



# (12)发明专利申请

(10)申请公布号 CN 111553473 A

(43)申请公布日 2020.08.18

(21)申请号 202010617408.9

周聖元

(22)申请日 2018.07.05

(74)专利代理机构 广州三环专利商标代理有限公司 44202

(66)本国优先权数据

代理人 熊永强

- 201710541899.1 2017.07.05 CN
- 201710578784.X 2017.07.14 CN
- 201710677922.X 2017.08.09 CN
- 201710793531.4 2017.09.06 CN
- 201710910124.7 2017.09.29 CN

(51)Int.Cl.

- G06N 3/063(2006.01)
- G06F 9/30(2006.01)

(62)分案原申请数据

201880001841.0 2018.07.05

(71)申请人 上海寒武纪信息科技有限公司

地址 201306 上海市浦东新区同汇路168号 B座6层

(72)发明人 陈天石 张磊 刘少礼 王在

周徐达 杜子东 胡帅 何皓源

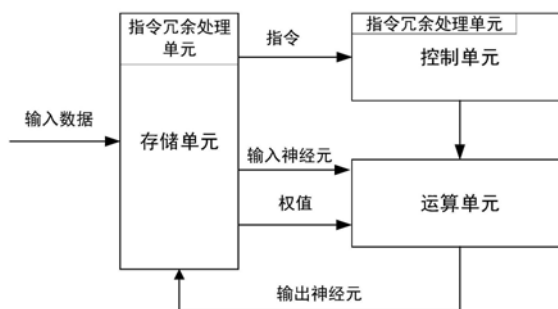
权利要求书3页 说明书6页 附图3页

(54)发明名称

数据冗余方法及执行数据冗余方法的神经网络处理器

(57)摘要

本申请公开了一种数据冗余方法和神经网络处理器,包括:存储单元,控制单元、运算单元和指令冗余处理单元。



1. 一种执行数据冗余方法的神经网络处理器,其特征在于,包括:存储单元,控制单元、运算单元和指令冗余处理单元;

存储单元,用于接收外界输入的数据,存储神经网络的神经元、权值和指令,并将指令发送给指令控制单元,将神经元和权值发送给运算单元;

控制单元,接收存储单元发送的指令,经过译码后生成控制信息控制运算单元;

运算单元,用于接收存储单元发送的权值和神经元,完成神经网络训练运算并将输出神经元重新传输给存储单元存储;

指令冗余处理单元,用于对指令进行数据冗余处理。

2. 根据权利要求1所述的神经网络处理器,其特征在于,

所述指令冗余处理单元,具体用于对所述指令数据进行读操作时,读取原始数据和备份的两份数据,若出现不一致,则取相同的两份数据作为最终读取的数据,同时将第三份不一致的数据进行修复;当对所述指令数据进行写操作时,同时将所述指令数据写回到两个备份地址,同时保证原始数据和两个备份数据一致。

3. 根据权利要求1所述的神经网络处理器,其特征在于,所述指令冗余处理单元分别内嵌至存储单元和控制单元中。

4. 根据权利要求1所述的神经网络处理器,其特征在于,所述数据包括:神经网络参数,按照神经网络参数的绝对值划分M个重要等级,按照由大到小的顺序设定M+1个阈值 $T_0, T_1, T_2, \dots$ 及 $T_M$ ,当神经网络参数的绝对值D满足 $T_{i-1} > D > T_i$ 时,该数据归为第i个重要等级,其中 $i = 1, 2, \dots, M$ ,  $T_0, T_1, T_2, \dots$ 及 $T_M$ 均为实数,且 $T_0 > T_1 > T_2 > \dots > T_M \geq 0$ 。

5. 根据权利要求1所述的神经网络处理器,其特征在于,所述数据包括稀疏神经网络参数,所述稀疏神经网络参数包括非零参数和非零参数位置,所述非零参数位置设置为第1重要等级,非零参数按照参数绝对值划分M-1个重要等级,按照由大到小的顺序设定M个阈值 $T_1, T_2, \dots$ 及 $T_M$ ,当非零参数的绝对值D满足 $T_{i-1} > D > T_i$ 时,该数据归为第i个重要等级,其中 $i = 2, 3, \dots, M$ ,  $T_1, T_2, \dots$ 及 $T_M$ 均为实数,且 $T_1 > T_2 > \dots > T_M \geq 0$ 。

6. 根据权利要求1所述的神经网络处理器,其特征在于,所述数据包括图计算应用数据,包括顶点数据和边数据,所述顶点数据设置为第1重要等级,边数据按照访问频率划分M-1个重要等级,按照由大到小的顺序设定M个阈值 $T_1, T_2, \dots$ 及 $T_M$ ,当边数据按照访问频率F满足 $T_{i-1} > F > T_i$ 时,该数据归为第i个重要等级,其中 $i = 2, 3, \dots, M$ ,  $T_1, T_2, \dots$ 及 $T_{M-1}$ 均为实数,且 $T_1 > T_2 > \dots > T_M \geq 0$ 。

7. 根据权利要求1所述的神经网络处理器,其特征在于,

所述数据为浮点型参数,设置符号位、指数部分以及底数部分前y位指定为重要比特位,y为正整数;和/或

所述数据为定点型参数,设置符号位、数值部分前z位指定为重要比特位,z为正整数。

8. 根据权利要求1或2所述的神经网络处理器,其特征在于,所述运算单元包括一个主处理电路、多个从处理电路和分支处理电路;

所述主处理电路,具体用于将一个输入神经元分配成多个数据块,将所述多个数据块中的至少一个数据块、权值以及多个运算指令中的至少一个运算指令发送给所述分支处理电路;

