

基于上下文的情感词向量混合模型

霍丹¹ 张生杰² 万路军¹

1 空军工程大学 西安 710048

2 西安理工大学计算机科学与工程学院 西安 710048

(wan-95@163.com)



摘要 针对现有大多数基于词向量的学习方法只能对词语的语法语境建模,而忽略了词语的情感信息的问题,文中提出了基于上下文的情感词向量训练模型,使用了比较简单的方法来构建情感词向量的学习框架。该模型是能够获取情感的扩展混合模型在句子极性的情感信息和基于上下文级别词向量的融合方法,有效解决了具有相似上下文但相反情感极性的词被映射到相邻的词向量的问题。为验证学习到的情感词向量模型能准确包含情感和上下文词语的语义信息,分别在不同的语言和不同领域的数据集下训练情感词向量,并在词语级别进行了定量实验。结果表明,所提的情感词向量学习模型在情感词向量获取实验中,与传统的词向量学习模型相比,分类效果提升了14个百分点;在词语级别的情感分类实验中,与传统的词袋模型相比,准确性提升了10个百分点,从而也对产品提供商在大量的用户评价中得到有用的信息起到了指导性的作用。

关键词: 自然语言处理;词向量;语义信息;情感分类;神经网络

中图法分类号 TP389.1

Context-based Emotional Word Vector Hybrid Model

HUO Dan¹, ZHANG Sheng-jie² and WAN Lu-jun¹

1 Air Force Engineering University, Xi'an 710048, China

2 School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China

Abstract Most of the existing learning methods based on word vectors can only model the syntactic context of words, but ignore the emotional information of words. This paper proposes a context-based training model of emotional word vectors, and uses a relatively simple method to construct a learning framework of emotional word vectors. A fusion method is proposed to obtain the emotion information of the extended mixed model in the sentence polarity and the context-based word vectors. So as to solve the problem that words with similar contexts but opposite emotional polarity are mapped to adjacent word vectors, the adjacent words in the emotion vector space are semantically similar and have the same emotion polarity. In order to verify that the learned emotion word vector model can accurately contain the semantic information of emotion and context words, the emotion word vector is trained in different languages and data sets of different fields, and quantitative experiments are conducted at the word level. The results show that the classification effect of the proposed model is 14 percent higher than that of the traditional model. In the experiment of emotion classification at the word level, the accuracy is improved by 10 percentage points compared with the traditional word bag model. It also plays a guiding role for product providers to get useful information in a large number of user reviews.

Keywords Natural language processing, Word embeddings, Semantic information, Sentiment classification, Neural networks

1 引言

最早的词的表示方法为独热(one-hot)表示,该方法把每个词表示为一个长向量^[1]。词向量的维度是词表大小,其中绝大多数元素为0,只有一个维度的值为1,该维度即代表了当前词,这种独热表示方式如果采用稀疏方式存储,将会非常简洁,但也存在维度大以及“词汇鸿沟”现象,即所有词之间都是孤立的,仅从这两个向量中看不出两个词是否有关联^[2]。

而目前作为自然语言处理(Natural Language Processing, NLP)领域的热门研究方向,词向量越来越被研究者们所喜爱^[3]。

对于情感分类问题,构建特征的有效方法是利用句子或文档所含的词语或短语来表示它们的特征^[4-5],其中关键的一步是将词语或短语学习到一个低维、实数、非稀疏的词语向量,并且使用词向量进行表示(word vector representation)。现存的基于向量的词语学习方法^[6]均能获得词向量,但却不

本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61703452);陕西省自然科学基金发展计划(2016JQ6062)

This work was supported by the National Natural Science Foundation of China (61703452) and Research Plan of the National Natural Science and Development of Shaanxi Province, China (2016JQ6062).

通信作者:张生杰(754272317@qq.com)

能直接应用到情感分类。究其原因是一些传统方法,如词袋模型(Bag of Words, BoW)^[7],尽管其覆盖面大,但是主要是针对词语的语法、语境进行建模,无法捕捉许多蕴含情感倾向性的语言特征。这就会导致一些具有相同用法但情感极性却相反的词语被映射到“相近”的词向量中^[8],例如在词性标注和分词任务中,“good”与“bad”两个用法相同的词会映射到“相近”的词向量中^[9],然而,他们具有相反的情感极性,这对于情感分类任务将是一个失败的问题。

在许多自然语言处理的任务中,词向量都有着优异的表现,在情感分析任务中,词向量也被作为新颖的特征使用。例如,Luo等提出的基于n-gram及tf-idf特征的微博情感分类方法^[10];Chen提出的基于情感词典的中文微博情感倾向分析^[11]。

随着深度学习的发展,词特征向量的表示主要基于深度学习算法得到。比如Turian等^[12]将词分布表示方法分为3类:基于矩阵的分布表示、基于聚类的分布表示和基于神经网络的分布表示。在基于神经网络的文本分类方法中,最常使用的是基于矩阵的分布式表示和基于神经网络的分布式表示,Pennington等^[13]在2014年提出了Glove模型,该模型通过对“词-词”矩阵进行分解从而得到词表示。基于神经网络的词分布式表示通常被称为词嵌入或词向量,早期的词向量只是神经网络语言模型的副产品。Mikolov等^[14]在2013年提出了CBoW(Continuous Bag-of-Words)和Skip-gram模型,这两个模型直接以生成词向量为目的,能够快速高效地从大规模无标注语料中学习语义丰富的词向量。Irsay等采用循环神经网络(Recurrent Neural Network, RNN)对文本句子层级进行建模^[15];Pilehvar等使用更为复杂的长短期记忆模型(Long-Short Term Memory, LSTM)进行情感分析^[16]。词特征向量可以很好地体现词在统计语料中的语义分布特征,通过计算向量之间的距离来体现词语之间的相似性^[17]。然而,传统的词向量仅考虑了文本中的语义或语法信息,而忽略了词语所要表达的情感,即用户对不同主题的偏好及其他用户的评论,因此需要结合文本的上下文来反映不同的情感立场。因此,为能较准确地学习情感词向量,本文提出了基于上下文的情感词向量学习模型。为了验证情感词向量学习模型在情感分析任务中的有效性,文中在中文和英文数据对比试验下,定量评估了基于上下文的情感词向量在情感分类任务中的准确率。

