(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 102662642 A
(43)申请公布日 2012.09.12

(21)申请号 201210116428.3

(22)申请日 2012.04.20

(71)申请人 浪潮电子信息产业股份有限公司
地址 250014 山东省济南市高新区舜雅路 1036 号

(72)发明人 卢晓伟

(51)Int.Cl.
G06F 9/38(2006.01)
G06N 3/12(2006.01)

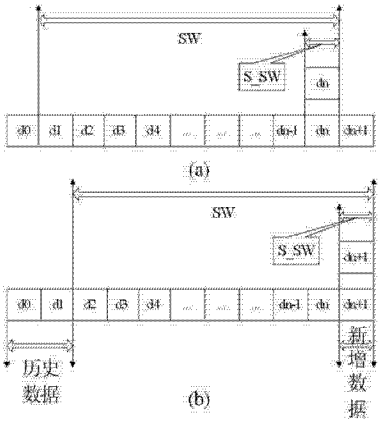
权利要求书 2 页 说明书 6 页 附图 7 页

(54)发明名称

一种基于嵌套滑动窗口和遗传算法的并行处理方法

(57)摘要

本发明提供一种基于嵌套滑动窗口和遗传算法的并行处理方法，采用基于滑动窗口之上的嵌套子窗口模型和利用遗传算法，根据数据流数据量大需要实时处理的特点，采用 GPU-CUDA 并行处理技术来进行动态挖掘出最新数据的频繁项集，综合处理滑动窗口内各嵌套子窗口中频繁项集，获得当前滑动窗口内数据的频繁项集，最后采用遗传算法的并行模式，得到数据流的频繁项集模式。



CN 102662642 A

1. 一种基于嵌套滑动窗口和遗传算法的并行处理方法， 其特征在于，采用基于滑动窗口之上的嵌套子窗口模型和利用遗传算法，根据数据流数据量大需要实时处理的特点，采用 GPU-CUDA 并行处理技术来进行动态挖掘出最新数据的频繁项集，综合处理滑动窗口内各嵌套子窗口中频繁项集，获得当前滑动窗口内数据的频繁项集，最后采用遗传算法的并行模式，得到数据流的频繁项集模式，获得当前滑动窗口内数据的频繁项集的步骤，包括：

1）滑动窗口内的数据分成 Z 段，将每个嵌套子窗口内的数据交给一个线程进行并行处理，获得初始种群，计算个体支持度值是初始种群内待考察频繁模式与实际事务匹配的过程，选择，交叉，变异，扫描确定变异后个体支持度值，判断结束条件；

2）获得的各个频繁项集模式与之前 U（$U=w_1/w_2-1$）次获得的频繁项集模式共同组成初始种群，进行一次搜索，最终满足条件的模式个体为滑动窗口内数据的频繁项集，随着数据流的流动，继续处理新接收到的数据，并抛弃最早的数据；

3） 利用遗传算法的并行性搜索嵌套子窗口内最新数据的频繁项集，从一组初始种群开始搜索过程，种群中的每个个体是一个可能的频繁模式，遗传算法通过交叉、变异、选择运算实现，经过若干代选择之后，得到最终频繁项集，其中变异操作是通过动态、随机改变个体中某些基因而产生新的个体，变异操作是产生全局最优的一个重要原因，有助于增加种群的多样性，但本算法中频繁项集产生所需的各对应非零基因都已存在，经交叉操作产生的基因基本上涵盖所有频繁项集，因此采用一个较低的变异率；

4） 综合处理滑动窗口内各嵌套子窗口中频繁项集，最终获得当前滑动窗口内数据的频繁项集；

5）随着新数据的流入，周期性删除过期流数据，并重复以上两部分操作；

具体步骤如下：

1）设定滑动窗口 SW 及子窗口 S_SW 大小，分别为 w_1、w_2 输入各类参数之后，根据数据流属性来确定窗口大小，SW 内容是根据当前多少条事务的频繁项集的兴趣度来决定的，子窗口是根据数据的处理能力以及被抛弃的旧数据条数来确定，也决定了需求所要求统计的频率；

2）给定支持度阈值 S，若某个个体 i， 其适应度为 F_i ，当 $F_i >= S$，事务 i 即为滑动窗口内数据集的频繁项集模式；

3）事务的属性种数、各属性的取值范围以及生成原始种群大小来确定最大迭代次数 T，处理方法是采用子窗口模型，避免在旧数据被淘汰之后，对滑动窗口 SW 内存在的数据进行多次重复处理；

4）设定交叉概率 P，个体变异概率 Q，子窗口内的数据分成 Z 段并行计算，函数采用 GPU CUDA 并行技术，将每个子窗口内的数据交给一个线程进行并行处理；

5）获得初始种群，数据在流动过程中，获取子窗口内最新到来的数据，同时得到此数据的频繁 1- 项集，将频繁 1- 项集编码为实数串，并将频繁 1- 项集非零项按原来所在位置随机组合编码，共同组成嵌套子窗口内的初始种群，此种群中个体为待考察频繁项集模式；

具体过程如下：

（1 ）统计 A、B 和 C 的属性值为 V1，V2，V3 的个数分别作为第一列、第二列和第三列；

（2）大于等于阈值 N 的保留，并按其所对应的行进行赋值，小于 N 的赋值 0，并去掉；

（3）将每一个非 0 值单独成一行，并保持其原来所在行的位置，其余位置填 0；

（4）非零项按原来所在位置随机组合编码，共同组成初始种群；

（5）函数是采用GPU CUDA编程模式，采用流技术和共享存储器的优化手段，将每个属性的求解过程进行并行处理；

6）计算个体支持度值是初始种群内待考察频繁模式与实际事务匹配的过程，当个体支持度值大于 S 时，将该个体模式加入当前子窗口频繁项集内，$F_i=W_i/W_Z$，F_i 为事务 i 的支持度，W_i 为当前子窗口内具有相同属性值的事务条数，W_Z 为当前子窗口内事务总条数；

