文章序号: 1001-2486 (2000) 03-0039-05

全球气象资料客观分析系统的分布式并行计算*

赵军, 宋君强, 张理论, 朱小谦

(国防科技大学计算机学院, 湖南 长沙 410073)

摘 要: 对已有的串行算法进行并行化, 是一项很困难的工作。通过对全球气象资料客观分析系统串行算法的研究, 提出了一种静态分配数据的分布式并行算法。该算法通过间隔选取分析盒子和模式格点纬圈行, 将数据分配给不同的处理机实现分布式并行。该并行算法负载平衡好, 并行效率高, 而且并行化代价较低, 具有良好的可扩展性。

关键词: 分布式并行; 分析盒子; 数据分配; 客观分析

中图分类号: TP301.6 文献标识码: A

Distributed Parallel Computation of the Global Atmospheric Data Objective Analysis System

ZHAO Jun, SONG Jun-qiang, ZHANG Li-lun, ZHU Xiao-qian

(College of Computer, National Univ. of Defense Techndogy, Changsha 410073, China)

Abstract: It is a difficult task to parallelize the existing sequential algorithm. A distributed parallelized algorithm is described by means of static data distribution after researching into the sequential algorithm of the global atmospheric data objective analysis system. We have achieved the parallelization by extracting the analysis boxes and model grid point latitude rows with leaped step and distributing data to different processors. The parallelization algorithm has a good load balance, a highly parallelization efficiency and a lower parallelization cost. Good scalability can be got from the algorithm.

Key words: distributed parallelization; analysis box; data distribution; objective analysis

中期数值天气预报系统对于实现天气预报的客观化、定量化、自动化和提高天气预报准确率具有重要意义。所谓数值天气预报就是在给定初始条件和边界条件的情况下, 数值求解大气运动基本方程组, 由已知的初始时刻的大气状态预报未来时刻的大气状态[1]。同时, 中期数值天气预报又是一个典型的高性能计算问题。中期数值天气预报系统主要包括资料同化系统和数值天气预报模式两部分。客观分析系统是资料同化系统的主要组成部分, 其主要作用是为数值天气预报提供初始时刻的大气状态即初始场。它使用了所有适当类型的观测资料, 给出数值形式的全球分析场。

随着大气探测手段的发展, 除了有来自全球各地的探空站、地面站、测风站等大量资料, 还越来越多地收集到各种非常规探测资料, 如飞机、各种气象卫星、探空火箭等不同类型的实时或非实时的信息[2,3]。但是充分利用这些大量而又复杂的气象资料, 其计算量将是巨大的, 必须使用高性能的并行计算机, 这就要求对客观分析系统进行并行化。从世界范围看, 数值天气预报业务主机的主流机型将是分布式内存并行计算机, 1996 年 12 月欧洲中期数值天气预报中心 (ECMWF) 举办的“并行计算在气象中的应用”研讨会已表明了这一点。本文在详细分析基于最优插值客观分析系统串行算法的基础上, 提出了一种静态分配数据的分布式并行化算法, 并在国产高性能计算机上实现了全球气象资料客观分析系统的分布式并行计算。

1 客观分析系统串行算法

对气象资料进行客观分析时, 有多项式法、订正法、最优插值法、谱方法、变分方法等几种主要方法。在本系统中, 分析方法采用最优插值方法。

* 收稿日期: 1999-12-07。
基金项目: 国家 863 计划资助项目 (863-306-ZD11-03-8)
作者简介: 赵军 (1972-), 男, 助理研究员。

1.1 最优插值方法

最优插值方法即是从统计意义来说，均方内插误差最小的线性插值方法[4]。最优插值方法具有下列优点：从统计意义来说，选取的权重使得分析误差最小，从而它对测站密度敏感性最差，而且可以同时进行多元分析（如同时进行风场、位势场、温度场分析），并提供一种利用不同时刻非常规资料的行之有效的方法；选取的最优权重只依赖于所考虑的气象要素的统计结构、它们的观测精确度以及测站相对于所考虑点的位置，不依赖于观测值。

A 表示任一个标量，E 表示这个标量的均方根误差；上标 i, p, o, t 分别表示插值、预报值、观测值和真值。基本插值方程为：

$$\frac{A_k^i - A_k^p}{E_k^p} = \sum_{n=1}^{N} w_{kn} \frac{A_n^o - A_n^p}{E_n^p}$$

