**0 Survey and Tutorial**

**0.1 Survey**

## 0.1.1 Database Meets Deep Learning: Challenges and Opportunities（SIGMOD Record-2016）

（1）综述。主要讲数据库领域和深度学习领域交叉的研究问题；从数据库角度讨论深度学习系统可能的改进，以及数据库应用程序如何从深度学习技术中受益；

（2）在深度学习中，涉及较多的代价昂贵的线性运算，包括矩阵乘法；为了加速运算，将数据依赖性不强的操作放在执行机器上并行计算，如CUDA流和CPU线程等；也可以调整读写顺序动态完成；数据库在优化事务执行时和查询计划时，也可以用类似这种解决方案**（我的理解是在事务执行时使用多个机器将数据相关性不强的操作并行执行、在制定查询计划时构造包含执行器和内存等的计算资源成本模型得到最优查询计划）**；

（3）数据库系统通过日志记录和检查点保证持久性，深度学习系统的训练主要基于检查点文件从崩溃中恢复训练，但是频繁的检查点会带来巨大的开销；与数据库系统在事务中强制执行严格的一致性相比，深度学习训练系统使用的SGD算法能够容忍一点过的不一致性，因此日志记录不是必须的；

（4）在数据库上理解自然语言的语义是一项挑战，可以通过应用RNN模型来解析自然语言查询生成SQL查询，并使用现有的数据库方法对其进行改进。

（5）当前大多数数据库系统使用复杂的启发式和成本模型来生成查询计划，参数SQL查询模板的每个查询计划都有一个最优性区域，只要SQL查询在这个区域内，最优查询计划就不会改变。即查询计划对于输入的参数微小变化并不敏感，因此可以训练一个查询规划器，它从一组SQL查询和最优计划中学习；

（6）总结：主要讨论数据库优化系统性能的技术，在内存优化和并行上讨论一下可能的改进；