基于多策略机制和 BERT 的中医药问题生成

杨祖元 方思凡 陈禧琛 李珍妮

（广东工业大学自动化学院 广州 510006）

〔摘要〕 基于预训练模型 BERT 和 UniLM MASK 提出一个可应用于中医药问题生成的生成式 BERT，结合基于标签平滑、对抗扰动和知识蒸馏的多策略机制，以及多模型软投票的集成策略，提高生成式 BERT 的性能表现和泛化能力，有助于中医药问题生成任务取得更好效果以及中医药文本数据的充分利用。

〔关键词〕 问题生成；深度学习；预训练；中医药；BERT

〔中图分类号〕R－058 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2022.11.010

Question Generation about Chinese Medicine Based on Multi Strategy Mechanism and BERT *YANG Zuyuan*, *FANG Sifan*, *CHEN Xichen*, *LI Zhenni*, *School of Automation*, *Guangdong University of Technology*, *Guangzhou 510006*, *China*

〔**Abstract**〕 Based on the pre－training model BERT and UniLM MASK, the paper proposes a generative BERT which can be applied to the generation of traditional Chinese medicine questions. Combined with the multi－strategy mechanism based on label smoothing, anti－disturbance and knowledge distillation, and the integrated strategy based on multi－model soft voting, the performance and generalization ability of the generative BERT are further improved. It is helpful for the question generation task of traditional Chinese medicine to achieve better results and to make full use of traditional Chinese medicine text data.

〔**Keywords**〕 question generation; deep learning; pre－training; traditional Chinese medicine; BERT

1 引言

中医是中华民族的瑰宝，丰富的中医药文本数据蕴藏在中医药书籍和信息网站中。如何利用海量数据来促进人工智能以及下游智能医疗产业发展，逐渐成为自然语言处理（Natural Language Processing，NLP）领域的研究热点之一[1]。问题作为自然语言文本中最常见的句式之一，多与其他句式共同应用于医疗领域的各种自然语言处理任务中，例如医疗问答系统[2]、医疗信息检索系统[3]等。但现有的大多数中医药文本信息都是以陈述句形式存在，缺乏与之对应的问题，大量文本数据无法被充分利用。因此，需要借助问题自动生成技术来高效构建中医药领域的问题文本。

近 10 年来，深度学习技术在各个领域深入应用，例如医疗图像分割[4]、心电识别[5]、药物设计[6]等。在自然语言处理领域的问题生成任务中，基于卷积神经网络（Convolutional Neural Networks，CNN）[7]、长短时记忆（Long Short－Tterm Memory，LSTM）网络[8]等深度学习网络结构的问题生成模型都具有比较突出的性能表现。深度神经网络可通过自身多层神经结构来学习数据在不同层次上的表示[9]，从而避开传统算法的特征工程或者外部知识库的依赖性。2018 年由谷歌提出的基于 Transformer 的大型预训练模型 BERT[10]刷新了自然语言处理领

〔修回日期〕 2022－11－07

〔作者简介〕 杨祖元，博士，教授，发表论文 60 篇；通信作者：陈禧琛。

域各种任务的排行榜记录，通过使用掩码建模以及句子预测的预训练任务让模型习得词层次和句层次的表征信息，然后通过在一些下游的任务进行微调，在文本分类、机器阅读理解等，任务上取得最好的性能表现。但 BERT 自身不具有 seq2seq 的能力，BERT 内部的 Transformer Encoder 的注意力编码是一个双向编码，即在对问题这个生成对象进行编码时，可以看到来自未来的信息。本文将 UniLM MASK[11]技术引入到预训练模型 BERT 中，构建出可应用于中医药问题生成的生成式 BERT。同时提出基于标签平滑、对抗扰动和知识蒸馏的多策略机制，期望基于多策略机制的生成式 BERT 可以在中医药问题生成的测试集上有更好的性能表现及泛化能力。

2 相关文献

早期的问题生成算法研究大多数是基于规则的传统方法。这种类型的算法一般步骤是：先对文本进行语法或句法分析、词性标注等预处理操作，然后利用人工指定的转换规则或者模板来进行问题生成。例如 Heilman M[12]利用简化事实陈述和浅层语义分析来进行句子生成和排序，最后保留排名靠前的问句来限制冗余问句的生成。汪卫明等[13]通过构建满足医学论文的特定语义关系，人工定义基于语法结构和概念元素的医学问题模板来进行问题生成。Curto S 等[14]使用自上而下的递归方法来建立依存句法树，然后检测句法树中可被用作答案的块，最后使用疑问词去替换答案块中的部分单词，从而实现问题生成。Liu M 等[15]将结构化的中文语言知识引入到句子的简化和生成规则中，并结合机器学习和排序算法来解决中文事实性问题生成任务。虽然这些传统方法具有良好的可解释性，但在实际应用中依赖大量人工设计的特征以及先验知识，模型的规模和泛化能力十分有限。

