

贸易地图标签绘制的并行解决方案与实现

王树伟 赵天宇 纪文成 李维铭

(北京工业大学计算机学院 北京 100124)

摘要 就绘制气泡图中经常遇到的图元标注重叠、遮挡问题,结合贸易地图生成软件中绘制产品标签的具体实现,设计了基于分治思想的并行处理方案。该并行方案将绘图区进行分区,并行计算各个分区数据,每个分区中采用位图搜索匹配算法对标签进行调整,以获得高质量的贸易地图。系统测试结果表明,该方案解决了产品标签绘制的重叠、遮挡问题,改善了运行时间长、系统性能较低的状况。

关键词 贸易地图,并行处理,气泡图,搜索匹配算法

中图法分类号 TP311 文献标识码 A

Parallel Processing Solution and Implementation of Trade Map Label Drawing

WANG Shu-wei ZHAO Tian-yu JI Wen-cheng LI Wei-ming

(College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China)

Abstract In view of the problem of diagram elements annotation overlaps and mask in drawing bubble charts and relate with the specific implementation of label drawing in the trade map software, this article had designed the parallel processing method which is based on divide and conquer. In this parallel processing method the chart space was divided to several regions and the data in every region was parallely computed. Meanwhile, used the bitmap search matching algorithm to solve this problem and get high quality trade map. The system running result indicated that this method solved the label overlaps and mask problem, shorten system running time and improved system performance.

Keywords Trade map, Parallel processing, Bubble charts, Search matching algorithm

1 引言

贸易地图从市场供给与需求相结合的角度对贸易市场状况进行分析,并把供给与需求两种因素放在图表的同一坐标系内表现出来,使抽象的贸易和市场分析理论变得直观易懂。

贸易地图是一种描述复杂贸易数据的图表,为了自动批量生成高质量的贸易地图,其关键技术之一是解决气泡图中气泡标签和其它图元相互重叠、遮挡的问题。为解决该问题一般采用回溯算法调整气泡标签,但是当气泡图中气泡数量较多时运算量加大导致运算时间很长,影响系统性能。本文提出基于分治思想的并行解决方案,在保证贸易地图质量的同时缩短计算时间,提高系统性能。

2 方案设计背景

2.1 贸易地图生成及需解决的关键问题

贸易地图生成软件批量自动生成贸易地图的生成过程如图 1 所示。

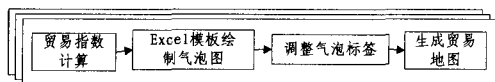


图 1 贸易地图批量自动生成示意图

Excel 模板生成的气泡图最初采用 Excel VBA 编程对气泡标签调整,由于 VBA 接口功能限制导致调整效果很差,重叠、遮挡严重(如图 2 所示),因此气泡图生成后要通过人工手动调整标签获得高质量无遮挡的贸易地图,当生成大批量贸易地图时工作量非常大,很难由人工完成。

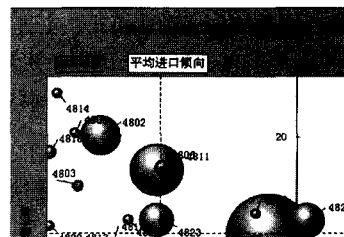


图 2 Excel VBA 调整结果图示

2.2 气泡图标签无遮挡放置方案

为了解决气泡图中标签无遮挡放置问题,核心思想是将贸易地图绘图区划分等大小的若干网格,其中每个空白格(格中不包含任何其他图元)的大小仅可以放置一个标签,且标签一旦放入,则该格不允许再放置其他标签,同时要求在放置时标签和标签对应的气泡尽可能地近,以确保指示清楚、美观,这样就可以避免标签的遮挡、重叠。基于上述思想由 Ex-

