

doi: 10.3969/j.issn.0490-6756.2020.01.009

基于增强问题重要性表示的答案选择算法研究

谢正文¹, 柏钧献², 熊熙³, 琚生根¹

(1. 四川大学计算机学院, 成都 610065; 2. 广东财经大学财政税务学院, 广州 510320;
3. 成都信息工程大学网络空间安全学院, 成都 610225)

摘要: 针对经典的文本匹配模型在问答系统中应用的缺陷和不足, 提出了一种基于增强问题重要性表示网络 BIWN 的答案选择算法. 目前, 现有的答案选择模型普遍将问题句子和答案句子直接进行匹配, 忽略了问题句子和答案句子中的噪声词对匹配的影响. 针对这个问题, 首先, 利用自注意力机制修改问题句子中各个词的权重, 生成“干净”的问题句子向量; 然后, 利用词级交互矩阵捕捉问题句子和答案句子之间的细粒度语义信息, 从而有效地弱化了噪声词对正确答案的影响; 最后, 利用多窗口 CNN 提取特征信息得到预测结果. 基准数据集上的对比实验表明, BIWN 模型在答案选择任务的性能优于主流的答案选择算法, MAP 值和 MRR 值提升了约 0.7%~6.1%.

关键词: 答案选择; 问题表示; 自注意力机制; 词级矩阵

中图分类号: TP391 **文献标识码:** A **文章编号:** 0490-6756(2020)01-0066-07

Question based importance weighting network for answer selection

XIE Zheng-Wen¹, BO Jun-Xian², XIONG Xi³, JU Sheng-Gen¹

(1. College of Computer Science, Sichuan University, Chengdu 610065, China;

2. Department of Public Finance and Taxation, Guangdong University of Finance and Economics, Guangzhou 510320, China;

3. School of Cybersecurity, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: According to the defects of the classic text matching model in the question and answer system, a question-based importance weighting network for answer selection is proposed. At present, the existing answer selection model generally matches the question sentence and the answer sentence directly, ignoring the influence of noise words in the question sentence and the answer sentence on the match. To solve this problem, the self-attention mechanism is firstly used to modify the weight of each word in the sentence to generate a "clean" question sentence vector. The word-level interaction matrix is then used to capture the fine-grained semantic information between the question sentence and the answer sentence. It weakens the influence of noise words on the correct answer. Finally, the multi-window CNN is used to extract the feature information to obtain the prediction result. The comparison experiments on benchmark datasets show that the performance of the BIWN model in the answer selection task is better than the mainstream answer selection algorithm, and the MAP value and MRR value are improved by about 0.7%~6.1%.

Keywords: Answer selection; Question representation; Self-attention; Word matrix

收稿日期: 2019-06-17

基金项目: 2018年四川省新一代人工智能重大专项(2018GZDZX0039)

作者简介: 谢正文(1994-), 男, 硕士生, 研究方向为数据科学.

通讯作者: 琚生根. E-mail: jsg@scu.edu.cn

1 引言

答案选择(answer selection, AS)是问答(question answer, QA)中的一个子任务,也是近年来信息检索(information retrieval, IR)中的一个热门话题.答案选择是根据输入问题从候选答案列表中选择最合适的答案.一般情况下,好的答案主要有以下两个特点.

(1) 问题影响答案,因此一个好的答案必须与问题相关;

(2) 好的答案并非要求严格的单词匹配,而是要表现出更好的语义相关性.

本文给出了一个从 WikiQA 数据集^[1]中提取的例子,如例 1 所示,正确的答案并没有直接提到丧生,而是用死亡,受伤等相关语义来代替.

例 1 WikiQA 数据集问答对

(1) Question:

有多少人在俄克拉何马城爆炸案中丧生?

(2) Postive Answer:

俄克拉荷马州的爆炸造成 168 人死亡,其中包括 19 名 6 岁以下儿童,并造成 680 多人受伤.

(3) Negative Answer:

据估计,炸弹造成至少 6.52 亿美元的损失.

传统的方法^[2-3]通常基于词法和句法特征,例如解析树之间的编辑距离^[2].这种方法在特征的设计和选择上非常耗时,而且由于语义的复杂性,特征选取更依赖于经验.端到端的神经网络模型可以自动选择特征^[4-6],避免了手动构造特征的麻烦,同时具有强大的拟合能力,被广泛应用.

