




李东方

哈尔滨工业大学（深圳），深圳，广东，中国

 crazyofapple@gmail.com  (86) 15762254498 (微信)  <https://crazyofapple.github.io/>

个人简介

我目前是哈尔滨工业大学（深圳）的四年级博士生，指导老师是陈清财教授和户保田副教授，研究方向是可解释自然语言处理模型与方法，博士入学已发表 11 篇论文，一作/共同一作 10 篇，其中 CCF-A, JCR-1 区论文 4 篇，领域顶会 3 篇，谷歌学术引用 229，并担任多个顶会与顶刊审稿人。本硕博均就读于 985 高校计算机系，曾在腾讯公司进行研究实习（3 次，2017 年 MIG, 2018 年 TEG, 2019 年 CDG），并被推荐为 CIPS 青年工作委员会 2019-2020 年学生委员会成员（全国 10 人），中国计算机学会-中文信息学会-百度千言技术专家。获得过腾讯犀牛鸟精英人才的资助，国家奖学金，华为奖学金等。在国际/国内重要技术评测或技术竞赛一等奖 3 项，二等奖 2 项，击败腾讯，阿里巴巴、海康威视，马萨诸塞大学等机构。申请发明专利 3 项，其中已授权第一发明人 1 项。在检索式对话系统的研究成果应用于日均回答百万量级的腾讯政务智能客服系统。博士带领团队构建的领域智能化产品服务多家大小公司，部署于银行，电网，检察院等机构。

教育背景

哈尔滨工业大学（深圳）

2019.03–如今

计算机应用技术，智能计算研究中心

研究兴趣：医疗与金融自然语言处理，模型鲁棒性和可解释性，信息抽取

导师：陈清财教授和户保田副教授

论文题目：自然语言处理可解释方法研究

哈尔滨工业大学（深圳）

2016.09–2019.01

计算机科学与技术，智能计算研究中心

GPA: 3.78/4.0 (前 4%)

研究主题：文本分类，检索式对话系统，迁移学习，文本匹配

导师：陈清财教授

论文题目：结合多种策略的文本语义匹配方法（校优秀毕业论文）

中国海洋大学

2012.09–2016.06

计算机科学与技术

CGPA: 前 10%

导师：仲国强教授

论文题目：面向社交媒体的文本关键词提取（校优秀毕业论文）

经历项目

哈工大-腾讯联合实验室 & 哈工大-图灵联合人工智能实验室

深圳

项目 Leader (已有 4 届硕博生，现如今组员 10+人)

2019.03–如今

– 智能客服，要素抽取，合同审核，OCR

– 文档信息抽取产品，舆情监控发掘产品等问题：金融领域模型标注数据少，标注难度大（多家上市公司）。解决：技术负责人和团队完成金融信息抽取、企业舆情分析系统的研发，有显著效益，服务 50+家券商基金银行等知名企业

(临床) 医疗可解释知识型智能诊疗 (百度、北京协和)	深圳
项目成员	2019.11 - 2022.05
核心骨干人员设计模型通过药综考试，构建医学知识图谱 CPubMed-KG，作为主要完成人申报教育部高等学校科研奖。基于知识图谱 (CMeKG CPubMed-KG1.0) 融合的问答法在药师药学综合考试中的准确率达到 61.7%，超过了基于 IR 的 30% 的基线	
腾讯 CDG 事业群-转化优化组	北京
应用研究实习生	2019.05-2020.08
实习期间，根据广告转化率数据，对不同目标转化率的优化方法进行比较，并进行广告转换延迟解决方案的研究，产出期刊论文 1 篇	
腾讯 TEG 事业群-数据平台中心	深圳
应用研究实习生	2018.06-2018.12
问题：政务客服系统学习样本少，系统要求高。核心骨干人员历时一年构建完成腾讯小知智能客服，日均回答问题百万级别，仍线上运行在深圳公安和王卡助手的微信公众号，有腾讯开具的应用证明	
腾讯 MIG 事业群-立知产品	深圳
实习生	2016.07 - 2016.09
- 参与立知产品内容建设、游戏网站、微博、知乎爬虫。	
微众银行、中信银行、招商银行 QA 项目研究	2017.06-2019.03
- 理解用户意图，并返回最相似的答案。	
- 基于深度学习的债券风险预测系统，包括事件抽取、事件预测和事件图谱	
- 低资源问答系统（意图识别，多轮 QA）迁移学习。探索监督和非监督方法的组合	

