1. 结论与展望

5.1 结论

答案选择排序是社区问答系统的重要环节，同时也是基于答案对的问答系统中的一个重要研究内容，作为自然语言处理领域中一个备受关注的研究方向，其研究目的是对问答系统中通过检索、匹配等获得的问题的多个候选答案进行排序，并选择出最优解作为问题的最佳答案；而深度学习方法，因为其可以在不依赖特征工程和外部资源的情况下，通过深度神经网络自动地学习问句和候选答案的特征，在答案选择排序任务中有着越来越广泛的应用；

1）构建了基于深度学习的答案选择排序算法框架，框架思路为：以问句和候选答案为输入，通过词向量计算，将问句和候选答案转化为词向量矩阵作为神经网络的输入端，然后通过深度神经网络结构对问题答案对进行句子语义建模，最后，将答案选择排序问题看做二分类为题，使用sigmoid函数将问题和答案的语义相关性得分转换为0到1之间的概率值，依此概率值对候选答案进行排序，将概率值最大的候选答案作为最佳答案；此框架为后续词向量计算和句子对建模明确了研究目标和研究思路；

2）答案选择排序模型中神经网络的输入端是词向量，目前所有的词向量计算工具及词向量模型都为了保证不同语言的普适性，大多是以词语为单位训练词向量的，本文分析了汉语言的特点，发现组成词语的字往往在词语的含义表达上有非常重要的影响和作用，因此，本文做了如下改进：提出了字和词语联合训练的词向量计算方法，将此方法应用于CBOW词向量模型的训练过程中；由于字的歧义性，设定了多标准的字向量确定方法，根据汉字在词语中的不同位置以及汉字所在的不同组合，构建了基于位置的字向量以及基于簇的字向量，并提出了将字的位置信息和簇信息进行融合来确定字向量的方法，最终完成词向量的计算；在数据集上对相关算法进行验证，实验结果表明，字和词语联合训练的CBOW词向量模型相比传统的CBOW词向量模型具有更好的性能，特别是基于位置和簇的方法，相比于基于位置的方法、基于簇的方法有更好的效果；

3）在得到问句和候选答案的词向量后，以词向量矩阵作为神经网络的输入，进而对问句和候选答案进行句子对建模，本文分别构建了基于CNN、BILSTM的句子对建模模型，发现BILSTM在处理自然语言序列信息时有着独特的优越性，所以在后续融合Atention机制的模型中都是以BILSTM为基础获得句子语义信息的，通过Attention计算可以得到句子中不同词语的权重以获得更好的语义特征表示，但是一般的Attention模型在处理答案选择排序问题时，句子中词语的权重的计算都是依赖于问句和候选答案的某种相似性或关联性，这对于答案选择排序中大量处理的非事实类问题而言，是不太合理的，所以，本文做了如下改进，提出了基于问句和答案独立性的Attention模型，认为句子中词语的重要性并不是建立在其和问句或答案有多少相似性或关联性的基础上，而是建立在其对于本身句子的语义特征表示所起的作用上，在相关数据集上的实验结果表明，在答案选择排序任务中，基于深度学习的方法比非深度学习的方法有更好的性能表现，特别是基于问句和候选答案独立性的Attention模型，在句子对建模时，根据词语的重要性赋予其不同的权重，获得了更好的句子语义特征向量，更好的提高了答案选择排序模型的性能；

5.2 展望

本文针对答案选择排序问题，研究了一套基于深度学习的较为有效的方法，改进的算法虽然得到了一定的效果提升，但仍有一部分细节问题需要进一步深入研究或改正以达到更好的效果，今后的努力方向主要有以下几个方面：

1）本文将字和词语联合训练的词向量计算方法融合于CBOW词向量模型，未来可以尝试将其应用于skip-gram或Glove，或许会有更好的效果；

1. 本文仅仅只是在问句和候选答案的文本内容上进行基于深度神经网络的句子对建模，而在实际的问答系统中中，答案获取渠道的可靠性、答案回答者的知识等级、问题的类型等对答案选择排序的效果都有一定的影响，今后可以进一步考虑将这些信息加入到训练过程中以提高模型的性能；
2. 本文中基于问句和答案独立性的Attention模型是通过双层BILSTM实现的，未来可以从self-Attention的角度出发进一步研究句子的内部结构，以及如何更好地将每个序列中的单元和该序列中的所有单元进行Attention计算；