出和情感符号集合全连接的输出分别输入到注意力模型中,采用双重注意力模型分别对微博文本和情感符号进行编码。注意力模型通过对每一个输入状态 h_t 进行加权,从而实现为

输入文本构造一个上下文相关的文本表示向量 \mathbf{v} ,

$$\mathbf{v} = \sum_{t=1}^T \alpha_t \mathbf{h}_t. \quad (4)$$

其中: T 为输入状态的数目, α_t 为输入状态 \mathbf{h}_t 的权重。与标准的注意力模型^[9]不同的是, 本文注意力模型中的 α_t 与每一时刻的输入状态和一个随机初始化的上下文向量 \mathbf{A} 有关, \mathbf{A} 可以看作是对输入的一种语义表示。 α_t 通过式(5)和(6)计算得到:

$$\alpha_t = \frac{\exp(\mathbf{e}_t^T \mathbf{A})}{\sum_{k=1}^T \exp(\mathbf{e}_k^T \mathbf{A})}, \quad (5)$$

$$\mathbf{e}_t = \text{Tanh}(\mathbf{W}\mathbf{h}_t + \mathbf{b}). \quad (6)$$

式(5)和(6)中: \mathbf{W} 为模型中的权重, \mathbf{b} 为模型中的偏置。两者和 \mathbf{A} 一起作为模型训练过程中的参数, 通过不断训练学习得到。通过以上注意力模型, 可以为输入构造一个定长的上下文相关的文本表示向量, 向量中的信息包含各个输入状态的重要程度。

2.3.3 语义合成层

语义合成层的主要任务是将 \mathbf{S}_T 和 \mathbf{S}_E 进行合并, 构建微博文本整体的语义表示向量。在本文中, 为了简化模型的计算量, 仅对 \mathbf{S}_T 和 \mathbf{S}_E 采用按行连接的方式进行语义合并, 构建一个 $(r_T + r_E) \times c$ 的矩阵 \mathbf{S} , 对微博文本进行语义表示。其中: r_T 和 r_E 表示 \mathbf{S}_T 和 \mathbf{S}_E 的行数, c 表示 \mathbf{S}_T 和 \mathbf{S}_E 的列数。

2.3.4 情感计算层

情感计算层的主要任务是构建情感分类器, 获取微博文本的语义表示 \mathbf{S} 对于每一个情感标签的得分向量, 并输出微博文本的最终情感标签。本文采用 Softmax 分类器构建每一个情感标签的得分向量, 并将其转化为条件概率分布,

$$P_i(x) = \frac{\exp(x_i)}{\sum_{j=1}^C \exp(x_j)}, \quad i = 1, 2, \dots, C. \quad (7)$$

其中 C 表示情感标签的数目。为了更好地训练模型, 本文采用交叉熵损失函数来衡量情感标签的真实概率分布 $P_i^1(d)$ 和预测的概率分布 $P_i^p(d)$ 之间的差距,

$$\text{loss} = - \sum_{d \in T} \sum_{i=1}^C P_i^1(d) \log_2(P_i^p(d)). \quad (8)$$

其中: T 为训练数据集, d 为 T 中的一篇微博文本。本文利用式(8)的交叉熵损失函数, 采用反向传播机制对模型中的参数进行训练和更新。

3 实验与分析

3.1 数据集

本文的实验数据采用 NLPCC 的 2013 年和 2014 年中文微博情感评测任务的公共数据集 NLPCC2013 和 NLPCC2014, 其微博文本来源于新浪微博平台, 每个数据集分为训练集和测试集两部分, 每一条微博文本标注为一个主要的情感标签和多个次要的情感标签。情感标签总共分为 8 种, 采用文[20]的分类标准, 并将无情感的微博文本标注为情感标签 none。以各个类别作为主要情感标签的微博文本的数目如表 2 所示。

表 2 NLPCC2013 和 NLPCC2014 数据集中微博文本的数目

情感标签	NLPCC2013		NLPCC2014	
	训练集	测试集	训练集	测试集
happiness	370	1 116	1 459	441
like	595	1 558	2 204	1 042
anger	235	436	669	128
sadness	385	744	1 173	189
fear	49	102	148	46
disgust	425	935	1 392	389
surprise	113	236	362	162
none	1 828	4 873	6 591	3 603
总计	4 000	10 000	13 998	6 000

