

基于 ALBERT 与双向 GRU 的中医脏腑定位模型

张德政^{1,2)}, 范欣欣^{1,2)}, 谢永红^{1,2)}✉, 蒋彦钊^{1,2)}

1) 北京科技大学计算机与通信工程学院, 北京 100083 2) 材料领域知识工程北京市重点实验室, 北京 100083

✉通信作者, E-mail: xieyh@ustb.edu.cn

摘 要 脏腑定位, 即明确病变所在的脏腑, 是中医脏腑辨证的重要阶段. 本文旨在通过神经网络模型搭建中医脏腑定位模型, 输入症状文本信息, 输出对应的病变脏腑标签, 为实现中医辅助诊疗的脏腑辨证提供支持. 将中医的脏腑定位问题建模为自然语言处理中的多标签文本分类问题, 基于中医的医案数据, 提出一种基于预训练模型 ALBERT 和双向门控循环单元 (Bi-GRU) 的脏腑定位模型. 对比实验和消融实验的结果表明, 本文提出的方法在中医脏腑定位的问题上相比于多层感知机模型、决策树模型具有更高的准确性, 与 Word2Vec 文本表示方法相比, 本文使用的 ALBERT 预训练模型的文本表示方法有效提升了模型的准确率. 在模型参数上, ALBERT 预训练模型相比 BERT 模型降低了模型参数量, 有效减小了模型大小. 最终, 本文提出的脏腑定位模型在测试集上 F1 值达到了 0.8013.

关键词 多标签文本分类; ALBERT; 门控循环单元; 脏腑定位; 中医

分类号 TP391.1

Localization model of traditional Chinese medicine Zang-fu based on ALBERT and Bi-GRU

ZHANG De-zheng^{1,2)}, FAN Xin-xin^{1,2)}, XIE Yong-hong^{1,2)}✉, JIANG Yan-zhao^{1,2)}

1) School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

2) Beijing Key Laboratory of Knowledge Engineering for Materials Science, Beijing 100083, China

✉ Corresponding author, E-mail: xieyh@ustb.edu.cn

ABSTRACT The rapid development of artificial intelligence (AI) has injected new vitality into various industries and provided new ideas for the development of traditional Chinese medicine (TCM). The combination of AI and TCM provides more technical support for TCM auxiliary diagnosis and treatment. In the history of TCM, many methods of syndrome differentiation have been observed, among which the differentiation of Zang-fu organs is one of the important methods. The purpose of this paper is to provide support for the localization of Zang-fu in TCM through AI technology. Localization of Zang-fu organs is a method of determining the location of lesions in such organs and is an important stage in the differentiation of Zang-fu organs in TCM. In this paper, the localization model of TCM Zang-fu organs through the neural network model was established. Through the input of symptom text information, the corresponding Zang-fu label for a lesion could be output to provide support for the realization of Zang-fu syndrome differentiation in TCM-assisted diagnosis and treatment. In this paper, the localization of Zang-fu organs was abstracted as multi-label text classification in natural language processing. Using the medical record data of TCM, a Zang-fu localization model based on pretraining models a lite BERT (ALBERT) and bidirectional gated recurrent unit (Bi-GRU) was proposed. Comparison and ablation experiments finally show that the proposed method is more accurate than multilayer perceptron and the decision tree. Moreover, using an ALBERT pretraining model for text representation effectively improves the accuracy of the localization model. In terms of model parameters, the ALBERT

收稿日期: 2021-01-13

基金项目: 国家重点研发计划云计算和大数据专项资助项目(2017YFB1002304)

pretraining model greatly reduces the number of model parameters compared with the BERT model and effectively reduces the model size. Finally, the F1-value of the Zang-fu localization model proposed in this paper reaches 0.8013 on the test set, which provided certain support for the TCM auxiliary diagnosis and treatment.

