

引用格式:Liu Longli,Xue Yong,Guang Jie,*et al.*Remote Sensing Information Service Grid Node and the Research of Data Compression and Task Allocation[J].Remote Sensing Technology and Application,2016,31(2):247-254.[刘龙历,薛勇,光洁,等.遥感网格的数据压缩与任务分配方法研究[J].遥感技术与应用,2016,31(2):247-254.]
doi:10.11873/j.issn.1004-0323.2016.2.0247

遥感网格的数据压缩与任务分配方法研究

刘龙历^{1,2},薛 勇¹,光 洁¹,刘 佳^{1,2}

(1.中国科学院遥感与数字地球研究所数字地球重点实验室,北京 100094;
2.中国科学院大学,北京 100049)

摘要:网格计算作为分布式计算在科学计算领域的发展方向,可以为对地观测数据的处理提供强大的计算力。在分析遥感信息服务网格节点(Remote Sensing Information Service Grid Nodes, RSSN)中网络数据传输和负载均衡两个关键问题的基础上,提出了一种有效的基于游程编码和 Huffman 编码的数据压缩方法和基于“计算端元”的任务分配策略,该方法针对遥感影像特点进行有效数据压缩,具有较好的压缩比,达到了 17%,且能实现任务负载均衡。并在遥感信息服务网格节点计算平台上,以中国范围内 1 km 分辨率气溶胶光学厚度(Aerosol Optical Depth, AOD)反演计算为例,从压缩率和计算时间效率方面验证和分析了上述方法的有效性。

关 键 词:分布式计算;网格计算;气溶胶反演;遥感信息网格节点;数据压缩;任务分配

中图分类号:TP 79 **文献标志码:**A **文章编号:**1004-0323(2016)02-0247-08

1 引 言

近年来,不断发射的卫星传感器积累了爆炸式增长的对地观测数据。传感器的质量和分辨率在不断地提高,对快速处理和分析遥感数据,并有效地从中提取有用的信息提出了急切的需求^[1]。例如 17 年的全球范围 Advanced Very High Resolution Radiometer(AVHRR)历史数据有 4.65 TB,在单机情况下处理完这些历史数据,应用于全球变化,需要花费近 3.5 a^[2];在地球资源与科学数据中心(Earth Resources Observation and Science,EROS)需要处理、归档和发布不同传感器的数据,仅仅是 Landsat 数据就有 1 PB,而且以每天 300 GB 的速度增长。这对数据的存储和处理都提出了很高的要求^[3]。数据密集型计算是一类使用数据并行方式处理大量数据(TB 或 PB)的并行计算应用,其典型特点是数据结构复杂、数据量大或者获取速度慢。解决日益增

长的数据量和复杂性要求在软件、硬件和算法开发方面具有划时代的进步。相反,计算密集型计算则把大部分时间集中在计算处理上^[4]。遥感反演问题常常既是数据密集型问题,又是计算密集型问题。且在整个反演流程中,由于反演步骤之间存在复杂的依赖关系,导致大量的数据来回移动,这对有效率的网络数据传输提出了要求^[5]。

并行机和超级计算机是解决计算问题的有效途径之一,但由于价格高昂,且需要专业人员的日常使用和维护,使得如何使用松散耦合的网格平台来完成快速处理大量遥感数据,解决遥感反演中计算耗时问题,达到高吞吐量和高性能计算的目的,在分布式并行计算、动态环境监测和全球变化研究中具有重大意义^[6]。构建了一个高通量遥感信息网格服务平台(Remote Sensing Information Service Grid Node, RSSN),用于气溶胶光学厚度(Aerosol Opti-

收稿日期:2015-01-27;修订日期:2015-04-28
基金项目:国家自然科学基金委面上项目(41271371);中国科学院知识创新工程项目(Y3SG0300CX)。
作者简介:刘龙历(1988—),男,湖南衡阳人,硕士研究生,主要从事网格计算、空间信息服务研究。E-mail:liulongli001@126.com。
通讯作者:薛 勇(1965—),男,江苏徐州人,研究员、博士生导师,主要从事定量遥感、高性能地学计算、空间信息服务等研究。
E-mail: xueyong@radi.ac.cn。