随着深度学习技术蓬勃发展，越来越多的学者将深度学习应用到问题生成任务中。目前大多数用于问题生成的深度学习框架都是基于 encoder - decoder 架构，用一个深度学习模型作为 encoder 来提取文本语义信息，以另一个深度学习模型作为 decoder 并利用 encoder 提取的特征信息进行问题生成。例如 Du X 等[16]建立两个双向 LSTM 分别作为 encoder 和 decoder，然后使用注意力机制来对 encoder 的编码向量进行注意力关注，再传给 decoder 进行问题生成。Sun X 等[17]提出一个基于答案关注和位置关注的问题生成模型，该模型对上下文单词和答案单词的相对位置进行建模，并在问题单词的生成过程中引入答案嵌入信息，以此来解决生成问题与答案类型不匹配以及复制的上下文单词与答案无关的问题。Zhao Y 等[18]提出一个新的 encoder - decoder 模型，encoder 由双向 LSTM 和门控自注意力机制组成，decoder 由另一个双向 LSTM 和最大值指针机制组成，该模型解决了问题生成中上下文篇幅过长的问题，能够充分利用篇章级别的上下文信息来进行问题生成。Kim Y 等[19]提出一个基于答案分离的 encoder - decoder 生成模型，该模型使用特殊字符去掩码上下文的答案部分，然后使用两个 encoder 分别对掩码后的上下文和答案进行编码，而 decoder 利用关键词网络来捕获答案中关键的信息，并结合 decoder 的编码信息来进行问题生成。谭红叶等[20]提出一个基于答案和上下文的问题生成模型，首先利用基于注意力机制的 CNN 来捕获答案和上下文关系并确定问题中的疑问词，然后用双向 LSTM 和命名实体识别（Conditional Random Field，CRF）提出上下文中的问题相关词，最后基于疑问词、相关词和上下文进行问题生成。虽然 encoder - decoder 的深度学习结构可以较好地解决问题生成任务，但要生成高质量的问题，则需要对 encoder 和 decoder 中的模型以及两者之间的交互关系进行精心的设计以及大量的实验验证，而且大多数 encoder 和 decoder 依赖于循环神经网络，而循环神经网络在实际应用中存在无法并行化计算的问题。

本文避免采用以往 encoder - decoder 结构以及对循环神经网络的依赖，利用预训练语言模

型 BERT 和 UniLM MASK 设计出一个生成式 BERT。生成式 BERT 无需进行复杂的模型和交互关系设计，就可以使模型具有 encoder 的文本编码能力以及 decoder 的问题生成能力。同时，本文基于标签平滑、对抗扰动和知识蒸馏的多策略机制，将生成式 BERT 应用于中医药问题生成任务。

3 模型设计

3.1 生成式 BERT

针对问题生成任务，本文使用大型预训练模型 BERT，并结合 2019 年提出的 UniLM MASK 来实现一个非 seq2seq 结构的生成式 BERT，见图 1。

