



(12)发明专利申请

(10)申请公布号 CN 110322010 A

(43)申请公布日 2019.10.11

(21)申请号 201910588964.5

(22)申请日 2019.07.02

(71)申请人 深圳忆海原识科技有限公司

地址 518116 广东省深圳市龙岗区坂田街
道岗头社区天安云谷产业园一期2栋
201

(72)发明人 任化龙

(51)Int.Cl.

G06N 3/063(2006.01)

G06N 3/08(2006.01)

G06N 3/10(2006.01)

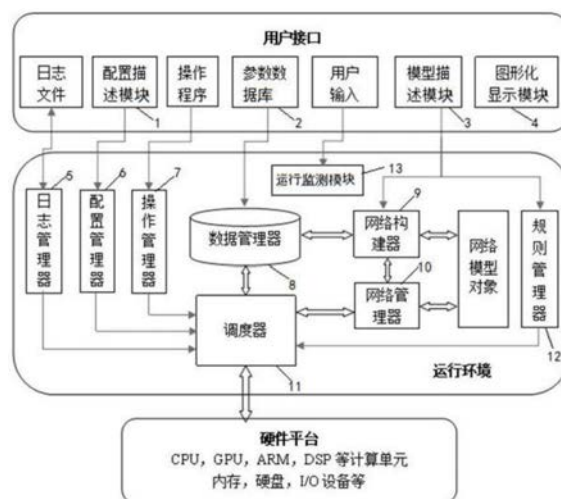
权利要求书6页 说明书16页 附图3页

(54)发明名称

用于类脑智能与认知计算的脉冲神经网络
运算系统及方法

(57)摘要

本发明公开了一种用于类脑智能与认知计算的脉冲神经网络运算系统及方法,该系统包括模型描述模块、参数数据库、配置描述模块、配置管理器、规则管理器、数据管理器、网络构建器、网络管理器、操作管理器、调度器、日志管理器、运行监测模块和图像化显示模块,该系统提供了按照一定条件和规则自动执行突触和神经元裁剪与新生的功能,并且提供了多种灵活的启动裁剪和新生过程的触发条件与执行过程的规则,免去了神经网络开发者需要自行编写突触和神经元裁剪与新生程序的负担,从而有效解决了现有的类脑脉冲神经网络运算框架存在的一些问题。



1. 一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,包括:
模型描述模块,用于供用户设计和描述网络模型的接口,为网络模型对象指定需要执行的操作和规则;
参数数据库,用于以参数数据库的形式存储网络模型的各项参数数据;
配置描述模块,用于描述当前网络运行环境的配置参数,启动突触和神经元的裁剪与新生过程的条件;
配置管理器,用于从所述配置描述模块中调取相关配置参数;
网络构建器,用于读取模型描述模块,解析网络的拓扑结构,并通过数据管理器读取数据文件,在内存中构建网络模型对象;
网络管理器,用于构建、遍历、访问和更新网络模型对象;
规则管理器,用于读取所述模型描述模块中用户声明的规则,并在调度器调度所述网络模型对象的运算时对用户声明的规则进行解释,并仲裁规则间的冲突;
数据管理器,用于读取并解析参数数据库,转换数据格式以及将数据序列化;
调度器,用于分配硬件资源和调度运算过程,调度执行相应的操作;
操作管理器,用于管理所有运行的操作;
日志管理器,用于记录系统运行时生成的日志,记录系统的工作状态,并对异常状态进行提示;
运行监测模块,用于接收并响应用户输入指令,管理整个系统的运行状态;以及
图形化显示模块,用于读取并显示网络数据。
2. 根据权利要求1所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述模型描述模块包括网络描述单元、汇聚描述单元和流通描述单元;
所述网络描述单元用于描述网络容器和一般参数容器,描述整个网络的参数与运行规则,并通过链接指向一个或多个汇聚描述单元和流通描述单元;
所述汇聚描述单元用于描述汇聚容器、模块容器、层组容器、节点容器、节点参数容器和一般参数容器,并用于描述网络中节点的模块与层组划分关系、各个网络模型对象的参数与运行时的规则和命令;
所述流通描述单元用于描述流通容器、通道容器、连接容器、边容器、边参数容器和一般参数容器,并用于描述网络中边的连接关系、各个网络模型对象的参数与运行时的规则和命令。
3. 根据权利要求2所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述模型描述单元采用模拟生物脑神经系统组织方式的多层级树状结构的网络描述方式;所述汇聚描述单元支持将网络中的节点按照一定层组和模块排列组织,用于表征生物脑中神经元及相关胶质细胞的多层级组织方式;所述流通描述单元支持将网络中的边按照拓扑相似性分组和层级排列组织,用于表征生物脑中神经突触的多种组织方式及相关胶质细胞的突起的组织方式。
4. 根据权利要求2所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述网络描述单元、汇聚描述单元和流通描述单元均选用XML或JSON的文件格式并采用嵌套式语法。
5. 根据权利要求1所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其

特征在于,所述参数数据包括初始化参数数据和运行时参数数据。

6. 根据权利要求1所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述参数数据库为二进制文件或文本文件,所述文本文件采用CSV文件格式或以其他字符为分隔数据的文件格式。

7. 根据权利要求1所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述网络模型对象包括容器、拓扑关系和参数数据,所述网络模型对象是调度器调度运行的对象。

8. 根据权利要求7所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述容器包括编号和名称,用于在多层级树状结构中索引。

9. 根据权利要求7所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述容器具有一控制块,用于存储统计和控制信息。

10. 根据权利要求9所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述控制块包括网络的遍历次序与规则、已参与遍历运算的次数、数据是否已入驻主内存、数据是否已入驻协处理器内存、硬盘读写的频次,并由所述规则管理器和调度器管理和更新。

11. 根据权利要求8或9所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述容器包括:

网络容器,位于树状结构的最顶级,用于表征全脑和行为尺度的模型;

汇聚容器,位于树状结构的第二层级,用于表征脑区尺度的模型;

模块容器,位于树状结构的第三层级,用于表征神经核团尺度的模型;

