doi:10.6041/j.issn.1000-1298.2022.01.027

基于 BERT - Attention - DenseBiGRU 的农业问答社区问句相似度匹配

王郝日钦[1,2]　王晓敏[1,3]　缪祎晟[1,3]　许童羽[4]　刘志超[1,3]　吴华瑞[1,3]

(1. 国家农业信息化工程技术研究中心, 北京 100097; 2. 内蒙古民族大学计算机科学与技术学院, 通辽 028043;
3. 北京市农林科学院信息技术研究中心, 北京 100097; 4. 沈阳农业大学信息与电气工程学院, 沈阳 110866)

摘要: 为了解决问答社区中相同语义问句文本的快速自动检测,提出一种基于 BERT 的 Attention - DenseBiGRU 农业问句相似度匹配模型。针对农业文本具备的特征,采用 12 层的中文 BERT 文本预训练模型对文本数据进行向量化处理,并与 Word2Vec、Glove、TF - IDF 方法进行对比分析,得出 BERT 方法能够有效地解决农业文本的高维性和稀疏性问题,并且解决多义词在不同语境下具有不同含义的问题。该网络的每一层都使用注意特征的连接信息以及前面所有递归层的隐藏特征,为了缓解由于密集拼接而导致特征向量尺寸不断增大的问题,在模型的最后使用自动编码器进行特征降维。试验结果表明:基于 BERT 的 Attention - DenseBiGRU 农业问句相似度匹配模型可以提高文本特征的利用率,减少特征丢失,能够实现快速及准确的农业问句文本相似度匹配,在本文所构建的农业问句相似对数据集上精确率及 F1 值达到 97.2% 和 97.6%,与其他 6 种问句相似度匹配模型相比,效果提升明显。

关键词: 问答社区; 农业问句相似度匹配; 自然语言处理; 密集连接 BiGRU; 协同注意力机制

中图分类号: TP183　　**文献标识码**: A　　**文章编号**: 1000-1298(2022)01-0244-09　　**OSID**:

Densely Connected BiGRU Neural Network Based on BERT and Attention Mechanism for Chinese Agriculture-related Question Similarity Matching

WANG Haoriqin[1,2]　WANG Xiaomin[1,3]　MIAO Yisheng[1,3]　XU Tongyu[4]　LIU Zhichao[1,3]　WU Huarui[1,3]

(1. *National Engineering Research Center for Information Technology in Agriculture*, *Beijing* 100097, *China*
2. *College of Computer Science and Technology*, *Inner Mongolia Minzu University*, *Tongliao* 028043, *China*
3. *Information Technology Research Center*, *Beijing Academy of Agriculture and Forestry Sciences*, *Beijing* 100097, *China*
4. *School of Information and Electrical Engineering*, *Shenyang Agricultural University*, *Shenyang* 110866, *China*)

Abstract: To allow fast and automatic detection of the same semantic agriculture-related questions, a method based on BERT - Attention - DenseGRU (gated recurrent unit) was proposed. According to the agriculture question characteristics, twelve layers of the Chinese BERT model method were applied to process and analyze the text data and compare it with the Word2Vec, Glove, and TF - IDF methods, effectively solving the problem of high dimension and sparse data in the agriculture-related text. Each network layer employed the connection information of features and all previous recursive layers' hidden features. To alleviate the problem of feature vector size increasing due to dense splicing, an autoencoder was used after dense concatenation. The experimental results showed that agriculture-related question similarity matching based on BERT - Attention - DenseBiGRU can improve the utilization of text features, reduce the loss of features, and achieve fast and accurate similarity matching of the agriculture-related question dataset. The precision and F1 values of the proposed model were 97.2% and 97.6%. Compared with six other kinds of question similarity matching models, a state-of-the-art method with the agriculture-related question dataset was presented.

Key words: question-and-answering communities; agriculture-related question similarity matching; natural language processing; densely connected BiGRU; coattention mechanism

收稿日期: 2021 - 09 - 13　修回日期: 2021 - 10 - 26

基金项目: 国家重点研发计划项目(2019YFD1101105)、财政部和农业农村部:国家现代农业产业技术体系项目(CARS - 23 - C06)、北京市农林科学院青年基金项目(QNJJ202030)和内蒙古民族大学教育教学研究项目(QN2021013)

作者简介: 王郝日钦(1988—),男,博士生,内蒙古民族大学实验师,主要从事农业智能系统研究,E-mail: wanghrq007@nercita.org.cn

通信作者: 吴华瑞(1975—),男,研究员,博士,主要从事农业智能系统研究,E-mail: wuhr@nercita.org.cn

0　引言

随着互联网时代的快速发展，在网络问答社区[1]提出问题、回答问题和讨论问题已经成为人们日常寻求问题解答，满足自身信息需求的重要方法。“中国农技推广信息平台”是一个专业提供农技问答、专家指导、在线学习、成果速递、技术交流等的综合性服务平台，其中农技问答模块在帮助农户找到问题的解决方案方面发挥着重要作用。用户每天在问答模块[2]提出的问题有千万余条，农业专家会及时对所提问题进行解答，但是由于中文语义表达的复杂性，会出现很多描述方式不同但语义相同的问句，对这类问题重复进行解答会耗费大量的人力和物力。这类农业文本具有冗余性、稀疏性[3]，以及规范性差等特点，导致了文本特征提取不准确，难以挖掘特征之间的关系，从文本数据集中快速、自动、准确地检测出语义相同的问句并将相同语义问句对应的正确答案返回给用户是实现农业智能问答[4]的关键技术环节。传统的文本语义相似度判断[5]依靠人工筛查很难高效地完成文本数据的处理。目前常用的关键词查询及浅层神经网络学习模型[6]虽然能够辅助完成相似问句判断等工作，但是由于采用人工特征选择的方式，不具备从大量的农业文本数据中自动、准确地判断相同语义问句的功能。因此利用深度学习[7]和自然语言处理技术[8]实现农业相同语义问句智能检索是“中国农技推广信息平台”需要解决的一个重要难题。

