

赵怀鹏

✉ huaipengzhao@gmail.com ☎ (+86) 130-6987-3530

🎓 教育背景

- | | |
|--|-------------|
| 哈尔滨工业大学 | 2016 – 至今 |
| 在读硕士研究生 社会计算与信息检索中心 (SCIR), 导师: 车万翔, 预计 2018 年 6 月毕业 | |
| 哈尔滨工业大学 | 2012 – 2016 |
| 学士 计算机科学与技术, 学分绩排名前 10% | |

👨‍💻 项目/实习经历

- | | |
|--------------|-----------------|
| 中文顺滑 | 2016 年 9 月 – 至今 |
| 语音后处理研究 主要项目 | |

主要研究中文语音识别之后文本不规范的问题。主要尝试的方法有以下几种:

- 序列标注模型, 这里尝试了 CRF, BI-LSTM, BI-LSTM-CRF 三种模型。实验结果来看后两种模型效果相似, 但明显好于 CRF, 达到了 83% 左右的 F 值。
- 级联模型, 将 LSTM 的结果作为 CRF 的特征, 做了 Stacking 操作, 比单纯的 CRF 要好一些。
- 基于生成的 Pointer Network 模型, 这是基于 seq2seq 的改版, 在生成词的时候用一个指针去和原文做 windows attention, 选出 windows 下分数最高的词。效果要好于 BI-LSTM-CRF。
- 最后尝试了 Transition-based 模型, 效果基本和 Pointer Network 持平。

- | | |
|-----------------|-------------------------|
| 讯飞研究院实习, 北京 | 2016 年 5 月 – 2016 年 9 月 |
| 研究员 负责语文主观题评测工作 | |

对高考语文主观题进行机器打分, 在定标实验下达到老师评分水平:

- 抽取了一些相似度离散特征, 例如基于 embedding 的余弦相似度, 编辑距离等。
- 采用了一些回归模型预测分数: SVR, Attention-Based-LSTM, 并且也做了联合 Stacking, 从结果来看, 联合模型得到的结果在 0 分一致率, 1 分一致率等指标上已经全面超过老师评分水平。

- | | |
|------------------|-------------------------|
| 语文高考题干分析, 863 项目 | 2016 年 3 月 – 2016 年 7 月 |
| 本科毕设 哈工大百优论文候选 | |

对高考语文题干进行分析, 用来抽取答题引擎需要的信息。

- 抽取答案个数, 触发词, 约束条件等语义槽。
- 采用了 SVM 和 CRF 模型, 实验结果证明该问题更适合序列标注模型。

🔧 IT 技能

- 编程语言: 常用语言: Python, 有一定 C++ 基础
- 研究领域: 自然语言处理, 深度学习, 中文顺滑

♥ 获奖情况

- | | |
|-------|------|
| 国家奖学金 | 2013 |
|-------|------|

📄 其他

- 技术博客: <http://hpzhao.com>
- 翻译书籍: Deep Learning and Neural Networks
- 外语: 英语六级