KEY WORDS multi-label text classification; ALBERT; GRU; localization of Zang-fu; traditional Chinese medicine (TCM)

人工智能技术的快速发展为各个行业注入了新的活力,也为中医学的发展提供了新的思路。随着人工智能与中医的结合,也为中医的辅助诊疗提供了更多可能。中医学在历史上形成了多种辨证方法,其中脏腑辨证是根据脏腑的生理功能,病理表现,对疾病证候进行归纳,借以推究病机,判断病变的部位、性质、正邪盛衰情况的一种辨证方法,是辨证体系中的重要组成部分。当某一脏腑发生病变时,反映出的临床症状也各不相同,且依据脏腑辨证特点,一个症状可涉及多个脏腑,所以本文将中医中的脏腑定位问题抽象为自然语言处理领域的多标签文本分类问题,将医案数据中的症状类文本信息作为输入,将具体病变脏腑作为多标签文本分类问题中的标签。

1 研究现状

在中医的辨证推理方面,已经有大量研究人员进行过相关研究。许强^[1]通过对经典证素辨证体系相关内容的总结,分析经典证素辨证体系存在的不足,并基于有向图这种数据结构重新对证素概念进行定义,对证素内容进行分类,并重新对辨证医理模型进行构建。尹丹等^[2]采用领域本体七步法构建中医经方知识图谱,并通过 Cypher 设计一种模拟人联想思维进行检索的经方知识检索框架。刘超等^[3]基于 BP 神经网络方法,建立冠状动脉临界病变患者证候要素及其常见组合的中医辨证诊断模型。褚娜^[4]在中医辨证的现状研究和相关智能算法的基础上,提出适合于中医辨证的理论方法和系统实施方案。杨开明^[5]将 85 例糖尿病患者结合 C4.5 决策树算法建立了中医辨证决策树,提取出糖尿病的 8 种中医证候分类规则。周璐等^[6]将 C4.5 决策树、随机森林、支持向量机、BP 神经网络算法进行融合,最终得到一种复合结构的智能化辨证选方模型。舒鑫等^[7]基于神经网络构建了 AIS 气虚证预测模型,并展现出了良好的准确率。在此类方法中,尚未充分使用运用文本的高层语义信息,可以尝试在此类方法基础上融入高层语义信息进行效果提升。

自然语言处理技术的发展为中医的辨证论治

提供了新的方法,多标签文本分类即为其中的重要方法。多标签文本分类是指一个样本可以属于多个类别(或标签)。比如一条症状同时与“肾”和“肝”两个标签关联,即该症状表现与“肾”和“肝”两个脏腑相关。有研究人员从聚类的角度对文本进行多标签分类,申超波等^[8]提出了一种基于标签聚类的标签幂集方法,通过改进平衡 k -means 聚类来发现训练集中潜在的重要标签集合,并用于形成新的训练集进行多标签分类。黄志强^[9]对基于 k -means 的多标签分类及标签补全算法进行了研究。

在基于神经网络对文本特征进行提取表示方面,研究人员也进行了大量研究。李德玉等^[10]提出一种基于标签特征的卷积神经网络(CNN)情绪多标签分类方法,使用 Word2Vec 进行词向量表示,然后利用标签特征来强化文本情绪和标签之间的联系,融合入 CNN 模型对文本情绪进行深层次的表示和多标签分类。Joulin 等^[11]基于浅层神经网络设计出了文本分类工具 FastText。易士翔等^[12]在中文突发事件数据集上验证了双向 LSTM 模型在文本表征中的有效性。Chen 等^[13]在获得文本的整体语义(Text feature vector)后,将文本整体语义输入到一个递归神经网络(RNN)的序列中作为初始值,每一时刻输入是上一时刻的输出,将卷积神经网络和递归神经网络进行集成应用,以获得全局和局部文本语义。Yogatama 等^[14]针对文本分类问题构建了基于 LSTM(Long short-term memory,长短期记忆网络)的生成模型和判别模型。Wang^[15]提出了中断递归神经网络(DRNN),将位置不变性引入 RNN 中。该网络通过限制 RNN 中的信息流的距离,将每个时间步长的隐藏状态限制为表示当前位置附近的单词。Kim^[16]提出 TextCNN,将卷积神经网络应用到文本分类任务,利用多个不同大小的 kernel 提取句子中的关键信息,这种方法类似于多窗口大小的 n -gram,能够更好地捕捉局部相关性。

在对多标签文本分类效果的提升过程中,输入高质量的文本表征对模型效果有巨大的作用。基于大规模语料库的预训练模型已经成为自然语

言处理技术的重要研究方向. 这些预训练模型是以庞大的语料库作为基础训练的, 因此相较于 onehot、Word2Vec^[17]、Glove^[18] 等文本表征方式, 预训练模型具有更好的表征能力, 可以更好地适用于不同场景下的自然语言处理任务. 为了更多地保留句子中词与词之间的内在联系, Google 提出 Transformer 编码器模型^[19], 在此基础上 Google 提出了 BERT 模型^[20](Bidirectional encoder representations from transformers), 利用了随机遮盖语言模型 (Masked LM) 和预测下一句 (Next sentence prediction, NSP) 2 种方法分别捕捉词语和句子级别的表示. 在 BERT 出现后, 又相继出现了很多在 BERT 基础上进一步改进和提升的模型, 如 XLNet^[21], RoBERTa^[22], DistilBERT^[23] 等.

