

## 北京大学第十九届“江泽涵杯”数学建模竞赛题目

---

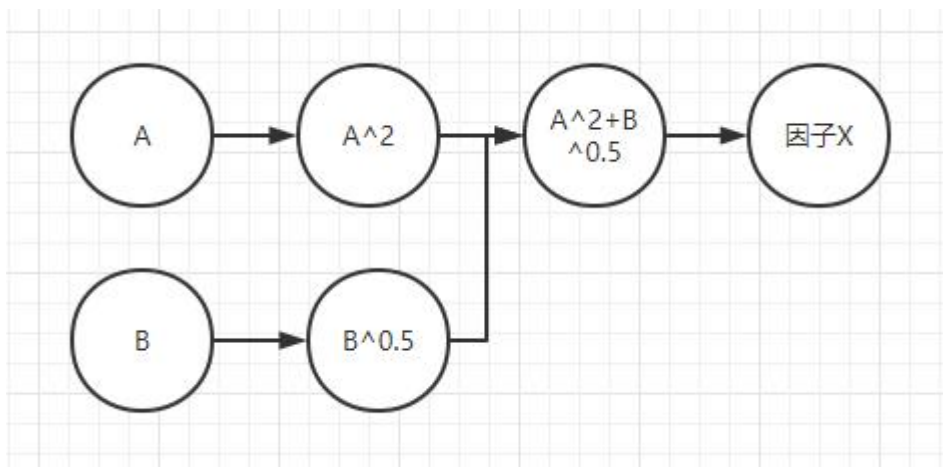
请先阅读竞赛规则：

<http://portal.math.pku.edu.cn/htdocs/showarticle.php?id=16675>

### B 题：因子高速计算优化问题

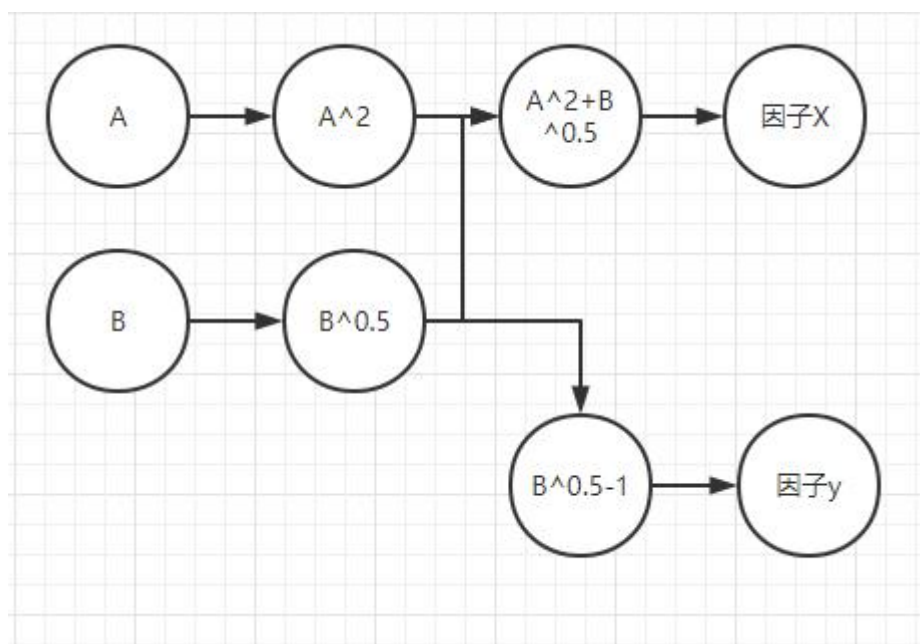
宽德投资作为一家量化基金每天要测试海量规模的策略，其中涉及到大量的因子计算。因此在有限的算力下，如何提高计算速度跟计算资源利用率是一个巨大的挑战。

每个因子的计算过程都可以被抽象为一个单独的有向无环图。例如：



（可以对应于因子  $1/(A^2+B^{0.5})$ ）

但由于不同因子可能会用到相同的中间变量，这就可以对这个图进行合并比如变成：



但现实情况会比这个稍微复杂一点，例如存在有：

1. 不同节点的计算时间不一致
2. 不同节点消耗内存不一致的问题
3. 内存是有存储上限（因此涉及到考虑，是否需要将节点切换到磁盘或者数据库上，或者重新计算）
4. Worker 数量限制（worker：同时工作的节点个数）
5. 同一任务在不同处理单元上的用时不一致；
6. 计算结果在不同处理单元之间传递的时间等等。

.....

这里为了简化问题，不妨假设

1. 每个计算节点消耗的时间都是相同的，设为单位 1。
2. 每个计算节点所需内存都是单位 1。

3. 总内存为  $T$ 。中间使用内存不允许超过  $T$ 。
4. 总 worker 数量为  $W$  个。(worker: 同时工作的节点个数)
5. 不考虑数据在不同 worker 之间通信的花费时间。

即：考虑在 worker、内存约束条件下的调度问题。(注：这里保证了，每个计算节点上输入，输出占用的总内存数量不超过总的内存限值)  
在这种情况下，输入任意一个有向无环图，在给定的约束下的，我们希望以尽可能少的计算时间完成计算。

我们考虑 2 个对模型的评判指标：

1. **任务调度成功率**。(可能出现的情况：因为内存的限值，有可能出现内存已经被占满，但之后的节点无法进行计算)
2. **任务调度 speedup 数值 (关键指标)**。Speedup 数值定义为每个节点的耗时之和/调度算法成功时的总耗时。关注 speedup 的均值，标准差，均值/标准差等指标。

数据读取方式：利用 python 中的 pickle 进行读取即可。