文章编号：2095-2163(2022)07-0035-05 中图分类号：TP391.1 文献标志码：A

基于 K-BERT 的情感分析模型

王桂江，黄润才
（上海工程技术大学 电子电气工程学院，上海 201620）

摘 要：利用预训练模型对中文文本进行情感分析是目前的主流方式，K-BERT 模型的提出克服了 BERT 模型不具备背景知识的问题。本文通过在 K-BERT 的基础上引入双向长短时记忆网络，提出了 KB-BERT 情感分析优化模型。首先，通过 K-BERT 预训练模型，对输入的内容进行背景丰富，获取包含背景知识的语义特征向量；其次，利用长短时记忆网络提取上下文的相关特征，进行文本情感分析。实验结果表明，使用 KB-BERT 的准确率优于 K-BERT，在 Book_review 和 Weibo 两个数据集上的准确率，分别达到了 87.97%和 98.33%。

关键词：K-BERT；长短时记忆网络；情感分析

Sentiment analysis model based on K-BERT

WANG Guijiang, HUANG Runcai
(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

【Abstract】 Using pretrained model to perform sentiment analysis on Chinese text is the current mainstream way. The proposal of the K-BERT model overcomes the problem that the BERT model does not have background knowledge. By introducing a two-way LSTM network on the basis of K-BERT, the KB- BERT sentiment analysis optimization model is proposed. Through the K-BERT pretrained model, the input content is enriched in the background, and the feature vector containing the contextual semantic information is obtained. Then LSTM is used to extract the relevant features of the context, and finally the text sentiment analysis is performed. The experimental results show that the effect of using KB-BERT is better than that of K-BERT, and the accuracy on the two data 87.97% and 98.33%, respectively.

【Key words】 K-BERT; Long Short-Term Memory; sentiment analysis

0 引 言

大数据时代的社交媒体为用户提供了反馈和信息交流的平台，对于一件商品，不同的人有不同的看法，了解用户的看法和态度，是改进和优化的重要途径。情感分析是对用户观点的凝练，代表着用户的实际感受。

情感分析的发展经历了 3 个主要阶段，基于情感词典、基于机器学习和基于深度学习。情感词典作为最早的情感分析方式，通过将人们可能的观点评价构建一个字典，进行内容的匹配，以此来获得用户的情感倾向，这类的方法简单直接，不需要太复杂的方法就能获取到结果，但是情感词典的构建却需要大量的人力、物力和精力。而且随着社会的发展，基于情感词典已经无法跟上时代的变化。机器学习的出现，一定程度上解决了情感词典构建的问题，基于机器学习的方法根据文本提取的特征进行分类，即利用支持向量机(Support Vector Machines，SVM)等分类器，但是这类方法仍然需要人工标注的数据，分类器的结果也取决于数据的标记效果，泛化能力并不强。

随着深度学习的发展，自然语言处理进入了新的发展阶段。基于深度学习的情感分析有 3 个典型代表：

(1)利用神经网络训练词向量。利用神经网络训练得到词向量，之后将词向量的结果应用到下游任务中。比较典型的方法是利用 Word2Vec 训练词向量，将训练好的词向量送入循环神经网络(RNN)进行分析。

(2)利用循环神经网络 RNN。RNN 是处理时序问题的关键技术，基于 RNN 改进的长短时记忆网络(Long Short-Term Memory，LSTM)、门控循环单元(Gated Recurrent Unit，GRU)。陈帆[1]利用 LSTM 对微博情感进行分析，并用于微博特定主题的谣言识别；李辉等[2]利用 GRU 学习文本词语，并引入注意力机制实

作者简介：王桂江(1990-)，男，硕士研究生，主要研究方向：自然语言处理；黄润才(1966-)，男，博士，副教授，主要研究方向：机器学习、自然语言处理、计算机网络和大数据等。
通讯作者：黄润才 Email：hrc@ sues.edu.cn

收稿日期：2021-09-13

现了比 LSTM 有竞争力的效果。但是对于情感分析,获取上下文非常有必要,张俊飞等[3]用 BiLSTM 来获取上下文信息,将提取到的信息送入分类器,对评教评语进行情感分析。这类方法的效果依赖于特征提取的效果,而且激活函数的选择也关系到最终的分类效果。

