计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(7):149−157 [doi: 10.15888/j.cnki.csa.008598]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

多神经网络协作的电力文本类型识别①

陈　鹏[1], 吴旻荣[1], 蔡　冰[1], 何晓勇[2], 金兆轩[2], 金志刚[2], 侯　瑞[3,4]

[1](国网宁夏电力有限公司, 银川 750001)
[2](天津大学 电气自动化与信息工程学院, 天津 300072)
[3](华北电力大学 苏州研究院, 苏州 215123)
[4](华北电力大学 经济与管理学院, 北京 102206)
通信作者: 陈　鹏, E-mail: chenpengtougao2021@163.com

摘　要: 电力企业为实现数字资产管理, 提高行业运行效率, 促进电力信息化的融合, 需要实施有效的数据组织管理方法. 针对电力行业中的数据, 提出了基于字级别特征的高效文本类型识别模型. 在该模型中, 将字符通过 BERT 预训练模型生成电力客服文本动态的高效字向量, 字向量序列输入利用融合注意力机制的双向长短期记忆网络 (BiLSTM), 通过注意力机制有效捕捉文本中帮助实现类型识别的潜在特征, 最终利用 Softmax 层实现对电力文本的类型识别任务. 本文提出的模型在电力客服文本数据集上达到了 98.81% 的准确率, 优于 CNN, BiLSTM 等传统神经网络识别方法, 增强了 BERT 模型的应用, 并有效解决了电力文本类型识别任务中语义的长距离依赖问题.

关键词: BERT 模型; 双向长短期记忆网络 (BiLSTM); 注意力机制; 深度学习; 自然语言处理

引用格式: 陈鹏,吴旻荣,蔡冰,何晓勇,金兆轩,金志刚,侯瑞.多神经网络协作的电力文本类型识别.计算机系统应用,2022,31(7):149–157. http://www.c-s-a.org.cn/1003-3254/8598.html

Power Text Type Recognition Based on Multi-neural Network Cooperation

CHEN Peng[1], WU Min-Rong[1], CAI Bing[1], HE Xiao-Yong[2], JIN Zhao-Xuan[2], JIN Zhi-Gang[2], HOU Rui[3,4]

[1](State Grid Ningxia Electric Power Co. Ltd., Yinchuan 750001, China)
[2](School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)
[3](Suzhou Institute, North China Electric Power University, Suzhou 215123, China)
[4](School of Economics and Management, North China Electric Power University, Beijing 102206, China)

Abstract: To realize digital asset management, improve industry operation efficiency, and promote the integration of power informationization, power companies need to implement effective data organization and management methods. This study proposes an efficient text type recognition model based on character-level features for the data in the electric power industry. In this model, characters are put through the BERT pre-training model to generate dynamic and efficient character vectors of the power customer service text. A BiLSTM network with the attention mechanism is used for the input of character vector sequences. The attention mechanism enables the effective capture of the latent features helpful for type recognition. Finally, we use the Softmax layer to recognize the power text type. The model proposed in this study achieves an accuracy of 98.81% on a data set of power customer service text, which is better than traditional neural network methods such as CNN and BiLSTM. It enhances the application of the BERT model and effectively solves the problem of semantic long-distance dependence in power text type recognition.

Key words: BERT model; bi-directional LSTM (BiLSTM); attention mechanism; deep learning; natural language processing (NLP)

① 基金项目: 国家重点研发计划 (2021YFE0102400)
收稿时间: 2021-10-25; 修改时间: 2021-11-29; 采用时间: 2021-12-08; csa 在线出版时间: 2022-05-31

近年来, 为响应国家对于传统产业数字化发展规划的号召, 电力企业采取了一系列措施来实现从传统电力企业到能源互联网[1] 的转型升级, 并将数字化转型作为企业管理和运营的工作重心. 数字化转型是利用新一代信息技术和人工智能技术对管理和运营过程中产生的一系列数据进行采集、传输、存储、处理和反馈. 数字化转型可以提高各行业内和相互间的数据流动, 从而提高行业的运行效率, 有利于促进国家构建全新的数字经济体系. 在数字化管理运营中产生的数据中存在一部分数据拥有更大的价值等待被发掘, 我们将其称之为数字资产. 对于电力行业而言, 在电力运行的过程中积累了大量的电力生产、运营数据, 可以被发掘出更多的电力数字资产. 电力行业运营的过程中, 客服是电力公司和电力用户沟通交流的重要途径, 客服交流中产生了一系列重要的数据信息. 这些重要的数据信息常常以文本的形式存在, 文本数据记录着电力用户对电力企业的相关需求, 根据将客户对电力企业的反馈进行有效识别, 可以进一步提升电力企业的自动化水平, 加速实现数字型转型.

将不同种类电力相关文本进行合理分类可以为电力数字资产化提供高效的组织管理方式. 电力企业通常采用人工方式来区分这些文本数据, 这种方式不仅极大地消耗了电力公司的人力和财力, 并且难以为电力客户提供足够高效的服务效率. 因此自动识别电力客服文本类型的研究成为电力企业所关注的问题, 利用先进信息技术对文本数据进行分析与智能计算[1] 不仅旨在缓解电力公司客服人员的工作压力, 还将有效地促进后续对电力数字资产的发掘. 不仅如此, 实现电力企业客服文本的高效分类还可以加速电力客户得到相关客服的及时反馈, 提升用户满意度.

