



基于多角度信息交互的文本语义匹配模型

翁兆琦,张琳

(上海海事大学 信息工程学院,上海 201306)

摘要: 现有的文本语义匹配方法大多基于简单的注意力机制进行交互,较少考虑文本自身结构信息和文本之间原始信息的交互。针对2个中文文本的语义匹配问题,构建一个多角度信息交互的文本匹配模型MAII。分别从颗粒、局部、全局3个角度计算2个文本深层次的语义交互矩阵,同时考虑语序信息之间和结构信息之间的交互以及文本内部的依赖关系,从而得到含有丰富信息的语义向量,并通过语义推理计算出两文本之间的语义匹配度。实验结果表明,相比在英文数据集上表现良好的DSSM、ESIM和DIIN模型,MAII模型在CCKS 2018问句匹配大赛的中文数据集上达到77.77%的准确率,表现出更好的匹配性能。

关键词: 信息交互;语义匹配;注意力机制;深度神经网络;中文数据集

开放科学(资源服务)标志码(OSID):



中文引用格式:翁兆琦,张琳.基于多角度信息交互的文本语义匹配模型[J].计算机工程,2021,47(10):97-102.

英文引用格式:WENG Z Q,ZHANG L.Text semantic matching model based on multi-angles information interaction[J].Computer Engineering,2021,47(10):97-102.

Text Semantic Matching Model Based on Multi-Angles Information Interaction

WENG Zhaoqi,ZHANG Lin

(College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China)

[Abstract] The existing semantic text matching methods are mostly based on a simple attention mechanism for interactions, and less consideration is given to the structural information of the text itself and the original information interactions between the texts. To address the semantic matching between two Chinese texts, a text matching model, MAII, is proposed based on multi-angle information interactions. The model calculates the deep-level semantic interaction matrix of the two texts at the granularity level, local level and global level respectively. At the same time, the interactions between the word order information, the interactions between structural information, and the internal dependencies of the two texts are also considered. On this basis, a semantic vector with rich information is obtained, and then the semantic matching degree between the two texts is calculated through semantic reasoning. Experimental results show that compared to DSSM, ESIM and DIIN models that perform well on the English data sets, the MAII model exhibits better performance on the Chinese data set of CCKS 2018 Question Matching Competition with the accuracy reaching 77.77%.

[Key words] information interaction; semantic matching; attention mechanism; Deep Neural Network(DNN); Chinese dataset
DOI: 10.19678/j.issn.1000-3428.0058964

0 概述

文本语义匹配是自然语言处理中众多任务的基础,也是自然语言处理的一个重要研究方向。文本语义匹配的应用场景十分广泛,例如信息检索中查询语句和检索内容之间的语义匹配度、自动问答中查询问句与候选答案之间的语义匹配度等。随着深度学习的不断发展,关于文本语义匹配问题的研究已经逐渐从传统基于统计的方法转变为深度文本语义匹配模型。目前,文本语

义匹配研究存在的挑战主要包括词语匹配的多元性、短语匹配的结构性和文本匹配的层次性3个方面^[1]。

本文构建文本匹配模型MAII,分别从颗粒、局部、全局3个角度进行语义信息交互,以获得更丰富的语义信息。其中,颗粒角度的交互用于匹配词语之间的语义,局部角度的交互用于匹配短语结构,全局角度的交互用于匹配文本层次性。此外,在CCKS 2018问句匹配大赛的中文数据集上进行实验,以验证模型的有效性。

基金项目:国家自然科学基金青年科学基金项目“视频测量用于多源船舶航迹融合的若干关键问题研究”(41701523)。

作者简介:翁兆琦(1996—),女,硕士研究生,主研方向为自然语言处理、文本语义匹配、知识库问答;张琳,副教授、博士。

收稿日期:2020-07-16 **修回日期:**2020-10-09 **E-mail:** QiQiWeng@163.com

1 相关工作

自然语言是一种包含复杂结构和丰富含义的语言。因此,包含丰富语义知识的文本表示是文本语义匹配的基础。深度文本匹配模型的发展主要可归纳为以下3个阶段:

1)单语义模型阶段,即先利用简单的神经网络编码文本,再计算文本之间的相似度。如DSSM模型^[2]先使用词哈希方法简化网络的输入向量维度,再利用深度神经网络将2个文本表示成同一语义空间中的低纬度语义向量,最后计算它们之间的语义相似度。

2)多语义模型阶段,即考虑到文本的局部结构,从文本的多颗粒角度解读文本。如MV-LSTM模型^[3]采用双向长短时记忆网络Bi-LSTM处理2个文本,为考察每个单词在不同语境下的含义,对长短时记忆网络LSTM^[4]隐藏层的输出进行语义匹配度计算。

3)深层次的句间交互模型阶段,即从文本不同的方面考虑深层次的交互,得到更丰富的语义表示,如BIMPM模型、DIIN模型、MatchPyramid模型和ESIM模型。BIMPM模型^[5]分别在句子的2个方向进行语义编码,并结合4种匹配方式进行文本的语义相似度匹配;DIIN模型^[6]先通过高速网络对2个句子进行语义编码,得到同一个句子之间不同单词的交互信息,再对之前的语义编码进行文本内部的注意力交互,最后通过DenseNet^[7]网络保留特征信息;MatchPyramid模型^[8]通过指示函数、余弦相似度、点积3种方法构建相似度矩阵,并用卷积神经网络对矩阵进行特征提取,最后进行分类得到句子的相似度;ESIM模型^[9]结合了BiLSTM和注意力机制,经过局部推理和推理组合后得到2个文本之间的语义匹配关系。

上述模型在很多高质量、大规模的英文开源数据集(如Quora、SNLI等)上都表现出较强的竞争力,但是由于中文在语法结构和语境表达上较英文更为复杂,并且中文文本语义匹配数据集相对缺乏,因此目前对中文文本语义匹配的研究较少。而这些在英文数据集上表现良好的文本语义匹配模型可能在中文数据集上表现并不出色。DSSM模型中使用的词哈希方法并不适用于中文,并且DSSM采用了词袋模型,丧失了文本的语序信息和上下文信息;ESIM模型和DIIN模型虽

然考虑了文本的语序信息和上下文信息,并且提取了这些信息的交互特征,但是对文本更多细节信息的交互仍考虑不够。

本文构建基于多角度信息交互的文本语义匹配模型MAII,针对词语匹配、短语结构匹配、文本层次性匹配这3个方面,分别从颗粒、局部、全局3个角度来充分利用文本原始的交互信息。首先得到词粒度的交互矩阵,然后通过卷积神经网络提取文本结构信息,进一步得到短语结构的交互矩阵,之后通过自注意力机制得到文本内部信息的交互矩阵,将所有交互矩阵进行拼接并送入Bi-LSTM中提取特征,分别得到2个文本最终的语义表示,经过平均池化后再由一个多层感知机进行分类,输出2个文本的匹配度。

2 多角度信息交互模型MAII

2.1 任务定义

对于给定的文本语义匹配样本 $sample = \{score, P, H\}$ (其中P、H表示需要进行语义匹配的2个文本,score为2个文本的匹配度,score可以为简单的0/1标签,也可以为连续的数值),文本语义匹配就是使P、H通过文本语义匹配模型得到它们的匹配度得分score。本文实验所采用数据集的score就是简单的0/1标签,以如下数据集样本为例,通过MAII模型判断score是否等于1。当score=1时,表示2个文本表达了同一语义,构成复述关系;当score=0时,表示2个文本表达了不同的语义,不构成复述关系。

例1

短文本P:卡里有钱为什么显示还款失败?

短文本H:为啥账户有钱却还款失败?

标签 score:1

例2

短文本P:一般多久能收到审核电话?