分 Z 段并行匹配，虽然增大了内存开销，但大量减少运行时间，对于数据流频繁项集挖掘具有很大意义；

7）选择：将种群中个体按支持度值进行轮盘选择；

8）交叉：以交叉概率 P 进行一次交叉；

9）变异：个体按变异概率 Q 进行基本位变异；

10）扫描确定变异后个体支持度值，新增的满足条件的个体添加到频繁项集中；

11）判断结束条件，迭代次数小于 T，转步骤 3，T 次迭代运算后，则终止迭代并获得当前嵌套子窗口内数据的频繁项集；

12）随着数据流的流动，继续处理新接收到的数据，并抛弃最早的数据，转步骤 S102 继续以上操作，至数据流结束为止。

一种基于嵌套滑动窗口和遗传算法的并行处理方法

技术领域

[0001] 本发明涉及近期数据流频繁项集挖掘的实现方法，具体地说是一种基于嵌套子窗口模型和遗传算法的近期数据流频繁项集挖掘并行处理方法，。

背景技术

[0002] 数据流实际上就是连续移动的元素队伍，其中的元素是由相关数据的集合组成。令 t 表示任一时间戳，at 表示在该时间戳到达的数据，流数据可以表示成 {…，at 1，at，at+1，…}. 区别于传统应用模型，流数据模型具有以下 4 点共性：(1)数据实时到达；(2)数据到达次序独立，不受应用系统所控制；(3)数据规模宏大且不能预知其最大值；(4)数据一经处理，除非特意保存，否则不能被再次取出处理，或者再次提取数据代价昂贵。

[0003] 滑动窗口(sliding window)模型：滑动窗口对窗口起点和终点都没有明确给定，只明确给定窗口的长度 W。窗口保持一定长度在数据流 D={ d_0, d_1 …, d_n } 上滑动，处理的数据流范围就由该窗口确定，随着窗口的滑动不断地把得到的结果输出。滑动窗口 SW 的长度既可由一个时间区间确定，也可由窗口所包含数据流元素个数确定；

嵌套子窗口模型：某时刻 T ，窗口长度为 W 的滑动窗口 SW 内最新数据集 d_n 落入到窗口大小为 W2 的嵌套子窗口 S_SW 中，称窗口 SW 为嵌套子窗口。

[0004] 如图 1 所示，应用滑动窗口对动态更新数据集进行说明。窗口数据集为图 1 (a) 中标识所示。当新增数据集到达时，滑动窗口向前移动一个单位，如图 1 (b) 所示。

[0005] 滑动窗口的频繁项集：对于当前滑动窗口内数据，设 I={i_1, i_2, …, i_n} 是项的集合，事务数据集 S={ s_0, s_1 …, s_n }，其中，数据集中每个事务 s 是项的集合，s I。如果 X s，则称 X 是个项集。如果 X 中有 k 个元素，则称 X 为 k- 项集。对于一个项集 X，如果其支持度大于等于用户给定的最小支持度阈值，则 X 为频繁项集。

[0006] 遗传算法：一种基于随机搜索的优化算法，已成功应用于函数优化、自动控制、生产调度、机器人学、图像处理、人工生命、机器学习和数据挖掘等领域。从代表问题可能潜在的解集的一个种群开始的，而一个种群则由经过基因编码的一定数目的个体组成。每个个体实际上是染色体带有特征的实体。染色体作为遗传物质的主要载体，即多个基因的集合，其内部表现(即基因型)是某种基因组合，它决定了个体的形状的外部表现，如黑头发的特征是由染色体中控制这一特征的某种基因组合决定的。因此，在一开始需要实现从表现型到基因型的映射即编码工作。由于仿照基因编码的工作很复杂，我们往往进行简化，如二进制编码，初代种群产生之后，按照适者生存和优胜劣汰的原理，逐代演化产生出越来越好的近似解，在每一代，根据问题域中个体的适应度大小选择个体，并借助于自然遗传学的遗传算子进行组合交叉和变异，产生出代表新的解集的种群。这个过程将导致种群像自然进化一样的后生代种群比前代更加适应于环境，末代种群中的最优个体经过解码，可以作为问题近似最优解。

[0007] 遗传算法的基本运算过程如下，算法流程示意图如图 2：

a) 初始化：设置进化代数计数器 t=0，设置最大进化代数 T，随机生成 M 个个体作为初

始群体 $P_{(0)}$。

b) 个体评价：计算群体 $P_{(t)}$ 中各个个体的适应度。

c) 选择运算：将选择算子作用于群体。选择的目的是把优化的个体直接遗传到下一代或通过配对交叉产生新的个体再遗传到下一代。选择操作是建立在群体中个体的适应度评估基础上的。

d) 交叉运算：将交叉算子作用于群体。所谓交叉是指把两个父代个体的部分结构加以替换重组而生成新个体的操作。遗传算法中起核心作用的就是交叉算子。

e) 变异运算：将变异算子作用于群体。即是对群体中的个体串的某些基因座上的基因值作变动。群体 $P_{(t)}$ 经过选择、交叉、变异运算之后得到下一代群体 $P_{(t+1)}$。

f) 终止条件判断：若 t=T，则以进化过程中所得到的具有最大适应度个体作为最优解输出，终止计算。

[0008] CUDA 是一种并行编程模型和软件环境，采用 C 语言等标准编程语言进行操作。该技术封装了 GPU 的硬件细节，CUDA 的核心有三个重要抽象概念：线程组层次结构、共享存储器、屏蔽同步(barrier synchronization)。

[0009] 这些抽象提供了细粒度的数据并行化和线程并行化，嵌套于粗粒度的数据并行化和任务并行化之中，将问题分解为更小的片段，以便通过协作的方法并行解决。这样的分解保留了语言表达，允许线程在解决各子问题时协作，同时支持透明的可伸缩性。因而，该技术可以利用 GPU 的众核特性，大幅加速并行化的应用。

[0010] 但是，目前还没有一种能快速有效地获得流数据频繁项集的技术来提高用户的操作体验。

发明内容

[0011] 本发明所要解决的技术问题是需要提供一种适应流数据的流动性特点，采用遗传算法的并行形式，得到一种并行处理的理论依据和解决方法及以快速有效地获得流数据的频繁项集。

[0012] 本发明的目的是按以下方式实现的：

采用基于滑动窗口之上的嵌套子窗口模型和利用遗传算法，根据数据流数据量大需要实时处理的特点，采用 GPU-CUDA 并行处理技术来进行动态挖掘出最新数据的频繁项集，综合处理滑动窗口内各嵌套子窗口中频繁项集，获得当前滑动窗口内数据的频繁项集，最后采用遗传算法的并行模式，得到数据流的频繁项集模式，获得当前滑动窗口内数据的频繁项集的步骤，包括：