学术论文

1. **Li D**, Hu B, Chen Q, et al. Unifying Model Explainability and Robustness for Joint Text Classification and Rationale Extraction. In AAAI 2022. (口头报告, 录用率: 4.5%)
2. Chen J*, **Li D***, Chen Q, et al. Diaformer: Automatic Diagnosis via Symptoms Sequence Generation. In AAAI 2022. (15% 录用率)
3. **Li D**, Hu B, Chen Q. Calibration Meets Explanation: A Simple and Effective Approach for Model Confidence Estimates. In EMNLP 2022. (12% 录用率)
4. **Li D**, Hu B, Chen Q, et al. Towards Explainable Medical Machine Reading Comprehension with Rationale Generation.
5. **Li D**, Hu B, Chen Q, et al. Towards Faithful Explanations for Text Classification with Robustness Improvement and Explanation Guided Training.
6. **Li D**, et al. Prompt-based Text Entailment for Low-Resource Named Entity Recognition. In COLING 2022.
7. **Li D**, Hu B, Chen Q, et al. Towards Medical Machine Reading Comprehension with Structural Knowledge and Plain Text, In EMNLP 2020.
8. **Li D**, Hu B, Chen Q, et al. Attentive capsule network for click-through rate and conversion rate prediction in online advertising. Knowledge-Based Systems, 2021, 211: 106522. (JCR Q1, 8.038)
10. **Li D**, Yu Y, Chen Q, et al. BERTSel: Answer Selection with Pre-trained Models. arXiv preprint arXiv:1905.07588, 2019.
11. **Li D**, Chen Q, Chen S, et al. Multi-strategies Method for Cold-Start Stage Question Matching of rQA Task, CCF International Conference on Natural Language Processing and Chinese Computing. Springer, Cham, 2019: 29-41.

12. Fang Y*, **Li D***, et al. Toward better prediction of recurrence for Cushing' s Disease: a factorization-machine based neural approach, International Journal of Machine Learning and Cybernetics, 2020. (JCR Q2, 4.012. 北京协和医学院合作)

13. **Li D**, Xiong Y, et al. Drug knowledge discovery via multi-task learning and pre-trained models. BMC Medical Informatics and Decision Making, 2021 (JCR Q3, 2.796)

14. Zhang W*, **Li, D***, et al. Electronic Medical Records as Input in Predicting Postoperative Immediate Remission of Cushing's Disease: Application of Word Embedding, 2021, Frontiers in Oncology (JCR Q2, 6.244)

技能

课程: 数据结构与算法, 计算机网络, 计算机组织, 操作系统, 数据库

Github: <https://github.com/crazyofapple/>

编程语言: Python, Java, C/ c++, SQL, HTML

深度学习框架: PyTorch, TensorFlow

主要工具: Linux, vim, git, L^AT_EX, VS Code

语言: 英语 (流利, 六级 478, 四级 582)

奖项

第一名, CCKS 2022: 金融 NL2SQL 评测	2022.08
第二名, 2022 语言与智能技术竞赛, 情感可解释	2022.08
AAAI 2022 Conference Student Grant	2022.01
第二名, CCL 2021: 智能诊疗对话评测任务	2021.12
国际评测第一名, EMNLP 2019 BioNLP-OST AGAC track	2019.08
第一名, CCKS 2019 评测金融实体抽取任务	2019.06
哈尔滨工业大学校级优秀毕业论文	2019.03
腾讯犀牛鸟精英人才计划	2018.08
国家奖学金, 一等奖学金	2018.05
二等奖, 微软开放学术精准画像大赛	2017.10
中国海洋大学校级优秀毕业生	2016.06
国家励志奖学金 (3 次)	2012.09-2016.06
ACM/ICPC 国际大学生程序设计竞赛亚洲区赛牡丹江站银牌	2014.10
山东省大学生程序设计竞赛银牌	2014.05

学术服务

期刊审稿人: IEEE Transactions on Neural Networks and Learning Systems, Applied Intelligence, TASLP

会议审稿人: ACL 2021-2023, EMNLP 2020-2022, EACL 2023, AACL-IJCNLP 2022, AAAI 2023, IJCAI 2023, CCKS 2021, 2022, NLPCC 2021, 2022, COLING 2020, 2022, CCL 2021, NAACL 2022, ACL Rolling Review