短文本H:一直都是等待电话通知

标签 score:0

2.2 MAII模型

本文首先对文本P、H做向量化处理,然后将其输入多角度信息交互模型MAII中进行端到端的训练,最后输出分类结果,得到2个文本的匹配度。MAII模型架构如图1所示。

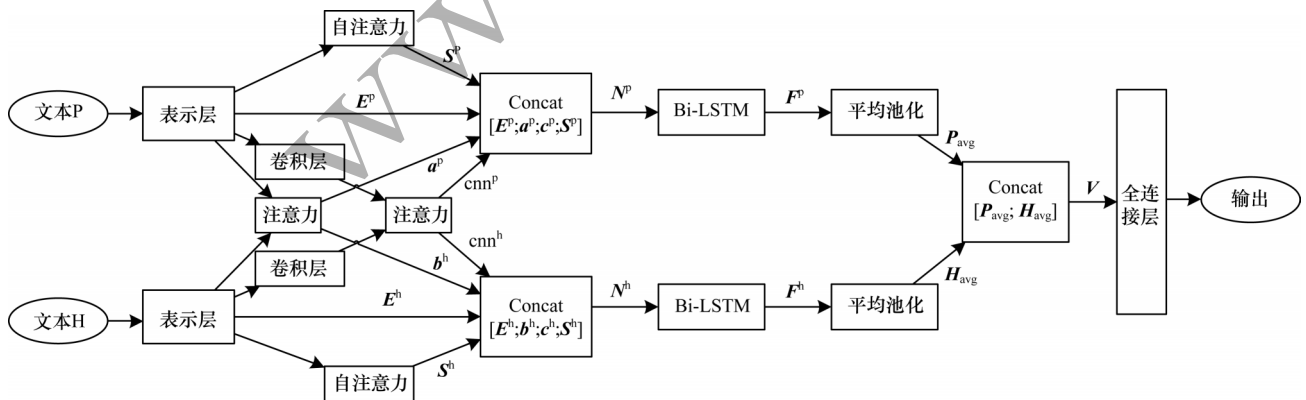


图1 MAII模型架构

Fig.1 Framework of MAII model

MAII模型主要由以下4个部分组成:

1)表示层:将自然语言形式表示的文本P、H向量化,以方便后续传入MAII模型中训练。

2)多角度注意力交互层:分别从3个角度计算不同层次的信息交互,得到3个注意力交互矩阵。

3)语义组合层:对得到的交互矩阵进行拼接组合后传入BiLSTM中提取特征,得到最终的文本语义表示。

4)预测层:将最终的文本语义表示经过平均池化后输入多层感知机进行分类。

2.2.1 表示层

表示层将自然语言形式表达的2个短文本向量化,便于后续语义向量的表达。利用Word2Vec^[10]、GloVe^[11]等预训练的词向量能够使语义相近的2个词语在维度上也更靠近,给模型带来更多的语义知识,使其具有更好的匹配性能。

当前较流行且取得了较好效果的预训练语言模型有GPT模型^[12]、BERT模型^[13-14]等,这些语言模型参数较多,加载速度慢,内存、时间等消耗大。Word2Vec是基于局部语料训练的,其特征提取基于滑动窗;而GloVe是基于全局词频统计的词表征工具,虽然利用到了全局信息,但是共现矩阵大,加载速度慢。综合考虑到实验效果和内存消耗等原因,本文选择Word2Vec语言模型来训练词向量^[15]。Word2Vec根据上下文之间的出现关系训练词向量,有Skip Gram、CBOW(Continuous Bags of Words)2种训练模式,Skip Gram根据目标单词预测上下文,CBOW则根据上下文预测目标单词,最后使用模型的部分参数作为词向量。

首先对短文本P、H进行分词,然后将每个单词用Word2Vec预训练词向量表示,对于词表中未出现的单词进行随机初始化。对于P、H这2个短文本,其词嵌入表示如式(1)和式(2)所示:

$$E^p = [E_1^p, E_2^p, \dots, E_i^p, \dots, E_m^p]^T, \forall i \in [1, 2, \dots, m] \quad (1)$$

$$E^h = [E_1^h, E_2^h, \dots, E_j^h, \dots, E_n^h]^T, \forall j \in [1, 2, \dots, n] \quad (2)$$

其中: $E^p \in \mathbb{R}^{m \times d}$; $E^h \in \mathbb{R}^{n \times d}$; m 、 n 分别表示短文本P、H中词的个数; d 表示词向量的维度。

2.2.2 多角度注意力交互层

一个好的语义匹配模型需要进行丰富的信息交互。注意力机制可以帮助模型对输入的每个部分赋予不同的权重,抽取出更关键和重要的信息,使模型做出更准确的判断,同时又不会为模型的计算和存储带来更大的开销。本文分别从颗粒、局部、全局3个角度进行注意力交互。