雷景生和钱叶^[24]在公开数据集上进行了实验, 验证了使用 ERNIE 预训练模型进行文本分类的有效性. 本文使用的 A lite BERT 模型 (ALBERT)^[25]是 Google 在 BERT 之后提出一种轻量级 BERT, 相比于 BERT, ALBERT 模型采用解绑输入的词向量大小和隐藏层大小关系的方式降低了模型参数量, 并实现了模型跨层的参数共享, 在不损失模型效果的前提下有效缩小了模型大小.

2 模型及算法描述

2.1 模型说明

本文提出的基于 ALBERT 和双向门控循环单元 (Bidirectional gated recurrent unit, Bi-GRU) 的多标签文本分类模型整体结果如图 1 所示, 整个模型可以分为 4 个部分: 输入层, 预训练层, Bi-GRU 层和输出层.

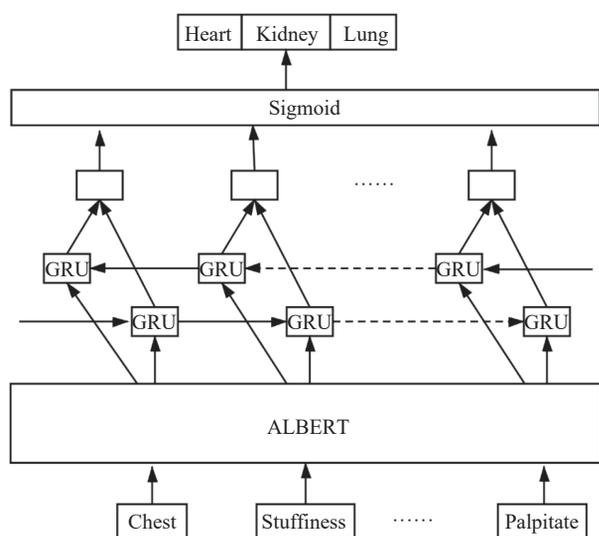


图 1 脏腑定位模型结构
Fig.1 Zang-fu localization model structure

输入层将症状文本进行预处理后输入模型中. 预训练层将输入的文本通过 ALBERT 预训练语言模型获得文本的向量表征.

Bi-GRU 层接收经过 ALBERT 预训练模型后的向量序列, 通过 Bi-GRU 进一步编码获取高层语义.

输出层通过以 Sigmoid 为激活函数的全连接层将各标签对应输出值压缩至 0~1 之间, 并将输出值大于等于设定阈值的所有病位标签输出作为最终脏腑定位结果.

2.2 ALBERT 预训练模型

自 BERT 的成功以来, 预训练模型都采用了很大的参数量以取得更好的模型表现. 但是模型参数量越来越大也带来了很多问题, 比如对算力要求越来越高、模型体积越来越大、模型训练时间越来越长等. 为了解决目前预训练模型参数量过大的问题, Lan 等^[25]提出了 ALBERT 模型, 相比于 BERT, ALBERT 使用了 2 种能够大幅减少预训练模型参数量的方法, 并提出用语句顺序预测 (Sentence order prediction, SOP) 任务代替 BERT 中的 NSP 任务. ALBERT 是一种轻量级的基于 Transformer 的双向编码器表示模型 (A lite bidirectional encoder representations from transformers). 通过对部分语义单元的随机掩码, 使得模型能够学习到完整的语义表示. ALBERT 模型结构如图 2 所示, 其中 $[E_1, E_2, \dots, E_{n-1}, E_n]$ 为原始文本向量, trm 为 Transformer 模型^[19].

