

基于注意力的 BiLSTM-CNN 中文微博立场检测模型

白 静 李 霏 姬东鸿
(武汉大学计算机学院 湖北 武汉 430072)

摘 要 针对海量社交网络数据,挖掘其中蕴含的立场信息逐渐成为一个重要的研究方向。第五届自然语言处理与中文计算会议(Nlpcc2016)提出了针对中文微博的立场检测任务。已有的立场检测任务工作中,研究者主要通过手工构建特征工程,添加情感词典和专家知识等方式挖掘语义特征,但这种方式需要花费大量人力在特征设计上。另一些研究者将深度学习应用于立场检测领域,但是没有考虑到句子中不同词对立场倾向有不同影响力。注意力机制由于能够凸显出有价值的特征常常被用于优化神经网络模型。提出一种基于注意力的 BiLSTM-CNN 中文微博立场检测方法,首先使用双向(Bi-directional)长短期记忆神经网络(LSTM)和卷积神经网络(CNN)分别获取文本表示向量和局部卷积特征,然后通过注意力机制(Attention Mechanisms)在局部卷积特征中加入影响力权重信息,最终将两种特征融合进行分类。针对 Nlpcc 语料的实验表明,该方法取得了较好的立场检测效果,注意力机制的添加可以有效地提升立场检测的准确性。

关键词 立场检测 微博 神经网络 注意力机制

中图分类号 TP391 文献标识码 A DOI:10.3969/j.issn.1000-386x.2018.03.051

ATTENTION BASED BILSTM-CNN CHINESE
MICROBLOGGING POSITION DETECTION MODEL

Bai Jing Li Fei Ji Donghong
(School of Computer, Wuhan University, Wuhan 430072, Hubei, China)

Abstract For a large number of social network data, mining the information contained in the position has gradually become an important research direction. The Fifth Session of Natural Language Processing and Chinese Computing (Nlpcc2016) proposed the stance detection task for Chinese microblogging. In existing studies, researchers always use feature engineering such as adding emotional dictionaries and expert knowledge. But this approach requires a lot of manpower. Other researchers use the deep learning to detect the stance information. But they do not consider the fact that different words in sentences have different influences on position tendencies. Attention mechanisms are often used to optimize neural network models because they highlight the valuable features. This paper presented an attention-based Chinese microblogging position detection method for BiLSTM-CNN. Firstly, the text representation vector and the local convolution feature were obtained by Bi-directional long-short memory neural network (LSTM) and convolutional neural network (CNN) respectively. Then we added influence weight information to the local convolution features through Attention Mechanisms, and finally combined the two features to classify them. Experiments on Nlpcc corpus showed that the method of this article had achieved a good effect of position detection. The addition of attention mechanism could effectively improve the accuracy of position detection.

Keywords Stance detection Micro-blog Neural networks Attention mechanism

0 引言

随着 Web2.0 的蓬勃发展,互联网逐渐倡导用户为中心,由被动地接收互联网信息向主动创造互联网信息迈进。因此,互联网上产生了大量用户参与的、对于诸如人物和事件等有价值的评论信息。这些评论信息表达了人们的各种情感色彩和立场倾向性,如批评、赞扬等。随着社交媒体信息规模的迅速膨胀,仅靠人工的方法难以应对网上海量信息的收集和处理,因此迫切需要计算机帮助用户快速获取和整理这些相关评价信息^[1],对其中蕴含的隐藏社会现象进行挖掘^[2]。立场检测任务受到了越来越多研究者的关注,同时在2016年的自然语言处理与中文计算会议上也提出了关于中文微博的立场检测任务。足可以看出基于社交文本进行立场信息检测的迫切需要。

立场检测任务希望可以从用户产生的文本中挖掘用户的立场信息。立场信息是针对某个描述主体的情感倾向,描述主体可以是某一产品、某一事件或者某个人等。在已有的研究中,一般将立场检测抽象为分类问题。立场检测问题需要探查文本蕴含的立场信息分别是:支持(FAVOR);反对(AGAINST);无立场(NONE)。如描述句“很想尝试下新出的 iPhone SE...”,其中描述主体为“iPhone SE”,句子中蕴含了观点持有者对于描述主体“iPhone SE”的立场信息。观点持有者针对 iPhone SE 产品富含兴趣想要尝试,立场标签为 FAVOR。与立场检测问题相似,情感分析问题也是针对文本倾向性进行分析。两者的主要区别在于立场倾向是针对某一描述主体产生的;而情感分析的关注重点是文本中的正面或负面的情感特征,比如是否包含具备强烈情感信息的词汇等特征,并不需要牵涉到描述主体这一信息。

目前,在立场检测领域,研究者们提出了许多行之有效的研究方法,其中绝大多数研究方法是采用特征工程 and 传统机器学习相结合的方式^[3-5]。研究者通过人工挖掘领域内知识,手工构建与任务相关的语义特征,添加领域相关的词典资源等方式,来挖掘文本中的立场信息。为了保证特征的质量,传统的文本特征设计需要耗费大量的专家领域知识与高昂的人工成本,并且使用特征工程和传统机器学习模型方法的效果,严重依赖于特征选择策略和模型参数的调优。

基于上述特征工程方法的固有缺陷,研究者有针对性地提出将深度学习方法应用在立场检测领域,从而高效地提取出有价值的特征。深度学习使用神经网络模型自动学习出可以描述数据本质的特征表达,将

原始数据通过一些简单的非线性模型变换为更高层次的抽象表达,再组合多层变换,学习出复杂的函数特征。与传统机器学习相比,深度学习通过自学习特征机制避免了人工设计特征的局限性并降低了人工设计特征的巨大成本,同时深度学习模型还具备更优秀的泛化性和可移植性。