近几年的研究表明^[7],答案选择任务的难点在于捕捉问题句子和答案句子之间复杂的语义关系.而问题和答案之间属于单向逻辑关系,答案需要根据问题来回答,因此问题表示在答案选择任务至关重要.鉴于此,本文提出一种基于增强问题重要性表示网络(question based importance weighting network, BIWN)的答案选择算法.该算法包括以下机制:(1) 本文在问题和答案进行交互前通过自注意力机制赋予问题句子各个词不同的权重从而缓解问题句子中的“无用词”噪声;(2) 本文构建词级矩阵,通过词级矩阵将问题和答案进行细粒度的交互,从而根据问题句子找出答案中的关键词;(3) 本文通过 CNN 捕获进行交互后的问题和答案的特征,进行匹配.

2 相关工作

答案选择通常被表述为一个文本匹配问题.以往的工作大致可以分为两个分支:基于特征选择的方法和基于神经网络的方法.传统的方法倾向于使用特征工程、语言工具或外部资源,这种方法费时且数据稀疏.前人的研究^[2,8]通常利用句法解析树上的信息来匹配问题和答案,而端到端的神经网络模型可以自动选择特征,避免了手动构造特征的麻烦;同时具有强大的拟合能力,更重要的是,在答案选择任务上神经网络模型的效果媲美于传统手动构建特征的方法,甚至更优于传统方法,因此如今被广泛应用. Feng 等人^[7]利用卷积神经网络(CNN)搭建 6 种不同的模型分别学习问题和答案的语义表示; Wang 和 Nyberg 等人^[9]利用循环神经网络(RNN)更容易捕捉全局序列信息的特点,将问题和答案句中的单词进行编码,然后输出它们的相关性得分;然而这些模型都缺乏问题和答案句子的交互信息.注意力机制旨在关注与问题最相关的候选答案的片段^[10-11], Wang 等人^[11]在词嵌入层利用门控注意力机制获取句子细粒度的语义表示; Nam 等人^[12]提出迭代模型 MAN,该模型利用动态记忆网络^[13]的思想,先用问题向量推导出第一个注意层中的答案向量,然后在下一步中对问题向量进行递归,学习第二个注意层中的答案向量.最后,与答案向量进行多步匹配得出分数总和.

以往的模型虽然取得了不错的效果,仍然存在噪声词的影响.本文把噪声词总结为以下两种情况.

(1) 如果问题句子中的词(例如一些 is, are, the 等词)与对应的答案句子中的词高度相关,但该词不是问题中的关键字,它应该被赋予更小的权重甚至是被忽略,因为它扰乱了加入问题信息后的答案句子的单词权重分布,从而产生噪声.

(2) 若一个单词在答案句子中是局部重要的,但与对应问题语义不相关,则需要赋予它在答案句子中较少的重要性,因其在语义匹配中无用.

本文模型利用自注意力在问题和答案进行交互前缓解问题噪声词的影响,同时构建词级交互矩阵,通过问题句子的信息来获取答案句子中重要的部分,从而更精确地完成问题和答案之间的语义匹配.

本文的工作很大程度上受到 MAN 模型进行逐步学习的思想.首先,模型利用自注意力机制将

编码后的问题句子重新分配各个词的权重,从而产生“干净”的问题句子,然后将“干净”的问题句子通过词级矩阵与答案进行交互,最后通过 CNN 获取问题和答案之间的多个特征进行匹配. 实验表明,在 WikiQA 数据集中,本文模型明显优于以往忽略了问题噪声词的模型,并且在 MAP, MRR 这两个评价指标上,本文模型超过 MAN 模型约 0.7%.

3 本文模型

本节将详细描述本文提出的 BIWN 模型,模型结构图如图 1 所示. 3.1 节介绍编码层中 BiLSTM 的工作原理;3.2 节介绍转化层中的自注意力机制;3.3 节介绍对齐层的相似度词级矩阵交互;3.4 节介绍特征融合层的多窗口 CNN;模型最后将 CNN 提取的特征用一个多层感知机转换成最后的预测向量.