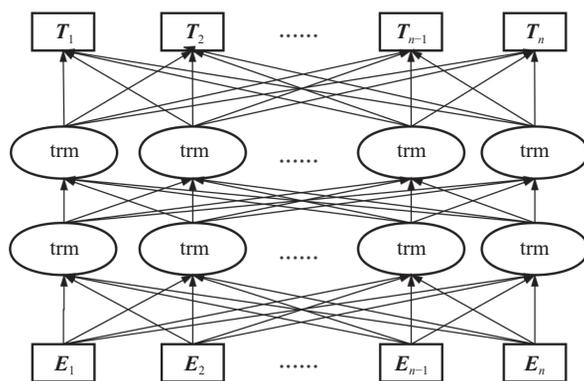


图 2 ALBERT 模型结构
Fig.2 ALBERT model structure

由图 2 可以看出, 模型的输出包含文本上下文信息的向量表示, 且每个向量 $[T_1, T_2, \dots, T_{n-1}, T_n]$ 都包含了整个序列的文本信息.

在 ALBERT 中, 提出了 2 种可以减少 BERT 参数量的方法: (1) 解绑 WordPiece 词向量大小 E 和隐层大小 H , 减小参数量; (2) 层之间参数共享, 即在模型多个层之间使用相同的参数. 第一种方法

对词嵌入参数进行了因式分解, 将它们分解为两个小矩阵. 不再将 One-hot 向量直接映射到大小为 H 的隐藏空间, 而是先将它们映射到一个低维词嵌入空间 E , 再映射到隐藏空间. 通过这种分解, 将词嵌入参数从 $o(V \times H)$ 降低到 $o(V \times E + E \times H)$, 在隐层大小 H 远大于 E 时, 会有非常明显的参数量减少. 第 2 种方法通过跨层参数共享, 有效降低了参数量, 并且通过参数共享, 提升了模型的鲁棒性.

此外, 对于 BERT 中的 NSP 任务, ALBERT 进行了改进. 由于 BERT 中 NSP 任务是预测两个句子在原文本中是否连续出现的二分类任务, 其正例为文章中连续出现的两个句子, 负例由两篇文档中各选一句构造而成. 由于负例的构造使得模型可能更多倾向于预测句子主题关系而非句子连续性, 这将 NSP 任务简单化. 所以 ALBERT 提出了句子顺序预测任务(SOP), 即正例选择方式不变, 与 NSP 任务中的正例选择方式相同, 负例选择一篇文档中两个连续的句子并将它们顺序交换后构造得到. 在 NSP 和 SOP 有效性对比实验中, NSP 在解决 SOP 问题时精度仅为 52%, 而 SOP 在 NSP 任务上精度为 78.9%, 且 SOP 任务在几乎所有下游任务中都相比 NSP 有提升^[25], 基于此证明了 SOP 相比于 NSP 的有效性.

经过上述处理, 在大规模语料上训练得到的 ALBERT 预训练模型可以凭借较少的模型参数表达丰富的语义信息, 有效缩减了模型体积.

2.3 Bi-GRU 层

门控循环单元是 LSTM 的变种, 是相比于 LSTM 网络更简单的循环神经网络^[26].

GRU 网络通过引入门控机制来进行信息的更新. 在 LSTM 网络中, 输入门和遗忘门是互补关系, 直接使用两个门比较冗余. GRU 将输入门与和遗忘门合并成更新门(Update gate), 并引入了重置门(Reset gate). 同时, GRU 也不引入额外的记忆单元, 直接在当前状态 h_t 和历史状态 h_{t-1} 之间引入线性依赖关系.

GRU 模型更新方式如图 3 所示.

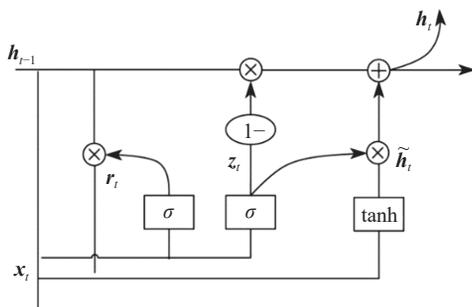


图 3 GRU 单元
Fig.3 GRU unit

按照流程, 可将图 3 中的 GRU 分成 4 部分: 更新门、重置门、产生新记忆 \tilde{h}_t 和产生新隐藏层状态 h_t , 下面将具体说明.

更新门: 用过去的隐藏层状态与当前输入决定如何产生新记忆.

重置门: 用过去的隐藏层状态与当前输入决定要有多少之前的记忆被遗忘.

产生新记忆 \tilde{h}_t : 用过去的隐藏层状态与当前输入决定当前产生的新记忆, 并用重置门控制过去的隐藏层状态信息流入新记忆中.

产生新隐藏层状态 h_t : 用更新门从新记忆 \tilde{h}_t 与之前的隐藏层状态 h_{t-1} 中得到新的隐藏层状态.