cal Depth, AOD)反演,然而在实际的实验过程中,发现其中影像数据传输量较大,而且其任务分配方式不均等。

通常情况下有效的数据压缩是解决网络数据传输的有效方式^[7],能够有效地进行数据存储、分发和共享。压缩方法按照失真率通常可以分为有损压缩和无损压缩。有损压缩中重构数据与原来数据有所不同,但不影响原始资料表达和显示,压缩率一般相对较高。常用的有损压缩算法有预测编码、变换编码(如小波变换等)、插值和外推。通常无损算法压缩率比较低,一般为 2~5 倍,主要包括基于稀疏矩阵的三元组算法(3-Tuples)、游程编码(Run Length Encoded, RLE)、基于 LZW 的压缩算法和 Huffman 编码等^[8]。在遥感影像有损压缩方面,邱玉宝等^[9]研究了 ECW 算法在高分辨率遥感图像压缩与分发中的应用,提出了一种新的分发与浏览体系结构;卞燕山等^[10]设计了一种基于 EZW 算法的遥感影像压缩算法,用于遥感影像的网络传输。在遥感影像无损压缩方面,Wang 等^[11]基于上下文条件平均算法实现了高光谱影像的无损压缩;刘兴科等^[12]研究了 Huffman 编码在地图压缩中的应用,指出该方法具有无损、高效、压缩率高的特点;李云锦等^[13]基于算术编码,提出了一种新的栅格电子地图的无损压缩算法,具有较高的压缩比和解码速率。而在高性能气溶胶反演中,由于是针对遥感影像的信息处理过程,所以其中的信息不能丢失或者改变,所以一般来讲只能采用无损压缩方法。关于网络任务分配问题,学者也进行了相关的研究,提出了一些基于遗传的任务分配算法^[14-15]和基于蚂蚁的任务分配方法^[16],然而这些方法都只是考虑单纯的网格计算任务,没有考虑遥感影像处理的大数据量,强耦合的特点,并不能完全适应。而对于遥感影像的并行处理研究,大多采用均匀的数据分块方法^[6],存在任务量的不均等性。

在研究了遥感影像数据压缩方法和并行任务分配方法的基础上,提出了一种改进的有效的遥感影像数据压缩方法和基于“计算端元”的任务分配策略,拟解决遥感网格中的数据传输量大和任务分配不均等问题。并以中国范围 1 km 分辨率的气溶胶遥感反演为例,对上述方法的有效性进行了验证和分析。

2 RSSN 平台的系统体系结构

RSSN 是一个用于生产全亚洲范围的气溶胶光

学厚度产品的遥感网格服务节点(<http://www.tgp.ac.cn>),部署在中国科学院遥感与数字地球研究所,服务于全球变化或其他相近的研究领域的大量历史数据处理^[17]。RSSN 网格节点包含有 30 多台商用 PC 机,充分利用 HTCondor(<http://research.cs.wisc.edu/htcondor>)的有效利用闲散的桌面资源的特点,为全球变化研究中大量历史数据的处理提供强大的计算能力。其平台体系结构如图 1 所示。在基础物理层之上有一个网格工作流管理系统,为用户提供工作流服务。

在实际遥感反演试验过程中,由于其步骤之间的高度依赖关系,导致在 RSSN 平台计算过程中存在大量遥感数据的网络迁移,且存在长时间任务等待短时间任务的不均衡现象,所以研究遥感数据压缩和任务分配策略问题具有现实意义。

3 数据压缩与任务分配子模块研究

3.1 遥感反演影像数据特点分析

一般来讲,气溶胶反演中,其影像数据经过去噪、去云和几何纠正等处理之后,一般具有稀疏的特点,即在整幅遥感影像中,很少出现大面积区域全是有效数据点的情况,所以造成网络传输带宽的浪费。如图 2 几何纠正后的影像和融合后的影像所示(黑点均为无效数据)。

3.2 影像数据压缩方法

游程编码是把一系列的重复值用一个单独的值再加上一个计数值来取代,如字符串“PPPPPPPP-Pxxxxxyyyyyyyzzzzzz”,可表示为 9P4x7y6z。游程编码算法具有实现简单、压缩还原速度快等优点,只需扫描一次原始数据即可完成数据压缩,对具有长重复值串的信源压缩效果较好。而气溶胶反演中的影像数据具有简单聚集的特点,且需要较快的压缩和解压缩效率,可以适用游程编码的特点。游程编码的压缩效率主要取决于数据本身的连续重复率。很明显,在气溶胶反演中的遥感数据主要取决于无效冗余数据,如果无效数据(即默认值)越多,其压缩比越高。但这也同时存在一个问题,当无效数据重复数据超过 32 767 个时,原来的统计行程长度变量需要定义为长整型(long int),该变量需要 4 个字节的存储空间。而整幅影像中的绝大部分重复像元都有限,这样就会造成很大的存储空间浪费。所以需要大块无效数据进行特殊处理,以减少统计游程的长度。为此,根据校正后的影像有效数据集