所述分支处理电路,用于转发所述主处理电路与所述多个从处理电路之间的数据块、

权值以及运算指令；

所述多个从处理电路，用于依据该运算指令对接收到的数据块以及权值执行运算得到中间结果，并将中间结果传输给所述分支处理电路；

所述主处理电路，用于将分支处理电路发送的中间结果进行后续处理得到该计算指令的结果，将该计算指令的结果发送给所述控制单元。

9. 根据权利要求1或2所述的神经网络处理器，其特征在于，所述运算单元包括一个主处理电路和多个从处理电路；所述多个从处理电路呈阵列分布；每个从处理电路与相邻的其他从处理电路连接，所述主处理电路连接所述多个从处理电路中的k个从处理电路，所述k个从处理电路为：第1行的n个从处理电路、第m行的n个从处理电路以及第1列的m个从处理电路；

所述k个从处理电路，用于在所述主处理电路以及多个从处理电路之间的数据以及指令的转发；

所述主处理电路，用于将一个输入数据分配成多个数据块，将所述多个数据块中的至少一个数据块以及多个运算指令中的至少一个运算指令发送给所述k个从处理电路；

所述k个从处理电路，用于转换所述主处理电路与所述多个从处理电路之间的数据；

所述多个从处理电路，用于依据该运算指令对接收到的数据块执行运算得到中间结果，并将运算结果传输给所述k个从处理电路；

所述主处理电路，用于将所述k个从处理电路发送的中间结果进行后续处理得到该计算指令的结果，将该计算指令的结果发送给所述控制单元。

10. 一种数据冗余方法，其特征在于，所述方法应用于神经网络处理器，所述神经网络处理器包括：存储单元、控制单元、运算单元和指令冗余处理单元；所述方法包括：

存储单元接收外界输入的数据，存储神经网络的神经元、权值和指令，并将指令发送给指令控制单元，将神经元和权值发送给运算单元；

控制单元接收存储单元发送的指令，经过译码后生成控制信息控制运算单元；

运算单元接收存储单元发送的权值和神经元，完成神经网络训练运算并将输出神经元重新传输给存储单元存储；

指令冗余处理单元对指令进行数据冗余处理。

11. 根据权利要求10所述的方法，其特征在于，所述指令冗余处理单元对指令进行数据冗余处理具体包括：

对所述指令数据进行读操作时，读取原始数据和备份的两份数据，若出现数据不一致，则取相同的两份数据作为最终读取的数据，同时将第三份不一致的数据进行修复；当对所述指令数据进行写操作时，同时将所述指令数据写回到两个备份地址，同时保证原始数据和两个备份数据一致。

12. 根据权利要求10所述的方法，其特征在于，所述方法还包括：

若所述数据包括神经网络参数时，按照神经网络参数的绝对值划分M个重要等级，按照由大到小的顺序设定M+1个阈值 $T_0, T_1, T_2, \dots$ 及 $T_M$ ，当神经网络参数的绝对值D满足 $T_{i-1} > D > T_i$ 时，该数据归为第i个重要等级，其中 $i = 1, 2, \dots, M$ ， $T_0, T_1, T_2, \dots$ 及 $T_M$ 均为实数，且 $T_0 > T_1 > T_2 > \dots > T_M \geq 0$ ；

若所述数据包括稀疏神经网络参数时，所述稀疏神经网络参数包括非零参数和非零参

数位置,所述非零参数位置设置为第1重要等级,非零参数按照参数绝对值划分M-1个重要等级,按照由大到小的顺序设定M个阈值 $T_1, T_2, \dots$ 及 $T_M$ ,当非零参数的绝对值D满足 $T_{i-1} > D > T_i$ 时,该数据归为第i个重要等级,其中 $i=2, 3, \dots, M$ ,  $T_1, T_2, \dots$ 及 $T_M$ 均为实数,且 $T_1 > T_2 > \dots > T_M \geq 0$ ;

若所述数据包括图计算应用数据时,所述图计算应用数据包括顶点数据和边数据,所述顶点数据设置为第1重要等级,边数据按照访问频率划分M-1个重要等级,按照由大到小的顺序设定M个阈值 $T_1, T_2, \dots$ 及 $T_M$ ,当边数据按照访问频率F满足 $T_{i-1} > F > T_i$ 时,该数据归为第i个重要等级,其中 $i=2, 3, \dots, M$ ,  $T_1, T_2, \dots$ 及 $T_{M-1}$ 均为实数,且 $T_1 > T_2 > \dots > T_M \geq 0$ 。

13. 根据权利要求10所述的方法,其特征在于,

若所述数据为浮点型参数时,设置符号位、指数部分以及底数部分前y位指定为重要比特位,y为正整数;

若所述数据为定点型参数时,设置符号位、数值部分前z位指定为重要比特位,z为正整数。

14. 根据权利要求10所述的方法,其特征在于,所述方法还包括:

将数据划分为M个重要等级;提取每一重要等级中的每一数据的重要比特位;

以及针对所述重要比特位进行数据冗余处理,其中M为正整数。

15. 根据权利要求14所述的方法,其中,若所述数据冗余处理包括纠错编码处理;所述纠错编码处理包括:循环冗余校验CRC和/或错误检查和纠正ECC根据权利要求3所述的数据冗余方法,其中,所述CRC具体包括:当进行读操作时,读取存储循环冗余校验码,并对原始数据进行循环冗余校验码计算,如果两个循环冗余校验码不一致,则按照存储的循环冗余校验码对该数据进行修正,进行写操作时,同时存储该数据的原始数据和循环冗余校验码;

