

文章编号: 2095-2163(2020)02-0007-05

中图分类号: TP391.1

文献标志码: A

基于 BERT-CNN 的电商评论情感分析

史振杰, 董兆伟, 庞超逸, 张百灵, 孙立辉

(河北经贸大学 信息技术学院, 石家庄 050061)

摘要:在对电商评论进行情感分析中,为了使提取的情感特征能够更多地捕获句子中的情感信息,提出了一种基于预训练的 Bidirectional Encoder Representations from Transformers(BERT)网络与卷积神经网络(CNN)相结合的 BERT-CNN 网络模型。首先利用 BERT 结构表达句子语义作为文本向量,然后通过卷积神经网络抽取句子的局部特征,通过在有标签的京东某手机评论数据集上的实验,表明该方法在该领域具有良好的性能。

关键词: BERT; 卷积神经网络; 情感分析

Sentiment analysis of e-commerce reviews based on BERT-CNN

SHI Zhenjie, DONG Zhaowei, PANG Chaoyi, ZHANG Bailing, SUN Lihui

(School of Information Technology, Hebei University of Economics and Business, Shijiazhuang 050061, China)

[Abstract] In the sentiment analysis of e-commerce reviews, in order to make the extracted sentiment features capture more sentiment information in sentences, a pre-trained Bidirectional Encoder is proposed. BERT-CNN network model combines Representations from Transformers (BERT) network and convolutional neural network (CNN). First, use BERT structure to express sentence semantics as a text vector, and then extract the local features of the sentence through a convolutional neural network. Experiments on a labeled Jingdong mobile phone review dataset show that the method has good performance in this field.

[Key words] BERT; CNN; sentiment analysis

0 引言

随着信息技术的快速发展,网络已经融入到人们生活的方方面面。网络上具有平等、自由、共享、虚拟等特点,近年来,互联网上出现越来越多的网络应用,例如微博,论坛,电子商务等等,其中,电子商务指的是买卖双方在不相见的互联网平台进行交易的商贸活动。电商应用平台的飞速发展改变了社会的生活方式,给人们的生活带来了极大的便利。同时,各电商平台之间的竞争也尤为激烈,为了增强竞争力,各平台不仅保证商品的质量和价格合理外,还要了解消费者对商品和服务的反馈,以便更好地制定相应的营销策略。如今,很多电商平台都涉及了评论系统,消费者在系统中可以自由发表自己的观点与看法^[1],表达自己的情感,在这些评价、观点与情感中,包含着许多有价值的信息,这些信息反映了消费者对于商品的主观感受,对于其他消费者挑选符合心意的商品具有极大的参考价值^[2],同时,也是商家改进自身营销模式的重要依据。

近年来,随着深度学习的发展以及自然语言领域的使用,越来越多的学者采用深度学习技术来处理海量的文本信息。深度学习可以经过多层次的学

习,自动学习到数据的本质特征,能够在预测与分类任务中获得较高的准确率。例如常用的网络有:长短期记忆网络^[3]、卷积神经网络^[4]、循环神经网络^[5]等。

Mikolov 等人^[6]提出一种使用 Skip-gram 和 CBOW 的方式学习词的分布式表示,通过这样的方式可以在一定程度上表示出词语之间的相关性,得到词语在更高维空间的映射。为了弥补 RNN 网络自身在梯度传递上的缺陷,Hochreiter 等人^[7]提出了基于循环神经网络的优化形式,长短时记忆网络 LSTM,在一定程度上解决了梯度爆炸或者梯度消失的问题。Denil 等人^[8]提出一种多层次的 CNN 模型来提取文本中的特征,通过多层次的卷积与池化结构,加强对句子中关键的局部特征的获取,以达到更好的效果。Kalchbrenner 等人^[9]将静态卷积网与动态卷积网相结合,对于解决不同文本的不同长度具有较好的效果。Zhou 等人^[10]提出了将 CNN 与 RNN 的特点相结合,先对文本用 CNN 进行分布特征的提取,再用 RNN 进行序列特征的提取,最后用于分类。Cao 等人^[11]改变了两种模型顺序,由此提取出来的特征用于文本分析。Yogatama 等人^[12]采

