

郭明灿

政治面貌：中共党员

电话：15917582764

邮箱：gmc9812@163.com

籍贯：广东湛江

出生日期：1998.12.11

身高：183cm

体重：66kg

个人网址：https://gmc.asia



求职意向：

教育背景

2022.09 - 2025.06 佛山大学 软件工程 085405(硕士研究生) 平均成绩 88.6 (专业排名 3/29)

主修课程：高级综合英语 自然辩证法概论 组合数学 算法设计与分析 软件体系结构 软件过程管理 机器学习 模式识别 自然语言处理

2017.09 - 2021.06 五邑大学 通信工程 080703(大学本科) GPA 3.55 (专业排名 21/201)

主修课程：大学英语 高等数学 计算机科学概论 C 语言程序设计 计算机组成原理 操作系统 web 开发技术 计算机网络原理 Linux 操作系统 微机系统与接口技术

主要成果 (详见第二页)

科研能力突出：★3 次国际评测获奖 ★1 次国际会议报告 ★4 篇论文

科研项目经验：★1 项国家级课题 ★1 项省级课题

工程经历丰富：★大型国企实习经验 ★4 项横向课题 ★2 件软件著作权授权 ★2 件发明专利申请

管理能力出色：★8 次被评为优秀干部、优秀学生 ★40 人学生研发队伍管理

学习能力优秀：★本硕期间专业排名第均在前 15% ★6 次奖学金

实习经历

2023.10 - 2024.05 中国联通佛山分公司 NLP 算法开发工程师

- 主要负责佛山 12345"坐席智能辅助开发", 内容包括热线工单分类算法、以及热线热点主题聚类算法
- 使借助 paddlepaddle、paddleNLP 框架和 enier 模型, 开发训练自反馈神经网络模型, 工单分类算法实现原始分类平均准确度达到 75%→92%的提升。完成一套完整的端到端管道推理架构, 有效分类 2022、2023、2024 年工单共计 300 万条。
- 使用 bertopic 聚类算法完成工单样本主题热词提取, 并借助词云等可视化工具展示当前诉求热点。

2021.03 - 2021.06 华软科技有限公司 前端开发工程师

- 负责公司核心产品的前端开发工作, 优化网页、小程序端性能, 提升用户体验。
- 月平均修复前端 Bug 数量 50 余个, 共解决需求近 100 个, 总代码量 1.1 万余行。

成果及证书

- 2018 获校级奖学金及三好学生、优秀共青团员
- 2019 获校级奖学金及三好学生、优秀团员标兵、优秀学生干部
- 2020 获校级奖学金及三好学生
- 2022 获研究生一等学业奖学金;
- 2023 获研究生一等学业奖学金、优秀研究生干部
- 2024 获研究生二等学业奖学金、获得优秀研究生
- 英语四级;
- 中级软件设计师
- Ms office 计算机二级
- 机动车驾驶证

校园经历

- 2022.09-至今 佛山大学专业班级 班长
- 2022.09-至今 佛山大学政企大模型及智能信息处理实验室 实验室负责人
- 2017.09-2018.09 五邑大学计算机学院 自律委委员
- 2018.09-2019.09 五邑大学计算机学院 自律部部长
- 2019.06-2019.08 五邑大学暑期三下乡 负责人

自我评价

为人诚实正直、稳重谦虚, 最大的优点是对待工作严谨认真, 一丝不苟, 拥有较强的学习能力同时注重工作效率和团队合作意识。日常生活中乐于沟通协调, 待人谦逊, 能够在各项事务性工作和日程安排下有较为出色的表现。注重学习, 乐于接受新事物, 有一定的创新创造能力。

科研经历及成果

科研课题

1. 2024.09-2025.09, 融合语言学知识的大语言模型生成文本检测研究 (批准号: 24BYY080, 国家社科基金, 纵向课题, 学生第一负责人)

负责内容: 针对机器生成文本与人类撰写文本高度相似难以区分的问题, 提出并使用一种融合传统语言学知识和预训练语言模型的检测模型。模型首先提取词频与困惑度特征, 通过 LSTM 编码获取特征向量, 同时借助预训练语言模型 (如 BERT、RoBERTa) 提取语义表示。两部分信息拼接后送入线性分类层输出判别概率, 并通过多模型集成实现加权融合。该方法在 PAN2024 AI 文本检测数据集中性能高达 96.6%

2. 2024.05-2025.05, 大语言模型生成文本的识别、事实性验证和价值观评价研究 (批准号: GD24CZY02, 广东社科基金, 纵向课题, 学生第一负责人)

负责内容: 主要负责大语言模型生成文本的识别研究, 基于半监督对比学习框架, 融合多种 BERT 与 RoBERTa 预训练模型, 并采用交叉验证策略。在 HC3 与 Detect-GPT 等数据集上开展实验, 结果显示平均准确率超过 90%。方法通过构建正负样本对并进行语义特征对比, 可以有效捕捉机器生成文本的潜在特征, 同时利用少量标注数据提升模型泛化能力。

