

何苏

13031037678 hesu@pku.edu.cn

教育经历

2016.09 - 至今 北京大学 软件与微电子学院 硕士

2012.09 - 2016.06 北京邮电大学 计算机学院 本科

项目经历

讯飞研究院--阅读理解组 nlp 算法实习生

2017.10 - 2018.04 智能客服项目--汽车说明书

1, 拒答模块：对于用户提出的问题，来确定是否给与回答，我们是用 CNN 层提取特征来做预测。调整正负比例、卷积核等后，最终准确率为 98%。

2, 搜索浏览模块：对于用户提出的问题，找到相应的文档。将 query、文档的 title 分别用 CNN 处理，文档的 document 用 bi-lstm 处理。再分别计算 query 和 title 的相似度，query 和 doc 的相似度，用这两个相似度做预测。我们在实验中尝试了添加 word 共现特征、共享 CNN、使用 mask 层等。在 181 个文档上的准确率为 95%，在 511 个文档上的准确率为 86%。

3, 内容详情模块：对于文档中的若干个句子，找到连续的几句来回答问题。对 query 和 doc 中的每一个句子都用 lstm 处理，计算他们的 word 级别的 attention。对于 doc 的 emb 表示用 CNN 处理，经过 pool 层后获得的新表示再与 query 算句子级别的 attention。然后对于 doc 用 CNN 获得的新表示还可以继续用 LSTM 处理，获得 doc 中每一个句子的上下文表示，进而与 query 获得 context 级别的 attention。最终用这三个 attention 进行预测。最终模型在 181 个文档的准确率为 80%，在 511 个文档上的准确率为 91%。

2018.04 -- 2018.06 参与多轮对话项目--ubuntu/douban 对话数据

我们将内容详情模块的模型移植到了 ubuntu/douban 对话数据上，经过近一个月的实验，模型参数精调、模型用 SGD 再训、多模型融合等，最终取得了 state-of-the-art 的效果，P@1 上达到了 47.3%。我作为论文作者之一正在投稿。

2018.07 -- 2018.12 选择题阅读理解项目 --ARC数据

具体功能：对于英文科学知识选择题问题，我们希望模型能自主回答选择题，从四个答案中选出正确的选项。

模型细节：我们先通过将 question 和每个选项 choice 结合起来，然后去外部知识库中用 elasticsearch 搜索相应的知识 corpus。然后针对每个 question、choice、corpus 来建模。先分别用 bi-lstm 处理输入获得他们的时序表示，然后分别计算 choice 内部和 choice 之间的 match，用 CNN 提取最大特征等。最后我们在 Challenge 数据集上准确率达到 39.76%，Easy 数据集上达到了 63.91%。我作为论文的第一作者正在投稿(已投过 AAAI，得分 6,6,4)。

自我评价

对于 nlp 中的阅读理解任务有一定的了解，有一定的论文阅读量。具有良好的团队沟通能力，能承受较大的压力。