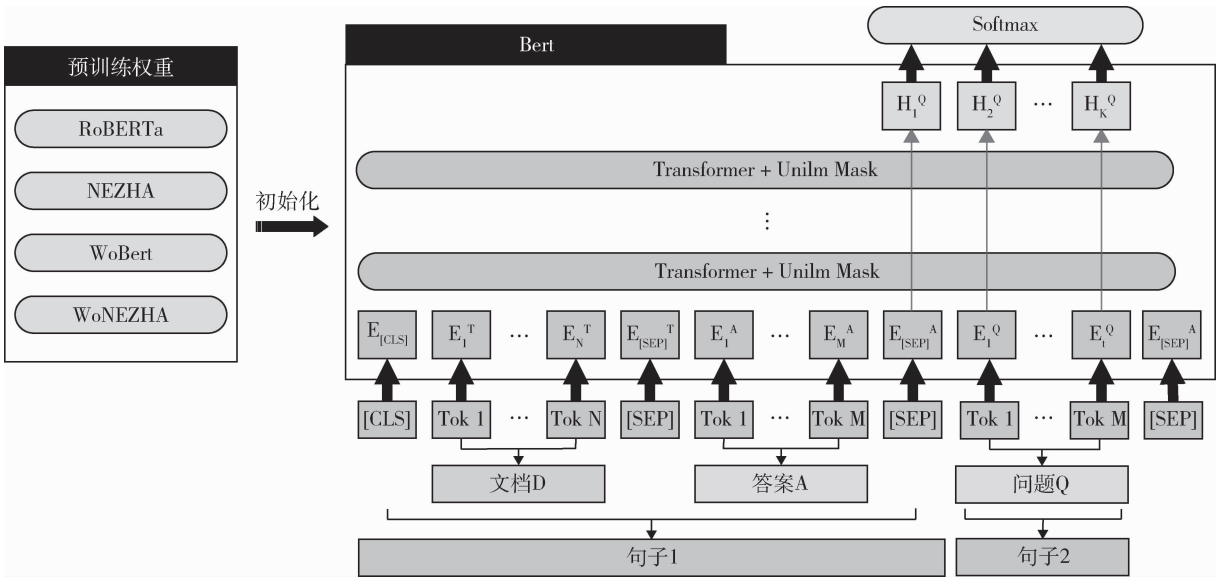

图 1 生成式 BERT 模型框架

在问题生成任务中，给定文档文本 $D = \{d_1, \cdots, d_N\}$ 和文档答案 $A = \{a_1, a_2, \ldots, a_M\}$，要求模型可以基于 D 和 A 来生成问题 $Q = \{q_1, q_2, \ldots, q_L\}$。其中 D、A 和 Q 都是以字为单词进行划分，每个字都来自 BERT 对应字表 V。基于 BERT 的通用输入构成，本文构建一个针对问题生成的输入序列作为生成式 BERT 的输入，该输入序列为：

$[[CLS], d_1, \ldots, d_N, [SEP], a_1, \ldots, a_M, [SEP], q_1, \ldots, q_L, [SEP]]$

生成式 BERT 的输入层与普通 BERT 的结构一致，其作用是将输入序列中的字 w_i 转换为嵌入向量 E_i。嵌入向量由字向量、位置向量和句子类别向量相加得到。字向量表示字的语义信息，通过字嵌入矩阵 $W_E \in E^{V \times d_k}$ 索引得到，字嵌入矩阵由大量文本数据预训练得到。因为 BERT 没有使用循环神经网络进行递归编码，所以需要引入位置向量来记录序列中每个字的位置信息。对于句子类别，将文档 D 和答案 A 的部分序列标记为句子 1，将问题 QD 的部分序列标记为句子 2。

BERT 的中间层由多个 Transformer Encoder 组成。Transformer Encoder 主要有两个子网络。第 1 个子网络是使用多头自注意力机制的上下文编码网络。上下文编码网络使用多头自注意力机制直接对序列中的每个字进行上下文编码，每个字的表示融合了上下文信息。多头自注意力机制由自注意力机制和多头机制组成。假设第 i 层 Transformer Encoder 的输入向量是 $H_i = [h_1, \ldots, h_T] \in E^{T \times d_k}$，其中 $T = N + M + L + 4$，d_k 是中间层的特征维度，则自注意力机制的具体计算如下：

$$Q = H_i \tag{1}$$

$$K = H_i \tag{2}$$

$$V = H_i \tag{3}$$

$$SelfAttention(Q, K, V) = softmax(\frac{Q K^T}{\sqrt{d_k}}) V \tag{4}$$

加入多头机制后，计算如下：

$$h_j = SelfAttention(Q W_j^Q, K W_j^K, V W_j^V) \quad (5)$$

$$MultiHead(Q,K,V) = Concat([h_1, \ldots, h_j]) W^o \quad (6)$$

第 2 个子网络由两层前馈神经网络组成，目的是巩固每个字自身固有的语义信息。假设上下文编码网络的输出是 O_i^1，则第 2 个子网络的具体计算如下：

$$FFN(O_i^1) = ReLu(O_i^1 W_i^1 + b_i^1) W_i^2 + b_i^2 \quad (7)$$

生成式 BERT 中间层的组成与 BERT 的中间层类似，不同之处在于引入 UniLM - MASK。原 BERT 中间层在使用 Transformer Encoder 对序列进行编码时，具有双向编码的特性，但在生成任务中模型对于问题序列的编码只能是单向的。因为问题序列是通过循环生成得到的，无法看到来自未来的信息。所以，本文引入 UniLM MASK 到生成式 BERT 的 Transformer Encoder 中，使得模型对文档序列和答案序列的编码是双向编码，对问题序列的编码是单向编码。具体计算如下：