层组容器,位于树状结构的第四层级,用于表征神经环路尺度的模型;

节点容器,位于树状结构的第五层级,用于表征神经元尺度或胶质细胞尺度的模型,也用于表征一群神经元或胶质细胞;

节点参数容器,位于树状结构的第最低级,用于表征分子尺度、受体尺度、神经递质或神经调质尺度的模型,也用于表征一群神经元或胶质细胞模型参数张量;

流通容器,位于树状结构的第二层级,用于表征联接脑区间的神经纤维束尺度的模型;

通道容器,位于树状结构的第三层级,用于表征由联接神经核团间的轴突构成的传导束的模型;

连接容器,位于树状结构的第四层级,用于表征神经环路中由轴突构成的神经通路的模型;

边容器,位于树状结构的第五层级,用于表征树突尺度或突触尺度的模型,也用于表征一群突触或胶质细胞的突起;

边参数容器,位于树状结构的最低级,用于表征分子尺度、神经递质或神经调质尺度、受体尺度的模型,也用于表征一群突触或胶质细胞的突起模型参数张量;

一般参数容器,属于辅助容器;

根据建模的需要,各个层级的容器都额外带有一个或多个一般参数容器,用于以张量的形式容纳参数数据。

12. 根据权利要求11所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述神经元模型的发放特性构建为紧张性发放、快速发放、爆发式发放、峰发

放或相位性发放；

所述神经元模型对于上游输入信号的响应可以构建为不同的神经适应性或敏感性曲线；

所述神经元模型对于下游的作用可以构建为兴奋型的、抑制型的、调制型的或中性的；

所述神经元模型构建为脉冲神经元模型或传统神经元模型；

所述胶质细胞模型构建为星形胶质细胞模型、少突胶质细胞模型、小胶质细胞模型、施万细胞模型和卫星细胞模型。

13. 根据权利要求11所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，所述神经递质或神经调质模型构建为兴奋型的、抑制型的或调制型的；

所述受体模型可构建为离子型的或代谢型的；

所述受体模型对于神经递质或神经调质的响应效果构建为兴奋型的、抑制型的、调制型的或中性的。

14. 根据权利要求11所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，所述树突尺度的模型构建为顶树突模型、基树突模型或突棘模型；

所述突触模型构建为兴奋型的、抑制型的、调制型的或中性的。

15. 根据权利要求11所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，所述分子尺度的模型构建为细胞内分子模型、细胞膜中分子模型和细胞间隙分子模型。

16. 根据权利要求11所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，所述节点参数容器、边参数容器以及一般参数容器内部可采用张量的形式容纳参数。

17. 根据权利要求16所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，所述张量的维度可以是一维或多维，具体排列和使用方式由用户指定。

18. 根据权利要求17所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，所述张量配置为四维，各个参数在张量中的位置由坐标 (x, y, z, t) 表示，其中， x 、 y 、 z 三个维度对应于父级容器中表征的各个神经组织的空间排布位置； t 表示时间维度，表征时序信息的缓存和延迟，用于模拟神经调质对神经元和突触的长时间作用机制；

所述张量中的参数可以由父级容器中全体或部分神经组织共享，用于模拟神经调质对目标区域中的全部神经组织的大面积作用。

19. 根据权利要求11所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，所述流通容器及其所有子级容器均对应于一个上游容器和一个下游容器，并通过所述上游容器和所述下游容器的编号或名称对其进行索引访问。

20. 根据权利要求19所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，所述上游容器和下游容器均是任意层级的容器，二者为同一个或者不同的容器。

21. 根据权利要求11所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，所述流通容器及其所有子级容器均与其上游容器和下游容器构成信息流动通路，表征信息在两个信息源间的流动和处理过程，网络中的多个容器间可构成信息流动的任意拓扑结构。

22. 根据权利要求21所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统, 其特征在于, 所述信息的流动和处理过程用于实现多种生物脑神经机制。

23. 根据权利要求22所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统, 其特征在于, 多种所述生物脑神经机制包括神经脉冲通过突触在神经元间传导、突触和突触之间交换信息以及神经元与突触可塑性。

24. 根据权利要求21所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统, 其特征在于, 所述信息流动的任意拓扑结构, 用于实现脑神经系统中任意的神经环路连接方式, 包括支持同一个神经元连回自己的反馈连接、同一群的神经元互相之间的连接、不同群神经元之间的任意连接, 以及突触和突触的直接连接, 且允许反馈型连接的无限次循环计算。

25. 根据权利要求1所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统, 其特征在于, 所述模型描述模块支持将任何层级的模型分解为数据和操作两部分的建模设计方式,

所述数据可由节点参数容器、边参数容器或一般参数容器容纳, 并由对应的参数数据库存储;

所述操作用于访问和更新所述数据的可执行程序, 所述操作运行在通用CPU、ARM、DSP、GPU或其它处理器, 保证系统具有一定跨硬件平台通用性。

26. 根据权利要求1所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统, 其特征在于, 所述模型描述模块支持用户通过定义一种或多种操作, 使同一个节点容器中的各个神经元直接互相访问和更新彼此的数据, 以实现信息的快速交换, 用于模拟生物脑神经网络系统中的电突触。

27. 根据权利要求1所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统, 其特征在于, 所述模型描述模块支持用户通过定义一种或多种操作, 使同一个边容器中的各个突触直接互相访问和更新彼此的数据, 以实现信息的快速交换, 用于模拟生物脑神经网络中同一个神经元的树突上多个突触间互相交换信息并进行逻辑运算的情况, 包括分流抑制机制。

28. 根据权利要求1所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统, 其特征在于, 所述系统支持按照一定触发条件和执行规则自动执行突触和神经元的裁剪与新生的功能;

所述触发条件由用户在所述配置描述模块中指定;

所述执行规则由用户在所述模型描述模块中指定;

所述执行规则作用于整个网络模型对象, 也作用于子网络或特定容器;

所述突触和神经元的裁剪与新生过程由调度器调度执行, 在网络运行状态时执行或在网络暂停状态时执行。

29. 根据权利要求28所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统, 其特征在于, 所述触发条件包括下面的一种或多种:

用户命令: 用户通过键盘或鼠标及其它方式输入命令给系统, 系统接收到命令后即刻或稍后执行裁剪或新生过程;

持续执行: 只要网络模型或其子区域达到符合裁剪或新生过程的规则就执行裁剪或新

生过程；

间隔执行：系统每间隔一定时间或遍历周期，适时地自动启动裁剪或新生过程。

30. 根据权利要求28所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，所述裁剪过程的执行规则包括突触裁剪规则和神经元裁剪规则；

所述突触裁剪规则包括下面的一种或任几种：

某个突触的参数和指定参照突触集合中所有突触参数的统计量达到一定数值关系，则该突触应被裁剪；

某个突触的参数和指定阈值达到一定数值关系则该突触应被裁剪；

某个突触超过一定时间或遍历周期没有被触发，则该突触应被裁剪；

某个突触被其它运算过程标记为可以被裁剪，则该突触应被裁剪；

所述神经元裁剪规则包括下面的一种或任几种：

某个神经元没有输入的突触，则该神经元应被裁剪；

某个神经元没有输出的突触，则该神经元应被裁剪；

某个神经元没有输入及输出的突触，则该神经元应被裁剪；

某个神经元的参数和指定参照神经元集合中所有神经元参数的统计量达到一定数值关系，则该神经元应被裁剪；

某个神经元的参数和指定阈值达到一定数值关系，则该神经元应被裁剪；

某个神经元超过一定时间或遍历周期没有发放，则该神经元应被裁剪；

某个神经元被其它运算过程标记为可以被裁剪，则该神经元应被裁剪。

31. 根据权利要求28所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，所述新生过程的执行规则，包括神经元新生规则和突触新生规则；

所述神经元新生规则包括下面的一种或任几种：

某个节点容器的现存神经元数量和该节点容器的总容量达到一定比例或数值关系，则以其总容量的一定比例或数量新生神经元；

某个节点容器按照一定速率，即每间隔一定时间或遍历周期以其总容量的一定比例或数量新生神经元；

某个节点容器被其它运算过程标记为需要新生神经元，并按照一定速率新生神经元；

所述突触新生规则包括下面的一种或任几种：

某个边容器的现存突触数量和该边容器的总容量达到一定比例或数值关系，则以其总容量的一定比例或数量新生突触；

某个边容器按照一定速率新生突触；

某个边容器被其它运算过程标记为需要新生突触，并按照一定速率新生突触；

某个节点容器存在没有输入或输出突触的神经元，则在对应的各个边容器中分别为其新生输入突触或输出突触。

32. 根据权利要求1所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，系统的运行状态包括默认状态、网络构建状态、网络运行状态和网络暂停状态。

33. 根据权利要求1所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统，其特征在于，所述模型描述模型中，用户为各个容器指定一条或多条规则，全部规则构成了一个规则库；

所述规则管理器对所述规则库中的规则按照一定优先级进行排序,当作用于一个容器的多条规则互相之间发生冲突时,只有最高优先级的规则被执行,当一个容器没有指定任何规则时,规则管理器采用默认规则执行;

所述规则库中的规则包括:遍历规则、内存使用规则、数据I/O规则、突触与神经元的裁剪与新生规则;

所述遍历规则用于指导调度器按照一定时间间隔或遍历周期重复遍历或跳过遍历网络的全部或特定容器,以将计算资源集中在需要计算密集的子网络,提高数据利用效率;

所述内存使用规则可以用于指导调度器合理安排主内存和协处理器内存的使用;

所述数据I/O规则用于指导所述调度器调度数据在主内存和协处理器内存间,以及内存与硬盘间交换的频次。

34. 根据权利要求1所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述调度器管理一个主内存池和一个或多个设备内存池;

所述主内存池用于管理主内存的使用;

所述设备内存池对应于各个协处理器,用于管理相应设备内存的使用;

所述主内存池和设备内存池的容量上限和下限由用户通过配置描述模块指定。

35. 根据权利要求1或34所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述调度器管理一个线程池,用于动态安排子线程参与多线程运算,以合理安排主计算单元和协处理器以及I/O设备的运算负载。

36. 根据权利要求35所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,所述调度器管理一个节点数据输入缓存器、一个节点数据输出缓存器、一个边数据输入缓存器以及一个边数据输出缓存器,用于缓存需要读写硬盘或I/O设备的数据,以使调度器能够根据处理器、硬盘和I/O设备的负载,适时地安排硬盘、I/O设备读写,避免I/O阻塞。

37. 根据权利要求36所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统,其特征在于,各个所述缓存器的容量、读写硬盘或I/O设备的频次上限和下限、读写硬盘或I/O设备的吞吐量上限和下限,由用户通过配置描述模块指定。

38. 一种用于类脑智能与认知计算的脉冲神经网络运算方法,其特征在于,所述方法使用如权利要求1至37任一项所述的一种用于类脑智能与认知计算的脉冲神经网络运算系统。

用于类脑智能与认知计算的脉冲神经网络运算系统及方法

技术领域

[0001] 本发明涉及类脑脉冲神经网络模拟和高性能计算技术领域,更具体的说是涉及一种用于类脑智能与认知计算的脉冲神经网络运算系统及方法。

背景技术

[0002] 目前,类脑智能和认知计算以脉冲神经网络为基础,结合生物脑中的多种神经递质、神经调质、受体、电突触、化学突触、树突、神经元、胶质细胞的丰富工作机制进行计算建模,构造成的神经环路、神经核团、脑区和全脑模型,能够模拟生物脑的诸多认知机制和行为,如记忆与学习、模拟情感、导航与计划、运动控制、类脑视觉和类脑听觉、注意力、决策等等,为人工智能系统的发展提供了更广阔的路线。

[0003] 但是,现有的类脑脉冲神经网络运算框架存在着如下问题:

[0004] 1、没有提供统一而灵活的建模方式,也不能支持灵活的网络拓扑结构,因而不能在建模尺度和建模丰富度之间取得统一。例如,在建模尺度方面,现有的类脑脉冲神经网络运算框架大多不能支持并统一在分子尺度、神经递质与神经调质尺度、受体尺度、突触尺度、树突尺度、神经元尺度、胶质细胞尺度、神经环路尺度、神经核团尺度、神经纤维束尺度、脑区尺度、全脑尺度和行为尺度的建模。又如,在建模丰富度方面,现有的类脑脉冲神经网络运算框架往往忽略了支持电突触的建模,也忽略了支持模拟神经调质的工作机制,还忽略了支持模拟树突中多个突触间互相交换信息并进行逻辑运算的机制,也不能支持突触和突触直接连接的拓扑结构等等。

[0005] 2、没有提供模仿生物脑神经系统组织方式的描述方法。

[0006] 3、没有集成足够丰富的内置功能和灵活的用户接口,如不能支持自动执行突触与神经元的裁剪与新生,需要用户自行编写程序实现相关功能。

[0007] 4、不能将脉冲神经网络、传统神经网络、以及其它以张量为主要数据表征方式的算法有效地统一结合进行混合运算。

[0008] 5、不能兼容CPU、ARM、GPU、DSP等芯片架构,对硬件资源的利用也大多不够优化,不适合将中等至大规模类脑神经网络高效运行于通用计算机和嵌入式设备。

[0009] 上述问题的存在导致现有的类脑脉冲神经网络运算框架硬件平台建模范围有限、兼容性差、运算效率低、开发使用不便捷,很难实现将类脑智能和认知计算低成本大规模地部署于智能玩具、机器人、无人机、IOT设备、智能家居、车载系统等产品中。

[0010] 因此,如何提供一种灵活高效的用于类脑智能与认知计算的脉冲神经网络运算系统及方法是本领域技术人员亟需解决的问题。

发明内容

[0011] 有鉴于此,本发明提供了一种用于类脑智能与认知计算的脉冲神经网络运算系统及方法,该系统提供了统一而灵活的建模方式,而且提供的多层级树状结构的网络描述方式支持生物脑神经系统全尺度的建模和灵活的网络拓扑结构,从而将建模尺度和建模丰富

度有机统一,并将各个尺度的全部模型融合为统一的神经网络进行运算,并且支持将数据以张量的形式表征和存储,这也使本系统既可以支持脉冲神经网络,也可以兼容传统神经网络(深度学习)以及其它以张量为主要数据表征方式的算法,还提供了按照一定条件和规则自动执行突触和神经元裁剪与新生的功能,免去了神经网络开发者需要自行实现相关功能的负担,从而有效解决了现有的类脑脉冲神经网络运算框架存在的上述问题。

[0012] 为了实现上述目的,本发明采用如下技术方案:

[0013] 一种用于类脑智能与认知计算的脉冲神经网络运算系统,所述系统包括:模型描述模块、参数数据库、配置描述模块、配置管理器、规则管理器、数据管理器、网络构建器、网络管理器、操作管理器、调度器、日志管理器、运行监测模块和图像化显示模块;

[0014] 所述模型描述模块用于提供用户设计和描述网络模型的接口;

[0015] 所述参数数据库用于存储网络的各项参数数据,包括初始化参数和运行时参数;参数数据库可选为二进制文件或文本文件;文本文件可采用CSV文件格式或以其它字符分隔数据的文件格式;

[0016] 所述配置描述模块用于描述当前网络运行环境的配置参数和启动突触和新神经元的裁剪与新生的条件;

[0017] 所述配置管理器用于读取上述配置描述模块,以获取系统配置参数;

[0018] 所述网络模型对象由网络构建器构造并驻留在内存中,用于表征整个网络,包括所有容器、拓扑关系和参数数据,是调度器调度运行的对象;

[0019] 所述规则管理器用于读取模型描述模块中用户声明的规则,并在调度器调度网络模型对象的运算时对这些规则进行解释以及仲裁规则间的冲突;

[0020] 所述数据管理器包括一至多种解码器和编码器,用于读取、解析参数数据库,转换数据格式,以及将数据序列化;用户可以在数据管理器中添加自定义的解码器和编码器,以读写自定义格式的文件;

[0021] 所述网络构建器用于读取模型描述模块,解析网络的拓扑结构,并通过数据管理器读取数据文件,在内存中构建网络模型对象;

[0022] 所述网络管理器用于构建、遍历、访问和更新网络模型对象;

[0023] 所述操作管理器用于管理所有可在本系统上运行的操作;所有的操作构成了操作库;用户可在模型描述模块中为各个容器指定需要执行的操作,运行时由调度器调度执行相应的操作;

[0024] 所述调度器用于分配硬件资源和调度运算过程,优化运算效率;

[0025] 所述日志管理器用于记录系统运行时生成的日志,提示用户系统的工作状态与异常,以便于调试和维护;

[0026] 所述运行监测模块用于接收和响应用户输入并管理整个系统的运行状态,包含有默认状态、网络构建状态、网络运行状态和网络暂停状态;

[0027] 所述图像化显示模块用于读取网络数据,并将其显示给用户,以便于开发、监测和调试。

[0028] 进一步地,所述模型描述模块包括网络描述单元、汇聚描述单元、流通描述单元,共同用于描述整个网络的各个构成部分与拓扑结构;它们可以选为文本文件,采用嵌套式语法,可以选用XML或JSON的文件格式。

[0029] 更进一步地,所述模型描述模块采用模拟生物脑神经系统组织方式的多层级树状结构的网络描述方式;

[0030] 所述汇聚描述单元支持将网络中的节点按照一定层组和模块排列组织,用于表征生物脑中神经元及相关胶质细胞的多层级组织方式(如,核团→脑区→全脑);

[0031] 所述流通描述单元支持将网络中的边按照拓扑(连接关系)相似性分组和层级排列组织,用于表征生物脑中神经突触的多种组织方式(如树突、神经通路中的投射、神经纤维束等)及相关胶质细胞的突起的组织方式。

[0032] 进一步地,所述网络描述单元用于描述Network及Param等容器、描述整个网络的参数与运行规则,并通过链接指向一至多个汇聚描述单元和流通描述单元;