表 2 中, NLPCC2014 的训练数据中的 happiness 情感类别和 sadness 情感类别中各有一条微博内容为空, 本文选择将其删除。为了进一步验证本文提出的 DAM 模型的性能, 还设计了微博文本主客观分类任务和微博文本情感正负极性分类任务。在主客观分类任务中, 将情感标签 none 作为客观标签, 其他 7 类情感标签作为主观标签; 在正负极性分类任务中, 将 happiness、like 作为正向标签, 将 anger、sadness、fear、disgust、surprise 作为负向标签, 具体数据集大小如表 3 所示。

表 3 情感二分类数据集中微博文本的数目

年份	客观	主观	正向情感	负向情感
2013	6 701	7 299	3 639	3 660
2014	10 194	9 804	5 146	4 658
总计	16 895	17 103	8 785	8 318

为了构建词语和表情符号的词向量, 本文采用开源的爬虫框架 WebCollector 实现了一个微博爬虫

程序,并采用 NLPPIR 工具对爬取的微博文本进行分词,在分词的过程中,将构建的微博情感符号库中的情感符号作为分词词典,使其在分词后作为一个基本的语言单元存在。此外,本文还进行了一系列的微博文本预处理,如繁简转换、url 替换、短小无意义微博的剔除等。最终,构建了一个 Word2Vec 词向量训练语料,包含微博 40 292 997 条,词语 1 342 646 318 个。训练的过程中采用 Skip-gram 模型,其他相关参数均采用默认设置。经过训练,得到了一个包含 850 599 个词的词向量空间,其中每个词的词向量维数为 200。

3.2 实验参数设置

本文采用 Hyperopt 库进行分布式参数调节,获取模型的最优参数集合,模型参数调节列表如表 4 所示。

表 4 模型参数调节列表

参数	说明	取值范围	取值
optimizer	模型优化器	Adadelta, RMSprop	Adadelta
lstm_units	LSTM 输出维度	100, 150, 200, 250	200
batch_size	梯度下降时,每个 batch 样本数	8, 16, 32, 64	32
dropout	每次更新参数时随机断开输入神经元的比例,用于防止过拟合	[0.05, 0.95]	在输入层之后和 Softmax 层之前分别使用了 dropout,取值分别为 0.30、0.50

3.3 实验对比模型

MNB 模型(multinomial naïve Bayes, MNB)^[21]: 作为传统机器学习的代表, MNB 在大量多分类的情感分类任务中都取得了优秀的效果。

MCNN 模型(multichannel convolutional neural network, MCNN)^[6]: 该模型采用多通道的卷积神经网络对微博文本的词向量表示进行特征学习,从而构建文本的语义表示,实现文本的情感分类,这是较早采用深度学习模型进行情感分析的模型之一。

ESM 模型^[15]: 该模型通过计算微博文本中词语的词向量与微博表情符号词向量的余弦函数值,实现词语到情感空间的映射,采用 SVM 模型进行情感分类。该模型分两种: 1) 对映射后的词语直接求和, 2) 对映射后的词语采用求最大、最小和求和策略。本文选择效果较好的上述第 2 种方法作为对照模型。

EMCNN 模型(emotion-semantic-enhanced multi-channel convolutional neural network, EMCNN)^[16]: 该模型为常见微博表情符号构建语义特征表示矩阵,通过矩阵乘积运算,使用表情符号的语义特征表示矩阵对微博文本进行语义增强,采用多通道的卷积神经网络进行特征学习,实现情感分类。

BLSTM 模型: 该模型不引入注意力模型和情感符号,直接采用 BLSTM 模型对微博文本建模,并进行情感分析,为本文模型的基准模型。

SAM 模型(single attention model, SAM): 只考虑对微博文本采用注意力模型进行相应编码的一个对照模型,用于与本文提出 DAM 模型进行对比,以体现出引入微博情感符号的 DAM 模型的优越性。

3.4 实验结果分析

3.4.1 NLPCC 微博情感测评任务

按照 NLPCC 的中文微博情感评测任务的相关要求和评价指标,本文分别对 NLPCC2013 和 NLPCC2014 数据集进行了实验分析,结果如表 5 和 6 所示。