本文的贡献在于提出了一种面向电力文本类型识别任务的新方法, 记为 BW_BiLSTM_ATTN. 该方法基于 BERT 全词遮盖 (whole word masking, WWM) 中文预训练模型的动态语义表示, 这种动态语义表示可以使电力文本具有更加符合语义上下文的向量表示, 通过利用融合注意力机制的 BiLSTM 作为编码层来对电力文本进行编码并且捕捉其潜在的文本特征, 利用注意力机制可以使得模型自动提高对电力文本类型识别有帮助的特征权重, 最终利用 Softmax 解码器实现对电力文本的高效自动化分类.

1 相关工作

文本分类[2] 本身是自然语言处理领域中的一个重要分支, 它的任务是根据文本的特征将其划分到预先被定义好的固定类别中. 常见的文本分类的应用有垃圾邮件判定、情感分类等. 最早的文本分类是通过判断文本中是否出现了与类名相同或相近的词汇来进行划分类别, 但这种方法难以在大体量的数据上得到一个好的分类效果[2]. 随后出现了基于知识工程的文本分类方法和基于机器学习的文本分类方法, 前者是利用专业人员的经验、人工提取规则来对文本进行类别的划分, 后者是利用计算机高效的自主学习能力、提取一定的规则来进行类别的划分. 现在, 人工智能的快速发展也使得文本分类有了新的研究方向, 人们开始把深度学习神经网络运用到文本分类[3,4] 中, 有效缓解了传统文本分类中耗费人力、分类效果差等问题.

电力行业作为社会的必需产业, 电力客服作为数字化转型中的重要数据来源途径, 其客服文本在文本分类研究中一直作为重点研究对象. 杨鹏等人[5] 利用循环神经网络实现对电力客服文本的层次语义理解, 关注词汇和字符的语义, 最后提升了客服文本类别划分效果. 朱龙珠等人[6] 利用门控循环神经网络 (GRU) 对国网客服中心的重大服务事件进行了有效的分类. 刘梓权等人[7] 针对电力设备缺陷文本的特点, 构建了基于卷积神经网络 (convolutional neural network, CNN) 对缺陷文本分类. 李灿等人[8] 利用双向长短期记忆网络 (bi-directional LSTM, BiLSTM) 和 CNN 相结合的优化模型对电力客服工单进行了类别划分实验, 得到了较好的效果. 肖禹等人[9] 提出一种基于融合注意力机制的 BiLSTM, 残差网络和 CNN 的模型来实现对中文文本分类. 顾亦然等人[10] 利用 BERT 增强了字符的语义表示并利用 BERT 生成的动态自向量实现高效的电机领域实体识别. 然而现存的一些工作很少有研究充分发挥 BERT 预训练模型作为嵌入层用于电力文本分类任务的模型. 本文提出的方法考量了 BERT 预训练模型对于电力领域的文本高效向量表达, 使用 BERT 输出的字级别的向量表达可以有效避免由分词带来的词表溢出问题, 使用融合注意力机制的 BiLSTM 对电力文本的高效向量表达进行有效的特征提取从而得到对分类任务有作用的特征, 通过使用本文提出的模型对最终使用 Softmax 函数得到文本类别的概率分布.

2 多神经网络协作的电力文本类型识别模型

2.1 BW_BiLSTM_ATTN 模型结构

BERT[11] 由具有多头自注意机制的 Transformer 结构[12] 堆叠组成, 通过预训练将大量语料利用自监督任务得到细粒度的语义信息, 由于 BERT 使用海量的文本进行训练, 所以 BERT 本身蕴含这些文本中的语义信息, 通过在模型中应用预训练模型可以将电力相关文本输入 BERT 得到具有上下文语义信息的词向量从而实现迁移学习. 得到电力文本内容关于 BERT 上下文语义表达的向量表示后, 利用具有注意力机制的 BiLSTM 可以有效捕捉对于文本类型识别的潜在特征, 相较于传统的 BiLSTM 可以让模型通过施加注意力机制的权重比例从而使模型更加自主地学习到对电力文本类型识别有效的特征. 最后通过使用 Softmax 分类结构解码输出文本类别的标签. 本文实现对电力客服文本类型识别的具体网络结构 BW_BiLSTM_ATTN 模型如图 1 所示.

