2018/10/12 会议记录

孙焱

QA（问答）系统代码编写：

代码结构

pycharm连接git，同步代码；GayHub插件显示侧边栏；pycharm远程代码调试；导入包，jieba；ujson用c编写，python封装，速度比json快。

文本处理过程，处理数据然后序列化；加载数据，pandas统计；分词；去掉停止词；用的哈工大停止词库，用了python里的列表推导式；剪枝；去掉停止词；加载预训练词向量，刚开始用的人民日报的词向量，后来用了百度百科的词向量；词向量可视化；训练模型。

如何评估结果好坏？

1.得到一个词语top-N的相关词，看其是否合理；

2.

杨帆

1. 机器学习复习

Word2vec词向量模型与N-gram模型。Word2vec词向量考虑上下文，N-gram模型考虑前N-1个词。

1. 实验方案进度

提出tree-tensor-convLSTM模型。

1. 拓展思路创新

细粒度情感分析：解决反讽问题，Aspect-level情感分类。