到稿日期:2008-07-10 本文受北京市科委项目:中国贸易业绩评价体系的研究与建立(课题编号:Z0004008040211)资助。

王树伟(1978-),男,硕士生,研究方向为并行计算和数据仓库, E-mail: wsw132@emails.bjut.edu.cn; 赵天宇(1987-),男,本科生; 纪文成(1984-),男,硕士生,研究方向为 J2EE 和软件工程; 李维铭(1962-),男,高级工程师,主要研究方向为软件工程。

cel 调用 C 语言编写的自动调整程序,解决标签的重叠、遮挡问题,处理过程如图 3 所示。

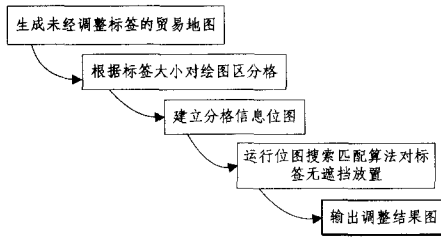


图 3 标签无遮挡放置的处理过程

该方案实施的结果表明核心思想正确,解决了气泡图中标签的重叠、遮挡问题。但是当绘图区中气泡的个数很多时,运行搜索匹配算法回溯计算量非常大,运行时间长,系统性能较低。

2.3 并行方案设计

为了解决系统性能问题,采用分治思想将贸易地图绘图区进行分区,每个分区内的气泡标签位置调整任务分配给一个计算节点进行独立计算,这样多个节点并行运行标签位置调整算法,最后再将计算结果合并输出。处理过程如图 4 所示。

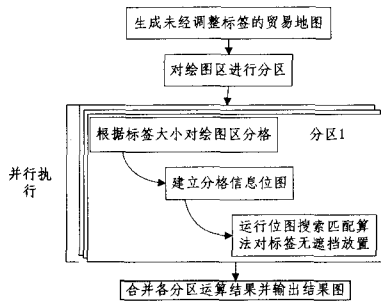


图 4 标签无遮挡放置的并行处理过程

贸易地图绘图区中气泡标签无遮挡分布并行程序设计和实施采用基于消息传递的主从设计模式。主进程负责分派任务、接收合并子问题的解。每个从进程负责解决一个分区的子问题。主进程和各从进程之间有通信,而各从进程彼此没有通信。

3 气泡标签无遮挡分布策略及核心算法

3.1 分区策略

绘图区分区状况的好坏将直接影响到整个程序性能的优劣和能否找到气泡标签无遮挡分布可行解。因此,绘图区分区时要综合考虑以下 3 个因素。

(1) 分区的大小

贸易地图中每个气泡都有一定的大小,最大气泡约占绘图区的 1/16,因此每个分区的大小不应小于整个绘图区的 1/16。

(2) 分区的形状

分区的形状不应出现长宽相差悬殊的分区。对于一个待分区域,算法尝试将其分为两个子区域时,尽量使子区域趋向于正方形,以增加可能解个数。

(3) 分区内气泡的数量

分区的目的是让问题的规模变小,而决定问题规模的指

标是区内的气泡的数量。所以尽量使各分区的气泡数量均匀,以保证各计算节点的负载均衡。

3.2 分区的网格划分

对每个分区进行网格划分时,划分网格的大小以一个标签的大小为基准,每个网格中至多放一个标签,如图 5 所示。

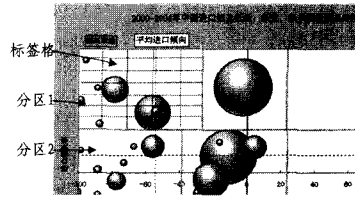


图 5 分区 1 内分格示意图

3.3 用位图表示网格信息

放置标签的搜索算法都是在一张“位图”上进行的。位图是一系列的整数,每个数据信息对应一个网格的标签放置情况,位图每个整数对应的二进制的最后一位表示该网格是否被气泡、坐标轴这些不可移动的对象占用,若未被占用则最后一位是 1。二进制的其余每一位表示该网格能否作为分区中气泡标签的一个可选放置位置。举例来说,某个区域被分成了 7×9 的 63 个网格,区内有 5 个气泡的标签需要调整。将整个位图映射成一个 7×9 的二维数组(行主序,起始下标为 0),5 个气泡按一定的规则编号为 1-5,若有 $\text{bitmap}(3,5) = 39$ (二进制数为 100111),其意义是:第 4 行第 6 列的网格可以作为气泡 1,2,5 的标签的放入可选位置,如图 6 所示。

