

# 基于 BERT-BiGRU 模型的智慧医疗问答系统

黄 涌, 蒯娟霞, 关成斌

(广东东软学院信息管理与工程学院, 广东 佛山 528000)

✉ 1427217679@qq.com; xijuanxia@nuit.edu.cn; guanchengbin@nuit.edu.cn



**摘 要:**针对医疗问答系统在处理复杂问题时面临上下文语义理解的局限,提出一种基于 BERT-BiGRU 的模型。通过预训练语言模型 BERT 和双向门控循环单元 BiGRU 建立医疗问答系统,其中 BERT 提取文本语义特征, BiGRU 学习文本的顺序依赖信息,进而全面表示文本语义结构信息。在 CBLUE 医疗问答数据集上与基准方法相比,该模型在意图识别任务上的精确率提高到 79.22%,召回率提高到 81.23%,F1 值(精确率和召回率的调和平均值)提高到 79.82%。研究表明,结合 BERT 和 BiGRU 的模型可以更好地理解医疗问句的语义和结构信息,显著地提升了医疗问答系统的性能。

**关键词:**BERT; BiGRU; 知识图谱; 意图识别; 医疗问答系统

**中图分类号:**TP391.1 **文献标志码:**A

## An Intelligent Medical Q&A System Based on BERT-BiGRU Model

HUANG Yong, XI Juanxia, GUAN Chengbin

(School of Information Management and Engineering, Guangdong Neusoft University, Foshan 528000, China)

✉ 1427217679@qq.com; xijuanxia@nuit.edu.cn; guanchengbin@nuit.edu.cn

**Abstract:** This paper proposes a model based on BERT-BiGRU to address the limitations of contextual semantic understanding in medical question answering (Q&A) systems when dealing with complex problems. A pre-trained language model BERT and a Bidirectional Gated Recurrent Unit (BiGRU) are used to establish a medical Q&A system, where BERT extracts semantic features of the text, while BiGRU learns the order dependency information of the text, thereby comprehensively representing the semantic structure information of the text. Compared with the benchmark method on the CBLUE medical Q&A dataset, the proposed model achieves an accuracy improvement of 79.22%, a recall improvement of 81.23%, and an F1 value (harmonic mean of accuracy and recall) improvement of 79.82% in intention recognition tasks. Research has shown that combining BERT and BiGRU models can better understand the semantic and structural information of medical questions, significantly improving the performance of medical Q&A systems.

**Key words:** BERT; BiGRU; knowledge mapping; intention recognition; medical Q&A system

## 0 引言(Introduction)

智慧医疗已成为缓解医疗资源短缺的重要手段,智慧医疗通过人工智能<sup>[1]</sup>、大数据<sup>[2]</sup>、云计算等技术手段对医疗信息进行深度挖掘和分析,可以提高医疗服务的效率和质量,并缓解医疗资源短缺的问题。知识图谱作为智慧医疗的核心技术之

一,能够将不同来源、不同结构、散乱复杂的医疗数据联合起来<sup>[3]</sup>,构建一个结构化和标准化的医疗知识网络。

与前人研究不同,本研究构建了基于 BERT 和 BiGRU 的智慧医疗问答系统。该系统充分利用 BERT 学习语义特征,利用 BiGRU 学习序列特征的优势,可以更准确地表示文本语义



类层,以输出最终的多分类判断结果。

## 2.2 BERT-BiGRU 模型的结构

本研究提出了一种基于 BERT-BiGRU 的医疗意图识别方法。BERT 是一个具有 Transformer 结构的预训练语言表示模型,它利用双向注意力机制学习输入序列的上下文语义信息。BERT 可以动态生成输入词的向量表示,更好地表示同一词在不同上下文中的不同含义。与静态词向量模型(Word2Vec)相比,BERT 具有更强的上下文感知和词义消歧能力。BERT 可以根据输入序列的上下文动态调整每个词的向量表示,从而在处理多义词和短语时表现更佳<sup>[9]</sup>。

BERT 通过在大规模未标记语料上进行预训练,学习到通用的语言知识和语义表示能力,使 BERT 可以进行迁移学习,大大减少任务对专有数据的需求,提高模型性能和训练效率<sup>[10]</sup>。

BERT-BiGRU 模型在 BERT 的基础上加入 GRU 层。GRU 可以更深入地学习序列的上下文依赖关系,增强模型的语义理解能力。BERT 的词向量输入被用作 GRU 输入,GRU 输出用于各下游任务(如文本分类和序列标注)。