2 情感词向量模型的建立

在建立情感词向量的模型时,首先,将基于上下文的深度学习用于建立词向量模型;接着,介绍了融合情感的扩展混合模型对句子极性的编码方式和上下文词语的相关情感词向量模型;最后,描述词语级词向量的集成向量学习方法。

2.1 相关工作

目前词向量学习的主流算法以分布假说为基础^[18-19],其中词的表示可以由它们的上下文反映。本节首先创建一个基于上下文的预测模型(Prediction Model)和排名模型(Ranking Model)进行编码,以学习词向量的上下文;然后,以这些模型为基本模型,通过抽取这些模型中的情感信息来学习形成基于上下文的情感词向量模型;最后,将学习到的基于上下文词向量与具有情感信息的词向量模型融合,构建了混合情

感词向量模型,可以输出整句的情感信息。

本节使用以下变量: w_i 表示在句子中索引为*i*的单词, h_i 是一个句子中 w_i 的上下文词, e_i 是 w_i 的词向量。然后,采用具有基本神经元要素的深度学习方法^[20],各层依次为lookup, hTanh, linear和softmax。每个神经元中的 O_{layer} 指输出向量。

2.2 词的上下文模型

在现存方法中,单词上下文编码的有效表示方法大部分采用“上下文预测”模型^[21]。给定一个目标词 w_i 及其上下文词 h_i ，“上下文预测”可以基于 h_i 预测 w_i ,并将 h_i 视为语言模型。目标词的上下文可以在前文、后文或者周围词的文本中。本文并不关注语言模型的准确性,所以可以假设文本中目标词周围的词为 $h_i = \{w_{i-c}, w_{i-c+1}, \dots, w_{i-1}, w_{i+1}, \dots, w_{i+c-1}, w_{i+c}\}$,这个方法可以自然地扩展到预测前面或后面的单词。构建一个预测模型类似于Bengio等^[20-21]提出的具有代表性“上下文预测”的神经语言模型,其模型中的条件概率就是基于上下文词 h_i 预测目标词 w_i 的概率 $P(w_i | h_i)$,而其中的得分函数是一个由lookup→linear→hTanh→linear→softmax层组成的前馈神经网络。

lookup层(也称为查找层)包含一个查找表 $LT \in R^{d \times |V|}$,将每个词映射到查找表中让其成为连续的向量。其中*d*是每个词向量的维度,*V*是词汇量的大小。查找操作可以被视为投影函数的计算,使用一个除了第*i*个索引之外的所有位置为零的二进制列向量 idx_i ,如式(1)所示。

$$e_i = LT \cdot idx_i \in R^{d \times 1} \quad (1)$$

连接上下文的词向量 $\{w_{i-c}, w_{i-c+1}, \dots, w_{i-1}, w_{i+1}, \dots, w_{i+c-1}, w_{i+c}\}$ 作为查找层的输出,形式化的表示如式(2)所示。

$$O_{lookup} = [e_{i-c}, \dots, e_{i-1}, e_{i+1}, \dots, e_{i+c}] \in R^{1 \times d \cdot 2c} \quad (2)$$

然后, O_{lookup} 被送到线性层进行维度变换,如式(3)所示。

$$O_{l1} = W_{l1} \cdot O_{lookup} + b_{l1} \quad (3)$$

其中, $W_{l1} \in R^{2c \times len}$ 是位置相关的权值, $b_{l1} \in R^{1 \times len}$ 是线性层的偏置,*len*是线性层输出向量 O_{l1} 的长度。

线性层之后是非线性的功能层,用于添加非线性的元素。标准的非线性层包括双曲正切函数、hTanh、sigmoid和Rectifier函数^[24]。本文选用hTanh函数计算效率和有效性。用 $O_{hTanh} \in R^{1 \times len}$ 表示hTanh的输出向量,如式(4)所示。

$$hTanh(x) = \begin{cases} -1, & \text{if } x < -1 \\ x, & \text{if } -1 \leq x \leq 1 \\ 1, & \text{if } x > 1 \end{cases} \quad (4)$$

神经语言模型的输出层是一个softmax层,其输出长度为词汇大小。Softmax^[25]适用于这种情况,因为其输出可以解释为条件概率,计算如式(5)所示。

$$softmax_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \quad (5)$$

条件概率 $P(w_i | h_i)$ 是softmax输出层的第 idx_{w_i} 的值,其中 idx_{w_i} 是词汇表 w_i 的索引。给定上下文词 (w_i, h_i) 的优化目标就是对其概率 $P(w_i | h_i)$ 求最大化。

真实计算过程中,直接预测概率 $P(w_i | h_i)$ 是非常耗时的,作为输出层softmax,它的长度是词汇表的长度,通常能达到 10^3 的数量级。为了加快训练速度,本文使用噪声对比估计函数^[26]来优化“上下文预测”的模型,通过逻辑回归算法对词语语境的真实情况或人工噪声加以区分,取自给定样本数

