

区域水土流失模型的网络并行计算研究

张宏鸣¹, 杨勤科², 李锐³, 刘晴蕊¹, 王春梅³, 杜直前⁴

(1. 西北农林科技大学 信息工程学院, 杨凌 陕西 712100; 2. 西北大学 城市与环境学院, 陕西 西安 710069;
3. 中国科学院 水利部 水土保持研究所, 杨凌 陕西 712100; 4. 中国西安人才市场, 陕西 西安 710061)

摘要: 区域水土流失模型为区域尺度土壤侵蚀模型研究奠定了基础, 可用于区域降雨产流产沙的预报, 为区域水土保持和环境问题的宏观决策提供支持。目前该模型在实际运行存在计算速度慢, 效率低, 难以进行大尺度的快捷运算等问题。对模型过程进行了分析, 采用网络并行计算方法及相应的任务调度规则来提高模型运行效率。结果表明, 模型的运行时间大幅减少, 计算效率显著提升, 该方法为解决海量数据的 GIS 模型计算问题提供了一个可行方案。

关键词: 区域水土流失模型; 并行计算; GIS; 数据分解

文献标识码: A

文章编号: 1000-288X(2011)01-0141-04

中图分类号: TP302.1

Network Parallel Computing of Regional Soil and Water Erosion Model

ZHANG Hong-ming¹, YANG Qin-ke², LI Rui³, LIU Qing-rui¹, WANG Chun-mei³, DU Zhi-qian⁴

(1. College of Informodion Engineering, Northwest A & F University, Yangling, Shaanxi 712100, China;
2. Department of Urbanology and Resource Science, Northwest University, Xi'an, Shaanxi 710069; 3. Institute of Soil and Water Conservation, CAS & MWR, Yangling, Shaanxi 712100, China; 4. China Xi'an Human Resources, Xi'an, Shaanxi 710061, China)

Abstract: As the foundation for regional soil erosion model, Regional Soil Lose Model can be used for regional rainfall runoff and sediment forecasting, soil and water conservation and environmental issues to support regional macro decision-making. Currently, the model cannot be used effectively for large areas due to its slow computing speed. Aiming at the problems, this paper analyzed the model process and incorporated network parallel computing methodology and corresponding task scheduling rules into the model. The results show that the running time greatly reduced in the revised model and the computational efficiency significantly improved. This method provides a way to solve the problem caused by huge amount of data in GIS model with a feasible scheme.

Keywords: regional soil and water erosion; parallel computing; GIS; distributed hydrologic model; data partition

在区域水资源环境研究中, GIS 已经越来越多地被作为一种实用工具用来分析研究资源环境要素的空间变化^[1]。近年来, 随着计算机处理速度的不断提高, 在很多应用领域计算速度已经成为越来越小的影响因素, 但在基于 GIS 的大尺度应用领域里, 算法的复杂化、数据的海量化依然是其制约因素之一^[2-3]。网络计算技术的发展为 GIS 模型的实现提供了越来越多的基于网络计算技术的解决方案。

区域水土流失模型是大区域土壤侵蚀普查和水土保持宏观决策的支持工具, 土壤侵蚀模型的研发是

土壤侵蚀学科的前沿领域^[4]。姚志红、杨勤科^[5]等设计了考虑“剥蚀—搬运—沉积”这一完整过程和关系的区域水土流失模型。该模型运行结果令人满意, 但模型输入、输出、计算的数据量较大, 同时模型模拟实际降雨过程, 计算复杂度较高。目前大尺度问题的研究, 如气象^[6]、海洋波浪模拟^[7]、大地震模拟^[8]等都采用了并行计算的策略。水资源研究中, 李铁键等^[9-10]采用集群计算机进行并行处理。为了提高区域水土流失模型运行效率, 使其能够在区域尺度下高效、快捷地获取计算结果, 本文尝试采用基于网络的