在已有的应用深度学习进行立场检测的研究工作中,研究者没有考虑到不同的特征对于句子整体的立场信息具有不同的影响力^[6]。同时,微博文本天然具有信息稀疏的特性。微博文本的长度通常只有 1 ~ 140 词,并且常常包含短链接、表情符号等非正式文本,所以单条包含的信息十分有限。通过对具体例子的分析发现,微博中少数几个关键的词就可以大致反映出句子的立场类别。比如虚词中的“的”,“着”就对句子的立场检测没有实际作用,而“赞”,“福利”,“差评”等词汇则可以很好地反映出句子的立场倾向。因此,受限与微博文本的稀疏性和不规范性,并且考虑到在已有的使用深度学习方法中研究者并未对特征的重要程度加以区分而是一视同仁的思路,如何从有限的句子信息中获得更丰富的文本特征表示是本文需要解决的问题。

CNN 已经被证实在文本建模方面具有强大的特征提取能力。CNN 中池化操作主要目的是有效减少模型中的参数并固定输出特征的维数,是 CNN 中必不可少的步骤。常用的池化策略分为最大池化和平均池化两种。最大池化虽然能减少模型参数数量,有利于缓解模型过拟合问题,但是丢失了文本的顺序信息与特征的强度信息。平均池化策略对领域内特征求平均,丢失了特征的强度信息。由此,本文借鉴注意力机制的思想来改进池化策略,通过计算不同特征的影响力权重来衡量特征的重要程度,既能够凸显出重要的卷积特征,又可以改善传统池化策略存在的固有问题。

具体而言,本文提出了一个基于注意力的 BiLSTM-CNN 混合网络模型,模型中使用了 Augenstein I 等^[7]针对英文立场检测任务提出的改进 BiLSTM 获得的中间文本表示作为 Yong 等^[8]提出的在 CNN 池化层结合注意力方法的部分输入数据,通过计算局部卷积特征和 BiLSTM 产生的中间文本表示的余弦相似度,从而产生针对局部卷积特征的影响力权重,达到凸显有效特征的目的,最终将 CNN 学习得到的句子表示和 BiLSTM 产生的中间文本表示一同通过 Softmax 进行分类获得立场标签。本文的主要贡献在于将注意力机制应用到了中文微博立场检测领域。针对 Nlpcc 语料的实验表明,本文方法取得了较好的效果,通过添加注意力机制可以有效提升立场检测的准确性。

1 相关工作

目前立场检测工作主要分为基于特征工程和深度学习两大类。

1) 基于特征工程方法的重点是如何选取合适的特征表示文本。Xu 等^[3]基于中文微博使用了丰富的语义表示技术(如 Para2vec, LDA(Latent Dirichlet Allocation), LSA(Latent Semantic Analysis)等)来获取原始文本中蕴含的语义信息。同时通过研究主观性词典构建专门立场词典特征库,并使用随机森林,基于线性核函数的支持向量机(SVM-Linear)以及基于 RBF 核函数的支持向量机(SVM-RBF)等机器学习算法研究不同语言特征和不同分类模型在立场检测任务中的表现。Liu 等^[4]通过加入情感词典等资源构建文本特征,并使用重复采样技术处理不平衡数据集,同时结合随机森林,朴素贝叶斯等多种分类器进行立场检测。此外, Sun 等^[5]探讨了基于语义特征,形态特征,情感特征和语法特征等多种特征在立场检测任务中的作用。可以看出,以上基于特征工程与传统机器学习分类算法的方法通常需要依赖于构造富含领域知识的复杂特征和分类器的参数调优。

2) 基于深度学习的方法。相比传统基于特征工程的方法深度学习模型拥有优秀的自特征提取能力。在针对中文的立场检测领域中 Yu 等^[6]使用改进的混合网络模型对立场检测问题进行建模。通过构建 CNN 中嵌套 LSTM 的网络学习文本表示取得了不错的效果。针对 Twitter 语料, Wei 等^[9]首先使用 google news 语料训练 word embedding 词向量作为句子的词表示,随后使用 CNN 学习特征提取并进行分类。Zarrella 等^[10]采用循环神经网络模型,使用迁移学习的思路将其他领域知识迁移到立场检测任务上。Augenstein 等^[7]采用了条件编码思想改进长短期记忆神经网络(LSTM)完成对描述主体和描述文本联合建模。

循环神经网络 RNN(Recurrent Neural Networks)首先被 Socher 在处理句法分析问题时采用^[11]。RNN 被证明在解决序列化问题上效果突出,在机器翻译等领域取得了不错的效果,但是在求解过程中存在梯度下降和梯度消失的问题。由此 Hinton^[12]提出的长短期记忆单元 LSTM(Long Short Term Memory)有效地解决了上述问题。卷积神经网络 CNN(Convolutional Neural Networks)是深层神经网络中的一种, Lecun 等^[13]采用 CNN 实现了一个多层结构的神经网络学习算法。由于 CNN 利用空间相对关系减少参数数目从而提高了反向传播 BP 算法的训练性能。CNN 近年来被广泛应

用于自然语言处理领域上并取得了良好的效果^[14]。Kim 提出了一个可以同时利用基于特定任务调整的词向量卷积神经网络模型,用于句子级别的文本分类任务^[15]。