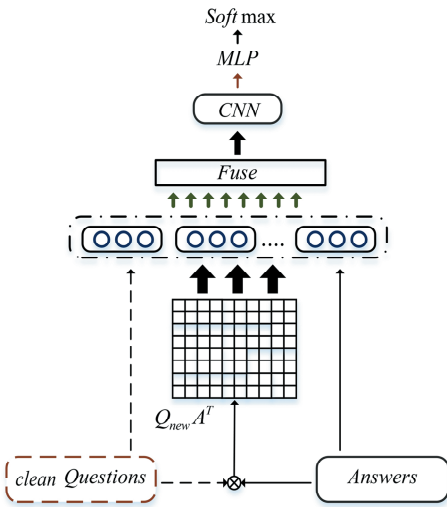


图 1 BIWN 模型总体结构图

Fig. 1 The network structure of BIWN

3.1 BiLSTM 编码层

输入两个文本序列,表示为问题 Q 和答案 A . 通过预先训练的 d 维词嵌入,用 $H_q = \{h_{q1}, \dots, h_{qn}\}$ 和 $H_a = \{h_{a1}, \dots, h_{an}\}$ 分别表示问题句向量和答案句向量. 其中, $h_{qi} \in R^d$ 是句子 H_q 的第 i 个词嵌入. m 和 n 分别表示问题和答案的长度. 为了捕捉句子上下文的信息,本文采用 BiLSTM 来对问题和答案进行编码. LSTM 的隐层维度为 u ,在 t 时刻的词嵌入为 x_t ,前一时刻的隐层和记忆单元分别为 h_{t-1} 和 c_{t-1} ,下一时刻的隐层 h_t 和记忆单元 c_t 计算如下.

$$g_t = \varphi(W_g x_t + V_g h_{t-1} + b_g),$$

$$\begin{aligned} i_t &= \sigma(W_i x_t + W_i h_{t-1} + b_i), \\ f_t &= \sigma(W_f x_t + W_f h_{t-1} + b_f), \\ o_t &= \sigma(W_o x_t + W_o h_{t-1} + b_o), \\ c_t &= g_t \odot i_t + c_{t-1} \odot f_t, \\ h_t &= c_t \odot o_t. \end{aligned}$$

其中, $W \in R^{u \times d}$; $V \in R^{u \times u}$; $b \in R^d$; σ 和 φ 分别是 sigmoid 函数和 tanh 函数; \odot 表示两个向量进行元素相乘,输入门 i ,遗忘门 f 和输出门 o 能够自动控制信息的流动. 同时记忆单元 c_t 可以记住长距离的信息, h_t 是 t 时刻的向量表示.

本文将 H_q 和 H_a 分别送入 LSTM 层,同时从两个方向进行计算,因而得到了前向向量 $T_{fh} = [t_{fh}^1, \dots, t_{fh}^m]$ 和后向向量 $T_{bh} = [t_{bh}^1, \dots, t_{bh}^n]$,然后将两个方向的向量进行拼接得到 $T_q^i = [t_{fh}^i, t_{bh}^i]$.

3.2 基于自注意力机制的问题向量重生成

在经过 BiLSTM 层后,分别得到问题的句子表示 $T_q = \{t_{q1}, \dots, t_{qn}\}$ 和答案的句子表示 $T_a = \{t_{a1}, \dots, t_{an}\}$,其中, $T_q \in R^{m \times 2u}$; $T_a \in R^{n \times 2u}$, u 表示隐层维度. 为了获得问题的表示,先前的工作^[9]通常使用以下三种机制之一: Last、Mean 或 Max Pooling. Last 取最后一个向量为 t_{qn} , Mean Pooling 取所有向量的平均值为 $\frac{1}{m} \sum T_q$, Max Pooling 取所有向量的最大值为 $\max T_q$. 这三种机制将句子压缩成单个向量表示,为了提取整个句子的信息,缓解噪声词现象,本文采用自注意力机制来为问题中的每个词计算权重并更新权重以忽略问题中的噪声词. 该层模型结构具体如图 2 所示.

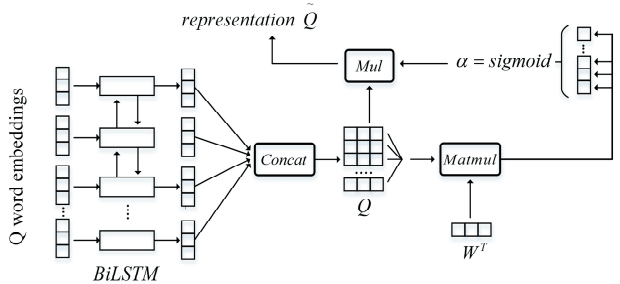


图 2 基于自注意力机制的问题向量重生成图

Fig. 2 The graph of re-generate the question vector based on Self-Attention Mechanism.

