

融合歧义感知的检索式问答方法*

蒲晓¹, 何睿¹, 王志文¹, 黄珊珊², 袁霖¹, 吴渝^{1†}

(1. 重庆邮电大学 网络空间安全与信息法学院, 重庆 400065; 2. 武昌首义学院 马克思主义学院, 湖北 武汉 430064)

摘要: 针对多义词在不同上下文中语义表达不一致的问题, 提出了一个融合歧义感知的问答模型, 即模型在问题-候选答案的语义匹配过程中, 与外部知识源相结合, 动态识别并检测出每个多义词在不同场景下的语义, 并将检测到的语义信息进行特征编码后融合到语义匹配任务中, 使模型能够更为准确地理解每个词的精准含义, 从而做出更为精准的匹配判断。在歧义感知模型的设计上, 采用基于Transformer的深度语义编码器, 使其能够更加全方位地抓取待分析歧义词以及知识源的深度语义特征, 从而做出更加准确的语义消歧。在标准检索式问答数据集上(WikiQA和TrecQA)的实验结果表明, 所提出的歧义感知的问答方法能够有效融合到多个基线模型中, 并捕捉到多义词在不同语境中的精准语义, 使其在包含公开数据集上的问答性能MAP评估高于对应基线模型约1%, 且该语义特征使得基于BERT的文本相似性匹配模型的性能优于当前先进的其它模型。

关键词: 语义消歧; 歧义感知; 智能问答; 语义匹配; 信息检索

DOI: 10.13568/j.cnki.651094.651316.2023.07.10.0001

中图分类号: TP391.9 **文献标识码:** A **文章编号:** 2096-7675(2024)01-0027-010

引文格式: 蒲晓, 何睿, 王志文, 黄珊珊, 袁霖, 吴渝. 融合歧义感知的检索式问答方法[J]. 新疆大学学报(自然科学版)(中英文), 2024, 41(1): 27-36.

英文引文格式: PU Xiao, HE Rui, WANG Zhiwen, HUANG Shanshan, YUAN Lin, WU Yu. Sense-aware retrieval-based question answering via word ambiguity induction[J]. Journal of Xinjiang University(Natural Science Edition in Chinese and English), 2024, 41(1): 27-36.

Sense-Aware Retrieval-Based Question Answering via Word Ambiguity Induction

PU Xiao¹, HE Rui¹, WANG Zhiwen¹, HUANG Shanshan², YUAN Lin¹, WU Yu¹

(1. School of Cyber Security and Information Law, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. School of Marxism, Wuchang Shouyi University, Wuhan Hubei 430064, China)

Abstract: To solve the problem of inconsistent semantic expression of polysemous words in different contexts, we propose a sense-aware question-answer model. During the semantic matching process of questions and candidate answers, the model integrates with external knowledge sources to dynamically identify and detect the semantics of each polysemous word in different scenarios. The detected semantic information is encoded as features and then integrated into the semantic matching task, enabling the model to capture the exact meaning of each word and achieve better matching performance. In the design of the ambiguity perception model, we adopt a deep semantic encoder based on the Transformer, which enables it to capture more comprehensive depth semantic features of the analyzed ambiguous words and knowledge sources, making more accurate semantic disambiguation. Experimental results on standard retrieval-based Q&A datasets (WikiQA and TrecQA) demonstrate that the proposed sense-aware Q&A method can effectively be integrated into multiple baseline models, capturing the precise semantics of polysemous words in different contexts. This approach achieves a MAP evaluation performance improvement of approximately 1% compared to corresponding baselines on public datasets. Moreover, this semantic feature enables a BERT-based text matching approach to outperform other state-of-the-art models.

Key words: semantic disambiguation; sense-aware mechanism; retrieval-based Q&A; semantic matching; information retrieval

* 收稿日期: 2023-07-10

基金项目: 重庆市自然科学基金面上项目“融合弱监督歧义感知的跨模态图文检索方法研究”(CSTB2022NSCQ-MSX1342); 重庆市教育委员会科学技术研究项目“融合弱监督语义消歧的文本匹配方法研究”(KJQN202300619)。

作者简介: 蒲晓(1987—), 女, 博士, 讲师, 从事信息检索、跨模态语义理解和机器学习的研究, E-mail: puxiao@cqupt.edu.cn.

† 通讯作者: 吴渝(1970—), 女, 博士, 教授, 从事网络智能、数字媒体、数据挖掘的研究, E-mail: wuyu@cqupt.edu.cn.

0 引言

随着信息化程度的提升与人工智能的飞速发展,自动问答(Question Answering)得到了较为广泛的关注和应用,如阿里巴巴的AliMe^[1]、微软的SuperAgent^[2]以及各类移动设备上的语音助手等.从答案获取的途径来看,自动问答可以分为检索式(Retrieval-based)问答方法^[3]与生成式(Generation-based)问答方法^[4]两大类.本文主要围绕检索式问答方法展开研究.检索式自动问答尽管有不同任务,但是构建流程类似,即针对用户给定的问题,对问题与知识库中的候选答案进行语义相似性匹配,并通过计算结果对候选答案进行重新排序,最终找到针对该问题的最佳答案.

