

陈佳豪

手机：18166891323 | 邮箱：2808581543@qq.com

个人主页：<http://blog.tabchen.com>



自我评价

具有较扎实的机器学习和深度学习理论基础，工作期间以第一作者发表3篇国际论文，获得2项世界比赛冠军。富有创新意识，以第一发明人向专利局提交5篇发明专利，1个已经授权，其余4项正在审核中。有云计算实践经验，有多个服务运行在自行构建的docker集群中。

教育背景



吉林大学 985

数学学院 信息与计算科学 本科

2014-09 至 2018-06

工作经历

北京世纪好未来教育科技有限公司 (NYSE : TAL)

自然语言处理高级工程师(集团AI中台)

2018-02 至今

工作职责和贡献：

- 负责关键词识别、文本分类等任务。完成了通用文本分类框架(ATC)的开发，可一键运行30+模型。
- 应用机器学习/深度学习等技术解决教育场景中问答识别、课堂精彩片段捕获等任务。重新定义教育场景中问句类型，可以捕获更细粒度的问句；实现4种课堂片段提取，可用于评估师生上课表现。
- 文本生成前沿探索和应用。利用gpt2/3实现了作文素材生成系统，素材人工审核通过率达到90%。

项目经历

作文素材生成系统

2020-12 至 2021-04

- 项目背景：根据用户输入关键词或者标题信息，自动生成高质量作文素材。挑战是如何生成长度、题材可控，且符合常识、逻辑的片段。
- 主要贡献：构建基于关键词和标题的作文素材自动生成系统，对GPT2/3进行post-train，然后进行fine-tuning，并提出逻辑、切题、符合常识三个后处理模块，人工审核通过率从32%提升到90%。

通用文本分类工具ATC

2020-08 至 2020-12

- 项目背景：开发通用的文本分类工具，覆盖学术界领先模型和工业界常用模型。挑战，需要对所有模型提供统一环境、接口、工程代码。
- 主要贡献：完成框架开发，可以一键运行30+模型。目前支持3+线上服务，凭借工具获得“CCL2020 幽默识别”冠军。

课堂精彩片段提取

2020-06 至 2020-12

- 项目背景：自动提取1对1课堂中师生互动的精彩片段，包含4种主要场景。挑战有如何精准识别开始点、如何客观对片段打分。
- 主要贡献：针对四个场景，利用BERT及其变种构建片段开始点检测模块，在4个场景识别准确率均达到90%以上。片段质量评估模块，针对标注主观性较强问题，提出基于pair的标注方法获得更客观的标签。片段质量评估模块，首先构建片段级特征，然后利用GBDT模型进行打分，pair一致率93.4%。

职业技能

自然语言处理 熟练

Python 熟练

机器学习 熟练