近年来,注意力机制逐渐成为了深度学习领域研究的一个热点,被广泛应用于 QA、sequence to sequence、机器翻译^[16]、句子级别摘要^[17]等领域,均取得了良好的效果。Luong 等^[18]最早在文本表示领域中提出了 global attention 和 local attention 模型用于机器翻译,二者的区别在于计算注意力概率分布的范围不同,针对机器翻译的任务特性,对不同的输入采用不同的方法来计算注意力概率。Rush 等^[17]将注意力机制应用于自动摘要领域,并给出可视化模型,直观地体现了注意力机制对模型的优化。

标准的 CNN 结构通常由一个或多个卷积层和池化层组成,池化层由于可以有效减少模型中的参数并固定输出特征的维数是 CNN 中必不可少的步骤。池化策略通常分为最大池化和平均池化两种。最大池化虽然能减少模型参数数量有利于缓解模型过拟合问题,但是丢失了文本的顺序信息与特征的强度信息。平均池化策略由于是对领域内特征求平均,丢失了特征的强度信息。针对 CNN 池化层的注意力机制工作中, Wang 等^[19]针对关系分类问题,在 CNN 的基础上提出了 input attention 和 pooling attention 两种层面的注意力机制。Yong 等^[8]针对 CNN 的池化层使用注意力机制进行改进,缓解了传统池化策略存在的信息丢失等问题。

现阶段针对微博文本的语言特性,如何合理科学地设计出可以学习到更丰富的文本特征表示的模型是本文研究的重点。

2 本文工作

本文提出一种基于注意力的 BiLSTM-CNN 混合网络立场检测模型。首先,本文使用 word2vec 获取文本的词向量表示。然后将词向量分别输入到 BiLSTM 和 CNN 中,获得中间文本表示和卷积特征。为了缓解 CNN 池化层的信息丢失问题,并进一步优化卷积特征,本文将基于注意力的池化策略应用到模型中,通过使用 BiLSTM 学习得到的中间文本表示用于指导卷积特征加权。最后将 CNN 学习得到的最终文本表示和 BiLSTM 获得中间文本表示进行拼接,一同输入 Softmax 层得到三类立场标签。模型结构如图 1 所示。接下来将对模型中各个部分的具体实现进行介绍。

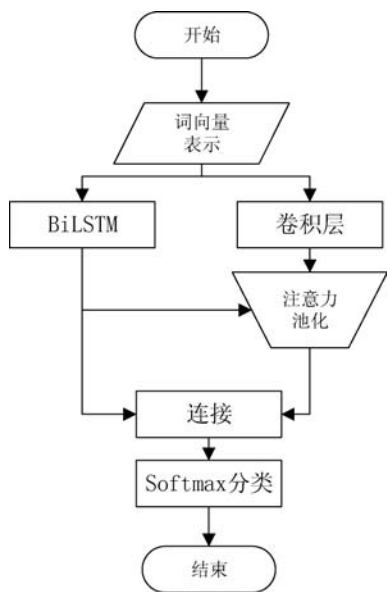


图1 基于注意力的 BiLSTM-CNN 混合网络流程图

2.1 微博词向量表示

本文使用 google 开源工具 word2vec^[20] 获得文本词向量表示。word2vec 提供了 skip-gram 和 CBOW 两种学习算法,本文使用 skip-gram 算法来获得微博的词向量表示。在实际应用中,训练所需语料的规模、词向量的维数等因素都会对词向量质量产生影响。本文除了使用 Nlpcc 提供的语料之外,还增加了收集到的 1 000 万条微博一同进行词向量训练。将获得的描述主体词向量表示和描述文本(微博)词向量表示作为 BiLSTM 的输入数据。将描述文本词向量表示作为 CNN 的输入数据。

由于使用 CNN 处理自然语言问题,需要将句子转换为矩阵形式的数值表示作为输入数据。本文在此通过对句子中每个词的词向量表示进行拼接获得句子的矩阵数值表示。假设微博的中文本长度为 n ,词向量维度为 d ,通过句子中的各个词的词向量堆叠,从而获得维度为 $n \times d$ 的矩阵 X 作为 CNN 的输入数据。

2.2 BiLSTM

在基于注意力的 BiLSTM-CNN 混合网络模型中 BiLSTM 模块的实现部分本文使用了 Augenstein 等^[7] 提出的改进的 BiLSTM 模型,该模型在 Semeval 发布的针对英文 twitter 的立场检测任务获得了非常不错的效果,以下将介绍该模型的具体实现。

由于基于 LSTM 单元的循环神经网络可以有效利用历史标记信息,由此可以认为描述主体对于描述文本有一定指导意义,所以通过建立了两个 LSTM 模型分别是针对描述主体训练生成的 LSTM-target 以及针对描述文本训练生成的 LSTM-weibo。模型内部使用 LSTM-target 输出的最终隐含单元状态作为 LSTM-wei-

bo 隐含单元的初始状态。

在图 2 中我们详细描述了模型状态传递的过程,图 2 中序列 w_1, w_2 表示描述主体,序列 w_3, w_4, \dots, w_7 表示描述文本, c_{ir}, c_{il} 分别表示右向和左向的隐含单元状态, h_{ir} 和 h_{il} 分别表示右向和左向的隐含层输出, C 表示 BiLSTM 最终输出的中间文本表示向量。图中粗箭头所示, c_{il} 的状态传递给了 c_{7l} , c_{2r} 的状态传递给了 c_{3r} 。图中最终产生两个方向的输出向量分别为 h_{7r} 和 h_{3l} ,将其首尾连接获得最终输出向量 C 。