[0033] 所述汇聚描述单元用于描述Confluence、Module、Layer、Node、NodeParam和Param等容器,并用于描述网络中节点的模块与层组划分关系、各个容器的参数与运行时的规则和命令;

[0034] 所述流通描述单元用于描述Flow、Channel、Link、Edge、EdgeParam和Param等容器,并用于描述网络中边的连接(拓扑)关系、各个容器的参数与运行时的规则和命令。

[0035] 更进一步地,所述Network表示网络容器,位于树状结构的第一层级(最顶级),用于表征全脑和行为尺度的模型,每个Network可以容纳一至多个Confluence和Flow;

[0036] 所述Confluence表示汇聚容器,位于树状结构的第二层级,可用于表征脑区尺度的模型,每个Confluence可以容纳一至多个Module;

[0037] 所述Module表示模块容器,位于树状结构的第三层级,可用于表征神经核团尺度的模型,每个Module可以容纳一至多个Layer;

[0038] 所述Layer表示层组容器,位于树状结构的第四层级,可用于表征神经环路尺度的模型,每个Layer可以容纳一至多个Node;

[0039] 所述Node表示节点容器,位于树状结构的第五层级,可用于表征神经元尺度或胶质细胞尺度的模型,也可用于表征一群神经元或胶质细胞,每个Node可以容纳一至多个NodeParam;

[0040] 所述Node还可用于表征输入、输出节点,用于对接系统的I/O设备,如摄像头输入、音频输入、传感器输入、控制输出等,读写I/O设备的数据通过该Node的各个NodeParam动态更新;

[0041] 所述NodeParam表示节点参数容器,位于树状结构的第六层级(最低级),可用于表征分子尺度、受体尺度、神经递质或神经调质尺度的模型,也可用于表征一群神经元或胶质细胞模型的参数张量;

[0042] 所述Flow表示流通容器,位于树状结构的第二层级,可用于表征联接脑区间的神经纤维束尺度的模型,每个Flow可以容纳一至多个Channel;

[0043] 所述Channel表示通道容器,位于树状结构的第三层级,可用于表征由联接神经核团间的轴突构成的传导束的模型,每个Channel可以容纳一至多个Link;

[0044] 所述Link表示连接容器,位于树状结构的第四层级,可用于表征神经环路中由轴突构成的神经通路的模型,每个Link可以容纳一至多个Edge;

[0045] 所述Edge表示边容器,位于树状结构的第五层级,可用于表征树突尺度或突触尺度的模型,也可用于表征一群突触或胶质细胞的突起,每个Edge可以容纳一至多个

EdgeParam;

[0046] 所述EdgeParam表示边参数容器,位于树状结构的第六层级(最低级),可用于表征分子尺度、神经递质或神经调质尺度、受体尺度的模型,也可用于表征一群突触或胶质细胞的突起模型的参数张量;

[0047] 所述Param表示一般参数容器,属于辅助容器。根据建模的需要,上述各层级的容器都可以额外带有一至多个Param,用于以张量的形式容纳参数数据;

[0048] 上述每个容器都有编号和名称,用于在多层级树状结构中索引;

[0049] 上述每个容器都有一个控制块(Control Block)用于存储统计和控制信息,包括网络的遍历次序与规则、已参与遍历运算的次数、数据是否已入驻主内存、数据是否已入驻协处理器内存、硬盘读写的频次等,并由规则管理器和调度器管理和更新。

[0050] 更进一步地,所述神经元模型的发放特性可以构建为紧张性发放、快速发放、爆发式发放、峰发放或相位性发放等;

[0051] 所述神经元模型对于上游输入信号的响应可以构建为不同的神经适应性或敏感性曲线;

[0052] 所述神经元模型对于下游的作用可以构建为兴奋型的、抑制型的、调制型的或中性的;

[0053] 所述神经元模型可以构建为脉冲神经元模型,也可以构建为传统神经元模型;

[0054] 所述胶质细胞模型可以构建为星形胶质细胞、少突胶质细胞、小胶质细胞、施万细胞和卫星细胞模型。

[0055] 更进一步地,所述神经递质或神经调质模型可以构建为兴奋型的、抑制型的或调制型的;

[0056] 所述受体模型可构建为离子型的或代谢型的;

[0057] 所述受体模型对于神经递质或神经调质的响应效果可以构建为兴奋型的、抑制型的、调制型的或中性的。

[0058] 更进一步地,所述树突尺度的模型可以构建为顶树突模型、基树突模型或突棘模型;

[0059] 所述突触模型可以构建为兴奋型的、抑制型的、调制型的或中性的。

[0060] 更进一步地,所述分子尺度的模型可以构建为细胞内分子模型、细胞膜中分子模型和细胞间隙分子模型。

[0061] 更进一步地,所述NodeParam、EdgeParam以及Param内部可采用张量(即多维矩阵)的形式容纳参数,

[0062] 所述张量的维度可以是1维至多维,具体排列和使用方式由用户指定;

[0063] 所述张量可以配置为4维,各个参数在张量中的位置可以由坐标(x,y,z,t)表示,其中x,y,z三个维度对应于父级容器中表征的各个神经组织(如神经元或突触等)的空间分布位置,t代表时间维度,可表征时序信息的缓存和延迟,可用于模拟神经调质对神经元和突触的长时间作用机制(具有延迟性);

[0064] 所述张量中的参数可以由父级容器中全体或部分神经组织(如神经元或突触等)共享,可用于模拟神经调质对目标区域中的全部神经组织的大面积作用。

[0065] 进一步地,所述Flow及其所有子级容器都可以对应于一个上游容器和一个下游容

器,并通过上、下游容器的编号或名称对它们进行索引访问,

[0066] 所述上游容器和下游容器都可以是任意层级的容器,二者可以是同一个或者不同的容器;

[0067] 所述Flow及其所有子级容器都可与其上、下游容器构成信息流动通路,表征信息在两个信息源(如上游容器和下游容器)间的(单向或双向)流动和处理过程,网络中的多个容器间可构成信息流动的任意拓扑结构。