BERT-BiGRU 模型结合 BERT 强大的上下文语义表示学习能力和 GRU 对长距离依赖关系的建模优势,它先用 BERT 进行预训练学习泛化语言知识,然后将 GRU 参数初始化为 BERT 输出权重。在下游任务中微调 GRU,以实现有效的迁移学习。

首先,定义一个问句文本序列  $X$ ,在其前后加入特殊标记 [CLS] 和 [SEP],具体如公式(1)所示:

$$X = [\text{CLS}] x_1 x_2 \cdots x_n [\text{SEP}] \quad (1)$$

其次,将  $X$  输入 BERT 模型中得到句子的上下文语义,表示为  $h$ 。其中: $x_n$  表示句子长度,具体如公式(2)所示:

$$h = \text{BERT}(X) \quad (2)$$

最后,在 BERT 模型的基础上添加 GRU 层,并用于各个下游任务,如文本分类、序列标注等。这种做法可以进一步增强模型对长距离依赖关系的建模能力和语义理解能力。同时,由于 BERT 提供了输入词向量并学习到了泛化语言知识和语义表示能力,因此可以有效地进行迁移学习以获得更好的效果。

## 2.3 Encoder 层和 Decoder 层

Encoder 和 Decoder 结构图如图 4 所示。

使用 Transformer 模型对 BERT 模型的编码层进行编码,具体为使用多层的 Self-Attention 和前馈神经网络模块实现对输入序列的高级语义信息的抽象和压缩。通过嵌入层和分段/位置嵌入层,将输入序列映射为词向量序列,并为每个词赋予语义信息和位置信息。这种做法可以更好地利用词汇之间的语义相关性,并提高模型对序列中不同部分的理解能力。使用 LayerNorm 层对词向量序列进行归一化,使其均值为 0、标准差为 1,从而让模型更易于训练和收敛。使用 Self-Attention 层实现序列内部的注意力机制,获得整个序列的上下文信息和词与词之间的依赖关系。Self-Attention 层的输出会经过 GELU 函数激活,引入非线性,之后使用 Add & Norm 层将 Self-Attention 层的输出与输入序列相加,并重新归一化,以便模型学习输入序列的残差信息。使用前馈神经网络模块进一步抽象序列信息,得到输入序列的高级语义表达。通过重复使用 Self-Attention 和前馈神经网络模块,一般为 12~24 层,实现了多层抽象和压缩,从而获得输入序列的高级语义信息和上下文表示。这种基于 Transformer 模型的编码层设计,可以有效地提高 BERT 模型的性能和泛化能力,并在自然语言处理领域等

任务中取得了重要的进展。

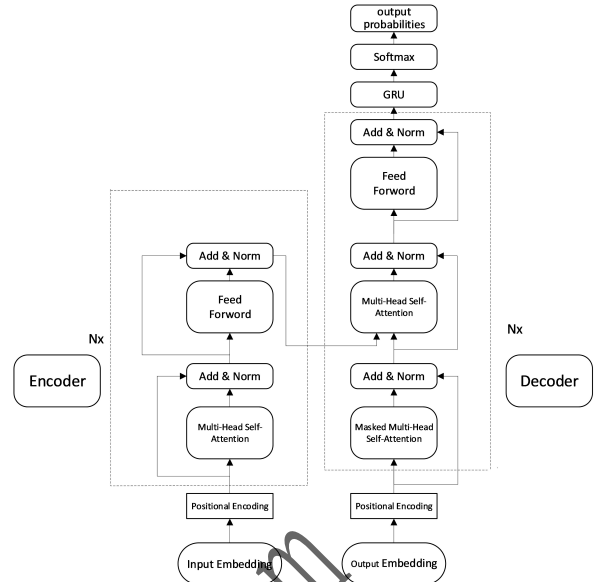


图 4 Encoder 和 Decoder 结构图

Fig. 4 Encoder and Decoder structure diagram

## 2.4 双向 GRU 层

GRU 通过重置门和更新门控制信息的存储与遗忘,可以有效解决 RNN(循环神经网络)的梯度弥散问题,它仅需要保留隐状态和候选隐状态两个向量,计算简单、高效。所以,GRU 非常适用于序列建模,可以学习长距离依赖关系。GRU 只包含重置门和更新门两个门控机制,LSTM 包含遗忘门、输入门和输出门三个门控机制,但是 GRU 简单且高效,只需要维护两个向量,计算过程简单<sup>[9]</sup>。