首先,计算问题向量 T_q 中每个单词的权重,具体计算公式如下.

$$v = T_q W_1 \quad (1)$$

$$\alpha_q = \text{sigmoid}(v) \quad (2)$$

其中, $W_1 \in R^{2u \times 1}$ 表示参数矩阵; $\alpha_q \in R^{m \times 1}$ 表示问

题句子中每个单词不同的权重分布; sigmoid 表示非线性激活函数, 在得到权重分布之后, 将权重 α_q 与问题向量 T_q 相乘, 生成新的问题向量表示 $U_q \in R^{m \times 2u}$, 具体计算公式如下。

$$U_q = \alpha_q \odot T_q \quad (3)$$

3.3 词级相似度矩阵对齐

词级相似度矩阵的主要思想是将两个句子拆分, 进行句子间的词级匹配, 词级矩阵中的元素是句子的局部相似度集合. Wang 等人^[14]直接用原始的词向量计算词级矩阵, 但词向量并未给句子带来上下文信息, 本文采用经过 BiLSTM 层后的句子向量获取上下文信息, 然后进行词级矩阵计算. 词级矩阵有两种计算方式, 一种是将两个句子直接相乘, 不带训练参数; 另一种是利用训练参数与两个句子共同作用. 本文采用两种词级矩阵进行建模, 同时在后续的实验(详见 4.2.3 节)比较两种词级矩阵的优劣. 词级矩阵的两种计算方式如下。

$$M(i, j) = U_q(i) T_a(j)^T \quad (4)$$

$$\tilde{M}(i, j) = U_q(i) W_p T_a(j)^T \quad (5)$$

其中, 矩阵 $W_p \in R^{2u \times 2u}$ 是模型训练的参数; $M(i, j) \in R^{m \times n}$, 词级矩阵的每一行是问题中的词对答案中的每个词的影响. 类似地, 词级矩阵的每一列看成答案中的词对问题中的每个词的影响. 分别对词级矩阵的行和列用 softmax 函数进行归一化, 得到互信息影响因子 $\lambda_q(i, j)$ 和 $\lambda_a(i, j)$, 其中, $\lambda_q(i, j)$ 和 $\lambda_a(i, j)$ 的取值范围均在 $[0, 1]$, 最后将问题向量和答案向量与对应的影响因子相乘得到两个新的向量 E_q 和 E_a . 具体计算公式如下。

$$\lambda_q(i, j) = \frac{\exp(M_{i,j})}{\sum_{i=1}^n \exp(M_{i,j})} \quad (6)$$

$$\lambda_a(i, j) = \frac{\exp(M_{i,j})}{\sum_{i=1}^m \exp(M_{i,j})} \quad (7)$$

$$E_a = \sum_{i=1}^m \lambda_a(i, j) \cdot U_q \quad (8)$$

$$E_q = \sum_{j=1}^n \lambda_q(i, j) \cdot T_a \quad (9)$$

其中, $E_q \in R^{m \times 2u}$; $E_a \in R^{n \times 2u}$.

3.4 层级信息融合

层级信息融合旨在捕捉多粒度的语义信息. 原始向量和注意力向量结合可以反映问题和答案不同粒度的语义表示, 本文将未经过注意力层的原始向量和经过注意力层的对齐向量融合. 问题原始向量表示为 Q , 经过注意力对齐层的向量表示为 \tilde{Q} ;

答案原始向量为 A , 经过注意力对齐层的向量表示为 \tilde{A} . 本文将不同粒度的向量进行比较, 其中向量相减表示两个向量之间的欧式距离, 向量相乘近似于两个向量之间的余弦距离, 具体计算公式如下。

$$K_q = [Q, \tilde{Q}, Q - \tilde{Q}, Q \odot \tilde{Q}] \quad (10)$$

$$K_a = [A, \tilde{A}, A - \tilde{A}, A \odot \tilde{A}] \quad (11)$$

其中, $K_q \in R^{m \times 8u}$; $K_a \in R^{n \times 8u}$.