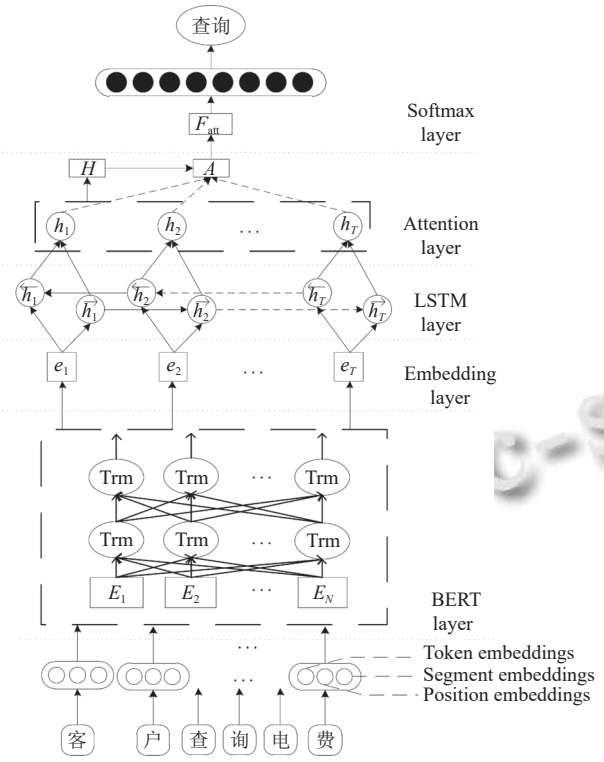

图 1 BW_BiLSTM_ATTN 模型

首先将电力客服文本数据进行预处理, 得到每条客服文本的字符序列, 将得到的电力客服文本字符序列进行位置编码和分段编码, BERT-WWM 将字符的静态嵌入表达, 位置向量和语句的分段信息进行整合, 通过使用堆叠 Transformer 结构来为每个字符计算其关于上下文的语义向量, 得到每个字符的向量表达后使用融合注意力机制的 BiLSTM 网络来捕捉文本分类中潜在的特征, 注意力机制通过矩阵的点积计算求得判断文本类别的特征权重, 从而使模型更加侧重有效的特征表达. 最终通过使用 Softmax 层计算每个类别的概率, 得到模型判断的最大可能性的文本种类, 实现对电力客服文本数据的多分类目标.

本文使用的特征与传统方法相比在于本文使用字符级别特征得到句子的上下文表示, 通过使用字符级别特征可以避免分词过程中低频率出现的词组导致词表溢出的问题, 此外利用 BERT-WWM 预训练模型可以通过计算每个字符和前后字符之间相互的作用来模拟句子中的上下文语义影响, 从而得到更加符合语义表达的字符表示. BiLSTM 使用逻辑门控的方式来更新和修改细胞的状态, 从理论上解决了传统循环神经网络 (recurrent neural network, RNN) 无法保留较早循环神经网络单元的输出结果. 不过 BiLSTM 仍有梯度消失和梯度爆炸的现象从而容易损失一些长程的信息, 使用注意力机制可以动态调整对 BiLSTM 的隐状态权重从而捕捉文本中对类型识别有效的潜在语义特征, 最终使 Softmax 层解码得到更加合理的文本类别.

2.2 BERT 嵌入层

本文使用 BERT 预训练模型作为词嵌入层, BERT[11] 最早由谷歌提出的利用海量的语料, 通过使用大量算力资源进行自监督训练任务训练出来的含有词组的通用语义表达的模型. BERT 的模型结构由多层 Transformer 结构[12] 堆叠而成, Transformer 是一种由多头自注意力机制和全连接神经网络相互叠加, 融合层归一化和残差连接的一种高效神经网络模型. BERT 的模型结构如图 2 所示.

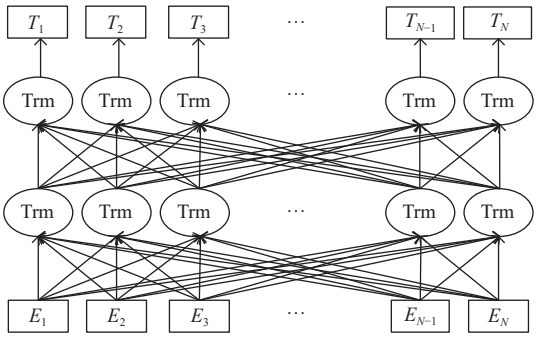

图 2 BERT 结构

BERT 使用遮盖语言模型的任务和下一句预测的任务一同进行自监督训练, 遮盖语言模型是指将文本中随机遮盖 15% 的子词, 其中将 80% 替换成 [MASK] 标签, 10% 替换成随机词汇, 10% 不进行改变. 通过让模型预测遮盖住的 15% 的这些子词从而让模型学习到文本的向量表达, 中文预训练模型 BERT-WWM 模型[13] 与原 BERT 不同的是遮盖的是中文中的整个词的, 通过使用 LTP 工具分割语料从而遮盖完整的词, 通过增进复杂的自训练任务来加大模型预测难度使模型学习到更加符合语义的表达. 通过遮盖词汇例如“线路短路故障和负荷的急剧多变, 使变压器发生故障”中的“变压器”, 模型需要通过理解上下文的语义在自监督的训练中填出“变压器”这一词, 通过这样的自监督训练可以使模型具有丰富的语义知识. 下一句预测的任务目的是学习到句子之间的关系, 通过 1:1 的比例设置正确句子顺序和错误句子顺序, 在让模型判断该顺序是否正确的训练过程中学习到两段句子的关联, 从而使 BERT-WWM 具有判断句子之间关系的能力.

由于采用多层的 Transformer 结构, 因此 BERT 利用多层累加的多头自注意力机制来捕捉文本之间的关联, 得到动态的上下文词向量表达. 通过使用 BERT 可以将预训练模型从海量语料中学习的知识进行迁移, 从而计算出的电网客服文本的词向量都具有上下文的语义特征, 高效语义特征也便于编码层捕捉合适的分类特征. 例如图 1 中句子“客户查询电费”, 这句话中每个字符都有一个固定的字向量, 即 token embedding, token embedding 与传统的词向量 Word2Vec 具有相似的性质, 都是静态的向量表达, 不能很好解决一词多义的问题, 比如“查询”和“查干湖”中的“查”如果用静态向量表达则会使用相同的向量, 这会使模型无法很好捕捉文本的特征, 而通过使用 BERT 进行上下文语义计算, 每个字都和句中的其他字做交互计算, 得到上下文相关的字符向量表达. 由于 BERT 模型使用堆叠 Transformer 结构, 使用全自注意力机制需要引入显式的位置信息, 例如“客户不满意, 举报”和“客户满意, 不举报”具有不同含义, 因此本文使用 BERT 中训练的位置信息, 经过与字符的字向量相加后输入 BERT 中即可让 BERT 模型捕捉到句子中字符出现的前后顺序.