GRU 单元结构如图 5 所示。GRU 的更新门和重置门对应于“ $z$ ”门和“ $r$ ”门。

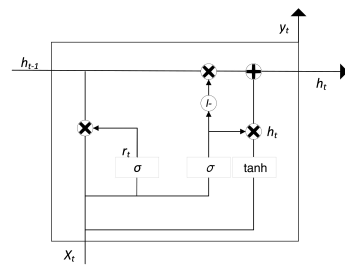


图 5 GRU 单元结构

Fig. 5 GRU unit structure

更新门控制前一隐状态的信息流入当前隐状态  $h_t$  的量。更新门的值越大,表示前一隐状态的信息对当前隐状态的影响越大。更新门可以有效地存储与传递长期依赖信息,但也可能导致历史无关信息的累积。

重置门控制先前隐状态信息被忽略的量。重置门的值越小,表示更多无关历史信息被忽略。重置门可以有效地遗忘历史无关信息,但也可能割断长期依赖关系。

通过更新门和重置门的协同工作,GRU 可以在遗忘无关历史信息 and 维持长期依赖信息之间达到平衡。更新门确保重要历史信息得以保留和传递,重置门确保遗忘无关历史信息,使隐状态仅保留相关信息。更新门和重置门之间的比例关系决定了历史信息在当前隐状态中的比例,实现了对信息的精细



控制。公式(3)至公式(6)说明了 GRU 单元的计算过程。

重置门:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (3)$$

更新门:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (4)$$

候选隐状态:

$$h_t = \tanh(W_h \cdot [h_{t-1} \cdot r_t, x_t]) \quad (5)$$

隐状态:

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot h_t' \quad (6)$$

### 3 实验与结果分析(Experiment and result analysis)

#### 3.1 实验数据集选择

为了评估本文提出的模型的性能,使用 CBLUE 中的 QIC 数据集对意图识别模型进行实验。意图识别模型使用的是一个包含病情诊断、病因分析、治疗方案、就医建议、指标解读、疾病描述、后果表述、注意事项、功效作用和医疗费用等数据的中文医学意图数据集。该数据集总共包含 6 931 个样本,涵盖医学领域的很多知识。

#### 3.2 评价指标

为了客观地评价意图识别模型与槽位填充模型的性能,本研究采用了三个广泛应用的评价指标:精确率(Precision)、召回率(Recall)与 F1 值。

精确率是正确识别的标签数与识别出的总标签数之比,反映模型识别结果的准确度,其计算公式如下:

$$Precision = \frac{\text{正确识别的标签数}}{\text{识别出的总标签数}} \times 100\% \quad (7)$$

召回率是正确识别的标签数与全部相关标签数之比,反映模型识别出的相关结果占全部相关结果的比例,其计算公式如下:

$$Recall = \frac{\text{正确识别的标签数}}{\text{全部相关标签数}} \times 100\% \quad (8)$$

F1 值是精确率与召回率的调和平均值,综合考虑了模型的准确度与覆盖度,其计算公式如下:

$$F1 = \frac{2 \times Precision \times Recall}{(Precision + Recall)} \times 100\% \quad (9)$$

F1 值同时考量了精确率与召回率,它是评价模型性能的重要指标。一般来说,性能优异的模型具有较高的精确率、召回率与 F1 值,因此本研究采用这三个指标评估意图识别模型的性能。

#### 3.3 实验设置及结果分析

在实验中,采用 BERT Base 预训练模型,该模型包含 12 层的编码器、768 个隐藏单元、12 个注意力头、110 MB 个参数。在 PyTorch 框架下,通过采用 GPU 加速来搭建意图分类模型,以提升模型训练的速度和效率。在模型参数设置环节,将序列长度设定为 128。为了进一步优化模型的训练效果,选用了 Adamax 优化算法,并设定学习率为  $2.5e-5$ ,权重衰减为  $1e-4$ 。此外,在训练过程中,还采用了线性学习率调度器,旨在提高模型的最终效果。