上述流程描述, 可以得到如下 4 个公式, 其中包括重置门向量、更新门向量、新记忆和经过门控处理后的新的记忆(新的隐藏层状态 \tilde{h}_t):

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (1)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (2)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t] + b_h) \quad (3)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (4)$$

式中, r_t 为 t 时刻的重置门, z_t 表示 t 时刻的更新门, \tilde{h}_t 表示 t 时刻的候选激活状态, h_t 表示 t 时刻的激活状态, h_{t-1} 表示 $(t-1)$ 时刻的隐藏层状态, $[W_r, W_z, W_h]$ 表示各部分对应的权重系数矩阵, (b_r, b_z, b_h) 为偏移向量, σ 表示 sigmoid 激活函数, $*$ 表示哈达玛积(Hadamard product). 相较于 RNN 模型, GRU 能够较好地解决梯度消失问题. 而与 LSTM 相比, GRU 模型更为简单, 需要的参数更少, 而且不容易产生过拟合现象.

GRU 作为 RNN 的变种同 RNN 一样, 均具有遗忘性. 通常来说, 在训练过程中, 最后一个状态所包含的信息是有损失的, 而且越靠前, 损失越严重. 对于包含很多序列的训练任务而言, 获取前后背景信息对增加模型准确度来说有重要意义. 因此, 本文通过训练正向学习和反向学习两个 GRU 神经网络, 能够把正向和反向学习过程中最后状态对应的向量进行连接, 将连接后得到的向量作为最后的结果, 这个网络就可以利用前后所有信息. 本文中使用的双向 GRU 模型如图 4 所示.

2.4 输出层

依据中医一个症状涉及多个脏腑的特点, 该模型的输出为多标签分类输出而非多分类输出. 多标签分类与多分类的区别在于, 多标签分类中一个样本可以同时属于多个类别, 而多分类中, 一个样本属于且只属于一个类别. 如症状“胸闷, 气

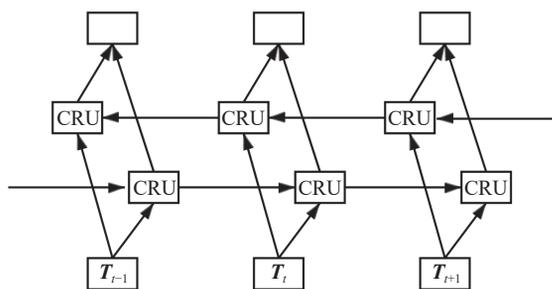


图 4 双向 GRU 模型示意图

Fig.4 Bi-GRU model diagram

短, 头晕, 头痛, 夜寐不宁, 舌暗, 苔薄, 苔白, 苔腻, 脉弦, 脉细, 脉沉”同时发生在“心, 肾”两个病位, 即同时属于“心”, “肾”两个类别, 所以该问题为多标签分类问题。

基于脏腑定位模型具有多标签分类的特点, 在输出层使用 Sigmoid 函数作为激活函数预测各个标签的概率, 区别于单标签模型常用的 Softmax 函数。Sigmoid 函数是一类 S 型曲线函数, 为两端饱和函数。可以看成是一个“挤压”函数, 把一个实数域的输入“挤压”到 (0, 1) 之间。当输入值在 0 附近时, Sigmoid 函数近似为线性函数; 当输入值靠近两端时, 对输入进行抑制。输入越小, 越接近于 0, 输入越大, 越接近于 1。

在模型训练过程中, 损失函数调整为带有 Sigmoid 函数的二元交叉熵损失函数, 样本损失计算方式如式 (5) 所示:

$$\text{Loss} = -\frac{1}{n} \sum_{i=1}^n y_i \cdot \ln \hat{y}_i + (1 - y_i) \cdot \ln(1 - \hat{y}_i) \quad (5)$$

式中, n 为类别总数, \hat{y}_i 是模型的输出值, 表示预测

样本属于某一类别的概率; y_i 是样本在某一类别下的真实标签。区别于 Softmax 交叉熵损失, Sigmoid 函数使样本属于各类别的概率分布在 (0, 1) 之间。

