**林恒旭**

138-5776-9582 | [linhx21@m.fudan.edu.cn](mailto:linhx21@m.fudan.edu.cn)

GitHub: [KomorebiLHX](https://github.com/KomorebiLHX)

# 教育经历

## 复旦大学 大数据学院 | 应用统计专业 2021年9月 - 2023年6月（预计）

* **GPA：**3.6 / 4.0
* **实验室：**[复旦大学数据智能与社会计算实验室](http://www.fudan-disc.com/)（导师：[魏忠钰](http://www.sdspeople.fudan.edu.cn/zywei/)）
* **研究方向：**自然语言处理

## 浙江大学 管理学院 | 信息管理与信息系统专业 2017年8月 - 2021年6月

* **GPA：**4.44 / 5.0（名次：2/25）
* **海外交流：**2019年7-8月，参加新加坡国立大学计算机学院暑期研讨会，学习机器学习的相关知识
* **主要荣誉：**浙江大学优秀毕业生、浙江大学一等奖学金、华为财务精英挑战赛一等奖、美国大学生数学建模比赛H奖

# 科研经历

## EASED: An Emotion-Aware Stratified Encoder-Decoder for Dialogue Strategy Prediction 2022年3月 – 至今

* COLING 2022一作在投；该论文提出了一种情绪感知的、采用分层编码器-解码器架构的情感抚慰对话策略预测模型EASED
* 该模型采用多任务训练框架，对求助者的情绪状态进行显式建模，并考虑到对话的交互信息，对对话双方进行交互式的分层建模
* EASED在ESConv数据集上的F1得分达到36.71，相较最先进的情感抚慰对话策略预测模型提升约21.9%

## 复旦-中电金信智能金融科技联合研究中心项目：中电金信智能外呼 2021年9月 – 2022年6月

* 采用自然语言处理技术实现的自动对话系统，是助力传统客服中心数字化转型的解决方案，包含规则引擎、检索引擎及闲聊引擎
* 负责规则引擎的开发，包含采用Poly-Encoder对用户进行意图识别、语料库数据增强、构建规则逻辑树、Flask框架构建算法API
* 负责银行场景命名实体识别的算法实现，在小样本场景下采用BERT+CRF模型识别8种银行场景命名实体，离线测试F1得分为97.45

## 推荐系统前沿方法探索与实践 2021年11月 - 2022年1月

* 担任项目队长，负责在MovieLens数据集上构建四种神经协同过滤模型（MF, GMF, MLP, NeuMF）以及Bert4Rec模型
* 基于协同过滤思想计算用户的关系距离，并采用d3.js构建Force-Directed Graph可视化用户关系

## [学士学位论文：《对话文本中的用户情感识别与回应生成算法研究》](https://github.com/KomorebiLHX/Emotion-Recognition-in-Conversations) 2020年9月 - 2021年6月

* 使用带关系位置编码的图注意网络（RGAT）等四种模型进行对话中的情感识别，并对IEMOCAP数据集进行了统计分析
* 在IEMOCAP数据集上对比了四种模型的性能，并对RGAT中的关系位置编码进行重新设计，提出了新的位置编码
* 实验证明，新提出的多维位置编码可以为RGAT提供更丰富的序列信息，其性能优于其他三种位置编码

## Kaggle：Airbnb New User Bookings 2019年9月 - 2020年1月

* 利用Airbnb用户数据，预测新用户首次预订国家，采用NDCG作为推荐任务的衡量指标，最终结果排名前20%
* 担任项目队长，负责特征工程、XGBoost模型的训练与调参，在特征工程阶段采用特征选择对高维数据进行降维

# 实习经历

## 字节跳动科技有限公司（Data-推荐） 2022年5月 – 至今

推荐算法工程师

* 参与今日头条的推送业务，通过前沿的召回、精排、频控算法优化推送消息的点击率，进而直接增长头条相关app的DAU
* 扩增推送候选，增加微头条体裁分发功能+省份推送分发功能；开展反转实验证明扩增候选有效增长CTR+0.835%，DAU+0.2%
* 采用batch softmax、增加真实负例等方法在召回阶段对高热文章消偏，在有效降低高热文章分发量的基础上CTR+0.3%

## 中电海康集团有限公司（研究院） 2020年11月 – 2021年4月

算法工程师-自然语言处理

* 参与浙江省重点研发计划项目“基于人工智能的数字诊疗设备及系统研发”，研发有关眼科疾病的医疗对话系统
* 使用Python Scrapy爬取眼科疾病实体信息和关系信息，并使用Neo4j数据库构建医疗知识图谱，在此基础上搭建对话系统Demo
* 采用BERT+CRF构建命名实体识别模型，在MSRA语料库达到94.18的F1得分

# 技能与兴趣

* **技能：**Python, R, SQL, C++, d3.js
* **兴趣特长：**英语（托福100分）, 弹吉他