随着深度学习技术的快速发展，国内外学者主要使用卷积神经网络[9]和循环神经网络[10]等模型研究文本相似度计算。文献[11]提出的 DSSM (Deep structured semantic models)模型是最早将孪生网络架构作为基础模型用于语义文本相似度计算的方法，DSSM 模型在文本匹配任务上取得了良好的效果，但是忽略了文本语序和上下文信息。文献[12]将卷积神经网络加入到 DSSM 模型以保留更多文本语序和上下文信息，在表示层中增加了卷积层和池化层，提升了文本匹配效果，但是由于卷积神经网络的限制，仍会丢失一些距离较远的文本特征。文献[13]将长短期记忆网络引入其中，LSTM[14]模型可以很好地解决上述问题，使得文本匹配效果提升。随着自注意力机制在图像和自然语言处理领域的应用，文献[15]将双向长短期记忆网络和自注意力机制技术相结合用来提取文本的特征向量，提高了文本匹配精确度。由于基于孪生网络的方法在提取文本特征时是互相独立的，而基于交互模型的神经网络可以解决上述问题，它是在编码层增加 2 个网络之间的交互作用，从而提取句子对之间的特征关联来提高文本匹配的精确率。文献[16]提出了基于注意力机制的卷积神经网络模型，首先使用 Word2Vec 对文本进行向量化的基础上通过卷积神经网络对句子进行特征提取，再分别对文本对进行卷积和池化操作的同时，使用注意力机制对 2 个中间过程进行交互，提升了文本匹配的精确率。以上研究表明，交互模型对文本相似度匹配具有更好的效果。文献[17－24]为深度学习和自然语言处理技术在农业文本处理方面提供了可行性依据和参考。但是在农业问句相似度匹配过程中，仍存在无法解决多义词在不同语境下具有不同含义的问题和文本特征提取不精确等问题。同时由于农业领域一直缺乏大规模可用的数据库，因此关于农业文本相似度计算的研究还鲜有报道。

为了实现农业问答社区提问数据的快速自动重复语义检测，本文首先利用 12 层的中文 BERT[25]预训练模型获得文本的上下文特征表示，使用密集连接的双向 GRU 网络进一步提取农业问句对的文本特征，使用连接操作将注意力机制对 2 个问句交互的信息合并到密集连接的 BiGRU 中用于问句相似度匹配，并进一步针对神经网络的重要参数进行优化和改进，提出基于 BERT－Attention－DenseBiGRU 的农业文本相似度匹配模型，以期实现农业问答社区相同语义问句自动、精确识别。

1　数据集构建

从“中国农技推广信息平台”问答社区后台服务器中导出农业问答文本数据，共提取到涉及 8 个类别的 30 000 对问答数据，针对 8 个类别内的每条提问文本数据使用 Simhash 算法[26]检索相同语义文档，并进行人工筛选及校对，语义表达相同的问句对记作标签 1，语义表达不相同的问句对记作标签 0，这样可以尽量避免使神经网络认为 2 个句子相同关键词越多越相似，得到农业文本相似问句对共 17 000 对，其中病虫草害、动物疫病、栽培管理、市场销售、养殖管理、土壤肥料、储运保鲜、其他 8 个类别问句对数量分别为 5 890、2 100、3 870、760、1 210、765、217、2 188。训练集样本示例如表 1 所示。

2　模型构建

本文提出的 BERT－Attention－DenseBiGRU 模型如图 1 所示。该模型主要由文本预处理层、密集连接 BiGRU(DenseBiGRU)、注意力机制层和交互分类层 4 部分组成。与传统深度学习模型相比，首先本文使用 12 层的中文 BERT 预训练模型得到问句

表 1 训练集样本示例

Tab. 1 Training set sample example

编号	问句 1	问句 2	标签
1	果树底肥用哪种工艺的肥料好一些？	用哪种工艺的肥料给果树施底肥好一些？	1
2	玉米倒伏有哪些原因？	玉米倒伏原因都有什么？	1
3	马铃薯拱棚栽培技术要点是什么？	马铃薯拱棚栽培如何管理？	0
4	草莓褐斑病发病条件是什么？	草莓褐斑病如何防治？	0
5	请问高粱锈病如何防治？	请问应该怎么避免高粱锈病发生？	1
6	防治农业二化螟的物理防治法都采取哪些措施？	都有哪些措施防治农业二化螟的物理防治法？	1
7	为什么成熟的香蕉会裂开？	成熟的香蕉会裂开的原因是什么？	1
8	五味子果腐病的症状是什么？	五味子黑斑病的发病原因是什么？	0
9	甜菜的主要病害有哪些？	甜菜的形态特征是什么？	0

特征向量化表示；其次本文提出的模型利用 DenseBiGRU 和协同注意力机制提取文本不同粒度的局部特征；最后将提取的特征向量输入到交互分类层。

2.1 文本预处理

首先使用 12 层的中文 BERT 预训练模型作为词向量转换工具获得问句特征表示，既能获得农业问句文本语法、语义特征，又能解决 Word2Vec[27] 忽略词语多义性的问题。BERT 是一个预训练的语言表征模型，不同于传统的单向语言表征模型，BERT 使用双向的 Transformer[28] 对语言模型进行训练，可以获得更深层次的语言表征信息，Transformer 模型结构如图 2 所示。

Transformer 模型中包含 2 个相同的编码器，每个编码单元都由自注意力机制和前向传播网络构成，并且在前向传播的过程中使用了残差连接，将之前层的输入和本层的输出合并之后向后传递，最后进行求和归一化操作。解码单元跟编码单元相比，增加了 1 层注意力机制。BERT 作为文本特征的提取器优于传统的卷积神经网络和循环神经网络，这是因为 BERT 采用了双向的 Transformer 获取词向量，可以充分考虑每个词的上下文信息，从而提取到更加准确的词向量表示，解决了中文一词多义的现象。