作者简介: 史振杰(1994-),男,硕士研究生,主要研究方向:计算机应用技术;董兆伟(1967-),男,博士,教授,主要研究方向:机械学与职业教育学、计算机应用。

收稿日期: 2019-12-10

用生成对抗模型,其中一部分模型用于生成数据,另一部分通过预测数据来进行分析。

李然^[13]在进行本文情感分析时采用了神经网络语言模型,通过自适应的学习短文本向量,提取更深层次的语义信息,在大规模的商品评论数据集上表明了深度学习的分类性能更好。胡朝举等人^[14]提出了一种将 CNN 与 GRU 相结合的方法,不仅能采用 CNN 捕捉句子的局部特征,而且能够较好地获取句子的上下文信息,实验获得了较高的 F 值。陈葛恒^[15]针对 GRU 网络只能前向学习而不能预测到后面的信息问题,提出了双向的 GRU 来进行句子前后信息的学习,有效地解决了上下文之间的关联性。冯兴杰等人^[16]使用注意力模型与 CNN 相结合的方法,能够减少人为对于特征的构造,在相关数据集上结果表明,与传统的机器学习方法和卷积神经网络相比有着明显的提升。马思丹等人^[17]根据词向量的特点,提出了一种加权 Word2Vec 的文本分类方法,通过设置文本词语相似度阈值,分为加权的部分与不加权的部分,在此基础上进行分类,实验表明该方法比传统的 TF-IDF 效果要好。

在以上的研究中,都使用了基于深度学习的方法进行情感分析,这种方法虽然在一定程度上表现优异,但是要进行长时间的训练,而且需要大量的语料用于学习,语料不足时很难达到满意的效果。

综上,针对现有的文本情感分析方法特征提取不充分,难以表达句子的复杂语义,不能关注上下文信息等问题,提出了一种基于 BERT-CNN 的网络结构,通过 BERT 结构进行语义的向量化表达,运用卷积网络结构来进一步提取局部特征,最后使用 Softmax 分类器进行文本的情感分类,在某手机的评论数据集中,相比于其他模型,准确率有一定的提升。

本文的主要工作如下:

(1) 利用 BERT 对评论信息的句子进行向量化表达,充分考虑了句子中每一个词语对其他上下文中词语的影响,以及同一个词语在不同语境中的不同含义表达。

(2) 先利用 BERT 网络结构来处理文本中上下文特征的提取,再对已经提取的特征使用卷积神经网络 CNN 进行局部语义特征提取,既能同时利用 BERT 与 CNN 特征提取的优势,又能很好地解释要处理文本的语义,从而提高文本情感分析的准确率。

1 BERT 模型

BERT 的出现在自然处理领域带来了很大的提升,之前的模型是从左向右或者是将从左向右和从右

向左的训练结合起来,而 BERT 使用了多层 Transformer 模型^[18],实现了将句子中的每个词的信息都涵盖进词向量中去,实验的结果表明,双向训练的语言模型对语境的理解会比单向的语言模型更深刻,其中文本分类中,用到了 Transformer 中的 Encoder。

BERT 中的 Transformer 由 6 个 Encoder-Decoder 叠加组成,在结构上是相同的,但是彼此间却不共享权重^[19]。注意力机制运算如图 1 所示。图 1 即为一层编码器 (Encoder) 和对应的一层解码器 (Decoder),在 Encoder 中,输入 (input) 经过 Embedding 后,要做位置嵌入 (positional Encoding),然后是多头注意力机制 (multi-head Attention),再经过全连接层^[20],每个子层之间都有残差连接。

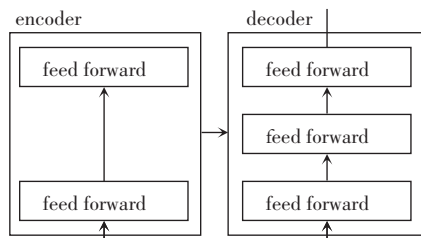


图 1 注意力机制计算图

Fig. 1 Attention mechanism calculation diagram

多头注意力机制就是将一个词的向量切分成 h 个维度,求 Attention 的相似度时每个 h 维度计算,由于每个单词在高维空间表示唯一的一个向量,每一维空间都可以学到不同的特征,相邻空间所学结果更相似,相较于全体空间放到一起对应更加合理。比如对于 $vector - size = 512$ 的词向量,取 $h = 8$,每 64 个空间做一个 Attention,学到的结果更加准确。