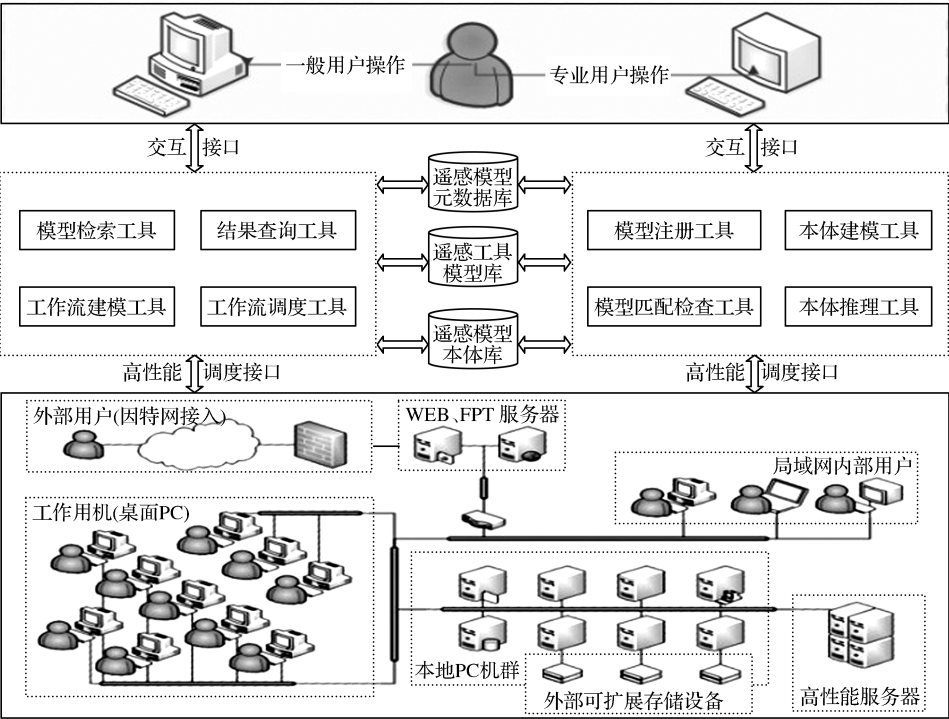


图 1 RSSN 平台体系结构图

Fig.1 RSSN platform system architecture

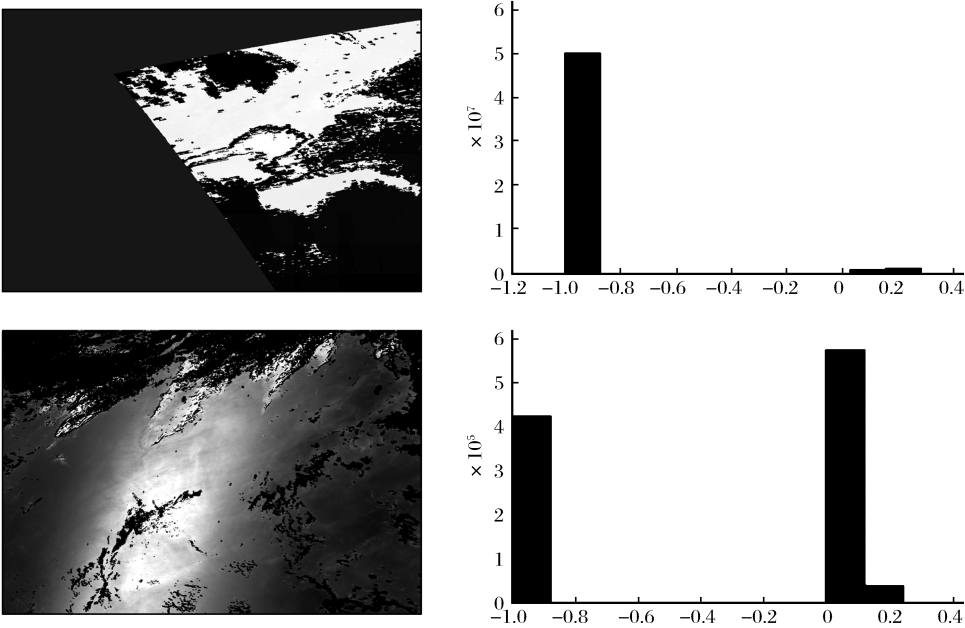


图 2 几何纠正影像和融合后影像及统计分析

Fig.2 Statistical analysis of geometric correction image and fusion image

法,用最小外包矩形将有效数据的最大范围表示出来,然后针对这个小区域进行游程编码,在该小区域中通常不会出现很长的游程统计,有效避免统计游程过长的缺点。同时,将最小外包矩形外的无效数据隐含在最小外包矩形关系中表达,而不需要实际编码,可进一步压缩数据。另外,对于最小外包矩形

中的有效数据部分,其有效数据通常来讲是不连续的,特别地类似于“ABCD”,其游程编码为“1A1B1C1D”,这样其编码的大小反而增加了一倍。为了解决这一“病态”游程情况,在气溶胶反演遥感影像中,其有效数据值非负,所以可取消行程编中区分压缩数据的最高标志位并将非连续重复的数据取