1)基于颗粒角度的注意力交互

为使2个文本互相匹配到对方低层次的语义信息,首先计算文本中词级粒度的交互信息。

基于颗粒、局部角度的注意力交互层以相同的方式分别计算文本P、H的注意力机制表示。以计算基于颗粒角度的注意力机制表示为例:

(1)计算文本P中第*i*个词和文本H中第*j*个词之间的点乘相似度 e_{ij} ,如式(3)所示:

$$e_{ij} = E_i^p \cdot E_j^h \quad (3)$$

其中: $e_{ij} \in \mathbb{R}^{m \times n}$ 。

(2)分别得到文本P、H的注意力表示 α^p 、 β^h ,如式(4)和式(5)所示:

$$\alpha_i^p = \frac{\exp(e_{ij})}{\sum_{k=1}^n \exp(e_{ik})} E_j^h, \forall i \in [1, 2, \dots, m] \quad (4)$$

$$\beta_j^h = \frac{\exp(e_{ij})}{\sum_{k=1}^m \exp(e_{kj})} E_i^p, \forall j \in [1, 2, \dots, n] \quad (5)$$

其中: $\alpha_i^p \in \mathbb{R}^{m \times d}$; $\beta_j^h \in \mathbb{R}^{n \times d}$; α_i^p 是对文本H中的每个词进行加权求和得到的,表示文本P中第*i*个单词与文本H中每个单词的匹配信息。 β_j^h 计算同理。

2)基于局部角度的注意力交互

相较于利用N-gram算法获取文本的结构信息,卷积神经网络由于局部连接和权值共享而计算量小,因此具有高效性。此外,利用卷积神经网络的位置不变性更易得到文本P、H的局部结构信息。针对判断2个文本之间语义是否相似的问题,以文本P1、H1为例进行说明。文本P1表示的是询问微粒贷分期还款功能的页面在哪里;文本H1则表示的是询问微粒贷分期还款所需要的月数页面在哪里。如果舍弃了文本头部和尾部的信息,而只用文本中间的信息来判断这2个文本语义是否相似,则会得到错误的结果。

P1:哪有页面显示微粒贷的分期还款啊?

H1:没有页面显示微粒贷分期还款的月数?

无论是文本头部、尾部的语义信息量还是文本中间的语义信息量,在对2个文本进行语义匹配度计算时都具有相同的重要性。因此,本文选用宽卷积^[16-18],具体步骤如下:

(1)将经过表示层的词嵌入 E^p 、 E^h 输入卷积神经网络中提取局部结构信息。经过Wide_CNN结构得到的文本语义矩阵可以表示如式(6)和式(7)所示:

$$C^p = \text{Wide_CNN}(E^p) \quad (6)$$

$$C^h = \text{Wide_CNN}(E^h) \quad (7)$$

其中: $C^p \in \mathbb{R}^{m \times 1 \times c_k}$; $C^h \in \mathbb{R}^{n \times 1 \times c_k}$; m 、 n 分别为P、H中词的个数; c_k 为卷积核的数目。 C^p 为词嵌入 E^p 经过Wide_CNN结构后得到的结果, C^h 同理。Wide_CNN结构如图2所示。假设文本P、H分别有6个单词,若卷积核的宽度为3,则在P、H的头部和尾部分别填充2列零向量后再进行卷积操作。为了后续交互信息的融合,对卷积的结果进行平均池化,使文本P、H的长度恢复到原始输入长度6。