收稿日期: 2010-07-02

修回日期: 2010-08-18

资助项目: 国家自然科学基金项目“中低分辨率坡度变换方法研究”(40971173); 黄河水利委员会项目“黄土高原水土流失数学模型研发”(黄水保 200651-08); “973”国家重点基础研究发展计划项目“区域水土流失模型研究”(2007CB407203)

作者简介: 张宏鸣(1979—), 男(蒙古族), 内蒙古赤峰市人, 讲师, 在读博士, 主要从事网络计算技术与 GIS 研究。E-mail: zhm@nwsuaf.edu.cn。

通信作者: 杨勤科(1962—), 男(汉族), 陕西省陇县人, 博士, 教授, 主要从事基于遥感和地理信息系统技术的区域水土流失评价研究。E-mail: qkyang@nwu.edu.cn。

并行计算的方法,从算法并行化及空间数据分解分析,将模型运行效率大幅提高,为区域水土流失模型进一步应用提供了一个可行方案,也为网络并行计算在区域水资源研究的应用进行了一定的探索。

1 GIS 模型网络并行计算原理

1.1 网络并行计算

网络并行计算是一种将需要进行大量计算的工程任务分割成小块,分发给网络中的节点,由多台计算机接收并完成处理任务,再把每个节点机的处理结果上传、汇总,而后再统一合并处理得出结论的计算方法^[11-12]。

进行网络计算,必须将计算任务进行分解,或称为计算功能分解,分解后的任务被分配到不同的节点,达到多机网络并行处理的目的。任务划分是实现任务处理并行化的基础,将串行进行的任务划分成多个独立的几乎等尺度的子任务,用多台处理机并行解决诸子任务。任务划分过程如图 1 所示。

GIS 模型的实现是复杂的,它的过程可能有离散的影响因素,也可能有连续变化的成分;有线性的,也有非线性的部分。GIS 中常见的模型计算任务分解方法可分为计算模型的串行算法并行化和空间数据分解。

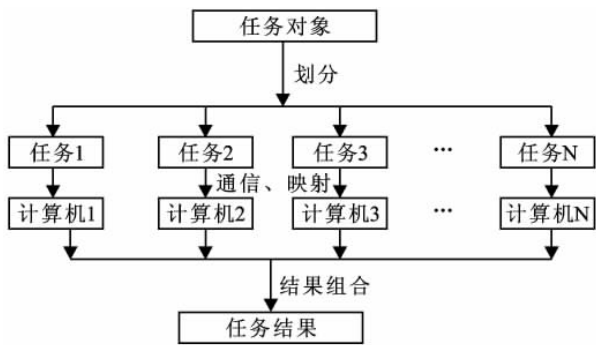


图 1 网络并行计算任务执行过程

1.2 计算模型的串行算法并行化

对模型的串行算法进行分析、分解使模型分解为多个可并行计算任务的子任务,从而达到网络处理目的。根据多处理机并行计算过程对处理机间信息交换的依赖方式的不同可分为同步并行计算方法和异步并行计算方法^[13]。

1.2.1 同步并行计算方法 同步并行计算方法强调负载均衡,是指算法中有一个可以为所有节点提供的同步点,当网络中的所有节点都达到该同步点后才完成下一步的计算工作。

例如计算: $X=AB$ (1)

$$\text{其中 } A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix},$$

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

计算 X 的同步并行算法可以由 4 个节点组成,如图 2 所示。当节点 P_1, P_2, P_3 分别完成 X_{11}, X_{21}, X_{31} 的计算,到达同步点后将结果汇总并获得下一步计算所需数据,进行计算,直到最终完成计算任务。