BERT 是一种遮蔽语言模型，在获取词向量的过程中随机遮蔽一些词语，然后在预训练过程中在原始词汇的位置进行预测。对于 BERT 模型的输

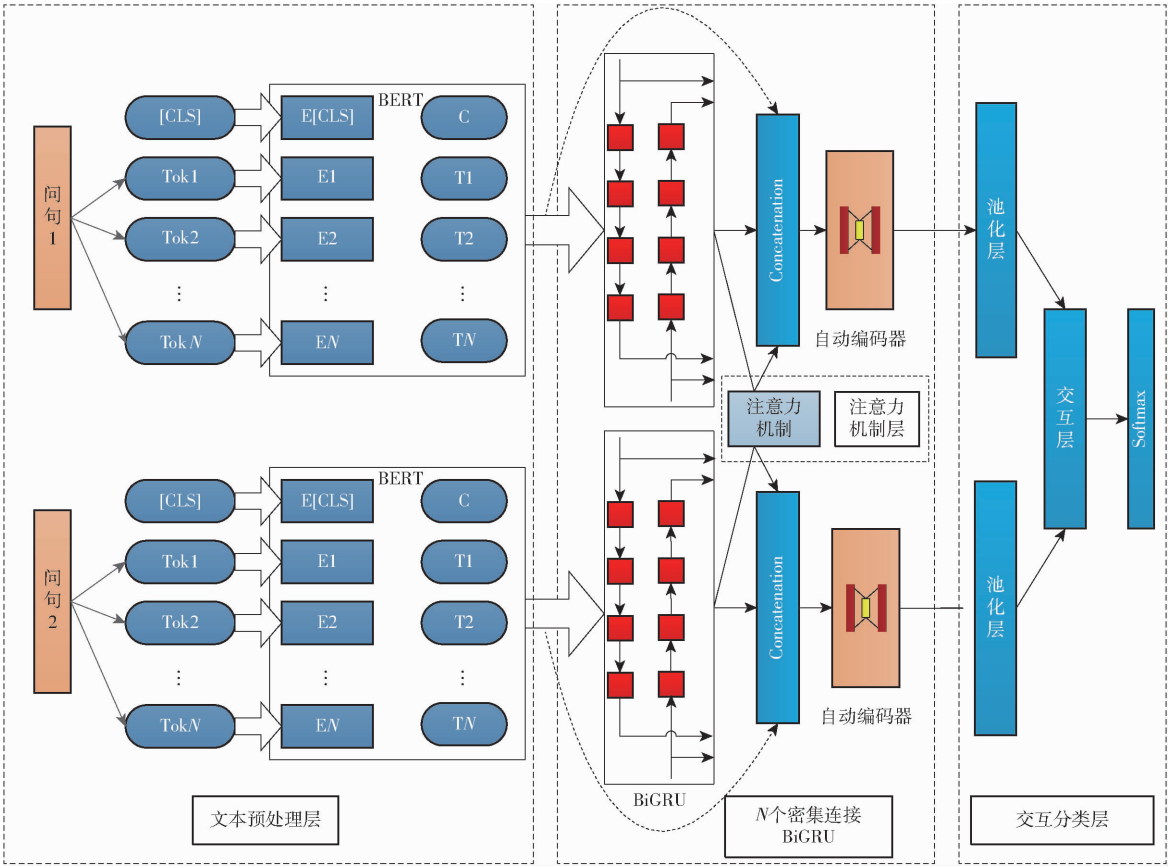


图 1 模型架构图

Fig. 1 Model architecture diagram

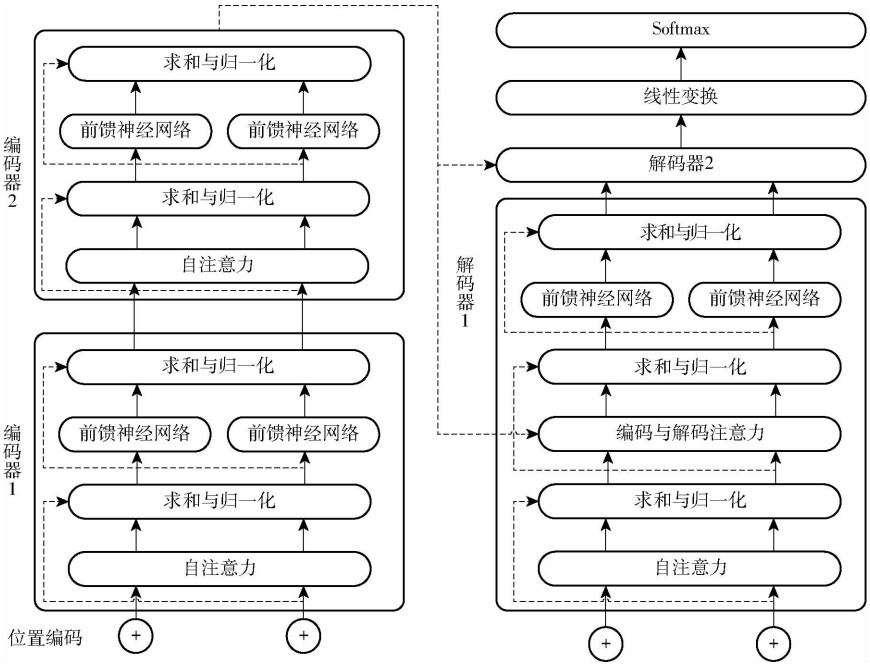


图 2　Transformer 模型结构

Fig. 2　Transformer model structure

入，每个词语的表示都由词语向量、段向量和位置向量共同组成，其中，标记[CLS]代表句子的开始，标记[SEP]代表句子的结束，如图 3 所示。

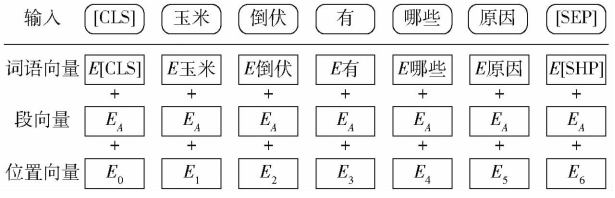


图 3　BERT 输入示例

Fig. 3　BERT input example

使用哈尔滨工业大学开发的语言技术平台(LTP)工具[29]作为分词工具。它会屏蔽组成同一个单词的所有汉字，然后训练模型以获得分段单词。

2.2　密集连接 BiGRU

使用双向 GRU 获取文本的特征向量。GRU[30]是一种特殊的循环神经网络(RNN)，RNN 是一种对序列数据建模的神经网络。它以有序的方式向网络传输文字存储之前的单词信息，可以有效地解决长期依赖关系。然而 RNN 存在消失梯度问题，文献[14]针对上述问题提出了长期短期记忆网络(LSTM)改进了 RNN。LSTM 通过遗忘门和输入门持续更新内存内容，以便 LSTM 可以有效地获得长期依赖关系，解决梯度消失和梯度爆炸问题，文献[24]提出了门控循环单元网络(GRU)，将 LSTM 的输入门和遗忘门合并形成了更新门，形成了更加流线型的门结构。