$$M = \begin{cases} 0, & if \quad 允许注意力关注 \\ -\infty, & if \quad 阻止注意力关注 \end{cases} \in R^{T \times T} \quad (8)$$

$$MaskSelfAttention(Q,K,V) = softmax(\frac{Q K^T}{\sqrt{d_k}} + M) V \quad (9)$$

生成式 BERT 的输出层是一个带有 softmax 的线性映射层，计算出字表中每个字的预测概率。假设第 j 个样本输入序列的第 j 个字在最后一层 Transformer Encoder 的输出向量为 $o_{ji} \in R^{d_k}$，则输出层计算如下：

$$l_{j,i} = ReLu(o_{j,i} W_f^1 + b_f) W_f^2 \in R^V \quad (10)$$

$$p_{j,i} = softmax(l_{j,i}) \in R^V \quad (11)$$

其中 $W_f^1 \in R^{d_k \times d_o}$，$W_f^2 \in R^{d_o \times V}$，$d_o$ 是输出层的神经元个数。生成式 BERT 的训练过程就是最大化生成问题序列中每个字的概率，所以模型的损失函数计算如下：

$$Loss_{CE} = -\frac{1}{N} \sum_{j=1}^{N} \sum_{i=N+M+4}^{N+M+4+L} y_{j,i} log(p_{j,i-1}^T) \quad (12)$$

其中 N 是训练样本个数，$y_{j,i} \in R^V$ 是第 j 个输入序列的第 i 个字在字表 V 上的真实 one - hot 标签向量。当 $i = N + M + 4$ 时，预测输入序列中问题序列的第 1 个单词 q_1，当 $i = N + M + 3 + L$ 时，预测输入序列中问题序列的最后一个单词 q_K。

解码阶段：在生成式 BERT 的解码阶段，本文使用集束搜索（beam search）来进行问题生成，见图 2。首先将文档和答案进行拼接得到一个输入序列，然后输入到已训练的生成式 BERT 中预测出第 1 个问题单词，将该单词复制 P 份，作为 P 个 beam 的第 1 个问题单词。后续使用某个 beam 已得到的问题单词序列追加到由文档和答案组成的序列后面，进行下一个问题单词的预测，直到预测出终止符或者达到问题的最大长度。最后选择综合概率最大的 beam 所对应的问题单词序列作为最终生成的问题。