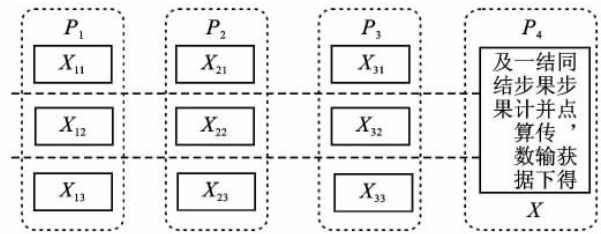


图 2 四节点同步并行算法示意图

1.2.2 异步并行计算方法 异步并行算法是指计算节点间的计算任务不具有同步点的并行算法,也就是用于等待同步操作的时间内并行计算机系统内每一处理器各自完成自己的有用工作。与同步计算相比,当某节点等待其它节点到达某一同步点时,异步并行计算提供该节点有用的计算供其执行。例如计算:

$$X = (A/B) \times (C+D) + E \times F + G \quad (2)$$

计算 X 的异步并行算法可以由 3 个节点完成 P_1, P_2, P_3 (如图 3)。节点 P_2 的部分计算任务要等待节点 P_1 的计算任务结束,当节点 P_2 第二步运算时,节点 P_3 不需要考虑同步问题,完成自身计算任务,最终等待 P_2 的计算任务完成后完成全部计算工作。

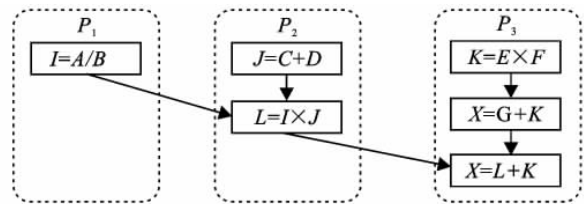


图 3 三节点异步并行算法示意图

2 计算模型的空间数据分解

将空间数据分解为等粒度的数据块,或按照模型特征分解数据,达到多节点并行计算的目的。

2.1 将空间数据分层

主要针对具有层次特征的空间信息,将多源异构的信息集成任务划分为不同的独立对象完成,子任务

之间的协同模式为并行计算模式。如姚艺强等^[14]利用图层式并行算法实现了武汉城市地图中一条道路缓冲区分析的并行计算。当然不同图层的任务计算完成时间可能是不等的,在实际划分中为了平衡运算时间,可能多个图层被分配到一个节点。其算法示意图如图 4 所示。

X 代表运算结果, F 代表不同图层内的运算规则, A 代表不同图层, 则其数学表达形式可写为:

$$X = FA \quad (3)$$

其中 $F = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}$, $A = [A_1, A_2, \dots, A_n]$

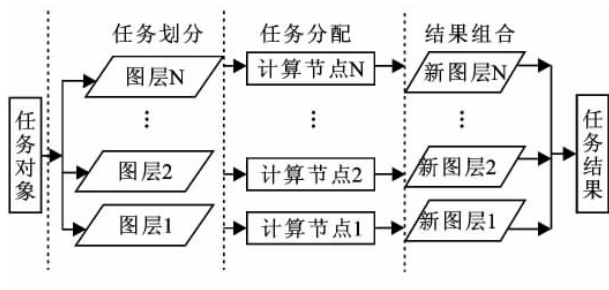


图 4 图层并行计算示意图

2.2 将空间数据分块

对于海量空间数据而言,除了可以将空间数据按照图层的方式来完成以外,还可以将数据分块来完成空间数据的划分。易法令等^[15]在将 GIS 空间数据格式进行并行转换的调度算法中就采用该方法,尤其在图层间数据存在一定的计算关系的情况下,该分块方法可用性较高。如图 5 所示,以栅格数据为例,将其分块。

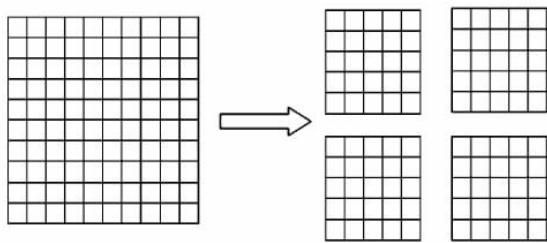


图 5 栅格数据分块示意图

其算法步骤可表达如下: n 代表 GRID 的数量, 划分成 m 个块, $x = n/m$, 即将

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \quad (4)$$

划分为:

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mm} \end{bmatrix} \quad (5)$$

其中:

$$A_{11} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1x} \\ a_{21} & a_{22} & \dots & a_{2x} \\ \vdots & \vdots & \vdots & \vdots \\ a_{x1} & a_{x2} & \dots & a_{xx} \end{bmatrix}$$

$$A_{12} = \begin{bmatrix} a_{1x+1} & a_{1x+2} & \dots & a_{12x} \\ a_{2x+1} & a_{2x+2} & \dots & a_{22x} \\ \vdots & \vdots & \vdots & \vdots \\ a_{xx+1} & a_{xx+2} & \dots & a_{x2x} \end{bmatrix} \dots$$

$$A_{1m} = \begin{bmatrix} a_{1n-x+1} & a_{1n-x+2} & \dots & a_{1n} \\ a_{2n-x+1} & a_{2n-x+2} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{xn-x+1} & a_{xn-x+2} & \dots & a_{xn} \end{bmatrix} \dots$$

$$A_{mm} = \begin{bmatrix} a_{n-x+1n-x+1} & a_{1n-x+2} & \dots & a_{1n} \\ a_{n-x+2n-x+1} & a_{2n-x+2} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m-x+1} & a_{m-x+2} & \dots & a_{m} \end{bmatrix}$$

3 区域水土流失模型的网络并行实现

3.1 区域尺度土壤侵蚀过程

根据对坡面、小流域和区域土壤侵蚀模型的分析总结,参考分布式水文模型对大区域水文过程的描述和水土流失治理的过程,区域尺度的土壤侵蚀过程表现为 3 个方面,即产流产沙过程,水沙物质汇集和传输过程,水土流失治理过程。

区域水土流失模型研究并设计了产流产沙过程、水沙物质汇集和传输过程,利用 NET 框架下的 C# 语言,基于 AO 组件完成初步设计,并行数据传输采用 NET Remoting 方法。模型以姚志宏等区域尺度降雨径流估算方法为基础,其计算过程参考文献^[5]。

3.2 研究区概况

延河是黄河右岸,中游区上段的河口镇至龙门段的一级支流,位于北纬 $36^{\circ}21' - 37^{\circ}19'$,东经 $108^{\circ}38' - 110^{\circ}29'$,发源于陕西省靖边县东南天赐湾乡周山,由西北向东南流经志丹、安塞、宝塔、延长等 4 县区,在延长县南河沟乡凉水岸附近汇入黄河。延河干流全长 284 km,河道平均比降 3.29%,流域面积 7 725 km^2 ,属于温暖半干旱大陆性气候,年平均气温 9.3 $^{\circ}\text{C}$,多年平均降雨为 511 mm。

3.3 网络并行计算方法设计及结果

3.3.1 任务划分 对于区域水土流失模型网络并行的实现过程而言,其算法复杂度并不高,但数据串行化计算要求较多,计算过程产生的 9 个新图层中只有初始计算的 3 个图层是依据不同数据获得的,可用于网络并行计算,时段径流深后的计算则依据上一次计算结果运算得到。更关键的是,如果将算法进行并行划分,因模型计算使用的数据存在串行关系,那么就使得不同节点间的数据传输量增大,不易用于网络并行实现^[16]。与现有的 GIS 模型并行算法相比,数据分解的 2 种方法在此均不完全适用,区域水土流失模型的数据层间存在运算关系,故不易分层;数据只能分块来进行计算,直接分块会将关联数据分离开,导致模型运算无法进行。表 1 给出了以沿河流域数据为基础,基于 3 个节点按植被截留、入渗量、填洼同时进行计算所需时间关系。对于在试验中单机运行需要近 41 min 的计算来讲,利用算法的可划分部分进行并行化处理的效果并不十分理想。数据分块必须保证分开的数据块间无相互影响,而产流产沙过程、水沙物质汇集和传输过程在河流集水区间的分水线达到划分条件的要求,因此,模型的并行任务划分就以分水线为分界线将大区域划分为若干个流域。

