

张 智

香港中文 计算机 硕士

+86-15976870920
✉ cheungZeeCn@gmail.com



教育背景

- 2013.09–2014.07 理学硕士, 计算机科学与工程系, 香港中文大学 (Top 30 CSE).
曾获得 优秀学生奖学金(约前五)
- 2006.09–2010.07 学士, 计算机科学与技术, 哈尔滨工业大学 (985).
曾获得 三等奖学金(25/180) 和 单项奖学金(优秀班干)

工作经历

- 2018.09–至今 资深算法工程师, 人工智能研发团队, 平安寿险, 深圳.
知识库知识挖掘, 智能陪练机器人, 知识中台等;曾获最佳协作奖,香蜜湖金融科技创新奖等
- 2016.11–2018.03 大数据分析师, 招银前海金融资产交易中心, 深圳.
资讯系统和舆情系统开发,大数据系统构建,运营数据分析
- 2015.07–2016.11 合伙人, 浮图科技(深圳)有限公司, 深圳.
数据系统,数据采集/存储/分析/展示,辅助投资决策
- 2015.02–2015.07 系统工程师SE, 大数据解决方案与设计部, 华为, 深圳.
负责大数据架构类技术设计
- 2010.07–2013.05 高级运维工程师, 百度, 北京.
流量接入和调度, 用户访问质量统计分析和调度;曾获最佳质量奖

技术特长

- 数据挖掘与机器学习 熟练使用Python工具栈来进行分析与挖掘, 掌握常用的机器学习理论和应用,如TEXTCNN, TRANSFORMER, BERT, LDA, HMM等.
掌握Pytorch, Tensorflow等学习框架的使用
- 数据分析和报表 业务Sense强,可快速理解和把握业务需求并进行分析抽象, 可在大数据环境下进行数据分析和展示
- WEB/LINUX 熟悉Web各个层次的技术和原理与Web开发(Php,Laravel), 熟悉Linux系统管理与编程,有独自开发稳健系统的能力
- 编程语言 熟悉Python,C,掌握Scala,Java及Web相关语言

项目经验

公司项目(平安寿险): rebert模型, 基于bert + Levenshtein Transformer的纠错

2020.10-至今 **[1人 自发项目|非自回归生成模型用于端到端离线纠错]**
对标soft-masked bert和小米最近对外分享的语音识别后的文字纠错;将bert的语义信息和pinyin的发音信息结合在一起,借鉴levenshtein-transformer的生成架构来支持输入输出长度不一的纠错情况;目前完成如下几个工作:1.fairseq框架掌握和改造;2.bert集成到fairseq中,并公开给网友使用;3.rebert模型开发,目前调试中. 参考
<https://github.com/CheungZeeCn/fairseq/blob/master/rebert.md>

公司项目(平安寿险): 知识中台相关

2020.04-至今 **[5人, 参与者 |内容理解,图谱构建,概念关联等]**
概念图谱:创新型基础服务,用于支持推荐/检索;分内容理解,图谱构建,概念关联等三个主要模块.
贡献主要有1.抽象业务规则 and 核心场景,构建taxonomy;2.用多种类型的边来表达业务规则;3.基于规则和概念相似度的概念关联推理;4.整体代码框架.
目前无监督内容理解服务和概念兴趣扩展已落地于信息流召回,CTR相对提升8%和10%.0到1落地基于产品推荐物料场景;系统技术方案参考Probase,AliCoco.
另:知识星图(围绕产品的知识探索)和知识服务重构等工作;

公司项目(平安寿险): 智能陪练

2019.07-2020.04 **[3-5人, owner | 语音纠错,语义相似,对话管理等]**
通过模拟客户来培训代理人的话术掌握能力,为部门流量最大的服务.
分阶段贡献有:1.接手并快速提升旧版本(背诵式)效果,关键指标(关键词命中率)提升13%+,距离天花板还剩3%(详细分析定位+内容质检+匹配算法优化+效果监控);2.积极从更好的剧本(写)和更像真实场景(练)的两个方面开展技术预研,并推动服务从背诵式升级到场景模拟演练(实况DM;语义相似模型+语料增强;剧本场景化定制生成).并完成新版本落地应用,用户满意度达百分之97,成功为部门翻盘.
带领团队在项目中综合应用了如下技术方案: 语音纠错(LCS+发音距离), 同义词挖掘(同义词表融合+bert召回/过滤), 文本挖掘(坐席通话记录挖掘, 通顺度过滤+相似度召回), 语义相似(esimcnn,albert-tiny孪生网络等), 文本增强(EDA+句法依存), 辅助写作(bert+crf 关键词预测, 文本质检, 剧本生成(规则)).
参考 https://www.sohu.com/a/395105197_114877
以及 <https://github.com/CheungZeeCn/littleRoundTable>

公司项目(平安寿险): 知识挖掘

- 2018.10-2019.07 **[4-5人, owner | 文本通顺度, 文本分聚类, 语义相似]**
从0到1设计和构建问答知识库挖掘模块,挖掘和分析海量FAQ引擎未应答问题.
pipeline如下:1.过滤出其中业务相关问题, 2.将其合理分类和聚类, 以及属性预测, 3.最后将问题簇分发到对应知识属主完成回答入库, 提升标注效率;
涉及技术点有基于bert的通顺度模型, 孪生网络, Textcnn分类, 图聚类, 语料清洗优化等;
其它挖掘工作:险条款信息提取,结合疾病,工种等实体信息; 通过长文本相似度,层次聚类等技术手段开展事件挖掘和追踪预研; 并代表部门在paper weekly 上做分享.
参考 <https://www.bilibili.com/video/av51991812/>
以及 <https://zhuanlan.zhihu.com/p/72945036>

公司项目(招银前海): 大数据系统构建和招招理财资讯模块等

- 2016.11-2018.03 **[1-3人, owner |CDH,报表,爬虫,文本数据处理分析, 资讯推荐,spark]**
大数据套件和系统选型部署, 基于tableau+impala的报表系统开发;
资讯模块:爬虫和基础资讯系统+个性化推荐和主题热点等模块;
个性化推荐模块主要基于业务规则,主题热点模块基于louvain聚类.

业余项目: i啄米, 基于语音识别技术的点读机

- 2016.02-2016.05 **[爬虫开发, 语音识别, 网站开发]**
个人项目,一个外语阅读学习工具,针对'复读机永远倒带不到刚才听不清的那个词上'的体验痛点开发. 点击页面中的文字,录音会精确定位到文字对应的发音处,精确到单词. 实现: 爬虫抓取html和对应录音, 再由语音识别模块(Htk)对语料文字和音频时间轴进行对齐,从而产生点读依赖的index文件, 最后将所得的数据载入数据库和页面, 完成发布;
目前已下线,核心体验见(单击点读,双击或长按查词,手机打开体验更佳):<http://old.izhuomi.com/Article/view/230>

个人特点

- | | |
|---------|--|
| 带人做事能力强 | 对团队目标明确清晰,技术沟通和协作能力优秀,有包容和利他精神,有带领小队攻坚成长经验 |
| 综合能力强 | 技术扎实视野开阔,业务和产品理解抽象能力突出,可快速结合现状和最佳实践来制定,拆分方案和解决问题 |
| 兴趣爱好 | 冲浪游泳足球等体育运动,热爱探索 |