CNN 和 LSTM 经常被用于特征融合层, LSTM 通过计算每个隐藏向量的上下文感知权重来捕捉句子的全局序列信息, 但是 LSTM 的串行工作机制使得模型运行的时间较长. 而 CNN 可以并行快速计算, 并且 CNN 捕捉特征能力比 LSTM 要强, 同时多窗口的 CNN 也在一定程度上弥补了其处理长序列信息能力较弱的缺点. 综合上述原因, 本文采用多层的 CNN 来提取融合特征. 具体计算公式如下。

$$u = \text{CNN}(\text{Fuse}) \quad (12)$$

其中, Fuse 表示融合向量 K_q 或者融合向量 K_a .

本文将 CNN 的输出 u 通过最大池化和平均池化得到 $S_{q, \max}$, $S_{a, \max}$, $S_{q, \text{mean}}$ 和 $S_{a, \text{mean}}$, 然后将其拼接成一个向量 S , 送入多层感知器中 (MLP) 得到最后的预测向量 $\text{Score} \in R^N$. 本文选用 Listwise 方法来建模答案排序任务, 具体计算公式如下。

$$G = \text{soft max}(\text{Score}) \quad (13)$$

$$Y = \frac{Y}{\sum_{i=1}^N y_i} \quad (14)$$

其中, G 表示得分向量; Y 表示目标标签归一化后的标签向量; N 表示 Listwise 的大小. 本文使用 KL 散度作为模型的损失函数, 该函数可以使得预测值的概率分布接近于标签值的概率分布. 损失函数公式具体如下。

$$\text{Loss} = \frac{1}{n} \sum_{i=1}^n \text{KL}(G \parallel Y) \quad (15)$$

4 实验说明

4.1 实验设置

4.1.1 数据集 本文采用两个答案选择领域的基准数据集进行实验, 数据集具体信息见表 1.

WikiQA 数据集^[1]: WikiQA 数据集是由微软雷德蒙研究院在 2015 年发布, 主要针对英文领域的答案句子选择任务. 该数据集包括 3 047 个问题, 每个问题从微软必应搜索引擎的查询日志中采

样获得,并保证该问题以疑问词开头、以问号结尾、并且点击过至少一个维基百科页面.数据集中将能够回答该问题的句子标注为 1,将其他句子标注为 0.

TREC-QA 数据集^[15]:TREC-QA 数据集是从 TREC 问答网站收集的基准数据集.数据集包含一组事实型问题,其中候选答案仅限于一句话.

表 1 问答数据集描述

Tab. 1 The statistics of QA datasets

数据集	WikiQA	TREC-QA
训练集/验证集/测试集	873/126/243	1162/65/68
问题平均长度	6	8
答案平均长度	25	28
候选答案平均长度	9	38

4.1.2 实验参数设置集 本文实验算法基于 Tensorflow 框架实现,BIWN 模型采用 300 d 的 Glove 词向量^[16]作为词嵌入;采用 mini-batch 的训练方式训练模型,采用用 Padding 的方式处理数据集,问题句子的 padding 长度设置为 25,答案句子的 padding 长度设置为 140;训练采用 Listwise 的数据格式,列表尺寸设置为 15,CNN 的窗口尺寸设置为[1,2,3,5].采用学习率衰减的方式训练模型,初始学习率设置为 0.001,采用 Adam 优化器来进行梯度下降,为了避免模型过拟合,采用 L2 正则化来优化模型.

4.1.3 本文对比算法 (1) APLSTM^[17]:该模型利用交互注意力获取问题和答案的交互信息;(2) MP-CNN^[18]:该模型在 CNN 的基础上利用全连接层融合问题和答案的表示;(3) PWIM^[19]:该模型将问题和答案的进行显式交互以计算问题和答案之间的重要部分;(4) BiMPM^[20]:Wang 于 2017 年提出的多尺度匹配模型,该模型将问题和答案进行词级,句子级的多尺度匹配,取得不错的效果;(5) MAN^[12]:Nam 于 2018 年提出的迭代匹配模型,该模型用顺序注意力机制将问题与答案进行多步匹配,最后将各个部分的分数相加进行预测.