据的概率是 $P(D|\mathbf{w}, \theta)$ 。 D 表示数据集, \mathbf{w} 是数据集中的词向量, θ 是函数的参数, 如式(6)所示。

$$P(D|\mathbf{w}, \theta) = \frac{\exp(f_\theta(\mathbf{w}_i, \mathbf{h}_i))}{\exp(f_\theta(\mathbf{w}_i, \mathbf{h}_i)) + k \cdot \exp(f_\theta(\omega^n, \mathbf{h}_i))} \quad (6)$$

其中, ω^n 是人为噪声, 例如从词表中随机选择的词, 缩放因子 k 说明噪声样本比数据样本的频率高 k 倍; 得分函数 $f_\theta(\mathbf{w}_i, \mathbf{h}_i)$ 可以量化处理上下文词 \mathbf{h}_i 和目标词 \mathbf{w}_i 。通过上述处理方法, 可以很自然地将前馈神经网络变为 $\text{lookup} \rightarrow \text{linear} \rightarrow \text{hTanh} \rightarrow \text{linear}$ 。查找层的输入是目标词 \mathbf{w} 和上下文词 \mathbf{h} 的连接, 输出是长度为 1 的线性层, 表示上下文词 \mathbf{h} 和目标词 \mathbf{w} 之间的相关性。用 softmax 层来实现概率 $P(D|\mathbf{w}, \theta)$, 并且将 softmax 最大值的对数概率用来进行参数估计。故, 可得到如式(7)所示的基于上下文预测模型的损失函数计算公式。

$$\text{loss}_{\text{cPred}} = - \sum_{\mathbf{w} \in T} \log P(D|\mathbf{w}, \theta) \quad (7)$$

C&W(Collobert & Weston)模型使用成对排序的方法来获取词语的上下文进行词向量的学习, 它使用噪声对比估计函数实现这个过程, 但是优化的目标是通过将真实上下文词语对 $(\mathbf{w}_i, \mathbf{h}_i)$ 和人工噪声词语对 (ω^n, \mathbf{h}_i) 进行边缘化的比对, 然后获取一个更高的分数。用合页损失函数进行最小化处理^[27], 得到如式(8)所示的基于上下文排名模型, 其中 T 是训练语料。

$$\text{loss}_{\text{cRank}} = \sum_{(\mathbf{w}_i, \mathbf{h}_i) \in T} \max(0, 1 - f_\theta(\mathbf{w}_i, \mathbf{h}_i) + f_\theta(\omega^n, \mathbf{h}_i)) \quad (8)$$

其中, 得分函数 $f_\theta(\mathbf{w}_i, \mathbf{h}_i)$ 通过前馈神经网络实现, 其输入是当前词 \mathbf{w}_i 和上下文词 \mathbf{h}_i 的连接, 输出是仅有一个结点的线性层, 代表了 \mathbf{w} 和 \mathbf{h} 之间的相关性。训练期间, 在均匀分布下的词汇表上随机选择人工噪声 ω^n 。

2.3 情感极性模型

具有情感信息的预测模型 (Structural Emotional-Structural Prediction Model, SE-SPred) 的基本思想是将情感预测作为多分类任务, 将词向量作为参数来预测词序列的积极/消极分类概率。特别地, 给定一个可变大小的句子, 滑过一个句子的固定窗口的词, 并基于当前词向量预测每个窗口句子的情感极性。

预测模型包含 5 层, 分别是 $\text{lookup} \rightarrow \text{linear} \rightarrow \text{hTanh} \rightarrow \text{linear} \rightarrow \text{softmax}$ 。其输入是固定长度的词序列 $\{\mathbf{w}_{i-c}, \mathbf{w}_{i-c+1}, \dots, \mathbf{w}_{i-1}, \mathbf{w}_{i+1}, \dots, \mathbf{w}_{i+c-1}, \mathbf{w}_{i+c}\}$, \mathbf{w}_i 是当前词, c 是窗口的大小。上一节详细介绍了 Lookup, linear 和 hTanh 层, 输出层 hTanh 用来预测输入特征的消极与积极概率, 因为神经网络中较高层的连续表示形式可以将输入的词向量进行抽象化, 以此来区分输入数据的特征。为了预测消极和积极的概率, 本文将 hTanh 层的输出传送到线性层, 向量的长度转换为类别数 C , 其在二进制的分类情况下分为两类。第二个线性层的参数为: $\mathbf{w}_{l2} \in R^{C \times len}$ 和 $b_{l2} \in R^{1 \times C}$, 再添加一个 softmax 层作为输出层, 以产生消极和积极类别的条件概率。

训练样本期间, 把具有显著性情感极性的词语作为输入。令 $f^s(t) \in R^{1 \times C}$ 为显著性的分布函数, 其中的一个输入为 t , C 是情感极性的标签数量, 在此只有积极和消极两类。 f^s 具有 $1-C$ 的编码形式, 与类别具有相同的维数, 并且只有同维度对应的真实值是 1, 其他的都是 0。例如, $f^s(t) = [1, 0]$ 代表一个句子是积极的属性, 而 $f^s(t) = [0, 1]$ 代表句子是消极的属性。使用显著性情感分布和预测分布之间的交叉熵误差作为 softmax 层的损失函数^[28]。对于语料库 T , 如式(9)给出的损

失函数, 可以利用标准的反向传播来学习参数。

$$\text{loss}_{\text{sPred}} = - \sum_t \sum_{k=(0,1)} f_k^s(t) \cdot \log(f_k^{\text{pred}}(t)) \quad (9)$$