使用 BERT 预训练模型获得的词向量作为 BiGRU 层的输入。BiGRU 层包含 2 部分，同时读取向前和向后方向的单词向量。然后 GRU 计算传递的向量并输出固定维度的向量。GRU 包括 4 部分：

(1)重置门。GRU 使用重置门进行选择在前一刻要放弃哪些信息。

(2)更新门。GRU 通过更新门选择并更新当前时刻的信息，重置门和更新门计算公式为

$$R_t = \sigma(\boldsymbol{W}_r \boldsymbol{S}_t + \boldsymbol{U}_r \boldsymbol{h}_{t-1} + \boldsymbol{B}_r) \tag{1}$$

$$Z_t = \sigma(\boldsymbol{W}_z \boldsymbol{S}_t + \boldsymbol{U}_z \boldsymbol{h}_{t-1} + \boldsymbol{B}_z) \tag{2}$$

式中　$\boldsymbol{W}_r$、$\boldsymbol{U}_r$、$\boldsymbol{W}_z$、$\boldsymbol{U}_z$——权重

R_t——t 时刻的重置门

$\boldsymbol{h}_{t-1}$——$t-1$ 时刻的隐藏状态

$\boldsymbol{B}_r$、$\boldsymbol{B}_z$——偏差

σ——Sigmoid 函数

$\boldsymbol{S}_t$——输入的词向量

Z_t——t 时刻的更新门

(3)GRU 计算候选内存内容，这是计算当前时刻输出的重要步骤，计算公式为

$$\hat{H}_t = \tanh(\boldsymbol{W}\boldsymbol{S}_t + \boldsymbol{U}R_t \boldsymbol{h}_{t-1} + \boldsymbol{B}) \tag{3}$$

式中　$\boldsymbol{W}$、$\boldsymbol{U}$——权重　　$\boldsymbol{B}$——偏差

$\hat{H}_t$——候选内存内容

(4)GRU 根据上述结果计算输出，计算公式为

$$H_t = (1 - Z_t)\hat{H}_t + Z_t H_{t-1} \tag{4}$$

式中　H_t、H_{t-1}——t、$t-1$ 时刻的输出

农业问句语义的表达与当前文字及其上下文均有一定的关联，由于 GRU 的单向特性，不能从后向前编码文本信息，会影响提取文本上下文特征信息。因此，本文采用了 BiGRU，而不是直接使用 GRU。BiGRU 通过 2 个子层从 2 个方向处理输入序列，以

获取完整的农业问句文本特征。这 2 个子层分别计算前向隐藏序列$\overrightarrow{h_t}$和后向隐藏序列$\overleftarrow{h_t}$。然后将它们组合起来计算当前隐藏状态和 BiGRU 的输出。具体计算公式为

$$\overrightarrow{h_t}=\mathrm{GRU}(\boldsymbol{c}_{tij},\overrightarrow{h}_{t-1}) \tag{5}$$

$$\overleftarrow{h_t}=\mathrm{GRU}(\boldsymbol{c}_{tij},\overleftarrow{h}_{t-1}) \tag{6}$$

$$h_t=w_t\overrightarrow{h}_t+v_t\overleftarrow{h}_t+b_t \tag{7}$$

式中 w_t——t 时刻前向隐藏状态权重

v_t——t 时刻反向隐藏状态权重

b_t——t 时刻隐藏层的偏置

$\boldsymbol{c}_{tij}$——t 时刻第 i 个句子第 j 个词的向量

本层是该模型的关键所在,采用了多层 BiGRU 堆叠在一起的结构并循环了 5 次,在每个密集连接的 BiGRU 层都将之前层的输入和本层的输出合并之后向后传递,H_l 代表的是 BiGRU,l 表示的是 BiGRU 的层数,t 表示的是时刻,其隐藏状态值计算公式为

$$h_t^l=H_l(h_t^{l-1},h_{t-1}^l) \tag{8}$$

式中 h_t^l、h_{t-1}^l——t、$t-1$ 时刻 l 层的隐藏状态

H_l——第 l 层的 BiGRU

h_t^{l-1}——t 时刻 $l-1$ 层的隐藏状态

2.3 协同注意力机制

注意力机制在许多领域都取得了很好的应用效果,是一种有效学习上下文向量在特定序列上匹配的技术。给定 2 个问句,即在每个 BiGRU 层基于协同注意力机制的 2 句话,计算上下文向量。计算出的注意信息值代表了两句话之间的对齐关系。本文使用连接操作将注意力机制对 2 个问句交互的信息合并到密集连接的重复特征中。这种从最下层到最上层的密集连接特征所获得的串联循环和协同注意特征,丰富了词汇的语义特征。与密集连接的 BiGRU 隐藏特征类似,本文将注意力上下文向量与向量$\boldsymbol{h}_{hi}$连接,保留注意信息作为下一层的输入。对句子 Q 的第 i 个位置词的注意信息σ_{pi}计算为与$\boldsymbol{h}_{hi}$的加权和,其加权值为 Softmax 的权值,计算公式为

$$e_{i,j}=\cos(\boldsymbol{h}_{pi},\boldsymbol{h}_{hi}) \tag{9}$$

$$\alpha_{i,j}=\frac{\exp(e_{i,j})}{\sum_{k=1}^{J}\exp(e_{i,k})} \tag{10}$$

$$\sigma_{pi}=\sum_{j=1}^{J}\alpha_{i,j}\boldsymbol{h}_{hj} \tag{11}$$

式中 $e_{i,j}$——2 个问句向量的相似度

$\boldsymbol{h}_{pi}$、$\boldsymbol{h}_{hi}$——问句向量表示

$\cos(\cdot)$——相似度函数

$\alpha_{i,j}$——注意力权重

2.4 交互分类层

本文提出的模型使用所有层的输出作为一个语义知识的社区。然而,该网络是一种随着层的加深而增加输入特征的结构,并且具有大量的参数,尤其是在全连接层。为了解决这个问题,使用了一个自动编码器作为减少特征数量同时能维持原有信息的结构。此外,该部分在试验中作为一个正则化,从而提高了测试性能。