所述ECC具体包括:对数据进行纠错编码处理包括:对该数据采用错误检查和纠正内存进行冗余存储,在进行读写操作时,错误检查和纠正内存自动进行检错和纠错处理。

16. 根据权利要求14所述的方法,其中,若所述数据冗余处理包括副本冗余处理,所述副本冗余处理包括同时备份N份副本,其中 $N \geq 2$ ,且N为正整数;其中,对数据的重要比特位进行副本冗余处理包括:进行读操作时,对于重要比特位同时读取原始数据和备份的N份数据,若出现数据不一致,则取相同的N份数据作为最终读取的数据,同时将第N+1份不一致数据修复,当进行写操作时,对于重要比特位同时写回到N个备份地址,同时保证原始数据中的数据 and N个备份数据一致。

17. 根据权利要求14-16中任一所述的方法,其中,所述将数据划分为M个重要等级包括根据数据的大小、数据绝对值的大小、数据的类型、数据的读操作频率及数据的写操作频率中的至少一个来进行划分。

## 数据冗余方法及执行数据冗余方法的神经网络处理器

### 技术领域

[0001] 本申请涉及数据处理领域,更具体地涉及一种数据冗余方法及执行数据冗余方法的神经网络处理器。

### 背景技术

[0002] 数据冗余技术采用副本冗余或者纠错编码的方式提高数据的安全性和可靠性,但是数据冗余技术带来巨大的存储容量开销和访存功耗开销,针对大规模的数据,这个问题会更加严重。因此如何结合数据的特性进行数据冗余成为一个亟待解决的问题。

### 发明内容

[0003] 鉴于现有方案存在的问题,为了克服上述现有技术方案的不足,本申请提出了一种数据冗余方法及神经网络处理器。

[0004] 根据本申请的一个方面,提供了一种执行数据冗余方法的神经网络处理器,包括:存储单元,控制单元、运算单元和指令冗余处理单元;

[0005] 存储单元,用于接收外界输入的数据,存储神经网络的神经元、权值和指令,并将指令发送给指令控制单元,将神经元和权值发送给运算单元;

[0006] 控制单元,接收存储单元发送的指令,经过译码后生成控制信息控制运算单元;

[0007] 运算单元,用于接收存储单元发送的权值和神经元,完成神经网络训练运算并将输出神经元重新传输给存储单元存储;

[0008] 指令冗余处理单元,用于对指令进行数据冗余处理。

[0009] 本申请另一方面提供一种数据冗余方法,所述方法应用于神经网络处理器,所述神经网络处理器包括:存储单元,控制单元、运算单元和指令冗余处理单元;所述方法包括:

[0010] 存储单元接收外界输入的数据,存储神经网络的神经元、权值和指令,并将指令发送给指令控制单元,将神经元和权值发送给运算单元;

[0011] 控制单元接收存储单元发送的指令,经过译码后生成控制信息控制运算单元;

[0012] 运算单元接收存储单元发送的权值和神经元,完成神经网络训练运算并将输出神经元重新传输给存储单元存储;

[0013] 指令冗余处理单元对指令进行数据冗余处理。

[0014] 从上述技术方案可以看出,本申请具有至少以下有益效果:

[0015] 对数据选择性的进行冗余处理,减少存储容量开销和访存功耗开销;

[0016] 区分数据的重要等级,对不同重要等级的数据进行不同的冗余处理,保证存储数据的安全性和可靠性的同时,减少存储容量占用和访存功耗。

### 附图说明

[0017] 图1为本申请一实施例中数据冗余方法的流程图;

[0018] 图2为本申请另一实施例数据冗余装置的结构框图;

- [0019] 图3为本申请一实施例神经网络处理器；
- [0020] 图4为本申请实例提供的运算单元的结构示意图；
- [0021] 图5为本申请实施例提供的运算单元的另一结构示意图；
- [0022] 图6为本申请实施例提供的运算单元的又一结构示意图。

### 具体实施方式

[0023] 本申请某些实施例于后方将参照所附图做更全面性地描述,其中一些但并非全部的实施例将被示出。实际上,本申请的各种实施例可以许多不同形式实现,而不应被解释为限于此数所阐述的实施例;相对地,提供这些实施例使得本申请满足适用的法律要求。

[0024] 在本说明书中,下述用于描述本申请原理的各种实施例只是说明,不应该以任何方式解释为限制申请的范围。参照附图的下述描述用于帮助全面理解由权利要求及其等同物限定的本申请的示例性实施例。

[0025] 为使本申请的目的、技术方案和优点更加清楚明白,以下结合具体实施例,并参照附图,对本申请进一步详细说明。

[0026] 本申请实施例提供了一种数据冗余方法,通过将数据划分为多个重要等级,并针对不同的重要等级的数据进行不同的数据冗余处理,在保证存储数据的安全性和可靠性的基础上,减少存储容量开销和访存功耗开销。

[0027] 具体的,图1示出了数据冗余方法的流程图,如附图1所示,数据冗余方法包括以下具体步骤:

[0028] 步骤S101:将数据划分为M个重要等级,M为正整数;

[0029] 具体地,数据的重要等级可以综合考虑数据的大小,数据绝对值的大小,数据的类型(浮点型、定点型),数据的读操作频率,数据写操作频率等因素进行设置。

[0030] 步骤S102:提取每一重要等级中的每一数据的重要比特位;

[0031] 具体地,数据中的比特位分为重要比特位和非重要比特位。若数据共有x个比特位,其中y个比特位是重要比特位,(x-y)个比特位是不重要比特位,其中x,y为正整数。后续仅对数据的y个重要比特位进行处理,y个重要比特位的位置可以是连续的,也可以是不连续的。