1) 滑动窗口内的数据分成 Z 段，将每个嵌套子窗口内的数据交给一个线程进行并行处理，获得初始种群，计算个体支持度值是初始种群内待考察频繁模式与实际事务匹配的过程，选择，交叉，变异，扫描确定变异后个体支持度值，判断结束条件；

2) 获得的各个频繁项集模式与之前 U（$U=w_1/w_2-1$）次获得的频繁项集模式共同组成初始种群，进行一次搜索，最终满足条件的模式个体为滑动窗口内数据的频繁项集，随着数据流的流动，继续处理新接收到的数据，并抛弃最早的数据；

3) 利用遗传算法的并行性搜索嵌套子窗口内最新数据的频繁项集，从一组初始种群开始搜索过程，种群中的每个个体是一个可能的频繁模式，遗传算法通过交叉、变异、选择

运算实现，经过若干代选择之后，得到最终频繁项集，其中变异操作是通过动态、随机改变个体中某些基因而产生新的个体，变异操作是产生全局最优的一个重要原因，有助于增加种群的多样性，但本算法中频繁项集产生所需的各对应非零基因都已存在，经交叉操作产生的基因基本上涵盖所有频繁项集，因此采用一个较低的变异率；

4）综合处理滑动窗口内各嵌套子窗口中频繁项集，最终获得当前滑动窗口内数据的频繁项集；

5）随着新数据的流入，周期性删除过期流数据，并重复以上两部分操作；

具体步骤如下：

1）设定滑动窗口 SW 及子窗口 S_SW 大小，分别为 w_1、w_2 输入各类参数之后，根据数据流属性来确定窗口大小，SW 内容是根据当前多少条事务的频繁项集的兴趣度来决定的，子窗口是根据数据的处理能力以及被抛弃的旧数据条数来确定，也决定了需求所要求统计的频率；

2）给定支持度阈值 S，若某个个体 i， 其适应度为 F_i ，当 $F_i >= S$，事务 i 即为滑动窗口内数据集的频繁项集模式；

3）事务的属性种数、各属性的取值范围以及生成原始种群大小来确定最大迭代次数 T，处理方法是采用子窗口模型，避免在旧数据被淘汰之后，对滑动窗口 SW 内存在的数据进行多次重复处理；

4）设定交叉概率 P，个体变异概率 Q，子窗口内的数据分成 Z 段并行计算，函数采用 GPU CUDA 并行技术，将每个子窗口内的数据交给一个线程进行并行处理；

5）获得初始种群，数据在流动过程中，获取子窗口内最新到来的数据，同时得到此数据的频繁 1- 项集，将频繁 1- 项集编码为实数串，并将频繁 1- 项集非零项按原来所在位置随机组合编码，共同组成嵌套子窗口内的初始种群，此种群中个体为待考察频繁项集模式；

具体过程如下：

（1 ）统计 A、B 和 C 的属性值为 V1，V2，V3 的个数分别作为第一列、第二列和第三列；

（2）大于等于阈值 N 的保留，并按其所对应的行进行赋值，小于 N 的赋值 0，并去掉；

（3）将每一个非 0 值单独成一行，并保持其原来所在行的位置，其余位置填 0；

（4）非零项按原来所在位置随机组合编码，共同组成初始种群；

（5）函数是采用 GPU CUDA 编程模式，采用流技术和共享存储器的优化手段，将每个属性的求解过程进行并行处理；

（6）计算个体支持度值是初始种群内待考察频繁模式与实际事务匹配的过程，当个体支持度值大于 S 时，将该个体模式加入当前子窗口频繁项集内，$F_i=W_i/W_Z$ ，F_i 为事务 i 的支持度，W_i 为当前子窗口内具有相同属性值的事务条数，W_Z 为当前子窗口内事务总条数；

分 Z 段并行匹配，虽然增大了内存开销，但大量减少运行时间，对于数据流频繁项集挖掘具有很大意义；

（7）选择：将种群中个体按支持度值进行轮盘选择；

（8）交叉：以交叉概率 P 进行一次交叉；

（9）变异：个体按变异概率 Q 进行基本位变异；

（10）扫描确定变异后个体支持度值，新增的满足条件的个体添加到频繁项集中；

（11）判断结束条件，迭代次数小于 T，转步骤 3，T 次迭代运算后，则终止迭代并获得当

前嵌套子窗口内数据的频繁项集；

（12）随着数据流的流动，继续处理新接收到的数据，并抛弃最早的数据，转步骤 S102 继续以上操作，至数据流结束为止。

[0013] 本发明与现有技术相比，本发明技术方案通过 GPU 强大的浮点计算能力和在 GPU 上进行编程的 CUDA 加速技术，处理流数据的频繁项集，可以采用遗传算法的并行形式进行建模，提升了用户操作体验。

[0014] 本领域的技术人员应该明白，上述的本发明的各模块或各步骤可以用通用的计算装置来实现，它们可以集中在单个的计算装置上，或者分布在多个计算装置所组成的网络上，可选地，它们可以用计算装置可执行的程序代码来实现，从而，可以将它们存储在存储装置中由计算装置来执行，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样，本发明不限制于任何特定的硬件和软件结合。

[0015] 虽然本发明所揭露的实施方式如上，但所述的内容只是为了便于理解本发明而采用的实施方式，并非用以限定本发明。任何本发明所属技术领域内的技术人员，在不脱离本发明所揭露的精神和范围的前提下，可以在实施的形式上及细节上作任何的修改与变化，但本发明的专利保护范围，仍须以所附的权利要求书所界定的范围为准。

附图说明

[0016] 图 1 是窗口数据集中的数据更新过程示意图；

图 2 是遗传算法流程示意图示意图；

图 3 是获取当前子窗口内频繁项集的流程图示意图；

图 4 初始种群生成示意图；

图 5 计算个体支持度；

图 6 频繁模式群形成；

图 7 扫描获得当前窗口最终频繁项集。

具体实施方式

[0017] 参照说明书附图对本发明的方法作以下详细地说明。

[0018] 以下将结合附图及实施例来详细说明本发明的实施方式，借此对本发明如何应用理论模型和技术手段来解决技术问题，并达成技术效果的实现过程能充分理解并据以实施。

[0019] 首先，如果不冲突，本发明实施例以及实施例中的各个特征的相互结合，均在本发明的保护范围之内。另外，在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行，并且，虽然在流程图中示出了逻辑顺序，但是在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤。