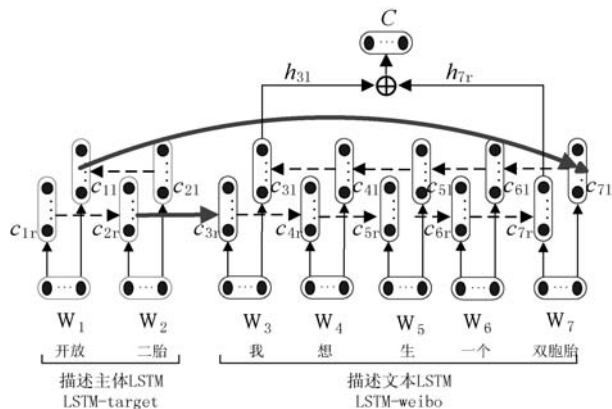


图2 BiLSTM 模块

通过状态的传递,增强了 LSTM-weibo 模型中描述主体的信息。下式中给出了单向 LSTM 的状态传递过程,其中序列 w_1, w_2, \dots, w_T 表示描述主体,序列 w_{T+1}, \dots, w_N 表示描述文本, $[h_0, c_0]$ 的初始状态为 0:

$$[h_1, c_1] = LSTM^{TARGET}(w_1, h_0, c_0)$$

...

$$[h_T, c_T] = LSTM^{TARGET}(w_T, h_{T-1}, c_{T-1})$$

$$[h_{T+1}, c_{T+1}] = LSTM^{Weibo}(w_{T+1}, h_0, c_T)$$

...

$$[h_N, c_N] = LSTM^{Weibo}(w_N, h_{N-1}, c_{N-1}) \quad (1)$$

BiLSTM 最终将产生两个方向的输出向量 $h_r \in \mathbb{R}^{1 \times H}$ 和 $h_l \in \mathbb{R}^{1 \times H}$,将二者顺序连接获得中间文本表示向量 $C \in \mathbb{R}^{1 \times 2H}$,其中 H 表示预设的隐含单元数目:

$$C = h_r \oplus h_l \quad (2)$$

2.3 CNN

2.3.1 卷积层

如图 3 所示,卷积层接收大小为 $n \times d$ 的微博词向量特征矩阵 X ,矩阵包含 n 个词,词向量维度为 d 。矩阵 X 的每一行为句子中一个词的词向量。选取尺寸为 $m \times d$ 的卷积核 $w \in \mathbb{R}^{m \times d}$ 对输入矩阵 X 进行卷积操作获得特征值 g_i 。本文使用了“same”模式进行卷积,该模式将获得和输入矩阵规模相同的输出,卷积过程如公式所示:

$$g_i = f(w \times X_{i:i-m+1} + b) \quad (3)$$

式中, f 表示 ReLu (Rectified Linear Units) 激活函数。为了加快训练收敛速度, 这里使用的激活函数是 ReLu 函数。 m 为卷积计算滑动窗口的大小, b 表示偏至项。 $X_{i:i+m-1}$ 表示在 X 的第 i 行到第 $i+m-1$ 行范围内抽取的局部特征。

如图 3 所示, 本文将卷积核的提取结果组成一个 $n \times T$ 的特征矩阵 S 。

$$S = [v_1, v_2, \dots, v_n] = \begin{bmatrix} g_1^1 & g_1^2 & \dots & g_1^T \\ g_2^1 & g_2^2 & \dots & g_2^T \\ \vdots & \vdots & & \vdots \\ g_n^1 & g_n^2 & \dots & g_n^T \end{bmatrix} \in \mathbb{R}^{n \times T} \quad (4)$$

式中: T 表示卷积核个数; n 表示句子长度。矩阵 S 中的每一列是经过卷积操作后得到的特征向量:

$$G = [g_1, g_2, \dots, g_n] \in \mathbb{R}^n \quad (5)$$

矩阵中的每一行表示 T 个卷积核在句子矩阵相同位置上的提取结果。由于汇集了 T 个卷积核的提取结果, 所以 S 中行向量 v_i 表示了针对句子某个位置的提取出的所有卷积特征。

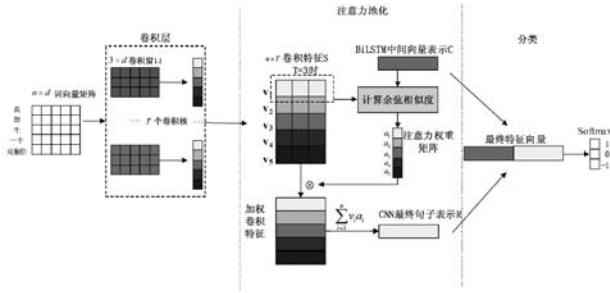


图 3 基于注意力的 BiLSTM-CNN 混合网络模型原理图

2.3.2 基于注意力的池化机制

由于微博存在文本短小、信息稀疏的特性, 并且传统池化策略会产生信息丢失的问题, 如何从微博中学习更丰富的文本表示, 是本文的关注点。神经网络的注意力机制借鉴了人类视觉系统中的“注意力”原理, 通过将注意力集中在目标对象 (如图像、文本) 中最相关的部分而不是所有信息。注意力机制应用到学习句子表示的相关任务时, 主要思想是句子中不同词包含的语义信息不同, 包含更多语义信息的词将产生一个高的影响力权重, 由此注意力机制可以判断出句子中哪些词相对重要, 达到优化模型的目的。所以本文有针对性的采用基于注意力机制池化策略, 既使得重要的特征得以凸显, 又缓解了传统池化策略造成的信息丢失问题。本文在此借鉴了 Yong 等^[8]提出的 CNN 的注意力池化机制, 并将上一步获得的改进 BiLSTM 模型作为注意力池化机制的输入数据。