(3)无监督学习,并充分考虑上下文信息。基于注意力机制的 Transformer 模型,使用编码和解码的机制,通过对注意力机制进行不同形式的构造,取得了比 RNN 更强的效果;基于 Transformer 的 BERT 模型,在传统的分类、问答和翻译等十多项任务中取得了历史最好的成绩,郝彦辉等[4]在 BERT 模型的基础上引入 BiLSTM,根据上下文判断情感倾向不明显的内容的真实情感倾向;李文亮等[5]在 BERT 的基础上融合多层注意力机制,在方面级情感分析上取得了不错的效果。

基于 BERT 预训练模型,升级和改造出了如 ALBERT、XLNET 等表现不俗的模型,基于 transformer-XL 的 XLNET 使用相对小的数据规模实现了接近 BERT 的效果;ALBERT 使用相对小的模型实现了与 BERT 接近的表现,甚至在部分场景下效果更好。尽管这一类的预训练模型在特征提取和词向量构建上表现出了较好的效果,但却存在无法理解语义背景的问题,比如:"基督山伯爵在巴黎的住处位于香榭丽舍大街,他很期待在这里遇见莫雷尔先生"是一句包含了地点、人物和社会关系的句子,而且带有开心的语气,如果不能理解其背景,只能感觉他在会见朋友。

综上所述,基于前人的研究成果和优化策略,为更好的获取文本信息的语义特征,增强对于语义信息的理解,提升模型对文本的情感分析能力,本文提出了结合 K-BERT 和 BILSTM 的情感分析模型,使用带有知识图谱的 K-BERT 代替 BERT,丰富句子的背景信息,有利于组合句子内容,提高特征提取能力;在 K-BERT 基础上引入 BILSTM,进一步增强对于上下文之间的语义提取;模型在 NLP 中文文本任务情感分类数据集上表现出了有竞争力的效果。

1 K-BERT 语言模型概述

K-BERT 是融合知识图谱的语言训练模型,该模型在开放域的 8 个中文 NLP 任务上超过了 Google BERT,模型由 Knowledge layer、Embedding layer、Seeing layer、Mask-Transformer Encoder 组成。

1.1 Knowledge layer

Knowledge layer 的作用是将知识图谱关联到句子中,形成一个包含背景知识的句子树(Sentence Tree)。知识嵌入句子的过程可以分为知识图谱查询(K-Query)和知识谱图嵌入(K-Inject)。K-Query 从知识图谱中查询句子所涉及到的命名实体,K-Inject 将查询到的命名实体相关的三元组嵌入到句子中合适的位置上,形成句子树。假设给定句子 $s = \{w_0, w_1, \cdots, w_n\}$ 和知识图谱(knowledge graph,KG),知识层输出的句子树结构为 $t = \{w_0, w_1, \cdots, w_i\{(r_{i0}, w_{i0}), \cdots, (r_{ik}, w_{ik})\}, \cdots, w_n\}$,句子树形状如图 1 所示。

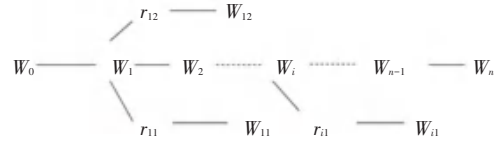

图 1 句子树结构

Fig. 1 Sentence tree structure

其中 w_i 是从知识谱图中查询到的命名实体;r_{ij} 是与 w_i 相关的第 j 个分支;w_{ij} 是与 w_i 相关的第 j 个分支对应的值。

1.2 Embedding layer

Embedding layer 层包含 token embedding、soft-position embedding 和 segment embedding,其作用是将句子树转换成为序列,同时要保留句子树的结构信息。

Token-embedding 主要用于实现句子树的序列化,将句子中的每个 token 映射成为一个 H 维度的向量表示,并在每个句子的开头添加一个[CLS]标记;

