



马学威

电话: (+86) 188-4454-9019
邮箱: xuew_ma@163.com
博客: maxuewei2.cnblogs.com
GitHub: github.com/maxuewei2

教育背景

2017.09 – 2020.07 吉林大学 计算机应用技术 硕士 (推免)
2020 届优秀毕业生; 连续三年获得学业奖学金及优秀奖学金
2013.09 – 2017.07 吉林大学 计算机科学与技术 学士
2016 年获校优秀学生; 连续三年获得一等奖学金
GPA: 3.49/4.0 GPA Ranking: 11/355

研究方向

图计算/图数据挖掘、网络表示学习/图嵌入、复杂网络

论文

<https://github.com/maxuewei2/RiWalk>

- RiWalk: Fast Structural Node Embedding via Role Identification
ICDM - IEEE International Conference on Data Mining, 2019 (一作, CCF B 类会议, 接收率 9.08%)
 - 网络节点角色表示学习, 输入网络 (如社交网络), 仅依据网络拓扑结构推测各节点的角色, 并学习节点的向量表示, 使角色相近的节点的表示向量距离更近.
 - 算法将角色表示学习问题解耦为重标识过程和网络表示学习过程. 重标识过程改变上下文节点的标签, 使角色相似的中心节点有更多重合的上下文节点, 从而可使用网络表示学习方法学习角色表示.
 - 使用 C 和 Python 实现. 算法时空复杂度比基线算法 (struc2vec, GraphWave) 低 100 倍以上.

项目 (部分)

<https://github.com/maxuewei2>

- 知乎爬虫 2017.09
 - 为构建社交网络, 爬取知乎用户数据. 爬虫支持多线程、账户切换、代理切换, 自动爬取免费代理.
 - 未使用任何框架, 使用 Python 实现 (后又用 Java 实现了一遍).
 - 流水线结构, 队列作为传送带, 上游线程组装 URL, 中游线程负责网络请求, 下游负责解析及组装数据.
 - 每隔几秒请求一次, 两周爬取 11 万用户数据, 数据量达 60GB.
- 结合知识图谱的推荐系统 2017.03 – 2017.06
 - 使用贝叶斯个性化排序优化矩阵分解, 结合知识图谱提升精确度并避免冷启动.
 - 使用 C++ 实现, 使用知识图谱表示学习 (TransR) 等技术, 在电影和书籍推荐任务上评测.
- 遗传算法生成正则表达式对文本分类 2017.10
 - 使用遗传算法生成正则表达式集合, 使匹配较多正类文本, 而尽量少匹配负类文本.
 - 将正则表达式映射为编码, 通过编码的杂交、变异生成新的编码, 通过优胜劣汰逐代筛选即可得到更好的正则表达式.
- 其他 2015-2020
 - 学位论文 LaTeX 模板、局域网象棋游戏、迷宫、五子棋、校园导游系统、Flappy Bird 等.

技能

专业技能: (大规模) 网络表示学习/图嵌入, 图数据挖掘, 社交网络挖掘, 机器学习, 推荐系统.
具有良好的数据结构、算法、计算机组成与体系结构、计算机网络等专业基础.

编程技能: Python, C, Java; NetworkX, PyTorch; GNU/Linux, \LaTeX .

语言: 英语 (六级 511), 可流畅阅读英文文档.

学术技能: 具有较好的英文专业文献阅读能力及写作能力.