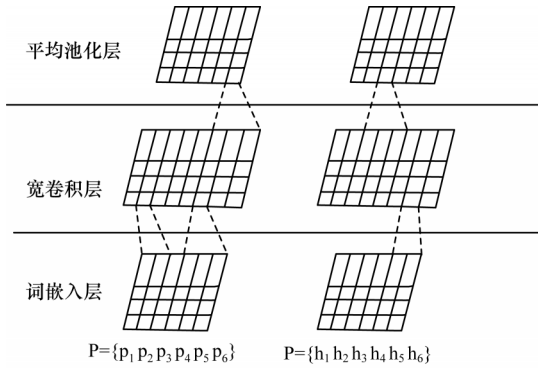


图2 Wide_CNN网络结构

Fig.2 Structure of Wide_CNN network

(2)对经过Wide_CNN结构的含有文本局部信息的编码层输出进行注意力交互计算,分别得到文本P、H最终的注意力表示 c^p 、 c^h 。

3) 基于全局角度的注意力交互

为了捕获文本的长距离依赖,考虑文本的语序和上下文信息,得到深层次语义信息,需要在文本内部进行交互,寻找文本序列内部的联系。以文本P为例,计算P的自注意力表示 S^p 如式(8)所示:

$$S_i^p = \frac{\exp(\alpha_{ij})}{\sum_{k=1}^m \exp(\alpha_{kj})} E_j^p, \forall i, j \in [1, 2, \dots, m] \quad (8)$$

其中: $\alpha_{ij} = \frac{E_i^{pT} E_j^p}{\sqrt{d}}$; d 为词向量的维度。同理,得到H的自注意力表示 S^h 。

2.2.3 语义组合层

经过多角度的交互信息计算,可以得到3种注意力交互矩阵。将上文得到的4个语义矩阵进行拼接组合得到新的语义表示,即:

$$N^p = \text{Concat}[E^p; \alpha^p; c^p; S^p] \quad (9)$$

$$N^h = \text{Concat}[E^h; \beta^h; c^h; S^h] \quad (10)$$

在此基础上,将新的语义表示输入Bi-LSTM提取文本语义特征,在LSTM中包含3个门控单元(输入门 I_t 、遗忘门 F_t 、输出门 O_t)和2个记忆单元(长期记忆、短期记忆 h_t)。在 t 时刻,LSTM的3个门单元和2个记忆单元的计算方法如下:

$$I_t = \sigma(W_I \cdot [h_{t-1}, x_t] + b_I) \quad (11)$$

$$F_t = \sigma(W_F \cdot [h_{t-1}, x_t] + b_F) \quad (12)$$

$$c_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (13)$$

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \quad (14)$$

$$C_t = F_t * C_{t-1} + I_t * c_t \quad (15)$$

$$h_t = O_t * \tanh(C_t) \quad (16)$$

其中: h_{t-1} 为 $t-1$ 时刻隐藏层的输出; x_t 为当前时刻(即 t 时刻)的输入; W_I 、 W_F 、 W_C 、 W_O 表示不同的权重矩阵; b_I 、 b_F 、 b_C 、 b_O 表示不同的偏置矩阵; σ 表示

Sigmoid函数; h_t 表示 t 时刻LSTM单元的输出。

最终的文本语义表示如式(17)和式(18)所示:

$$F_t^p = \text{Bi-LSTM}(F_{t-1}^p, N^p) \quad (17)$$

$$F_t^h = \text{Bi-LSTM}(F_{t-1}^h, N^h) \quad (18)$$

其中: F_t^p 、 F_t^h 分别为经过Bi-LSTM编码后2个文本在 t 时刻的语义表示。

2.2.4 预测层

为将整个文本的语义聚合成一个向量,对Bi-LSTM的输出结果进行平均池化操作来提取语义特征:

$$P_{\text{avg}} = \text{Avg_Pooling}(F^p) \quad (19)$$

$$H_{\text{avg}} = \text{Avg_Pooling}(F^h) \quad (20)$$

$$V = \text{Concat}[P_{\text{avg}}; H_{\text{avg}}] \quad (21)$$

最后,将语义向量 V 传入带有ReLU激活函数的多层感知机的分类器中进行分类。本文使用softmax交叉熵损失函数,整个网络采用反向传播进行端到端的训练。

3 实验与结果分析

3.1 数据集和模型参数设置

使用CCKS 2018微众银行智能客服问句匹配大赛^[19]提供的数据集进行实验,验证本文MAII模型的有效性。该数据集提供了10万对训练文本对、1万对验证文本对和1万对测试文本对,其中包括同义对和非同义对,同义对和非同义对的比例接近1:1。

