

# 一种增强的多粒度特征融合语义匹配模型

尚福华, 蒋毅文\*, 曹茂俊

(东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318)

**摘要:**语义匹配作为自然语言处理任务中重要的一环,直接制约问答系统、信息检索等任务的效率。针对现有语义模型大多只以词为基本语义单元进行注意力交互,较少考虑中文中的词边界模糊和字符信息获取不足而带来的语言颗粒度对整体建模忽略的问题,提出一种增强的多粒度特征融合语义匹配模型 EMGFM。首先结合 BERT 模型和 word2vec 以获得增强的字符向量表示,然后从字、词、句三种粒度进行注意力的交互,并对交互结果进行加权融合,以突出不同交互信息对整体建模的贡献。为减少交互过程中产生的信息损失,通过构造差异性来对交互信息进行信息增强。最后通过最大池化、平均池化两种方式获得文本的最终语义表示以进行匹配度的计算。该模型在 CCKS 问句匹配大赛中文数据集上达到了 87% 的正确率,相比于一些语义匹配的经典模型准确率均有提升,证明该方法确实能有效提升问句语义匹配的准确性。

**关键词:**语义匹配;语言颗粒度;Siamese 网络;可分解注意力机制;BERT 模型

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2022)07-0028-06

doi:10.3969/j.issn.1673-629X.2022.07.005

## An Enhanced Multi Granularity Feature Fusion Model for Semantic Matching

SHANG Fu-hua, JIANG Yi-wen\*, CAO Mao-jun

(School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)

**Abstract:**As an important part of natural language processing tasks, semantic matching directly restricts the efficiency of question answering system, information retrieval and other tasks. Most of the existing semantic models only take words as the basic semantic unit for attention interaction, and less take into account the problem of language granularity ignoring the overall modeling caused by the fuzzy word boundary and insufficient acquisition of character information in Chinese. Therefore, an enhanced multi granularity feature fusion semantic matching model EMGFM is proposed. Firstly, the BERT model and word2vec are combined to obtain the enhanced character vector representation, then the attention interaction is carried out from the three granularity of words, phrases and sentences, and the interaction results are weighted fused to highlight the contribution of different interaction information to the overall modeling. In order to reduce the information loss in the interactive process, the interactive information is enhanced by constructing differences. Finally, the final semantic representation of the text is obtained by maximum pooling and average pooling to calculate the matching degree. The model achieves 87% accuracy on the Chinese data set of CCKS question matching competition. Compared with some classical models of semantic matching, the accuracy is improved. It proves that the proposed method can effectively improve the accuracy of question semantic matching.

**Key words:**semantic matching; language granularity; Siamese network; decomposable attention mechanism; BERT model

## 0 引言

文本语义匹配是目前自然语言处理领域中一个重要的任务分支,也是自然语言处理中重要的研究方向和研究热点,在信息检索、问答系统、机器翻译等任务中起到了核心作用。

正因为语义匹配的重要性,国内外许多学者都在语义匹配上作了许多研究。传统的匹配模型主要基于人工提取的特征以及基于统计机器学习的方法,如 BM25 模型、PLS、SMT 模型、WTM 模型等。这些模型主要解决文本表层的匹配问题<sup>[1]</sup>,但难以捕捉文本所

收稿日期:2021-07-28

修回日期:2021-11-30

基金项目:黑龙江省自然科学基金(LH2019F004);东北石油大学青年科学基金(2018QNL-25);东北石油大学优秀中青年科研创新团队(KYCXTD201903)

作者简介:尚福华(1962-),男,博士,教授,CCF 会员(07338S),研究方向为人工智能、图像处理、油田智能软件;通信作者:蒋毅文(1997-),男,硕士生,研究方向为知识图谱、智能问答。