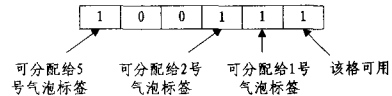


图 6 位图示意图

3.4 核心算法

主进程把分区信息发送给从进程,由进程执行标签调整算法。算法首先根据标签的大小将该分区划分为若干等大小的网格,然后利用相应的搜索算法,将标签放入合适的网格中。从进程算法的流程图如图 7 所示。

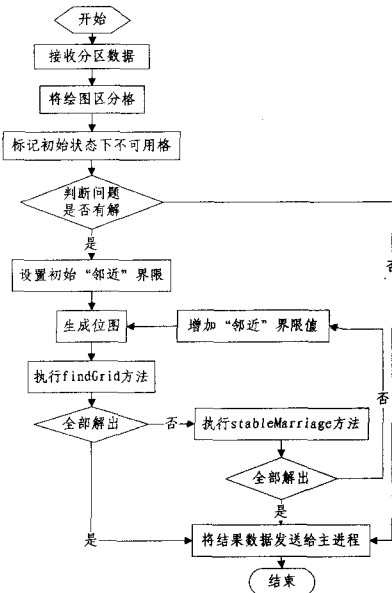


图 7 从节点算法流程图

(1)判定某个气泡是否在该分区内

算法中采用的是圆心判定法:只要气泡的圆心坐标在该分区内,就认为该气泡在分区内。

(2)建立初始状态位图

建立初始状态位图的目的是找出分区中没有被图元对象占用的所有网格。

有3种网格被占用的情况:

A)被圆心在该区内的气泡占用;

B)被圆心不在该分区内但有部分被划分到该分区的气泡占用;

C)被坐标轴或刻度线占用。

通过对以上3种情况的筛选,可以得到所有能放标签的“空白”网格。

(3)判断问题是否有解

得到所有“空白”网格的信息后,可以直接判断问题是否有解。如果给定的分区和其中待调整的气泡至少存在一种标签放置方法,则称该问题有解。

如果“空白”网格的数目小于分区内气泡数,就判定问题无解。当出现某个分区无解的情况时,需要对分区进行修改、合并,然后重新计算。

(4)“邻近”区域界限

每个气泡的标签都应在气泡的附近标注,如果“邻近”距离选得比较小,那么得到的解的质量比较高,所以在求解过程中,采用的是逐渐增大邻近区域界限的方法来寻找可行解,直到找到为止。

(5)生成位图

前面提到位向量的每一位(最低位除外)表示能否放入某个气泡的标签,所以对每个“空白”的网格,只要判断其是否在某个气泡的邻近区域内,就能确定整个位向量的值。因此,每个邻近界限的取值都唯一确定了一张位图,搜索算法的搜索过程都是基于该位图的。

(6)搜索匹配算法

在某个分区位图中,如何选择合适的网格直接影响到最终标签的放置效果,设计搜索算法时需要考虑以下几个因素:

1)网格优先级:若一个气泡的标签有多个“空白”网格可选,但这些网格与气泡的距离不一定相等。如果邻近区域界限比较小,这时把所有网格可以均等看待,但如果邻近区域界限已经比较大,该气泡的邻近区域内所有可以放标签的网格应该设定优先级。

2)网格竞争:一个气泡的标签从它能放置的若干网格中随机选取一个,有可能造成其它气泡的标签无法放置。比如气泡A的标签可以放在网格1,2,3中,气泡B的标签只能放在网格2中。如果气泡A选中网格2将会造成气泡B的标签无法放置。

3)无竞争网格:气泡标签间不存在竞争同一网格的情况:比如标签A可以放在网格1中,而其它所有标签可用网格不包括网格1,这样A就可以独立选取。

考虑以上因素,设计和采用了两种搜索算法如下:

第一种:findGrid 搜索

findGrid 搜索是处理无竞争网格的标签放置算法,为了标签指示美观、清楚,同时考虑标签放置网格和圆心的距离以及各标签间的距离。

要寻找只能作为一个气泡的标签放置位置的网格,只要找出位向量中除最后一位以外只有一个1的即可,也就是位向量值减1后恰好是2的n次幂,同时可知n就是相应的气泡序号。通过扫描所有网格的位图向量,就可以找到有哪些气泡的标签可以放进哪个网格中,然后选取气泡的标签放入相应网格中。