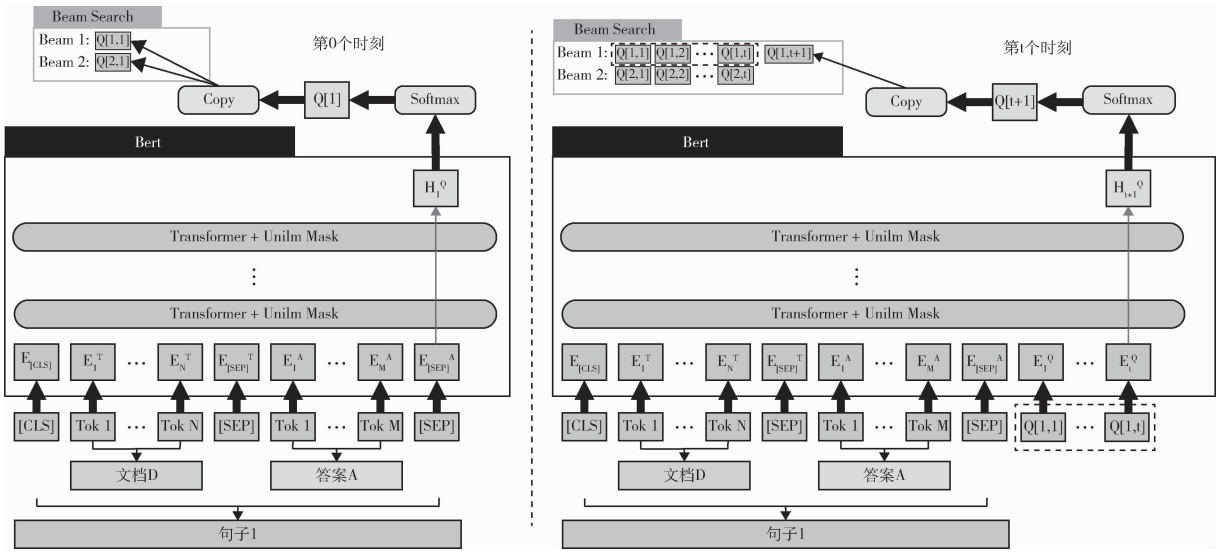

图 2 生成式 BERT 推理阶段

3.2 标签平滑策略

标签平滑策略是一种正则化方法，通常用于分类任务。虽然生成式 BERT 是用于问题生成任务的模型，但单词预测阶段本质上就是一个基于字表的多分类。当模型训练时，单词的标签是一个 one - hot 向量，除了真实标签为 1，其他标签都为 0，即模型需要让 1 标签和 0 标签所属类别和其他类别之间的差距尽可能加大，但由梯度有界可知，这种情况很难适应，会造成模型过于相信预测的类别。为了防止此类问题，改善泛化能力差的问题，本文引入标签平滑策略，来对真实标签进行尺度上的缩放，具体实现如下：

$$\hat{y}_i = y_i * (1 - \alpha) + \frac{\alpha}{K} \in R^V \tag{13}$$

其中 y_i 是第 i 个样本的 one - hot 标签向量，α 是平滑因子，通常是 0.1，K 是字表的字个数，$\hat{y}_i$ 是平滑后的标签向量。

3.3 对抗扰动策略

随着深度学习在各个领域蓬勃发展，关于对抗样本的研究也受到越来越多的关注。在计算机视觉领域，可通过对深度学习模型进行对抗攻击或者防御来提高模型的鲁棒性。而在自然语言处理领域，对抗训练更多是作为一种正则化的方法来提高模型的泛化性能。在计算机视觉领域中，图像可以看作是一个连续实数向量，因此可以很容易加上一个很小的实数向量作为扰动，形成一个对抗样本。但 NLP 的输入是文本，本质上就是 one - hot 向量，因此不存在所谓的小扰动。因此，本文直接对生成式 BERT 中的字嵌入矩阵 W_E 进行对抗扰动，让索引后的字嵌入向量发生变化。

假设生成式 BERT 在输入层的字嵌入表示为 $X \in R^{(N+M+L+4)\times d_k}$，真实标签为问题序列 $y \in R^L$，则对抗扰动策略的具体实现如下：

$$\min_{\theta} E_{(X,y)\in D}\left[\max_{\Delta X \in \Omega} Loss(X + \Delta X, y; \theta)\right] \tag{14}$$

首先，对字嵌入表示 X 加入对抗扰动 ΔX，使得生成式 BERT 的损失值增大，同时 ΔX 受限于约束空间 Ω。对抗样本 $X + \Delta X$ 构建完成后，再次输入到生成式 BERT 来最小化模型的损失值，更新模型参数。本文使用快速梯度方法计算字嵌入矩阵的梯度 ΔW_E，然后根据得到的梯度对字嵌入矩阵 W_E 进行对抗扰动。输入序列通过已被对抗扰动的字嵌入矩阵获得新字嵌入表示，这个新字嵌入表示间接作为关于字向量表示的对抗样本 $X + \Delta X$：

$$\Delta W_E = \epsilon \frac{\nabla_{W_E} Loss(X, y; \theta)}{|| \nabla_{W_E} Loss(X, y; \theta) ||} \tag{15}$$

$$W_E = W_E + \Delta W_E \tag{16}$$

ϵ 是一个超参数，一般取 0.1。同时，本文对梯度进行标准化，防止计算出来的梯度过大。

3.4 知识蒸馏策略

知识蒸馏策略就是通过引入教师模型和学生模型的概念，使用教师模型来指导学生模型的训练，从而实现知识迁移的目的。本文的教师模型和学生模型都是同种结构的生成式 BERT。

该策略的核心思想就是用教师模型预测的软标签概率来辅助学生模型的训练。首先，本文先训练一个基于生成式 BERT 的教师模型。然后构建一个与教师模型结构相同的学生模型。在对学生模型进行训练时，先计算出教师模型的软标签概率，接着与学生模型输出的概率进行知识蒸馏（Knowledge Distillation，KD）损失函数计算。但是，直接使用教师模型预测的软标签概率是一种有缺陷的做法。因为一个网络训练好后，对正标签有很高的置信度，负标签的值都很接近 0，对损失函数的贡献非常小，小到可以忽略不计。所以，需要引入一个温度变量 T 来让教师模型和学生模型输出的概率分布更加平滑：

$$\hat{t}_{j,i} = softmax(\frac{l^t_{j,i}}{T}) \tag{17}$$

$$\hat{s}_{j,i} = softmax(\frac{l^s_{j,i}}{T}) \tag{18}$$

其中 $l^t_{j,i}, l^s_{j,i} \in R^V$ 分别是教师模型和学生模型对第 j 个输入序列的第 i 个单词的输出概率。放大负标签概率可以让学生模型学习到不同负标签与正标签之间的关系。