为了提取每个句子的适当表示,对密集连接的 BiGRU 和注意特征采用逐级最大池化运算。具体来说,如果最终 BiGRU 层的输出是句子 30 个单词的 100 维向量,得到了一个 30×100 的矩阵,使得合成的向量 $\boldsymbol{p}$ 或 $\boldsymbol{q}$ 的维度为 100。然后,在交互层中以各种方式对 2 个句子 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 的表示形式 $\boldsymbol{p}$ 和 $\boldsymbol{q}$ 进行聚合,最终得到语义句匹配的特征向量 $\boldsymbol{v}$,计算公式为

$$\boldsymbol{v}=[\boldsymbol{p};\boldsymbol{q};\boldsymbol{p}+\boldsymbol{q};\boldsymbol{p}-\boldsymbol{q};|\boldsymbol{p}-\boldsymbol{q}|] \tag{12}$$

这里,从元素的角度执行操作 +、-、|·|来推断 2个句子之间的关系。元素相减 $\boldsymbol{p}-\boldsymbol{q}$ 是用于单向类型任务的非对称操作符。提取完特征 $\boldsymbol{v}$ 之后使用了 2 个全连接层,激活函数为 ReLU,后面是一个完全连接的输出层。最后应用 Softmax 函数得到概率分布。

3 试验与结果分析

3.1 评价指标

本研究使用 Pytorch 深度学习框架构建神经网络。试验把 17 000 个问题对按 9∶1的比例划分为训练集和测试集,使用随机梯度下降算法对模型权重进行更新,训练集共 15 300 条,测试集 1 700 条,本文以精确率、召回率和 F1 值作为评价指标。

3.2 文本向量化处理与分析

采用 12 层的中文 BERT 模型对农业问句文本数据进行向量化处理。同时与 Glove[31]、TF - IDF[32]、Word2Vec 向量化模型进行对比分析。对 4 种模型训练得到的文本特征通过一个全连接层直接输入到 Softmax 分类器中,由表 2 可知,在嵌入层使用 4 种不同词向量转换工具,BERT 预训练模型取得最高的精确率和 F1 值,分别达到 85.3%、81.2%,TF - IDF 方法的效果最差,这是由于 TF - IDF 主要考虑词频的重要性,忽略了词与词的位置信息以及词与词之间的相互关系。而经过 BERT 预训练模型方法比 Word2Vec 精确率和 F1 值分别提高 2.7、3.5 个百分点,这是因为 Word2Vec 虽然考虑了词的周边信息,但是却忽略了词序问题,以及受限于窗口尺寸的限制,不能考虑整个句子中所有词

的相关性。而经过 BERT 模型在农业问句语料库上预训练得到文本特征后，实现了同时考虑上下文及词序信息，提高了神经网络的精确率和 F1 值，说明 BERT 可以很好地解决词语在不同语境下具有不同含义的问题。因此本文采用 BERT 预训练模型，将农业问句转换为词向量输入到神经网络模型中。

表 2　不同嵌入层下模型匹配效果

Tab. 2　Model matching effect under different embedding layers　%

向量化模型	精确率	召回率	F1 值
TF－IDF	72.7	63.7	67.9
Glove	81.8	72.6	76.9
Word2Vec	82.6	73.3	77.7
BERT	85.3	77.6	81.2

3.3　参数设置

本文设置模型迭代训练次数为 50，Batch-Size 设置为 64，设置密集连接的 BiGRU 循环 5 次，每个密集连接的 BiGRU 有 100 个隐藏单元，将全连接层的隐藏单元设置为 1 000。在单词和字符嵌入层之后，将 Dropout 设置为 0.5。对于自动编码器，设置 200 个隐藏单元作为自动编码器的编码特征，Dropout 设置为 0.2。本文使用初始学习率为 0.001 的 RMSProp 优化器。所有权值都被 L2 正则化约束，正则化常数 $\lambda = 10^{-6}$。

3.4　试验结果与分析

使用本文提出的 BERT－Attention－DenseBiGRU 模型与其他 6 种文本相似度匹配模型：BiLSTM[33]、Selfattention－BiLSTM[34]、TextCNN[35]、ABCNN[36]、BiGRU、Attention－BiGRU[37] 在农业问句相似对文本数据集上进行对比试验，均使用 12 层的中文 BERT 模型对文本进行向量化，表 3 展示了 7 种不同深度学习模型在精确率、召回率、F1 值的比较。基于自注意力机制的 BiLSTM 相比于 BiLSTM 的精确率和 F1 值分别提高 4.7、4.4 个百分点，基于注意力机制的卷积神经网络模型（ABCNN）的精确率和 F1 值相比于 TextCNN 模型精确率和 F1 值提高 3.2、3.9 个百分点，基于注意力机制的 BiGRU 相比于 BiGRU 模型的精确率和 F1 值提高了 2.6、1.9 个百分点。通过 3 组对比试验可以看出，加入注意力机制的深度学习模型可以明显提升模型效果，这是由于注意力机制在农业问句关键信息上分配更多的权重、突出局部的重要信息，有利于精确提取文本特征。本文提出的模型 Attention－DenseBiGRU 获得了最高的精确率和 F1 值，达到 97.2% 和 97.6%，精确率、召回率、F1 值明显优于其他 6 种深度学习模型，相比于 Attention－BiGRU 的精确率和 F1 值分别提高 5.5、5.3 个百分点，原因为：①通过密集连接的 BiGRU 可以加强特征的传递和提取，减少特征损失。②Attention－BiGRU 只是在网络后面加入注意力机制，而本文提出的 Attention－DenseBiGRU 是在编码层通过注意力机制增加 2 个网络之间的交互作用，从而提取句子对之间的特征关联来提高文本匹配的精确率。

表 3　不同模型在农业问句数据集上的效果

Tab. 3　Effects of different models on agricultural question data sets　%

模型	精确率	召回率	F1 值
BiLSTM	86.5	88.6	87.5
Selfattention－BiLSTM	91.2	92.6	91.9
TextCNN	87.4	86.7	87.0
ABCNN	90.6	91.3	90.9
BiGRU	89.1	91.7	90.4
Attention－BiGRU	91.7	92.9	92.3
Attention－DenseBiGRU	97.2	98.1	97.6