在确定将某个气泡的标签放入后,需要对整个位图进行更新,以避免重复放入其它标签。重复该搜索过程直到再也找不到满足条件的网格为止。

findGrid 搜索的流程图如图8所示。

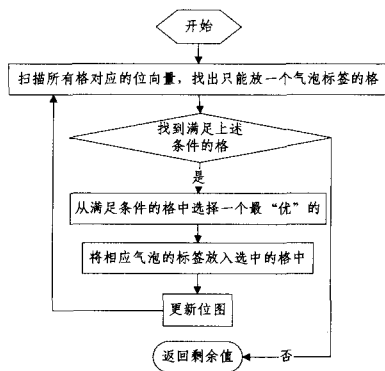


图8 findGrid 搜索的流程图

第二种:stableMarriage 搜索

stableMarriage 搜索用来处理网格竞争的情况。流程图如图9所示,该搜索算法的理论基础是匹配问题中的“稳定婚姻”问题,稳定婚姻问题是由两个美国数学家 Gale 和 Shapley 提出的匹配问题,有n位男士和n位女士,每位男士按照自己喜欢程度从高到低依次向每位女士主动求婚直到有一个女士接受他,对于每个女士,如果当前向她求婚的男士比她现有的配偶好则抛弃当前配偶,否则不予理睬,循环往复直到所有人都有配偶。算法一定得到稳定匹配,并且时间复杂度是 $O(n^2)$ 。

把稳定婚姻问题向标签放置问题进行映射:首先,将标签映射为男士,网格映射为女士。将原问题中的“n位男士和n位女士”改为“m个标签和n个网格(m≤n时才继续)”。

其次,将婚姻问题中的“每人对每个异性喜好度”进行映射,即要实现标签对网格的“喜好程度”。标签在选择网格时的依据是网格距气泡的距离。越小的距离数值对应越大的喜好程度,因此,增加一个网格与气泡距离权值。

这样,就完成了从稳定婚姻问题向标签放置问题的映射,修改后的算法的时间复杂度为 $O(n * m)$ 。

根据以上思想,稳定婚姻搜索过程具体设计如下:首先建立一张二维表,表的行、列分别表示尚未放置的标签和当前可用网格,如果某个气泡的标签能放在某个网格中,则将表中相应位置的值为该网格与相应气泡圆心的距离,表中其余位置用特殊值填充,表示相应标签不可放入相应网格中,整个搜索过程都是依据这张表的。图9给出了整个过程的流程图。

