李建 简历

(+86) 13541051104 ✉ lijianjack@gmail.com 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

华为诺亚方舟实验室–语音语义实验室–高级研究员 *2020.6–至今*

工作课题: 基于预训练语言模型的搜索与推荐

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，使用BERT编码新闻，并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)，论文已投稿至ACL 2022 (论文[1]，第一作者)。

目前该算法已**落地**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)平均提升2%

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，并提出算法DIGAT交互地更新用户图和新闻图，在MIND数据集上取得显著提升，论文已投稿至ACL 2022 (论文[2]，指导实习生完成)。

3. 研究基于隐变量建模的文本生成技术, 提出算法DPrior将传统VAE中的连续隐变量替换成一系列可学习的离散点, 并与预训练语言模型结合, 在VAE benchmark上取得显著提升, 论文已投稿至ACL 2022 (论文[3], 指导实习生完成)。

4. 参加NeurIPS 2020 开放域问答(Open-Domain QA)比赛，改进文档检索模块，搭建基于BERT的双塔模型，并使用模型蒸馏技术，从BERT单塔模型蒸馏排序能力到双塔，最终在榜单上机构排名第五(Link)。

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: 800+ (2021.12)

[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 submission to *ACL 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 submission to *ACL 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 submission to *ACL 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 submission to *ACL 2022*.

[5] **Jian Li**, Xing Wang, Zhaopeng Tu, Michael R. Lyu. On the Diversity of Multi-Head Attention. In *Journal of Neurocomputing 2021 (SCI)*.

[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] **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)*.

[10] **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)*.

[11] **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)*.

[12] 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)*.

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