

武汉大学2017级硕士毕业论文开题报告答辩

融合情感信息的机器阅读理解方法研究

答辩人：郭天翼 指导老师：彭敏

2019.11.26

武汉大学计算机学院自然语言处理研究中心



选题背景与研究现状

相关工作

研究内容与方法

进度安排

选题背景与研究现状



领域背景

- 互联网文本数据资源容量巨大，增长速度快，使用人工处理数据过于消耗人力资源
- 人工智能技术的发展为机器抽取大量文本中的有效信息提供了高效便捷的方法
- 人类需求与机器之间的人性化交流，因此机器学习并运用情感非常有必要

国内外研究现状

- 机器阅读理解的研究趋于多样化与丰富化：单文档、多文档；抽取式、生成式；基于文档本身与引入外部知识。
- 机器阅读理解的模型与评测较多，包括DrQA、DuReader等模型与SQuAD、CoQA、QuAC、XQA、Cosmos QA等数据集
- 融合情感的对话系统研究也在迅速发展



学术研究意义

- 融合情感信息与机器阅读理解的研究点是先前学者较少研究的，具有创新意义
- 情感作为人类生理的一大特征，对它的研究可以有效地推进人工智能的研究发展

经济与社会效益

- 具有情感特征的阅读理解问答机器人，更加人性化，具有使用价值
- 能够与用户进行人际交流的问答机器人，必须具备感知和表达情感的能力
- 使用智能问答机器人，可以在一定程度上改变社会服务业现状，促进形成新的服务业体系

相关工作

文本情感分析方法

文本情感分析方法：基于字典、基于深度学习

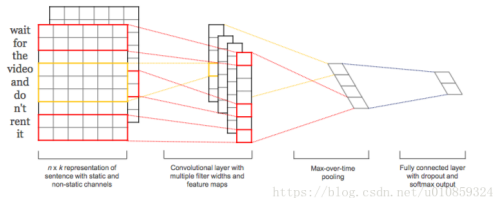


图: Convolutional Neural Networks for Sentence Classification(2014)

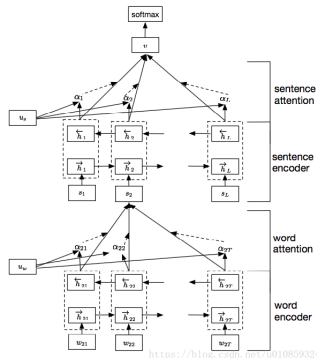


图: Multilingual Hierarchical Attention Networks for Document Classification(2017)



Seq2Seq自然语言生成方法

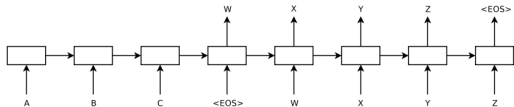


图: Sequence to Sequence Learning with Neural Networks(2014)

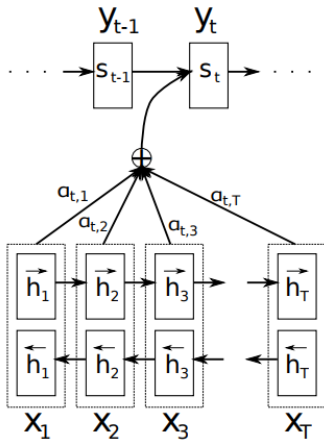


图: NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE(2014)

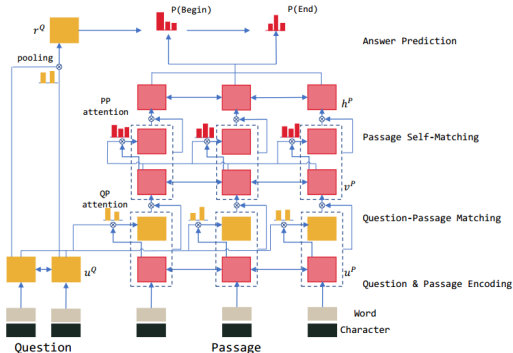


图: R-NET: MACHINE READING COMPREHENSION WITH SELF-MATCHING NETWORKS(2016)

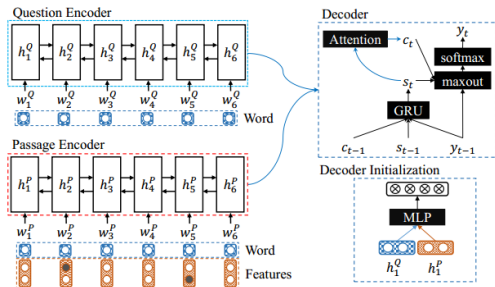


图: S-NET: FROM ANSWER EXTRACTION TO ANSWER GENERATION FOR MACHINE READING COMPREHENSION(2017)

研究内容与方法

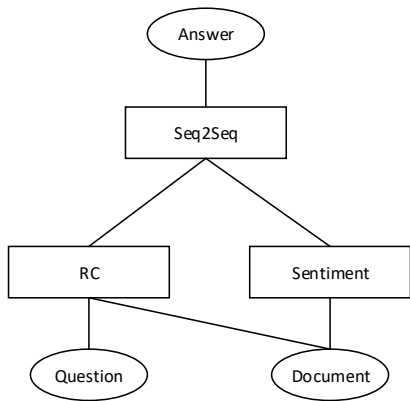


难点

- 如何将从文本中提取出文本的情感特征与阅读理解的答案进行融合
- 针对不同的问题，可能需要表达出不同的情感，如何根据问题与答案表达出复杂多变的情感

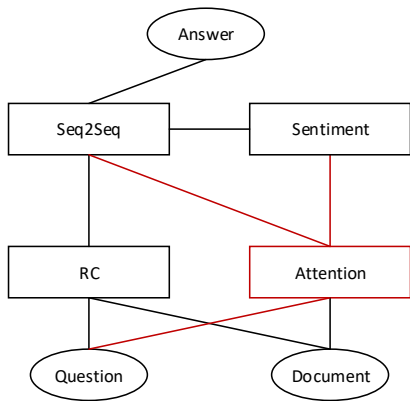
主要贡献与创新

- 提出了融合情感嵌入的机器阅读理解方法
- 提出了基于注意力机制的多层次情感融合方法



- 通过情感分析模块提取文档情感嵌入
- 通过阅读理解模块抽取不含情感因素的答案
- 将情感嵌入与答案送入Seq2Seq模型，得到融合情感嵌入的答案

图：融合情感嵌入的机器阅读理解方法



- 通过文档与问题的Attention获取问题与文档的相关度权值
- 通过带权文档提取情感嵌入
- 通过阅读理解模块抽取不含情感因素的答案
- 将情感信息、Attention信息与答案共同送入Seq2Seq模型，得到多层次情感融合答案

图: 基于注意力机制的多层次情感融合阅读理解方法



数据集选择

- 情感数据集：Chinese Sentiment set
- 阅读理解数据集：CMRC2018数据集

对比模型

- 基于注意力机制的多层次情感阅读理解模型（本文模型）
- 融合情感嵌入的机器阅读理解模型（本文模型）
- BERT-base
- XLNet-mid



评价指标

- 阅读理解评价指标：F1、EM、BLEU、ROUGE-L
- 情感评价指标：人工评估情感检测准确率

进度安排



时间安排	工作安排
2019.10-2019.11	论文选题，查阅相关文献资料，撰写开题报告
2019.12-2020.2	设计模型，完成实验
2020.3-2020.4	撰写论文
2020.5-2020.6	论文答辩