**《并行计算》**

**实验指导书**

**主编：王云岚**

西北工业大学

**2024 年10月**

**MPI、OpenMP并行程序设计**

**一、实验目的**

熟悉和掌握基于共享内存的多线程并行编程和基于分布内存的多进程并行编程，包括：1）熟悉Pthread多线程并行编程，掌握临界区、信号量、互斥量等线程同步方法。

1. 熟悉OpenMP多线程并行编程，掌握基于OpenMP的多线程应用程序开发，掌握OpenMP相关函数以及数据作用域机制、多线程同步机制等。
2. 熟悉MPI多进程并行编程环境，掌握MPI编程基本函数及MPI点对点通信及集合通信函数用法，掌握MPI的主从模式及对等模式编程。

**二、实验内容**

1. **斐波那契数列**

[数列](https://baike.so.com/doc/5389111-5625689.html) 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233，377，...为斐波那契数列。斐波那契数列满足：

**要求：**

1. 采用Pthread多线程并行方法实现计算的并行程序；
2. 测试分析不同n值和线程数P时的程序运行时间和并行效率；
3. 测试并比较不同的线程同步方法对性能的影响。
4. **Floyd算法的并行化**

Floyd算法又称为插点法，是一种用于寻找给定的加权图中多源点之间最短路径的算法，Floyd算法选择的数据结构为邻接矩阵。

**要求**：

1）进（线）程p-1负责读入和分发矩阵数据。

2）假设矩阵行数*n*可以被进（线）程数P整除，采用按行分块的方案。

3）所有进（线）程依次将结果数据写入数据文件。进（线）程0先写，进（线）程1再写，进（线）程p-1最后写，最终完成结果输出。

4）分析不同n值、P值以及不同矩阵稠密度时算法的运行时间，进行算法并行性能和可扩展性分析。

1. **并行k-means算法**
2. means是最常用的基于欧式距离的[聚类算法](https://www.zhihu.com/search?q=%E8%81%9A%E7%B1%BB%E7%AE%97%E6%B3%95&search_source=Entity&hybrid_search_source=Entity&hybrid_search_extra=%7b)，其认为两个目标的距离越近，相似度越大。
3. K-means有一个著名的解释：牧师—村民模型：

有四个牧师去郊区布道，一开始牧师们随意选了几个[布道点](https://www.zhihu.com/search?q=%E5%B8%83%E9%81%93%E7%82%B9&search_source=Entity&hybrid_search_source=Entity&hybrid_search_extra=%7b)，并且把这几个布道点的情况公告给了郊区所有的村民，于是每个村民到离自己家最近的布道点去听课。

听课之后，大家觉得距离太远了，于是每个牧师统计了一下自己的课上所有的村民的地址，搬到了所有地址的中心地带，并且在海报上更新了自己的布道点的位置。

牧师每一次移动不可能离所有人都更近，有的人发现A牧师移动以后自己还不如去B牧师处听课更近，于是每个村民又去了离自己最近的布道点……

就这样，牧师每个礼拜更新自己的位置，村民根据自己的情况选择布道点，最终稳定了下来。

K-means算法的步骤为：

1. 选择初始化的k个样本作为初始聚类中心；
2. 针对数据集中每个样本，计算它到k个聚类中心的距离，并将其分到距离最小的聚类中心所对应的类中；
3. 针对每个类别，重新计算它的聚类中心（即属于该类的所有样本的质心）；
4. 重复上面的2)、3)两步操作，直到达到某个终止条件（迭代次数、最小误差变化等）。

**要求：**

1）设计并实现计算k-means并行程序。

2）可采用UCI的用于聚类的数据集（https://archive.ics.uci.edu/ml/index.php），也可以自己写程序生成数据集。

3）测试并分析并行程序的时间性能。

**（4）其他可选题目**

1）2-N之间素数个数统计。

1. N-皇后问题。
2. 矩阵乘法。

4）分布式并行训练。

......