反,即变为负数,这样就可以通过正负来区分压缩数据和非压缩数据,以此达到进一步优化游程编码的目的。通过此方法处理,上面的字符串“ABCD”可编码为“-A-B-C-D”,还原时判断若为负值则取绝对值为“ABCD”,保证信息完全还原,解决了病态游程问题。其算法流程如图 3 所示。

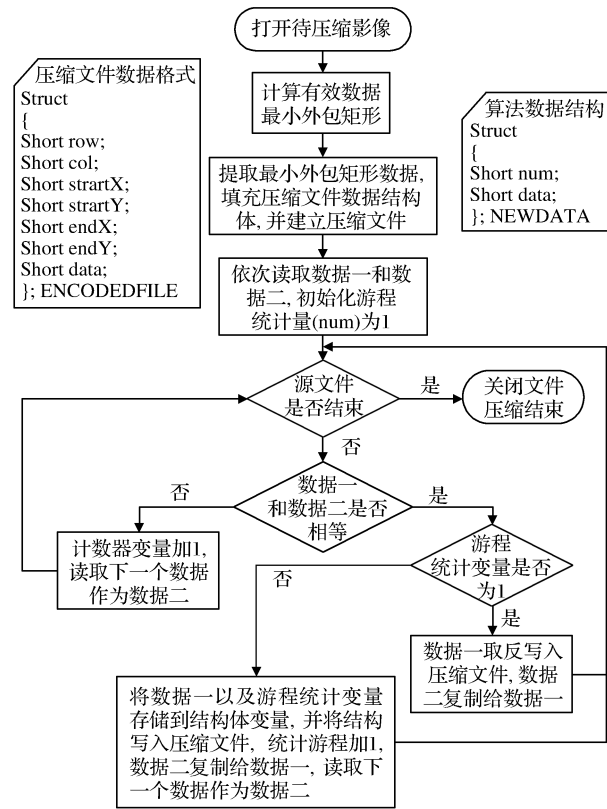


图 3 优化游程编码(ORLE)算法流程
Fig.3 Flow chart of ORLE method

通过优化游程编码将原始数据进行了大幅压缩,但仍然可能会在不同区域出现相同的行程以及相同的数据,从而产生数据冗余。如数据 A 在某区域出现 n 次,数据 B、C 也可重复 n 次,出现行程相同;而且这些数据 A、B、C 本身还可在其他区域再次出现,这样仍然存在大量冗余信息,在此基础上,进行数据概率的统计即哈夫曼编码压缩,充分实现了原始数据的压缩,对于优化游程编码中产生的中间数据,再进行哈夫曼编码后可随即删除,从而尽可能减小存储空间,相应的算法流程如图 4 所示。

3.3 基于“计算端元”的任务分配策略

王岩广总结了遥感图像处理可采用的并行模式大致有 3 种^[18]:管道并行、功能并行和数据并行。其中数据并行具有直观和天然的并行性特点。在进行遥感数据并行处理时,一个关键问题就是并行任务的分配。由于通常情况下,把各计算节点看成是

同等的角色,所以常采用均等的数据划分方式。但是这种看似均等的数据划分方式,在实际的遥感应应用中存在严重的不均等现象^[19],我们称其为“假均等”。如气溶胶遥感反演中对预处理后的影像进行数据划分^[16],如图 5 所示(其中亮点为待计算的有效数据点)。

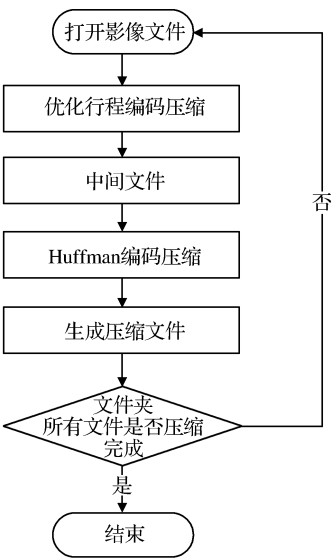


图 4 ORLE 与 Huffman 结合编码(ORLE-Huff)流程
Fig.4 Flow chart of ORLE-Huff encoded method

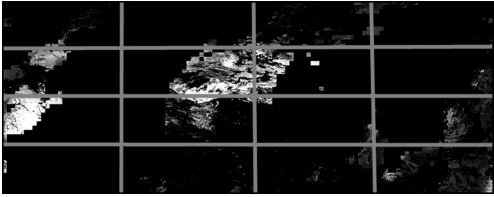


图 5 影像任务分块示意图

Fig.5 Task allocation schematic map of remote sensing image

按照 4×4 的规则格网划分,可以明显看出每一个分块中的有效数据点数(即计算量)不均等。从并行计算的关键路径的角度分析,这种传统的均等数据划分方式将造成负载不均衡现象,严重影响并行效率。