由于 BiLSTM 可以很好地保留文本的历史、未来信息和时序信息, 所以本文认为在第 2.2 节中介绍的

改进 BiLSTM 模型产生的中间文本表示向量是对原始文本的一个高质量表示, 可以代表句子的整体信息。相对于 BiLSTM 而言, CNN 可以更好地提取出句子的局部上下文信息。所以, 本文通过计算中间文本表示和局部卷积特征的余弦相似度来衡量出局部卷积特征对句子整体语义信息的贡献程度。

在卷积层得到卷积特征矩阵 S 中, $v_i \in \mathbb{R}^{1 \times T}$ 表示了针对句子矩阵某一位置所有卷积核提取的特征。本文在具体工作中通过设置隐含单元数目 H 和卷积核个数 T 的数量关系 $T = 2H$, 从而将 $v_i \in \mathbb{R}^{1 \times T}$ 与 BiLSTM 获得的中间文本表示 $C \in \mathbb{R}^{1 \times 2H}$ 映射到相同的维度空间中, 进而计算余弦相似度。可以认为如果局部卷积特征 v_i 和中间文本表示 C 具有高相似度, 那么意味着 v_i 包含更强的语义信息, 所以将会得到 v_i 的一个高的注意力权重。

如图 3 中, 首先由局部卷积特征 v_i 与中间文本表示 C 计算余弦相似度 e_i , 随后通过 e_i 获得注意力权重值 a_i :

$$e_i = \text{cossim}(C, v_i) \quad (6)$$

$$a_i = \frac{\exp(e_i)}{\sum_{i=1}^n \exp(e_i)} \quad (7)$$

在计算完所有的局部卷积特征的注意力权重后获得注意力权重矩阵, 逐一将局部卷积特征 v_i 点乘与之相对应的注意力权重值 a_i 完成加权。最后对加权后的局部卷积特征进行累加操作, 获得 CNN 最终句子表示 R , 计算方式如公式所示:

$$R = \sum_{i=1}^n a_i v_i \in \mathbb{R}^{1 \times T} \quad (8)$$

基于注意力的池化策略区别于最大池化策略, 并非简单的选择卷积特征中的最大值, 而是通过计算局部特征对于整个句子的影响力权重来衡量局部卷积特征的重要性, 相比之下保留了更多的信息。注意力池化策略与平均池化策略相比, 保留了局部特征的强度信息。

2.3.3 Softmax 层

如图 3 中, 在这里将 CNN 最终句子表示 R 与 BiLSTM 产生的中间文本表示 C 进行首尾连接, 作为最终的立场检测特征向量, 利用 Softmax 输出立场检测判定结果。依据训练数据中的样本标签, 采用反向传播算法来对模型参数进行梯度更新。分类过程如公式所示:

$$y = w_s([C \oplus R] \odot p) + b_s \quad (9)$$

$$\text{prob}_c = \frac{\exp(y_c)}{\sum_{c' \in L} \exp(y_{c'})} \quad (10)$$

为了防止模型过拟合,本文在 Softmax 层中采用了 Dropout 策略和 L_2 正则矩阵。Dropout 策略通过随机丢弃一部分模型参数,从而降低模型复杂度。其中 p 表示 Dropout 概率, w_s 表示 L_2 范式参数矩阵, L 表示立场类别标签, b_s 表示偏至项。

2.4 模型训练

模型的训练方法主要采用梯度下降法。在梯度下降法中主要分为批量梯度下降法和随机梯度下降两种。二者的主要区别在于收敛速度方面。批量梯度下降法虽然能够很好地找到最优解,但是由于每次更新权值需要全部样本参与运算,收敛速度较慢。随机梯度下降法每次更新权值只需要一个样本参与运算。随机梯度下降法可以显著加快收敛速度,但是容易造成收敛到局部最优解的情况。为了能兼顾两种方法的优点,本文采用 mini-batch 的梯度下降法进行训练,即每次更新权值只需要一部分样本参与计算,这样能够在保证寻找最优解的同时加快训练速度。实验中, Batch 的大小被设置为 45 能够获得比较好的折中效果。本文使用了三种不同尺寸的卷积核以获得多种卷积特征。在实验中设置卷积窗口尺寸分别为 $m=3$ 、 $m=4$ 、 $m=5$ 。

3 实验

3.1 数据集

本文选用了 NLPCC 发布的中文微博立场检测任务提供的公开语料数据集来验证本文方法。数据集提供了 3 000 条已标注立场类别的训练数据,1 000 条已知立场类别标签的测试数据。每条数据包含微博正文,描述主体,立场标签。数据集约定立场类别标签分别是:FAVOR, AGAINST, NONE,表示支持,反对,无立场三种立场。数据集共分为五个描述主体,分别是:#春节放鞭炮,#Iphone Se,#俄罗斯在叙利亚的反恐行动,#开放二胎,#深圳禁摩限电。每个立场标签和描述主体的分布统计结果如表 1 – 表 2 所示。

表 1 Nlpcc 数据集样例

描述主体	微博正文	立场类别
春节放鞭炮	考虑到空气质量的问题,咱们还是少放点鞭炮吧。	支持
	过年放鞭炮是多年来的传统习俗。	反对
	今年春节,您打算放鞭炮吗?	无立场
俄罗斯在叙利亚的反恐行动	支持俄罗斯反恐!	支持
IphoneSE	讲真,对 iphoneSE 很期待。	支持

续表 1

描述主体	微博正文	立场类别
开放二胎	@姚晨 用实际行动响应国家二胎政策,自曝将在 11 月给小土豆“添加”一个小伙伴	支持
深圳禁摩限电	没有考虑老百姓的心声,没有关注各个阶层的现状。这样的一刀切是不能解决问题的。	反对