表 1 三节点不同任务执行时间对比表 s

节点	计算结果	执行时间	数据接收时间	结果返回时间
P_1	植被截留	235	6	7
P_2	入渗量	783	9	11
P_3	填洼	233	7	8

3.3.2 调度策略 任务调度是网络并行计算中最基本的一个问题,当计算开始后,系统完成任务的划分,同时要将任务恰当地分配给节点机,完成任务调度,使得每一个节点机得到有效的利用。任务调度问题是一个 NP 问题^[17],一般采用启发式的方法解决,这些方法可以分成许多类,如列表调度算法、基于聚类的算法和基于复制的算法,它们主要应用于同构的处理器系统中^[18]。本文的调度算法采用列表调度算法,通过对节点分配优先级别来构造一个调度列表,优先权是由计算花费和通信花费给出,然后重复从调度列表中顺序取出最优先节点,将节点分配到使它的启动时间最早的处理器上,直到任务图中所有节点被调度完毕。

表 2 给出了按流域不同月份的网络并行运算结果参数,可以看出网络损耗时间要远远小于单机模型计算时间。

表 2 三节点不同月份执行时间对比表 s

节点	计算月份	执行时间	数据接收时间	结果返回时间
P_1	4	2 431	21	20
P_2	5	2 523	23	21
P_3	6	3 002	26	25

4 结论

(1) 依靠计算机网络,完成区域水土流失模型的并行计算,能够大幅提高计算效率,对于解决区域尺度的计算提供了有效的方案。

(2) 区域水土流失模型为分布式模型,植被截留、入渗、填洼、流速、携沙能力、径流剥蚀量、泥沙沉积等计算过程经过连续的多次迭代,具有高耦合性,适合将数据按分水线进行划分,进行网络并行计算。

(3) 数据量大小决定了网络延时和计算时间,多少栅格的数据才适合网络并行计算,如何在主控机自动将数据划分,并将计算结果回收后拼接、恢复原始区域的大小还需要进一步的研究和验证。

[参 考 文 献]

- [1] 杨勤科, MeVicar T. GIS 及其在水土保持研究中的应用[J]. 水土保持通报, 2005, 25(6): 55-58.
- [2] Kenneth A H, Coddington P D, James H A. Distributed frameworks and parallel algorithms for processing large-scale geographic data [J]. Parallel Computing, 2003, 29(10): 1297-1333.
- [3] Park J, Sandhu R. The UCONABC usage control model [J]. ACM Transactions on Information and System Security, 2004, 7(1): 128-174.
- [4] 杨勤科, 李锐, 曹明明. 区域土壤侵蚀定量研究的国内外进展[J]. 地球科学进展, 2006, 21(9): 31-38.
- [5] 姚志宏, 杨勤科, 吴喆, 等. 区域尺度侵蚀产沙估算方法研究[J]. 中国水土保持科学, 2007, 5(4): 13-17.
- [6] 袁金南, 王在志, 薛纪善. 广州区域数值预报模式并行化计算[J]. 应用气象学报, 2004, 15(5): 556-563.
- [7] Kashiyama K, Saitoh K, Behr M, et al. Parallel finite element methods for large scale computation of storm surges and tidal flows[J]. International Journal for Numerical Methods in Fluids, 1997, 24(12): 1371-1389.
- [8] 李丽君, 金先龙. 空间多点地震动模拟的并行计算方法及软件[J]. 系统仿真学报, 2005, 17(5): 1068-1071.
- [9] 李铁键, 刘家宏, 和杨, 等. 集群计算在数字流域模型中的应用[J]. 水科学进展, 2006, 17(6): 841-846.