Soft-position embedding 在 BERT 中,所有句子的输入信息都对应一个位置信息,在 K-BERT 中,将句子树的内容平铺以后,当分支中的 token 插入到对应的主干节点之后,主干节点后续的 token 会发生移动,导致原有的位置信息发生变化,软位置(Soft position)通过对句子树的位置进行二次编码,将其原有的顺序信息进行恢复,理顺了句子的结构。

Segment embedding 该层用以区分一个句子对中的两个句子,当包含多个句子时,第一个句子中的各个 token 被赋值为 A,第二个句子中的各个 token 被赋值为 B,当只有一个句子时,segment embedding 为 A。

1.3 Seeing layer

Seeing layer 层的作用是通过一个可视化矩阵来限制词与词之间的关系,解决句子树软位置编码后的一对多现象。对于一个可视化矩阵 $\boldsymbol{M}$,相互可见的取值为 0,互不可见的取值为 $-\infty$,$\boldsymbol{M}$ 定义如式(1):

$$\boldsymbol{M}_{ij} = \begin{cases} 0, w_i, w_j \text{ 相互可见} \\ -¥, w_i, w_j \text{ 相互不可见} \end{cases} \quad (1)$$

1.4 Mask-Transformer

Mask-Transformer 的核心思想是让一个词的嵌

入只来源于其同一个枝干的上下文，而不同枝干的词之间相互不受影响，可视化矩阵 $\boldsymbol{M}$ 解决了句子树位置不同但编码相同的问题，通过在 softmax 函数中添加可见矩阵 $\boldsymbol{M}$，控制注意力的影响系数。Mask-Transformer由 12 层 mask-self-attention 堆叠，mask-self-attention 的定义如式(2)~式(4)：

$$Q^{i+1},K^{i+1},V^{i+1} = h^iW_q,h^iW_k,h^iW_v \tag{2}$$

$$S^{i+1} = \mathrm{softmax}\left(\frac{Q^{i+1}K^{i+1} + M}{\sqrt{d_k}}\right) \tag{3}$$

$$h^{i+1} = S^{i+1}V^{i+1} \tag{4}$$

其中：W_q,W_k,W_v 是需要学习的模型参数；h^i 是隐状态的第 i 个 mask-self-attention 块；d_k 是缩放因子；$\boldsymbol{M}$ 为可见矩阵。

Embedding layer、seeing layer、句子树和可见矩阵是 K-BERT 的处理的关键技术，四者之间的关系如图 2 所示。从 knowledge layer 得到句子树后，对句子树同时构建可视化矩阵和送入 embedding layer 编码，这两个过程得到的信息归并后输入到 mask-selft-attention 中进行计算。