然而在当前问答任务中,模型通常用同样的词向量来表示每个词在该文本数据中所有出现语义.但是,文本数据中存在语义多元化的词(或称多义词),即一个相同的词在不同的语境下可能表达不同的语义.用相同的词嵌入表达多义词的所有出现会导致其语义的多样性被隐藏,甚至导致当前语境的语义被错误理解.与此同时,检索式自动问答中,文本长度本身较短,所包含的语义信息较为缺乏,模型根据问题的文本特征进行问题-答案相似性匹配时,多义词语义的错误表达会严重影响模型对该句的理解,从而做出错误的判断.

针对上述问题,本文提出了一种融合歧义感知的检索式自动问答方法.模型由两部分组成:歧义感知模块和问题-答案语义相似性匹配模块.首先,我们将问题-候选答案文本对输入到歧义感知模块中,该模型通过双向循环神经网络(Bidirectional Recurrent Neural Network, BRNN)^[5]以及Transformer^[6]的协同分析,挖掘出输入文本更为细粒度的上下文语义信息,并结合外部词典WordNet的语义注解信息,生成该词在当前场景中的精准语义特征表达.最后将生成的语义特征融合到相似性匹配模块,使模型在计算问题-答案语义相似性的同时,能够更加全面地理解该句的当前语义,从而作出更加精准判断.本文的主要贡献如下:

1) 提出一个针对输入问题-答案文本对的歧义感知方法,该方法主要由Bi-GRU以及Transformer构成,通过不同维度的上下文语义联合分析,挖掘出文本中更深层次的语义信息.同时结合WordNet知识源提供的多义词相关信息,对多义词在当前语境下的精准语义进行判断和表达.

2) 在三个经典语义匹配基线模型(BCNN、RE2和BERT)上进行融合对比,将歧义感知模块融合到这三个基线模型中,融合了歧义感知的检索问答模型在计算文本相似度的同时,对文本中每个多义词的语义进行检测和特征生成.

3) 将设计的融合歧义感知的检索问答模型与其基线模型,以及当前具有代表性的语义相似性匹配方法进行了性能评估,通过使用WikiQA、TrecQA两个基准数据集,对提出的算法性能进行了全面的评估,实验结果表明,所提方法优于基线算法,同时也优于当前其它算法.

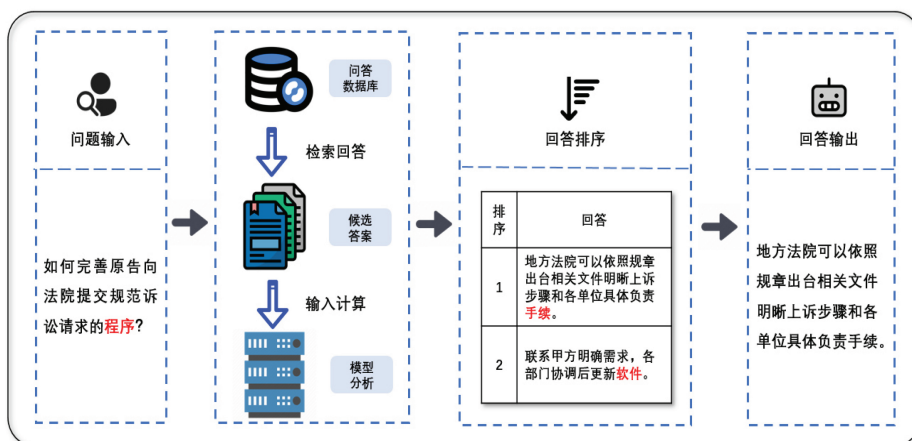


图 1 检索式自动问答示例

1 相关工作

问答任务中,用户输入问题时不可避免地出现多义词.而在文书数据中,多义词出现也十分频繁.大部分NLP任务都使用相同的词特征向量表示每个在文本中出现的多义词^[7-9].使用该种方式强制将多义词映射为

同一向量会丧失该词在当前维度下的丰富语义信息,导致模型对每个词的理解不充分,从而对整个句子的理解产生歧义^[10]。如图1所示,针对给定问题,共有两个候选答案。问题文本中含有“程序”一词,该词可表达为“手续、步骤”,也可表达为“應用程式”。该词在第一句中表达的是“手续、步骤”的含义,模型应判断两者的关联性较高,而在第二句中表达的是“电脑软件”含义,该样本的关联度与源问题较低。由此可见,如果我们能够将文本中的多义词在不同场景下的语义进行精准表达,则可使问答模型能够更加全面和准确地理解样本含义,从而作出更为精准判断。

1.1 检索式自动问答

检索式自动问答方法旨在通过对问题及候选答案的语义相似性计算,找到语义最为匹配的对应答案。其本质是对问题-候选答案对之间进行相似度匹配的过程。近几年深度学习引入后,检索式自动问答任务得到了飞速的发展。目前的方法主要采用注意力机制、多通道等方式。如Chen等^[11]通过计算文本对之间的注意力矩阵,从而获得问题与候选答案的语义相似性。Kundu等^[12]在获取问题与候选答案的相似矩阵后进行归一化处理,依次得到问题与回答的相关性。Du等^[13]利用时态卷积网络(Temporal Convolutional Network)得到输入问题的主题转移特征,结合注意力机制进行模型融合。对于多通道机制,Yan等^[14]利用融合了输入文本上下文特征的多通道机制,分别计算候选答案与各通道问题的语义相似性,根据获取分数进行排名,而Wu等^[15]结合多通道与注意力机制,在使用多通道获取上下文特征的同时,结合注意力机制捕获局部信息,以此检索回答。