表达的深层含义。随着神经网络的热火发展,人们发现基于深度学习的模型在语义匹配任务中取得了更好的效果。因此关于语义匹配的研究热点已逐渐转移到深度学习语义匹配模型。庞亮等人<sup>[2]</sup>将目前的基于深度学习的语义匹配模型分为三种,第一种是基于单语义文档表达的,主要将待匹配的两个文本通过深度学习的方式表达成两个向量,再通过计算两个向量之间的相似度便可得到匹配度。如 ARC-I、DSSM 模型。第二种是基于多语义文档表达的,综合考虑文本的局部性表达(词粒度信息、短语粒度信息)和全局性表达(句粒度信息),通过这样多层次多粒度的匹配可以很好地弥补基于单语义文档表达的深度学习模型在压缩整个句子过程中的信息损失,而达到更好的效果。如 MultiGranCNN、MV-LSTM 等。第三种则是直接建模匹配模式的。这种模型从多角度提取文本间的深层交互信息,并认为对文本间的交互学习应该早于文本的表示学习之前,一方面是为了避免损失信息,另一方面通过这种方式能够得到更丰富的语义表示,利于后面的学习。典型模型有 ARC-II、ABCNN、BiMPM、ESIM、DIIN 等。Pair-CNN 模型是文献[3]所提出的,利用 CNN 完成短文本的匹配任务。ABCNN<sup>[4]</sup>则是在此基础上加入注意力机制形成的。文献提出双向多视角匹配模型 BIMPM<sup>[5]</sup>模型,对比了多种不同注意力机制的匹配策略。ESIM<sup>[6]</sup>模型本身是用于自然语言推理任务,但稍加改造后就能用于文本匹配等任务。

目前这些模型表现优异,但仍存在一定的不足。首先,这些模型大多是以词作为语义单元并基于静态词向量(Word2Vec、GloVe 等)来获取句子表示,对于英语来说,它的最小语义单元是单词,因此适用于这样的方法。而中文语言显然更加复杂,中文分词、语义信息等要素直接影响了语义匹配的效果,而汉语中单个字也承载着一定的语义<sup>[7]</sup>,因此只考虑词作为语义单元不一定能取得更好的效果,需要考虑多种语义单元在句子建模中的作用。此外静态词向量一旦训练完成,相同的词的向量表示是不会再随语境变化而变化的,因此这一类词向量难以解决一词多义问题<sup>[8]</sup>,且不能表征词性等语法特征。因此如何获得更准确的句子嵌入表示仍是现在的研究重点。其次,模型在交互时往往都是在词级别上进行交互,如 ESIM 模型在提取文本的交互特征时只是基于词进行注意力交互,未能挖掘文本更深层的交互信息。而句子是具有一定的层级关系的,如果忽略了语言颗粒度对句子建模的影响,那么得到的交互信息也是不充分的。最后,一些模型通过使用多种注意力机制来提升模型效果,如 BIMPM 采用了四种注意力匹配策略来提取更多的信息,但是对于多种交互结果往往只是采用简单的 concat 方式。

对于如何正确组合优化多种注意力信息考虑甚少。

针对以上问题,该文提出一种增强的多粒度特征融合语义匹配模型 EMGMF。主要工作如下:

(1)为避免由于中文词边界模糊而带来的影响,同时考虑中文中单个字所包含的语义信息,该文将字向量与词向量进行混合使用,在字向量的获得上,利用 BERT 模型强大的语义表达能力获取中文字符向量,从而更好地表征字符的多义性,而词向量选用 Word2Vec 进行获取。最后将两种向量进行拼接得到增强的字符向量。(2)为解决只考虑词交互带来的交互特征单一问题,对文本的不同的粒度(字、词、句)进行注意力交互特征提取,因为不同粒度的信息提取关注重点各有不同,为了综合利用多种方法,最大程度上完全提取信息,设计了一种融合多种交互信息的方法。该方法利用注意力思想衡量每种交互信息的重要性,而且不会因为直接连接导致向量维度过高,可以有效减少模型计算的复杂性。并通过原始信息与融合后的信息进行点乘相减等操作来构造差异性,以达到语义增强的效果。(3)针对语义匹配提出了增强的多粒度特征融合语义匹配模型,实验结果表明,提出的方法能有效提升问句语义匹配的准确性。