表 2 Nlpcc 数据集语料分布

描述主体	FAVOR	ANAINST	NONE
#春节放鞭炮	250	250	100
#Iphone Se	245	209	146
#俄罗斯在叙利亚的反恐行动	250	250	100
#开放二胎	260	240	240
#深圳禁摩限电	160	300	126

由此可见,训练数据集中存在 14 条漏标记数据,本文将漏标记数据移出训练集。总计 2 986 条语料构成训练集。使用 Nlpcc 公开标签的 1 000 条测试数据作为测试集。

3.2 数据预处理

由于微博文本中常常包含表情,URL,用户名和微博语料特有的话题标签#Hashtag 等非正式文本。这些非正式文本给立场检测任务带来了很大的挑战。为了去除不必要的噪声干扰,本文预先对微博文本以下预处理:

- 1) 过滤所有标点符号和特殊字符,只保留具有语义价值信息的中英文文本。
- 2) 格式转换。将原始数据中全角英文字符转换为半角英文字符。英文字符全部转换为小写字符。
- 3) 移除微博中的短连接以及@ 标签。
- 4) 分词处理。中文中词与词之间没有办法直接区分,所以需要对中国文本进行分词处理。分词的效果会直接影响实验结果。本文使用 jieba 分词进行中文分词。

3.3 评价指标

评价分类器的主要评价指标为准确率和召回率。对于给定的测试数据集,准确率是指分类器正确分类的样本数与总样本数之比,召回率是指被正确判定的正例占总正例的比重,二者相互制约。为平衡准确率和召回率之间的关系,引入综合衡量指标 F 度量值 (F-measure) 作为分类器评价指标。

在 Nlpcc 提供的官方评测标准中,需计算支持 (FAVOR) 标签的 F 度量 (F_{FAVOR}),反对 (AGAINST) 标

签的 F 度量 (F_{AGAINST}),使用二者的宏平均值 (F_{avg}) 作为最终评价指标:

$$F_{\text{avg}} = \frac{F_{\text{FAVOR}} + F_{\text{AGAINST}}}{2}$$

(11)

$$F_{\text{FAVOR}} = \frac{2P_{\text{FAVOR}}R_{\text{FAVOR}}}{P_{\text{FAVOR}} + R_{\text{FAVOR}}}$$

(12)

$$F_{\text{AGAINST}} = \frac{2P_{\text{AGAINST}}R_{\text{AGAINST}}}{P_{\text{AGAINST}} + R_{\text{AGAINST}}}$$

(13)

3.4 基 线

本文选择在 Nlpcc2016 会议上发表的 Nanyu 的实验结果作为基线。Nanyu 提出了 CNN-LSTM 混合神经网络模型针对中文微博进行立场检测,取得了很好的效果。Nanyu 模型结构如图 4 所示。在 CNN 的卷积层和池化层中间嵌入 LSTM 层构建混合网络模型。首先使用卷积层对文本特征进行初步提取,随后对卷积层的结果进行时序展开,使用 LSTM 神经网络再次学习文本表示,最后通过池层,Softmax 层输出立场分类标签。



图 4 Nanyu 混合神经网络结构

与 Nanyu 混合网络结构有所不同,本文考虑卷积操作之后有关原始文本的时序信息可能已经不再完整,或者已经被破坏。在卷积层之后加入 LSTM 提取特征可能不能很好地捕获到原始文本的时序信息,没有发挥出 LSTM 的优势。所以本文使用并行特征提取结构,有别于 Nanyu 的串行结构。由于 BiLSTM 可以很好地提取上下文信息与文本时序信息,本文不但使用 BiLSTM 学习得到的深层文本表示来指导卷积特征进行加权,也将它与加权后的卷积特征一同输入到分类器中进行分类。此外,区别于 Nanyu 没有考虑到不同特征的影响力不同,本文通过在池化层中添加注意力机制,不但减少了传统池化策略的信息丢失问题,而且在卷积特征中加入了影响力权重信息,从而优化模型。

3.5 参数设置

实验中涉及的参数和函数设置如表 3 所示。使用 ReLu 作为激活函数。在实验中所涉及的偏至项 b 与参数矩阵 W ,均采用随机初始化的方式进行初始化。实验采用表 2 中的训练集和测试集,使用本文方法测试在不同滤波器个数 T 和不同 Dropout 比例 p 下的 F_{avg} 值。

表 3 模型超参数设置

参数	参数/函数名称	值
H	隐藏层单元数	T/2
d	词向量维度	200
m	滤波器窗口大小	3,4,5
T	卷积核个数	50,80,100,150
p	Dropout 比率	0.1,0.3,0.5,0.7,0.9
Batch size	批量梯度下降样本数	45
----	激活函数	Relu

表 4 给出了不同 Dropout 比率 p 和不同滤波器数量 T 下的 F_{avg} 值。从实验结果可以看出,当 Dropout 值 p 大于 0.5 时,立场检测的分类效果明显下降。当滤波器的数量 T 增多时,立场检测的 F_{avg} 有所提升,当 T 达到 150 个时,对立场检测最好。在计算相似性时,为了将 BiLSTM 文本表示和局部卷积特征映射到同一维度,我们令 $T = 2H$,所以卷积核个数 T 会影响 BiLSTM 模块的复杂度。综上,本文在模型训练复杂度和训练时间没有显著增加的前提下,最终选择 150 个滤波器数量,Dropout 比率 p 为 0.5 在此条件下,本文设计的立场检测方法准确性可以达到 0.579 1。