以损失值和准确率2个指标来评价测试模型。实验由300维的Word2Vec预训练词向量工具更新分好词的文本对,对于词表中未出现的词采用随机初始化方式表达。Bi-LSTM的隐藏层大小为150,卷积核的宽度设置为3,dropout比例为0.5,batch的大小为128,实验选用Adam作为优化器。

3.2 匹配效果

作为实验对比,根据相关的论文进行复现或使用开源代码,将3个在英文数据集上表示良好的语义匹配模型应用到本次实验的数据集中。实验中所有模型均不进行任何特征工程^[20-21]。各模型的实验对比结果如表1所示。实验结果表明,本文提出的MAII模型在数据均衡的语义匹配中文数据集上表现较好。

表1 不同文本语义匹配模型的实验结果

Table 1 Experimental results of different text semantic matching models

模型	损失值	准确率/%
DSSM模型	0.65	69.15
ESIM模型	0.52	75.08
DIIN模型	0.48	76.39
MAII模型	0.46	77.77

为探索3种角度的交互信息对于文本语义匹配的有效性,本文另做了3次实验,在保证其他模块和参数不变的情况下,分别去除基于颗粒角度的信息交互模块、基于局部角度的信息交互模块和基于全局角度的信息交互模块,在测试集上得到如表2所示的准确率。可以看出,在去除任一个角度的交互信息模块后,模型在测试集上的准确率都有所下降,这说明3种角度的交互信息都具有重要的作用。其中,其中基于颗粒角度的交互信息模块对MAII模型的影响最大,缺少该模块的模型准确率比MAII模型低2.78%。

表2 3种角度的交互信息对模型性能的影响

Table 2 Influence of three-angles interactive information on the performance %

交互信息	准确率
去除基于颗粒角度的交互信息	74.99
去除基于局部角度的交互信息	76.77
去除基于全局角度的交互信息	77.03

3.3 分析与讨论

文本语义匹配实际上包含了很多类似的任务场景,如文本相似度计算、复述识别、问答匹配、自然语言推理、文本蕴含识别、机器阅读理解等。虽然在这些任务中,文本的语义匹配根据场景的不同而有略微的差异,但是很多情况下文本语义匹配模型在这些类似的任务上都具有通用性。本文实验选择了样本标签为0/1的短文本对。从同义文本对P:“卡里有钱为什么显示还款失败”和“H为啥账户有钱却还款失败”中可以看到,P中“卡里/有钱”“还款/失败”与H中“账户/有钱”“还款/失败”的语义是高度契合的。这些词语或短语的匹配可以通过本文提出的基于颗粒角度的交互和基于局部角度的交互得到很好实现。

在预测层,为了能够得到更多的语义特征,本文考虑了除平均池化外的最大池化,以期通过拼接平均池化和最大池化的结果来获取更多的信息,但笔者经过实验发现,添加了最大池化并没有提高测试集的准确率反而增加了计算量,因此,本文没有具体叙述。

此外,深度学习是端到端的训练难以解释,而本文通过提出的3个角度给很好地解释了这一过程,并且本文对提出的3个角度分别进行了实验,结果证明3个角度的提出都是非常有意义的。

4 结束语

本文提出基于多角度信息交互的文本语义匹配模型MAII,从3个不同的角度考虑文本信息的交互以获

得含有丰富语义知识的文本表示。实验结果表明,MAII模型较其他3个在英文数据集上表现良好的模型,在中文数据集上准确率更高。从MAII模型各模块拆分实验结果可见,从不同角度考虑文本的交互信息能够为文本的语义表示引入更丰富的知识,提升文本匹配任务的准确率。MAII模型虽然为中间的多角度交互层提供了更为丰富的语义知识,但是在开始的表示层仅使用了Word2Vec预训练词向量工具对文本进行向量化,而Word2Vec用法虽然简单易懂,却不能解决一词多义的问题。因此,下一步将使用BERT、EMLO等动态词向量模型向量化文本,使文本在进入多角度交互层之前可以为模型引入更多的语义信息。此外,在多角度交互层本文仅简单地对4种交互矩阵进行了拼接,没有更深层次地考虑4种交互矩阵之间的差异性。因此,后续将进行语义的各种推理和组合,加深对语义信息的理解和建模。