[0020] 利用遗传算法动态挖掘出最新数据的频繁项集，从一组初始种群开始搜索过程，种群中的每个个体是一个可能的频繁模式。遗传算法主要通过交叉、变异、选择运算实现。经过若干代选择之后，得到最终频繁项集。其中变异操作是通过动态、随机改变个体中某些基因而产生新的个体，变异操作是产生全局最优的一个重要原因，有助于增加种群的多样

性，但本算法中频繁项集产生所需的各对应非零基因都已存在，经交叉操作产生的基因基本上可以涵盖所有频繁项集，因此采用一个较低的变异率。

[0021] 本专利的方法分为三大部分：如图 1、2 所示；

1）利用遗传算法的并行性搜索嵌套子窗口内最新数据的频繁项集；

2）综合处理滑动窗口内各嵌套子窗口中频繁项集，最终获得当前滑动窗口内数据的频繁项集；

3）随着新数据的流入，周期性删除过期流数据，并重复以上两部分操作。

[0022] 实施例 1

利用遗传算法的并行性搜索嵌套子窗口内最新数据的频繁项集；如图 3 所示，本实施例主要包括如下步骤：

1）设定滑动窗口 SW 及子窗口 S_SW 大小，分别为 w_1、w_2 输入各类参数之后，根据数据流属性来确定窗口大小，SW 内容是根据当前多少条事务的频繁项集的兴趣度来决定的，子窗口是根据数据的处理能力以及被抛弃的旧数据条数来确定，也决定了需求所要求统计的频率；

2）给定支持度阈值 S，若某个个体 i， 其适应度为 F_i ，当 $F_i > = S$，事务 i 即为滑动窗口内数据集的频繁项集模式；

3）事务的属性种数、各属性的取值范围以及生成原始种群大小来确定最大迭代次数 T。本处理方法是采用子窗口模型，避免在旧数据被淘汰之后，对滑动窗口 SW 内存在的数据进行多次重复处理；

4）设定交叉概率 P，个体变异概率 Q，子窗口内的数据分成 Z 段并行计算。该处的函数采用 GPU CUDA 并行技术，将每个子窗口内的数据交给一个线程进行并行处理；

5）获得初始种群。数据在流动过程中，获取子窗口内最新到来的数据，同时得到此数据的频繁 1- 项集，将频繁 1- 项集编码为实数串，并将频繁 1- 项集非零项按原来所在位置随机组合编码，共同组成嵌套子窗口内的初始种群，此种群中个体为待考察频繁项集模式。具体过程如下：

（1 ）统计 A、B 和 C 的属性值为 V1，V2，V3 的个数分别作为第一列、第二列和第三列；

（2）大于等于阈值 N 的保留，并按其所对应的行进行赋值，小于 N 的赋值 0，并去掉（本例 N 取 3）；

（3）将每一个非 0 值单独成一行，并保持其原来所在行的位置，其余位置填 0；

（4）非零项按原来所在位置随机组合编码，共同组成初始种群；

过程如图 4 所示，该步骤的函数是采用 GPU CUDA 编程模式，采用流技术和共享存储器等优化手段，将每个属性的求解过程进行并行处理；

5）计算个体支持度值是初始种群内待考察频繁模式与实际事务匹配的过程。当个体支持度值大于 S 时，将该个体模式加入当前子窗口频繁项集内。$F_i=W_i/W_Z$ ，F_i 为事务 i 的支持度，W_i 为当前子窗口内具有相同属性值的事务条数，W_Z 为当前子窗口内事务总条数；

6）分 Z 段并行匹配，虽然增大了内存开销，但大量减少运行时间，对于数据流频繁项集挖掘具有很大意义。并行匹配，过程如图 5 所示；

7）选择。将种群中个体按支持度值进行轮盘选择；

8）交叉。以交叉概率 P 进行一次交叉；

9）变异。个体按变异概率 Q 进行基本位变异；

10）扫描确定变异后个体支持度值，新增的满足条件的个体添加到频繁项集中；

11）判断结束条件。如迭代次数小于 T，转步骤 3，T 次迭代运算后，则终止迭代并获得当前嵌套子窗口内数据的频繁项集；

12）随着数据流的流动，继续处理新接收到的数据，并抛弃最早的数据，转步骤 S102 继续以上操作，至数据流结束为止；

实施例 2

获得当前滑动窗口内数据的频繁项集，步骤如下：

1）本次获得的各个频繁项集模式与之前 U（$U=w_1/w_2-1$）次获得的频繁项集模式共同组成初始种群，进行一次搜索，最终满足条件的模式个体为滑动窗口内数据的频繁项集。过程如图 6 和图 7 所示；

（1）For i=1 ;U+1 ;

（2）将各段得到的频繁模式组合成为频繁模式群；

（3）End ;