所以，基于知识蒸馏的损失函数如下：

$$Loss_{KD} = -\frac{1}{N}\sum_{j=1}^{N}\sum_{i=N+M+3}^{N+M+3+L} \hat{t}_{j,i} log(\hat{s}^T_{j,i}) \tag{19}$$

而学生模型的总损失函数如下：

$$Loss = \beta Loss_{KD} + (1-\beta) Loss_{CE} \quad (20)$$

其中，β 是一个超参数，用于调节两个损失的比重，$Loss_{CE}$ 是交叉熵损失函数（Cross EntropyLoss，CE）。

4 实验与结果

4.1 数据集描述

本文实验所使用的数据集来自于第六届中国健康信息处理大会发布的“中医文献问题生成”评测任务。数据集中的标记数据全部来源于中医药领域文本，包括《黄帝内经翻译版》《名医百科中医篇》《中成药用药卷》《慢性病养生保健科普知识》4 个主要来源和部分中医论坛文本，从 5 000 篇文档中共标注了 13 000 对答案、问题对，其中每篇文档由人工构造产生 1 ~4 对答案、问题对，“答案”均为文档中的连续片段，“问题”均由人工标注产生。该评测任务由阿里云天池平台提供技术支持，其中 3 500 篇语料及其标注数据用作训练数据，750 篇测试数据用于决赛阶段的线上评测。同时，本文将训练数据按照 9∶1 的比例进行训练集和验证集的划分。

4.2 评测指标

本文使用的评测指标是 Rouge - L，通过计算生成问题和原问题之间的最长公共子序列，以此来衡量自动生成的问题与原问题之间的匹配度。使用最长公共子序列的一个优点是它不需要连续匹配，而且反映了句子级词序的顺序匹配。由于自动包含最长的 n - gram，不需要预定义的 n - gram 长度。Rouge - L 定义公式为：

$$R_{lcs} = \frac{LCS(X,Y)}{m} \quad (21)$$

$$P_{lcs} = \frac{LCS(X,Y)}{n} \quad (22)$$

$$Rouge - L = \frac{(1+\lambda^2) R_{lcs} P_{lcs}}{R_{lcs} + \lambda^2 P_{lcs}} \quad (23)$$

其中，LCS 是计算最长公共子序列的函数；m 是原问题的长度；n 是生成问题的长度；R_{lcs} 和 P_{lcs} 分别表示召回率和精准率；λ 为一个超参数，用于调节召回率和精准率之间的比重。

4.3 参数设置

本文实验所使用的 BERT 预训练权重文件有：哈尔滨工业大学开源的 RoBERTa、华为开源的 NEZHA，以及追一科技开源的 WoBERT 和 WoNEZHA。文档的最大长度为 384，答案的最大长度为 96，问题的最大长度为 32。训练轮数为 5，批大小为 4，梯度累积更新步数为 8。标签平滑的平滑因子 α 为 0.1。如果预训练权重文件为 RoBERTa 或者 WoBERT，则对抗训练的 ϵ 参数为 0.3，如果预训练权重文件为 NEZHA 或者 WoNEZHA，则对抗训练的 ϵ 参数为 0.1。知识蒸馏的 β 参数是 0.5，温度系数 T 为 10。Rouge - L 指标的 λ 为 1。Beam Search 的参数 P 为 5。

4.4 结果与分析

使用标签平滑策略的 RoBERTa 比没有使用该策略的 RoBERTa 在决赛线上测评时 Rouge - L 分数提高 0.001 8，见表 1，说明引入标签平滑，可以防止模型过度自信地进行标签预测，从而达到提高模型泛化能力的效果。而结合标签平滑和对抗扰动策略的 RoBERTa 模型比单纯使用标签平滑策略的 RoBERTa 模型的 Rouge - L 提高 0.003 3，因为该策略通过对模型的 Embedding 层进行对抗扰动来产生对抗样本，模型在训练时受到对抗样本的攻击可以在一定程度上提高模型的鲁棒性，从而达到提高模型表现能力的目的。此外，使用结合标签平滑和对抗扰动策略的 RoBERTa 模型构建了教师模型和学生模型，利用知识蒸馏策略让教师模型去指导学生模型的训练过程，同时让学生模型习得教师模型软标签中正标签和不同负标签的关系，表 1 中结果证明了基于知识蒸馏策略的 RoBERTa 比只使用标签平滑和对抗扰动策略的 RoBERTa 的 Rouge - L 提高 0.001 4。最后，使用多策略机制（标签平滑 + 对抗扰动 + 知识蒸馏）的 RoBERTa 比没有使用多策略机制的 RoBERTa 在决赛线上测评时 Rouge - L 提高 0.006 5，这表明结合多策略机制的 BERT 模型在中医药问题生成任务上的有效性。