## 1 增强的多粒度特征融合语义匹配模型

### 1.1 问题定义与分析

给定待匹配样本  $\text{Sample} = \{H, T, \text{similarity}\}$ ,其中  $H$ 、 $T$  为待匹配的两个短文本,  $\text{similarity}$  为两个文本之间的相似度,一般为 0 或者 1。当  $\text{similarity} = 1$  时表示两个文本语义相似,  $\text{similarity} = 0$  则反之。该文的目标就是训练一个语义匹配模型  $M$ ,使得对任意的两个句子,  $M$  能够合理给出其相似度的评判。

### 1.2 模型结构

整个模型结构如图 1 所示,模型主要包含编码层、交互层、语义组合层、输出层 4 个主要部分。

首先在嵌入层将待匹配的两段文本转换为文本向量,在交互层利用可分解注意力机制,从三种粒度(字、词、句)分别计算两段文本的注意力权重,从而得到两段文本在不同粒度下的注意力加权表示。然后对不同粒度下的注意力加权表示进行融合,并利用原始信息对其进行信息增强,最后经过平均池化和最大池化处理后传入一个多层感知机进行输出。

#### 1.2.1 编码层

(1)BERT 预训练模型。

BERT 是一个语言表征模型(language representation model)<sup>[9]</sup>,基于多层 Transform 实现,通过 Transform 结构获取到字符向量可以充分利用上下文信息,更能表达多义性。因此,选用 BERT 预训练模

