

智能系统软件的机遇和挑战

李国良

(清华大学计算机科学与技术系 北京 100084)

由于数据的增多、算力的提升、算法的优化，机器学习技术近年来得到了快速发展，也为系统软件的自动化、智能化、自治化带来了很多机会。第一，传统基于手工的优化方法可以转变为自动的优化方法，例如系统软件参数的调优和诊断，传统手工方法需要花费数周来完成系统优化，而机器学习可以通过数据驱动方法来自动调整参数，提升调参性能。第二，传统启发式方法逐渐转变为智能化的方法，例如传统的基于均匀性分布和独立性假设的估计问题导致估计不准，可以通过机器学习算法进行智能化的估计来提升估计准确率。第三，传统经验式软件设计方法可以走向自治化的设计方法，例如传统的系统软件设计都是依赖于专家的经验进行设计，由于专家的经验往往受限，难以探索未知空间，而基于机器学习的方法可以探索更加广泛的空间，可以自治化地设计系统软件。

论文“机器学习方法赋能系统软件：挑战、实践与展望”介绍了上海交通大学陈海波团队在并发索引结构、键值存储系统、并发控制协议等方面应用机器学习方法的工作。

1) 基于机器学习模型的索引设计，包括利用简单的线性模型来降低复杂模型执行开销大的问题、设计后备机制来解决模型预测误差的问题、两阶段压缩方法实现新索引构建来解决并发更新和查询的问题。

2) 基于机器学习的键值存储系统，包括基于机器学习方法替代客户端的索引结构缓存以提供高效查询性能的同时减少索引结构所需内存空间、设计了一种混合的系统架构并构建了原型系统 XStore。

3) 基于机器学习的并发控制协议，包括能够精确区分事务执行的建模方式来降低执行开销、基于 one-hot 编码构建了行动空间来覆盖现有并发控制算法、后备机制来保证事务并发正确性。

该文还分析了机器学习赋能系统软件的经验和教训，建议通过简单的学习模型以应对复杂的推理时延、降低后备机制开销以应对机器学习带来的错误、深入理解系统瓶颈再引入学习模型。

机器学习为系统软件自动化、智能化、自治化带来了很多机遇，但是在适应性、昂贵代价、在线系统部署等方面还存在众多挑战。此论文是一个很好的开端，在此文的基础上，建议持续深入研究，推动机器学习在系统软件的发展和落地。

评述专家：



李国良，清华大学计算机科学与技术系教授，IEEE Fellow. 主要研究方向为数据库和智能数据管理。

亮点论文：

唐楚哲, 王肇国, 陈海波. 机器学习方法赋能系统软件：挑战、实践与展望 [J]. 计算机研究与发展, 2023, 60(5): 964–973.

DOI: 10.7544/issn1000-1239.202330127