[0032] 步骤S103:针对所述重要比特位进行数据冗余处理。

[0033] 具体地,所述数据冗余处理包括副本冗余处理和/或纠错编码处理,可以根据不同的重要性来进行不同的处理,例如可以,当一数据中的所有比特位均为重要比特位时,对该数据的所有比特位进行纠错编码处理,当一数据中的部分比特位为重要比特位时,对该数据的重要比特位进行副本冗余处理。

[0034] 副本冗余可以在同一块存储介质备份冗余,也可以在不同存储介质备份冗余。数据可以同时备份N份,其中N是大于0的正整数。纠错编码方式包括但不限于循环冗余校验(Cyclic Redundancy Check,CRC),错误检查和纠正(Error Correcting Code,ECC)。

[0035] 以下通过几个实例具体介绍本实施例中的数据冗余方法。

[0036] 实例0:对神经网络指令做冗余,对参数不做冗余

[0037] 将神经网络指令设定为第1重要等级,将神经网络的参数设定为第2重要等级。其中神经网络参数包括拓扑结构信息,神经元数据和权值数据。对第1重要等级的数据进行冗

余存储,对第2重要等级的数据不进行冗余存储。当对第1重要等级数据进行读操作时,读取原始数据和备份的两份数据,若出现相应数据不一致,则取相同的两份数据作为最终读取的数据,同时将第三份不一致的数据进行修复。当对第1重要等级数据进行写操作时,同时写回到两个备份地址,同时保证原始数据和两个备份数据一致。

[0038] 如图3所示,对神经网络处理器由存储单元,控制单元和运算单元组成。存储单元接收外界的输入数据,存储神经网络的神经元、权值和指令,并将指令发送给指令控制单元,将神经元和权值发送给运算单元。

[0039] 指令控制单元,接收存储单元发送的指令,经过译码后生成控制信息控制运算单元。

[0040] 运算单元,用于接收存储单元发送的权值和神经元,完成神经网络训练运算并将输出神经元重新传输给存储单元存储。

[0041] 神经网络处理器还包括指令冗余处理单元,分别内嵌至存储单元和指令控制单元中,对指令进行数据冗余处理。

[0042] 在本公开的一些实施例中,运算单元的拓扑结构可以如图4所示,运算单元可以包括一个主处理电路和多个从处理电路。图4所示的拓扑结构为树型模块,所述树型模块包括:一个根端口和多个支端口,所述树型模块的根端口连接所述主处理电路,所述树型模块的多个支端口分别连接多个从处理电路中的一个从处理电路;所述树型模块,用于转发所述主处理电路与所述多个从处理电路之间的数据块、权值以及运算指令。如图4所示,树型模块可以包括多层节点结构,节点为具有转发功能的结构,该节点本身可以不具有计算功能。

[0043] 在本公开的一些实施例中,运算单元的拓扑结构可以如图5所示,运算单元可以包括一个主处理电路、多个从处理电路、和分支处理电路。所述主处理电路,具体用于将一个输入神经元分配成多个数据块,将所述多个数据块中的至少一个数据块、权值以及多个运算指令中的至少一个运算指令发送给所述分支处理电路;

[0044] 所述分支处理电路,用于转发所述主处理电路与所述多个从处理电路之间的数据块、权值以及运算指令;

[0045] 所述多个从处理电路,用于依据该运算指令对接收到的数据块以及权值执行运算得到中间结果,并将中间结果传输给所述分支处理电路;

[0046] 所述主处理电路,用于将分支处理电路发送的中间结果进行后续处理得到该计算指令的结果,将该计算指令的结果发送给所述控制器单元。

[0047] 在本公开的一些实施例中,运算单元的拓扑结构可以如图6所示,运算单元可以包括一个主处理电路和多个从处理电路。所述多个从处理电路呈阵列分布;每个从处理电路与相邻的其他从处理电路连接,所述主处理电路连接所述多个从处理电路中的k个从处理电路,所述k个从处理电路为:第1行的n个从处理电路、第m行的n个从处理电路以及第1列的m个从处理电路;

[0048] 所述k个从处理电路,用于在所述主处理电路以及多个从处理电路之间的数据以及指令的转发;

[0049] 所述主处理电路,用于将一个输入数据分配成多个数据块,将所述多个数据块中的至少一个数据块以及多个运算指令中的至少一个运算指令发送给所述k个从处理电路;

[0050] 所述k个从处理电路,用于转换所述主处理电路与所述多个从处理电路之间的数据;

[0051] 所述多个从处理电路,用于依据该运算指令对接收到的数据块执行运算得到中间结果,并将运算结果传输给所述K个从处理电路;

[0052] 所述主处理电路,用于将所述K个从处理电路发送的中间结果进行后续处理得到该计算指令的结果,将该计算指令的结果发送给所述控制器单元。在一些实施例中,所述主处理电路,具体用于将多个从处理电路发送的中间结果进行组合排序得到该计算指令的结果;

[0053] 在一些实施例中,主处理电路,具体用于将多个处理电路的发送的中间结果进行组合排序以及激活处理后得到该计算指令的结果。

[0054] 在一些实施例中,所述主处理电路包括:转换处理电路、激活处理电路、加法处理电路中的一种或任意组合;

[0055] 所述转换处理电路,用于对所述数据执行前序处理,具体为:将主处理电路接收的数据或中间结果执行第一数据结构与第二数据结构之间的互换;或将主处理电路接收的数据或中间结果执行第一数据类型与第二数据类型之间的互换;