型来获取文本的字向量表示。

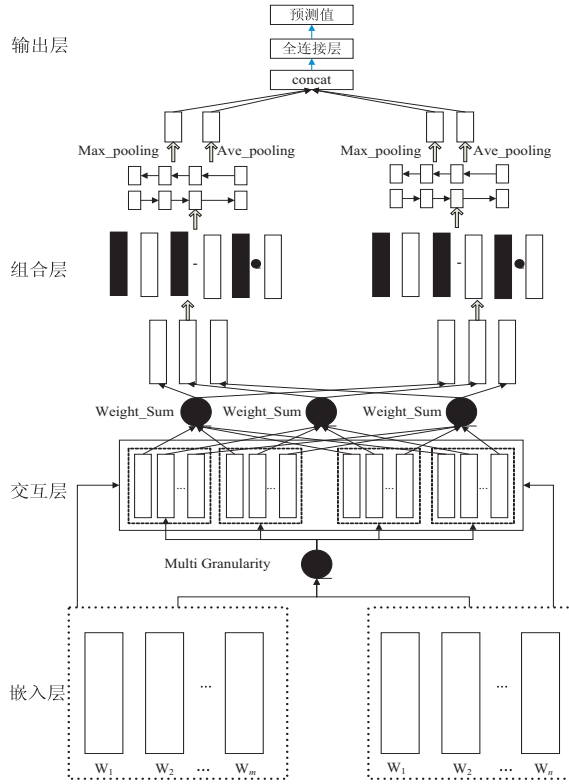


图 1 模型结构

## (2) 输入表示。

与其他传统词向量模型输入不同,BERT 模型对

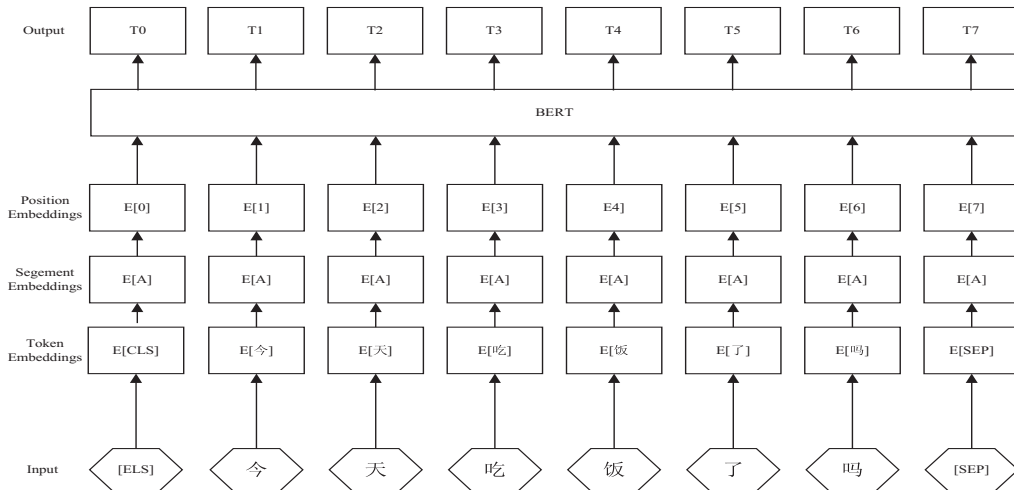


图 2 BERT 输入表示

## (1) 可分解注意力机制。

综上所述,该文在将特征向量送入后续网络之前,参考可分解注意力机制<sup>[11]</sup>的做法,对句子进行不同粒度的对齐,使得句子间的关系特征得以保留。以下先介绍可分解注意力的计算过程:

Step1:首先对句子 A、B 中的每个词,通过  $F$  函数计算它们的注意力权重矩阵  $e_{ij}$ ,  $F$  函数选用的是向量的点积。

$$e_{ij} = F(\rightarrow_{a_i}, \rightarrow_{b_j}) = \rightarrow_{a_i} \cdot (\rightarrow_{b_j})^T \quad (4)$$

中文和英文的处理都是基于字的,其输入输出如图 2 所示。为对 BERT 输出的字符向量进行语义增强,受李伟康等人<sup>[7]</sup>研究的启发,采用分词工具 jieba 对输入的文本进行分词,并使用 Word2Vec 对分词结果进行编码,并将字向量与词向量进行拼接。用  $v_i^c$  表示一个句子中第  $i$  个位置的字的初始字向量,  $v_i^w$  表示该句中该字所在词的词向量,计算过程如下式所示:

$$v_i^e = \text{concat}(v_i^c, v_i^w) \quad (1)$$

经过上述操作,最终将输入的句子 A、B 嵌入表示为:

$$E^A = [T_1^A, T_2^A, \dots, T_m^A] \quad (2)$$

$$E^B = [T_1^B, T_2^B, \dots, T_n^B] \quad (3)$$

其中,  $T_i^A, T_i^B \in R^{1 \times d}$ ,  $d$  为字向量的维度,  $m$ 、 $n$  分别为句子 A、B 的长度。

## 1.2.2 交互层

大量的实验证明在句子特征提取完成后才进行句子之间的交互,势必会浪费一些特征信息,而对待匹配的文本先进行交互再进行后续学习会取得更好的效果,并且更丰富的语义交互能为后续网络学习带来更多信息,匹配效果更好。如吴少洪等人<sup>[10]</sup>从不同的方面提取文本更深层的交互信息,取得了不错的效果。受这些工作的启发,增加更细粒度的文本交互,旨在获得更丰富的语义交互信息。

Step2:使用注意力权重矩阵对句子  $a$ 、 $b$  中的每个词进行注意力加权求和。

$$\rightarrow_{a_i} = \sum_{j=1}^{l_b} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_b} \exp(e_{ik})} \rightarrow_{b_j}, \forall i \in [1, \dots, l_a] \quad (5)$$

$$\rightarrow_{b_j} = \sum_{i=1}^{l_a} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_a} \exp(e_{kj})} \rightarrow_{a_i}, \forall j \in [1, \dots, l_b] \quad (6)$$

