

文章编号:1001-9081(2006)03-0645-02

## 基于 MPI 的并行小波聚类算法在曙光 TC1700 上的实现

赵光嗣,李伟生

(北京交通大学 计算机与信息技术学院,北京 100044)

(zguangsi@163.com)

**摘 要:**针对我们在第 22 届全国数据库学术会议中提出的基于 MPI 的改进小波聚类算法,利用消息传递 MPI 机制在曙光 TC1700 上实现了并行聚类。对编程涉及到的主要问题,寻找和标记网格连通区域进行了重点说明,并对程序实现使用的技巧给出了解释。最后对两组数据分别在 5 节点和 9 节点情况下聚类结果进行了比较,实验结果表明该算法是高效并且可行的。

**关键词:**小波聚类;并行;MPI;集群系统

**中图分类号:**TP391.41 **文献标识码:**A

## Implementation of MPI-based parallel wavecluster algorithm on Shuguang TC1700

ZHAO Guang-si, LI Wei-sheng

(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

**Abstract:** A MPI-based parallel wavecluster algorithm was used to achieve parallel clustering by message-passing model on TC1700. The main problem involved in programming, finding and labling the connected components was explained, especially on the skill of programme making. The experiment results on two data set by 5 nodes and 9 nodes respectively show that our algorithm is efficient and effective.

**Key words:** wavecluster; parallel; MPI; cluster system;

### 0 引言

串行聚类算法在统计和数据库领域得到大量的研究和应用,早期有 k-means, k-medoids, 以后又有面向大规模数据库系统的 BIRCH<sup>[2]</sup> 算法、处理非数值属性聚类的 CACTUS<sup>[3]</sup> 算法、处理空间数据的 STING<sup>[4]</sup> 算法,子空间聚类算法 ENCLUS<sup>[5]</sup> 等。

但早期的并行聚类算法并不完全适合集群系统。这些并行聚类算法是基于共享存储系统等低通信延迟系统的,在设计算法时没有把通信代价作为重点来考虑。因而这些算法往往采用计算并行的设计思想,把全局聚类模式分布在各个计算节点中,以便可以并行地进行聚类模式匹配。在这一过程中,匹配过程的启动及结果的收集都需要通信,也就是说,处理的每一步至少需要两次通信,而作为一种低成本的通用并行系统,集群系统的通信能力相对于节点的计算能力是一个瓶颈。因为集群系统的通信机制一般是 MPI、PVM 等,基于 TCP/IP 和 UDP 通信协议的消息传递机制,所以通信延迟非常大,一般达到 ms 级。因此在集群环境中开发并行算法时,如何有效地减少通信次数是提高算法效率的重点。所以采用计算并行的并行聚类算法不适合集群系统<sup>[6]</sup>。

在文献[1]中,我们提出了一种新的基于 MPI 的并行小波聚类算法,并在曙光 TC1700 并行机上实现。实验结果表明我们的算法是可行的高效算法。

### 1 MPI 的特性

由于集群系统的发展,近年来出现了多个并行程序开发

及运行系统,比如 PVM, MPI, ExPress, Linda, P4 等。这些系统的适应平台非常广,应用程序在这些系统上的可移植性较好,特别是 PVM 和 MPI。

MPI 的特点概括来说有三方面:较高的通信性能、较好的程序可移植性和强大的功能。具体来说,包括以下几个方面:

**通用性** MPI 是可移植的标准平台,其通讯单元包含上、下文和组的信息,以保证消息传递的安全性。

**点对点通讯** MPI 能有效地管理消息缓存区,具有结构化缓存、扩充数据类型及异构性, MPI 异步执行时能保护用户的其他软件不受影响,能实现完全的异步通信。立即发送与接受可完全与计算覆盖进行。

**数据汇集方式** 具有内定和用户自定义的数据汇集操作方式,可对大量数据进行整体传输,可直接或依托拓扑结构定义子组。

**MPI 的实现方式多样化** 同一编程界面可有多种开发工具,具有面向应用的信息传递拓扑结构:内定支持网格和图拓扑结构。

**良好的操作环境** 具有差错控制功能; MPI2 在 I/O、主动消息、进程启动、动态进程控制等方面有进一步增强; MPI 还有在 X 窗口下运行的 MPICH 版本,用户可以在 X 窗口下用 Upshot 或 Nupshot 直观地考察 MPI 应用程序运行过程中各处理器之间的同步、计算、消息发送、接收等情况,从而为程序的修改提供依据。