[0056] 所述激活处理电路,用于执行所述后续处理,具体为执行主处理电路内数据的激活运算;

[0057] 所述加法处理电路,用于执行所述后续处理,具体为执行加法运算或累加运算。

[0058] 在一些实施例中,所述从处理电路包括:乘法处理电路;

[0059] 所述乘法处理电路,用于对接收到的数据块执行乘积运算得到乘积结果。

[0060] 在一些实施例中,所述从处理电路还包括:累加处理电路,所述累加处理电路,用于对该乘积结果执行累加运算得到该中间结果。

[0061] 在一些实施例中,所述树型模块为n叉树结构,所述n为大于等于2的整数。

[0062] 实例1:对神经网络参数进行数据冗余。

[0063] 首先将神经网络参数按照参数绝对值大小确定M个重要等级,第1、2……M重要等级,并将参数对应划入各重要等级。

[0064] 具体地,设定M+1个阈值,按照从大到小排序后的分别记为 $T_0, T_1, T_2, \dots$ 及 $T_M$ 。当神经网络参数的绝对值D满足 $T_{i-1} > D > T_i$ 时,该数据归为第i个重要等级,其中 $i = 1, 2, \dots, M$ ,  $T_0, T_1, T_2, \dots$ 及 $T_M$ 均为实数,且 $T_0 > T_1 > T_2 > \dots > T_M \geq 0$ 。即神经网络参数的绝对值 $T_0 > D > T_1$ 时,划入第1重要等级,神经网络参数的绝对值 $T_1 > D > T_2$ 时,划入第2重要等级,依次类推。

[0065] 对于第i重要等级参数中浮点型的一参数,共有 $x_i$ 个比特位,设置符号位与指数部分以及底数部分前 $y_i$ 位指定为重要比特位,其中 $x_i, y_i$ 均为正整数,且 $0 < y_i \leq x_i$ 。

[0066] 对于第i重要等级参数中的定点型的参数,共有 $x_i$ 个比特位,设置符号位和数值部分前 $z_i$ 位是重要比特位,其中 $x_i, z_i$ 均为正整数,且 $0 < z_i \leq x_i$ 。

[0067] 对第i重要等级参数中重要比特位采用数据备份的方式进行数据冗余,并且备份两份,对不重要比特位不进行冗余存储。当对第i重要等级中的参数进行读操作时,对于重要比特位同时读取原始数据和备份的两份数据,若出现相应数据不一致,则取相同的两份数据作为最终读取的数据,同时将第三份不一致的数据进行修复。当对第i重要等级参数进行写操作时,对于重要比特位同时写回到两个备份地址,同时保证原始数据中的数据和两

个备份数据一致。

[0068] 实例2:对稀疏神经网络参数进行数据冗余。

[0069] 本实例中稀疏神经网络参数分为两部分,分别是非零参数和非零参数位置。

[0070] 将非零参数位置设置为第1重要等级,并且将其所有的比特位标记为重要比特性,采用CRC校验码的方式进行冗余存储。当进行读操作时,读取存储CRC校验码并对原始数据进行CRC校验码计算,如果两个CRC校验码不一致,则按照存储的CRC校验码对数据进行修正。进行写操作时,同时存储原始数据和CRC校验码。

[0071] 将神经网络的非零参数按照参数绝对值大小设置重要等级,从第2重要等级开始依次设置M-1个重要等级。设定M个阈值,按照从大到小排序后的分别记为 $T_1, T_2, \dots$ 及 $T_M$ 。当非零参数的绝对值D满足 $T_{i-1} > D > T_i$ 时,该数据归为第i个重要等级,其中 $i = 2, 3, \dots, M$ ,  $T_1, T_2, \dots$ 及 $T_M$ 均为实数,且 $T_1 > T_2 > \dots > T_M \geq 0$ 。即非零参数的绝对值满足 $T_1 > D > T_2$ 时,划入第2重要等级,非零参数的绝对值满足 $T_2 > D > T_3$ 时,划入第3重要等级依次类推。

[0072] 对于第i重要等级参数中浮点型的一参数,共有 $x_i$ 个比特位,设置符号位与指数部分以及底数部分前 $y_i$ 位指定为重要比特位,其中 $x_i, y_i$ 均为正整数,且 $0 < y_i \leq x_i$ 。

[0073] 对于第i重要等级参数中的定点型的参数,共有 $x_i$ 个比特位,设置符号位和数值部分前 $z_i$ 位是重要比特位,其中 $x_i, z_i$ 均为正整数,且 $0 < z_i \leq x_i$ 。

[0074] 对第i重要等级参数中重要比特位,采用数据备份的方式进行数据冗余,并且备份两份,对不重要比特位不进行冗余存储。当对第i重要等级中的参数进行读操作时,对于重要比特位同时读取原始数据和备份的两份数据,若出现相应数据不一致,则取相同的两份数据作为最终读取的数据,同时将第三份不一致的数据进行修复。当对第i重要等级参数进行写操作时,对于重要比特位同时写回到两个备份地址,同时保证原始数据中的数据和两个备份数据一致。

[0075] 实施例3:图计算的应用中数据冗余。

[0076] 本实施中图计算应用中数据分为两部分,包括顶点数据和边数据。

[0077] 将图计算应用中的顶点数据设置为第1重要等级,并且将所有的数据比特位标记为重要比特性,采用CRC校验码的方式进行冗余存储。当进行读操作时,读取存储CRC校验码并对原始数据进行CRC校验码计算,如果两个CRC校验码不一致,则按照存储的CRC校验码对数据进行修正。进行写操作时,同时存储原始数据和CRC校验码。

