分类号_____ 密级_____

UDC _____ 编号_____

中国科学院研究生院
博士学位论文

并行超松弛算法和GPU并行隐马尔可夫模型

何沧平

指导教师 唐贻发 (研究员, 博士)

中国科学院数学与系统科学研究院

申请学位级别 博士 学科专业名称 计算数学

论文提交日期 2011年4月 论文答辩日期 2011年5月

培养单位 中国科学院数学与系统科学研究院

学位授予单位 中国科学院研究生院

答辩委员会主席_____

摘　要

本文研究并行超松弛(SOR)算法和GPU并行隐马尔可夫模型算法.

分布式并行SOR的目标是收敛速度快和通信时间短, 本文用重复计算代替通信, 将经典的PSOR算法的通信由五次减为四次, 节省大约20%的通信时间. 为进一步减少通信时间, 设计了一个新的快速并行SOR算法(FSOR), 在二维网格上, 对五点格式和九点格式, FSOR的收敛速度与串行SOR相同, 通信次数也降到了最少, 而且几乎所有的通信都能被计算重叠.

GPU通用计算是当前热点, 本文在GPU上实现了红黑SOR算法和四色SOR算法, 在分析多种方法优缺点的基础上, 新的网格分割方法能避免线程空闲, 消除CUDA程序分支, 消除非合并访问, 减少非对齐访问; 在96核的GeForce显卡上实现了32倍的加速, 在240核的Tesla显卡上实现了127倍的加速.

GPU上实现隐马尔可夫模型时, 仔细研究了对数连加的数学性质, 进而找到适合串行计算的嵌套方法和适合GPU并行计算的折半相加方法. 设定正态分布方差的最小值来避免灾难性的向下溢出; 设计符合GPU硬件特性的数据结构、综合使用避免双精度运算、共享内存/寄存器代替全局内存、转置矩阵消除非对齐访问、外层次并行减少内核启动次数等多种手段, 一步步提高程序速度, 实现了60倍左右的加速.

关键词：并行SOR, FSOR, GPU并行SOR, GPU并行隐马模型