收稿日期:2005-09-15 修订日期:2005-12-02

作者简介:赵光嗣(1982-),男,山西太原人,硕士研究生,主要研究方向:并行算法、网络数据库技术; 李伟生(1945-),男,江苏金坛人,教授,主要研究方向:算法的设计与分析、网络数据库技术。

## 2 算法主要思想

我们在曙光 TC1700 集群系统上通过主/从模式实现并行小波聚类算法如图 1。我们设计的并行小波聚类算法,做了一些改进使它更适合并行。首先,我们使用局部查找表代替全局查找表,因为在对特征空间进行小波变换和将聚类结果转回特征空间都是在从节点完成的。其次,我们利用集群系统完成了查找和标记连通区域。现有的寻找和标记连通区域的并行算法并不适用于集群系统,通信是集群系统的主要瓶颈,但现在的并行算法都存在需要大量反复通信的缺点<sup>[7]</sup>。而我们设计的寻找和标记连通元素的并行算法,大大地减少了通信次数,通过仅有的几次通信实现了全局的寻找和标记连通区域。

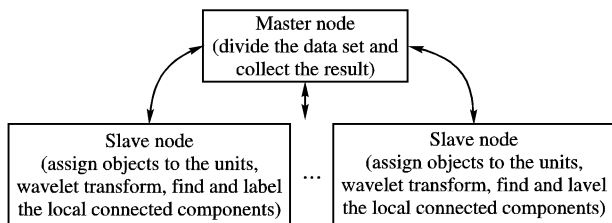


图1 程序结构

主要步骤如下:

Step1 量化特征空间,将数据划分到不同的网格

- 1) 主节点读取原始数据;
- 2) 主节点将数据集划分到不同的从节点上;
- 3) 从节点将收到的数据划分到不同的网格中;
- 4) 主节点将所有从节点的结果收集起来,确定最后所有数据的网格划分。

Step2 对特征空间进行小波变换

- 1) 主节点将网格划分到不同的从节点上;
- 2) 每个从节点对分到的局部网格进行小波变换,并建立局部查找表。

Step3 寻找连通区域

- 1) 每个从节点寻找局部连通区域;
- 2) 从节点将边界网格的标记发送回主节点;
- 3) 主节点将等价的网格标记合并,并生成等价标记表;
- 4) 主节点将等价标记表传给每个从节点;
- 5) 从节点根据等价标记表对连通区域重新标记。

Step4 将原数据进行聚类

- 1) 从节点根据局部查找表将网格转回特征空间;
- 2) 从节点将聚类好的结果传回主节点。

Step5 主节点将最后的结果输出。

## 3 重点技术问题

### 3.1 网络连通的寻找和标记

首先我们介绍一下网络连通,如图 2 表示。

0	1	1
0	1	0
0	0	1

图2 网络连通情况

对于中间网格,在四连通情况下,它与正上方都是 1,表明这两个网格连通要被标记为同一类,对于右下角的网格,由于它的正上方和左方都为 0,所以它与图 1 中的网格都不连

通,需要赋予新的标记。

对网格进行扫描是从左向右、从上向下进行的(起点在网格矩阵的左上方)。要标记当前正被扫描的网格需要检查它与在它之前扫描到的若干个近邻网格的连通性。这里有几种不同的情况要分别考虑。例如它与两个或多个目标相连通,则可以认为这些目标实际是同一个,并把它们连接起来。如果发现了值为 0 的网格到一个孤立的值为 1 的网格的过渡,就赋一个新的目标标记。

我们考虑四连通的情况。可按如下方法对网格矩阵进行标记。以上面的方式对网格逐个扫描,假如当前网格的值是 0,就移到下一个扫描位置。假如当前网格的值是 1,检查它左边和上边的两个近邻网格(根据所采用的扫描次序,当扫描到达当前网格时这两个近邻网格已被处理过了)。这两个网格值和标记的组合中有四种情况要考虑:1) 它们的值都是 0,给当前网格一个新的标记(根据已有信息,知道目前这是该连通区域第一次被扫描到);2) 它们中间只有一个值为 1,把该网格的标记赋给当前网格;3) 它们的值都为 1 且具有相同的标记,就将该标记赋给当前像素;4) 它们的值都为 1 但具有不同的标记,就将其中的一个标记赋给当前网格并做个记号表明这两个标记等价(两个近邻网格通过当前网格而连通)。在扫描终结时所有值为 1 的网格都已有了标记但有些标记可能是等价的。此时所需做的就是将所有等价的标记对合并,对各个等价对赋一个不同的标记(我们取所有等价标记中的最小值)。然后第 2 次扫描网格,将每个标记用它所在等价对的标记代替。