在训练阶段结束使用模型进行预测推理时, 设定一个阈值, 将概率大于阈值的所有标签输出, 作为多标签分类的预测结果。如输入症状“胸闷, 气短, 头晕, 头痛, 夜寐不宁, 舌暗, 苔薄, 苔白, 苔腻, 脉弦, 脉细, 脉沉”后, 输出各个标签概率为 [0.64240265, 0.6062219, 0.04209191, ..., 0.17620572], 分别对应 [‘心’, ‘肾’, ‘肺’, ..., ‘脾’], 通过网格调参的方法, 最终设定阈值为 0.5, 即大于等于 0.5 均为相关病位, 在上述例子中, 即“心”, “肾”为输入症状的对应病位标签。

3 实验及结果分析

3.1 实验数据

实验数据来源于名老中医医案库, 在经过专家指导后共筛选出 14821 份医案, 本文中需要用到其中的“症状”和“病位”信息。其中“症状”指疾病过程中机体内的一系列机能、代谢和形态结构异常变化所引起的病人主观上的异常感觉或某些客观病态改变; “病位”指与表现出的症状相关的部位, 即脏腑标签。在这些医案中存在症状信息不全、症状文本中存在无意义字符的情况, 通过专家指导和设置相应规则, 对症状文本中去噪处理和筛选后, 得到最终有效医案数据。将这些医案中“症状”信息和“病位”信息进行提取, 得到本实验中所用到的全部数据, 共计 12735 份。所得数据格式如表 1 所示。

表 1 脏腑定位数据格式

Table 1 Zang-fu location data format

No.	Symptoms	Tag
1	Legs ache, and wake up unable to sleep, along with hemoptysis and a sore throat	spleen, kidney, heart
2	The patient had high blood pressure, weakness in the right limb, and pain in the left upper arm	liver, kidney

3.2 实验参数设置

具体实验参数设置如表 2 所示。

3.3 评价标准

为了综合考虑模型的性能, 本文采用了精确率 (Precision)、召回率 (Recall)、F1 值 (F1-Value) 作为模型评价指标。

精确率, P : 指在被所有预测为正例中实际为正例的概率。

召回率, R : 指在所有实际为正例中被预测为正

例的概率。

F1 值: 指精确值和召回率的调和均值。

计算表达式如下:

$$P = \frac{\sum_i TP_i}{\sum_i TP_i + FP_i} \quad (6)$$

$$R = \frac{\sum_i TP_i}{\sum_i TP_i + FN_i} \quad (7)$$

$$F1 = 2 \frac{P \cdot R}{P + R} \quad (8)$$

表 2 训练过程中的参数

Table 2 Parameters in the training process

Parameter name	Parameter value
Max_seq_lenth	128
GRU_units	128
Dropout	0.4
Learning_rate	1×10^{-4}
Epochs	10
Batch_size	128

式中, TP_i 表示将类别 i 中正例预测为正例的数目; TN_i 表示将类别 i 中负例预测为负例的数目; FP_i 表示将类别 i 中负例预测为正例的数目; FN_i 表示将类别 i 中正例预测为负例的数目。

3.4 实验对比与分析

为了验证本文提出的脏腑定位模型对于实验的有效性, 采用多种模型与本文模型进行对比, 如表 3 和表 4 所示, 并为了验证本文提出模型的有效性进行了消融实验, 如表 5 所示。

表 3 多标签分类对比实验结果

Table 3 Comparative experimental results of multiple label classification

No.	Method	Precision	Recall	F1-value
1	Word2Vec+Bi-GRU	0.8015	0.7653	0.7830
2	MLP Classifier	0.7091	0.7067	0.7079
3	Decision Tree Classifier	0.6744	0.6633	0.6688
4	ALBERT+Bi-GRU	0.8301	0.7745	0.8013

表 4 BERT 与 ALBERT 对比实验结果

Table 4 Comparative experimental results of BERT and ALBERT

Id	Method	Precision	Recall	F1-value	Time/s	Model_ parameters/ MB
1	BERT+Bi-GRU	0.8253	0.7783	0.8011	99.8219	363.3
2	ALBERT+Bi-GRU	0.8301	0.7745	0.8013	84.7045	37.3

表 5 多标签分类消融实验结果

Table 5 Ablation experiment multiple label classification results

Method	Precision	Recall	F1-value
ALBERT	0.7711	0.7315	0.7508
ALBERT+Bi-GRU	0.8301	0.7745	0.8013