2019年,随着BERT模型(Bidirectional Encoder Representations from Transformers, BERT)的提出,文本语义相似度计算领域取得了巨大的突破。BERT是由Google公司基于Transformer模型改进且由大量数据训练而成,该模型很好地抓取了样本对之间以及自身之间的特征信息,并在NLP多个领域取得了较大的突破。但是,即便BERT拥有很好的语义抓取能力,它依旧存在多义词不同语义特征表示统一的问题,这其实是语义编码的一种错误。

1.2 语义消歧

歧义与消歧是自然语言理解中非常典型且普遍的问题,且在词义、语义、篇章含义层次都会出现语言根据上下文描述而语义不同的现象。消歧即根据上下文确定对象语义的过程^[16]。词义消歧即在词语层次上的语义消歧。根据所使用的资源类型不同,可以将词义消歧方法分为两类:基于知识的方法和基于监督的方法。

基于知识的方法依赖如WordNet^[17]这样的词典资源。该词典中,每个多义词可能的语义都配有相应的语义注解。该注解开始被应用在Lesk算法中^[18-19],然后被广泛运用到各类任务中^[20-21]。与此同时,语义结构模型也通过图模型算法被利用起来^[20]。随着深度学习在NLP领域的应用,基于深度学习方法的词义消歧成为这一领域的一大热点。深度学习算法自动提取分类需要的低层次或高层次特征,减少了很多特征工程方面的工作量。近几年由于BERT模型在NLP领域的巨大成功,一些研究尝试将BERT引入到语义消歧任务中^[21-22]。该类研究借助词典WordNet的帮助,将样本中多义词的上下文和词典中该多义词每个可能语义的注解进行相似度匹配。

2 融合歧义感知的检索式问答方法

如图2所示,该项工作整体分为两个模块:歧义感知与语义融合。歧义感知模块旨在检测文本中的多义词,通过将检测到的多义词的上下文信息与知识源中对应多义词注解信息之间的语义相似度计算,从而判断出该词在当前场景下的精准词义;语义融合模块旨在将前者生成的每个多义词在当前场景下的精准语义融合到主任务中,使模型在计算问题-答案语义相似性的同时,能够更加全面理解该句的精准语义,从而提高问答性能。

2.1 WordNet词汇知识库

WordNet提供了一种词汇知识:多义词语义注解。该词典列出所有多义词可能的语义,每个语义附带一句文字注解。如多义词“season”在WordNet知识库中主要包含两种含义,即season.n.01: a period of the year marked by special activities in some field (一年中以特殊事件的活动为标志的时期),season.n.02: time of year (一年中的一段时间)。本文基于WordNet中的多语义信息,针对输入文本中的每个单词进行多义词识别,即如果该词在知识库中存在至少一个以上可能的词义(Word Sense),则该词将被标注为多义词。并对该词进行之后的词义分析与选择(Sense Selection)。值得注意的是,在歧义感知模型中,我们新建了一个多义词语义向量词典,该词典会在训练迭代过程中进行更新优化。

2.2 语义消歧

歧义感知模块中,我们设计了一个孪生的文本编码器,即语义编码器和词义解释编码器.两个部分结构相似,但是不共享网络权重参数.两个语义编码器分别提取待分析多义词的上下文语义特征和该多义词在WordNet上罗列的所有可能词义的文本解释语义特征.最终获得当前场景下该多义词所有可能词义的权重分布.

针对数据集中的每一个名词和动词 W_t ,我们首先将该词所在的整个文本作为该词的上下文信息,即 $T=\{w_1, \dots, w_L\}$,其中 L 为该句的长度.然后通过NLTK^[23]工具抓取出WordNet中针对 W_t 罗列出的所有可能语义以及对应的解释 $G_t=\{G_{t,1}, \dots, G_{t,N}\}$,其中 N 为该词在词典中罗列的所有词义总数.每一个语义解释都是一句短文本.我们假设,该多义词所在的上下文语义与WordNet罗列出的可能语义所对应的解释越相近,则该词在当前场景下为该语义的可能性越高.

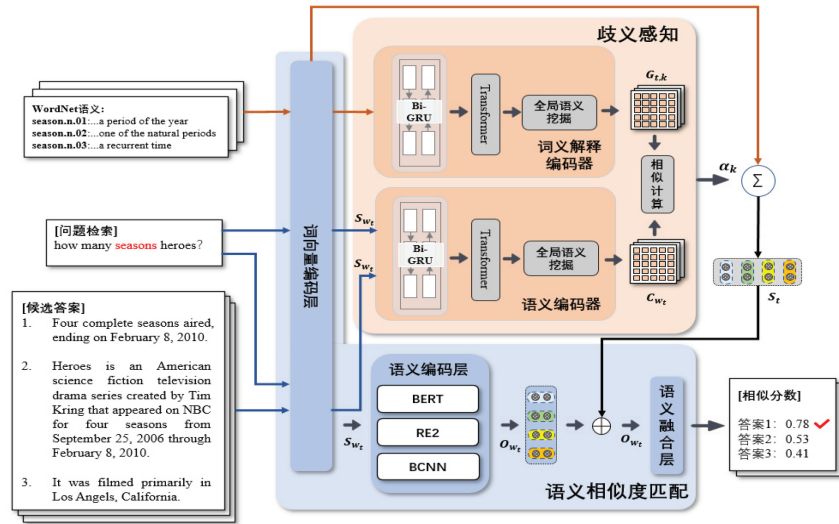


图 2 融合歧义感知的检索式自动问答模型框架

语义编码器.首先我们将整个句子输入到当前基线模型的词嵌入层中,获取初始文本向量.然后将该向量输入到语义编码器中,该编码器首先采用Bi-GRU模块,通过在每个时间步从前后两个方向捕获句子中的上下文信息:

$$\vec{h}_i = \overrightarrow{GRU}(w_i, \vec{h}_{i-1}) \quad (1)$$

$$\vec{h}_j = \overleftarrow{GRU}(w_j, \vec{h}_{j-1}) \quad (2)$$

其中每一个文字特征将由双向GRU的输出特征的平均值表示.通过该层后,我们获得具有初步上下文语义的文本表征 R_{w_t} .

Transformer编码层主要由自注意力机制(Self-attention)组成.其中包含多个编码器(Encoder),而每个编码器中包含一个自注意力机制单元和两层全连接网络.首先我们将上述Bi-GRU层获取的语义向量 R_{w_t} 进行转换并代入该模块中,通过Transformer强大的上下文语义挖掘能力,可获取更为细粒度的语义特征 Z_{W_t} :

$$Z_{W_t} = Attention(Q, K, V) = Softmax(QK^T / \sqrt{d_k})V \quad (3)$$

其中: $Q = K = V = R_{w_t}$, d_k 为缩放因子.得到自适应机制的输出 Z_{W_t} 后,它被送到Transformer中的下一个模块 F ,即前馈网络层(Feed Forward Network, FFN),可以表示为 $\mu_{w_t} = F(Z_{W_t})$.之后Transformer中该编码层的输出 μ_{w_t} 将作为下一个自注意力编码层的输入.经过多个编码层后的输出,为该词 w_i 在上下文解码器的语义向量输出:

$$\mu_{w_t}^t = Encoder_t(\mu_{w_t}^{t-1}, \Theta_t) \quad (4)$$

其中: $Encoder_t$ 为Transformer中第 t 个自注意力编码器, $\mu_{w_t}^{t-1}$ 为第 $t-1$ 个自注意力编码器的输出, Θ_t 为第 t 个自注意力编码器中所有待优化的参数集.

从之前的编码层中,针对每个多义词 w_t ,我们获得了该词所在句子的局部语义表征 μ_{w_t} .在该层将对该表征进行深度语义挖掘与融合,从而获得一个能够概括整个句子的全局语义向量 C_{w_t} .首先对局部语义特征通过内积计算获得权重分布,然后通过加权平均获得该句的全局语义信息:

$$C_{w_t} = \sum_{i=1}^T \text{Softmax}(\mu_{w_t,i}, \mu_{w_t,i}^T) \mu_{w_t,i} \quad (5)$$

接下来,我们将该语义融合层的结果输入到前馈神经网络中,其输出经过一个映射函数则得到该词最终的语义向量 $C_{w_t} = F(C_{w_t})$.

词义解释编码器.词义解释编码器在结构上与上述编码器相似,旨在获取WordNet中针对多义词 w_t 罗列出的所有可能词义对应的短文本解释的全局语义向量 G_t .

语义相似性计算层.该语义相似性计算层旨在计算多义词所在文本的全局语义与候选词义特征之间的相似度.该层我们采用了一个基于注意力机制的方法.该层的输入为词义解释编码器和语义编码器的输出.其目的是计算多义词语义信息与各注解特征的相似度.基于 C_{w_t} 和 $G_{t,k}$ 的相似分数计算方法为:

$$\alpha_{t,k} = (W_{C_{w_t}} C_{w_t})^T (W_{G_{t,k}} G_{t,k}) \quad (6)$$

其中 $W_{C_{w_t}}$ 和 $W_{G_{t,k}}$ 为可学习参数.最后通过Softmax归一化之后获得其权重分布.

2.3 语义消歧

从歧义感知模块我们获取到数据中每个多义词在当前场景下的所有可能词义的权重分布,下一步我们尝试将得到的词义信息融合到问答系统中的问题-答案语义相似性匹配模型中,通常语义匹配模型可分为三个部分,即词向量编码层、语义编码层以及语义融合层.本文针对每个问题与待匹配答案对,结合WordNet挖掘出存在的多义词 w_t ,然后将原始文本通过匹配模型中的向量编码层和语义编码层,获得该文本从基线语义匹配模型中挖掘出的语义特征 O_{w_t} .与此同时,歧义感知模块也学习出该词在上下文中的词义权重分布 $\alpha_{t,k}$ 以及对应每个词义的特征向量 $G_{t,k}$.本文采用基于注意力机制的加权平均方式来获得消歧模型产生的语义特征 S_t .针对每个被分析的多义词,我们的歧义感知模块会生成一个归一化后的所有可能语义的权重分布.最后用该词的语义向量加权平均作为该词的语义向量:

$$S_t = \sum_{1 \leq k \leq M} \alpha_{t,k} \odot S_{t,k} \quad (7)$$

其中 $S_{t,k}$ 为多义词 w_t 的语义特征表示.我们将上述得到的加权平均后的语义特征向量与基线模型的语义编码层获得的文本语义向量进行融合,其融合方式为语义叠加:

$$O_{w_t} = W_{w_t} \oplus F(S_t) \quad (8)$$

其中: O_{w_t} 是多义词 w_t 所在问题或候选答案经过基线模型编码层所获得的语义特征, F 为前馈神经网络.