明冬萍等^[20]在深入分析计算模型和地学计算特点的基础上总结出了地学计算内涵,提出了一种“深度计算”——“主动计算”的空间数据计算模式,并强调在各种领域的空间数据的计算与分析中,基本都存在最小的计算分析单元,如景观学中的景观单元、土地学中的土地单元、大气中的气团和海洋中的水团等。通过研究分析可以发现,在地学计算领域,很多应用分析模型都是基于这种最小的计算分析单元,且对于相同类型的最小计算分析单元由于

其应用分析模型的确定性,其计算方法基本上也一致。所以在进行并行任务分配时,可以根据应用模型需求先提取出这种最小计算分析单元(也叫“计算端元”),剔除干扰对象,然后基于“计算端元”进行任务分配,保证各计算节点的负载均衡,同时可以有效避免应用模型迭代计算中不必要的全局性像元搜索所带来的效率下降。本文中气溶胶遥感反演为多幅影像单个像元的迭代计算,所以以像元为对象,提取“计算端元”,将多幅影像的同一位置像元作为一个任务“计算端元”,并根据处理节点的个数进行任务均等划分,实现细粒度的任务划分。其算法描述如表 1 所示。

表 1 基于“计算端元”的任务分配算法描述

Table 1 Algorithm of task allocation based on “computing end”

(1)提取“计算端元”,根据应用模型输入而定;

(2)基于“计算端元”的任务划分:

设节点个数为 n ,提取得到的“计算端元”个数为 H ,第 i 个节点分配到的“计算端元”个数为 h_i 。当 H 整除 n 时,任务划分公式为:

$$h_i = \frac{H}{n}$$

式(1)

当 H 不整除 n 时,任务划分公式为:

$$h_i = \begin{cases} \frac{H}{n} + 1, 0 \leq i < r \\ \frac{H}{n}, r \leq i \leq n \end{cases} \quad r = H \bmod n$$

式(2)

4 气溶胶光学厚度(AOD)遥感定量反演实验

将上述改进的游程编码遥感影像压缩方法和基于“计算端元”的任务分配方法应用于 AOD 网格平台(RSSN)的遥感反演,为了实验的可靠性,采用 2008 年 1 月份 NASA 的 MODIS 数据,计算一个月的 AOD 反演结果。同时从数据压缩率和计算时间效率两个方面对实验结果进行阐述。

4.1 影像压缩效率分析

为了验证比较上述算法的效率,实验时与目前几种比较常用的无损压缩算法及目前网络中应用最广泛的也是压缩率最高的无损算法之一,WinRAR (其压缩算法未公布)工具进行了比较测试。测试采用的样本数据如表 2 所示。通过对比分析 7 种压缩算法对两类不同的影像数据的压缩率的统计分析(表 3、图 7),可以发现,虽然商业的 WinRAR 具有最优的压缩效果,但是由于其商业机密,其压缩算法

并未对外公开。而本文提出的优化游程编码也具有较好的压缩率,在所有 7 种压缩算法中仅次于 WinRAR 压缩算法。显然可以看出,7 种压缩算法对 Georef.img 的压缩效率明显高于 MCD.img,结合前文影像的统计分析,可知,MCD.img 影像的数据分布有效数据较多,所以其压缩率没有 Georef.img 压缩效率高,符合实际情况。尽管如此,压缩算法对 MCD.img 影像还是具有一定的效果,其压缩率最低也有 50%左右,而优化游程编码和优化游程编码与 Huffman 结合的方法的压缩率达到了 19% 和 17%,具有明显的优势。

表 2 测试数据属性信息

Table 2 The information of test data

文件名称	文件数	文件大小/Gb	日期
Georef.img	37	35.66	2008-01
MCD.img	50	5.89	2008-01

表 3 7 种算法的压缩效率统计结果

Table 3 Statistical results of compression efficiency

压缩算法	压缩文件大小/Gb		压缩率/%	
	Georef	MCD	Georef	MCD
3-Tuples	1.37	2.08	3.84	35.31
LZW	1.23	1.94	3.45	32.94
Huffman	2.52	3.18	7.09	53.99
RLE	1.13	1.83	3.17	31.07
ORLE	0.85	1.12	2.38	19.01
ORLE-Huff	0.76	1.03	2.13	17.49
WinRAR	0.47	0.65	1.32	11.04