(下转第 230 页)

点,为使土方量二次搬运最少,要准确计算复垦区挖方量和填方量。应用 CASS 软件,根据 DEM 求出该片土地的填方和挖方量,设计复垦地的高程为 25.036 m,此时挖填土方量达到平衡,节省成本。然后根据塌陷地的实际下沉值很大,考虑到最大程度的利用好每一块塌陷地,作者在复垦时在该块土地东南角设计了 3 个鱼塘的方式最大限度地对耕地加以修复。

4 结论

本文应用自主开发的煤矿开采地表移动的变形预计系统(MMSPS)计算软件对矿井地表沉降现状和动态变化趋势进行了计算预测,掌握了在规划期内的地表沉降动态变化规律,为沉陷区征迁规划研究提供了科学依据。经在采煤沉陷地复垦地的实地应用,证明本文所采用的方法和所获的效益是可行的,同时研究区任楼矿区属于大型现代化矿井,井田范围大,地物复杂,本文研究成果对于该矿井及其它矿井土地复垦征迁规划具有指导意义。

[参 考 文 献]

- [1] 赵长胜,周复旦.任楼煤矿地表塌陷规划应用研究[R].徐州师范大学测绘学院,2010.
- [2] 任楼矿区地质报告[R].皖北煤电集团有限责任公司任楼煤矿,2004.
- [3] 邹友峰,邓喀中,马伟民,等.矿山开采沉陷工程[D].江苏徐州:中国矿业大学出版社,2003.
- [4] 祝启坤,郭娜娜,黄昱清.赤马山铜矿 I 号塌陷区地质灾害综合防治对策研究[J].金属矿山,2010,35(4):153-157.
- [5] 李树志,高均海,鲁叶江,等.平原矿区采煤沉陷地复垦耕地生产力评价[J].矿山测量,2010,30(1):5-9.
- [6] 于丽梅,赵迎春.煤矸石及综合利用[J].煤炭技术,2008,27(11):127-128.
- [7] 张国良.矿区环境与土地复垦[M].徐州:中国矿业大学出版社,1997.
- [8] 张连贵.兖州矿区综放开采地表沉陷规律[J].煤炭科学技术,2010,38(2):89-92.
- [9] 何国清,杨伦,凌赓娣,等.矿山开采沉陷学[M].徐州:中国矿业大学出版社,2003.
- [10] 林振山,王国祥.矿区塌陷土地改造与构造湿地建设:以徐州煤矿矿区塌陷地改造为例[J].自然资源学报,2005,20(5):790-795.
- [11] 罗爱武.淮北市采煤塌陷区土地复垦研究[J].安徽师范大学学报:自然科学版,2002,25(3):288-289.
- [10] 王皓,傅旭东,孙其诚,等.大尺度流域水文并行计算的方法改进[J].应用基础与工程科学学报,2009,17(S1):1-8.
- [11] 沈志东,陈珉.分布式对象技术在分布式 GIS 方面的应用[J].计算机工程,2003,29(4):62-64.
- [12] Foster I, Kesselman C. The Grid:Blueprint for a New Computing Infrastructure[M]. San Francisco: Morgan Kaufmann Press, 1998.
- [13] 陈先桥,郭庆平.异步分布式并行 SRM 算法[J].武汉理工大学学报:交通科学与工程版,2003,27(5):614-617.
- [14] 姚艺强,高劲松,孟令奎,等.网格环境下缓冲区分析的并行计算[J].地理空间信息,2007,5(1):98-101.
- [15] 易法令,严圣华. GIS 空间数据格式并行转换的调度算法[J].计算机工程,2004,30(23):47-49.
- [16] Wang S W, Armstrong M P. A quadtree approach to domain decomposition for spatial interpolation in grid computing environments [J]. Parallel Computing, 2003,29(10):1481-1504.
- [17] Ranaweera A, Agrawal D P. A scalable task duplication based algorithm for heterogeneous systems[C]// Proc of the ICPP. Los Alamitos: IEEE Computer Society Press, 2000:445-450.
- [18] 张忠平,刘欣媛.网格环境下静态启发式任务调度算法[J].计算机研究与发展,2008,45(S):21-25.

(上接第 144 页)