2.3 BiLSTM 层

BiLSTM[14] 是在循环神经网络的基础上改进而来的门控神经网络, 通过使用输入门, 遗忘门, 记忆门和输出门来调整对上一时刻和这一时刻的状态比例, 需要模型记住并传递对电力分类文本有作用的信息并舍弃冗余信息, 从而使神经网络具有长期记忆的效果. LSTM 的网络模型如图 3 所示.

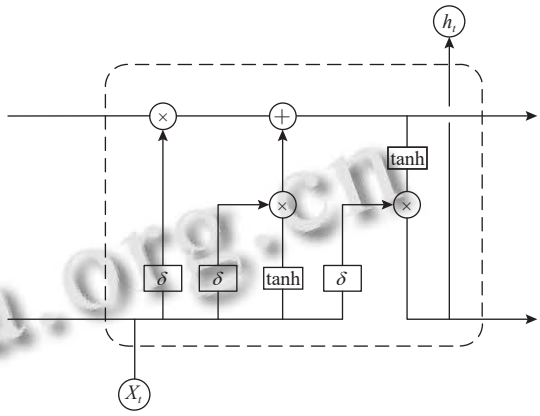

图 3 LSTM 结构

遗忘门的计算公式如式 (1) 所示:

$$\boldsymbol{f}_t = \delta\left(\boldsymbol{W}_f \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_f\right) \tag{1}$$

其中, $\boldsymbol{W}_f$和$\boldsymbol{b}_f$是模型中可训练的参数, $\boldsymbol{h}_{t-1}$是上一时刻输出的隐状态, $\boldsymbol{x}_t$是当前时刻的输入, 将电力分类文本输入进入 BERT 后输出电力文本中每个字符的上下文相关向量, 得到每个字符的向量作为 BiLSTM 的输入$\boldsymbol{x}_t$, 使用激活函数计算得到遗忘门的值$\boldsymbol{f}_t$, 遗忘门可以控制模型忘记电力文本中一些对类型识别无关的信息, 例如“客户查询电费”中“客户”对模型分类提供的信息非常有限, 遗忘门通过控制模型遗忘这类提供信息有限的向量来达到促进类型识别使用高效向量特征. 此外, BiLSTM 使用记忆门来控制模型记住一些对电力文本类型识别有帮助的内容, 记忆门的计算公式如式 (2) 所示:

$$\boldsymbol{i}_t = \delta(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i) \tag{2}$$

式 (2) 与式 (1) 相似, $\boldsymbol{W}_i$ 和 $\boldsymbol{b}_i$ 是模型中可训练的参数, $\boldsymbol{h}_{t-1}$ 是上一时刻输出的隐状态, $\boldsymbol{x}_t$ 是当前时刻的输入, 使用激活函数计算得到记忆门的值 $\boldsymbol{i}_t$, 例如“客户查询电费”中“查询”对模型分类到查询类别提供了决定性的信息, 因此通过记忆门来记住这类对分类提供有效信息的向量. 除了遗忘门和记忆门, BiLSTM 为了反映当前时刻的临时细胞状态, 定义了临时细胞状态$\tilde{C}_t$, $\tilde{C}_t$的计算公式如式 (3) 所示:

$$\tilde{\boldsymbol{C}}_t = \tanh(\boldsymbol{W}_C \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_C) \tag{3}$$

式 (3) 中, $\boldsymbol{W}_C$和$\boldsymbol{b}_C$是模型中可训练的参数, $\boldsymbol{h}_{t-1}$是前一时刻输出的隐状态, $\boldsymbol{x}_t$是当前时刻的输入, 使用双曲正切函数计算得到当前细胞的临时状态. 最终利用遗忘门和记忆门作为系数计算得到当前时刻的细胞状态. 细胞状态的计算公式如式 (4) 所示:

$$\boldsymbol{C}_t = \boldsymbol{f}_t \times \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \times \tilde{\boldsymbol{C}}_t \tag{4}$$

式 (4) 中, $\boldsymbol{C}_{t-1}$是前一时刻的细胞状态, 该式表示通过控制遗忘门和记忆门可以达到控制上一时刻的信息流对当前时刻的细胞状态的影响. 通过使用分别对上一时刻细胞状态和当前时刻的临时细胞状态的加权计算可以得到当前时刻得到这一时刻的细胞状态后需要通过输出门的计算来得到隐层的状态输出, 输出门的计算公式如式 (5) 所示:

$$\boldsymbol{o}_t = \delta(\boldsymbol{W}_o \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_o) \tag{5}$$

式 (5) 中, $\boldsymbol{W}_o$和$\boldsymbol{b}_o$是模型中可训练的参数, $\boldsymbol{h}_{t-1}$是前一时刻输出的隐状态, $\boldsymbol{x}_t$是当前时刻的输入, 使用激活函数得到输出门的值$\boldsymbol{o}_t$, 利用输出门的值可以计算得到隐层的输出, 隐层输出的计算公式如式 (6) 所示:

$$\boldsymbol{h}_t = \boldsymbol{o}_t \times \tanh(\boldsymbol{C}_t) \tag{6}$$

式 (6) 中, 将输出门与当前时刻细胞状态的双曲正切值相乘得到隐层的单向输出 , 隐层的输出则蕴含着当前时刻和之前的序列信息, 若 LSTM 为正向则蕴含着电力文本中正向的分类特征, 例如文本“客户查询电量”中, 正向 LSTM 获得“查询”的向量时只能获得“客户”的特征表达, 而使用反向 LSTM 时“查询”只能获得“电量”的特征表达. 因此只有单向的序列信息往往是不全面的, 在文本中往往一个词的含义与上下文都有关联, 因此采用 BiLSTM 可以将电力文本在分类任务中的特征表达更加全面.

