

冯小康

联系方式：13709142883

邮箱：research@xkfeng.com

最高学历：博士

出生年月：1990 年 12 月



能力概述

计算机基础全面、复杂软件工程研发经验、基础问题算法优化创新：

- 参与商业级数据库内核研发、问题分析定位熟练
- 高维向量搜索属于前沿基础问题，成果获得国内数据库会议最佳学生论文
- 注重软件工程与编码规范，交付代码通过高标准 review；注重测试，善于发现问题
- 对计算机科学、工程领域基础问题感兴趣

C/C++/Matlab/Linux 熟练；熟悉 Python

教育背景

- | | |
|-------------------------------|------------------|
| ● 2013.08-2020.06 西安电子科技大学 | 计算机系统结构专业博士 |
| ● 2018.09-2019.09 澳大利亚新南威尔士大学 | 教育部公派联合培养博士生 |
| ● 2009.09-2013.07 西安电子科技大学 | 计算机科学与技术专业学士（保研） |

工程经历

2020.9~. 华为云数据库内核研发工程师.

- 作为主力参与 Innodb 并行建索引特性研发、测试。（交付内核代码 4000+，测试代码 6000+、发现非本人引入 BUG 8 个）
- 独立完成“并行索引创建”与“算子下推”特性的结合，涉及设计、研发与测试（交付内核代码 1500+，测试代码 2700+）
- 独立调研 MySQL 社区版并行索引创建特性，通读代码，吃透算法，完成性能评估
- 分析定位多项数据库内核 BUG、给开源社区提单 1 个

学术成果

研究方向：高维特征向量索引与相似性搜索、高维数据分析

代表性成果：

1. 获得第 35 届中国数据库学术会议(NDBC 2018)萨师煊优秀论文奖(Best Student Papers)和最佳论文奖提名，并被国内计算机顶刊《计算机学报》录用。该工作研究了外存环境下不同空间曲线下最近邻候选点局部分部质量进行了全面的量化分析，最终确定出最优曲线进一步提升了海量高维最近邻搜索的 I/O 效率。
2. 最远邻搜索能够增加推荐系统推荐的新颖性，探索发现多个高维最远邻点的分布规律，综合聚类、远邻图技术构建出一种自适应的最远邻搜索策略，性能提升显著。

所有研究均手动实现，相关**核心代码 6000+**，详见：<https://github.com/xkfengxd227/libxkfeng>

详细成果：

- [1] **冯小康**，彭延国，崔江涛等. 基于最优排序的局部敏感哈希索引[J/OL]. **计算机学报**, 2020, 43(5): 930-947. (**CCF 中文核心期刊第一位**)
- [2] **Feng Xiaokang**, Cui Jiangtao, Liu Yingfan, et al. Effective optimizations of cluster-based nearest neighbor search in high-dimensional space[J/OL]. **Multimedia Systems**, 2017, 23(1): 139-153. (**CCF C 类, SCI 4 区**)
- [3] **Feng Xiaokang**, Cui Jiangtao, Li Hui, et al. An efficient LSH indexing on discriminative short codes for high-dimensional nearest neighbors[J/OL]. **Multimedia Tools and Applications**, 2019, 78(17): 24407-24429. (**CCF C 类, SCI 4 区**)
- [4] 崔江涛，**冯小康**，刘畅，侯勇超，蔡洋. 一种索引生成方法、数据检索方法和装置：中国，201611170581.9[P]. 2020-5-12. (已授权)

科研项目

主持西电研究生创新基金项目 1 项，**主持**省级科技项目立项 1 项，作为**研究骨干**参与国家自然科学基金面上项目 2 项：

- 西安电子科技大学研究生创新基金，海量非结构化数据近邻搜索关键技术研究，**主持**
- 陕西省国家电网科技项目，信息系统隐患挖掘与智能调控，**主持**，完成立项
- 国家自然科学基金面上项目，基于哈希的海量高维数据近似最近邻查询研究，负责人：崔江涛，个人负责：外存哈希索引研究与实现
- 国家自然科学基金面上项目，面向空间运动对象的位置服务技术研究，负责人：崔江涛，个人负责：相似性搜索算法研究设计