参考文献

- [1] 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.
PANG L, LAN Y Y, XU J, et al. A review of deep text matching[J]. Chinese Journal of Computers, 2017, 40(4): 985-1003. (in Chinese)
- [2] HUANG P S, HE X D, GAO J F, et al. Learning deep structured semantic models for Web search using clickth-rough data[C]//Proceedings of the 22nd ACM International Conference on Information and Knowledge Management. New York, USA: ACM Press, 2013: 2333-2338.
- [3] WAN S X, LAN Y Y, GUO J F, et al. A deep architecture for semantic matching with multiple positional sentence representations[C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2016: 2835-2841.
- [4] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [5] WANG Z G, HAMZA W, FLORIAN R. Bilateral multi-perspective matching for natural language sentences[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. New York, USA: ACM Press, 2017: 4144-4150.
- [6] GONG Y C, LUO H, ZHANG J. Natural language inference over interaction space[EB/OL]. (2018-05-26)[2020-05-10]. <https://arxiv.org/pdf/1709.04348.pdf>.
- [7] HUANG G, LIU Z, MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 1-5.
- [8] PANG L, LAN Y Y, GUO J F, et al. Text matching as image recognition[C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2016: 2793-2799.

- [9] CHEN Q, ZHU X D. Enhanced LSTM for natural language inference [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. San Diego, USA; Association for Computational Linguistics, 2017: 1657-1668.
- [10] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [J]. Advances in Neural Information Processing Systems, 2013, 26: 3111-3119.
- [11] PENNINGTON J, SOCHER R, MANNING C. GloVe: global vectors for word representation [C]//Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. San Diego, USA; Association for Computational Linguistics, 2014: 1532-1543.
- [12] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training [EB/OL]. [2020-05-10]. <http://www.nlp.ir.org/wordpress/wp-content/uploads/2019/06/Improving-language-understanding-by-generative-pre-training.pdf>.
- [13] 吴天波, 刘露平, 罗晓东, 等. 基于弱依赖信息的知识库问答 [J]. 计算机工程, 2021, 47(6): 76-82.
WU T B, LIU L P, LUO X D, et al. Knowledge base question and answer based on weak dependence information [J]. Computer Engineering, 2021, 47(6): 76-82. (in Chinese)
- [14] WOLF T. Some additional experiments extending the tech report "Assessing BERT's syntactic abilities" by Yoav Goldberg [EB/OL]. [2020-05-10]. <https://huggingface.co/bert-syntax/extending-bert-syntax.pdf>.
- [15] 徐菲菲, 冯东升. 文本词向量与预训练语言模型研究 [J]. 上海电力大学学报, 2020, 36(4): 320-328.
- XU F F, FENG D S. Research on text word vector and pre-training language model [J]. Journal of Shanghai Electric Power University, 2020, 36(4): 320-328. (in Chinese)
- [16] YIN W P, SCHÜTZE H, XIANG B, et al. ABCNN: attention-based convolutional neural network for modeling sentence pairs [EB/OL]. (2018-06-25) [2020-05-10]. <https://arxiv.org/pdf/1512.05193.pdf>.
- [17] HU B T, LU Z D, LI H, et al. Convolutional neural network architectures for matching natural language sentences [J]. Advances in Neural Information Processing Systems, 2015, 3: 2042-2050.
- [18] PENG S, CUI H B, XIE N T, et al. Enhanced-RCNN: an efficient method for learning sentence similarity [C]//Proceedings of 2020 Web Conference. Taipei, China: [s. n.], 2020: 2500-2506.
- [19] CCKS 2018 微众银行智能客服问句匹配大赛 [EB/OL]. (2018-04-01) [2020-05-10]. https://www.biendata.com/competition/CCKS2018_3/leaderboard/.
- [20] LAI H Y, TAO Y Z, WANG C L, et al. Bi-directional attention comparison for semantic sentence matching [J]. Multimedia Tools and Applications, 2020, 79: 14609-14624.
- [21] 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.
WU S D, PENG D L, YUAN W W, et al. MGSC: a short text semantic matching model with multi-granularity semantic intersection [J]. Journal of Chinese Mini-Micro Computer Systems, 2019, 40(6): 1148-1152. (in Chinese)

编辑 金胡考