训练过程中,我们定义了一个为检索问答系统的答案匹配任务收敛的损失函数.采用交叉熵(Cross-Entropy)作为其优化过程.收敛函数为:

$$\mathcal{L}_{\text{match}} = -(1/N) \sum_{i=1}^N y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i) \quad (9)$$

其中: y_i 是二元标签0或1, \hat{y}_i 代表当前样本属于某一类别的预测概率.通过上述方法获得融合了多义词的准确语义,将经过基线模型编码后的语义表征与该向量一同输入到如图2所示的融合层.该融合模型的优势为增强了多义词语义信息对检索问答任务的影响,本文尝试了相加或拼接的融合方式,由于并不是每个词都含有多义词的语义信息,拼接方式容易引入较多的噪音,而叠加融合方式可以很好地提高该细粒度特征信息对整个模型结果的影响,且对其它非多义词的语义影响较小.融合后的向量不仅具有该文本自有的语义向量,同时还具有多义词的语义信息.将融合后的向量再输入到语义匹配基线模型中的特征融合层进行语义融合及特征降维计算,然后得到该文本对的相似分数.最后利用该相似分数对所有候选答案进行排序,最终将最佳匹配答案输出.

2.4 检索问答系统的基线模型

检索式问答方法的核心是问题-答案的语义相似度计算. 因此本文选择了包括BERT在内的三个经典语义匹配模型进行实验对比, 从多个维度证明歧义感知对检索式问答领域的性能影响.

基线模型从采用的实现方法上考虑, 可以分为表示型 (Representation Based) 和交互型 (Interaction Based). 前者的特点是将样本对中的两个文本分别映射成两个向量, 再将两个向量通过神经网络, 最终输出一个相似性分数; 而后者则是将样本对中的两个文本映射成向量之后, 先进行交叉融合, 再将融合后的矩阵输入神经网络, 得到是否相似的结论. 两种类型的模型各有利弊. 因此我们各选择了一个交互型模型和一个表示型模型中的经典模型RE2和BCNN作为其中两个基线模型. 与此同时, 我们也选择了基于Transformer的BERT作为基线模型. 三个基准模型各自代表了三种不同的匹配思想.

BCNN^[24]是2016年提出的一个经典的表示型匹配模型, 该模型由两组卷积网络层组成, 每组卷积网络分别提取样本对中的两个文本特征向量, 然后通过池化层进行特征提取和降维计算, 最后再输入到融合层两个特征向量的相似度计算, 得到其匹配概率. BCNN的特点是模型简单且效率高, 其结果针对短文本匹配任务效果较好.

RE2^[25]是2018年由阿里巴巴达摩院提出的基于交互型的文本匹配模型, 作者在不同任务的公开数据集上分别获得了SOTA结果. 且RE2模型的速度相较其它模型更快, 鲁棒性也有所提升, 同时在不同的文本匹配任务上都取得了更好的成绩.

BERT^[26]是2018年Google AI研究员提出的一种预训练模型, 其采用Transformer Encoder Block进行连接, 是一个典型的双向编码模型. 该模型的预训练结果可以用到其它的任务中.

图3阐释了基于以上三种不同匹配思想的基线模型的歧义感知融合方法细节. 如图3所示, BCNN模型分为词向量编码层、语义编码层以及语义融合层. 我们将歧义感知模型提取的消歧信息 S_t 与编码层输出 O_{w_t} 进行融合, 并将融合后的特征作为其语义融合层的输入, 最终使模型基于歧义感知信息做出准确的匹配判断; RE2模型由词向量编码层、编码层、融合层以及语义融合层组成, 该模型我们选择在最终语义融合层之前将歧义感知信息与其输入相融合. 同样在BERT模型中, 我们将由Transformer组成的编码层输出 O_{w_t} 与歧义感知特征 S_t 融合, 使模型在最后的判断中能够更加精准地理解文本的细节语义, 从而做出更为准确的判断.

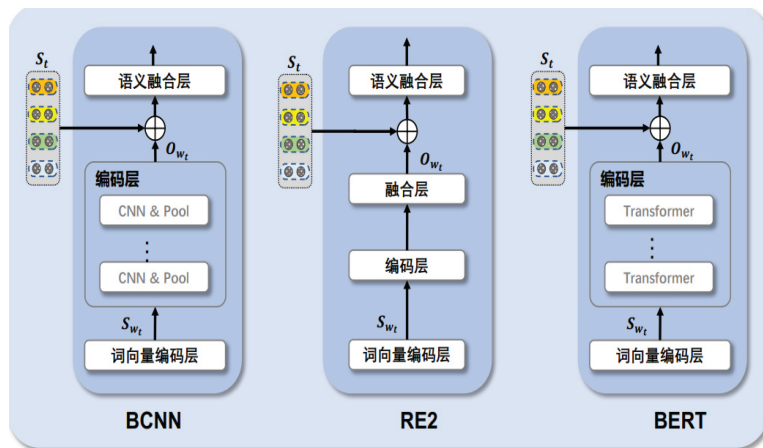


图 3 基于不同基线模型的歧义感知特征融合示例

3 实验

3.1 实验数据与评价指标

我们在检索式问答任务的公开数据集上进行实验, 包括WikiQA与TrecQA. 表1给出了数据集所对应的详细项目信息.