表 5 NLPCC2013 实验结果

模型	F_1	
	微平均	宏平均
MNB	0.320 8	0.310 5
MCNN	0.436 5	0.346 2
ESM	0.439 0	0.350 0
EMCNN	0.442 2	0.351 7
BLSTM	0.407 8	0.315 1
SAM	0.437 7	0.357 0
DAM	0.454 8	0.365 6

表 6 NLPCC2014 实验结果

模型	F_1	
	微平均	宏平均
MNB	0.359 5	0.278 3
MCNN	0.453 4	0.382 9
ESM	0.442 5	0.378 6
EMCNN	0.472 3	0.394 0
BLSTM	0.445 6	0.338 5
SAM	0.477 9	0.384 2
DAM	0.494 4	0.414 2

从表 5 和 6 可以看出, DAM 模型在所有的指标上都取得了最好的结果,相较于目前已知的最好模型 EMCNN,在 2013 年数据集上,微平均 F_1 和宏平均 F_1 分别提升了 1.26% 和 1.39%,在 2014

年数据集上,微平均 F_1 和宏平均 F_1 分别提升了 2.21% 和 2.02%,体现了个性化的情感符号的分类效果要好于全局化的情感符号,这与文[16]的结论是相一致的。同时,本文的 DAM 模型在 2014 年数据集上的性能提升要好于 2013 年的,这是由于 2014 年的训练集要比 2013 年的更大,模型得到了更加充分的训练。

SAM 模型在 2013 年数据集上的微平均 F_1 相对较低,排到了第 4 位,但是宏平均 F_1 指标却仅次于 DAM 模型,排在第 2 位;在 2014 年数据集上,微平均 F_1 排第 2 位,宏平均 F_1 排第 3 位。整体来说,SAM 模型的性能在 MCNN 模型之上,说明本文基于注意力机制对微博文本的情感分类是有效的。DAM 模型的性能明显高于 SAM 模型,说明本文的情感符号对于微博文本的情感分类具有很大的帮助。EMCNN 模型的性能稍好于 SAM,也验证了这一点。

3.4.2 情感二分类任务

在 NLPCC 情感测评任务的基础上,本文对表 3 所示的数据集,采用十折交叉验证的方法,设计了情感正负极性分类任务和主客观分类任务的实验,实验的结果采用准确率作为指标进行评价。具体的实验结果如图 2 和 3 所示。

在情感正负极性分类任务中,图 2 表明本文的 DAM 模型的准确率相较于 EMCNN 模型有稍微的提高,但是 SAM 模型的性能要差于 EMCNN 模型,这说明情感标签特别是微博中的表情符号对于情感正负极性分类任务具有较大贡献。