2.4 注意力层

注意力机制将对任务有帮助的内容进行高权重关注, 这与人类的感官注意力模式非常相像, 本文在分类模型中的 BiLSTM 层引入注意力层[15]. 传统方法通过将 BiLSTM 前向与后向传播的最终状态拼接作为句子本身的特征, 但由于 BiLSTM 仍存在梯度消失和梯度爆炸的问题, 这导致句中的信息不能有效的被输出表达, 若电力文本中对类型识别任务有决定作用的字符处于长句的中部, 则当模型训练不够充分时 LSTM 的正向和反向都难以捕捉到中部的有效信息. 本文将注意力机制作用到 BiLSTM 层上, 目的是将每个隐层的输出经过加权的拼接来看作句子整体的嵌入表达, 通过使用注意力机制的权重计算, 输出更符合分类器来判别电力客服文本类别的句子嵌入表达. 注意力层的计算过程如图 4 所示.

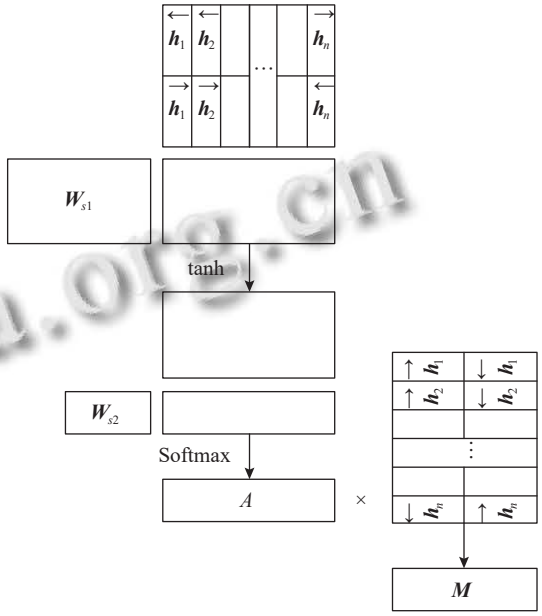

图 4 自注意力机制

图 4 中表示的运算由以下内容阐明: 对于一个含有 n 个字符的电力文本句子通过 BERT 捕捉上下文语义之后的表达$S = (x_1, x_2, \cdots, x_n)$, 其中$S$代表整个句子, x_t代表每个字符通过 BERT 之后需要输入到 BiLSTM 的词向量. 那么 LSTM 的隐层输出可由式 (7)–式 (10) 表示:

$$\overrightarrow{\boldsymbol{h}_t} = \overrightarrow{LSTM}\left(\boldsymbol{x}_t, \overrightarrow{\boldsymbol{h}_{t-1}}\right) \tag{7}$$

$$\overleftarrow{\boldsymbol{h}_t} = \overleftarrow{LSTM}\left(\boldsymbol{x}_t, \overleftarrow{\boldsymbol{h}_{t-1}}\right) \tag{8}$$

$$\boldsymbol{h}_i = \text{concat}\left(\overrightarrow{\boldsymbol{h}_i}, \overleftarrow{\boldsymbol{h}_i}\right) \tag{9}$$

$$\boldsymbol{H} = (\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_n) \tag{10}$$

其中, $\overrightarrow{\boldsymbol{h}_t}$表示 LSTM 的正向隐层输出, $\overleftarrow{\boldsymbol{h}_t}$为 LSTM 的反向隐层输出, $\boldsymbol{H}$ 表示将所有 LSTM 单元的隐层输出进行拼接的矩阵. 随后利用矩阵线性变换等多种运算求得在该句子的向量分布下注意力的权重, 由于一个句子中促进模型识别电力文本类型的重点字词可能会有多个, 因此本文采用了多点的自注意力机制来求得句子中的多种权重概率分布. 公式如式 (11) 所示:

$$\boldsymbol{A} = Softmax(\boldsymbol{W}_{s2} \tanh(\boldsymbol{W}_{s1} \boldsymbol{H}^{\mathrm{T}})) \tag{11}$$

式 (11) 中, $\boldsymbol{W}_{s2}$和$\boldsymbol{W}_{s1}$是可学习的矩阵, 通过一系列运算后可以求得隐状态矩阵$\boldsymbol{H}$所需的注意力权重, 例如“客户查询电费”中, “查询”的注意力权重在训练的过程中会增高, 通过 Softmax 函数来归一化概率分布. 最终通过$\boldsymbol{A}$矩阵与$\boldsymbol{H}$矩阵相乘得到最后使用自注意力机制后的句子级别的嵌入表示$\boldsymbol{M}$, 在得到的嵌入表示中, 对分类有效的信息已经通过注意力机制得到更充分的表达, 因此通过 BiLSTM 编码层可以获得电力文本分类任务更高效的特征表达.