为了评估本文提出的 BERT-BiGRU 模型在意图分类任务上的性能,选择多种文本分类模型作为基准,包括 FastText、LSTM 和 Text-CNN,与本文的 BERT-BiGRU 模型在相同数据集上进行比较。表 2 展示了本文提出的模型和其他基线模型在 Question Intent Classification (QIC) 数据集上的实验结果<sup>[7]</sup>。

表2 QIC 数据集的实验结果

Tab.2 Experimental results of QIC dataset

| 模型         | 精确率/% | 召回率/% | F1 值/% |
|------------|-------|-------|--------|
| Fasttext   | 70.84 | 62.15 | 66.21  |
| LSTM       | 72.72 | 68.63 | 70.61  |
| Text-CNN   | 73.40 | 70.23 | 71.78  |
| BERT-BiGRU | 79.22 | 81.23 | 79.82  |

从表 2 可以看出,在医疗短文本的意图识别任务上,LSTM 模型和 Text-CNN 模型的评价指标相近,表现较为稳定。其中,LSTM 模型的精确率和召回率分别达到 72.72% 和 68.63%,F1 值为 70.61%;Text-CNN 模型的精确率和召回率分别达到 73.40% 和 70.23%,F1 值稍高,达到 71.78%。

基于 BERT-BiGRU 模型的意图识别模型效果要显著优于 LSTM 模型和 Text-CNN 模型。BERT-BiGRU 模型的召回率高达 81.23%,精确率达到 79.22%,F1 值最高时达到 79.82%。这是因为 BERT-BiGRU 模型在大规模通用数据集上进行预训练,学习到通用的语言知识和丰富的语义信息。在 Domain Adaptation 阶段,BERT-BiGRU 模型通过添加分类器和进行微调,有效迁移预训练的知识用于下游的意图识别任务,明显提高了模型的泛化能力和语义理解效果。

对于医疗领域的意图识别任务,BERT-BiGRU 模型具有高召回率。高召回率就意味着模型可以正确识别出更多的正样本,减少漏报的样本数,这对于医疗领域的应用而言至关重要,因为只有尽可能识别出所有的潜在医疗意图,才能最大限度地避免遗漏任何医疗相关的词汇或短语。同时,BERT-BiGRU 模型通过其强大的语言表达能力和丰富的上下文语义信息,可以最大限度地识别出给定文本中的所有医疗相关信息,进而实现更高的召回率。

实验证明,本文提出的融合模型相比其他基线模型性能更优,即能更好地表示文本的语义信息和应用于医疗领域的意图识别任务。

基于 BERT-BiGRU 模型的智慧医疗问答系统输出结果展示图(部分)如图 6 所示,通过头痛问题的举例来看,该系统显示出较强的理性分析和知识运用能力。首先,系统准确地识别和概括了两种主要类型的头痛及其特征,展现出对头痛的准确分类,且对其成因有深入的理解;其次,针对不同的头痛类型,智慧医疗问答系统提出了具有系统性和针对性的预防和治疗方案。总之,智慧医疗问答系统在解答头痛相关问题时表现出较强的知识支撑能力和理性思维能力。这为该系统解答更加复杂的问题提供了基础与保障。

#### 智慧医疗问答

小满

欢迎使用医生知识图谱!

小满

头痛的常见原因包括:一、原发性头痛(如偏头痛、紧张性头痛);二、继发性头痛(如感染、外伤、药物副作用等)。头痛的预防和治疗需要根据具体情况进行。一、对于原发性头痛,建议保持规律的作息,避免过度劳累,适当进行体育锻炼,保持心情舒畅。二、对于继发性头痛,应积极治疗原发病,如感染、外伤等。同时,应避免使用可能导致头痛的药物,如某些抗生素、避孕药等。三、对于药物副作用引起的头痛,应咨询医生调整用药方案。四、对于紧张性头痛,建议放松心情,适当进行放松训练,如深呼吸、冥想等。五、对于感染引起的头痛,应积极治疗感染,如使用抗生素、抗病毒药物等。六、对于外伤引起的头痛,应进行对症处理,如冷敷、热敷、使用止痛药等。七、对于其他原因引起的头痛,应根据具体情况进行处理。总之,头痛的预防和治疗需要综合考虑多种因素,建议在医生的指导下进行。

请输入

发送

图 6 基于 BERT-BiGRU 模型的智慧医疗问答系统输出结果展示图(部分)

Fig. 6 Display of some output results of intelligent Q&A system based on BERT-BiGRU model

(下转第 25 页)