作为预测模型的变换, 学习融入情感信息的排名模型 (Structural Emotional-Structural Rank Model, SE-SRank) 是根据输出具有固定窗口大小单词序列的得分而进行情感分类的, 它的基本原理是: 如果单词序列的显著性情感极性为积极的, 则预测的积极性得分一定高于消极性的得分。例如, 如果词序列相关联的两个值 $[f_{\text{pos}}^{\text{rank}}, f_{\text{neg}}^{\text{rank}}]$, 那么 $[0.7, 0.1]$ 的结果可以表示为积极的情况, 因为积极得分 0.7 大于消极得分 0.1。通过这种类比, $[-0.2, 0.6]$ 的结果就表示为消极性的。

基于上面的讨论, 本文利用了基于神经网络的排名模型^[29], 如图 1(b)所示。该排序模型是由 4 层组成的前馈神经网络 $\text{lookup} \rightarrow \text{linear} \rightarrow \text{hTanh} \rightarrow \text{linear}$, 与图 1(a)的预测模型相比, 去除了 softmax 层, 因为目标 Z' 是不需要计算概率的。将排名模型的输出向量表示为 $f^{\text{rank}} \in R^{1 \times C}$, 其中 $C=2$, 表示二进制积极与消极的分类, 给训练模型设计了一个边缘排名 (Margin Ranking) 损失函数, 其描述如式(10)所示。

$$\text{loss}_{\text{sRank}} = \sum_t \max(0, 1 - \delta_s(t) f_0^{\text{rank}}(t) + \delta_s(t) f_1^{\text{rank}}(t)) \quad (10)$$

其中, T 是训练语料库; f_0^{rank} 是预测的积极性得分; f_1^{rank} 是消极性的得分; $\delta_s(t)$ 是反映句子显著性情感极性(消极或积极)的指标函数, 如式(11)所示。

$$\delta_s(t) = \begin{cases} 1, & \text{if } f^s(t) = [1, 0] \\ -1, & \text{if } f^s(t) = [0, 1] \end{cases} \quad (11)$$

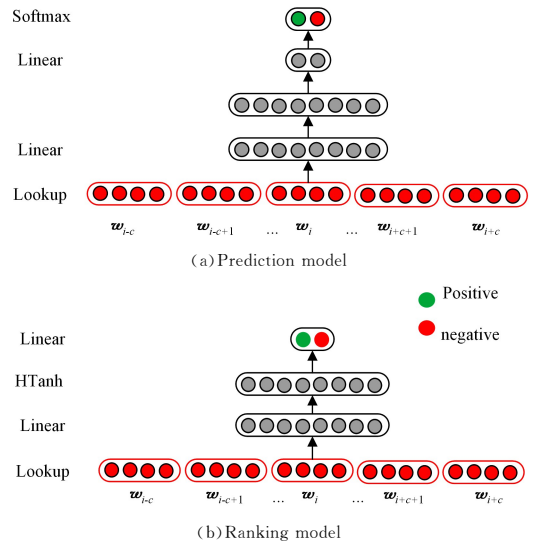


图 1 二元情感分类的结构模型图

Fig. 1 Structural model of binary emotional classification

2.4 混合情感极性模型

情感极性模型仅能够获取当前词和所在句子的情感信息, 通过神经网络的训练得到情感极性的输出, 而 MSEPM 在预测目标词的上下文的同时, 将滑动窗口内所有的词以及词的上下文通过加权的方式映射在一起, 再将整句的情感信息与以上方式融合, 更新词向量中的情感极性并输出。

结合基于上下文的预测模型和基于情感的预测模型, 文中提出了混合情感预测模型 (Structural Emotional-Hybrid Prediction Model, SE-HyPred)。由情感损失函数 $\text{loss}_{\text{sPred}}$ (式(9))和上下文预测函数 $\text{loss}_{\text{cPred}}$ (式(7))的加权线性组合而成, 可作为混合情感预测模型的损失函数, 来进行学习参数

的调整,其中权重 α_{pred} 的范围为 $[0, 1]$,如式(12)所示。

$$\text{loss}_{\text{hyPred}} = \alpha_{\text{pred}} \cdot \text{loss}_{\text{sPred}} + (1 - \alpha_{\text{pred}}) \cdot \text{loss}_{\text{cPred}} \quad (12)$$

此外,为了把词语和上下文级别的编码信息用词向量的形式表示出来,避免把情感特定的上下文作为线性层的参数,本节使用神经网络层 hTanh 作为共享的参数,混合情感预测模型由 5 层前馈神经网络组成,与图 2(a)的相比,混合情感排名模型去除了 softmax 层,因为目标 Z' 是不需要计算概率的。两种模型都是将词语的情感信息和上下文词向量融合作为网络结构的 Linear 的参数,通过各模型的损失函数学习参数的调整,最后输出词语的情感极性,如图 2 所示。

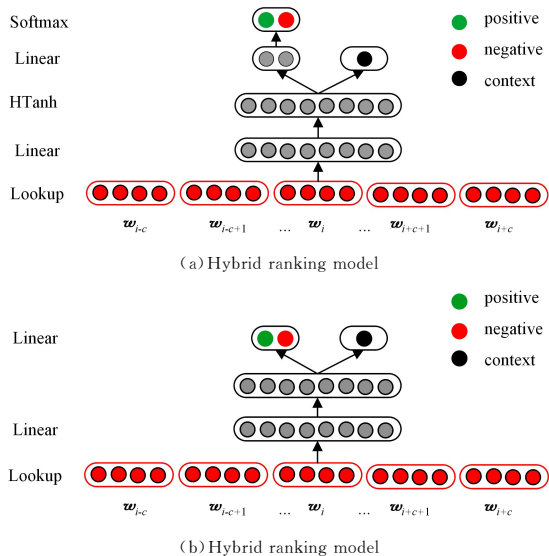


图 2 混合情感极性模型

Fig. 2 Mixed emotion polarity model

值得注意的是,人为生成的“噪声”仅用于计算基于上下文的损失函数 $\text{loss}_{\text{cPred}}$,而对于情感损失函数 $\text{loss}_{\text{sPred}}$ 没有影响。