令 $\alpha_n^o = \frac{A_n^o - A_n^t}{E_n^o}$, $\alpha_n^p = \frac{A_n^p - A_n^t}{E_n^p}$, $\alpha_k^i = \frac{A_k^i - A_k^t}{E_k^i}$, $\varepsilon_n^o = \frac{E_n^o}{E_n^p}$, $\varepsilon_k^i = \frac{E_k^i}{E_k^p}$。假定预报误差与观测资料误差相关为 0，采用使归一化内插误差方差的期望值达到最小的方法，确定每个观测资料的最优权重，最后得到最优插值方程：

$$\frac{A_k^i - A_k^p}{E_k^p} = \boldsymbol{B}^{\mathrm{T}} \boldsymbol{M}^{-1} \boldsymbol{P}_k = \boldsymbol{C}^{\mathrm{T}} \boldsymbol{P}_k$$

其中 $\boldsymbol{M} = \boldsymbol{P} + \boldsymbol{O} = [\langle \alpha_m^p, \alpha_n^p \rangle + \varepsilon_m^0 \langle \alpha_m^0, \alpha_n^0 \rangle \varepsilon_n^0]$, $\boldsymbol{B}$ 是归一化增量值 $\frac{A_n^o - A_n^p}{E_n^p}$ 的向量，$\boldsymbol{P}_k$ 是预报误差相关 $\langle \alpha_k^p, \alpha_n^p \rangle$ 的向量。

1.2 盒子树

客观分析系统中的很多计算都和盒子树有关，下面我们介绍构造盒子树的步骤[4]：

(a) 构造基本盒子树　将全球分成近似等边的基本盒子，如在两极之间有 32 排，边长为 5.62 经、纬度的基本盒子。对每个基本盒子，指定其最大、最小资料选取区域，在最小资料选取区域内，按每个垂直层计算资料个数。

(b) 生成子盒子树　如果在最小资料选取区域内的资料数目超过允许的最大矩阵阶数（一个常数，如取 451），则将此盒子分裂成四个具有相同经、纬度边长的子盒子。对每一个新的盒子，定义相应的最大、最小资料选取区域。如果在新的最小资料选取区域内的资料数目又超过了允许的最大矩阵阶数，则再把其对应的盒子分裂成四个。直至新盒子的最小资料选取区域内的资料个数小于允许的最大矩阵阶数为止。所生成的新盒子就构成了子盒子树。

在客观分析系统中对盒子树的计算，就是对盒子树的每片“叶子”即最终的分析盒子的计算。

1.3 客观分析系统串行算法

客观分析系统由：初始值、背景场的勒让德逆变换、观测资料加工扫描、质量场与风场分析、湿度分析、分析场的正谱变换等模块组成。其中，质量场与风场分析模块主要包括四个方面：

(a) 形成复合观测资料。

(b) 资料检查，剔除错误资料。主要算法简述如下：

```
为资料检查构造盒子树
for 每个最终分析盒子 do
    计算观测对观测的误差相关矩阵；误差相关矩阵求逆；
    计算观测对误差格点的误差相关矩阵；估算分析误差；
end for
```

(c) 估算分析系数。主要算法简述如下：

```
为估算分析构造盒子树
for 每个最终分析盒子 do
    计算观测对观测的误差相关矩阵；为分析系数求解线性方程组
```

end for

(d) 产生网格点分析值。主要算法简述如下:

for 每个模式格点纬圈行 do

找出本行的所有影响盒子; 在地面气压初估值上计算观测对格点的误差相关矩阵;

观测对格点的误差相关矩阵; 计算虚温的分析值;

计算 u、v 的分析值; 计算行的地面气压对数;

end for

湿度分析模块: 类似于质量场与风场分析, 不同的是分析变量是相对湿度;

2 客观分析系统的并行化

2.1 并行化的软件工程问题

在对客观分析系统进行并行化的过程中, 需要解决两个软件工程方面的问题: (a) 并行化后的客观分析系统应该能在尽可能多的不同类型机器上运行; (b) 现有的客观分析系统程序中所使用的FORTRAN 语言是标准 FORTRAN 语言的一种扩展。

对已有的程序代码进行并行化是一项很困难的工作, 其工作量是巨大的。因为现有的程序代码是为专用的向量计算机设计的, 而且经过了许多程序员的修改和长时间的发展, 其结构非常复杂。为了尽可能地减少并行化的代价以及考虑到系统的可移植性, 我们在并行化过程中使用通用的编程语言和工具:

FORTRAN 77: 对于主要的程序代码我们选用 FORTRAN 77 编程语言。因为现有的程序代码所用语言是在FORTRAN 77基础上扩展的, 选用FORTRAN 77 可减少大量的程序改写工作, 而且FORTRAN 77 的通用性很好。

标准 C 语言: 对于现有程序代码中在 FORTRAN 77 基础上扩展的部分, 采用标准 C 语言进行改写; 并对调用原有的向量机上的库函数, 使用标准 C 语言编写相应的高效函数代替。对于标准 C 语言大多数高性能计算机都是支持的, 而且 FORTRAN77 与标准 C 语言之间的接口的编写也是很容易的。

消息传递库: 可供选择的通用消息传递库有 MPI、PVM 等几种, 选用 MPI 作为消息传递界面[5]。因为 MPI 通信子程序很丰富, 使用起来很方便, 并且在同构机上可获得更高的速度[6,7]。

经过上述选择, 并行化过程中的两个软件工程问题可以得到很好的解决。

2.2 客观分析系统的并行化

在 1.3 中, 我们简单介绍了客观分析系统的串行算法。我们对客观分析系统的串行程序运行时间的组成进行分析后发现, 质量场与风场分析部分和湿度分析部分的运行时间占总运行时间的 90% 以上, 而且随着气象资料数目的增加, 这个比例还将增大。而这两部分的主要计算时间 (95%) 在于其资料检查、估算分析系数、产生网格点值部分。为了减少并行化的代价, 只对这几部分进行并行化工作。

并行化中的一个很重要的问题就是负载平衡问题, 它决定了数据的分配方式。注意到资料检查和估算分析系数两部分的计算单位为分析盒子, 如果各个分析盒子是独立的, 则可以进行并行计算; 而产生网格点值部分的计算单位是模式格点纬圈行, 且其产生的网格点值也是按纬圈行存储的, 因此, 这部分可按纬圈行进行并行化。

2.2.1 资料检查和估算分析系数两部分的并行化

这两部分的计算单位都是分析盒子, 只要在每个处理机上分配一部分分析盒子就可以进行并行计算。但是, 如何分配分析盒子才能使各个处理机上的负载达到平衡?

一般进行数据分配的方法有两种: 静态分配和动态分配。动态分配可以根据各个处理机的计算情况进行分配任务, 可达到较好的负载平衡。但是, 动态分配编程复杂, 且占用一定的开销, 对于 MPP 型的计算机来说占据的开销将是巨大的。因此, 在并行化过程中, 采用静态分配数据。

我们对全球的气象资料分布进行了研究，发现气象资料在全球的分布总体上是不均匀的：陆地密集，海洋及两极地区极为稀少。但是，在陆地上的分布却具有一定规律：美国、欧洲大陆、亚洲等主要大陆气象资料相对集中，并且对于资料检查和估算分析系数中构造的盒子树中的分析盒子来说，邻近盒子中的资料数相差很小，这就意味着邻近盒子的计算量几乎相同。这样，只要将资料密集地区分析盒子进行间隔选取分配，就可使各个处理机上的负载比较平衡。而对于海洋及两极这样资料极为稀少的地区，由于在这些地区的分析盒子中的资料一般只有几个，即计算量极小，对整个的负载平衡没有什么影响。

资料检查部分的并行化算法为：

采用广播的方式将复合观测资料发送给各个处理机
每个处理机上为资料检查构造盒子树（盒子树计算的代价小于通信的代价）
在每个处理机上 for 分析盒子号= 处理机序号，最大分析盒子号，步长= 处理机数 do
　计算观测对观测的误差相关矩阵；误差相关矩阵求逆
　计算观测对误差格点的误差相关矩阵；估算分析误差
end for
采用归约方式收集各个处理机的计算结果进行“拼合”，然后广播给各个处理机。

估算分析系数部分并行化算法为：

每个处理机上为估算分析构造盒子树（盒子树计算的代价小于通信的代价）
在每个处理机上 for 分析盒子号= 处理机序号，最大分析盒子号，步长= 处理机数 do
　计算观测对观测的误差相关矩阵；为分析系数求解线性方程组；
end for
采用归约方式收集各个处理机的计算结果进行“拼合”，然后广播给各个处理机。

这样，我们就对资料检查和估算分析部分进行了并行化。经过试验后，发现使用这种数据分配方法，各个处理机的负载是基本平衡的，其负载差别大约在 1% 左右。而且通信量也不大，并且随着处理机数目的增加，通信量增加很少。因此，通信量对并行效率的影响很小。