表 4 展示了 7 种神经网络模型在 BERT 文本向量化表示方法和 Word2Vec 文本向量化表示方法下的农业问句相似度匹配精确率。本文提出的 BERT 文本向量化表示方法在 7 种神经网络模型上精确率均高于 Word2Vec 文本向量化表示方法。Attention－DenseBiGRU 模型在 BERT 文本向量化表示和 Word2Vec 文本向量化表示方法下均取得了最高的精确率，分别达到 97.2% 和 94.3%，问句相似度匹配效果明显优于其他 6 种神经网络模型。从表 4 可以看出，BERT 文本向量化表示方法在每组对比试验中都提高了精确率，这是由于 Word2Vec 文本向量化表示方法忽略了不同语境下多义词的问题以及长距离语义关联信息，而 BERT 文本向量化表示方法可以很好地解决上述问题，从而提高问句相似度匹配的精确率。

表 4　不同模型在 BERT 和 Word2Vec 文本向量化表示方法下问句相似度匹配精确率

Tab. 4　Question similarity matching precision of different models under BERT and Word2Vec text vectorization representation methods　%

模型	Word2Vec	BERT
BiLSTM	83.6	86.5
Selfattention－BiLSTM	87.5	91.2
TextCNN	82.1	87.4
ABCNN	87.5	90.6
BiGRU	86.5	89.1
Attention－BiGRU	88.7	91.7
Attention－DenseBiGRU	94.3	97.2

由表 5 可知，与 BiLSTM、Selfattention－BiLSTM、TextCNN、ABCNN、BiGRU、Attention－BiGRU 6 种文

本相似度计算模型相比,Attention - DenseBiGRU 在病虫草害、动物疫病、栽培管理、市场销售、养殖管理、土壤肥料、储运保鲜、其他共 8 个类别的农业问句对数据集上均具有最高的匹配精确率,对农业问句相似对匹配的精准率大于 93.7%,整体效果明显优于其他模型。在病虫草害、栽培管理 2 个类别试验数据量充足的数据集中,本文模型的精确率达到 99.1%、98.7%,明显高于其他 6 种深度学习模型。这是因为深度学习模型在不断迭代训练的过程中,数据集越大,模型的训练效果越好;在市场营销、土壤肥料、储运保鲜 3 个类别数据量较少的数据集中,本文模型的精确率分别达到 97.3%、96.6%、93.7%,明显高于其他模型,说明 Attention - DenseBiGRU 模型在数据量不充足的情况下,仍能够有效提取短文本的特征进行文本相似度计算,也说明了该模型具有很好的鲁棒性。

表 5 不同模型在农业相似问句数据集不同类别的精确率

Tab. 5 Precision of different models in different categories of agricultural similar question data sets %

模型	病虫草害	栽培管理	动物疫病	市场销售	养殖管理	土壤肥料	储运保鲜	其他
BiLSTM	90.1	93.7	82.5	89.7	79.7	87.5	79.1	89.7
Selfattention - BiLSTM	94.5	92.7	91.9	92.1	91.6	91.9	89.5	85.5
TextCNN	94.7	89.7	86.5	81.1	88.4	89.2	80.2	89.1
ABCNN	96.6	96.9	95.3	92.6	90.3	95.3	89.2	92.6
BiGRU	92.3	91.5	87.7	88.7	87.7	87.7	85.2	92.3
Attention - BiGRU	93.3	92.0	89.9	93.3	89.9	89.9	92.0	93.3
Attention - DenseBiGRU	99.1	98.7	98.1	97.3	97.1	96.6	93.7	97.3

通过一组试验来证明本文所提出 Attention - DenseBiGRU 模型中每个模块的有效性。首先将模型中自编码器(autoencoder)删除后得到模型 2,由表 6 可以看出,模型 2 的精确率和 F1 值与模型 1 相比下降了 2、1.7 个百分点,验证了自动编码器的有效性。自动编码器可以降低文本特征维度,有效提高文本相似度匹配效果。然后分别删除 BiGRU 之间的密集连接和注意力机制,得到模型 3、4,可以看出模型 3、4 的精确率和 F1 值与模型 1 相比分别下降了 2.9、2.9 个百分点和 2.1、1.8 个百分点,表明 BiGRU 之间的密集连接比协同注意力机制更能提高模型的效果。模型 5、6 分别是基于注意力机制的 5 层普通连接 BiGRU 模型和不具有注意力机制的 5 层普通连接 BiGRU 模型,表 6 中可以看出注意力机制可以提高模型的效果,这是因为注意力机制可以强化关键词在问句相似度匹配中的权重,提升文本匹配精确率。

表 6 不同模型在农业问句数据集上的效果

Tab. 6 Effects of different models on agricultural question data sets %

模型序号	模型名称	精确率	召回率	F1 值
1	Attention - DenseBiGRU	97.2	98.1	97.6
2	删除 autoencoder	95.2	96.7	95.9
3	删除 Dense(Att)	94.3	95.1	94.7
4	删除 Dense(Rec)	95.1	96.5	95.8
5	BiGRU + Attention	91.7	92.9	92.3
6	BiGRU	89.1	91.7	90.4

表 7 展示了模型 1 ~ 6 在不同层数 BiGRU 时农业问句数据集上的分类效果,由表 7 可知,Attention - DenseBiGRU 在 5 层 BiGRU 时精确率达到最高的 97.2%,这是由于通过层数的增加和密集连接,可以有效提高文本特征的提取,减少特征丢失,提高文本匹配效率。模型 5、6 在第 2 层时精确率达到最高,之后开始逐渐下降。这可能是因为没有通过密集连接的 BiGRU 在提取特征时会导致特征丢失。进一步验证了通过深层次的神经网络可以有效获取文本特征及表示,有利于提升文本相似度匹配效果。

表 7 不同条件下模型在每层网络的匹配精确率

Tab. 7 Matching effect of model in each layer of network under different conditions %

序号	模型	1	2	3	4	5
1	Attention - DenseBiGRU	92.1	93.8	94.9	96.2	97.2
2	删除 autoencoder	91.9	92.3	92.5	93.8	95.2
3	删除 Dense(Att)	91.8	91.9	92.1	92.5	94.3
4	删除 Dense(Rec)	91.7	92.1	92.8	93.9	95.1
5	BiGRU + Attention	90.1	91.7	90.8	89.6	89.1
6	BiGRU	87.2	89.1	88.7	87.9	86.6