## (2) 字粒度的交互。

通过 1.2.1 的操作即可将输入的两个句子 A 和 B

各自转化为  $m * n$  的嵌入矩阵  $E^A$ 、 $E^B$ , 其中  $m$  为句子的字数,  $n$  为字向量的维度。因为 BERT 是基于字的模型, 因此输出即是每个字的字向量, 因此可直接对其进行字粒度的交互, 使用上文提及的注意力交互方式可得到关于句子 A 和 B 的基于字粒度的交互矩阵 CIA、CIB。

### (3) 短语粒度的交互。

大量实验证明长短语级别和短语级别的特征对句子的建模十分关键, CNN 因其独特的网络结构在捕捉句子的局部特征上具有天然优势。参考文献[12]中的做法, 该文在抽取局部特征时兼顾不同数量字组成的短语, 选择二元(two-gram)、三元(tri-gram)进行特征提取。具体做法如下:

#### Step1: 不同长度短语特征提取。

将 1.2.1 得到的句子嵌入矩阵 A、B 分别输入卷积神经网络进行卷积操作, 卷积核的宽度与字嵌入维度一致。为保证卷积操作前后维度保持一致以便于后续网络学习, 对 A、B 进行零向量填充后再进行卷积, 以二元提取特征为例, 卷积核的高度设置为 2, 得到的卷积结果可以表示为:

$$\text{Two\_C}^A = \text{two\_gram\_CNN}(A) \quad (7)$$

$$\text{Two\_C}^B = \text{two\_gram\_CNN}(B) \quad (8)$$

其中,  $C^A \in \mathbb{R}^{l_a \times 1 \times K}$ ,  $C^B \in \mathbb{R}^{l_b \times 1 \times K}$ ,  $l_a$ 、 $l_b$  分别为句子 A、B 的长度,  $K$  为卷积核的数目。

#### Step2: 提取交互信息。

对  $\text{Two\_C}^A$ 、 $\text{Two\_C}^B$  进行注意力交互计算, 分别得到文本 A、B 关于不同长度短语粒度的特征提取矩阵  $\text{Two\_C}^A$ 、 $\text{Two\_C}^B$ 、 $\text{Tri\_C}^A$ 、 $\text{Tri\_C}^B$ 。

### (4) 句子粒度的交互。

注意力机制能突出每个句子语义特征最明显的部分<sup>[13]</sup>, 该文应用注意力机制来捕捉文本的上下文信息特征, 最基本的注意力机制主要分为两种, 一种是由 Bahdanau<sup>[14]</sup>提出的加法注意力机制, 另外一种是由 Luong<sup>[15]</sup>提出的乘法注意力机制。

该文对编码层输出采用 Bahdanau 提出的加法注意力机制, 具体计算主要分为以下三步:

Step1: 将 query 和每个 key 进行计算得到注意力得分 score, 以某个时间步输出  $h_t$  为例, 计算注意力得分  $a_t$ :

$$s_t = (s_1^t, \dots, s_k^t) \quad (9)$$

$$s_k^t = \tanh(h_t + h_k) \quad (10)$$

Step2: 使用一个 SoftMax 函数对这些权重进行归一化, 如下式所示:

$$a_t = \frac{\exp(s_t)}{\sum_{j=1}^k \exp(s_j^t)} \quad (11)$$

Step3: 将权重和相应的键值 value 进行加权求和得到最后的 attention, 如下式所示:

$$c_t = \sum_{j=1}^k h_j a_j^t \quad (12)$$

目前在 NLP 研究中, key 和 value 常常都是同一个, 即 key = value。在文中, key 和 value 取值为编码层的字向量输出, 通过上述计算得到基于注意力加权的表示  $\text{SA\_A}$ 、 $\text{SA\_B}$ , 再将  $\text{SA\_A}$ 、 $\text{SA\_B}$  经过注意力交互计算即可得到句子 A、B 关于句子粒度交互的特征矩阵  $\text{sen\_A}$ 、 $\text{sen\_B}$ 。