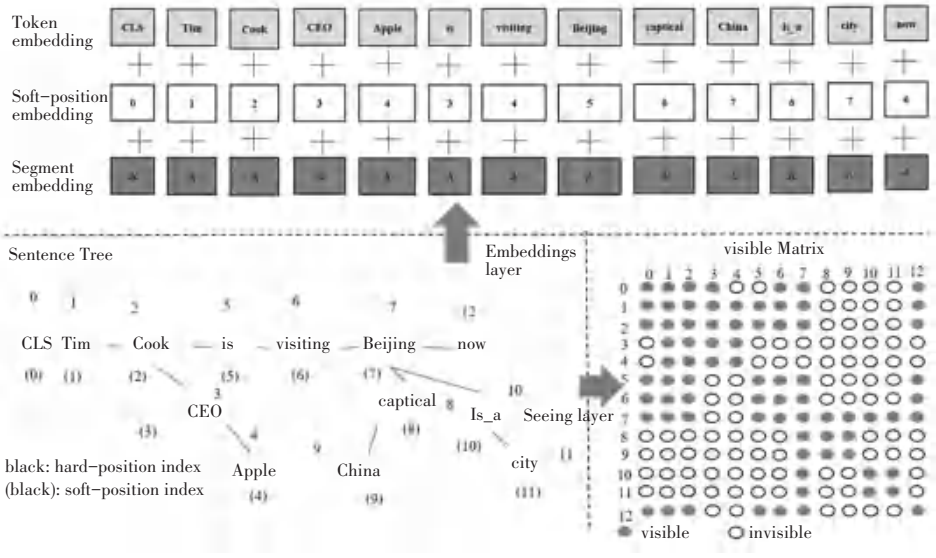

图 2　K-BERT 处理的关键技术

Fig. 2　Key technology of K-BERT processing

2　一种改进 K-BERT 的情感分析模型

本文在 K-BERT 的基础上，通过引入双向 LSTM，增强模型对于上下文的语义关联能力，使模型既有丰富的背景知识，又能很好的关联上下文，获取更多的语义信息，从而实现情感分类效果的提升。本文的模型如图 3 所示，称其为 KB-BERT。

K-BERT 层：K-BERT 是一种基于知识图谱的语言表示模型，在原有 BERT 模型的基础上引入了知识图谱的表示方式，输入的文本经过 K-BERT 后包含了原来文本没有的背景知识，输出包含丰富背景信息的词向量；

BILSTM 层：双向 LSTM 层的目的是学习文本所含特征，K-BERT 层计算输出的词向量在 LSTM 层进行再次学习，获取句子的上下文信息，对语义信息进一步增强；

Softmax 层：经过双向 LSTM 提取到的特征信息被输入到 softmax 层中进行分类，将情感分为正面和负面。

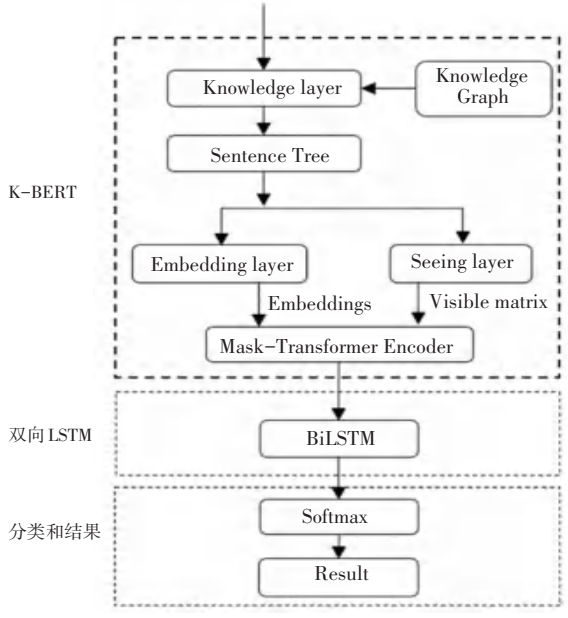

图 3　KB-BERT 情感分析模型

Fig. 3　KB-BERT sentiment analysis model

2.1 BiLSTM 层

LSTM 是一种基于 RNN 的网络结构,LSTM 由输入门、遗忘门、记忆单元和输出门 4 部分组成,LSTM 结构如图 4 所示。

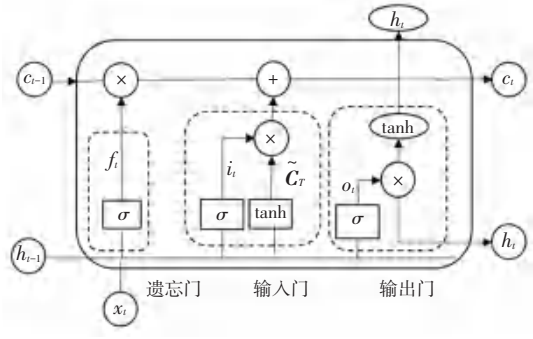

图 4 LSTM 网络结构

Fig. 4 LSTM network structure

其中,h_{t-1} 为上一个单元输出;h_t 为当前单元输出;x_t 为当前输入;σ 为 *sigmoid* 函数;f_t 为遗忘门输出;i_t 与 $\widetilde{\boldsymbol{C}}_t$ 的乘积为输入门输出;o_t 为输出门输出。

