

中国科学技术大学连德富教授学术报告：面向向量检索的可学习索引

应齐鲁工业大学（山东省科学院）计算机科学与技术学部邀请，中国科学技术大学大数据学院博士生导师、国家优秀青年基金获得者连德富教授将于6月1日访问校（院），并做《面向向量检索的可学习索引》的学术报告。欢迎各位老师和同学参加。

报告主题：面向向量检索的可学习索引

报告时间：2023年6月1日周四 15:00-16:00

报告地点：机电楼 B310 会议室

报告简介：向量检索专注于从一堆数据点中找到 k 个最近邻，并在信息检索和推荐系统等多个领域得到广泛应用。当前的先进方法如 HNSW 生成的索引会占用大量内存，限制了它们能处理的数据规模，除非借助外部存储的混合索引。占用小内存的空间划分学习索引近年来取得了重要进展。然而，这些方法依赖大量标签数据进行有监督学习，模型复杂度限制了其泛化能力。为了解决这个问题，我们提出了一种基于平衡 K 叉树的轻量级可学习索引 BATL，针对数据点集进行分层的学习划分，其中相同的数据点桶由根到对应叶的路径表示。与其将每个查询点映射到一个桶中，BATL 将其分类到一系列分支上（即一条路径），这大大减少了类的数量并有可能提升泛化能力。BATL 以交替方式更新分类器和平衡树。在更新分类器时，我们创新性地利用序列到序列的学习范例，学习如何将每个查询点路由到平衡树上的实际叶节点上。向量检索就被简化为序列（即路径）的生成任务，这可以通过对编码器-解码器进行束搜索来简单地实现。在更新平衡树

时，我们在平衡约束下自顶向下地应用分类解码器将每个数据点导航到树节点上。在多个大规模向量数据集的实验表明了 BATL 在延迟、准确性和内存成本之间的权衡优于 SOTA 基线的优越性。

报告专家姓名：连德富、特任教授、国家优青、博士生导师

报告专家简介：连德富，国家优青，中国科学技术大学特任教授，博士生导师。他曾任电子科技大学副教授、悉尼科技大学访问学者，曾入选微软亚洲研究院铸星计划。他的主要研究方向包括深度学习和推荐系统，主持了国家自然科学基金优秀青年科学基金、面上项目、科技部科技创新 2030 重大项目课题、国防科技 173 重点项目等。他在 KDD、NeurIPS、TPAMI、TKDE 等高水平会议和期刊上发表论文 100 余篇。他曾获得教育部自然科学一等奖、安徽省教学成果一等奖、四川省教学成果奖二等奖、APWeb 最佳学生论文、WISE 2022 最佳论文、WWW 2021 最佳论文候选。

报告专家主页：<http://staff.ustc.edu.cn/~liandefu/>

报告专家照片：



备注：本校师生，欢迎直接参会！外校师生，请联系鹿文鹏教授
(lwp@qlu.edu.cn) 办理入校手续。