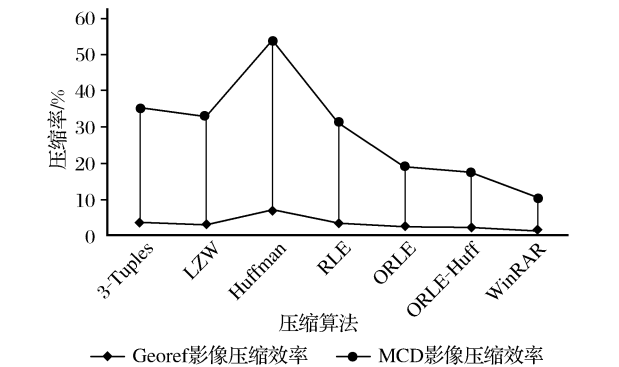


图 6 7 种算法对两种不同数据的压缩率比较

Fig.6 Thecompression efficiency comparison of 7 algorithms

4.2 计算时间效率分析

在气溶胶遥感反演计算中,对于每一天的数据,输入数据大小大约为 6.8~9.21 Gb。在 RSSN 网格平台上(根据平台中的 30 个节点进行任务划分),在

数据压缩前和传统均等分块方式情况下,其每天的 AOD 反演计算时间最少需要 4.29 h,最多需要 7.64 h,平均需要 6.39 h;而采用数据压缩方法和基于“计算端元”的任务分配方法情况下,用同样的数据进行计算,其计算时间最少仅需要 3 h,最多也只需要 4 h,平均需要 3.5 h,其效率提升将近 1 倍。

为了更进一步验证,还使用 PHPC 高性能服务器对同样的数据进行了处理,利用其中一个核心(两个至强四核 CPU,2.4 GHz,8 GB 内存)运行与 PC 机上相同的处理程序,结果最少需要 25.87 h,最大

则需要 38.17 h,平均达到 33.03 h。计算时间对比统计分析如表 4 和图 7 所示。

表 4 计算时间统计结果(取前 4 d 为例)

Table 4 Statistical results of calculation time

日期	数据量	时间消耗/h			
		单机	PHPC	改进前	改进后
0801	/Gb				
01	6.23	49.13	38.17	6.94	4.93
02	6.8	51.67	36.8	5.50	4.28
03	6.47	46.78	37.97	5.13	5.18
04	6.79	50.87	35.18	5.83	4.88

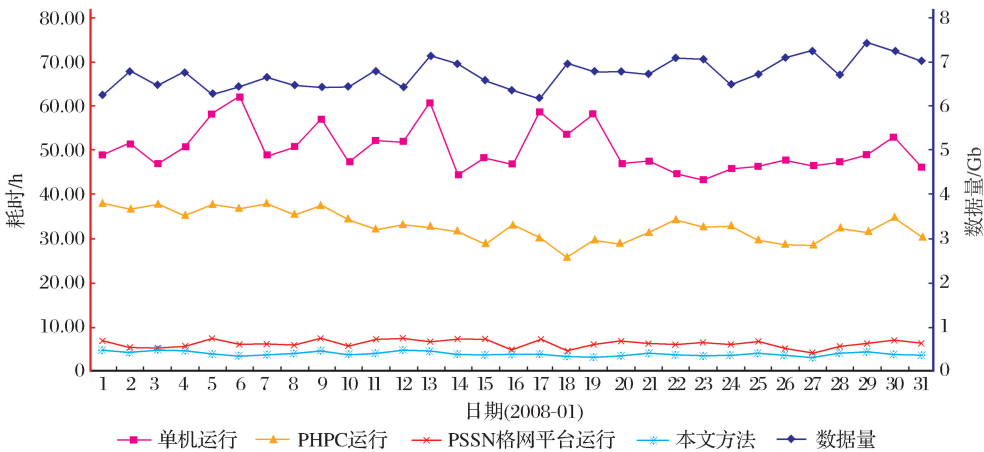


图 7 计算时间对比分析

Fig.7 Analysis of calculation time

经过对比分析,发现数据压缩方法和基于“计算端元”的任务分配方法能够有效提高分布式网络环境下气溶胶遥感反演的效率,提供一个有效解决分布式环境下大量遥感影像网络传输问题和任务分配不均等方法。同时,对整个反演流程中需要分布式计算的 AOD 反演迭代计算步骤(如图 8 所示的 AOD 反演,将整个反演计算任务根据节点数划分为 30 个

子任务)进行分析,从图中可以看出,基于“计算端元”的任务分配方法可以有效减少这个步骤的计算时间,进一步分析发现每一个节点的任务执行时间也基本相等,更进一步说明该任务分配方法的有效性。其任务执行时间不是完全一样,主要是因为每一个计算节点的性能存在差异,同时每一个“计算端元”的迭代计算次数也存在着差异,造成在计算任务