[0068] 更进一步地,所述信息的流动和处理过程可用于实现神经脉冲通过突触在神经元间传导、突触和突触之间交换信息、神经元与突触可塑性等多种生物脑神经机制。

[0069] 进一步地,所述信息流动的任意拓扑结构,可用于实现脑神经系统中任意的神经环路连接方式,包括支持同一个神经元连回自己的反馈连接、同一群(层)的神经元互相之间的连接、不同群(层)神经元之间的任意连接(顺次/前馈、跨层、反馈等),以及突触和突触的直接连接,并且允许反馈型连接的无限次循环计算。

[0070] 更进一步地,所述系统支持将任何层级(或尺度)的模型分解为数据(data)和操作(operation)两部分的建模设计方式,

[0071] 所述数据可由NodeParam、EdgeParam或Param容纳,并由对应的参数数据库存储;

[0072] 所述操作是用于访问和更新前述数据的可执行程序(如函数及包含函数的类),操作可以运行在通用CPU、ARM、DSP、GPU或其它处理器,以保证所述系统具有一定跨硬件平台通用性。

[0073] 更进一步地,所述系统支持用户通过定义一至多种操作,使同一个Node中的各个神经元(无需通过Edge)直接互相访问和更新彼此的数据,以实现信息的快速交换,用于模拟生物脑神经系统中的电突触。

[0074] 更进一步地,所述系统支持用户通过定义一至多种操作,使同一个Edge中的各个突触直接互相访问和更新彼此的数据,以实现信息的快速交换,用于模拟生物脑神经系统中同一个神经元的树突上多个突触间互相交换信息并进行逻辑运算的情况,包括分流抑制(shunting inhibition)等机制。

[0075] 进一步地,所述系统支持按照一定触发条件和执行规则自动执行突触和神经元的裁剪与新生的功能,

[0076] 所述触发条件可以由用户在所述配置描述模块中指定;

[0077] 所述执行规则可以由用户在所述模型描述模块中指定;

[0078] 所述执行规则可以作用于整个网络模型对象,也可以作用于子网络或特定容器;

[0079] 所述突触和神经元的裁剪与新生过程由调度器调度执行,可以在网络运行状态时执行,也可在网络暂停状态时执行。

[0080] 更进一步地,所述触发条件包括下面的一种或任几种:

[0081] 用户命令,即用户通过键盘或鼠标及其它方式输入命令给本系统,本系统接收到命令后即刻或稍后执行裁剪或新生过程;

[0082] 持续执行,即只要网络模型或其子区域达到符合裁剪或新生过程的规则就执行裁剪或新生过程;

[0083] 间隔执行,即系统每间隔一定时间或遍历周期,适时地自动启动裁剪或新生过程。

[0084] 更进一步地,所述裁剪过程的执行规则,分为突触裁剪规则和神经元裁剪规则;

- [0085] 所述突触裁剪规则包括下面的一种或任几种：
- [0086] 某个突触的参数和指定参照突触集合中所有突触参数的统计量达到一定数值关系(如某突触的权重小于指定边中全部突触权重平均值的1%)，则该突触应被裁剪；
- [0087] 某个突触的参数和指定阈值达到一定数值关系(如某突触的权重小于10.0)，则该突触应被裁剪；
- [0088] 某个突触超过一定时间或遍历周期没有被触发，则该突触应被裁剪；
- [0089] 某个突触被其它运算过程标记为可以被裁剪，则该突触应被裁剪；
- [0090] 所述神经元裁剪规则包括下面的一种或任几种：
- [0091] 某个神经元没有输入的突触，则该神经元应被裁剪；
- [0092] 某个神经元没有输出的突触，则该神经元应被裁剪；
- [0093] 某个神经元没有输入及输出的突触，则该神经元应被裁剪；
- [0094] 某个神经元的参数和指定参照神经元集合中所有神经元参数的统计量达到一定数值关系(如某神经元的阈值大于指定节点中全部神经元阈值最大值)，则该神经元应被裁剪；
- [0095] 某个神经元的参数和指定阈值达到一定数值关系(如某神经元的阈值大于1000.0)，则该神经元应被裁剪；
- [0096] 某个神经元超过一定时间或遍历周期没有发放，则该神经元应被裁剪；某个神经元被其它运算过程标记为可以被裁剪，则该神经元应被裁剪。
- [0097] 更进一步地，所述新生过程的执行规则，分为神经元新生规则和突触新生规则；
- [0098] 所述神经元新生规则包括下面的一种或任几种：
- [0099] 某个节点容器的现存神经元数量和该节点容器的总容量达到一定比例或数值关系，则以其总容量的一定比例或数量新生神经元；
- [0100] 某个节点容器按照一定速率，即每间隔一定时间或遍历周期以其总容量的一定比例或数量新生神经元；
- [0101] 某个节点容器被其它运算过程标记为需要新生神经元，并按照一定速率(即每间隔一定时间或遍历周期以其总容量的一定比例或数量)新生神经元；
- [0102] 所述突触新生规则包括下面的一种或任几种：
- [0103] 某个边容器的现存突触数量和该边容器的总容量达到一定比例或数值关系，则以其总容量的一定比例或数量新生突触；
- [0104] 某个边容器按照一定速率(即每间隔一定时间或遍历周期以其总容量的一定比例或数量)新生突触；
- [0105] 某个边容器被其它运算过程标记为需要新生突触，并按照一定速率(即每间隔一定时间或遍历周期以其总容量的一定比例或数量)新生突触；
- [0106] 某个节点容器存在没有输入或输出突触的神经元，则在对应的各个边容器中分别为其新生输入突触或输出突触。
- [0107] 进一步地，所述模型描述模块中，用户可以为各个容器指定一至多条规则，全部规则构成了一个规则库，
- [0108] 所述规则管理器对所述规则库中的规则按照一定优先级进行排序，当作用于一个容器的多条规则互相之间发生冲突时，只有最高优先级的规则被执行，当一个容器没有指

定任何规则时,规则管理器采用默认规则执行;

[0109] 所述规则库中的规则包括:遍历规则、内存使用规则、数据I/O规则、突触与神经元的裁剪与新生规则;

