

# 李建 简历

☎ (+86) 13541051104

✉ lijian-jack@qq.com

📄 [jack57lee.github.io](https://jack57lee.github.io)

## 教育背景

- 香港中文大学** (The Chinese University of Hong Kong) 中国香港, 2015.8–2020.6  
博士学位, 计算机科学与工程, 导师: 吕荣聪 (Michael R. Lyu) 教授  
研究方向: 自然语言处理与信息检索
- 斯坦福大学** (Stanford University) 美国旧金山, 2019.7–2020.3  
访学博士, 计算机科学, 导师: Monica S. Lam 教授
- 电子科技大学** (UESTC) 中国成都, 2011.9–2015.6  
学士学位, 电子信息工程, 排名 2/360, 两获国家奖学金

## 工作经历

- 腾讯广告-广告数据部-高级研究员** 2022.3–至今  
工作主题: 人群标签, 用户画像, 推荐模型  
工作内容:  
1. 人群标签挖掘: 意向标签, 即通过种子人群进行意向建模, 用以人群扩量 (Lookalike); 事实标签, 即通过用户的广告域行为、在微信的浏览和搜索行为等, 为用户打上具有意义的标签。挖掘出的人群一方面可直接用于广告系统人群提价, 另一方面可建设成特征应用于广告系统的召回、粗排和精排等阶段。  
2. 精排助攻模型: 即个性化出价, 根据客户提供的一方后链路数据, 训练并部署推荐模型产出一个系数, 实时个性化地为目标广告与用户提价或降价, 用以提高广告投放的后端效果。
- 华为研究院-诺亚方舟 AI 实验室-高级研究员** 2020.6–2022.2  
工作主题: 基于预训练语言模型的文本匹配  
工作内容:  
1. 研究基于内容匹配的信息流推荐技术, 即根据用户的新闻浏览历史, 匹配内容最相关的新闻。创新性提出多兴趣用户建模框架 MINER, 从多角度建模用户的历史行为和阅读兴趣。该方法在微软新闻推荐榜单 MIND 上**排名第一** ([Link](#), 2021 年 9 月), 论文发表在顶会 ACL 2022 (论文 [1], 第一作者)。  
该算法已应用至华为浏览器新闻信息流服务, 通过与精排模型 DCN 融合, 线上用户点击率 (CTR) 相对提升 2.5%  
2. 研究基于图学习的新闻推荐方法, 分别把用户阅读历史和候选新闻建模成两张图, 提出算法 DIGAT 交互地更新用户图和新闻图, 论文已发表在顶会 EMNLP 2022 (论文 [2], 指导实习生完成)。  
3. 研究文本生成与开放域问答技术, 提出文本生成算法 DPrior 发表在顶会 ACL 2022 (论文 [3]); 参加 NeurIPS 2020 开放域问答比赛, 改进基于 BERT 的文档检索模块, 在榜单上机构排名第五。
- 腾讯 AI Lab-NLP 中心-研究实习生** 2018.3–2019.4  
研究课题: 基于深度学习的机器翻译

## 论文发表

谷歌学术 [链接: Link](#) 引用: 1100+

- [1] Jian Li, Jieming Zhu, Guohao Cai, Lifeng Shang, Zhenhua Dong, Xin Jiang, Qun Liu. [MINER: Multi-Interest Matching over BERT for News Recommendation](#). In ACL findings 2022.
- [2] Zhiming Mao, Jian Li (Corresponding Author), Xingshan Zeng, Lifeng Shang, Xin Jiang, Kam-Fai Wong, Qun Liu. [DIGAT: Modeling News Recommendation with Dual Graph Interaction](#). In EMNLP findings 2022.
- [3] Xianghong Fang, Jian Li (Corresponding Author), Lifeng Shang, Xin Jiang, Qun Liu, Dit-Yan Yeung. [Controlled Text Generation Using Dictionary Prior in Variational Autoencoders](#). In ACL findings 2022.
- [4] Qiwei Bi, Jian Li (Corresponding Author), Lifeng Shang, Xin Jiang, Qun Liu, Hanfang Yang. [MTRec: Multi-Task Learning over BERT for News Recommendation](#). In ACL findings 2022.
- [5] Jian Li, Xing Wang, Zhaopeng Tu, Michael R. Lyu. [On the Diversity of Multi-Head Attention](#). In Journal of Neurocomputing 2021 (SCI Journal).
- [6] Silei Xu, Giovanni Campagna, Jian Li, Monica S. Lam. [Schema2QA: Answering Complex Queries on the Structured Web with a Neural Model](#). In 2020 International Conference on Information and Knowledge Management (CIKM 2020).

- [7] Jian Li, Xing Wang, Baosong Yang, Michael R. Lyu, Zhaopeng Tu. [Neuron Interaction Based Representation Composition for Neural Machine Translation](#). In 2020 AAAI Conference on Artificial Intelligence (AAAI 2020).
- [8] Jian Li, Baosong Yang, Zi-Yi Dou, Xing Wang, Michael R. Lyu, Zhaopeng Tu. [Information Aggregation for Multi-Head Attention with Routing-by-Agreement](#). In 2019 Conference of North American Chapter of the Association for Computational Linguistics (NAACL 2019).
- [9] Baosong Yang, Jian Li, Derek Wong, Lidia S. Chao, Xing Wang, Zhaopeng Tu. [Context-Aware Self-Attention Networks](#). In 2019 AAAI Conference on Artificial Intelligence (AAAI 2019).
- [10] Jian Li, Zhaopeng Tu, Baosong Yang, Michael R. Lyu, Tong Zhang. [Multi-Head Attention with Disagreement Regularization](#). In 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018).
- [11] Jian Li, Yue Wang, Irwin King, Michael R. Lyu. [Code Completion with Neural Attention and Pointer Networks](#). In 2018 International Joint Conference on Artificial Intelligence (IJCAI 2018).
- [12] Jian Li, Pinjia He, Jieming Zhu, Michael R. Lyu. [Software Defect Prediction via Convolutional Neural Network](#). In 2017 IEEE International Conference on Software Quality, Reliability and Security (QRS 2017).
- [13] Pinjia He, Jieming Zhu, Shilin He, Jian Li, Michael R. Lyu. [An Evaluation Study on Log Parsing and Its Use in Log Mining](#). In 2016 IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2016).
- [14] Pinjia He, Jieming Zhu, Shilin He, Jian Li, Michael R. Lyu. [Towards Automated Log Parsing for Large-Scale Log Data Analysis](#). In IEEE Transactions on Dependable and Secure Computing (TDSC 2017).

## 获奖与荣誉

---

CUHK Global Scholarship Programme for Research Excellence	2019–2020
CUHK Overseas Research Attachment Programme Scholarship	2019–2020
CUHK Full Postgraduate Studentship	2015–2019
国家奖学金 (Top 2%)	中国教育部, 2012 & 2013
唐立新奖学金 (Top 1%)	电子科大, 2014
成电杰出学生	电子科大, 2015
四川省优秀毕业生	四川省, 2015
美国大学生数学建模竞赛 (ICM) 一等奖	COMAP, 2014

## 技能

- 
- 编程语言: Python, Scala, C/C++.
  - 工具: PyTorch, Tensorflow, Spark, Linux, Git, Vim.
  - 托福英语: 102 (R29 L24 S22 W27)    GRE: 321 (V151 Q170) + 3.5