**三、实验仪器、设备**

**Linux 操作系统**

MPICH2或者OpenMPI

支持OpenMP的GCC、ICC等编译器

**四、实验原理**

Pthread（POSIX Threads Programming）是一组C语言编程类型和过程调用, 通过pthread.h 头文件和线程库实现, Pthread API分为线程管理、互斥量、条件变量、同步等类型，可支持实现多线程并行程序。

OpenMP应用编程接口API是在共享存储体系结构上的一个编程模型，包含编译制导(Compiler Directive)、运行库例程(Runtime Library)和环境变量(Environment Variables)，支持增量并行化(Incremental Parallelization) ，是C/C++ 和Fortan等的应用编程接口，已经被大多数计算机硬件和软件厂家所标准化。

MPI为程序员提供一个并行环境库，以实现分布内存环境下的并行编程，程序员通过调用MPI的库函数来达到程序员所要达到的并行目的，可以只使用其中的6个最基本的函数就能编写一个完整的MPI程序去求解很多问题。

OpenMP、MPI混合编程结合两种并行编程模式的优点，用于集群环境下的并行编程，在集群节点之间采用MPI进行消息传递，节点内部则采用OpenMP实现并行，从而充分发挥集群系统的计算能力。

**五、试验步骤**

1. 编写Pthread并行程序，设计实现斐波那契数列的并行程序。
2. 编写OpenMP或MPI并行程序，设计实现Floyd算法的并行程序；
3. 编写MPI或OpenMP并行程序，设计实现并行k-means算法。

**六、实验报告要求**

写清实验目的、实验内容、实验环境、实验步骤、实验结果（**要求有屏幕截图**）、实验结果分析。

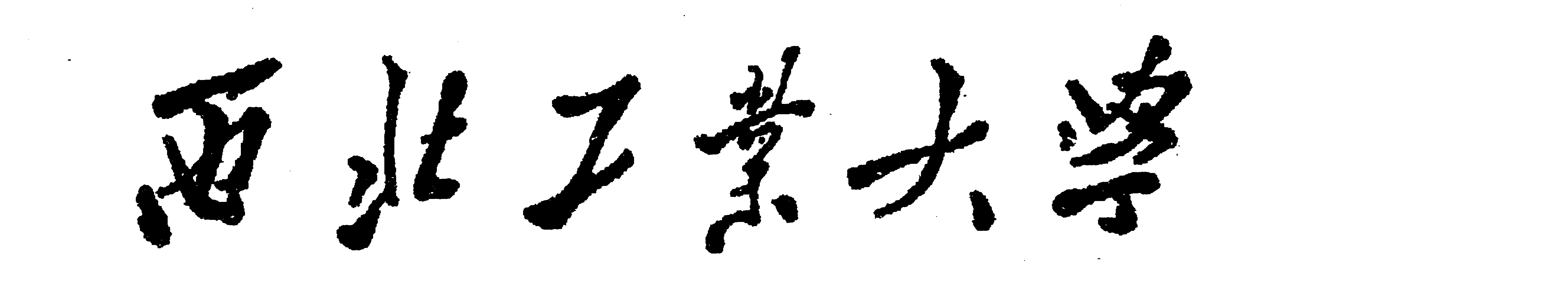
要求：

1. 封面写清楚课程名称、学院、学号、姓名等信息。
2. 说明软硬件环境；
3. 编写、编译、运行、调试代码；统计程序运行时间；
4. 说明算法设计思路，试着进行并行算法性能分析和可扩展性分析；
5. 提交源程序代码，要求能够运行通过，并得到正确结果。

**七、实验注意事项**

要求独立完成算法设计、编程实现，**切勿抄袭**。

附件：

本科生课程编程实验报告

|  |
| --- |
| **得 分：** |

**学 号**

**姓 名**

**考试课目 并行计算**

**考试日期 2024年11月**

**西北工业大学计算机学院**