在表 3 对比实验中可以看出, 方法 1 与方法 4 相比, 在其他条件完全相同的情况下, 本文提出

的基于 ALBERT 预训练模型相比于 Word2Vec 的表示方式, 能更好地完成文本表征任务, F1 值相比提升了 0.0183. 将方法 4 与方法 2、方法 3 相比, 本文提出的模型比多层感知机 (MLP classifier, No.2) 模型和决策树 (Decision tree classifier, No.3) 方法 F1 值分别提升了 0.0934 和 0.1325, 均有大幅提升. 表 4 的对比实验结果表明, 相比于使用 BERT 预训练模型, 采用 ALBERT 预训练模型的方法 F1 值提升了 0.0002, 且在测试集上预测时间有了明显减少, 模型参数大幅减少, 模型体积缩小为原来的近十分之一。

通过表 5 消融实验可以看出, 使用 ALBERT 预训练模型后不增加 Bi-GRU 层, 最终在测试集上的 F1 值为 0.7508. 与之相比, 在 ALBERT 层后加入 Bi-GRU 层的模型 F1 值提升了 0.0505. 通过该消融实验, 可以得出在本文的脏腑定位模型中加入 Bi-GRU 层确实有效捕捉了上下文信息, 生成了高层语义表示, 有效提升模型效果。

4 总结与展望

针对中医中的脏腑定位方法, 本文提出了一种基于 ALBERT 和 Bi-GRU 的多标签文本分类模型. 该模型采用 ALBERT 表示作为输入, 通过 Bi-GRU 有效地捕捉上下文信息生成高层次的文本语义信息表示, 之后再对高层次信息表示进行分类. 该方法从医案数据的角度对中医的脏腑定位方法的探索, 相较于决策树、感知机等模型 F1 值有较大的提升. 使用的 ALBERT 向量化表示方式与 Word2Vec 相比有很大的提升, 相比于 BERT 方法, 一定程度上缩短了预测时间, 并大幅缩减了模型体积. 在未来的工作中, 将在进一步提升模型效果的基础上, 考虑融入更多中医中脏腑定位的知识, 将知识图谱中的知识融入到脏腑定位模型中, 进行融合知识的脏腑辨证. 探索在中医知识的约束下更准确地进行中医证素的推理工作, 在提高模型效果的同时也为模型加入更多的可解释性。

参 考 文 献

- [1] Xu Q. Mining the Syndrome Factor Distribution of AECOPD by the Attribution Model Built by Directed Graph [Dissertation]. Chengdu: Chengdu University of TCM, 2017
(许强. 基于有向图的证素归因模型挖掘 AECOPD 的证素分布规律[学位论文]. 成都: 成都中医药大学, 2017)
- [2] Yin D, Zhou L, Zhou Y M, et al. Study on design of graph search