与混合预测模型类似,合并上下文排名模型和情感排名模型,以得到这部分的混合排名模型 (Structural Emotional-Hybrid Rank Model, SE-HyRank),模型如图 2(b)所示。其损失函数是由情感排名损失函数 $\text{loss}_{\text{sRank}}$ (式 10)和上下文排名函数 $\text{look}_{\text{cRank}}$ (式 7)的加权线性组合而成,来调整混合排名模型的学习参数,如式(13)所示。

$$\text{loss}_{\text{hyRank}} = \alpha_{\text{rank}} \cdot \text{loss}_{\text{sRank}} + (1 - \alpha_{\text{rank}}) \cdot \text{loss}_{\text{cRank}} \quad (13)$$

其中,权重 α_{rank} 的范围为 $[0, 1]$ 。同理,人为生成的“噪声”仅用于计算基于上下文的损失函数 $\text{loss}_{\text{cRank}}$,而对于情感损失函数 $\text{loss}_{\text{sRank}}$ 没有影响。

3 模型的实验设置

3.1 参数学习设置

对于整个参数集而言,可通过反向传播获得损失函数的导数的方式来学习情感的嵌入,从均匀分布 $U(-0.01, 0.01)$ 来对词向量的值进行初始化,并且以 fan-in^[30] 输入长度来初始化线性层,即为均匀分布 $U\left(\frac{-0.01}{\text{InputLength}}, \frac{0.01}{\text{InputLength}}\right)$ 。根据初步实验结果,将窗口大小设置为 7(当前词的前 3 个字和后 3 个字),向量长度为 50,隐藏层的长度为 20。对于参数的更新,使用广泛应用于深度学习的 AdaGrad^[31],它的学习速率针对不同的参数可以自适应地改变,如式(14)所示。

$$\theta_t = \theta_{t-1} - \frac{\epsilon}{\sqrt{\sum_{\tau=1}^t g_{\tau}^2}} g_t \quad (14)$$

其中, θ_t 是在时间步长 t 下参数的值, g_t 是在时间步长 t 下的梯度, ϵ 是学习率并且设置为 0.1。

表 1 中英文情感词典描述

Table 1 Chinese and English emotional dictionary description

情感词典	积极倾向	消极倾向	总计
BL(英文)	1052	1201	2253
NRC(英文)	1157	2215	3372
中文	1285	2804	4089

3.2 实验数据集

对于中文情感词向量定量评估,本实验选取中文极性情感词典 NTUSD(National Taiwan University Sentiment Dictionary)中的情感词作为候选词集合。词向量训练数据集取自于 CORA2014 任务手机领域的评论,选择其中 2232 条积极评论和 2146 条消极评论。考虑到词向量的训练语料库不能完全覆盖 NTUSD 中的所有词,实验选用词向量集合与 NTUSD 的交集作为中文实验数据集,最终得到积极词 1285 个,消极词 2804 个。

对于英文情感词向量定量评估,词向量训练的数据集取自 NRC(National Research Council)推文。使用两个人工标注的情感词典,即 BL(Basic Language)情感词典与 NRC 情感词典。同样地,取词向量集合与两个情感词典的交集作为英文实验数据集。

在将数据输入到情感词向量训练模型之前,首先对数据进行预处理工作:

- (1)对于英文评论数据,将所有的词转化为小写;
- (2)对于中文评论数据,将所有的评论文本进行分词处理,对于中英文数据,分别使用 NLPIR(Natural Language Processing & Information Retrieval Sharing Platform)和 Stanford CoreNLP 提供的分词工具进行分词。

在实验中,情感词典与词向量词典的交集的情感极性分布如表 2 所列。

表 2 情感极性分布数据

Table 2 Emotional polarity distribution data

类别	名称	词数	描述
英文	BL	6789	2003 条表示肯定,4783 条表示否定
	NRC	62467	得分从 -5 到 5 的实数值,-5 表示极端否定,5 表示极端肯定
中文	NTUSD	11088	2810 条表示肯定,8276 条表示否定

4 实验过程与结果

4.1 情感词获取实验

一个好的情感分析任务,应该具有将积极词归并到其接近的向量中的能力,同样能将消极的词归并到其接近的向量中,并将这两类词明确地分开。因此,在情感嵌入的向量空间中,诸如“good”这样带有积极色彩词的相邻词应当由像“cool”“awesome”“great”等积极性的词占据;而诸如“bad”带有消极色彩的词的相邻位置应当是由像“terrible”“nasty”等消极性的词包围。基于这个思想,设计本实验来查找现有情感词典中的相邻情感词,以探究情感向量是否有助于发现情感词之间的相似性。

具体来说,给定一个情感词作为输入,首先,找出情感词

典中最接近该词的前 N_w 个词。本文将 N_w 设置为 10 和 30 来评估情感词向量的性能。两个词的邻近度通过词向量之间的相似性(例如余弦)来测量。然后,计算这些邻近词与目标情感词具有相同的情感极性的百分比。

极性一致性的百分比对应于精度评估度量,其计算如式(15)所示。

$$Accuracy = \frac{\sum_{i=1}^{\#Lex} \sum_{j=1}^{N_w} \delta_w(w_i, c_{ij})}{\#Lex \times N_w} \quad (15)$$