3. 2022.10-2023.10, 2022.新闻纠错系统 (佛山大学实验室开放创新基金项目, 纵向课题, 主要负责人)

主要内容: 采用具有词法 CSC、语法 CGEC 和敏感词检测功能。核心算法参考了最新的自然语言处理研究成果, 采用 CRASpell+Seq2Edit 等多模型融合, 通过海量中文数据训练与微调, 能够自动检测和纠正新闻文本中的错误, 并辅以敏感词模块识别过滤。与传统的人工编辑和校对相比, 该系统具有速度快、准确性高、可扩展性好等优点, 能够大幅提高新闻发布的质量, 降低人力成本。

4. 2023.10 - 2025.03, 热线系统工单数据分类模型 (佛山 12345 便民热线, 横向课题, 主要负责人)

主要内容: 基于佛山 12345 便民热线提供标注样本和待分类数据, 设计与训练基于自反馈神经网络的层级文本分类模型。使用多种 HTC 层级分类与训练模型包括 Bert、Roberta、ERNIE 等主流模型。针对包含市场主体、违章建筑举报、噪声污染、招生考试、拖欠工资、售后服务、社保统筹服务 7 大民生热点以及 14 项细分类别, 设计多层级分类模型, 并使用 paddle 框架进行训练, 平均准确率从最初 70%多大幅提升至 92%。

成果: 佛山 12345 项目已立项; 已申请提交发明专利

5. 2022.06-2022.12, 法律合同比对项目 (佛山盈科律师事务所, 横向课题, 负责人)

主要内容: 具有文档、图像、PDF 比对以及印章识别等功能。

使用包括 diff 在内的多种比对算法, 和 OCR 在内的多种图像处理技术, 在 uni-app 上进行跨平台开发, 实现文档、图像、PDF 比对以及印章识别等功能。已在微信小程序上线。

成果: 已申请软著 2 项 (文档比对系统 V1.0, 图片比对系统 V1.0)

6. 2022.03-2022.06, 业绩管理系统 (横向课题, 独立开发)

主要内容: 个人独立开发的统计理疗店铺业绩管理系统, 应用 SpringBoot、Vue、Uni-app、SQL、elementUI、Redis 等多种技术栈。功能包括排工、结算、统计汇总、单日管理、单月管理、会员管理、员工管理, 项目管理等核心功能。部署包括 web 和小程序两端。

成果: 已部署于真实企业并落地应用

发表论文

7. EI 会议论文, Guo M, Han Z, Chen H, et al. A Contrastive Learning of Sample Pairs for Authorship Verification[C]//Aliannejadi M, Faggioli G, Ferro N, et al. Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023), Thessaloniki, Greece, September 18th to 21st, 2023: Vol. 3497. CEUR-WS.org, 2023: 2608-2612.
8. EI 会议论文, EI 会议论文 Guo M, Han Z, Chen H, et al. A Machine-Generated Text Detection Model Based on Text Multi-Feature Fusion[C]//Faggioli G, Ferro N, Galuscáková P, et al. Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024: Vol. 3740. CEUR-WS.org, 2024: 2593-2602.
9. EI 会议论文, Guo M, Han Z, Kong L, et al. Advantages of XLM-R Model for Urdu Sentiment Multi-Classification[C]//Ghosh K, Mandl T, Majumder P, et al. Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation, Kolkata, India, December 9-13, 2022: Vol. 3395. CEUR-WS.org, 2022: 291-297.
10. EI 会议论文, Guo M, Han Z, Wang X, et al. Multidimensional Text Feature Analysis: Unveiling the Veil of Automatically Generated Text[C]//Jiménez-Zafra S M, Chiruzzo L, Rangel F, et al. Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), Valladolid, Spain, September 24, 2024: Vol. 3756. CEUR-WS.org, 2024.

比赛获奖

11. 国际评测比赛 EmoThreat@FIRE 2022 国际评测情感分类赛道, 获得第一名(1/8)
12. PAN@CLEF 2023 国际评测作者识别赛道, 获得第二名(2/20)
13. IberAuTextification@IberLEF 2024 国际评测 AI 生成文本识别和归因, 子赛道 1 获得第二名(2/54), 子赛道 2 获得第一名(1/14)
14. 2022.12.09-2022.12.13, 国际会议口头报告, Forum for Information Retrieval Evaluation, Indian Statistical Institute, Kolkata, India

软著成果

15. 2023.05.24, 文档比对系统 V1.0, 开发人 (登记号: 2023SR0563463)
16. 2023.03.13, 基于预训练语言模型的文本对齐研究系统 V1.0, 著作权人 (登记号: 2023SR0323004)