使用记忆单元 $\widetilde{\boldsymbol{C}}_t$ 解决长距离依赖和梯度爆炸的问题,使用 C_t 避免梯度消失的问题。在 LSTM 中,将文本看成一个文本序列,上一个过程处理的结果经过输入门进行输入,通过遗忘门决定哪些信息需要丢弃,遗忘门的取值范围介于 0~1 之间,随着信息的输入量大小而变化,并将处理后的信息融入到 $\boldsymbol{C}$ 中;输入门中经过激活函数处理后的 i_t 和记忆单元计算后的数据传递到 $\boldsymbol{C}$ 中,实现对信息的更新,使得向下传递的信息进一步增多;输出门对遗忘门和输入门更新后的信息做一次激活,再将激活后的信息与 o_t 进行矩阵运算,得到当前单元隐藏层的输出。LSTM 解决了学习能力弱化的问题,避免了预测信息与相关信息距离过大而导致的信息丢失,使用双向 LSTM 解决了上下文的语义关联问题。门控单元的计算公式分别如式(5)~式(10):

遗忘门:

$$f_t = \sigma(\boldsymbol{W}_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$

输入门:

$$i_t = \sigma(\boldsymbol{W}_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\tilde{C}_t = \tanh(\boldsymbol{W}_c \cdot [h_{t-1}, x_t] + b_c) \tag{7}$$

$$C_t = f_t C_{t-1} + i_t \tilde{\boldsymbol{C}}_t \tag{8}$$

输出门:

$$O_t = \sigma(\boldsymbol{W}_o \cdot [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = O_t \tanh(C_t) \tag{10}$$

其中,$\boldsymbol{W}_f$ 和 b_f 分别为遗忘门的权重矩阵和偏置;$\boldsymbol{W}_i$ 和 b_i 分别为输入门的权重矩阵和偏置;$\widetilde{\boldsymbol{C}}_t$ 为候选向量;$\boldsymbol{W}_c$ 和 b_c 分别为输出门的权重矩阵和偏置;$\boldsymbol{W}_o$ 和 b_o 分别是计算单元的权重矩阵和偏置。

2.2 Softmax 分类层

经过处理后的信息使用 softmax 层进行情感分类。softmax 为每个输出分类的结果均赋值一个概率,表示每个类别的可能性,式(11):

$$\mathrm{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{c=1}^{C} e^{z_c}} \tag{11}$$

其中,z_i 为第 i 个结点的输出值;C 为输出结点的个数;s_{z_t} 是当前元素与所有元素的比值,即当前元素 i 的概率。

3 实验

3.1 实验环境及数据集

本实验环境:处理器:E3-1281-V3 3.7 GHz 八核;内存:16 GB 1 600 MHz DDR3;GPU:华硕 1070Ti 8G;系统环境:Ubuntu 18.04 LTS;编程语言:python3.7,pycharm 开发环境,深度学习库为 Pytorch。

本文使用 Book_review 和 Weibo 两个情感数据集,正面情绪标签为 1,负面情绪标签为 0。Book_review 从豆瓣获取,包含正负情绪各 20 000 条;Weibo 从新浪微博获取,包含正负情绪各60 000条。

3.2 评价指标

为了验证模型的有效性,采用准确率(*Accuracy*)对测试集和验证集进行分别验证,准确率的计算公式(12):

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{12}$$

其中,T_p 表示正面评价样本中被预测为正面的样本总数;T_n 表示负面评价样本中被预测为负面的样本总数;F_p 表示负面评价样本中被预测为正面的样本总数;F_n 表示正面评价样本中被预测为负面的样本总数。

3.3 对比实验设置

为证明本文方法的有效性,取以下对比方法进行验证:

(1) Google BERT。首先将输入的文本进行词向量编码,对于获取到的词向量进行信息提取,之后运用分类器进行结果分类。

(2) K-BERT。首先对输入的句子进行命名实体识别,之后对识别到的命名实体从知识图谱中查

询关联词，将查询到的关联词插入到句子中形成包含背景知识的句子树，对输入的句子树编码，得到信息丰富的词向量，将得到的词向量直接送入分类器进行结果分类。

(3)KB-BERT。首先使用 K-BERT 获取信息丰富的词向量，将得到的词向量送入 LSTM 循环网络二次特征提取，丰富上下文提取，最后将得到的词向量送入分类器进行结果分类。