（4）将频繁模式群在滑动窗口 SW 内进行一次并行搜索；

（5）支持度大于 S 的最终确定为频繁模式；函数采用 OpenMP 共享编程模式进行多线程并行处理；

（6）随着数据流的流动，继续处理新接收到的数据，并抛弃最早的数据，转步骤 S102 继续以上操作，至数据流结束为止。

[0023] 除说明书所述的技术特征外，均为本专业技术人员的已知技术。

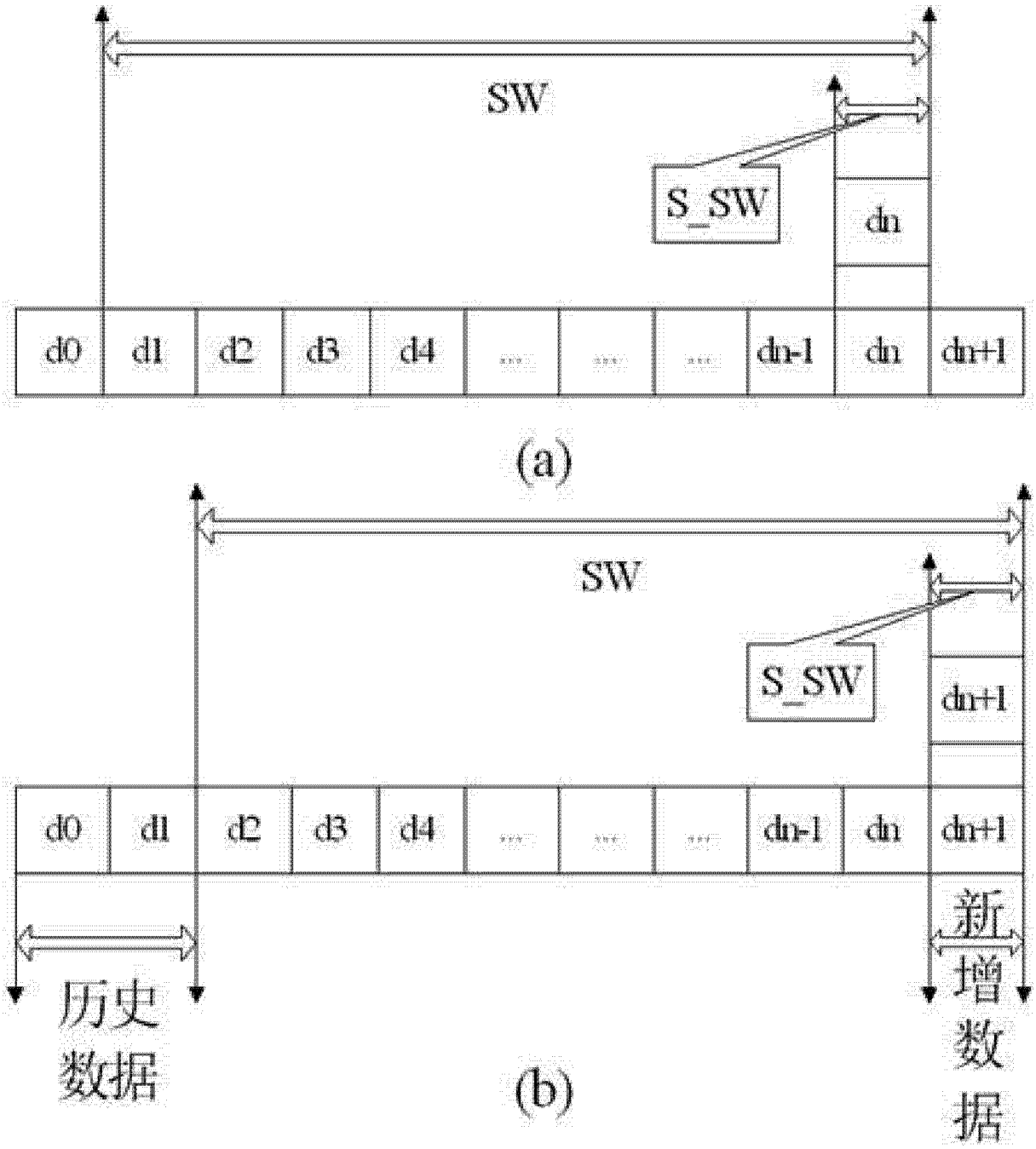

SW
S_SW
dn
d0
d1
d2
d3
d4
...
...
...
dn-1
dn
dn+1
(a)
SW
S_SW
dn+1
d0
d1
d2
d3
d4
...
...
...
dn-1
dn
dn+1
历史数据
新增数据
(b)

