**论文题目：基于深度学习的答案选择排序研究**

**类别（领域）：工程硕士（软件工程）**

**学位申请人：高潇洒**

**指导教师：刘跃虎教授**

摘要

随着人工智能时代的到来，对各类智能问答系统的研究成为实现人类与机器无障碍交流的一个重要切入点，社区问答的出现，为基于问答对的问答系统提供了稳定可靠的问答数据来源，答案选择排序作为社区问答系统中的关键任务以及基于问答对的问答系统中的重要研究内容，主要是通过对检索得到的候选答案进行排序，选择最佳答案返回给用户，现阶段对于此问题的研究大多是通过提取问句与候选答案的句法及语义信息，然后计算问题与候选答案的相关性，特别是近几年，相比于句法结构相关性，研究人员更多关注的是语义相关性；然而，传统的答案选择排序研究，大多需要依赖大量的语言工具、外部资源或特征工程，因此，研究基于深度学习的答案选择排序具有重要的意义。

本文针对模型输入端词向量的计算问题，首先提出了一种字和词语联合训练的词向量计算方法，并根据汉语语言的特点，提出将基于位置的方法和基于簇的方法结合以确定字向量，最后，将字和词语联合训练的方法应用于CBOW词向量模型中，并训练改进的词向量模型，得到问句和候选答案语料库中所有词语的词向量；以问句和候选答案的词向量矩阵作为答案选择排序模型的输入对句子对进行建模，构建了基于CNN、BILSTM的句子对建模模型，并在BILSTM模型的基础上融合Attention 机制计算句子中各个词的权重，由于考虑到非事实类问题存在如下特性：大部分的候选答案和问句之间并没有很高的相似性或关联性；和问句有很多相似词语的候选答案并不一定是问题的最佳答案；提出了基于问句和候选答案独立性的Attention模型self Attention，该模型利用一个附加的BILSTM进行Attention计算，以词语在句子本身语义表达中所起的作用来衡量词语的重要性，并赋予其不同的权重，最终，获得更好的句子语义特征向量；

综上，在相关数据集上对本文提出的算法进行实验验证；基于字和词语联合训练的CBOW词向量模型相比传统的CBOW词向量模型表现出了更好的性能，基于深度学习的答案选择排序方法相比非深度学习方法性能有了很大的提升，特别是基于问句和候选答案独立性的self Attention模型，在MAP和MRR两个评价指标上达到。。。和。。。，均高出已有研究结果，充分证明了本文提出的方法的合理性和有效性；

关键字：答案选择排序 深度学习 词向量 句子对建模 self Attention

论文类型：应用研究