[0078] 将图计算应用中边数据按照边的访问频率设置重要等级,从第2重要等级开始依次设置M-1个重要等级。按照从大到小排序后的分别记为 $T_1, T_2, \dots$ 及 $T_M$ 。当边数据访问频率F满足 $T_{i-1} > F > T_i$ 时,该数据归为第i个重要等级,其中 $i = 2, 3, \dots, M$ ,  $T_1, T_2, \dots$ 及 $T_M$ 均为实数,且 $T_1 > T_2 > \dots > T_M \geq 0$ ,即边数据访问频率满足 $T_1 > F > T_2$ 时,划入第2重要等级,边数据访问频率满足 $T_2 > F > T_3$ 时,划入第3重要等级依次类推。

[0079] 对于第i重要等级中浮点型的边数据,共有 $x_i$ 个比特位,设置符号位与指数部分以及底数部分前 $y_i$ 位指定为重要比特位,其中 $x_i, y_i$ 均为正整数,且 $0 < y_i \leq x_i$ 。

[0080] 对于第i重要等级参数中的定点型的边数据,共有 $x_i$ 个比特位,设置符号位和数值部分前 $z_i$ 位是重要比特位,其中 $x_i, z_i$ 均为正整数,且 $0 < z_i \leq x_i$ 。

[0081] 第i重要等级边数据中重要比特位采用数据备份的方式进行数据冗余,并且备份两份,对不重要比特位不进行冗余存储。当对第i重要等级边数据进行读操作时,对于重要

比特位同时读取原始数据和备份的两份数据,若出现数据不一致,则取相同的两份数据作为最终读取的数据,同时将第三份不一致数据修复。当对第*i*重要等级边数据进行写操作时,对于重要比特位同时写回到两个备份地址,同时保证原始数据中的数据和两个备份数据一致。

[0082] 本申请另一实施例提供一种数据冗余装置100,图2示出了数据冗余装置的结构框图,如图2所示,数据冗余装置100包括重要等级划分单元10、重要比特位提取单元20以及数据冗余处理单元30。

[0083] 其中,重要等级划分单元10用于根据重要性将数据划分为M个重要等级,M为正整数,具体地,数据的重要等级可以综合考虑数据的大小,数据绝对值的大小,数据的类型(浮点型、定点型),数据的读操作频率,数据写操作频率等因素进行设置。

[0084] 重要比特位提取单元20用于提取每一重要等级中的每一数据的重要比特位。重要比特位提取单元20会识别不同重要等级的数据,并将数据比特位分为重要数据比特位和不重要数据比特位。同时对每一个重要等级的每一数据,提取重要比特位。

[0085] 数据冗余处理单元30用于针对所述重要比特位进行数据冗余处理。

[0086] 如图2所示,数据冗余处理单元30包括冗余存储单元31和读写控制单元32。

[0087] 冗余存储单元31可以对原始数据进行存储,并且对数据中的重要比特位进行数据冗余存储。数据冗余可以是副本备份或者纠错编码。副本可以同时备份N份,其中N是大于0的正整数。纠错编码方式包括但不限于CRC校验,ECC校验。冗余存储单元31可以是硬盘,DRAM,SRAM,ECC-DRAM,ECC-SRAM和非易失性内存。

[0088] 读写控制单元32能够对冗余的数据进行读写操作,保证读写数据的一致性。

[0089] 前面的附图中所描绘的进程或方法可通过包括硬件(例如,电路、专用逻辑等)、固件、软件(例如,被承载在非瞬态计算机可读介质上的软件),或两者的组合的处理逻辑来执行。虽然上文按照某些顺序操作描述了进程或方法,但是,应该理解,所描述的某些操作能以不同顺序来执行。此外,可并行地而非顺序地执行一些操作。

[0090] 需要说明的是,在附图或说明书正文中,未绘示或描述的实现方式,均为所属技术领域中普通技术人员所知的形式,并未进行详细说明。此外,上述对各元件和方法的定义并不仅限于实施例中提到的各种具体结构、形状或方式,本领域普通技术人员可对其进行简单地更改或替换。

[0091] 以上所述的具体实施例,对本申请的目的、技术方案和有益效果进行了进一步详细说明,应理解的是,以上所述仅为本申请的具体实施例而已,并不用于限制本申请,凡在本申请的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