其中, $\#Lex$ 是情感词典中的词的数量; w_i 为词典中第 i 个词; c_{ij} 是词典中与 w_i 最接近的第 j 个词; $\delta_w(w_i, c_{ij})$ 是一个指标函数,如果 w_i 和 c_{ij} 具有相同的情感极性,则其值为 1,若具有相反的情感极性,则其值为 0。准确度越高,说明在计算情感词向量之间的相似性时,情感与词向量嵌入地更好。

为了验证情感词向量模型确实能够挖掘评论中的情感信息,本文在分析实验中构建了以下几个对比词向量。

(1) C&W: 该模型是一种用于学习词向量的代表性算法,是在 1.1 节中描述的上下文排名(context-ranking)模型。该模型并没有求一个字符串的概率,而是求窗口连续 n 个词的分值 f ,其值越高说明该 n 个词的正确性较高;其值越低说明词语不太合理;如果随机把几个词堆积在一起,则肯定为负分(差评)。

(2) Word2vec: Mikolov 等开发的 CBoW 和 Skip-gram 是学习连续词向量的模型,Google 将这两个模型组合到一个广泛使用的词向量工具包“word2vec”,它是用来产生词向量的相关模型。这些模型为双层的神经网络,通过训练以重新构建词语文本。其网络以词表现,并且需猜测相邻位置的输入词,在 word2vec 中词袋模型假设下,词的顺序是不重要的。训练完成之后,word2vec 模型可用于映射每个词到一个向量,或用来表示词对词之间的关系,该向量为神经网络的隐藏层。实验使用第 2 节介绍的 CBoW 模型,其类似于上下文预测(context-prediction)模型。

(3) ReEmbed: Labutov 等^[30]提出的基于已存在背景下的向量与情感语料库学习任务的特定向量。Re(C&W)和 Re(Word2vec)代表使用不同的词向量作为基准向量。

(4) SE-Spred 和 SE-Srank: 分别为第 2 节中介绍的情感预测模型和情感排名模型。

根据上述的参数设置,表 3 列出了在不同情感数据集和不同邻近词个数 N_w 之下,不同词向量在单词级别的情感极性分类任务中的表现。其中,加粗的结果为对应情感词典数据集下表现最好的词向量。

表 3 不同词向量模型下的情感分类结果

Table 3 Emotional classification results under different word vector models

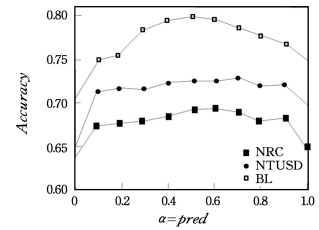
Embedding	$N_w=10$			$N_w=30$		
	BL	NRC	NTUSD	BL	NRC	NTUSD
C&W	65.5	58.8	60.1	64.4	58.1	59.4
Word2vec	70.8	63.0	65.0	68.8	61.3	63.2
Re(C&W)	67.1	59.6	61.8	65.9	58.6	60.9
Re(Word2vec)	71.2	63.4	65.7	69.6	61.8	64.1
SE-Spred	75.0	65.5	69.3	74.6	65.1	69.0
SE-SRank	73.6	66.2	68.4	73.8	65.9	68.3
SE-HyPred	79.7	70.1	74.4	77.4	67.9	72.1
SE-HyRank	78.3	69.4	69.4	77.9	68.9	72.6

(单位: %)

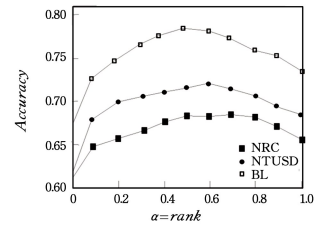
从表 3 可以看出,不同的邻近词个数 N_w 之间的实验结果几乎是一致的,在所有的词向量模型中,基于上下文的向量学习算法(C&W, Word2vec)的分类效果低于其他情感特定模型。其原因在于,它们将词的上下文编码为词向量,而忽略了文本的情感信息,使得模型不能够区分具有类似上下文但情感极性相反的词,如“good”和“bad”。将文本的情感重新融合到词向量之后,情感词的分类准确度明显提高,这显示了提高词向量在获取情感词之间的相似性对文本情感的重要性。

通过以上的实验可以看出,情感模型(SE-Spred, SE-SRank)和混合模型(SE-HyPred, SE-HyRank)的准确率胜过基准的方法。不同于基于上下文的模型,本文提出的情感词向量捕获了文本的情感信息,为情感词之间的相似性提供了依据。混合模型可以被视为 ReEmbed 的“融合”模型,从训练开始就将单词的上下文和整句的情感信息编码为词向量的表示形式。之所以产生了最佳性能,是因为两个混合模型不仅能够获取词语的上下文,而且能获取句子的情感信息。

接下来,通过设置 Accuracy 从 0 递增至 1(每次增加 0.1)来研究上下文部分和情感信息部分在混合模型中的相互影响。对 3 个情感词典 BL, NRC 和 NTUSD 进行试验,并设置 $N_w=10$, 实验结果如图 3 所示。



(a) SE-HyPred with different $\alpha = \text{pred}$



(b) SE-HyRank with different $\alpha = \text{rank}$

图 3 SE-HyPred 和 SE-HyRank 在不同的 α 变化曲线

Fig. 3 Comparison of SE-HyPred and SE-HyRank under different alpha curves

从图 3 可以看出:

(1) 当 $\alpha=0.1$ 时,模型的准确率有一个明显的性能提升,说明文本的情感信息被集成到基于上下文的模型中,表明情感信息在发现情感词之间的相似性是很重要的;