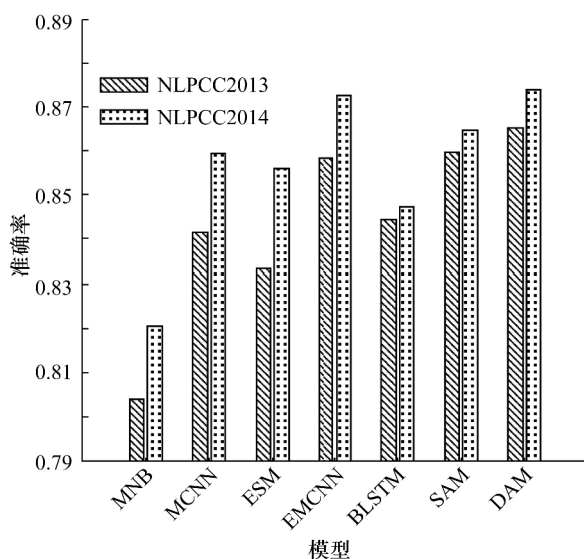


图2 正负极性分类结果对比图

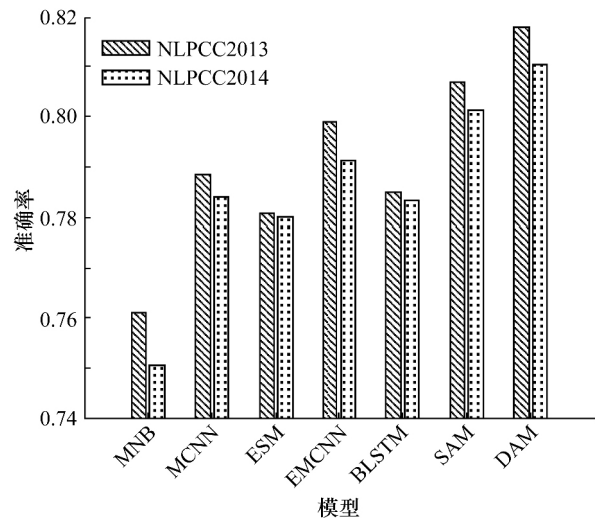


图3 主客观分类结果对比图

图 3 表明在主客观分类任务中本文的 DAM 模型仍然取得了最好的准确率,并且 SAM 模型优于 EMCNN 模型取得了次佳的准确率,这说明注意力模型的引入对于主客观分类任务具有很大的帮助。虽然 2013 年的训练数据量要少于 2014 年(见表 3),但是所有模型 2013 年的性能要好于 2014 年的,说明 2014 年数据集在主客观分类任务上要更加困难,这与文[15-16]的结论基本一致。

3.4.3 情感符号性能分析

为了验证各种情感符号对于微博文本情感分析的影响,本文在 SAM 模型的基础上,分别在 NLPCC2014 数据集的测评任务和情感正负极性分类任务中添加各种情感符号及其组合,设置了 6 组对照实验,分别是 S+S(SAM+情感词)、S+E(SAM+表情符号)、S+N(SAM+网络用语)、S+EN(SAM+表情符号+网络用语)、S+SEN(SAM+情感词+表情符号+网络用语)和 S+SNA(SAM+情感词+否定词+程度副词)。由于否定词和程度副词只有作用于情感词时,才会表现出其作用,且两者在数据集中较少,因此未单独与 SAM 模型进行组合,仅设置了与情感词搭配的对照组。6 组对照实验与 SAM 和 DAM 的对比结果如图 4 和 5 所示。

从图 4 和 5 中可以看出,在 SAM 模型的基础上,单独添加情感词、表情符号以及网络用语时,模型的性能都有提升,但是在测评任务中单独添加表情符号和网络用语时性能的提升不如情感正负分类任务中明显,在情感正负分类任务中单独添加表情符号性能的提升甚至要高于单独添加情感词,将表情符号和网络用语同时添加时,情感正负分类任

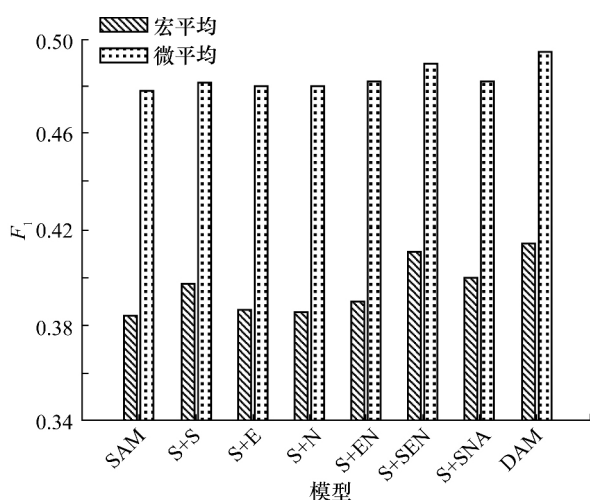


图4 测评任务中情感符号性能对比

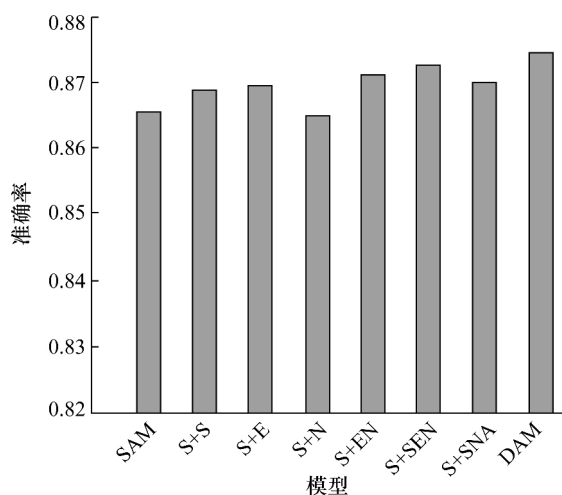


图5 正负极性分类任务中情感符号性能对比

务的性能提升也要好于测评任务。这说明了微博符号和网络用语对于正负情感的表现能力较多种复杂情感的表现能力要强一些。同时,否定词和程度副词对于性能的提升作用不是很明显,这主要与数据集中否定词与程度副词出现较少有关。