4.2 实验结果与实验分析

4.2.1 基线模型对比 答案选择任务文档检索(IR)任务相似,因此本文采用 IR 领域的评价指标 MAP(平均准确率),MRR(平均倒排准确率)来评估模型性能.实验结果如表 2 所示.

如表 2 所示,在 WikiQA 数据集上,BIWN 模型的 MAP 值和 MRR 相比于其他模型分别提高

了 0.7%~5.9%和 0.7%~6.1%.BIWN 模型与 APLSTM 模型均使用了词级矩阵结构,但是 BIWN 模型的效果却远远好于 APLSTM,原因在于 BIWN 模型在问题和答案进行交互前利用自注意力机制消除了噪声词.BIWN 模型与 MAN 模型都采用了逐步学习的方式进行句对匹配,然而 BIWN 模型的指标优于 MAN 模型,表明 BIWN 模型中的词级矩阵可以捕获更细粒度的语义信息.在 TREC-QA 数据集上,BIWN 模型的指标优于大部分基线模型,并且与 MAN 模型比较接近,原因在于 TREC-QA 数据集相比于 WikiQA 数据集更加规整,因此噪声词对模型的影响相对而言没有那么大,而本文针对噪声词提出的 BIWN 模型则不能发挥其最大的作用.

表 2 基于 WikiQA 和 TREC-QA 数据集的实验结果对比
Tab. 2 Comparison of experimental results based on the WikiQA dataset and TREC-QA dataset

模型	WikiQA		TREC-QA	
	MAP	MRR	MAP	MRR
APLSTM	0.670	0.684	0.713	0.803
MP-CNN	0.693	0.709	0.777	0.836
PWIM	0.709	0.723	—	—
BiMPM	0.718	0.731	0.802	0.875
MAN	0.722	0.738	0.813	0.893
BIWN	0.729	0.745	0.806	0.889

4.2.2 词级矩阵对比 为了探究不同词级矩阵对 BIWN 模型的影响,本文在不改变模型其他结构和模型参数的情况下进行对比实验,实验结果如表 3 所示.

表 3 词级矩阵实验对比

Tab. 3 Comparison of word-level matrix experiments

Model's Matrix Part	MAP	MRR
BIWN with W	0.729	0.745
BIWN without W	0.724	0.742

如表 3 所示,两种词级矩阵的效果没有太大的差别,但是使用参数 W 的相似度词级矩阵的效果要比两个向量直接相乘要好.本文对此进行分析得出原因:答案选择任务并不是完全意义上的句对匹配任务,相似的词在语义表示上有可能是不同的,但是可训练的参数 W 可以转换信息,将其映射到另一个相似的空间,因此效果更好.

4.2.3 消融实验对比 为了验证模型中各个部件

的有效性, 分别移除特定模块构建下列消融模型: (1) BIWN-Encoder 表示去除 Bi-LSTM 编码层; (2) BIWN-SelfAtten 表示去除自注意力层; (3) BIWN-CNN 表示去除词级相似度矩阵; (4) BIWN-WordMatrix 分表示去除 CNN 特征融合层的消融模型. 图 3 展示在 WikiQA 和 TREC-QA 的实验结果, 可以看出 BIWN 模型的每个模块都具有有效性, 其中自注意力机制和词级矩阵相对其中结构影响最大.

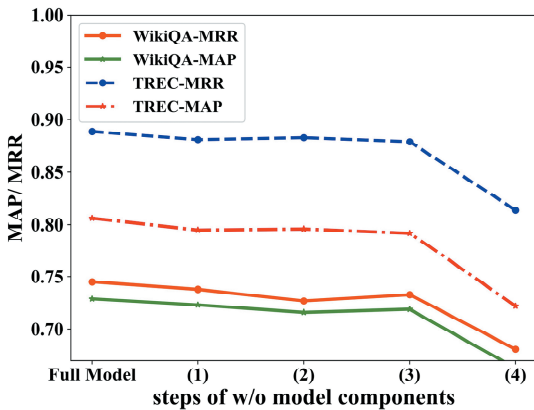


图 3 基于 WikiQA 和 TREC-QA 测试集的模型消融分析折线图

Fig. 3 Ablation analysis about different components of model on WikiQA and TREC-QA test set

5 结 论

本文针对句子中的噪声词, 提出了基于问题重要性表示网络的答案选择算法. 该模型利用自注意力机制重新赋予各个词不同的权重从而生成“干净”的问题句子向量; 利用词级交互矩阵捕捉问题句子和答案句子之间的细粒度语义信息, 缓解了答案句子中噪声词的影响. 基准数据集上的对比实验表明, BIWN 模型在答案选择任务的性能优于目前主流的答案选择算法. 在未来的工作中, 会将外部的知识加入 BIWN 模型中; 同时, 没有标签的数据相对与有标签的数据更容易获得, 将会把无监督学习应用到答案选择任务中.