2.2.2 产生网格点分析值

这部分的计算单位为模式格点纬圈行。由于对于各个纬圈行的处理是独立的，因此，我们可以将纬圈进行划分，不同的处理机处理不同的纬圈行。由于不同纬圈行所包含的资料数目不同，其计算量也不同。因此，不能将全球简单地划分成几个纬圈带。由于全球气象资料分布具有相对集中特点，我们采用一种间隔选取纬圈行的方法分配纬圈行，这样可以使各个处理机轮流分配到计算量大或小的纬圈行，从而可使各个处理机上的负载比较平衡。

产生网格点分析值部分并行化算法：

在每个处理机上 for 纬圈行号= 处理机序号，最大纬圈行号，步长= 处理机数 do
　找出本行的所有影响盒子；在地面气压初估值上计算观测对格点的误差相关矩阵；
　观测对格点的误差相关矩阵；计算虚温的分析值；
　计算 u、v 的分析值；计算行的地面气压对数；
end for
采用集约方式收集计算结果至 1 号处理机，“拼合”得到完整的计算结果。

由于最后得到的完整计算结果数组的大小是一个常数，不随处理机数的增加而改变，因此当处理机数增加时，虽然通信量有所增加，但增加的幅度很小，总通信量不会超过计算结果数组的大小。

3 数值结果分析与结论

采用四组资料对于并行化的客观分析系统在某国产高性能计算机上进行了并行效率测试，这四组资料选自不同的季节、不同的时次，具有一定的代表性，分别为 1998 年 10 月 15 日 18 时、1999 年 1 月 15 日 00 时、1999 年 4 月 15 日 06 时、1999 年 7 月 15 日 12 时。下表为并行化的资料检查、估算分

析系数、产生网格点分析值在 1 处理机、2 处理机、4 处理机、8 处理机和 16 处理机情况下的加速比和并行效率。

表 1

Tab. 1

处理机数目	1998 年 10 月 15 日 18 时		1999 年 1 月 15 日 00 时		1999 年 4 月 15 日 06 时		1999 年 7 月 15 日 12 时	
	加速比	效率	加速比	效率	加速比	效率	加速比	效率
1	1	100%	1	100%	1	100%	1	100%
2	1.8412	92.06%	1.893	94.65%	1.8348	91.74%	1.89	94.50%
4	3.6376	90.94%	3.6876	92.19%	3.614	90.35%	3.7044	92.61%
8	6.8568	85.71%	7.0048	87.56%	6.7968	84.96%	6.9008	86.26%
16	12.832	80.20%	13.144	82.15%	12.818	80.11%	13.232	82.70%

从上表中, 可以看出对于上述三部分采用的并行化方法是可行的, 负载基本上平衡的, 并行效率也很高, 并且在一定处理机数目内具有可扩展性。同时, 由于在 18 时次、06 时次的气象资料较 00 时次、12 时次的气象资料数目少, 即 18 时次、06 时次的计算量较 00 时次、12 时次小, 因此 18 时次、06 时次的并行效率低于 00 时次、12 时次的并行效率。在将来气象资料数目急剧增加的情况下, 本文中的分布式并行化方法将很容易扩展到具有 32 处理机、64 处理机的分布式并行系统上, 即具有良好的可扩展性。

参考文献:

[1] 田永祥, 沈桐立, 葛孝贞, 陆维松. 数值天气预报教程 [M], 北京: 气象出版社, 1995. 11.

[2] Daley R. Atmospheric Data Analysis [M] . Cambridge Atmospheric and Space Science Series, Cambridge University Press, 1991.

[3] Da Silva A, Pfaendner, J Sienkiewicz M, Cohn S E. Assessing the Effects of Data Selection with DAO' s Physical-space Statistical Analysis System [C] . In International Symposium on Assimilation of Observations, Tokyo, Japan, 1995.

[4] 国家气象中心编译. 资料同化和中期数值预报 [M] . 北京: 气象出版社, 1991.

[5] Message Passing Interface Forum. MPI: A Message-Passing Interface Standard [S]. (draft obtainable by ftp from info. mcs. anl. gov directory pub/mpi) 1994.

[6] 孟杰, 孙彤, 李三立. MPI 网络并行计算系统通信性能及并行计算性能研究 [J] . 小型微型计算机系统, 1997, 1.

[7] 秦忠国, 姜弘道. 消息传递界面 PVM 和 MPI 的现状与发展趋势 [J] . 计算机研究与发展, 35 (6).