2.5 Softmax 层

将通过自注意力机制进行权重计算后求得的句子级别嵌入表示输入 Softmax 层进行解码输出, 得到每条句子对应每个类别的概率. Softmax 层可以看做是一个单层的全连接神经网络, 当求得电力文本的句子嵌入表示 M 后, 通过矩阵的线性变换后利用 Softmax 函数求得概率分布. 具体计算公式如式 (12), 式 (13) 所示:

$$\boldsymbol{o} = \boldsymbol{W}_s\boldsymbol{M} + \boldsymbol{b}_s \tag{12}$$

$$\hat{y}_j = \frac{\exp(\boldsymbol{o}_j)}{\sum_k \exp(\boldsymbol{o}_k)} \tag{13}$$

其中, $\boldsymbol{W}_s$和$\boldsymbol{b}_s$是可学习的参数, $\boldsymbol{o}$是概率的权重, $\boldsymbol{k}$ 是分类的类别, 通过式 (13) 中的指数幂运算, 不仅可以将可能性较大的类别概率所占权重增大, 还可以保证所有类别的概率相加值为 1, 从而保证$\hat{y}_j$是合理的概率分布. 本文所针对电力客服文本分类任务一共有“报修”“查询”“以往业务”“投诉”“举报”“表扬”“建议”和“反映” 8 类文本类型. 因此 Softmax 层将电力文本的向量表示通过一层神经网络转化为符合概率分布的分类向量, 最终得到该句电力文本概率最大的类别.

3 实验部分

3.1 实验环境和数据介绍

实验使用计算机的系统配置和主要程序版本如下: Linux 操作系统, Python 3.7, PyTorch 1.2 深度学习框架, 16 GB 内存.

本文使用某电网信通公司的标注工单日志作为本实验的数据集进行实验, 文本由“反映, 建议, 表扬, 举报, 投诉, 以往业务, 查询, 报修” 8 种类别的工单日志组成. 首先对文本的预处理, 删除了一些乱码的文本与空文本. 然后经过对数据集进行随机划分与统计, 数据集的规模和文本最长字符数如表 1 所示.

表 1 数据集统计信息

参数	训练集	开发集	测试集
文本条数	2 706	677	846
文本最长字符数	327	455	327

实验所用划分后的数据集种类数量分布如图 5 所示, 其中, train 代表训练集, dev 代表开发集, test 代表测试集.

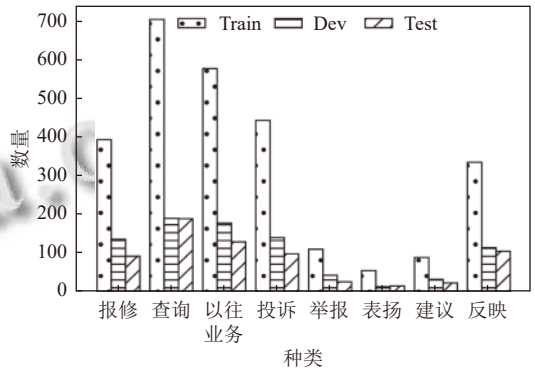

图 5 数据集分布图

如图 5 所示, 数据集经过随机划分后各种类的数目具有大致相同的分布, 训练集中查询种类的文本条目最多, 表扬种类的文本条目最少.

3.2 电力客服文本类型识别指标

本文实验采用的评价指标为类型识别的准确率, 精确率 (P), 召回率 (R) 和 F 值. 对于每个单独的类型来看, 其他类型相对于当前类型均为负类, 其指标的计算公式如式 (14)–式 (16) 所示:

$$P = \frac{T_P}{T_P + F_P} \tag{14}$$

$$R = \frac{T_P}{T_P + F_N} \tag{15}$$

$$F = \frac{2 \times P \times R}{P + R} \tag{16}$$

其中, T_P是当模型判断为正例中判断正确的数目, F_P是当模型判断为正例中判断错误的数目, F_N是模型判断为负例中错误的数目. 将每一个类别的指标相加求平均可得整体的宏平均值, 公式如式 (17)–式 (19) 所示:

$$MacroP = \frac{1}{n}\sum_{i=1}^{n} P_i \tag{17}$$

$$MacroR = \frac{1}{n}\sum_{i=1}^{n} R_i \tag{18}$$

$$MacroF = \frac{1}{n}\sum_{i=1}^{n} F_i \tag{19}$$

其中, P_i, R_i, F_i是每个种类所对应的精确率, 召回率和 F 值. 将所有分类正确的数目相加可得整体的准确率. 公式如式 (20) 所示:

$$accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{20}$$

其中, T_P, F_P和F_N与上文中将某一个类型当做正例时的含义相同, T_N是模型判断为负例中正确的数目, 由于将正例和负例均判断正确的数量之和在一个分类器中为定值, 样本总数不变, 因此对于每个类别求得的准确率均相同, 可认为*accuracy*是模型整体的分类准确率.

3.3 对比实验

为了验证本文所提出模型的有效性, 本文采用了多种模型的对比实验来进行比较. 在电力客服文本的类型识别任务中, 本文使用 BiLSTM 作为基线模型, 其使用复旦大学开源的 Word2Vec 静态预训练词向量[16]作为词嵌入. 本文还对比了使用 Word2Vec 作为词嵌入的 BiGRU 方法[17], BiGRU 融合注意力方法, CNN 方法[18]和 Transformer 方法[12], 对于使用预训练模型的方法, 本文使用基于全词遮盖自监督任务的 BERT-WWM 预训练模型的微调方法 (BW), BERT-WWM 融合卷积神经网络的 BW_CNN 方法和 BERT-WWM 类型识别结果如表 2 所示.