参考文献:

[1] Yang Y, Yih W, Meek C. Wikiqa: a challenge dataset for open-domain question answering [C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: EMNLP, 2015.

[2] Heilman M, Smith N A. Tree edit models for recognizing textual entailments, paraphrases, and answers to questions [C]// Proceedings of the Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. California: NAACL, 2010.

[3] 郑晓东, 潘敬敏, 胡汉辉, 等. 一种基于 SCAD 的改进谓词发现方法 [J]. 江苏大学学报: 自然科学版, 2018, 39: 576.

[4] 王霞, 孙界平, 琚生根, 等. 基于段落内部推理和联合问题答案匹配的选择型阅读理解模型 [J]. 四川大学学报: 自然科学版, 2019, 56: 423.

[5] 高云龙, 左万利, 王英, 等. 基于集成神经网络的短文本分类模型 [J]. 吉林大学学报: 理学版, 2018, 56: 933.

[6] 陈波. 基于循环结构的卷积神经网络文本分类方法 [J]. 重庆邮电大学学报: 自然科学版, 2018, 30: 705.

[7] Feng M, Xiang B, Glass M R, *et al.* Applying deep learning to answer selection: A study and an open task [C]// Proceedings of the 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). Hong Kong: IEEE, 2015.

[8] Severyn A, Moschitti A. Automatic feature engineering for answer selection and extraction [C]// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle: EMNLP, 2013.

[9] Wang D, Nyberg E. A long short-term memory model for answer sentence selection in question answering [C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Beijing: ACL, 2015.

[10] Tan M, Dos Santos C, Xiang B, *et al.* Improved representation learning for question answer matching [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin: ACL, 2016.

[11] Wang B, Liu K, Zhao J. Inner attention based recurrent neural networks for answer selection [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin: ACL, 2016.

[12] Tran N K, Niedereée C. Multihop attention networks for question answer matching [C] // Pro-

- ceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. Ann Arbor: ACM, 2018.
- [13] Kumar A, Irsoy O, Ondruska P, *et al.* Ask me anything: Dynamic memory networks for natural language processing [C]//Proceedings of the International conference on machine learning. Washington: ICML, 2016.
- [14] Wang S, Jiang J. A compare-aggregate model for matching text sequences [C]//Proceedings of the 5th International Conference on Learning Representations. Toulon: ICLR, 2017.
- [15] Wang M, Smith N A, Mitamura T. What is the Jeopardy model? A quasi-synchronous grammar for QA [C]//Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). Prague: EMNLP-CoNLL, 2007.
- [16] Lin W S, Dai H J, Jonnagaddala J, *et al.* Utilizing different word representation methods for twitter data in adverse drug reactions extraction [C]//Proceedings of the 2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI). Hong Kong: IEEE, 2015.
- [17] Santos C, Tan M, Xiang B, *et al.* Attentive pooling networks [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: ACL, 2016.
- [18] He H, Gimpel K, Lin J. Multi-perspective sentence similarity modeling with convolutional neural networks [C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: EMNLP, 2015.
- [19] He H, Lin J. Pairwise word interaction modeling with deep neural networks for semantic similarity measurement [C]//Proceedings of the The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. San Diego: NAACL-HLT, 2016.
- [20] Wang Z, Hamza W, Florian R. Bilateral multi-perspective matching for natural language sentences [C]//Proceedings of 9th International Joint Conference on Computational Intelligence. Madeira: IJCCI, 2017.

引用本文格式:

中文: 谢正文, 柏钧献, 熊熙, 等. 基于增强问题重要性表示的答案选择算法研究[J]. 四川大学学报: 自然科学版, 2020, 57: 66.

英文: Xie Z W, Bo J X, Xiong X, *et al.* Question based importance weighting network for answer selection [J]. J Sichuan Univ: Nat Sci Ed, 2020, 57: 66.