表 4 不同滤波器个数和 Dropout 比率下的立场检测效果

	$T = 50$	$T = 80$	$T = 100$	$T = 150$
$p = 0.1$	0.497 1	0.502 0	0.533 7	0.563 9
$p = 0.3$	0.493 3	0.497 3	0.537 2	0.570 1
$p = 0.5$	0.489 0	0.491 1	0.539 0	0.579 1
$p = 0.7$	0.464 2	0.470 3	0.491 5	0.497 2
$p = 0.9$	0.319 7	0.332 1	0.348 0	0.354 0

3.6 实验结果

为了验证本文提出的基于注意力的混合神经网络在中文微博立场检测方法的有效性。将本文方法与传统机器学习方法 (SVM), CNN, BiLSTM, 未添加注意力机制的 BiLSTM-CNN 混合网络以及基线对比。具体实验如下:

- 1) 传统机器学习:采用 word2vec 模型训练词向量作为输入数据,使用 SVM 分类。
- 2) CNN:采用 word2vec 训练词向量,训练 CNN 提取特征并分类。
- 3) BiLSTM:采用 word2vec 训练词向量,训练改进的 BiLSTM 提取特征并分类。
- 4) BiLSTM-CNN-max:采用 word2vec 训练的词向量,使用基于 1 - max pooling 策略的 CNN 提取特征,融合 BiLSTM 学习的文本表示进行分类。

5) Nanyu-NN:采用 Nanyu 提出的混合神经网络方法作为基线。

6) 本文方法 BiLSTM-CNN-ATT:采用本文提出的基于注意力的 BiLSTM-CNN 混合神经网络进行特征提取和分类。

表 5 为不同方法的分类准确性对比。

表 5 不同方法的分类准确性对比

方法	F_{FAVOR}	$F_{AGAINST}$	F_{avg}
SVM	0.263 3	0.325 1	0.294 2
CNN	0.310 7	0.443 7	0.377 2
BILSTM	0.400 3	0.586 2	0.493 3
BILSTM-CNN-1 max	0.631 7	0.498 9	0.565 3
Nanyu-NN	0.672 9	0.458 4	0.566 5
BiLSTM-CNN-ATT	0.663 0	0.495 2	0.579 1

1) 与基线 Nanyu-NN 对比: Nanyu 的思路是在 CNN 的卷积层与池化层之间加入 LSTM 再次学习文本表示,构建混合网络模型。通过和本文方法对比可以看出,本文使用 word2vec 作为基础词表示,构建 BiLSTM-CNN 混合网络模型,并通过注意力机制改进池化层优化特征。本文方法相对于 Nanyu 提出的混合网络模型的效果提升了 2.23%,能够有效提升立场检测的表现。

2) 通过对比单一结构的 CNN 和 BiLSTM 模型发现:相对于单一结构神经网络,混合网络模型的表现更好。理论上随着模型的深度增加,由于参数增加模型的表达能力更加优秀。但是模型更加复杂也更容易出现过拟合。本文通过 Dropout 策略和 L_2 正则进行平衡。

3) 通过和未添加注意力机制,采用普通池化策略的 BiLSTM-CNN 模型对比发现:注意力机制可以有效地优化特征,使重要的特征得以凸显。池化层通过注意力机制缓解了信息丢失问题,提升了模型的表现。

4) 通过与 SVM 模型对比发现:可以看出对比实验仅仅使用 word2vec 词向量作为输入是远远不够的。由于传统机器学习模型极大地依赖于特征工程的构建方式,需要大量的领域内知识和词典资源,需要丰富的特征选择策略以及大量的模型参数调优工作,所以仅仅使用词向量作为特征的 SVM 模型并没有获得很好的效果。也可以看出,深度学习方法减少了特征工程类方法地所需工作,模型本身已具备优秀的特征提取能力。在已有的立场检测任务中,研究者提出的基于特征工程的方法亦取得了不错的效果。

为了讨论本文提出的注意力机制对卷积神经网络

池化层的改进。将本文方法与采用传统池化策略的混合网络模型对比,具体实验如下:

1) BiLSTM-CNN-avg:采用 word2vec 训练的词向量,使用基于 average pooling 策略的 CNN 提取特征,融合 BiLSTM 学习的文本表示进行分类。

2) BiLSTM-CNN-1 max:采用 word2vec 训练的词向量,使用基于 1-max pooling 策略的 CNN 提取特征,融合 BiLSTM 学习的文本表示进行分类。

3) BiLSTM-CNN-kmax:采用 word2vec 训练的词向量,使用基于 k-max pooling 策略的 CNN 提取特征,融合 BiLSTM 学习的文本表示进行分类。

表 6 为不同池化策略的分类准确性对比。

表 6 不同池化策略的分类准确性对比

方法	F_{FAVOR}	$F_{AGAINST}$	F_{avg}
BILSTM-CNN-avg	0.613 1	0.461 7	0.537 4
BILSTM-CNN-1 max	0.631 7	0.498 9	0.565 3
BILSTM-CNN-kmax	0.647 9	0.478 3	0.563 1
BiLSTM-CNN-ATT	0.663 0	0.495 2	0.579 1

从与不同池化策略的混合网络对比实验可以看出,在传统的池化策略中 1-max pooling 池化策略比 k-max pooling 有略微的提升,但是均显著好于 average pooling 策略。本文提出的基于注意力的池化策略相对于表现最好的 1-max pooling 策略有明显的效果提升。与传统池化策略相比,基于注意力机制改进的池化操作可以明显提升实验效果,减少池化层的信息丢失。同时注意力机制通过凸显有效的立场特征达到优化模型的目的,提升了立场检测任务的准确率。