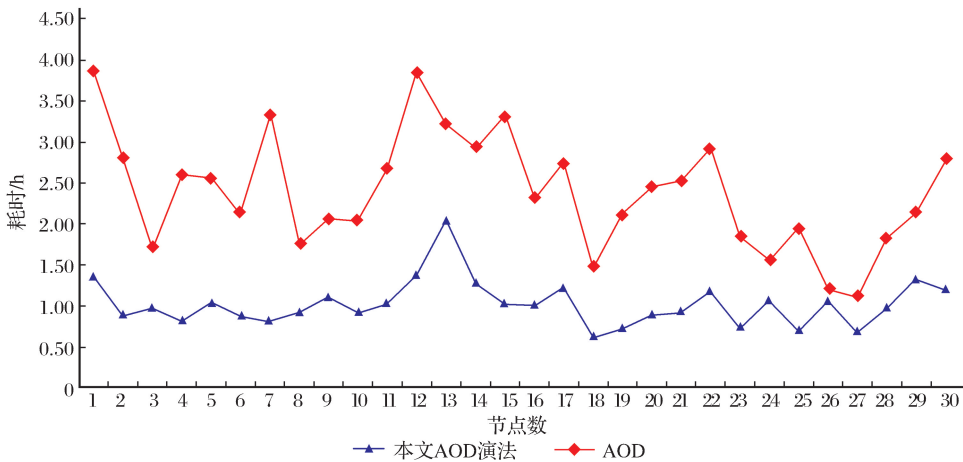


图 8 AOD 反演计算时间对比分析

Fig.8 Analysis of calculation time in AOD retrieval processing

均等的情况下计算时间的个别差异。

5 结 语

本文根据遥感影像数据处理中涉及网络传输的需求和遥感影像的统计特点分析,提出了优化游程编码和优化游程与 Huffman 结合的方法,并应用于遥感影像的压缩。实验结果表明,该方法能够取得很好的压缩比,能够有效解决分布式环境下的遥感影像传输问题。同时在网格环境下,对并行处理数据任务分配方式进行了研究,提出了基于“计算端元”的任务分配方式和实现原理,该方法能够有效地避免应用模型中不必要的全局性像元或者基元搜索带来的效率下降的弱点,通过剔除干扰对象,来保证任务分配均等和加快计算效率。在 RSSN 网格平台上用上述方法进行 AOD 反演实验,实验结果表明该方法能够显著提高总体反演效率,为加速遥感数据的处理提供了一种有效的手段,具有一定的实用价值。

参考文献 (References):