WikiQA^[27]是一个基于Wikipedia的答案选择数据集. 针对开放领域并使用Bing搜索日志作为问题来源, 它包含了问题与其多个候选答案, 使用二元标签判断候选回答是否为它所属的正确答案. 实验中使用该数据集的原始拆分比例, 使用平均精度均值MAP (Mean Average Precision)、平均倒数排名MRR (Mean Reciprocal Rank)

作为评价指标.

TrecQA^[28]是一个为开放域问题回答设置而设计的答案选择数据集.与WikiQA类似,每对数据包含一个问题、一个候选答案和一个二元标签.我们选择数据集的clean版本作为实验数据集,该版本排除了开发和测试集中没有答案和只有正面/负面答案的文本.实验中使用该数据集的原始拆分比例,且同样选择MAP和MRR作为该数据集的评价指标.

表 1 实验数据集

数据集	训练集	验证集	测试集	多义词数量	语义数量
WikiQA	20.4 k	2.7 k	6.2 k	128 k	6 k
TrecQA	53.4 k	1.1 k	1.4 k	200 k	10 k

表1展示了每个实验中用到的数据集的具体信息.其中“多义词数量”表示该数据集中所检测到且进行分析的多义词个数;“语义数量”表示该数据集中被提取到的多义词的不同语义的总数量.

3.2 实验设置

预处理.样本对中每个词将被输入到WordNet词库中查看该词是否为多义词.我们只考虑语义小于等于4个的多义词.运用mask机制表示每个多义词不同的语义个数.

歧义感知模型设置.我们使用基线模型中的文本特征向量层输出的词向量作为歧义感知模型的文本向量输入.与此同时,新建一个关于多义词语义的词汇表,并随机初始化每个语义的特征向量,该向量在之后的参数优化过程中会同时进行优化.为保持与原本词向量的一致性,我们设置该语义向量的维度和词向量的维度相同,如针对BERT基准模型融合,我们设置其语义向量维度为768;针对BCNN和RE2基准模型融合,我们设置其语义向量维度为300.

基线模型设置.针对BERT基准模型,我们的参数设置与文献[28]设置一致,即 $L=12$ 、 $H=768$,参数总数为110 M.批量大小设置为8、16、32,训练轮次为5~7次.针对RE2基线模型,我们的参数设置同样与文献[27]设置一致,如文本最大长度为20、Block最大设置为3等.

4 结果与分析

4.1 总体结果分析

我们分别在WikiQA和TrecQA两个答案选择任务的公开数据集上对设计的融合歧义感知的检索式问答模型进行性能验证.表2为本文融合模型与其它研究在相同数据集上的性能比较.由表2可知,我们设计的三个融合歧义感知模型在MAP与MRR评估下均优于其基线模型.其中基于BCNN模型的融合模型的MRR评估比BCNN基线模型提高近1%.融合了歧义感知的BERT模型,在测试数据上的MAP评估达到83.5,其性能优于BERT基线模型,在相同数据集下也优于近几年其它相关研究.

表 2 融合歧义感知的模型在答案选择任务数据集上的性能对比

模型	WikiQA		TrecQA		
	MAP	MRR	MAP	MRR	
其它相关研究	SS-BERT ^[29]	83.4	84.8	-	-
	Sem-BERT ^[30]	83.1	85.1	87.8	89.5
	BCNN-base	69.2	71.1	73.8	74.5
	BCNN-sense	69.8	71.9	74.3	75.1
歧义感知	RE2-base	74.5	76.2	75.8	77.1
	RE2-sense	75.2	76.7	76.4	77.5
	BERT-base	83.3	84.1	85.7	90.2
	BERT-sense	83.5	84.2	85.9	90.2

4.2 多义词的结果分析

为了验证融合歧义感知模型性能的改善,我们对被模型标记为多义词所在的样本进行了进一步的分析.选择WikiQA和TrecQA数据集,用多义词所在的文本构建了一个子集(称为多义词测试集),该子集仅包含被检测

到多义词的文本. 然后, 在多义词测试集和原始测试集上运行我们之前选择的三种基线模型以及对应的融合歧义感知模型, 并比较二者的性能差异. 由表3可知, 针对MAP评估标准, 每个语料库的多义词测试集上的性能低于基线模型性能, 这意味着数据集中语义模棱两可的单词会阻碍基线模型做出正确决策. 在歧义感知模块的作用下, 所提模型在多义词测试集上实现了比原始测试集更高的精度. 这意味着我们设计的歧义感知模块确实有助于自动问答模型正确区分句子之间的语义差异.

表 3 融合歧义感知的三个模型在多义词数据集与原始数据集上的性能对比

模型	模型	WikiQA		TrecQA	
		原始测试集	多义词测试集	原始测试集	多义词测试集
BCNN	基线模型	69.2	68.5(-0.7)	73.8	73.2(-0.6)
	歧义感知	69.8	70.2(+1.7)	74.3	74.7(+1.5)
RE2	基线模型	74.5	73.9(-0.6)	75.8	75.3(-0.5)
	歧义感知	75.2	75.3(+1.4)	76.4	76.5(+1.3)
BERT	基线模型	83.3	82.8(-0.5)	85.7	85.1(-0.6)
	歧义感知	83.5	84.4(+0.6)	85.9	85.9(+0.8)

4.3 案例分析

为了更好地理解歧义感知对自动问答的影响机理, 图4展示了WikiQA测试数据集中挑选的两个案例. 第一个案例中有两个候选答案, 基线模型错误的将答案2判断为最佳答案. 而通过歧义感知的语义增补, 我们的方法能够正确地将答案1判断为该问题的最佳答案. 同理, 在第二个案例中, 数据集提供了三个候选答案. 通过对问题和候选答案中的多义词(如“country”“bank”)的正确理解, 我们的方法能够正确预测出该问题的最佳答案为第3个答案.