在我们提出的并行算法中,当处理的目标数据量较大时,将上述网格划分为连续的几块;然后各个从节点同时对各自的一块执行上述标记,标记完成后,从节点将标记好的区域边界传回主节点;主节点将边界上连通的不同标记进行标记合并,生成全局等价标记表,然后将全局等价标记表传到各个从节点上;从节点根据等价标记表对其所处理的块进行重新标记(不需要重新寻找,只需标记替换即可)。从节点根据区域查找表将网格转回特征空间,最后从节点将聚类好的结果传回主节点,主节点搜集结果并输出。

### 3.2 网络等价合并和标记时所用技巧

在对等价网格合并时建立一个二维合并表,当两个标记的网格连通时以这两个标记值较小的作为横坐标,较大的作为纵坐标,同时将二维表相应位置置为 1,这样,对多个相同标记合并时,不需要多次判断当前标记是否已经合并,直接赋值 1 即可,减少了运算量。当对所有边界合并完后,遍历二维合并表,将所有等价关系合并,建立等价关系表,相同等价关系以标记最小的作为最后合并的标记。最后将等价关系表传给各个从节点进行重新标记。

每个节点扫描进行标记时,为了避免当前节点标记和其他节点标记发生重复,影响最后聚类结果,所以,当前区域的标识是当前进程号乘以系数  $c$  ( $c$  为每个节点内可能的最大非连通区域数),每发现一个新区域,将当前标识标记,并将当前标记加 1。这样就可以使各个节点标记的区域标号各不相同。

## 4 试验结果

我们使用了两组数据对结果进行测试。数据 1 区域划分

(下转第 654 页)

我们一共定义了 87 条未登录词词性猜测规则。

### 3 实验结果

为了测试上述基于决策树的词性标注算法的效果,我们通过半自动的方式对测试语料进行了词性标注,以便与经过算法标注的结果进行对比与分析。

首先对原始的未标注的测试语料通过查找词库进行初始标注,发现其中有 2752 个未登录词。然后再使用未登录词词性猜测规则,猜测出未登录词可能的词性,发现在所有未登录词中,词性猜测正确的有 2546 个(所谓猜测正确,是指猜测出的词性组合中包含其正确的词性),其正确率为 92.51%。其中猜测错误的主要是那些词性比较复杂的词,如 nominal,由于 nominally 在词库中存在,因而依据规则,nominal 加上 ly 后,其作为一个副词在词库中存在,因而判定 nominal 的词性为 JJ,而事实上 nominal 还可作 NN。

未登录词经过处理后,测试语料中的所有词均被标注上其可能的词性,现在再将测试语料通过经学习语料学习生成的决策树进行词性消歧,并最终生成每个词和符号只有一个词性标记的语料,发现共有 978 个标注不正确,标注正确率为 95.11%。

### 4 结语

显然 95.11% 的标注正确率与其他一些标注系统大多能达到 97% (最高近 99%) 以上的高正确率相比,还有不小的差距,然而由于我们既缺少一个大规模的标注语料库用于标注学习及决策树的修剪,且由于词库规模有限,有相当一部分未登录词只能通过词性猜测规则来猜测词性,能达到 95.11% 的正确率应该说还算是一个不错的结果。

#### 参考文献:

- [1] GREENE BB, RUBIN GM. Automatic Grammatical Tagging of English[ R]. Department of Linguistics, Brown University, 1971.
- [2] OOSTDIJK N. Corpus Linguistic and the automatic analysis of Eng-

lish[ Z]. Rodopi, Amsterdam, 1991.