### 1.2.3 组合层

在实际模型中, 经常会将不同的特征组合起来一同使用, 全面获取信息, 来达到提升模型性能的目的, 然而大多数模型在组合不同的语义信息时都使用了简单的 concat 方式, 使得维度过大, 模型计算量变大, 无法有效衡量各种语义信息的重要性。该文创新性提出一种融合注意力方法来解决这个问题, 具体内容如下文所述:

#### (1) 聚合交互信息。

将上述经过不同交互方式得到的四种矩阵在这一层进行聚合, 以聚合句子 A 的交互信息为例, 通过上文计算得到四个特征矩阵  $\text{CIA}$ 、 $\text{Two\_C}^A$ 、 $\text{Tri\_C}^A$ 、 $\text{sen\_A}$ , 其中某时刻输出分别为  $h_t^a$ 、 $h_t^b$ 、 $h_t^c$ 、 $h_t^d$ , 通过注意力加权求和得到表达式如下:

$$s_j = \tanh(W_1 h_t^j) \quad (j = a, b, c, d) \quad (13)$$

$$a_i = \frac{\exp(s_i)}{\sum_{j=(a,b,c,d)} s_j} \quad (14)$$

$$x_t = \sum_{i=(a,b,c,d)} a_i h_t^i \quad (15)$$

通过上述计算, 得到句子 A、B 的聚合交互信息表示  $\text{AG\_A}$ 、 $\text{AG\_B}$ 。

#### (2) 交互信息加强。

得到聚合交互信息表示后, 使用下式进行信息增强, 目的是为了构造与原有信息的差异性, 便于网络后续学习。

$$\text{eh\_A} = [E^A, \text{AG\_A}, E^A - \text{AG\_A}, E^A \odot \text{AG\_A}] \quad (16)$$

$$\text{eh\_B} = [E^B, \text{AG\_B}, E^B - \text{AG\_B}, E^B \odot \text{AG\_B}] \quad (17)$$

其中,  $E^A$ 、 $E^B$  分别为句子 A、B 的句嵌入原始矩阵,  $\text{AG\_A}$ 、 $\text{AG\_B}$  分别为其聚合交互信息表示,  $\text{eh\_A}$ 、 $\text{eh\_B}$  即为  $\text{AG\_A}$ 、 $\text{AG\_B}$  经过信息增强后得到的结果。

#### (3) 池化。

该文为最大限度保留上文提取到的特征, 综合使用平均和最大两种方式进行池化, 并将两种 pooling 得到的向量 concat 起来。池化过程如下式所示:



$$V_{a,ave} = \sum_{i=1}^{l_a} \frac{V_{a,i}}{l_a}, V_{a,max} = \max_{i=1}^{l_a} V_{a,i} \quad (18)$$

$$V_{b,ave} = \sum_{j=1}^{l_b} \frac{V_{b,j}}{l_b}, V_{b,max} = \max_{j=1}^{l_b} V_{b,j} \quad (19)$$

$$V = [V_{a,ave}, V_{a,max}, V_{b,ave}, V_{b,max}] \quad (20)$$

其中,  $V_{a,i}$  为增强信息经过 BILSTM 后每个时间步的输出。

#### 1.2.4 输出层

通过池化得到的向量经过全连接层作最后分类输出,其计算公式如下式所示:

$$\text{softmax}(y_i) = \frac{\exp(y_i)}{\sum_j \exp(j)} \quad (21)$$

$$\text{sim}(A, B) = \text{argmax}(y_i) \quad (22)$$

其中,  $y_i$  是模型对应每个分类的输出值,  $\text{argmax}$  则是取最大输出值对应的类别值,  $\text{sim}(A, B)$  即为两个句子最终的相似度。

## 2 实验设计

### 2.1 数据集构造

实验数据来源于 CCKS2018 评测项目中的微众银行客户问句匹配大赛数据集,共包含 182 478 个句子对,测试集包含 10 000 个句子对。若句子对语义信息相同,则对应的标签为 1,否则为 0,训练集样本实例如表 1 所示。