自注意力机制中,每个词都可以无视方向与距离,有机会直接和句子中的每个词进行编码。权重的大小代表了两者之间联系的深刻度,一般意义来说,模糊词所连的边都比较深。

位置嵌入能够表示句子的序列信息的顺序,对于模型学习句子的含义有重要的影响。Transformer 计算 token 的位置信息使用了正弦波,类似模拟信号的周期性变化,这样的循环函数在一定程度上能够增加模型的泛化能力。运算公式如下所示:

$$PE_{2i}(p) = \sin(pos/100\ 000^{2i/dpos}),$$

$$PE_{2i+1}(p) = \cos(pos/100\ 000^{2i/dpos}), \quad (1)$$

将 id 为 p 的位置映射为一个 d 维的位置向量,这个向量的第 i 个元素的数值就是 $E_i(p)$ 。

但 BERT 直接训练一个 position Embedding 来保留位置信息,每个位置随机初始化一个向量,加入模型进行训练,能够得到一个包含位置信息的

Embedding,最后这个 Position Embedding 与 Word Embedding 进行直接拼接。

此外,BERT 使用 masked language model 做到了真正意义上的双向编码,随机屏蔽预料中 15% 的 token,然后将被屏蔽的 token 位置输出的最终隐层向量输送到分类器,预测被屏蔽的 token,其中类似于完形填空,虽然能够看到所有的位置信息,但需要预测的词已经被特殊符号所代替,可以直接进行编码。但是确定屏蔽掉的单词后,并没有直接去掉,而是 80% 的会被直接替换,10% 会被替换为任意单词,10% 会保留原始 token。这是为了增强模型的鲁棒性,避免出现模型不认识的单词,增强其泛化能力。

2 CNN 模型

在传统的神经网络中,把每个神经元都连接到下一层的每个神经元上,这就是全连接,在 CNN 中,对输入层进行卷积得到输出,这就不是全部连接而成为了局部连接,即输入的局部区域连接到一个神经元上,每一层都用不同的卷积核,再将其组合起来。池化层是卷积神经网络中的一层重要结构,在卷积层之后应用,池化层对其输入进行下采样,最常用的方法就是保留最大信息,一般是经过窗口化的最大池化。

卷积的作用可以认为是发现一种特征,而池化的作用是减少输出维度同时保留显著的信息^[21]。在卷积神经网络中,卷积层中的卷积核类似于一个滑动的窗口,如图 2 所示,在整个输入图像中进行特定步长的滑动,经过卷积运算后,能够得到输入图像的特征图,这个特征图就是卷积层提取出来的局部特征,这个卷积核是参数共享。在整个网络的训练过程中,包含权值的卷积核也会随之更新,直到训练完成。

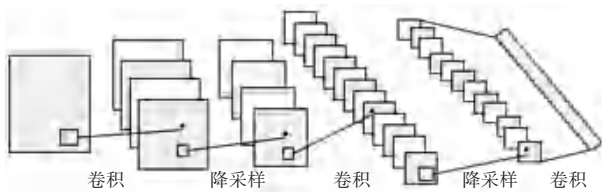


图 2 卷积神经网络结构图

Fig. 2 Convolutional neural network structure

全连接层的作用主要是对特征进行整合,池化层的输出以全连接的形式传递给全连接层,通过分类器得到分类,再将预测的结果与实际的结果进行比较,通过反向传播的方式更新网络参数。

在自然语言处理中,多数任务的输入不再是图片,而是以矩阵表示的句子或文档。矩阵的每一行对应一个 token(象征),一般是一个单词或字符,也即每

一行代表一个词向量。在图像中,卷积核划过的是图像中的一块区域,在自然语言领域一般用卷积核划过矩阵的一行,即单词。然后卷积核的宽度就是矩阵的宽度,而高度不是固定的,需要进行设置。

