

徐 璦

机器学习 · 数据挖掘

☎ 13812688705 | ✉ leoxc1571@163.com | 🌐 leoxc1571.github.io | 📧 leoxc1571



教育经历

南京信息工程大学-应用统计学（本科）

江苏南京

2017年9月-2021年6月

浙江工商大学-理学统计学（硕士）

浙江杭州

以第一作者发表JCR-Q2论文一篇（深度强化学习）和JCR-Q1(CCF-B)论文一篇（图神经网络推荐系统）

2021年9月-2024年1月

技能

技能 Python, PyTorch, PyG, DGL, Tensorflow, Git, Linux, SQL, \LaTeX
外语 CET4 (597), CET6 (597)

实习经历

之江实验室-图计算研究中心

图算法工程师

智能计算计算医药（大规模图预训练，分子性质预测，Diffusion-based分子生成）

2022年7月-至今

- 负责OGB-LSC NeurIPS22竞赛，构建模型实现超大规模分子图性质学习，取得全球第11名，参与OGB-DDI榜单打榜，取得榜单第一名；
- 复现、部署大规模图预训练模型及分子构象生成模型；
- 负责Diffusion-based生成模型研究，撰写Diffusion-based图生成领域综述；
- 创造性提出基于全局注意力的E(n)等变图网络，结合原子自身性质与分子间相互作用，实现时下最优的3D分子学习性能。同时将此图网络与Diffusion模型结合，实现时下最优的3D分子生成性能，相关文章正在撰写中。

项目经历

基于图对比学习的公平的标签感知推荐系统

JCR-Q1(TOP)&CCF-B

第一作者

2022年5月-2022年12月

- 基于GNN和对比学习范式，构建TAGCL模型进一步提升基于标签感知的个性化推荐性能，同时显著提升推荐结果的公平性；
- 分别构建<用户-标签>和<标签-商品>的子图，通过添加经过softmax归一化的随机扰动进行表征增强，用于建立对比学习任务；
- 分别基于交叉熵损失，基于Tag负采样的二部图损失和改进知识图谱算法TransT，构建目标函数，分别用于推动生成更公平的推荐和提升二部图之间的一致性；
- 实验显示模型推荐精度较标签感知推荐领域的SOTA模型至少提升5%，同时在歧视性较强的数据集上显著提升结果的公平性。

OGB-LSC NeurIPS 2022

竞赛

队长

2022年8月-2022年11月

- 对超大规模分子数据(超200W)进行表针学习，预测目标分子能量隙；
- 构建HFAGNN模型，运用Hybrid模块，学习原子自身化学性质，同时运用Bessel方程提取2元组与3元组间的原子3D信息，将分子2D拓扑信息与3D结构信息结合，在满足三维空间等变性的同时实现对分子性质的高效学习；
- 使用多卡并行实现高效运算，在较小规模参数数量的情况下取得榜单第11名。

基于扩散模型的3D分子生成

论文(撰写中)

负责人

2022年11月-至今

- 基于扩散生成模型，构建去噪扩散概率模型，生成创新且有效的3D分子结构；
- 将扩散模型自身特点与原子间相互作用力本质相结合，构建出既能学习不同原子化学特性，又能学习任意原子间的相互作用强弱的全局注意力机制，构建创新的等变图网络作为去噪过程的内核；
- 改进原有采样策略，引导模型在生成大分子过程中做出更多探索，提升生成分子的创新性。

电商用户商品价值评估与基于图神经网络的个性化推荐系统

全国统计专业研究生案例大赛

组长

2022年2月-2022年3月

- 基于真实电商数据，实现商品从盈利能力、畅销水平和退货率等方面的价值评估，同时基于价值对用户群体进行有效划分，最后对用户进行个性化推荐。
- 通过PCA降维与K-Means++聚类，实现对商品从盈利能力，畅销水平和退货率等多个维度的有效划分。
- 运用改进RFM模型，对用户特征做特征交叉，实现对用户从消费能力，下单频率和退货水平等多维度的有效划分。
- 基于LightGCN模型，创新性的提出了WideGCN模型，实现对基于商品和用户的16维特征这一旁信息的有效利用。推荐性能较SOTA模型有2%到5%的提升。
- 该项目获全国三等奖

荣誉奖励

2018.5 2018年全国大学生英语竞赛（NECCS）

全国一等奖

2019.5 南京信息工程大学第十三届数学建模竞赛

三等奖

2021.10 华为杯第十八届全国研究生数学建模竞赛

全国三等奖

2022.8 第五届全国统计专业研究生案例大赛

全国三等奖

2022.11 浙江工商大学研究生学业奖学金

一等奖