表 1 部分训练数据集样本

Sentence1	Sentence2	Similarity
第一次使用,额度多少	我的额度多少钱	1
我何时能通过微粒贷借钱	提前还清贷款还能再借吗	0
借款后多长时间给打电话	借款后多久打电话啊	1

### 2.2 实验设置

实验选用的深度网络框架是 tensorflow2.3, 字向量模型选用的是 BERT 模型中预训练好的 bert\_zh\_L-12\_H-768\_A-12 模型,句子最大允许长度为样本集中所有样本的最大长度, batch 大小设置为 20, epoch 设置为 15, 优化方式为 Adam, 损失函数选用的是多分类常用的交叉熵损失函数。静态词向量模型经由维基百科中文语料(1.42 G)训练而成,该语料包含 342 624 个词,维度为 300。实验选用准确率来衡量模型的性能,准确率的计算如下式:

$$\text{准确率 (Accuracy)} = \frac{\text{正确匹配的句子对数量}}{\text{句子对的总数量}}$$

### 2.3 模型验证

为了验证提出模型的有效性,对照实验模型采用 Siamese-LSTM、Siamese-CNN、ESIM、MatchPyramid。其中各模型最终评测结果如表 2 所示。

表 2 各模型准确率比较

模型	ACC
Siamese-CNN	0.72
Siamese-LSTM	0.81
ESIM	0.85
文中模型	0.87
MatchPyramid	0.71

各模型在各批次的损失及准确率变化如图 3、图 4 所示。

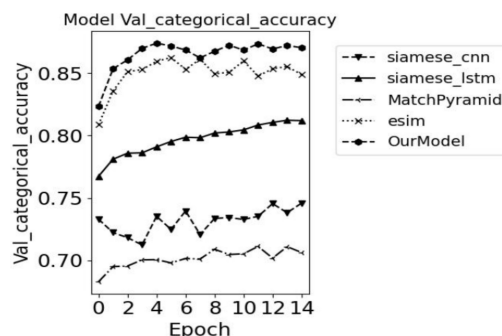


图 3 各模型准确率对比

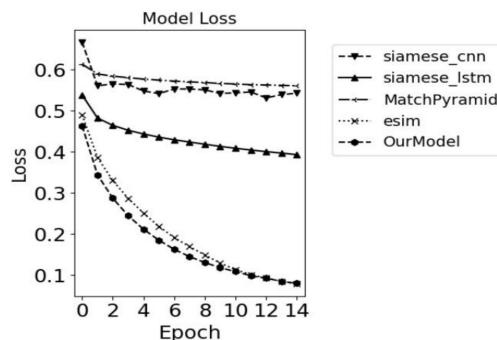


图 4 各模型损失值对比

### 2.4 实验结果分析

从表 2 可以看出,该文提出的模型相较于其他传统文本匹配模型表现较好,取得了 87% 的准确率,相较于对比模型中表现最好的 ESIM 模型大约有 2% 的提升,其中直接建模匹配的模型要比直接表示型的模型效果要好,这是因为 CNN、LSTM 等仅从文本本身进行语义建模,而忽略了文本之间的联系。而基于 LSTM 的 Siamese 模型要比基于 CNN 的 Siamese 特征提取能力强,这是因为 LSTM 在文本等序列建模问题上有一定优势,具有长时记忆功能,能够捕捉文本的长程依赖。ESIM 虽然在语义建模之前进行了注意力的交互以保留一些重要的文本特征值,并在后续中进行局部信息增强等操作来丰富语义信息,但总的来说 ESIM 只对词粒度进行了交互,而对中文来说考虑显然不够。从颗粒度特征提取来看,无论是 CNN、LSTM 还是 ESIM 都只是基于单一特征的提取,而该文提出的模型综合了多个粒度的交互信息,因而取得了更好的效果。