3.4　实验参数

实验一共训练 5 个 epoch，每次的信息输入量 batch_size 为 8，使用 dropout 防止过拟合，dropout 的值设置为 0.5，使用 12 层 mask-self-attention，学习率设置为0.000 02。

3.5　结果分析

在本地实验条件下，KB-BERT、K-BERT 和 Google BERT 在 Book_review 和 Weibo 数据集上的表现如图 5 和图 6 所示，其中图 5 是验证集上的效果，图 6 是测试集上的效果。

图 5　不同模型在验证集上的准确率(%)

Fig. 5　Accuracy of different models on the validation set (%)

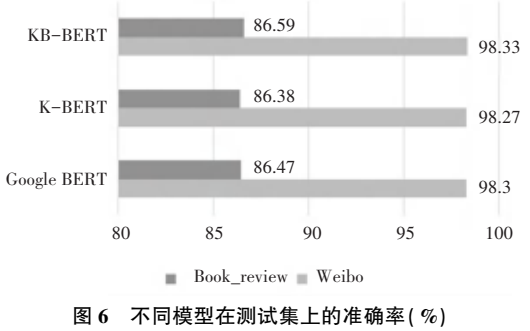

图 6　不同模型在测试集上的准确率(%)

Fig. 6　Accuracy of different models on the test set(%)

在 Book_review 数据集上，KB-BERT 的效果最好。在验证集上较 K-BERT 提升 0.6%，较 BERT 提升 0.9%；在测试集上，较 K-BERT 提升 0.2%，较 BERT提升0.1%。这说明，在数据内容为长文本的情况下，引入 LSTM 有助于对上下文信息的获取，本文优化后的模型在 Book_review 数据集上的表现最佳，准确率在验证集上达到 87.97%。

在 Weibo 数据集上，BERT、K-BERT 和 KB-BERT 表现近乎一致。这说明，在文本内容较为稀疏无规则的情况下，引入知识图谱不能很好的得到命名实体，但是简短的稀疏文本在使用 LSTM 后，对于上下文的语义获取有一定的提升，说明在简短稀疏的文本内容中，LSTM 网络对于增强语义获取仍旧发挥效果。尽管三者的区别不大，但本文所用方法在 Weibo 数据集上仍然取得最佳的效果，准确率在验证集上达到 98.28%。

引入双向 LSTM 的 KB-BERT 模型，由 Book_review 和 Weibo 数据集上的表现说明，对于增强上下文语义理解，提升准确率均有效果。在涉及专业知识或背景知识的情况下，对于长文本的分析结果表现更佳，对于短文本和稀疏文本，所提模型仍然有效。

4　结束语

本文给出了一种具有知识图谱背景的情感分析模型 KB-BERT。首先，通过 K-BERT 对输入的内容进行处理，丰富其知识背景，增强语义的理解能力；其次，引入双向 LSTM 网络，进一步增强对于语义的上下文内容理解。实验结果表明，改进后的 KB-BERT 在涉及背景信息的长文本数据上表现更好，在 Book_reivew 和 Weibo 两个中文数据集上，准确率分别达到 87.97%和 98.33%，证明了本文方法的有效性。

参考文献

[1] 陈帆. 基于 LSTM 情感分析模型的微博谣言识别方法研究[D]. 武汉：华中师范大学，2018.

[2] 李辉，郑媛媛，任鹏举. 基于 CNNGRU-Attention 模型的文本情感分析[J]. 制造业自动化，2019，41(9)：19-23，72.

[3] 张俊飞，毕志升，吴小玲. 基于词向量 Doc2vec 的双向 LSTM 情感分析[J]. 计算机与数字工程，2018，46(12)：2385-2389.

[4] 郝彦辉，王曦，陈铎. 基于 BERT-BiLSTM 模型的舆情监测方法及实证研究——以研究生招生考试为例[J]. 情报科学，2021，39(8)：78.

[5] 李文亮，杨秋翔，秦权. 多特征混合模型文本情感分析方法[J/OL]. 计算机工程与应用：1-12[2021-08-15]. http://kns.cnki.net/kcms/detail/11.2127.TP.20210621.1806.004.html.