3.4.4 微博文本语义表示的可视化分析

为了更清楚地展现本文模型对于微博文本的表现力,本文采用主成分分析(PCA)降维技术对DAM模型中微博文本的语义表示进行降维处理,并将降维后的结果进行可视化。具体来说,本文对NLPCC2014数据集的测评任务、情感正负分类任务中的微博文本语义表示向量采用PCA技术从103 401维降低为2维,在平面直角坐标中进行可视化展示,可视化后的结果如图6和7所示。

从图6和7中可以看出,本文的DAM模型能够较好地将相同情感的语义表示聚集在一起,同时也能较好地地区分不同情感的语义表示。在测评任务

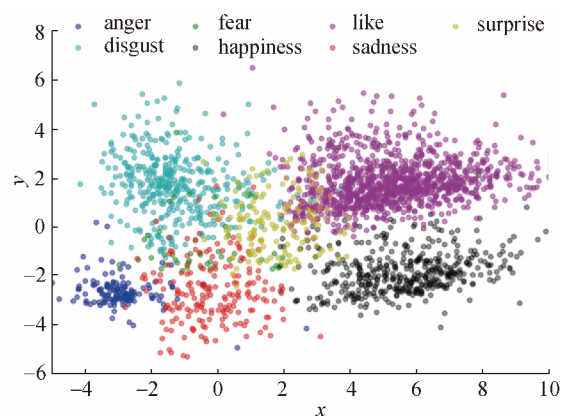


图6 (网络版彩图)2014年数据测评任务分析可视化结果

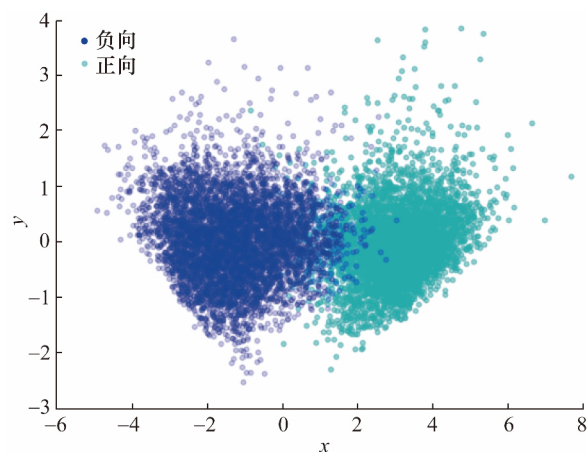


图7 2014年数据正负极性分类任务可视化结果

中,情感标签不同但是情感极性相近的标签,在图6中具有较近的距离,而极性相反的标签具有较远的距离。在情感正负极性分类任务中,各个类别的聚类效果相较于测评任务中各类别的聚类效果更好。这是由于在情感正负极性分类任务中,训练数据比较充足,模型训练得比较充分,这与测评任务中训练数据比较充分的大类情感标签聚类效果较好而训练数据比较少的小类情感标签聚类效果较差是相吻合的。同时,情感正负极性分类任务中的情感类别相对较少,这也使得数据聚集得比较充分。

4 总结

本文提出了一种基于双重注意力模型的微博情感分析方法,利用现有的情感分析资源构建了一个包含情感词、程度副词、否定词、微博表情符号和常用网络用语的微博情感符号库,引入注意力模型对微博文本和文本中包含的情感符号分别进行编码,构建微博文本的语义表示。本文方法通过将情感符号和注意力模型相结合,有效地增强了模型对于文本情感语义的捕获能力,进而提升了对微博文

本的情感分析性能。在 NLPCC 微博情感测评数据集上,本文提出的模型在多个任务上都表现出优越的性能,并在多个指标上超过了已知的最好模型。

由于数据的限制,本文的情感符号库中收集的情感符号还不够全面。同时,对于各种情感符号语义的捕捉能力的度量方法还有待研究。因此,下一步的工作是进一步扩展情感符号库,并探究各种情感符号对于语义表示的影响。

参考文献 (References)