- [3] KARLSSON F, VOUTILAINEN A, HEIKKILA J, *et al.* Constraint Grammar. A Language-Independent System for Parsing Unrestricted Text[ M]. Berlin, New Youk: Mouton de Gruyter, 1995.
- [4] GARSIDE R, LEECH G, SAMPSON G. The Computational Analysis of English[ M]. London and New York: Longman, 1987.
- [5] SCHMID H. Part-of-speech tagging with neural networks[ A]. Proceedings of 15th International Conference on Computational Linguistics[ C], 1994.
- [6] ROSEFELD R. Adaptive Statistical Language Modeling: A Maximum Entropy Approach[ D]. School of Computer Science, Carnegie Mellon University, 1994.
- [7] SAMUELSSON C, TAPANAINEN P, VOUTILAINEN A. Inducing Constraint Grammars[ A]. Proceedings of the 3rd International Colloquium on Grammatical Inference[ C], 1996.
- [8] DAELEMANS W, ZAVREL J, BERCK P, *et al.* MTB: A Memory-Based Part-of-speech Tagger Generator[ A]. Proceedings of 4th Workshop on Very Large Corpora[ C], 1996.
- [9] MARQUEZ L, RODRIGUEZ H. Part-of-Speech Tagging Using Decision Trees[ A]. Proceedings of the 10th European Conference on Machine Learning, ECML[ C]. Chemnitz, Germany, 1998.25 - 36.
- [10] QUINLAN JR. C4.5: Programs for Machine Learning[ M]. San Mateo, CA: Morgan Kaufmann, 1993.
- [11] MARCUS M, SANTORINI B, MARCINKIEWICZ MA. Building a large annotated corpus of English: The Penn Treebank[ J]. Computational Linguistics, 1993, 19(2): 313 - 330.
- [12] MITCHELL T. Machine Learning[ M]. McGraw Hill, 1997.
- [13] ERIC B. Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging[ J]. Computational Linguistics, 1995, 21(4): 543 - 565 .
- [14] VASILAKOPOULOS A. Improved Unknown Word Guessing by Decision Tree Induction for POS Tagging with TBL[ A]. Proceedings of CLUK 2003[ C], 2003.

(上接第 646 页)

单一,数据量较少,如图 3(a)。数据 2 区域划分较多,数据量比较大,如图 3(b)。

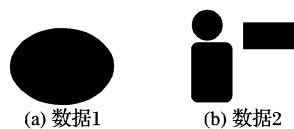


图 3 测试数据

在 5 节点(1 个主节点 4 个从节点)和 9 节点(1 个主节点 8 个从节点)情况下分别对两组数据并行处理。试验结果如表 1 ~ 表 2。

表 1 5 节点加速比

数据 区域	步骤			
	Step1	Step2	Step3	Step4
1	2.6	4.0	1.8	2.6
2	2.25	4.0	1.6	2.25

表 2 9 节点加速比

数据 区域	步骤			
	Step1	Step2	Step3	Step4
1	5.2	8.0	2.7	5.2
2	4.5	8.0	2.3	4.5

通过试验结果我们可以看出,在数据量较大、计算比较复杂的情况下,使用我们的基于 MPI 的并行小波聚类算法获得了较好的加速比,表明我们的算法具有较好的性能。

#### 参考文献:

- [1] LIANG H, ZHAO G-S, LI W-S. A New MPI-Based Parallel Wave-Cluster Algorithm[ A]. 第 22 届全国数据库学术会议 (NDBC2005)[ C], 2005.
- [2] ZHANG T, RAMAKRISHNAN R, LIVNY M. BIRCH: An Efficient Data Clustering Method for Very Large DataBases[ A]. ACM SIGMOD International Conference on Management of Data[ C], 1996.
- [3] GANTI V, GEHRKE J, RAMAKRISHNAN R. CACTUS - clustering Categorical Data Using Summaries[ A]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[ C], 1999.
- [4] WANG W, YANG J, MUNTZ R. STING: A Statistical Information Grid Approach to Spatial Data Mining[ A]. 23rd VLDB Conference [ C], 1997.
- [5] CHENG CH, FU AW, ZHANG Y. Entropy-based Subspace Clustering for Mining Numerical Data[ A]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[ C], 1999.
- [6] ZHOU B, SHEN JY, PENG QK. Parallel Clustering Algorithm for PCs Cluster[ J]. Computer Engineering, 2004, 30(4).
- [7] WANG K-B, CHIA T-L, CHEN Z, *et al.* Parallel Execution of a Connected component labeling Operation on a Linear Array Architecture[ J]. Journal of Information Science and Engineering, 2003, 19(2): 353 - 370.