4 结 语

近年来,识别文本内容的立场倾向性已经成为自然语言处理中较为重要的课题之一。区别于传统机器学习需要构建反映任务特性的复杂特征集合,同时有别于立场检测领域内其他工作对不同词一视同仁进行处理的方法,本文在已有工作的基础上提出了一种基于注意力的 BiLSTM-CNN 混合网络立场检测方法。首先使用改进的 BiLSTM, CNN 分别获得中间文本表示和局部卷积特征,进一步通过注意力机制得到不同局部卷积特征的不同影响力权重从而凸显出有效特征,最终将 BiLSTM 学习得到的中间文本表示和 CNN 学习得到的最终句子表示相结合得到最终句子特征向量通过 Softmax 进行分类。本文的贡献主要在于将注意力

机制应用于立场检测领域,在官方数据集上的实验结果显示,本文方法能够有效针对中文微博进行立场检测并取得了良好的表现。下一步的工作将针对一些复杂语法结构的短文本展开,进一步提高中文微博文本立场检测的准确性。

参 考 文 献

- [1] 魏犇, 向阳, 陈千. 中文文本情感分析综述[J]. 计算机应用, 2011, 31(12): 3321-3323.
- [2] 赵妍妍, 秦兵, 刘挺, 等. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848.
- [3] Xu J, Zheng S, Shi J, et al. Ensemble of Feature Sets and Classification Methods for Stance Detection[C]//International Conference on Computer Processing of Oriental Languages. Springer International Publishing, 2016: 679-688.
- [4] Liu L, Feng S, Wang D, et al. An Empirical Study on Chinese Microblog Stance Detection Using Supervised and Semi-supervised Machine Learning Methods[M]//Natural Language Understanding and Intelligent Applications. Springer International Publishing, 2016.
- [5] Sun Q, Wang Z, Zhu Q, et al. Exploring Various Linguistic Features for Stance Detection[M]//Natural Language Understanding and Intelligent Applications. Springer International Publishing, 2016: 840-847.
- [6] Yu N, Pan D, Zhang M, et al. Stance Detection in Chinese MicroBlogs with Neural Networks[M]//Natural Language Understanding and Intelligent Applications. Springer International Publishing, 2016: 893-900.
- [7] Augenstein I, Rocktäschel T, Vlachos A, et al. Stance Detection with Bidirectional Conditional Encoding[C]//Conference on Empirical Methods in Natural Language Processing. 2016: 876-885.
- [8] Meng J E, Zhang Y, Wang N, et al. Attention pooling-based convolutional neural network for sentence modelling[J]. Information Sciences An International Journal, 2016, 373(C): 388-403.
- [9] Wei W, Zhang X, Liu X, et al. pkudlab at SemEval-2016 Task 6: A Specific Convolutional Neural Network System for Effective Stance Detection[C]//International Workshop on Semantic Evaluation, 2016: 384-388.
- [10] Zarrella G, Marsh A. MITRE at SemEval-2016 Task 6: Transfer Learning for Stance Detection[C]//International Workshop on Semantic Evaluation, 2016: 458-463.
- [11] Socher R, Perelygin A, Wu J Y, et al. Recursive deep models for semantic compositionality over a sentiment treebank[C]//Proceedings of the conference on empirical methods in natural language processing (EMNLP), 2013: 1631-1642.
- [12] Hinton G E. Learning distributed representations of concepts[C]//Proceedings of the eighth annual conference of the cognitive science society, 1986: 1-12.
- [13] Lecun Y, Kavukcuoglu K, Farabet C C. Convolutional networks and applications in vision[C]//Proceedings of IEEE International Symposium on Circuits and Systems. Lisbon, Portugal; IEEE, 2010: 253-256.
- [14] 吴轲. 基于深度学习的中文自然语言处理[D]. 江苏南京: 东南大学, 2014.
- [15] Kim Y. Convolutional neural networks for sentence classification[C]//Proc of Conference on Empirical Methods in Natural Language Processing, 2014: 1746-1751.
- [16] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. Computer Science, 2014.
- [17] Rush A M, Chopra S, Weston J. A neural attention model for abstractive sentence summarization[J]. arXiv preprint arXiv:1509.00685, 2015.
- [18] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint arXiv:1508.04025, 2015.
- [19] Wang L, Cao Z, Melo G D, et al. Relation Classification via Multi-Level Attention CNNs[C]//Meeting of the Association for Computational Linguistics. 2016: 1298-1307.
- [20] Goldberg Y, Levy O. word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method[J]. arXiv preprint arXiv:1402.3722, 2014.

(上接第 214 页)

- [9] 杜宏博, 王丽会. 基于改进暗原色先验模型的快速图像去雾方法[J]. 计算机工程与应用, 2016, 52(1): 178-184.
- [10] 刘楠, 程咏梅, 赵永强. 基于加权暗通道的图像去雾方法[J]. 光子学报, 2012, 41(3): 320-325.
- [11] 蒋建国, 侯天峰. 改进的基于暗原色先验的图像去雾算法[J]. 电路与系统学报, 2011, 16(2): 7-12.
- [12] He K, Sun J, Tang X. Guided image filtering[C]//Proceedings of Computer Vision—ECCV 2010. Berlin Heidelberg: Springer, 2010: 1-14.
- [13] 汪荣贵, 傅剑锋, 杨志学. 基于暗原色先验模型的 Retinex 算法[J]. 光子学报, 2013, 41(6): 1188-1192.
- [14] 徐晶, 刘鹏, 唐降龙. 大气光估计对单幅图像去雾复原的影响[J]. 智能计算机与应用, 2012, 2(3): 68-71.
- [15] 陆海俊, 汪荣贵, 杨娟. 基于均值漂移的暗原色先验图像去雾算法[J]. 合肥工业大学学报, 2016, 39(9): 1205-1210.