3 BERT-CNN 模型建立

整个 BERT-CNN 神经网络模型主要分为 BERT 层、CNN 层和情感分类输出层,整体框架如图 3 所示。

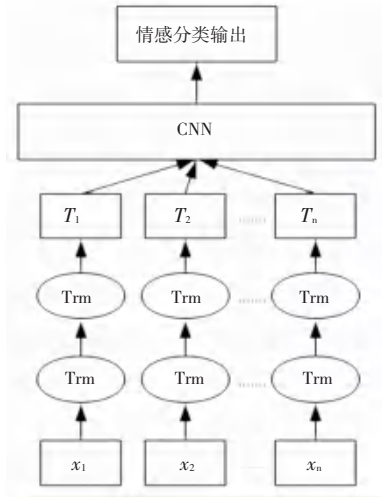


图 3 BERT-CNN 模型结构图

Fig. 3 BERT-CNN model structure

BERT-CNN 神经网络模型的情感分类输出层基本相似,但是词向量输入层有差别,BERT-CNN 神经网络模型采用预训练语言模型 BERT 作为文本信息的表示,而传统神经网络语言模型采用 Word2Vec 作为文本信息的表示。此外,BERT-CNN 将 BERT 输出的词向量经过卷积神经网络做进一步特征提取,增强了模型的健壮性。CNN 层分为卷积与池化两部分:

对 BERT 层的输出矩阵 $B = \{H_1, H_2, \dots, H_n\}$ 进行卷积操作,假设卷积核长度为 m ,即每次对 m 个分词向量进行卷积操作,卷积核滑动的步长一般设为 1,对文本矩阵进行上下滑动,则 B 可以分成 $\{H_{1:m}, H_{2:m+1}, \dots, H_{n-m+1:n}\}$,其中, $H_{i:j}$ 表示向量 H_i 到 H_j 的所有向量的级联,对于每一个分量执行卷积操作后得到向量 $C = \{c_1, c_2, \dots, c_{n-m+1}\}$,而 c_i 是对分量 $H_{i:i+m-1}$ 执行卷积操作后得到的值,称为一个局部特征映射,计算公式如下所示:

$$c_i = W^T H_{i:i+m-1} + b, \tag{2}$$

其中, W 是卷积核的参数,按照均匀分布随机初始化,并在模型训练过程中不断学习, b 是偏置变量。

接着对卷积捕获的文本特征映射向量 $C = \{c_1, c_2, \dots, c_{n-m+1}\}$ 进行池化操作,研究中采用了最大池

化方式,公式可表示为:

$$\hat{c}_m = \max\{C\}, \quad (3)$$

上述为 1 个卷积核经过卷积、池化操作得到的结果,对于 q 个卷积核得到的结果如下:

$$\hat{C}_m = (c_{m1}, c_{m2}, \dots, c_{mq}). \quad (4)$$

为了提取更好的特征,研究采用 3 种不同尺寸的卷积核去提取特征,经过池化后,再将不同尺寸的卷积核提取的特征拼接在一起作为输入到分类器中的特征向量。如图 4 所示。

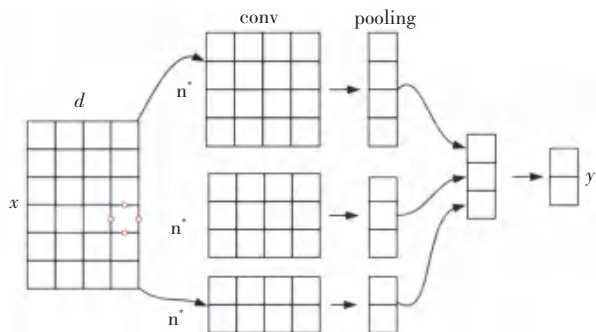


图 4 CNN 层结构图

Fig. 4 CNN layer structure

卷积操作实质上完成了对文本特征中表示局部重要信息的捕获,池化操作则完成了局部重要特征的提取。此后 CNN 层的输出向量经过拼接得到最终的向量,输入到情感分类器中进行分类,得到模型对于每条输入数据对应的情感类别。