图 1

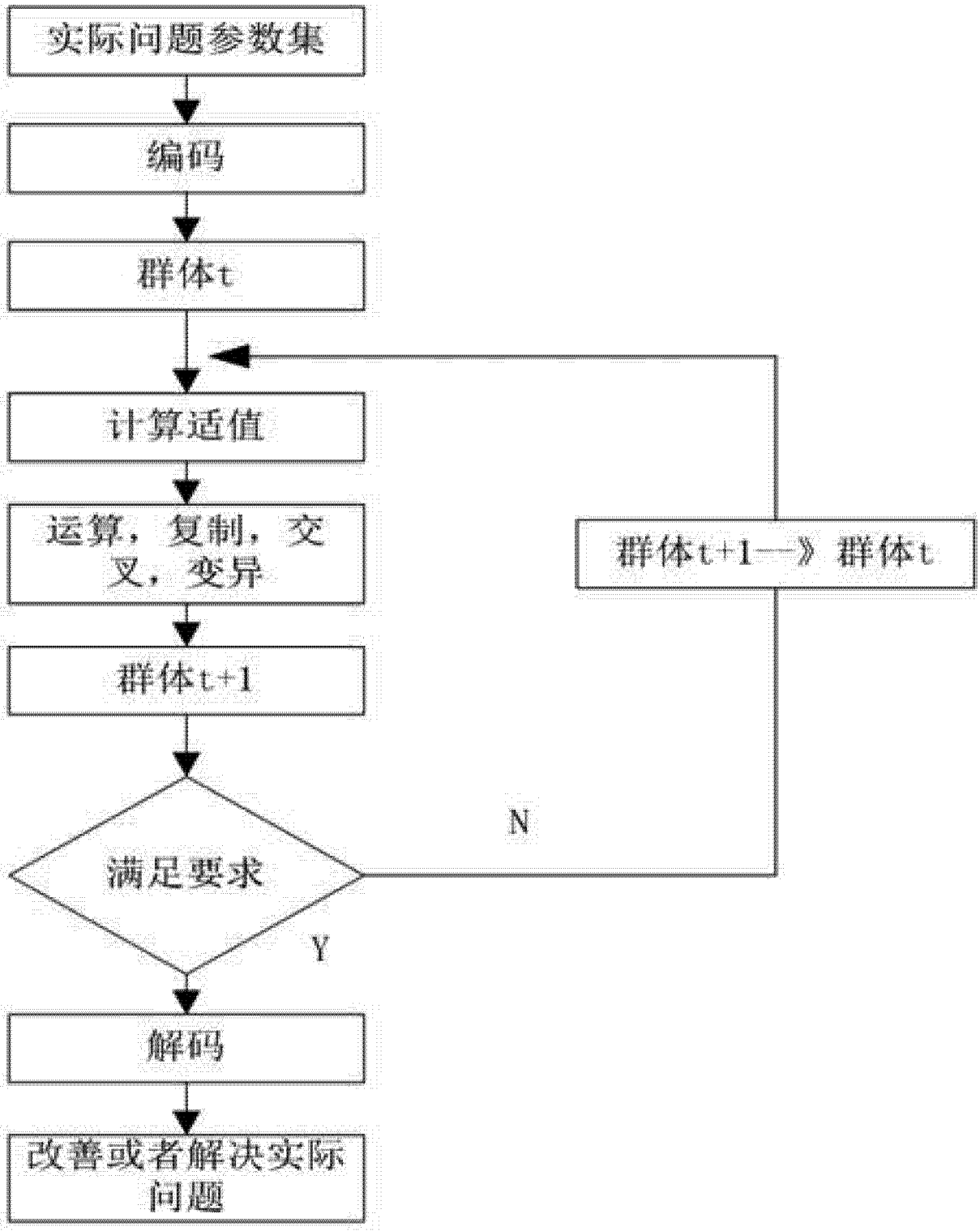

实际问题参数集
编码
群体t
计算适值
运算，复制，交叉，变异
群体t+1
满足要求
N
Y
解码
改善或者解决实际问题
群体t+1—》群体t

图 2

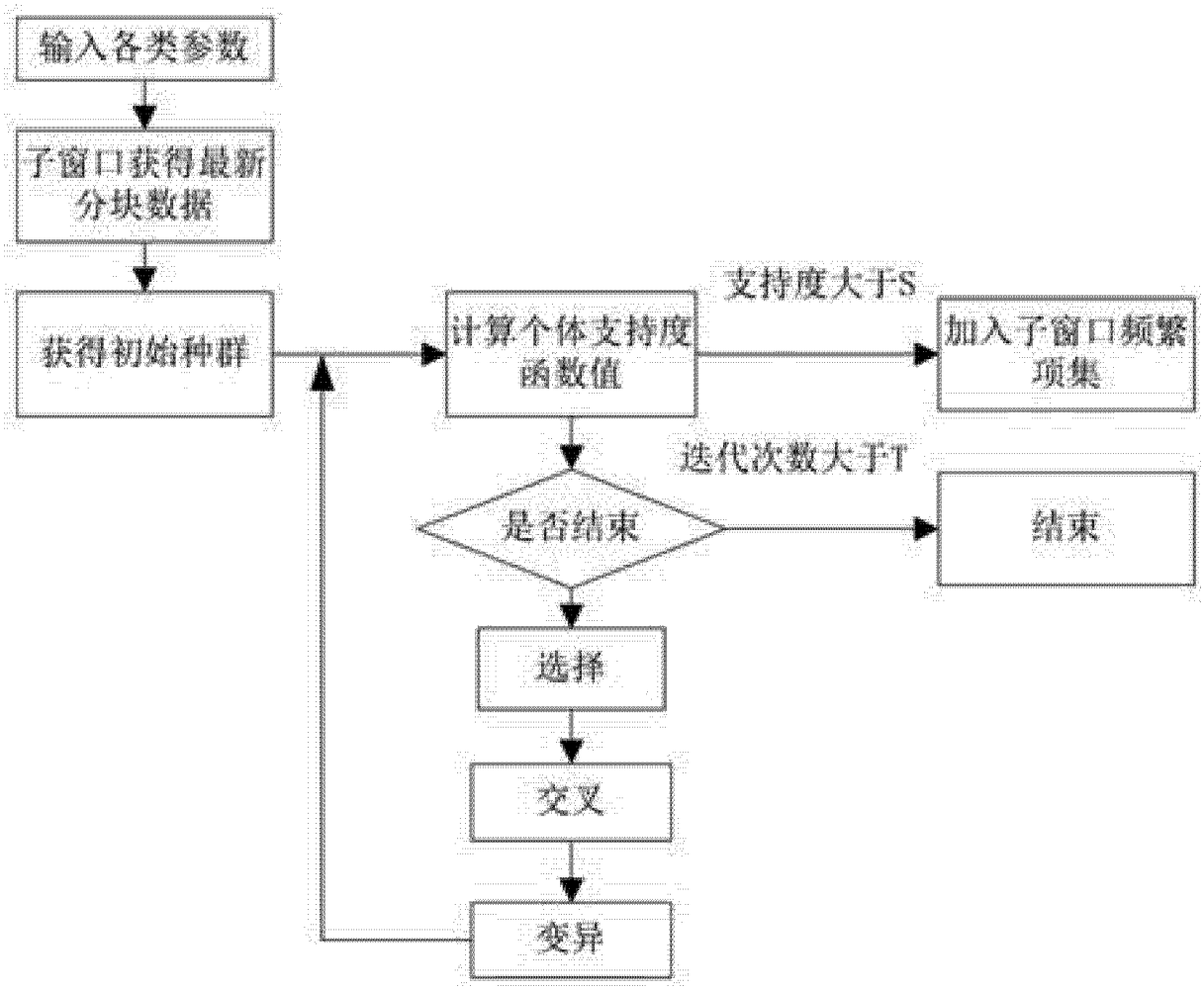

输入各类参数
子窗口获得最新分块数据
获得初始种群
计算个体支持度函数值
支持度大于S
加入子窗口频繁项集
迭代次数大于T
是否结束
结束
选择
交叉
变异

图 3

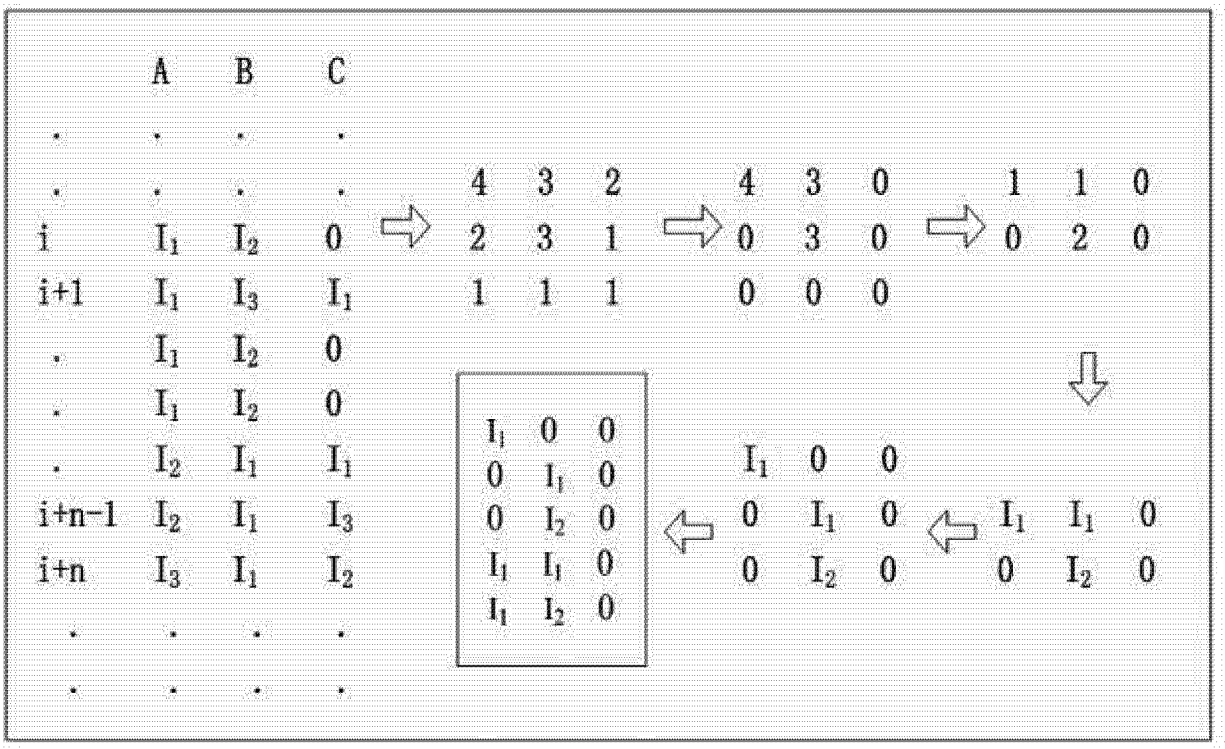

A B C
i I1 I2 0
i+1 I1 I3 I1
I1 I2 0
I1 I2 0
I2 I1 I1
