

王宇河

19921295690 | yuhewang1998@163.com | 上海
算法实习生



教育经历

华东师范大学	2020年09月 - 2023年03月
电子信息 硕士 数据科学与工程学院	上海
华东师范大学	2016年09月 - 2020年06月
数据科学与大数据技术 本科 数据科学与工程学院	上海
● 获得2017年国家数学建模大赛，全国二等奖	

项目经历

欧冶云商股份有限公司	2021年10月 - 2022年01月
算法实习生 数据部门	上海
● 项目描述：欧冶作为宝钢旗下在线钢铁交易平台，有庞大的用户量和交易量；该项目需要对用户业务属性做出预测，有利于后期消息的推送和商品的精准推荐；	
● 合作进行多平台数据的整合和处理，在Hive上进行数据初步的提取和整合；	
● 负责用户行为数据的预处理，并使用numpy和pandas进行数据的深度挖掘和充分的特征提取；	
● 负责模型算法（GBDT、DCN）的研究与落地等工作；与技术部门协同合作，推动身份预测算法上线，模型准确性达到71.3%；	
● 在该项目中，为了解决多分类数据不平衡问题，先使用欠采样方法避免极端情况，后又还原了预测概率方便多分类比较；	
● 通过分析LR的权重参数，得出模型的分析报告，并进而导出职业属性的用户画像；	
● 后续准备使用图学习方法，构建图结构，来补全缺失信息！	
上海收钱吧科技有限公司	2019年06月 - 2019年08月
算法实习生 算法小组	上海
● 协助数据小组，对商户行为数据、商户交易数据进行充分的数据挖掘和特征提取；	
● 完成商户流失预测模型上线，和初版模型对比，月平均准确率提高20%；	

论文发表

Meta-Learning Siamese Network for few-shot Text Classification	
● 该论文EMNLP在投，解决的是元学习小样本文本分类问题	
● 搭建了一个复杂的元学习、孪生网络模型，其中使用到了fast-text模型、MAML算法、孪生网络、度量学习，解决了PROTO模型中选取随机性问题；	
● 引入外部信息来却确定稳定的原型向量，再通过孪生网络和度量学习来拉近同类文本的低维映射，同时拉远不同类别的映射；	
● 利用MAML模型的特性，更快更好地找出最优初始点，让模型经过微调之后，可以对没见过数据表现出良好的性能，最终效果准确率比之前的方法高出了6.1%；	

专业技能

- 熟练掌握python, C, C++, 能在linux下进行熟练开发；
- 熟悉常见的机器学习和深度学习算法和原理；
- 能熟练使用深度学习框架pytorch、tensorflow；
- 能熟练使用hadoop、spark、hive等分布式开发工具；
- 能熟练使用git、svn等代码协同工具，能熟练掌握docker；
- 有良好的论文阅读能力，研究生期间阅读论文百余篇，涉及联邦学习、推荐系统、元学习、对比学习、图算法；
- 对推荐系统、联邦学习、元学习都有一定的了解和掌握；
- 具有良好的英文阅读能力，通过四六级考试；

个人总结

- 本人待人友善、做事认真负责、勤奋好学，踏实肯干，具备很强的分析问题、解决问题的能力；
- 喜欢持续学习、更新知识和技能，喜欢富有挑战性的竞争环境，敢于面对挑战；
- 具有良好的环境适应能力和团队合作意识，可以很好的与同事进行团队合作；能够虚心接受他人意见和建议，不断成长；