(2) 当 α 在 $[0.5, 0.6]$ 内时,两个混合模型都产生了最佳的性能,说明对平衡上下文和情感信息的效果达到最佳。

4.2 词级情感词向量实验

为了定量地评估本文提出的情感词向量学习模型,本实验比较在中英文数据下,不同的模型生成的词向量在词级别的情感分析中的效果。具体方法如下:从情感词典中选取部分词做训练集,训练分类模型,将 3 个情感词典(BL, NRC 和 NTUSD)进行二进制极性分类,预测其他词是表达积极还是消极的意义,衡量其准确性。其中,词的连续词向量可以作为其分类的特征。

然后,对每个数据集进行 N -Fold 交叉验证,将 $(N-1)$ 次作为训练数据,一部分作为测试数据,训练监督分类器 N 次,并把分类的平均精度作为最终的评估度量。表 4 列出了 $N=5$ 和 $N=10$ 的实验结果。

表 4 不同词向量模型下词级别的情感分类结果

Table 4 Emotional classification results at different word vector models

Embedding	(单位:%)					
	N-Fold=5			N-Fold=10		
	BL	NRC	NTUSD	BL	NRC	NTUSD
C&W	76.2	68.1	70.5	76.6	67.5	69.8
Word2vec	80.1	71.7	75.7	81.1	71.2	73.4
Re(C&W)	75.0	68.6	71.3	76.6	68.2	71.4
Re(Word2vec)	81.5	71.6	76.9	83.1	73.1	75.3
SE-SPred	80.8	73.3	78.2	80.4	72.4	76.0
SE-SRank	78.6	71.3	76.2	77.8	71.2	71.1
SE-HyPred	86.0	77.5	85.3	87.0	76.9	83.4
SE-HyRank	83.9	75.7	80.1	85.6	77.0	79.2

从表 4 可以看出:

(1)在不同的 N -Fold 下,分类的结果是近似一致的。基于上下文的词向量学习模型(C&W, Word2vec)分类效果较差,因为它们没有融合任何的情感信息。

(2)重新融合情感信息后,分类的准确性有了明显的提升。在 4 个情感向量学习算法中可以看到,混合模型(SE-HyPred 和 SE-HyRank)表现出比情感模型(SE-SPred 和 SE-SRank)更好的性能。这表明语境和情感信息都有助于发现情感词之间的相似性,并且词向量结合情感信息显著提高了准确率。其中,SE-HyPred 在此任务中表现最好。

(3)比较 SE-HyPred 和 SE-HyRank 可以发现,在大多数情况下 SE-HyPred 表现略好,这两种方法使用相同的信息,包括单词的语境和句子的情感。不同之处在于 SE-HyPred 以分类方式融合情感信息,而 SE-HyRank 以排名的方式融入情感信息。在这种二元情况下,认为 SE-HyPred 更适合作为情感标签(积极或消极)。这是因为在这些情感标签之间不存在明显的排序关系。而对于那些星级评论(例如亚马逊、淘宝和 Yelp)SE-HyRank 可能会更好。

结束语 为了弥补传统词向量在情感分析任务上的欠缺,本文提出了两种情感词向量学习模型,训练出含有情感和语义信息的词向量,并通过实验验证了情感词向量混合学习模型的有效性。实验表明,本文提出的情感词向量学习模型确实能够在大量的有标注的训练语料中捕获情感信息。

虽然基于上下文的混合情感词向量模型可以根据上下文较准确地获取到情感信息,但网络词语和较少出现的古代文体会对准确率造成一定影响。在今后的工作中,希望构建一个不需要分词的神经网络模型,处理上下文关联较弱的文本内容。

参考文献

[1] LAI S W. Word and Document Embeddings based on Neural Network Approaches[D]. Beijing: Institute of Automation Chinese Academy of Sciences, 2016.

[2] THOMAS K L, PETER W F, DARRELL L. An introduction to latent semantic analysis[J]. Discourse processes, 1988, 25(2/3): 259-284.

[3] MICHAEL N J, DOUGLAS J K M. Representing word meaning

and order information in a composite holographic lexicon[J]. Psychological review, 2007; 114(1): 301-311.

- [4] SOCHER R, PERELYGIN A, WU J. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank[C]// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013: 1631-1642.
- [5] YESSANALINA A, CARDIE C. Compositional matrix-space models for sentiment analysis[C]// Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. 2011: 172-182.
- [6] COLLOBERT R, WESTON J, BOTTOU L. Natural Language Processing (Almost) from Scratch[J]. Journal of Machine Learning Research, 2011: 2493-2537.
- [7] ZHENG X, CHEN H, XU T. Deep learning for Chinese word segmentation and POS tagging[C]// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013: 647-657.
- [8] TANG D, WEI F, YANG N, et al. Learning sentiment-specific word embedding for twitter sentiment classification[C]// Proceedings of the 2014/ 52th Annu. Meeting Assoc. Computer. Linguistics, 2014: 1555-1565.
- [9] BENGIO Y, DUCHARME R, VINCENT P. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003: 1137-1155.
- [10] LUO Y, LI L, TAN S, et al. Sentiment analysis on Chinese Micro-blog corpus[J]. Journal of Shandong University (Natural Science), 2014, 49(11): 1-7.
- [11] CHEN X D. Research on Sentiment Dictionary based Emotional Tendency Analysis of Chinese MicroBlog[D]. Wuhan: Huazhong University of Science & Technology, 2012.
- [12] TURIAN J, RATINOV L, BENGIO Y. Word representations: a simple and general method for semi-supervised learning[C]// ACL 2010, Proceedings of the Meeting of the Association for Computational Linguistics, Uppsala, Sweden. DBLP, 2010: 384-394.
- [13] PENNINGTON J, SOCHER R, MANNING C. Glove: Global Vectors for Word Representation[C]// Conference on Empirical Methods in Natural Language Processing. 2014: 1532-1543.
- [14] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient Estimation of Word Representations in Vector Space[J]. Computer Science, 2013: 566-569.
- [15] IRSOY O, CARDIE C. Opinion Mining with Deep Recurrent Neural Networks[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014: 720-728.
- [16] PILEHVAR M T, COLLIER N. Improved Semantic Representation for Domain-Specific Entities[C]// Proceedings of the 15th Workshop on Biomedical Natural Language Processing. 2016: 13-16.
- [17] PANG B, LEE L, VAITHYANATHAN S. Thumbs up Sentiment Classification using Machine Learning Techniques[C]// Proceedings of the 2002 Conference on Clinical Orthopaedics and Related Research. 2002: 79-86.
- [18] ANDREW L M, RAYMOND E D, PETER T P, et al. Learning word vectors for sentiment analysis[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Lin-