从图4可以看出,Siamese-CNN、Siamese-LSTM、MatchPyramid在训练过程中Loss下降缓慢并逐渐趋于平缓,这也导致正确率无法得到有效提升。提出的模型在训练中的Loss虽然最终和ESIM趋向一致,但是模型在训练过程中的下降速度上表现出色,在更少的批次就能达到比较好的效果,这也验证了文中模型结构的优越性。

## 2.5 模型应用

将提出的模型最终应用于CIFLog大型软件平台开发知识问答中,首先建立软件平台开发知识图谱作为问答数据支撑,同时人工构建问句模板库,然后通过命名实体识别等技术对用户的问句进行处理后,使用文中模型找到相似度最高的模板以完成答案的检索。整体效果如图5、图6所示。



图5 用户进行提问



图6 模型应用问答效果

## 3 结束语

针对传统模型存在的不足提出了相应的改进。探寻了更细的语言颗粒度在中文语义交互中的作用,较好地解决了句子交互特征不充分的缺点。同时在组合层中使用注意力机制和构造差异性方法来对交互信息进行融合增强,在最大程度保留交互特征的同时降低了模型的参数计算量,有效地提升了模型的性能。但是,从模型对语言的适用性来看,实验是在中文数据集上进行验证,因此在以后的工作中将尝试在英文数据集上进行实验,以验证模型是否具有普适性。同时,在实际的应用场景如问答中,发现模型的计算时间稍长,这不利于体验。因此在以后的工作中,也将进一步优化模型的执行效率,同时还需研究更多情形下的细粒

度特征抽取方法,以寻求更高性能的语义匹配模型。

## 参考文献:

- [1] 孙海霞,钱庆,成颖.基于本体的语义相似度计算方法研究综述[J].现代图书情报技术,2010(1):51-56.
- [2] 庞亮,兰艳艳,徐君,等.深度文本匹配综述[J].计算机学报,2017,40(4):985-1003.
- [3] SEVERYN A, MOSCHITTI A. Learning to rank short text pairs with convolutional deep neural networks[C]//Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval. [s. l.]: ACM, 2015: 373-382.
- [4] 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(4): 259-272.
- [5] WANG Z, HAMZA W, FLORIAN R. Bilateral multi-perspective matching for natural language sentences [C]//Twenty-sixth international joint conference on artificial intelligence. Melbourne: ACM, 2017: 4144-4150.
- [6] CHEN Q, ZHU X, LING Z, et al. Enhanced LSTM for natural language inference [C]//Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: long papers). Vancouver: ACL, 2016: 1657-1668.
- [7] 李伟康,李炜,吴云芳.深度学习中汉语字向量和词向量结合方式探究[J].中文信息学报,2017,31(6):140-146.
- [8] 王玉荣,林民,李艳玲. BERT跨语言词向量学习研究[J].计算机科学与探索,2021,15(8):1405-1417.
- [9] 吴炎,王儒敬.基于BERT的语义匹配算法在问答系统中的应用[J].仪表技术,2020(6):19-22.
- [10] 吴少洪,彭敦陆,苑威威,等. MGSC: 一种多粒度语义交叉的短文本语义匹配模型[J].小型微型计算机系统,2019,40(6):1148-1152.
- [11] PARIKH A P, TÄCKSTRÖM O, DAS D, et al. A decomposable attention model for natural language inference[J]. arXiv:1606.01933, 2016.
- [12] YIN W, SCHTZE H. MultiGranCNN: an architecture for general matching of text chunks on multiple levels of granularity [C]//Meeting of the association for computational linguistics & the international joint conference on natural language processing. Beijing: ACL, 2015: 63-73.
- [13] 朱张莉,饶元,吴渊,等.注意力机制在深度学习中的研究进展[J].中文信息学报,2019,33(6):1-11.
- [14] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [C]//Proceedings of the 3rd international conference on learning representations, San Diego: [s. n.], 2015: 1-15.
- [15] LUONG M T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation [C]//Proceedings of the 2015 conference on empirical methods in natural language processing. Lisbon: [s. n.], 2015: 1412-1421.