(7)合并计算结果

各从进程计算完成后,将每个分区的气泡标签匹配结果(坐标值)发送到主进程,主进程进行结果汇总并发送给 Excel,最后由 Excel VBA 完成绘图。

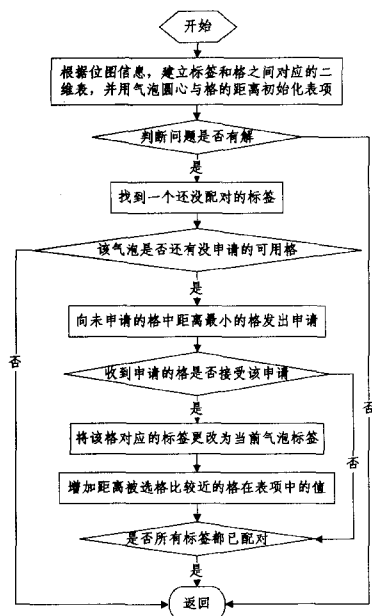


图9 stableMarriage 搜索流程图

4 贸易地图生成方案试验结果及性能评测

4.1 标签位置调整效果对比

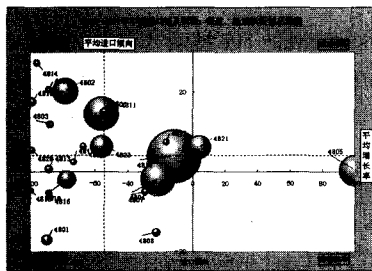


图10 进口贸易地图调整前的第四章四位码图

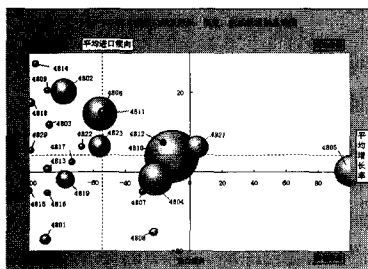


图11 进口贸易地图调整后的第四章四位码图

从图 10 可知,图中很多标签和其他的图元有重叠现象,位置不合适,运行并行标签调整程序后,该图效果如图 11 所示。

从图 11 可知,所有气泡的标签都被调整开了,没有标签和其它图元重叠现象,指示清晰、美观。

4.2 性能评测

贸易地图产品标签调整并行处理方案性能评测的主要内容是:测加速比和并行效率。并行算法加速比定义为: $S_P = T_S / T_P$,其中 T_S 为最佳串行算法在单处理机上的运行时间; T_P 为并行算法在并行机上使用 P 台处理机所需时间。相应并行效率定义为: $E_P = S_P / P$,其中 P 为处理器数。

本方案在测试时,保证了任务规模,机器硬件配置以及其它环境相同。调整 103 张出口产品贸易地图的标签,测试数据如表 1 所列。

表 1 并行方案性能测试表

	CPU(个)	时间 s	S_P	E_P
串行	1	254.2	1	1
并行	4	82.8	3.069	0.767
	8	44.3	5.738	0.717

从表 1 可知,并行计算随着节点数的增加完成整个作业的时间明显缩短,但由于网络存在通信和延时的问题,造成相应的加速比值低于理想状态,以及并行效率的下降。

结束语 在并行方案的设计过程中,从绘图区分区、划分网格、建立位图到搜索算法,在理论和实际上进行了验证,这种方法可以在类似的图元自动标注问题上得到应用和推广。从总体实现效果上看,该贸易地图气泡标签并行计算方案的设计目的已经达到:解决了标签自动放置的无遮挡分布问题,缩短了贸易地图生成时间,提高了效率。

参考文献

- [1] 陈泽星. 中国贸易地图(2006). 社会科学文献出版社,2006
- [2] 都志辉,李三立. 高性能计算并行编程技术-MPI 并行程序设计. 清华大学出版社,2001
- [3] 邵飞,邸瑞华. 贸易地图生成软件并行处理方案的研究与实现. 计算机科学,2008,35(3):267-270
- [4] 王建. 并行最优化算法与软件设计及数值软件移植[D]. 中国科学院研究生院(软件研究所),2005
- [5] Gusfield D, Irving R. The stable marriage problem: Structure and algorithms. The MIT Press, Cambridge, MA, USA, 1989

(上接第 252 页)

- [2] Bimbo A. Visual Information Retrieval[M]. San Francisco, CA, USA; Morgan Kaufmann Publishers, 1999
- [3] Zhao R, Grosky W I. Narrowing the semantic gap - improved text-based Web document retrieval using visual features [J]. IEEE Transactions on Multimedia, 2002, 4(2): 189-200
- [4] 徐银森. 篆刻艺术技法新编. 浙江人民美术出版社, 2002: 68-85
- [5] Gonzalez R C, Woods R E. Digital Image Processing Second Edition. Prentice Hall, 2002: 289
- [6] Chamberlin G J, Chamberlin D G. Colour: Its Measurement, Computation and Application. Heyden & Son Ltd, CIE 1960 and 1976 formulas, 1980: 60-61

- [7] Fortner B, Meyer T E. Number by Colors: A Guide to Using Color to Understand Technical Data. Excellent introduction to color, Springer-Verlag, 1997
- [8] Wyszecki, Günter, Stiles W S. Color Science -- Concepts and Methods, Quantitative Data and Formula. John Wiley, 1982: 139
- [9] Woods R E, Eddins S L. Digital Image Processing Using MATLAB. Prentice Hall, 2004
- [10] Hearn D, Baher M P. 计算机图形学. 蔡士杰, 吴春镛, 孙正兴, 等译. 电子工业出版社, 1998: 81-85
- [11] Gonzalez R C, Woods R E. Digital Image Processing Second Edition. 6. 2Color Models, p. 567; 10Image Segmentation, Prentice Hall, 2002: 289