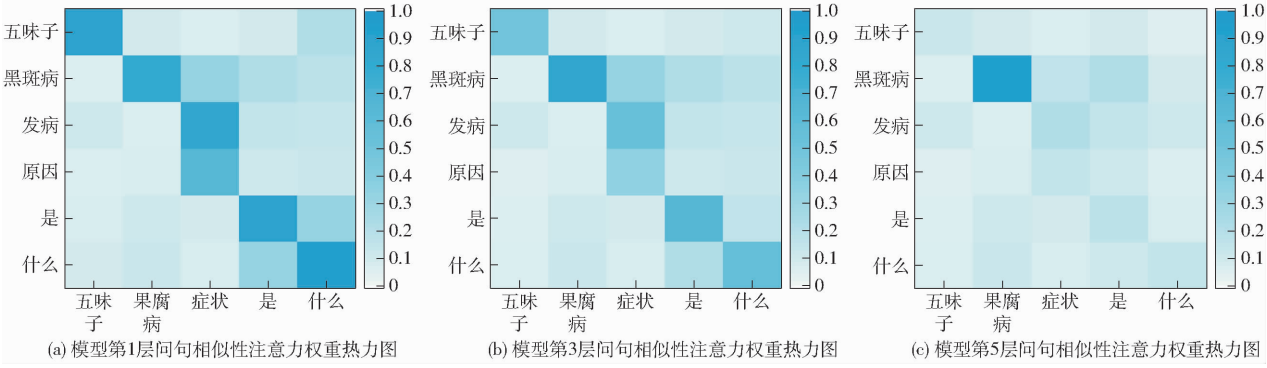
表 8 展示了基于 Attention 的 4 种深度学习模型在 1 700 对农业问句对测试集上的反应时间和精确率,ABCNN 速度最快,这是由于 ABCNN 模型主要采用卷积神经网络作为基础,模型结构简单,参数较少。本文提出的 Attention - DenseBiGRU 模型 7 s 可以完成 1 700 对农业问句相似性精准判断,精确率达到最高的 94.7%,满足了对农业问句对语义相似度快速、精确匹配的要求,在反应时间接近的情况下,Attention - DenseBiGRU 模型在文本语义相似度匹配精确率上取得了最好的效果。

表 8　4 种深度学习模型的反应时间和精确率

Tab. 8　Response time and precision of four deep learning models

测试模型	反应时间/s	精确率/%
ABCNN	6	92.8
Selfattention－BiLSTM	11	91.7
Attention－BiGRU	9	90.9
Attention－DenseBiGRU	7	94.7

通过密集连接 BiGRU 网络和注意力机制得到的特征是通过最大池化层连接到分类层，使每层的特征都影响损失函数，并进行深度监督学习。因此，本文使用注意力权重和最大池化位置来解释分类结果。注意力权重包含了两个句子对齐的信息，同时每个维度上最大池化位数在分类中起着重要作用。图 4 展示了本文提出的模型在不同层时注意力权重热力图，例如问句 1：五味子黑斑病发病原因是什么？问句 2：五味子果腐病症状是什么？除了黑斑病和果腐病外，问句 1 和问句 2 的大部分词同时存在。在第 1 层的注意力权重图中，每句话中相同或相似词的对应度较高。但是，随着层次的加深，黑斑病和果腐病的注意力权重在不断加深。因为黑斑病和果腐病虽然都是病虫害名称，但是在本质上有明显区别，在第 5 层时，问句 1 和问句 2 中除了黑斑病和果腐病，对应的其他词的注意权重都变得极小，由于“黑斑病”和“果腐病”2 个词在语义上区别明显，因此模型判断问句对为语义不相似，即标签为 0。



(a) 模型第1层问句相似性注意力权重热力图　(b) 模型第3层问句相似性注意力权重热力图　(c) 模型第5层问句相似性注意力权重热力图

图 4　农业问句相似度注意力权重可视化图

Fig. 4　Visualizations of similarity and attention weight of agricultural questions

4　结论

（1）为了解决中国农业技术推广问答社区无法自动准确地发现重复性语义问题的问题，构建了一个包含 8 个类别 17 000 对农业问句相似对语料库。针对农业相关问句的特点，提出了一种基于 BERT 的 Attention－DenseBiGRU 模型。

（2）利用基于协同注意机制的密集连接 BiGRU 模型对农业问答社区问句文本进行快速自动的重复语义检测。

（3）利用 DenseBiGRU 网络提取农业问句文本的特征表达用于问句相似度匹配。在此基础上，对其重要结构参数和训练策略进行了优化和改进，构建了基于协同注意力机制的农业文本相似度匹配算法，实现了问答社区中相似语义问句精确高效识别。与其他模型相比，该模型在农业问句相似对数据集上取得了最高的精确率和 F1 值，分别达到 97.2% 和 97.6%。

参　考　文　献

[1] LI M, LI Y, PENG Q, et al. Evaluating community question-answering websites using interval-valued intuitionistic fuzzy DANP and TODIM methods[J]. Applied Soft Computing, 2020, 99: 106918.

[2] SHEN H, LIU G, WANG H, et al. Social Q&A: an online social network based question and answer system[J]. IEEE Transactions on Big Data, 2016, 3: 91－106.

[3] YOGATAMA D, SMITH N A, Linguistic structured sparsity in text categorization[C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2014: 786－796.

[4] AHMED W, DASAN A, BABU A P. Developing an intelligent question answering system[J]. International Journal of Education and Management Engineering, 2017, 7: 50.

[5] LIU Y, TANG A, SUN Z, et al. An integrated retrieval framework for similar questions: word-semantic embedded label clustering-LDA with question life cycle[J]. Information Sciences, 2020, 537: 227－245.

[6] CHEN Y, ZHANG Z. Research on text sentiment analysis based on CNNs and SVM[C]// Proceedings of 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2018: 2731－2734.

[7] LECUN Y, BENGIO Y, HINTON G, et al. Deep learning[J]. Nature, 2015, 521(7553): 436－444.

[8] STRUBELL E, GANESH A, MCCALLUM A. Energy and policy considerations for deep learning in NLP[J]. arXiv preprint arXiv:1906.02243, 2019.