- [1] Xue Y, Guo H D, Wan W, *et al.* Overview and Preliminary Idea for the Building Digital Earth with Grid Computing Technology [J]. International Journal of Digital Earth, 2008, 1(2): 1-6.
- [2] Acharya A, Uysa M, Bennett R, *et al.* Tuning the Performance of I/O-intensive Parallel Applications [C]//Proceedings of the Fourth ACM Workshop on I/O in Parallel and Distributed Systems, Philadelphia, 1996, 27: 15-27.
- [3] NASA. Landsat 7 Science Data User's Handbook [M]. NASA, 2010.
- [4] Middleton A M, Furht B, Escalante A. Data-intensive Technologies for Cloud Computing [M]. New York: Handbook of Cloud Computing, Springer, 2010, 83-136.
- [5] Wan Wei. Study on Grid Enabled Aerosol Quantitative Remote Sensing Retrieval System [D]. Beijing: University of Chinese Academy of Sciences, 2008. [王伟. 基于网格的高性能气溶胶遥感定量反演研究 [D]. 北京: 中国科学院大学, 2008.]
- [6] Xue Y, Wang J Q, Wang Y, *et al.* Preliminary Study of Grid Computing for Remotely Sensed Information [J]. International Journal of Remote Sensing, 2005, 26(16): 3613-3630.
- [7] Kimura N, Latifi S. A Survey on Data Compression in Wireless Sensor Networks [C]//Information Technology: Coding and Computing, ITCC 2005. International Conference on 4-6 April 2005, 2: 8-13.
- [8] Liu Longli, Xue Yong, Zhang Jinzun, *et al.* Data Compression and the Application in High through Put Aerosol Retrieval [C]//19th Conference on Remote Sensing of China, Xi'an, September 21, 2014: 282-287. [刘龙历, 薛勇, 张敬尊, 等. 数据压缩及其在高性能气溶胶反演中的应用 [C]//中国遥感大会 19th, 西安, 2014: 282-287.]
- [9] Qiu Yubao, Hu Guangdao, Che Xuewen, *et al.* ECW Compress Technique in the Compression and Distribution Application of High Resolution Remote Sensed Image [J]. Remote Sensing Technology and Application, 2007, 22(6): 727-732. [邱玉宝, 胡光道, 车学文, 等. ECW 压缩技术在高分辨率遥感图像压缩与分发中的应用 [J]. 遥感技术与应用, 2007, 22(6): 727-732.]
- [10] Bian Yanshan, Zou Peng. Progressive Network Transmission of Remote Sensing Images based on EZW Algorithm [J]. Computer Engineering and Application, 2012, 48(20): 6-10. [卞燕山, 邹鹏. 基于 EZW 算法的遥感影像网络渐进传输 [J]. 计算机工程与应用, 2012, 48(20): 6-10.]
- [11] Wang H Q, Babacan S D, Khalid S. Lossless Hyperspectral-Image Compression Using Context-based Conditional Average [J]. IEEE Transactions on Geoscience and Remote Sensing, 2007, 45(12): 4187-4193.
- [12] Liu Xingke, Chen Ke, Yu Xiaoguang. Huffman Coding and Applications in Compression for Vector Maps [J]. Journal of Geomatics Science and Technology, 2014, 31(1): 89-92. [刘兴科, 陈轲, 于晓光. Huffman 编码在矢量地图压缩中的应用 [J]. 测绘科学技术学报, 2014, 31(1): 89-93.]
- [13] Li Yunjin, Liang Peng. New Lossless Compression Approaches for Raster Map [J]. Science of Surveying and Mapping, 2014, 39(9): 24-29. [李云锦, 梁鹏. 栅格电子地图的无损压缩新方法 [J]. 测绘科学, 2014, 39(9): 24-29.]
- [14] Chen Xin, Luo Siwei. Research of Grid Computing Task Assignment Algorithm based on Ant Algorithm [J]. Computer Technology and Development, 2006, 16(3): 98-100. [陈歆, 罗四维. 基于蚂蚁算法的网格任务分配算法研究 [J]. 计算机技术与发展, 2006, 16(3): 98-100.]
- [15] Liu Huiting, Jiang Xiaotao, Chen Jian. Research of Grid Task Scheduling Strategy based on Genetic Algorithm [J]. Computer Technology and Development, 2012, 22(4): 69-72. [刘慧婷, 姜晓涛, 陈健. 基于遗传算法的网格任务调度方法研究 [J]. 计算机技术与发展, 2012, 22(4): 69-72.]
- [16] Wei Dong, Wu Liangjie, Zuo Dan, *et al.* Grid Task Schedule based on Hybrid Ant Colony Algorithm [J]. Computer Engineering, 2010, 36(3): 215-217. [魏东, 吴良杰, 佐丹, 等. 基于混合蚁群算法的网格任务调度 [J]. 计算机工程, 2010, 36(3): 215-217.]
- [17] Xue Y, Chen Z Q, Xu H, *et al.* A High Throughput Geocomputing System for Remote Sensing Quantitative Retrieval and a Case Study [J]. International Journal of Applied Earth Observation and Geoinformation, 2011, 13(3): 902-911.
- [18] Wang Yanguang. The Design and Preliminary Realization of Remote Sensing Applications Grid Platform InterCondor [D]. Beijing: University of Chinese Academy of Sciences, 2005. [王岩广. 遥感应用网格平台 InterCondor 设计与初步实现 [D]. 北京: 中国科学院大学, 2005.]
- [19] Shen Zhanfeng, Luo Jiancheng, Chen Qinxiao, *et al.* High-efficiency Remotely Sensed Image Parallel Processing Method Study based on MPI [J]. Journal of Image and Graphics, 2007,

12(12):2132-2136.[沈占峰,骆剑承,陈秋晓,等.基于 MPI 的遥感影像高效能并行处理方法研究[J].中国图象图形学报,2007,12(12):2132-2136.]

[20] Ming Dongping, Luo Jiancheng, Zhou Chenghu, *et al.* The Preliminary Study of Spatial Data Computation Pattern based on Unit[J].Geo-information Science,2006,2(8):84-89.[明冬萍,骆剑承,周成虎,等.空间数据计算模型分析与应用[J].地球信息科学,2006,2(8):84-89.]

Remote Sensing Information Service Grid Node and the Research of Data Compression and Task Allocation

Liu Longli^{1,2},Xue Yong¹,Guang Jie¹,Liu Jia^{1,2}

(1.*Key Laboratory of Digital Earth Science,Institute of Remote Sensing and Digital Earth , Chinese Academy of Sciences,Beijing 100094,China ;*
2.*University of the Chinese Academy of Sciences,Beijing 100049,China*)

Abstract: Grid computing as a distributed computing in the field of scientific computing provides a powerful computation force for Earth observation data processing,based on the analysis of data transmission and load balance in grid computing,the paper puts forward a data compression method based on Run Length Encoded(RLE) and Huffman Encoded and a task allocation strategy considering the "computing end member". This combined method can effectively solve the problems of network data transmission by irregular extraction and the task load balancing. On Remote Sensing information service Grid Node(RSSN) computing platforms,we take the Aerosol Optical Depth(AOD) retrieval in China area of 1 km resolution for example to validate and analyze the effectiveness of this method.

Key words: Distributed computation;Grid computing; Aerosol retrieval; RSSN; Data compression; Task allocation