

第二届数学中国杯数学建模网络挑战赛

A 题: 串行算法的并行化处理

并行计算,是将一个计算任务分摊到多个处理器上并同时运行的计算方法。由于单个 CPU 的运行速度难以显著提高,所以计算机制造商试图将多个 CPU 联合起来使用。在巨型计算机上早已采用专用的多处理器设计,多台计算机通过网络互联而组成的并行工作站也在专业领域被广泛使用。台式机和笔记本电脑现在也已广泛地采用了双核或多核 CPU。双核 CPU 从外部看起来是一个 CPU,但是内部有两个运算核心,它们可以独立进行计算工作。在同时处理多个任务的时候,多核处理器可以自然地将不同的任务分配给不同的核心。但只运行一个以常规的串行代码写成的程序时,如何将计算任务拆分成多个部分并分解到多个核心上同时运行,是很困难的事情。有时甚至需要人工来读懂原来代码的含义,并以适合并行计算的语言重写程序。这需要耗费大量的人力物力。

最容易被并行化的计算任务称为“易并行”¹的,它可以直观地立即分解成为多个独立的部分,并同时执行计算。例如将一个数组里的所有元素求和。我们可以先将数组分成两段,对每段各自求和,最后再把结果相加。如果两段的大小划分得当,我们可以让双核 CPU 的每个核心的运算量相当,在数组规模很大时,总的运算速度比单核 CPU 能提高接近一倍。但并不是所有程序都能够分解成这种效果。

在软件行业中很自然地提出这个问题:希望设计一种方法,能够将一个常规的串程序分解成两个部分,使之能够在 CPU 的两个核心上并行运算,并且希望能够尽量发挥双核心的计算能力。一般来说,如果两个核心的运算量

¹有时也称作“自然并行”。

基本相当,那么总的运算效率较高²。如果出现某个核心空闲等待的状态,双核 CPU 的运算力就没有被充分的利用起来。

1 第一阶段问题：

问题： 请你们建立合理有效的模型,并依据模型对现成的串行算法进行处理。将能够使用双核心并行处理的部分分解开,并分配到两个核心上同时运行。以期达到比单核 CPU 处理更快速的目的。

具体要求： 假设算法是使用 C 语言写成的,代码里只有顺序执行、分支、循环三种结构。为简单起见,我们假设只对整型变量和整型数组进行操作,不需要调用已有的库函数。程序中所有的语句只包括简单的代数运算、赋值、条件分支语句 (**if-else** 语句),循环语句 (**while** 语句)。不包括其他语句。

这里的关键是如何通过分析代码,从总的计算任务中尽量识别可独立运算的部分,并估计每部分的计算量,来尽量使双核 CPU 发挥出超过单核的效能。

2 第二阶段问题：

问题： 在合理划分任务的前提下,双核处理器比单核处理器的运算效能要高。在算法级别上的划分,主要目的是进行数据的划分,以及顺序执行和循环的分解。分解的方法也就是第一阶段问题的目的。一个任务虽然可能能够分解到两个核心上分别处理,但不一定能较好地达到负载均衡的要求。由于算法的种类极多,所以分解方法也可能有适用范围的限制。请建立合理的模型,对你们设计的分解方法的效果作出评价,并根据你的评价,有针对性地对分解方法作出优化。

注： 解决这类问题,按照软件行业的要求,最终产品本应是一个代码预处理器,也就是可以把源代码进行分解的计算机程序。但完成本问题不需要进行编程,也不需要考虑语法分析等详细的问题。只需要在算法的层面上进行分析,给出确切的分析方法和模型即可。

²这个要求称为“负载均衡”或者“负载平衡”。