表 2 对比实验结果 (%)

模型	*MacroP*	*MacroR*	*MacroF*	*Acc*
BiLSTM	94.35	93.43	93.82	96.16
BiGRU	95.70	91.19	92.75	94.68
CNN	94.07	92.44	93.22	95.85
BiGRU_ATTN	94.16	94.27	94.21	96.30
Transformer	93.50	95.55	94.32	96.44
BW	97.89	95.68	96.64	97.28
BW_CNN	97.82	97.62	97.70	98.08
BW_BiGRU_ATTN	98.03	97.05	97.52	97.86
BW_BiLSTM_ATTN	99.05	97.87	98.43	98.81

从表 2 的实验结果中可以得出本文提出的模型 BW_BiLSTM_ATTN 在宏平均精确率, 宏平均召回率, 宏平均 F 值和整体准确率上都取得了最好的效果, 分别达到 99.05%, 97.87%, 98.43%, 98.81%. 均高于传统神经网络模型, 其中相比于 BiLSTM 基线模型在 *MacroF* 值上有 4.61% 的提高, 整体准确率较 BiLSTM 有 2.65% 的提升. BERT 模型本身作为在海量数据体量下预训练出的高效模型已经可以取得较好的效果, 通过级联传统神经网络后可以进一步提升对电力文本类型识别的特征提取与编码效果, 相比于 CNN和 BiGRU, 融合具有注意力机制的 BiLSTM 可以更加有效的提取文本的特征, 这是因为 BiLSTM 本身就利用门控单元求得电力文本每个字符输出的隐层状态, 通过使用注意力机制可以让模型自行选择需要的潜在特征从而得到更好的效果. 为了探究不同编码器对 BERT 的促进作用, 本文比较了在中文 BERT-WWM预训练模型上添加不同的编码器所取得的类型识别效果, 如图 6 所示.

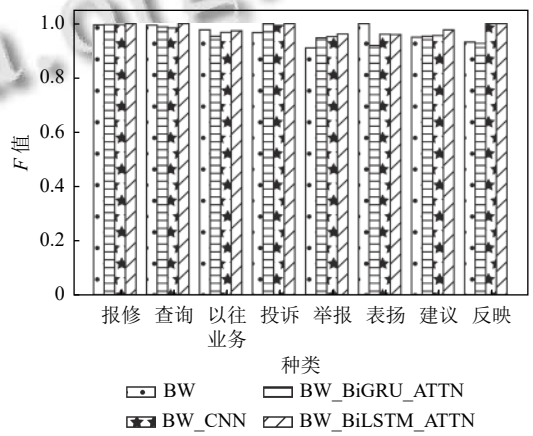

图 6 基于 BERT 的模型效果比较

通过实验统计了每个类别的 F 值作为比较的评价标准, 从图 6 可以看出本文提出的模型相较原生 BERT-WWM 的微调模型在每个类别的 F 值上均有提升, 并且相比于 BiLSTM 和 CNN 作为编码器, 融合注意力的 BiLSTM 模型在大多数类别中都有最好的分类表现. 实验结果表明了本文提出的使用 BERT-WWM 预训练模型作为嵌入层, 利用融合自注意力机制的 BiLSTM 作为编码层, Softmax 层作为解码层的模型在电力客服文本的类型识别任务上具有较好的分类效果.

3.4 消融实验

为了进一步说明本文提出的模型 BW_BiLSTM_ATTN 中每一部分对分类结果的作用, 本节做了多组消融实验来探究每一部分对结果的影响, 分别对比了基线模型 BiLSTM, 去除 BERT-WWM 预训练模型后使用中文 Word2Vec 作为嵌入层的 BiLSTM_ATTN 模型和去除注意力机制的 BW_BiLSTM 模型. 整体的指标比较如表 3 所示.

由表 3 可得去除 BERT-WWM 预训练模型对本文

所提出的模型效果损伤最大, 缺少 BERT 模型本身携带的有效语义信息限制了字符在模型中上下文的语义特征表达, 由于 Word2Vec 属于静态预训练模型, 难以动态为字符分配符合上下文的语义特征表达, 宏平均 *F* 降低了 3.03%, 整体的准确率下降了 1.92%. 去除注意力机制后对于模型捕捉潜在特征的能力也有所损伤, 不能使模型为 BiLSTM 输出的隐状态信息进行自动赋予权重, 从而造成电力文本类型识别时模型对分类信息有效的特征不够突出, 使得相较于原模型宏平均 *F* 下降了 2.29%, 整体的准确率下降了 1.17%.

表 3 消融实验结果 (%)

模型	*MacroP*	*MacroR*	*MacroF*	*Acc*
BiLSTM	94.35	93.43	93.82	96.16
BiLSTM_ATTN	96.21	94.69	95.40	96.89
BW_BiLSTM	97.78	94.72	96.14	97.64
BW_BiLSTM_ATTN	99.05	97.87	98.43	98.81

本文进一步分析在电力客服文本分类的消融实验中每个模型对各个种类文本的分类情况, 实验的结果如图 7 所示.

