



林廷恩

185-1197-6446 | sss950123@gmail.com | 北京

微信: s950123 | github.com/tnlin

2020应届生 | 求职意向: 算法工程师



个人总结

- 计算机专业, 研究方向是自然语言处理, 曾以一作身份发表论文于国际顶级会议ACL
- 熟悉跨行业数据挖掘标准流程(CRISP-DM), 3年数据科学经验, 完成超过10个项目, 可独立带领团队完成任务

教育经历

清华大学 2017年9月 - 2020年6月
计算机科学与技术系 硕士 智能技术与系统国家重点实验室(导师: 徐华 副教授) 北京

- GPA: 3.62 / 4.0 | 荣誉奖项: 2018, 2019 教育部港澳台侨奖学金 特等奖(全校8名)
- 相关课程: 大数据分析 & 处理(A), 高级机器学习(A-), 信息检索的前沿研究(A-)
- 助教课程: 数据挖掘: 方法与应用(2018-至今), 工业数据挖掘与分析(2018-至今)

台湾交通大学 2012年9月 - 2016年6月
电机工程学系 本科 台湾

- GPA: 3.65 / 4.0 | 荣誉奖项: 杰出贡献奖(全校毕业生前1%)
- 代表学校参与日本农工大学EDGE创业竞赛(亚军), 并获得全额奖学金前往美国斯坦福大学 SRI 培训一周

论文发表

研究内容: 基于深度学习的对话新意图发现

- Ting-En Lin, Hua Xu, Hanlei Zhang. Discovering New Intents via Constrained Deep Adaptive Clustering with Cluster Refinement, AAAI 2020.
- Ting-En Lin, Hua Xu. Deep Unknown Intent Detection with Margin Loss. ACL 2019.
- Ting-En Lin, Hua Xu. A Post-processing Method for Detecting Unknown Intent of Dialogue System via Pre-trained Deep Neural Network Classifier. Knowledge-based Systems 2019. (IF=5.1)

专业技能

自然语言处理: 多轮对话意图分类 / 情感分类 / 异常检测 / 无&半监督聚类 / 主动学习

- 程式语言: Python (5 years) / SQL (3y) / C++ (2y) / C (2y) / Java (1y) / Bash (1y)
- 常用框架: Pandas (4y) / Sklearn (4y) / Matplotlib (4y) / Keras (2y) / Pytorch (1y) / Tensorflow (1y)
- 其他工具: Linux (6y) / Git (4y) / MySQL (4y) / Docker (3y) / MongoDB (1y) / Tableau (1y)

工作经历

美商行动贝果 2016年9月 - 2017年12月
数据科学 实习生&顾问 台湾

- 参与世界百强企业客户GB~TB级别的数据挖掘项目, 包括打印机客户流失、发电机故障、餐点销售预测...等
- 参与自动化机器学习(AutoML)引擎Decanter项目, 包括测试工具、SDK开发、中间件、客户端开发...等

项目经历

跨模态情感分类(世界百强企业, 实验室横向项目) 2019年3月 - 2019年6月
负责人, 预测一段中文商品评论语音的情感为积极或消极 北京

- 音频模态: MLP + 音频情感特征
 - 比起LightGBM+MFCC特征, 提升了12个百分点的准确率(54.4% -> 67%)
 - 透过观察特征重要性、混淆矩阵等来进行错误分析
 - 使用Librosa将长音频切割为句子级音频(2~20秒) - 根据静默分贝阈值、最小静默时长、最短语句时长...等
 - 使用Opensmile抽取超过1500+个与音频情感相关的统计量特征, 包括响度、MFCC、有声概率...等 & 标准化
- 中文文本: 预训练双向语言模型(BERT)
 - 比起文本CNN, 提升了23个百分点的准确率(72.6% -> 96%)
 - 使用文本CNN(搭配不同卷积核大小)、BiLSTM(搭配dropout)作为baseline
 - 使用Jieba对中文文本进行分词(扩充外部辞典, 提升分词性能)、建立词表
 - 使用预训练中文词向量(CA8, Tencent)将权重加载到嵌入层, 并在训练过程中持续fine-tune
- 模态融合: 在决策层进行音频&文本模态融合, 提升模型鲁棒性
- 产出16页完整的英文项目报告, 将部分成果开发为网站工具, 供企业内部业务人员使用(使用flask开发后端)

其他

- 项目竞赛: 2018腾讯广告算法大赛-CTR预测(排名51/1500+), 2017 PokemonGo-TSP(用模拟退火解决游戏中的旅行商问题, 在Github获得100+ Stars), 2017 Booking.com黑客马拉松(冠军, 旅程规划协作平台), 2015 台湾交通大学黑客马拉松(冠军, 可视化操作的网页开发平台), 2014 台湾交通大学电机系专题竞赛(冠军, 数位教学平台)
- 社团经历: 清华大学研究生台生组(副组长), 清华大学在线教育协会(副会长), 台湾交通大学学生会(资讯部长)