- pattern of knowledge graph of TCM classic prescriptions. *Chin J Inf Tradit Chin Med*, 2019, 26(8): 94
(尹丹, 周璐, 周雨玫, 等. 中医经方知识图谱“图搜索模式”设计研究. 中国中医药信息杂志, 2019, 26(8): 94)
- [3] Liu C, Gao J L, Dong Y, et al. Study on TCM syndrome differentiation and diagnosis model based on BP neural network for syndrome elements and their common combinations in patients with borderline coronary lesion. *Chin J Inf Tradit Chin Med*, 2021, 28(3): 104
(刘超, 高嘉良, 董艳, 等. 基于BP神经网络的冠状动脉临界病变患者证候要素及其常见组合中医辨证诊断模型研究. 中国中医药信息杂志, 2021, 28(3): 104)
- [4] Chu N. *Research on Hybrid Intelligent Based Syndrome Differentiation System for Traditional Chinese Medicine* [Dissertation]. Shanghai: Shanghai Jiaotong University, 2012
(褚娜. 基于混合智能的中医辨证系统研究[学位论文]. 上海: 上海交通大学, 2012)
- [5] Yang K M. *Research on Clinical Data Mining Technology of Diabetes TCM* [Dissertation]. Kunming: Kunming University of Science and Technology, 2013
(杨开明. 糖尿病中医临床数据挖掘技术研究[学位论文]. 昆明: 昆明理工大学, 2013)
- [6] Zhou L, Li G G, Sun Y, et al. Construction of intelligent syndrome differentiation and formula selection of compound structure model. *World Chin Med*, 2018, 13(2): 479
(周璐, 李光庚, 孙燕, 等. 复合结构智能化辨证选方模型的构建. 世界中医药, 2018, 13(2): 479)
- [7] Shu X, Cao Y, Huang X, et al. Construction of prediction model of qi deficiency syndrome in acute ischemic stroke based on neural network analysis technique. *Glob Tradit Chin Med*, 2019, 12(11): 1650
(舒鑫, 曹云, 黄幸, 等. 基于神经网络分析技术的急性缺血性卒中气虚证预测模型构建的研究. 环球中医药, 2019, 12(11): 1650)
- [8] Shen C B, Wang Z H, Sun Y G. A multi-label classification algorithm based on label clustering. *Comput Eng Softw*, 2014, 35(8): 16
(申超波, 王志海, 孙艳歌. 基于标签聚类的多标签分类算法. 软件, 2014, 35(8): 16)
- [9] Huang Z Q. *Multi-Label Classification and Label Completion Algorithm Based on K-Means* [Dissertation]. Anqing: Anqing Normal University, 2020
(黄志强. 基于K-means的多标签分类及标签补全算法[学位论文]. 安庆: 安庆师范大学, 2020)
- [10] Li D Y, Luo F, Wang S G. A multi-label emotion classification method for Chinese text based on CNN and tag features. *J Shanxi Univ Nat Sci Ed*, 2020, 43(1): 65
(李德玉, 罗锋, 王素格. 融合CNN和标签特征的中文文本情绪多标签分类. 山西大学学报(自然科学版), 2020, 43(1): 65)
- [11] Joulin A, Grave E, Bojanowski P, et al. Bag of Tricks for Efficient Text Classification // *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Valencia, 2017: 427
- [12] Yi S X, Yin H P, Zheng H Y. Public security event trigger identification based on Bidirectional LSTM. *Chin J Eng*, 2019, 41(9): 1201
(易士翔, 尹宏鹏, 郑恒毅. 基于BiLSTM的公共安全事件触发词识别. 工程科学学报, 2019, 41(9): 1201)
- [13] Chen G B, Ye D H, Xing Z C, et al. Ensemble application of convolutional and recurrent neural networks for multi-label text categorization // *2017 International Joint Conference on Neural Networks (IJCNN)*. Anchorage, 2017: 2377
- [14] Yogatama D, Dyer C, Ling W, et al. Generative and discriminative text classification with recurrent neural networks[J/OL]. *ArXiv Preprint* (2017-03-06) [2020-12-29]. <https://arxiv.org/abs/1703.01898v1>
- [15] Wang B X. Disconnected Recurrent Neural Networks for Text Categorization // *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. Melbourne, 2018: 2311
- [16] Kim Y. Convolutional Neural Networks for Sentence Classification // *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, 2014: 1746
- [17] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[J/OL]. *arXiv preprint* (2013-10-16) [2021-5-22]. <https://arxiv.org/abs/1310.4546>
- [18] Pennington J, Socher R, Manning C. Glove: Global Vectors for Word Representation // *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, 2014: 1532
- [19] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J/OL]. *arXiv preprint* (2017-6-12) [2021-5-22]. <https://arxiv.org/abs/1706.03762>
- [20] Devlin J, Chang M W, 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*. Minneapolis, Minnesota, 2018: 4171
- [21] Yang Z L, Dai Z H, Yang Y M, et al. Xlnet: Generalized autoregressive pretraining for language understanding[J/OL].

- arXiv preprint* (2019-6-19) [2021-5-23]. <https://arxiv.org/abs/1906.08237>
- [22] Liu Y, Ott M, Goyal N, et al. Roberta: A robustly optimized bert pretraining approach[J/OL]. *arXiv preprint* (2019-07-26) [2020-12-29]. <http://arxiv.org/abs/1907.11692>
- [23] Sanh V, Debut L, Chaumond J, et al. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter[J/OL]. *arXiv preprint* (2019-10-02) [2020-12-29]. <http://arxiv.org/abs/1910.01108>
- [24] Lei J S, Qian Y. Chinese-text classification method based on ERNIE-BiGRU. *J Shanghai Univ Electr Power*, 2020, 36(4): 329
(雷景生, 钱叶. 基于ERNIE-BiGRU模型的中文文本分类方法. 上海电力大学学报, 2020, 36(4): 329)
- [25] Lan Z Z, Chen M, Goodman S, et al. ALBERT: A lite BERT for self-supervised learning of language representations. // *ICLR 2020 : Eighth International Conference on Learning Representations*. Addis Ababa, 2020
- [26] Chung J, Gulcehre C, Cho K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J/OL]. *ArXiv Preprint* (2018-08-13) [2020-12-29]. <http://arxiv.org/abs/1412.3555>