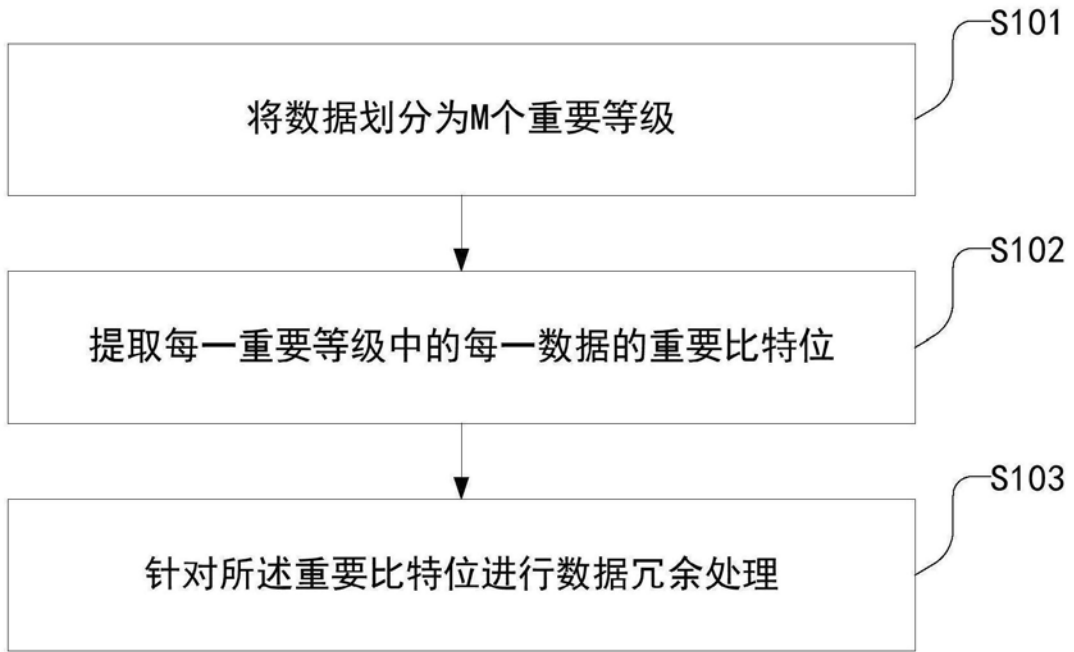


图1

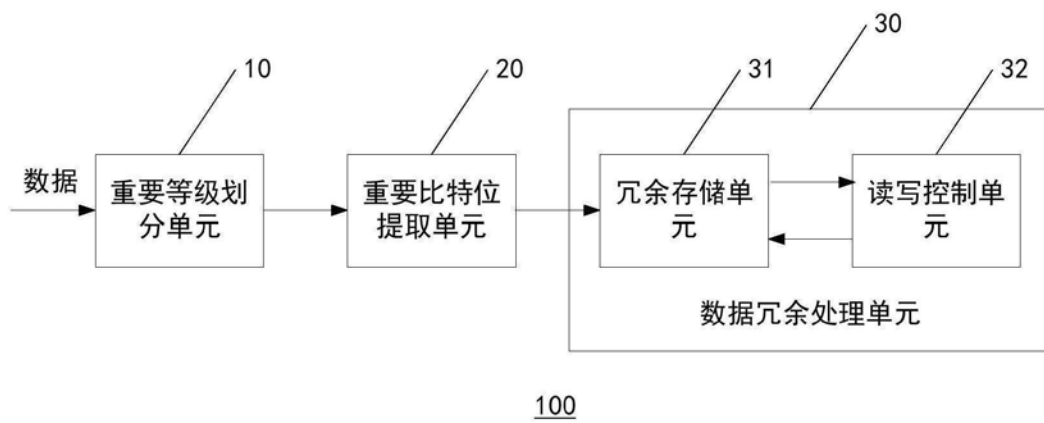


图2

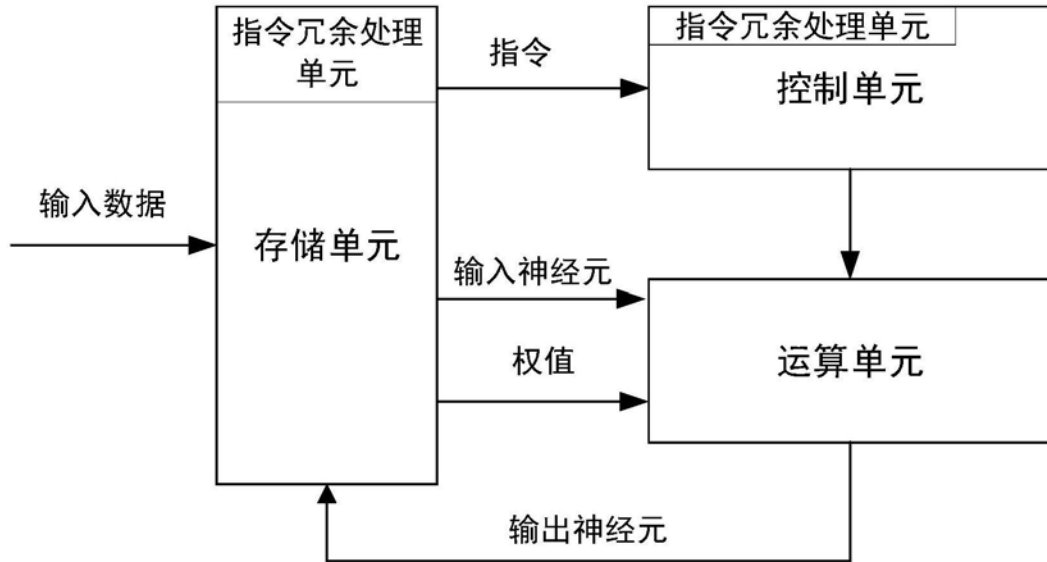


图3