表 1 生成式 BERT 在中医药问题生成数据集的实验结果

模型	线上测试集 Rouge - L
RoBERTa	0.612 6
RoBERTa + 标签平滑	0.614 4
RoBERTa + 标签平滑 + 对抗扰动	0.617 7
RoBERTa + label smoothing + 对抗扰动 + 知识蒸馏	0.619 1
[RoBERTa + NEZHA] + label smoothing + 对抗扰动 + 知识蒸馏（集成模型）	0.625 8
[WoBERT + WoNEZHA] + label smoothing + 对抗扰动 + 知识蒸馏（集成模型）	0.627 8

同时，本文还采用基于多模型软投票的集成策略来进一步提高模型性能表现，主要利用不同预训练模型之间对语言建模的差异性。比起 RoBERTa，NEZHA 使用了相对位置编码，所以更适合对长文本进行建模编码。从表 1 可以看出，基于多策略机制的 RoBERTa 和 NEZHA 的集成模型比起基于多策略机制的单模型 RoBERTa，Rouge - L 提高 0.006 7。使用基于字 + 词分别在 RoBERTa 和 NEZHA 继续预训练得到的 WoBERT 和 WoNEZHA，从结果可以得出，基于多策略机制的 WoBERT 和 WoNEZHA 的集成模型较之基于多策略机制的 RoBERTa 和 NEZHA 的集成模型，Rouge - L 提高 0.002 0，主要是因为 WoBERT 和 WoNEZHA 融合了中文词信息，使得模型在进行语义建模时更加明确。

5 结语

本文使用结合多策略机制的生成式 BERT 来解决中医药问题生成任务，并在中医药问题生成数据集上取得不错的性能表现。由 BERT + UniLM MASK 组成的生成式 BERT 的 Rouge - L 为 0.612 6，说明了非 encoder - decoder 结构的生成式 BERT 可以实现中医药问题生成。此外，本文还使用了基于标签平滑、对抗扰动和知识蒸馏的多策略机制来提高生成式 BERT 的性能表现和泛化能力，结合多策略机制的生成式 BERT 的 Rouge - L 为 0.619 1。最后使用基于多模型软投票的集成策略，让生成式 BERT 集成模型 Rouge - L 最高达到了 0.627 8。

从本文的实验效果来看，使用该代码可以在中医药问题生成任务上取得较好的效果。

参考文献

1 柴华，路海明，刘清晨. 中医自然语言处理研究方法综述 [J]. 医学信息学杂志，2015，36（10）：58 - 63.

2 马满福，刘元喆，李勇，等. 基于 LCN 的医疗知识问答模型 [J]. 西南大学学报（自然科学版），2020，42（10）：25 - 36.

3 Liu L, Liu L, Fu X, et al. A Cloud - based Framework for Large - scale Traditional Chinese Medical Record Retrieval [J]. Journal of Biomedical Informatics, 2018, 77 (1): 21 - 33.

4 Zhou S, Nie D, Adeli E, et al. High - resolution Encoder - decoder Networks for Low - contrast Medical Image Segmentation [J]. IEEE Transactions on Image Processing, 2019, 29 (1): 461 - 475.

5 Abdeldayem S S, Bourlai T. A Novel Approach for ECG - based Human Identification Using Spectral Correlation and Deep Learning [J]. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2019, 2 (1): 1 - 14.

6 Urban G, Bache K, Phan D T T, et al. Deep Learning for Drug Discovery and Cancer Research: Automated Analysis of Vascularization Images [J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2018, 16 (3): 1029 - 1035.

7 董孝政，洪宇，朱芬红，等. 基于密令位置信息特征的问题生成 [J]. 中文信息学报，2019，33（08）：93 - 100.

8 Song L, Wang Z, Hamza W, et al. Leveraging context information for natural question generation [C]. New Orleans: Association for Computational Linguistics, 2018, 569 - 574.