- [1] 丁兆云,贾焰,周斌. 微博数据挖掘研究综述 [J]. 计算机研究与发展, 2014, 51(4): 691-706.
DING Z Y, JIA Y, ZHOU B. Survey of data mining for microblogs [J]. Journal of Computer Research and Development, 2014, 51(4): 691-706. (in Chinese)
- [2] TABOADA M, BROOKE J, TOFILOSKI M, et al. Lexicon-based methods for sentiment analysis [J]. Computational Linguistics, 2011, 37(2): 267-307.
- [3] WIEBE J, WILSON T, CARDIE C. Annotating expressions of opinions and emotions in language [J]. Language Resources and Evaluation, 2005, 39(2): 165-210.
- [4] PANG B, LEE L. Opinion mining and sentiment analysis [J]. Foundations and Trends in Information Retrieval, 2008, 2(1/2): 1-135.
- [5] PANG B, LEE L, VAITHYANATHAN S. Thumbs up?: Sentiment classification using machine learning techniques [C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA, 2002: 79-86.
- [6] KIM Y. Convolutional neural networks for sentence classification [C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Doha, Qatar, 2014: 1746-1751.
- [7] ZHANG Y S, JIANG Y R, TONG Y X. Study of sentiment classification for Chinese microblog based on recurrent neural network [J]. Chinese Journal of Electronics, 2016, 25(4): 601-607.
- [8] WANG J, YU L C, LAI K R, et al. Dimensional sentiment analysis using a regional CNN-LSTM model [C]// Proceedings of the Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016: 225-230.
- [9] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [C]// Proceedings of the International Conference on Learning Representations. San Diego, USA, 2015.
- [10] LUONG M T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation [C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal, 2015: 1412-1421.
- [11] HERMANN K M, TOMÁS K, GREFFENSTETTE E, et al. Teaching machines to read and comprehend [C]// Proceedings of the International Conference on Neural Information Processing Systems. Montreal, Canada, 2015: 1693-1701.
- [12] RAFFEL C, ELLIS D P W. Feed-forward networks with attention can solve some long-term memory problems [C]// Proceedings of the International Conference on Learning Representations. San Juan, Puerto Rico, 2016.
- [13] 栗雨晴,礼欣,韩煦,等. 基于双语词典的微博多类情感分析方法 [J]. 电子学报, 2016, 44(9): 2068-2073.
LI Y Q, LI X, HAN X, et al. A bilingual lexicon-based multi-class semantic orientation analysis for microblogs [J]. Acta Electronica Sinica, 2016, 44(9): 2068-2073. (in Chinese)
- [14] BARBOSA L, FENG J. Robust sentiment detection on Twitter from biased and noisy data [C]// Proceedings of the International Conference on Computational Linguistics. Beijing, 2010: 36-44.
- [15] JIANG F, LIU Y Q, LUAN H B, et al. Microblog sentiment analysis with emoticon space model [J]. Journal of Computer Science and Technology, 2015, 30(5): 1120-1129.
- [16] 何炎祥,孙松涛,牛菲菲,等. 用于微博情感分析的一种情感语义增强的深度学习模型 [J]. 计算机学报, 2017, 40(4): 773-790.
HE Y X, SUN S T, NIU F F, et al. A deep learning model enhanced with emotion semantics for microblog sentiment analysis [J]. Chinese Journal of Computer, 2017, 40(4): 773-790. (in Chinese)
- [17] SONG K, FENG S, GAO W, et al. Build emotion lexicon from microblogs by combining effects of seed words and emoticons in a heterogeneous graph [C]// Proceedings of the ACM Conference on Hypertext and Social Media. New York, USA, 2015: 283-292.
- [18] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 2012, 9(8): 1735-1780.
- [19] GRAVES A, JAITLY N, MOHAMED A R. Hybrid speech recognition with deep bidirectional LSTM [C]// Proceedings of the Automatic Speech Recognition and Understanding. Olomouc, Czech Republic, 2013: 273-278.
- [20] 徐琳宏,林鸿飞,潘宇,等. 情感词汇本体的构造 [J]. 情报学报, 2008, 27(2): 180-185.
XU L H, LIN H F, PAN Y, et al. Constructing the affective lexicon ontology [J]. Journal of the China Society for Scientific and Technical Information, 2008, 27(2): 180-185. (in Chinese)
- [21] BERMINGHAM A, SMEATON A F. Classifying sentiment in microblogs: Is brevity an advantage? [C]// Proceedings of the ACM International Conference on Information and Knowledge Management. Toronto, Canada, 2010: 1833-1836.

(责任编辑 李丽)