[0110] 所述遍历规则可以用于指导调度器按照一定时间间隔或遍历周期重复遍历或跳过遍历网络的全部或特定容器,以将计算资源集中在需要计算密集的子网络,提高数据利用效率;

[0111] 所述内存使用规则可以用于指导调度器合理安排主内存和协处理器内存的使用;

[0112] 所述数据I/O规则可以用于指导调度器调度数据在主内存和协处理器内存间,以及内存与硬盘间交换的频次,节省I/O资源,提高整体运算效率。

[0113] 进一步地,所述调度器管理了一个主内存池和一至多个设备内存池,以合理分配网络模型对象在主内存和各个设备内存中的使用情况,

[0114] 所述主内存池用于管理主内存的使用;

[0115] 所述设备内存池对应于各个协处理器(可以是ARM,GPU,DSP,ASIC等),用于管理相应设备内存的使用;

[0116] 所述主内存池和设备内存池的容量上限和下限由用户通过配置描述模块指定。

[0117] 进一步地,所述调度器管理了一个线程池,用于动态安排子线程参与多线程运算,以合理安排主计算单元(可以是CPU、ARM等)和协处理器(可以是ARM、GPU、DSP等)以及I/O设备(硬盘、摄像头、音频输入、控制输出等)的运算负载。

[0118] 进一步地,所述调度器管理了一个节点数据输入缓存器、一个节点数据输出缓存器、一个边数据输入缓存器、一个边数据输出缓存器,用于缓存需要读写硬盘或I/O设备的数据,以使调度器能够根据处理器、硬盘和I/O设备的负载,适时地安排硬盘、I/O设备读写以避免I/O阻塞,

[0119] 所述各个缓存器的容量、读写硬盘或I/O设备的频次上限和下限、读写硬盘或I/O设备的吞吐量上限和下限,由用户通过配置描述模块指定。

[0120] 本发明还提供了一种用于类脑智能与认知计算的脉冲神经网络运算方法,该方法使用上述的一种用于类脑智能与认知计算的脉冲神经网络运算系统。

[0121] 经由上述的技术方案可知,与现有技术相比,本发明公开提供了一种用于类脑智能与认知计算的脉冲神经网络运算系统及方法,该系统提供了统一而灵活的建模方式,而且提供的多层级树状结构的网络描述方式支持生物脑神经系统全尺度的建模和灵活的网络拓扑结构,从而将建模尺度和建模丰富度有机统一,并将各个尺度的全部模型融合为统一的神经网络进行运算,并且支持将数据以张量的形式表征和存储,这也使本系统既可以支持脉冲神经网络,也可以兼容传统神经网络(深度学习)以及其它以张量为主要数据表征方式的算法,还提供了按照一定条件和规则自动执行突触和神经元裁剪与新生的功能,免去了神经网络开发者需要自行实现相关功能的负担。

[0122] 上述建模设计方式可以是将任何层级(或尺度)的模型分解为数据(data)和操作(operation)两部分。如前所述,数据可由NodeParam、EdgeParam或Param容纳,并由对应的参数数据库存储。操作是可以访问和更新前述数据的可执行程序(如函数及包含函数的类)。

附图说明

[0123] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据提供的附图获得其他的附图。

[0124] 图1附图为本发明提供的一种用于类脑智能与认知计算的脉冲神经网络运算系统的整体架构示意图;

[0125] 图2附图为本发明实施例中一种用于类脑智能与认知计算的脉冲神经网络运算系统的网络层级示意图;

[0126] 图3附图为本发明实施例中一种用于类脑智能与认知计算的脉冲神经网络运算系统的系统运行过程流程示意图。

具体实施方式

[0127] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0128] 参见附图1,本发明实施例公开了一种用于类脑智能与认知计算的脉冲神经网络运算系统,该系统包括:模型描述模块3、参数数据库2、配置描述模块1、配置管理器6、规则管理器12、数据管理器8、网络构建器9、网络管理器10、操作管理器7、调度器11、日志管理器5、运行监测模块13、图形化显示模块4。

[0129] 上述模型描述模块3包括:网络描述单元、汇聚描述单元、流通描述单元。它们一起描述了整个网络的各个构成部分与拓扑结构。它们优先采用嵌套式语法,可以选用XML或JSON的文件格式。

[0130] 网络描述单元可以用于描述Network及Param等容器,可用于描述整个网络的参数与运行规则,并通过链接指向一至多个汇聚描述单元和流通描述单元。

[0131] 汇聚描述单元可以用于描述Confluence、Module、Layer、Node、NodeParam和Param等容器,可用于描述网络中节点的模块与层组划分关系、各个容器的参数与运行时的规则和命令。

[0132] 流通描述单元可以用于描述Flow、Channel、Link、Edge、EdgeParam和Param等容器,可用于描述网络中边的连接(拓扑)关系、各个容器的参数与运行时的规则和命令。

[0133] 参见附图3,模型描述模块3中支持的网络描述方式优先选用多层次树状结构表示,以模仿生物脑神经系统的组织方式:

[0134] 1,上述汇聚描述单元支持将网络中的节点按照一定层组和模块排列组织,可以表征生物脑中神经元及相关胶质细胞的多层级组织方式(如,核团→脑区→全脑);

[0135] 2,上述流通描述单元支持将网络中的边按照拓扑(连接关系)相似性分组和层级排列组织,可以表征生物脑中神经突触的多种组织方式(如树突、神经通路中的投射、神经纤维束等)及相关胶质细胞的突起的组织方式。这样可使开发、调试、管理和调度运行大规模类脑神经网络更加直观、便捷。

[0136] 具体地,上述多层次树状结构中:

[0137] Network表示网络容器,位于树状结构的第一层级(最顶级),可用于表征全脑和行为尺度的模型。每个Network可以容纳一至多个Confluence和Flow。