9 LeCun Y, Bengio Y, Hinton G. Deep Learning [J]. Nature, 2015, 521 (7553): 436 - 444.

10 Devlin J, Chang M W, Lee K, et al. BERT: Pre - training

of Deep Bidirectional Transformers for Language Understanding [C]. Minneapolis: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019.

11 Dong L, Yang N, Wang W, et al. Unified Language Model Pre - training for Natural Language Understanding and Generation [C]. Vancouver: Advances in Neural Information Processing Systems, 2019.

12 Heilman M. Automatic Factual Question Generation from Text [D]. Pittsburgh: Carnegie Mellon University, 2011.

13 汪卫明, 陈世鸿, 王世同, 等. 基于语义模板的医学问答自动生成 [J]. 武汉大学学报 (理学版), 2009, 55 (2): 233 - 238.

14 Curto S, Mendes A C, Coheur L. Question Generation based on Lexico - syntactic Patterns Learned from the Web [J]. Dialogue & Discourse, 2012, 3 (2): 147 - 175.

15 Liu M, Rus V, Liu L. Automatic Chinese Factual Question Generation [J]. IEEE Transactions on Learning Technologies, 2016, 10 (2): 194 - 204.

16 Du X, Shao J, Cardie C. Learning to Ask: Neural Question Generation for Reading Comprehension [C]. Vancouver: Association for Computational Linguistics, 2017.

17 Sun X, Liu J, Lyu Y, et al. Answer - focused and Position - aware Neural Question Generation [C]. Brussels: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018.

18 Zhao Y, Ni X, Ding Y, et al. Paragraph - level Neural Question Generation with Maxout Pointer and Gated Self - attention Networks [C]. Brussels: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018.

19 Kim Y, Lee H, Shin J, et al. Improving Neural Question Generation Using Answer Separation [C]. Hawaiian: Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33.

20 谭红叶, 孙秀琴, 闫真. 基于答案及其上下文信息的问题生成模型 [J]. 中文信息学报, 2020, 34 (5): 74 - 81.

(上接第 54 页)

24 郭玉峰, 刘保延, 尹爱宁, 等. SNOMED CT 术语分类体系设定学科背景的探讨 [J]. 世界科学技术 - 中医药现代化, 2007 (4): 86 - 90.

25 钟伶, 林丹红, 林晓华. 临床医学系统术语 SNOMED CT 的特点及其应用 [J]. 中华医学图书情报杂志, 2007 (2): 58 - 60.

26 Cory M. What Is the Difference Between a Disease and a Disorder [EB/OL]. [2021 - 09 - 05]. https://www.verywellhealth.com/disease - vs - disorder - 5092243.

27 Williams N E, Sugden S. The Factory Model of Disease [J]. Monist, 2007, 90 (4): 555 - 584.

28 郭玉峰, 刘保延, 周雪忠. SNOMED CT 2007 的顶级概念分类详解 [J]. 中华中医药学刊, 2008 (9): 1928 - 1932.

29 郭玉峰, 刘保延, 周雪忠. SNOMED CT 的语义关系与连接概念 [J]. 中华中医药学刊, 2008 (10): 2206 - 2209.

30 郭玉峰, 刘保延, 崔蒙, 等. 借鉴 SNOMED CT 发展中医临床标准术语集 [C]. 北京: 中医药发展与人类健康——庆祝中国中医研究院成立 50 周年, 2005.

31 郭玉峰, 刘保延, 周雪忠. 面向中医临床科研需求的术语分类框架研究 [J]. 环球中医药, 2008 (2): 9 - 12.

32 郭玉峰, 尹爱宁, 周霞继, 等. 浅谈中医临床术语标准化工作现状及其深化推进 [J]. 中国中医药信息杂志, 2009, 16 (11): 3 - 4.

33 刘保延, 尹爱宁, 张润顺, 等. 中医规范术语在结构化电子病历中应用体系的研究 [J]. 中国数字医学, 2012, 7 (8): 41 - 44.

34 The OBO Foundary. Principle: Overview [EB/OL]. [2021 - 09 - 29]. http://www.obofoundry.org/principles/fp - 000 - summary.html.

35 郭玉峰, 刘保延, 李平, 等. 知识本体与中医临床术语规范化工作 [J]. 中华中医药学刊, 2007 (7): 1368 - 1370.

36 中国中医. 《中医病证分类与代码》和《中医临床诊疗术语》印发 [EB/OL]. [2021 - 09 - 19]. https://mp.weixin.qq.com/s/XtDrs8Z1XwhAFs_f85EwSA.