在前述流程后,研究中将通过平均池化得到的句向量利用全连接层来获取抽象特征。同时,该网络层使用 ReLU 激活函数,为了增强模型的泛化能力,再增加一层全连接层进行输出,由于本次分类是多标签分类任务,则需要经过分类器进行分类。

在模型中,BERT 作为评论文本的编码器,使用 BERT 语言模型的嵌入功能,将每一条评论编码到各词向量堆叠而成的句子。作为新的特征用作 CNN 层的输入,为了防止过拟合,在全连接层前面加入一层丢弃率为 0.5 的 dropout 层。

4 结果与分析

本文模拟实验环境配置见表 1。

本节在京东某手机评论数据集上对比了混合模型 BERT-CNN 与其他几种模型,不仅进行了模型准确性、稳定性的对比,同时对比了模型训练耗时程度。

采用几种不同的模型进行实验,研究得到的各指标的实验结果见表 2,时间对比结果见表 3。

表 1 实验环境配置表

Tab. 1 Experimental environment configuration table

实验环境	具体配置
操作系统	Windows 10(64 位)
CPU	Inter(R) Core(TM) i9900K
内存	64 G
硬盘	4 T
编程语言	Python3.5
词向量训练工具	Word2Vec
分词工具	Jieba

表 2 实验结果表

Tab. 2 Table of experimental results

模型	acc	precision	recall	f-beta
TextCNN	0.850 3	0.850 1	0.860 4	0.850 7
BiGRU-Attention	0.887 2	0.895 3	0.864 5	0.889 1
BERT-CNN	0.957 2	0.952 1	0.962 4	0.953 2
BERT	0.924 2	0.913 2	0.894 5	0.920 2
LSTM	0.845 2	0.830 4	0.828 5	0.837 6

表 3 时间表

Tab. 3 Time table

模型	时间
TextCNN	0.50
BiGRU-Attention	0.57
BERT-CNN	5.30
BERT	5.10
LSTM	1.10

比较表 2 中 BERT-CNN 模型和其他几种常用模型,例如 LSTM, BiGRU-Attention。可以看出,相比于这几种常用模型,混合模型 BERT-CNN 在准确率与 F 值上都有了明显的提升,这说明了 BERT-CNN 在该京东手机评论数据集上,情感分类效果比较好,同时,对比单独地使用 BERT 或者 TEXTCNN 进行文本分类,也得到了显著提升。

对比表 3 中 BERT-CNN 模型和其他模型所用的时间可以看出,除了 BERT 模型以外,其它模型虽然准确率与 F 值相对较低,但是所用时间相比混合模型 BERT-CNN 明显少很多,这说明了 BERT 预训练词向量模型,需要耗费大量的时间进行句子中词向量的表示。BERT-CNN 效果较好,但时间花费也很大。

5 结束语

近年来,电子商务发展渐趋火爆,电商评论情感分析已然成为一个研究热点,为了更好地进行研究,本文提出了结合 BERT 的词向量,与卷积神经网络相结合的 BERT-CNN 网络模型用于电商评论研究

和分析中。该模型先利用BERT进行词向量编码表示文本的语义信息,然后在此基础上使用卷积神经网络对文本特征进行更加深入提取,实现模型对于文本信息语义的深层理解,完成对电商评论的情感分析。使用某手机的评论数据集进行仿真实验,实验结果表明,BERT-CNN模型较大程度地提升了情感分类的性能。

BERT模型使用过程中,参数量较大,训练及加载都会耗费大量时间,所以,研究对BERT模型进行压缩,在模型精度不会受到较大损失条件下降低模型的复杂度这也是后续一项重要的研究工作。