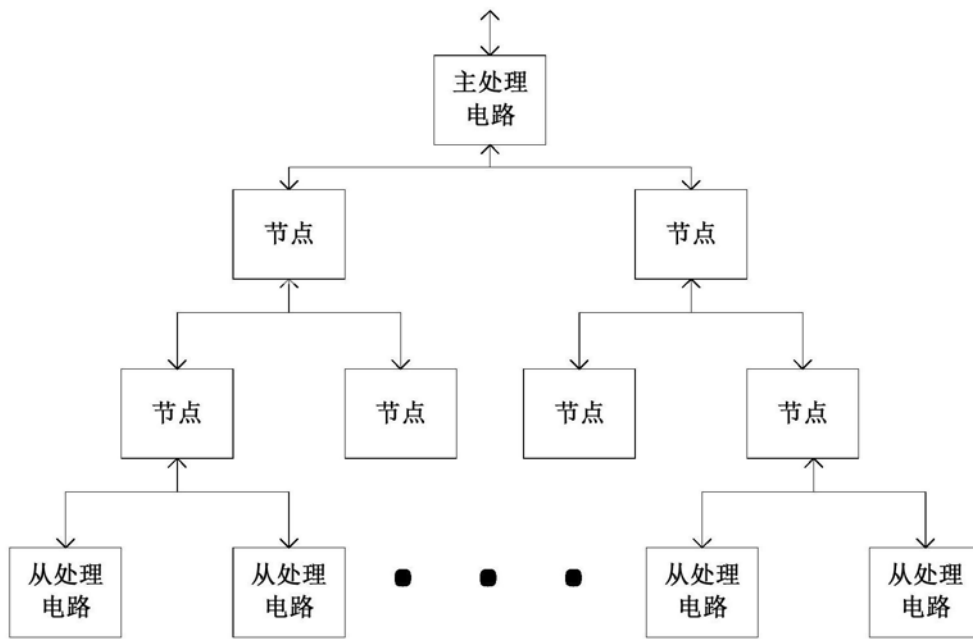


图4

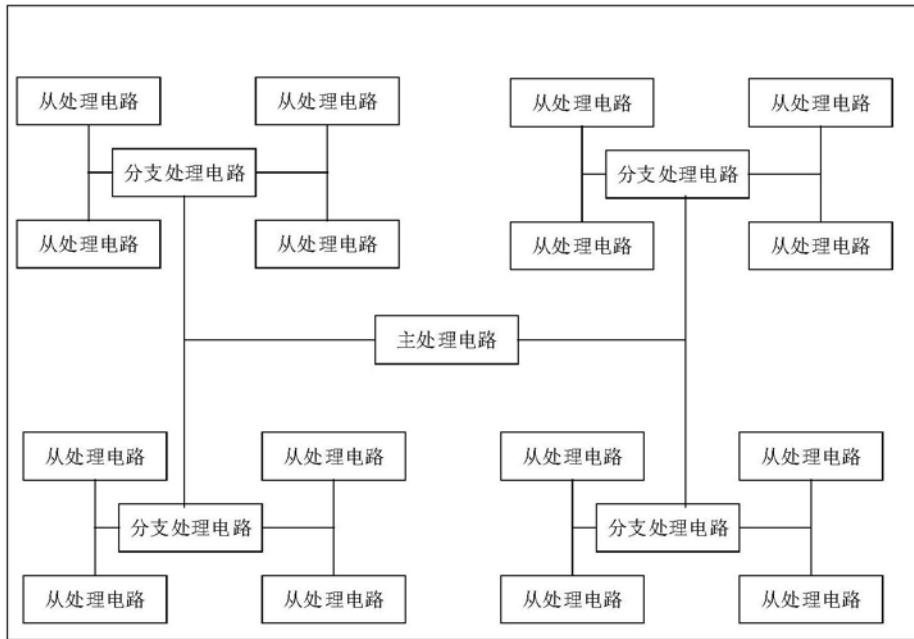


图5

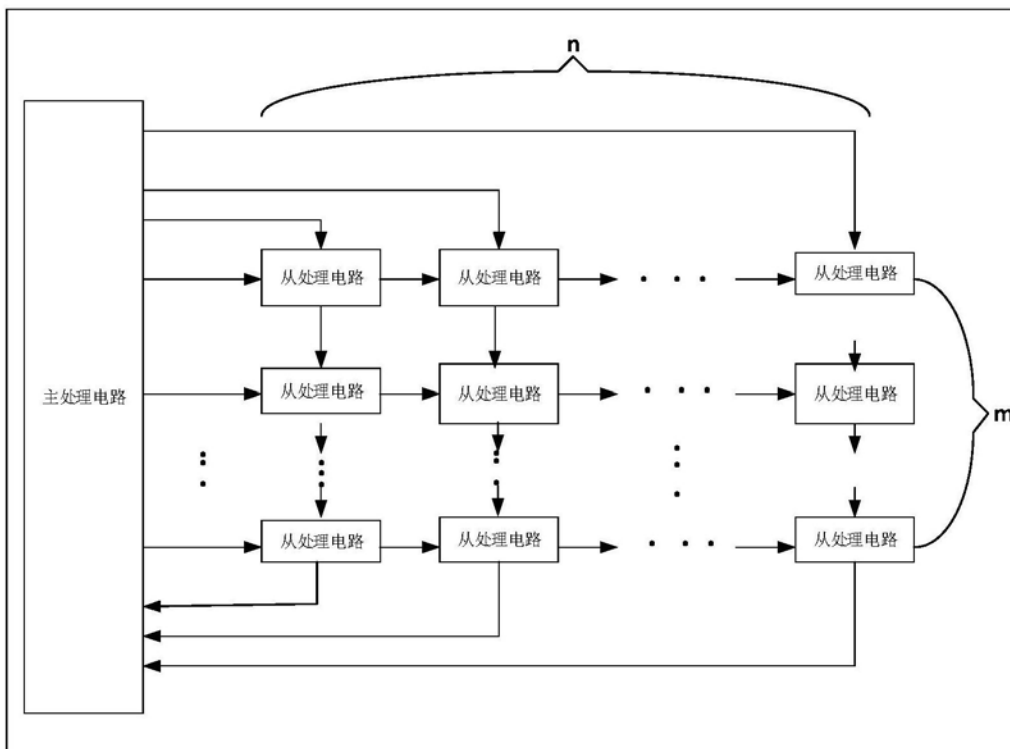


图6