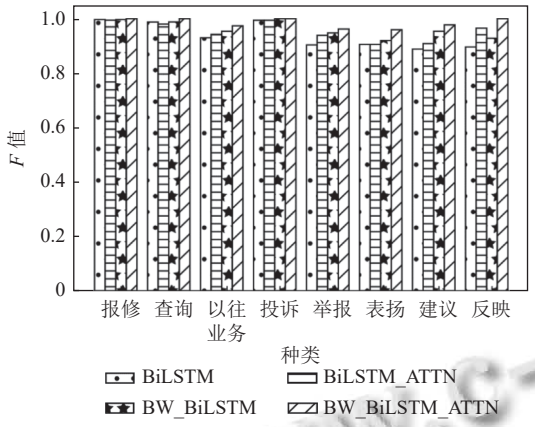

图 7 消融实验对比图

图 7 实验结果表示, 本文提出的模型 BW_BiLSTM_ATTN 与去掉模型中的任意结构的模型相对比, 在每一个类别的文本分类中都表现出最佳的效果. 本文提出的模型在各个种类上的分类结果如表 4 所示.

本文提出的模型在电力文本中的报修, 查询, 投诉, 反映的文本类型中分类指标都达到了 100%, 通过和其对应类别的数据条目对比分析, 以上 4 种分类效果较好的类别均为数据量较为充分的数据类别. 以往业务的种类数量也较多, 但是由于以往业务中的细分种类过于繁杂, 导致模型对其文本结构特征的捕捉能力较低. 模型在对举报, 表扬, 建议这样数据条目较少的类别分类的能力也略低于其他种类, 这是因为神经网络模型需要从大量的数据中捕捉其共同的特征与特点, 当文本特征不明显或者数据量过少, 都会影响模型的学习能力.

表 4 类型分类结果 (%)

种类	*MacroP*	*MacroR*	*MacroF*
报修	100.00	100.00	100.00
查询	100.00	100.00	100.00
以往业务	97.94	96.94	97.43
投诉	100.00	100.00	100.00
举报	94.50	98.10	96.26
表扬	100.00	92.31	96.00
建议	100.00	95.65	97.78
反映	100.00	100.00	100.00

4 结论与展望

本文提出了一种高效的神经网络模型 BW_BiLSTM_ATTN, 提升了对电力客服文本类型识别效果. 模型使用字符级别的向量作为特征, 通过使用 BERT-WWM 预训练模型得到动态的字符语义表达向量, 利用融合注意力机制的 BiLSTM 模型捕捉文本中潜在的语义特征并对文本的特征进行编码从而帮助类型识别, 最后使用 Softmax 层完成对电力客服文本的分类, 得到了优于现有方法的结果. 不过由于数据量有限, 在一些数据量较少的类别上模型的表现稍有下降, 下一步我们将研究在小数量类型上的分类方法. 另一方面, BERT-WWM 模型相较于传统模型的参数量巨大, 如何在保证模型分类质量效果的基础上减少模型参数, 这也是我们下一步研究的方向.

参考文献

1 田世明, 栾文鹏, 张东霞, 等. 能源互联网技术形态与关键技术. 中国电机工程学报, 2015, 35(14): 3482–3494.
2 王芝辉, 王晓东. 基于神经网络的文本分类方法研究. 计算机工程, 2020, 46(3): 11–17.
3 张云翔, 饶竹一. 基于 LSTM 神经网络的电网文本分类方法. 现代计算机, 2020, (2): 8–11. [doi: 10.3969/j.issn.1007-1423.2020.02.002]
4 刘婷婷, 朱文东, 刘广一. 基于深度学习的文本分类研究进展. 电力信息与通信技术, 2018, 16(3): 1–7.
5 杨鹏, 刘扬, 杨青. 基于层次语义理解的电力系统客服工单分类. 计算机应用与软件, 2019, 36(7): 231–235, 321. [doi:

10.3969/j.issn.1000-386x.2019.07.039]
6 朱龙珠, 徐宏, 刘莉莉. 基于深度学习的 95598 重大服务事件识别研究. 电力信息与通信技术, 2018, 16(11): 19–23.
7 刘梓权, 王慧芳, 曹靖, 等. 基于卷积神经网络的电力设备缺陷文本分类模型研究. 电网技术, 2018, 42(2): 644–650.
8 李灿, 田秀霞, 赵波. BiLSTM_DPCNN 模型在电力客服工单数据分类中的应用. 计算机系统应用, 2021, 30(2): 243–249. [doi: 10.15888/j.cnki.csa.007557]
9 肖禹, 王景中, 王宝成. 基于深度学习的中文文本分类方法. 计算机工程与设计, 2021, 42(4): 1014–1019.
10 顾亦然, 霍建霖, 杨海根, 等. 基于 BERT 的电机领域中文命名实体识别方法. 计算机工程, 2021, 47(8): 78–83, 92.
11 Devlin J, Chang MW, Lee K, *et al*. BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 4171–4186.
12 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2017. 6000–6010.
13 Cui YM, Che WX, Liu T, *et al*. Pre-training with whole word masking for Chinese BERT. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 29: 3504–3514.
14 Graves A. Supervised Sequence Labelling with Recurrent Neural Networks. Berlin, Heidelberg: Springer, 2012. 37–45.
15 Lin ZH, Feng MW, dos Santos CN, *et al*. A structured self-attentive sentence embedding. ICLR 2017. Toulon: OpenReview.net, 2017.
16 Yan H, Qiu XP, Huang XJ. A graph-based model for joint Chinese word segmentation and dependency parsing. Transactions of the Association for Computational Linguistics, Volume 8. Cambridge: MIT Press, 2020. 78–92.
17 梁志剑, 谢红宇, 安卫钢. 基于 BiGRU 和贝叶斯分类器的文本分类. 计算机工程与设计, 2020, 41(2): 381–385.
18 Zhang X, Zhao JB, LeCun Y. Character-level convolutional networks for text classification. Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2015. 649–657.

(校对责编: 孙君艳)