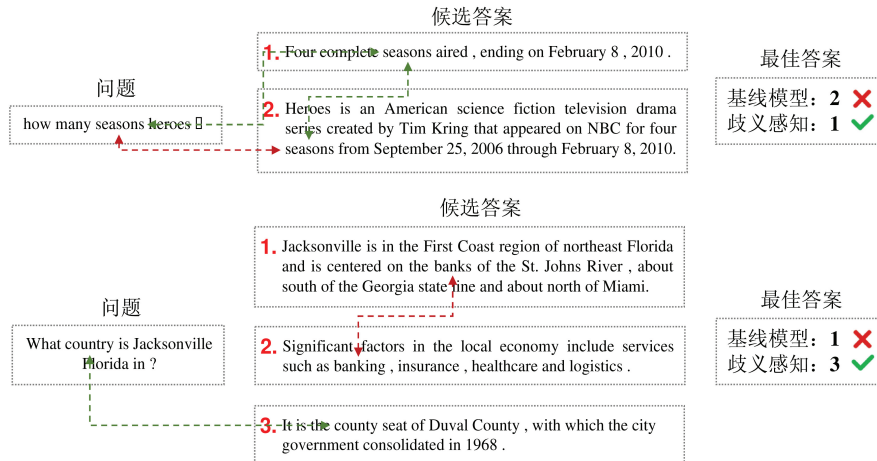


图 4 WikiQA数据集中包含多义词的示例

5 总结

针对多义词在检索式问答数据中的影响, 本文设计了三个融合歧义感知的检索式问答模型, 使问答模型在计算问题-答案语义相似度的过程中将文中多义词的语义进行检测, 并将检测得到的多义词的语义进行编码表达, 再融入到原任务中进行计算. 我们设计了一个基于Transformer的语义感知模型, 深度挖掘输入文本语义上下文特征表达, 并借助WordNet提供的语义注解信息, 判断该词在当前语境中的精准语义. 然后将消歧后的语义信息进行编码并融合到问题-答案语义相似性计算中. 设计的基于BCNN、RE2以及BERT的融合模型, 尤其是基于BERT的问答模型, 性能不仅优于其基线模型, 而且比近年来的相关研究性能更优.

参考文献:

- [1] LI F L, QIU M H, CHEN H Q, et al. *AliMe Assist: An intelligent assistant for creating an innovative e-commerce experience*[C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. Singapore, Singapore. ACM, 2017: 2495-2498.

- [2] CUI L, HUANG S H, WEI F R, et al. SuperAgent: A customer service chatbot for e-commerce websites[C]//Proceedings of ACL 2017, System Demonstrations. Vancouver, Canada. Stroudsburg, PA, USA: Association for Computational Linguistics, 2017: 97-102.
- [3] SONG S Y, WANG C, PU X, et al. An enhanced convolutional inference model with distillation for retrieval-based QA[C]//International Conference on Database Systems for Advanced Applications. Cham: Springer, 2021: 511-515.
- [4] HAO T Y, LI X X, HE Y L, et al. Recent progress in leveraging deep learning methods for question answering[J]. Neural Computing and Applications, 2022, 34(4): 2765-2783.
- [5] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. December 4-9, 2017, Long Beach, California, USA. ACM, 2017: 6000-6010.
- [6] MEDSKER L, JAIN L C. Recurrent neural networks: Design and applications[M]. New York: CRC Press, 2001.
- [7] LI F F, LU Y H, MAO X L, et al. Multi-task deep learning model based on hierarchical relations of address elements for semantic address matching[J]. Neural Computing and Applications, 2022, 34(11): 8919-8931.
- [8] HADIWINOTO C, NG H T, GAN W C. Improved word sense disambiguation using pre-trained contextualized word representations[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). HongKong, China. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019: 5297-5306.
- [9] SINGH V P, KUMAR P. Word sense disambiguation for Punjabi language using deep learning techniques[J]. Neural Computing and Applications, 2020, 32: 2963-2973.
- [10] WANG Y L, WANG M, FUJITA H. Word sense disambiguation: A comprehensive knowledge exploitation framework[J]. Knowledge-Based Systems, 2020, 190: 105030.
- [11] CHEN R C, YULIANTI E, SANDERSON M, et al. On the benefit of incorporating external features in a neural architecture for answer sentence selection[C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. August 7-11, 2017, Shinjuku, Tokyo, Japan. ACM, 2017: 1017-1020.
- [12] KUNDU S, NG H T. A question-focused multi-factor attention network for question answering[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 5828-5835.
- [13] DU X Y, SHAO J R, CARDIE C. Learning to ask: Neural question generation for reading comprehension[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver, Canada. Stroudsburg, PA, USA: Association for Computational Linguistics, 2017: 1342-1352.
- [14] YAN R, SONG Y P, WU H. Learning to respond with deep neural networks for retrieval-based human-computer conversation system[C]//Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. July 17-21, 2016, Pisa, Italy. ACM, 2016: 55-64.
- [15] WU B W, WANG B X, XUE H. Ranking responses oriented to conversational relevance in chat-bots[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics. December 11-17, 2016, Osaka, Japan. 2016: 652-662.
- [16] AGIRRE E, EDMONDS P. Word sense disambiguation: Algorithms and applications (Text, Speech and Language Technology)[M]. Berlin: Springer, 2007.
- [17] MILLER G A. WordNet: A lexical database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.
- [18] BANERJEE S, PEDERSEN T. An adapted Lesk algorithm for word sense disambiguation using WordNet[M]//Computational Linguistics and Intelligent Text Processing. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002: 136-145.
- [19] BASILE P, CAPUTO A, SEMERARO G. An enhanced Lesk word sense disambiguation algorithm through a distributional semantic model[J]. COLING 2014 - 25th International Conference on Computational Linguistics, Proceedings of COLING 2014: Technical Papers, 2014: 1591-1600.
- [20] MORO A, RAGANATO A, NAVIGLI R. Entity linking meets word sense disambiguation: A unified approach[J]. Transactions of the Association for Computational Linguistics, 2014, 2: 231-244.
- [21] HUANG L Y, SUN C, QIU X P, et al. GlossBERT: BERT for word sense disambiguation with gloss knowledge[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). HongKong, China. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019: 3509-3514.
- [22] YAP B P, KOH A, CHNG E S. Adapting BERT for word sense disambiguation with gloss selection objective and example sen-