[9] KALCHBRENNER N, GREFENSTETTE E, BLUNSOM P. A convolutional neural network for modelling sentences[J]. arXiv preprint arXiv:1404.2188,2014.
[10] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078,2014.
[11] HUANG P S, HE X, GAO J, et al. Learning deep structured semantic models for web search using clickthrough data[C]// Proceedings of the 22nd ACM International Conference on Information & Knowledge Management,2013:2333 - 2338.
[12] SHEN Y, HE X, GAO J, et al. A latent semantic model with convolutional-pooling structure for information retrieval[C]// Proceedings of the 23rd ACM International Conference on Information and Knowledge Management,2014:101 - 110.
[13] PALANGI H, DENG L, SHEN Y, et al. Semantic modelling with long-short-term memory for information retrieval[J]. arXiv preprint arXiv:1412.6629,2014.
[14] GREFF K, SRIVASTAVA R K, KOUTNÍK J, et al. LSTM: a search space odyssey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 28:2222 - 2232.
[15] LIU J, YANG Y, LV S, et al. Attention - based BiGRU - CNN for Chinese question classification[J/OL]. Journal of Ambient Intelligence and Humanized Computing, 2019. https://doi.org/10.1007/s12652 - 019 - 01344 - 9.
[16] MUELLER J, THYAGARAJAN A. Siamese recurrent architectures for learning sentence similarity[C]// Proceedings of the AAAI Conference on Artificial Intelligence,2016:2786 - 2792.
[17] WANG H, ZHU H, WU H, et al. A densely connected GRU neural network based on coattention mechanism for Chinese agriculture-related question similarity matching[J]. Agronomy, 2021,11(7):1307.
[18] WANG X, WANG H, ZHAO G, et al. ALBERT over match - LSTM network for intelligent questions classification in Chinese [J]. Agronomy, 2021, 11(8):1530.
[19] 王郝日钦,吴华瑞,冯帅,等. 基于 Attention_DenseCNN 的水稻问答系统问句分类[J]. 农业机械学报,2021,52(7):237 - 243.
WANG Haoriqin,WU Huarui,FENG Shuai,et al. Classification technology of rice questions in question answer system based on Attention_DenseCNN[J]. Transactions of the Chinese Society for Agricultural Machinery,2021,52(7):237 - 243. (in Chinese)
[20] 金宁,赵春江,吴华瑞,等. 基于 BiGRU_ MulCNN 的农业问答问句分类技术研究[J]. 农业机械学报,2020, 51(5):199 - 206.
JIN Ning, ZHAO Chunjiang, WU Huarui, et al. Classification technology of agricultural question based on BiGRU_MulCNN [J]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(5):199 - 206. (in Chinese)
[21] 张明岳,吴华瑞,朱华吉. 基于卷积模型的农业问答语性特征抽取分析[J]. 农业机械学报,, 2018, 49(12):203 - 210.
ZHANG Mingyue, WU Huarui, ZHU Huaji. Analysis of extraction of semantic feature in agricultural question and answer based on convolutional model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018,49(12):203 - 210. (in Chinese)
[22] 赵明,董翠翠,董乔雪,等. 基于 BIGRU 的番茄病虫害问答系统问句分类研究[J]. 农业机械学报, 2018,49(5):271 - 276.
ZHAO Ming, DONG Cuicui, DONG Qiaoxue, et al. Question classification of tomato pests and diseases question answering system based on BIGRU[J]. Transactions of the Chinese Society for Agricultural Machinery,2018,49(5):271 - 276. (in Chinese)
[23] 陈瑛,陈昂轩,董玉博,等. 基于 LSTM 的食品安全自动问答系统方法研究[J]. 农业机械学报,2019, 50(增刊):380 - 384.
CHEN Ying, CHEN Angxuan, DONG Yubo, et al. Methods of food safety question answering system based on LSTM[J]. Transactions of the Chinese Society for Agricultural Machinery,2019,50(Supp.):380 - 384. (in Chinese)
[24] 冯帅,许童羽,周云成,等. 基于深度卷积神经网络的水稻知识文本分类方法[J]. 农业机械学报,2021,52(3):257 - 264.
FENG Shuai, XU Tongyu, ZHOU Yuncheng, et al. Rice knowledge text classification based on deep convolution neural network[J]. Transactions of the Chinese Society for Agricultural Machinery,2021,52(3):257 - 264. (in Chinese)
[25] CUI Y, CHE W, LIU T, et al. Pre-training with whole word masking for Chinese bert[J]. arXiv preprint arXiv:1906.08101,2019.
[26] SADOWSKI C, LEVIN G. SimHash: Hash-based similarity detection[J/OL]. Technical Report, Google, 2007. https://www.webrankinfo.com/dossiers/wp-content/uploads/simhash.pdf.
[27] GOLDBERG Y, LEVY O, MIKOLOV, et al. Word2Vec explained: deriving negative-sampling word-embedding method[J]. arXiv preprint arXiv:1402.3722,2014.
[28] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of Advances in Neural Information Processing Systems, 2017:5998 - 6008.
[29] CHE W, LI Z, LIU T. LTP: a Chinese language technology platform[C]// Proceedings of Coling, 2010: 13 - 16.
[30] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555,2014.
[31] PENNINGTON J, SOCHER R, MANNING C D. Glove: global vectors for word representation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP),2014:1532 - 1543.
[32] QAISER S, ALI R. Text mining: use of TF - IDF to examine the relevance of words to documents[J]. International Journal of Computer Applications, 2018, 181: 25 - 29.
[33] XU G, MENG Y, QIU X, et al. Sentiment analysis of comment texts based on BiLSTM[J]. IEEE Access, 2019, 7:51522 - 51532.
[34] XIE J, CHEN B, GU X, et al. Self-attention-based BiLSTM model for short text fine-grained sentiment classification[J]. IEEE Access, 2019, 7:180558 - 180570.
[35] KIM Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv: 1408. 5882, 2014.
[36] YIN W, SCHÜTZE H, XIANG B, et al. ABCNN: attention - based convolutional neural network for modeling sentence pairs [J]. Transactions of the Association for Computational Linguistics, 2016, 4: 259 - 272.
[37] XIE J, CHEN B, GU X, et al. Self-attention-based BiLSTM model for short text fine-grained sentiment classification[J]. IEEE Access, 2019, 7:180558 - 180570.