i+n-1 I2 I1 I3
i+n I3 I1 I2
4 3 2
2 3 1
1 1 1
4 3 0
0 3 0
0 0 0
1 1 0
0 2 0
I1 0 0
0 I1 0
0 I2 0
I1 I1 0
I1 I2 0
I1 0 0
0 I1 0
0 I2 0
I1 I1 0
0 I2 0

图 4

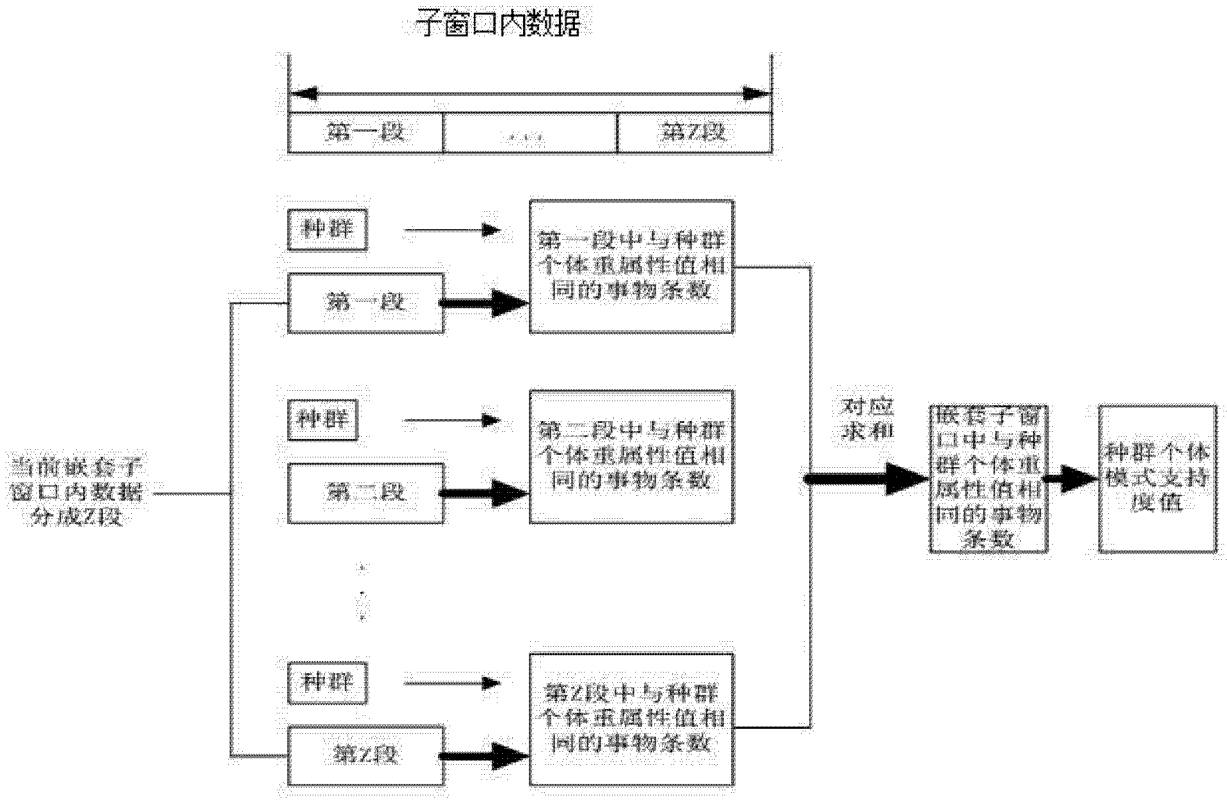

子窗口内数据
第一段
...
第Z段
当前嵌套子窗口内数据分成Z段
种群
第一段
第一段中与种群个体重属性值相同的事物条数
种群
第二段
第二段中与种群个体重属性值相同的事物条数
种群
第Z段
第Z段中与种群个体重属性值相同的事物条数
对应求和
嵌套子窗口中与种群个体重属性值相同的事物条数
种群个体模式支持度值