- tences[C]//Findings of the Association for Computational Linguistics: EMNLP 2020. Online. Stroudsburg, PA, USA: Association for Computational Linguistics, 2020: 41-46.
- [23] BIRD S. NLTK: The natural language toolkit[C]//Proceedings of the COLING/ACL on Interactive Presentation Sessions. ACM, 2006: 69-72.
- [24] YIN W P, 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.
- [25] YANG R Q, ZHANG J H, GAO X, et al. Simple and effective text matching with richer alignment features[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019: 4699-4709.
- [26] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[EB/OL]. 2018: arXiv: 1810.04805. <https://arxiv.org/abs/1810.04805.pdf>.
- [27] YANG Y, YIH W T, 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, Portugal. Stroudsburg, PA, USA: Association for Computational Linguistics, 2015: 2013-2018.
- [28] WANG M Q, SMITH N A, MITAMURA T. What is the Jeopardy model? A quasi-synchronous grammar for QA[C]//EMNLP-CoNLL 2007 - Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. June 28-30, 2007, Prague, Czech Republic. Stroudsburg: ACL, 2007: 22-32.
- [29] LIU T, WANG X, LYU C G, et al. Sentence matching with syntax- and semantics-aware BERT[C]//Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain(Online). Stroudsburg, PA, USA: International Committee on Computational Linguistics, 2020: 3302-3312.
- [30] ZHANG Z S, WU Y W, ZHAO H, et al. Semantics-aware BERT for language understanding[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(5): 9628-9635.

责任编辑: 张自强

(上接第 26 页)

- [2] LIU R F, ZHAI M Q, SHU J L. The least eigenvalue of unicyclic graphs with n vertices and k pendant vertices[J]. Linear Algebra and Its Applications, 2009, 431(5/6/7): 657-665.
- [3] YE M L, FAN Y Z, LIANG D. The least eigenvalue of graphs with given connectivity[J]. Linear Algebra and Its Applications, 2009, 430(4): 1375-1379.
- [4] FAN Y Z, ZHANG F F, WANG Y. The least eigenvalue of the complements of trees[J]. Linear Algebra and Its Applications, 2011, 435(9): 2150-2155.
- [5] JIANG G S, YU G D, SUN W, et al. The least eigenvalue of graphs whose complements have only two pendent vertices[J]. Applied Mathematics and Computation, 2018, 331: 112-119.
- [6] GUO J M, ZHANG G G, WANG Z W, et al. The least eigenvalue of unicyclic graphs with application to spectral spread[J]. Algebra Colloquium, 2022, 29(2): 265-272.
- [7] LI S C, WANG S J. The least eigenvalue of the signless Laplacian of the complements of trees[J]. Linear Algebra and Its Applications, 2012, 436(7): 2398-2405.
- [8] LI Y J, QIN R, LI D. On distance signless Laplacian spectrum of the complements of unicyclic graphs and trees[J]. Linear Algebra and Its Applications, 2021, 631: 235-253.
- [9] 卢志琴, 马小玲. 两种分裂点连接运算图的 Randić 谱[J]. 新疆大学学报(自然科学版)(中英文), 2022, 39(5): 550-559.
LU Z Q, MA X L. The Randić spectrum of two types of splitting-vertex join graphs[J]. Journal of Xinjiang University(Natural Science Edition in Chinese and English), 2022, 39(5): 550-559. (in Chinese)
- [10] QIN R, LI D, CHEN Y Y, et al. The distance eigenvalues of the complements of unicyclic graphs[J]. Linear Algebra and Its Applications, 2020, 598: 49-67.
- [11] HAEMERS W H. Interlacing eigenvalues and graphs[J]. Linear Algebra and Its Applications, 1995, 226/227/228: 593-616.

责任编辑: 赵新科