参考文献

- [1] 李丹阳. 面向中文评论的情感分析方法研究[D]. 西安:西安工业大学,2019.
- [2] 潘莹. 基于电商评论的情感分析方法研究[D]. 荆州:长江大学,2019.
- [3] 胡新辰. 基于LSTM的语义关系分类研究[D]. 哈尔滨:哈尔滨工业大学,2015.
- [4] KALCHBRENNER N, GREFFENSTETTE E, BLUNSOM P. A Convolutional neural network for modelling sentences[J]. arXiv preprint arXiv:1404.2188, 2014.
- [5] LAI Siwei, XU Liheng, LIU Kang, et al. Recurrent convolutional neural networks for text classification[C]//Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin, Texas, USA; AAAI, 2015;2267.
- [6] MIKOLOV T, SUTSKEVER I, CHEN Kai, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26: 3111.
- [7] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997,9(8):1735.
- [8] DENIL M, DEMIRAJ A, KALCHBRENNER N, et al. Modelling, visualising and summarising documents with a single convolutional neural network[J]. arXiv preprint arXiv:1406.3830, 2014.
- [9] KALCHBRENNER N, GREFFENSTETTE E, BLUNSOM P. A convolutional neural network for modelling sentences[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (ACL 2014). Baltimore:ACL, 2014: 655.
- [10] ZHOU Chunting, SUN Chonglin, LIU Zhiyuan, et al. A C-LSTM neural network for text classification[J]. arXiv Preprint arXiv:1511.08630,2015.
- [11] CAO Ziqiang, LI Sujian, LIU Yang, et al. A novel neural topic model and its supervised extension[C]// AAAI'15; Proceedings of the Twenty - Ninth AAAI Conference on Artificial Intelligent. Austin, Texas, USA; AAAI,2015;2210.
- [12] DANI Y, CHRIS D, WANG Ling, et al. Generative and discriminative text classification with Recurrent Neural Networks[J]. arXiv preprint arXiv:1703.01898,2017.
- [13] 李然. 基于深度学习的短文本情感倾向性研究[D]. 北京:北京理工大学,2015.
- [14] 胡朝举, 赵晓伟. 基于词向量技术和混合神经网络的情感分析[J]. 计算机应用研究, 2018,35(12):3556.
- [15] 陈葛恒. 基于极性转移和双向LSTM的文本情感分析[J]. 信息技术,2018(2):149.
- [16] 冯兴杰, 张志伟, 史金钊. 基于卷积神经网络和注意力模型的文本情感分析[J]. 计算机应用研究,2018,35(5):1434.
- [17] 马思丹, 刘东苏. 基于加权 Word2vec 的文本分类方法研究[J]. 情报科学,2019,37(11):38.
- [18] 袁和金, 张旭, 牛为华, 等. 融合注意力机制的多通道卷积与双向GRU模型的文本情感分析研究[J]. 中文信息学报,2019,33(10):109.
- [19] 段丹丹, 唐加山, 温勇, 等. 基于BERT的中文短文本分类算法的研究[J/OL]. 计算机工程;1-12[2019-12-30]. <https://doi.org/10.19678/j.issn.1000-3428.0056222>.
- [20] 胡春涛, 秦锦康, 陈静梅, 等. 基于BERT模型的舆情分类应用研究[J]. 网络安全技术与应用,2019(11):41.
- [21] 周锦峰, 叶施仁, 王晖. 卷积神经网络在短文本情感多分类标注应用[J]. 计算机工程与应用, 2018, 54(22):133.

(上接第6页)

- [5] TOIVONEN H. Sampling large databases for association rules[C]// Proceedings of the 22nd International Conference on Very Large Database. San Francisco, CA, USA; Morgan Kaufmann Publishers Inc., 1996; 134.
- [6] 惠晓滨, 张凤鸣, 虞健飞, 等. 一种基于栈变换的高效关联规则挖掘算法[J]. 计算机研究与发展, 2003, 40(2): 330.
- [7] LIU Bing, HSU W, MA Yiming. Integrating classification and association rule mining[C]// Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining. New York: AAAI, 1998; 1.
- [8] WANG K, ZHOU S, HE Y. Growing decision trees on support-less association rules[C]// KDD. Boston, MA, USA; ACM, 2000; 265.
- [9] XU Xiaoyuan, HAN Guoqiang, MIN Huaqing. Construct concise and accurate classifier by atomic association rules[C]// Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826). Shanghai, China; IEEE, 2004,3: 1604.
- [10] EKNOYAN G, HOSTETTER T, BAKRIS G L, et al. Proteinuria and other markers of chronic kidney disease: A position statement of the national kidney foundation (NKF) and the national institute of diabetes and digestive and kidney diseases (NIDDK) [J]. American Journal of Kidney Diseases, 2003, 42(4): 617.