图 5

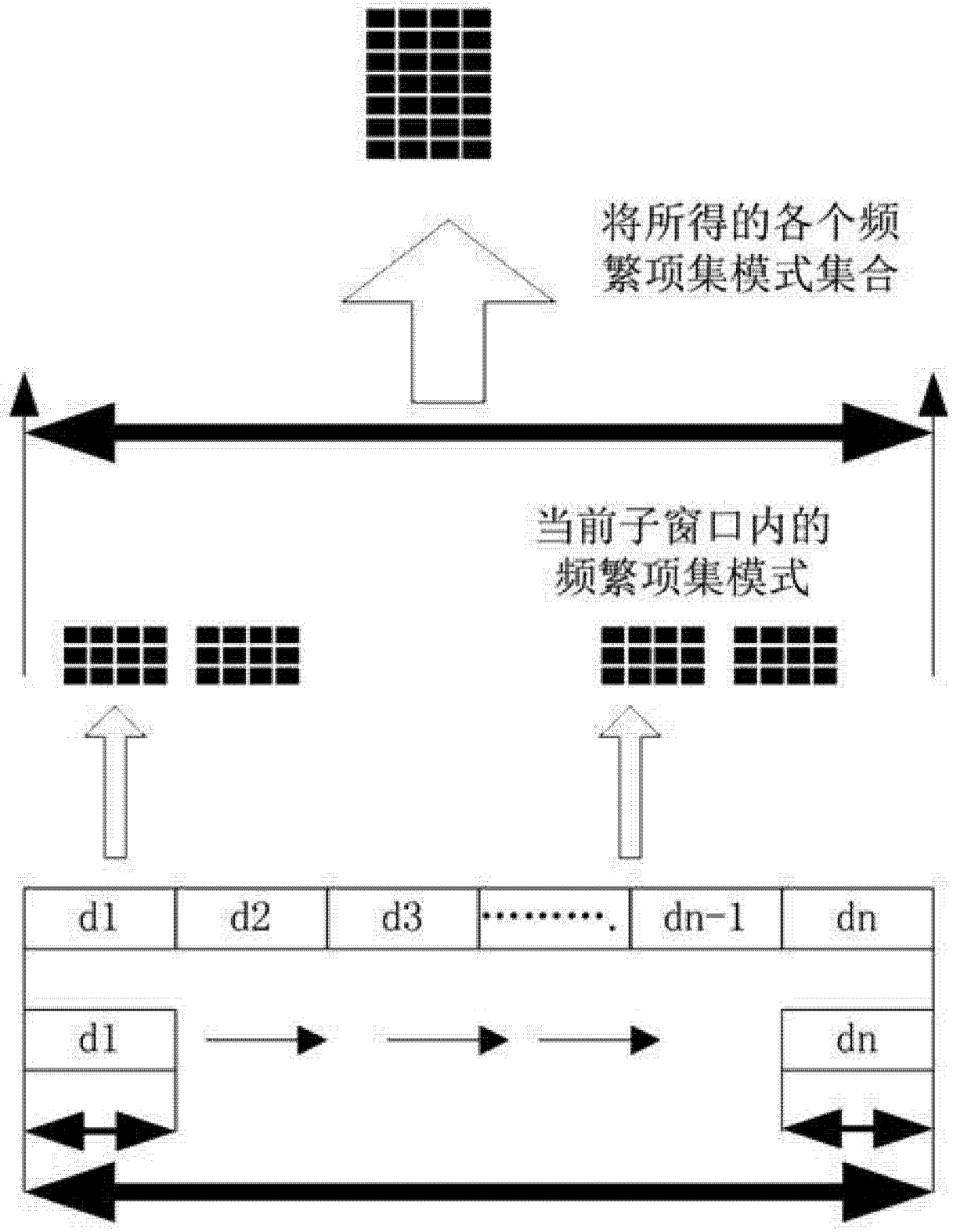


图 6

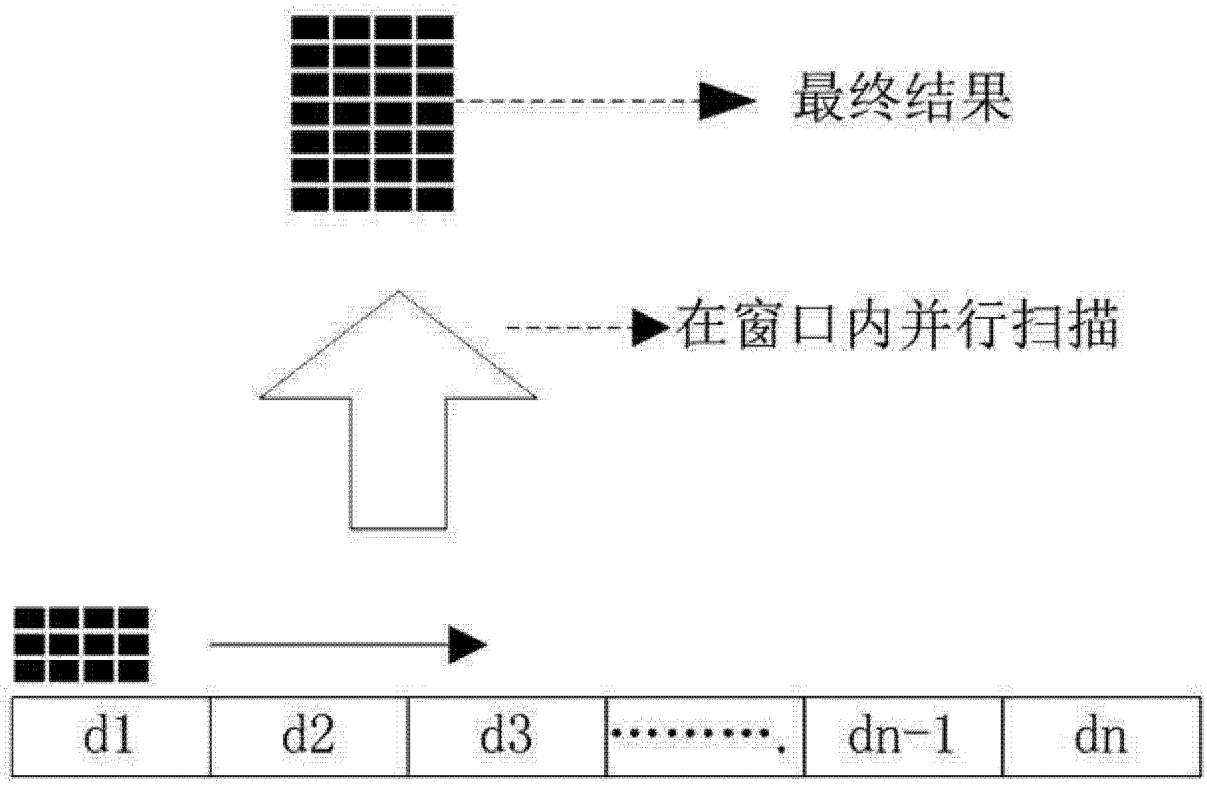


图 7