- guistics (ACL). 2011;142-150.
- [19] JEFFREY P, RICHARD S, CHRISTOPHER D M. GloVe: Global Vectors for Word Representation [C] // Proceedings of the 2014 Empirical Methods in Natural Language Processing. Washington, DC: IEEE Computer Society, 2014; 1201-1220.
- [20] BENGIO Y, COURVILLE A, VINCENT P. Representation Learning: A Review and New Perspectives [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35 (8) : 1798-1828.
- [21] MIKOLOV T, SUTSKEVER I, CHEN T, et al. Distributed representations of words and phrases and their compositionality [C] // Proceedings of the 2013/15th International Conference on Neural Information Processing Systems. 2013; 3111-3119.
- [22] BRIDLE J. Probabilistic interpretation of feed forward classification network outputs, with relationships to statistical pattern recognition in Neurocomputing [C] // Proceedings of the 1990 Neurocomputing: Algorithms, Architectures and Applications. Berlin: Springer, 1990; 227-236.
- [23] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch [J]. Journal of Machine Learning Research, 2011, 12; 2493-2537.
- [24] MNIH A, KAVUKCUOGLU K. Learning word embeddings efficiently with noise-contrastive estimation [J]. Advances in Neural Information Processing Systems, 2013; 2265-2273.
- [25] BARONI M, DINU G, KRUSZEWSKI G. Don't count, predict! A systematic comparison of Context-counting vs. Context-predicting semantic vectors [C] // Proceedings of the 2014/ 52th Annu. Meeting Assoc. Computer Linguistics. 2014; 238-247.
- [26] DUCHI J, HAZAN E, SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization [J]. Journal of Machine Learning Research, 2011, 12(2) : 2121-2159.
- [27] SOCHER R, PERELYGIN A, WU J, et al. Recursive deep models for semantic compositionality over a sentiment treebank [C] // Proceedings of the 2013 Conference Empirical Methods in Natural Language Processing. 2013; 1631-1642.
- [28] ZHANG W, SHI W X, LIU D N, et al. Improved Approach of Word Vector Learning Via Sentiment Information [J]. Application Research of Computers, 2016, 34(8) : 120-130.
- [29] RUMELHART D, HINTON G, WILLIAMS R. Learning representations by back-propagating errors [J]. Cognitive Modeling, 1988, 5; 1003-1120.
- [30] LABUTOV L, LIPSON H. Reembedding words [C] // Proceedings of the 2013 Meeting of Association for Computational Linguistics (ACL). 2013; 489-493.
- [31] COLLOBERT R, WESTON J. A unified architecture for natural language processing: Deep neural networks with multitask learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008; 160-167.
- [32] TANG D, QIN B, LIU T, et al. User modeling with neural network for review rating prediction [C] // Proceedings of the 2015/ 24th International Conference Analysis and Machine Intelligence. 2015; 1340-1346.



HUO Dan, born in 1990, M. S., lecturer. Her main research interests include natural language processing and so on.



ZHANG Sheng-jie, born in 1986, master, engineer. His main research interests include big data processing and content security.

(上接第 27 页)

- [16] LI X, ROTH D. Learning question classifiers [C] // Proceedings of the 19th international conference on Computational linguistics-Volume 1. Association for Computational Linguistics, 2002; 1-7.
- [17] HU M, LIU B. Mining and summarizing customer reviews [C] // Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2004; 168-177.
- [18] SOCHER R, PERELYGIN A, WU J, et al. Recursive deep models for semantic compositionality over a sentiment treebank [C] // Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013; 1631-1642.
- [19] KINGMA D P, BA J. Adam: A method for stochastic optimization [J]. arXiv:1412.6980, 2014.
- [20] YIN W, SCHÜTZE H. Multichannel variable-size convolution for sentence classification [J]. arXiv:1603.04513, 2016.
- [21] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [J]. arXiv:1406.1078, 2014.
- [22] ZHOU C, SUN C, LIU Z, et al. A C-LSTM Neural Network for Text Classification [J]. Computer Science, 2015, 1(4) : 39-44.
- [23] WANG C, JIANG F, YANG H. A hybrid framework for text modeling with convolutional RNN [C] // Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017; 2061-2069.
- [24] ZHOU P, QI Z, ZHENG S, et al. Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling [J]. arXiv:1611.06639, 2016.
- [25] ZHOU F, LI R Y. Convolutional Neural Network Model for Text Classification Based on BGRU Pooling [J]. Computer Science, 2018, 45(6) : 235-240.
- [26] ZHENG C, XUE M Y, HONG T T, et al. DC-BiGRU_CNN Model for Short text Classification [J]. Computer Science, 2019, 46(11) : 186-192.



WU Han-yu, born in 1996, M. S.. His main research interests include natural language processing, deep learning, etc.