[0138] Confluence表示汇聚容器,位于树状结构的第二层级,可用于表征脑区尺度的模型。每个Confluence可以容纳一至多个Module。

[0139] Module表示模块容器,位于树状结构的第三层级,可用于表征神经核团尺度的模型。每个Module可以容纳一至多个Layer。

[0140] Layer表示层组容器,位于树状结构的第四层级,可用于表征神经环路尺度的模型。每个Layer可以容纳一至多个Node。

[0141] Node表示节点容器,位于树状结构的第五层级,可用于表征神经元尺度或胶质细胞尺度的模型,也可用于表征一群神经元或胶质细胞。神经元模型的发放特性可以构建为紧张性发放、快速发放、爆发式发放、峰发放或相位性发放等,其对于上游输入信号的响应可以构建为不同的神经适应性或敏感性曲线,其对于下游的作用可以构建为兴奋型的、抑制型的、调制型的或中性的。胶质细胞模型可以构建为星形胶质细胞、少突胶质细胞、小胶质细胞、施万细胞和卫星细胞模型等。每个Node可以容纳一至多个NodeParam。当Node用于表征一群同种类神经元时,其容纳的NodeParam的数量由神经元模型的参数种类的数量决定,即每个种类对应一个NodeParam并将该Node中全部神经元的该类参数以张量排列保存。

[0142] Node还可用于表征输入、输出节点,用于对接系统的I/O设备,如摄像头输入、音频输入、传感器输入、控制输出等。读写I/O设备的数据通过该Node的各个NodeParam动态更新。

[0143] NodeParam表示节点参数容器,位于树状结构的第六层级(最低级),可用于表征分子尺度、受体尺度、神经递质或神经调质尺度的模型,也可用于表征一群神经元或胶质细胞模型的参数张量。神经递质或神经调质模型可以构建为兴奋型的、抑制型的或调制型的。受体模型可构建为离子型的或代谢型的,其对于神经递质或神经调质的响应效果可以构建为兴奋型的、抑制型的、调制型的或中性的。

[0144] Flow表示流通容器,位于树状结构的第二层级,可用于表征联接脑区间的神经纤维束尺度的模型。每个Flow可以容纳一至多个Channel。

[0145] Channel表示通道容器,位于树状结构的第三层级,可用于表征由联接神经核团间的轴突构成的传导束的模型。每个Channel可以容纳一至多个Link。

[0146] Link表示连接容器,位于树状结构的第四层级,可用于表征神经环路中由轴突构成的神经通路的模型。每个Link可以容纳一至多个Edge。

[0147] Edge表示边容器,位于树状结构的第五层级,可用于表征树突尺度或突触尺度的模型,也可用于表征一群突触或胶质细胞的突起。树突尺度的模型可以构建为顶树突模型、基树突模型或突棘模型。突触模型可以构建为兴奋型的、抑制型的、调制型的或中性的。每个Edge可以容纳一至多个EdgeParam。当Edge用于表征一群同种类突触时,其容纳的EdgeParam的数量由突触模型的参数种类的数量决定,每个种类对应一个EdgeParam将该Edge中全部突触的该类参数以张量排列保存。

[0148] EdgeParam表示边参数容器,位于树状结构的第六层级(最低级),可用于表征分子尺度、神经递质或神经调质尺度、受体尺度的模型,也可用于表征一群突触或胶质细胞的突

起模型的参数张量。分子尺度的模型可以构建为细胞内分子模型、细胞膜中分子模型和细胞间隙分子模型。神经递质或神经调质模型可以构建为兴奋型的、抑制型的或调制型的。受体模型可构建为离子型的或代谢型的,其对于神经递质或神经调质的响应效果可以构建为兴奋型的、抑制型的、调制型的或中性的。

[0149] Param表示一般参数容器,属于辅助容器。根据建模的需要,上述各层级的容器都可以额外带有一至多个Param,用于以张量的形式容纳参数数据。

[0150] NodeParam、EdgeParam以及Param内部可采用张量(即多维矩阵)的形式容纳参数。张量的维度可以是1维至多维,具体排列和使用方式由用户指定。例如,张量可以是4维,各个参数在张量中的位置可以由坐标(x,y,z,t)表示,其中x,y,z三个维度对应于父级容器中表征的各个神经组织(如神经元或突触等)的空间排布位置,t代表时间维度,可表征时序信息的缓存和延迟,可用于模拟神经调质对神经元和突触的长时间作用机制(具有延迟性)。再如,张量中的参数可以由父级容器中全体或部分神经组织(如神经元或突触等)共享,可用于模拟神经调质对目标区域中的全部神经组织的大面积作用。

[0151] 上述每个容器都有编号和名称,用于在多层级树状结构中索引。每个容器都有一个控制块(Control Block)用于存储统计和控制信息,如网络的遍历次序与规则、已参与遍历运算的次数、数据是否已入驻主内存、数据是否已入驻协处理器内存、硬盘读写的频次等,并由规则管理器和调度器管理和更新。

[0152] Flow及其所有子级容器都可以对应于一个上游容器和一个下游容器,并通过上、下游容器的编号或名称对它们进行索引访问。上游容器和下游容器都可以是任意层级的容器,二者可以是同一个或者不同的容器。因此Flow及其所有子级容器都可与其上、下游容器构成信息流动通路,表征信息在两个信息源(如上游容器和下游容器)间的(单向或双向)流动和处理过程。网络中的多个容器间可构成信息流动的任意拓扑结构。

[0153] 上述信息的流动和处理过程可用于实现神经脉冲通过突触在神经元间传导、突触和突触之间交换信息、神经元与突触可塑性等多种生物脑神经机制。

[0154] 上述信息流动的任意拓扑结构,可用于实现脑神经系统中任意的神经环路连接方式,包括支持同一个神经元连回自己的反馈连接、同一群(层)的神经元互相之间的连接、不同群(层)神经元之间的任意连接(顺次/前馈、跨层、反馈等),以及突触和突触的直接连接,并且允许反馈型连接的无限次循环计算。

[0155] 下面详细举例说明:

[0156] 当一个Edge用于表征一至多个突触,并且其对应的上、下游容器都是表征一至多个神经元的Node,那么:

[0157] 1,如果该Edge的上、下游容器是不同的Node,它们之间的拓扑关系(例如,Node 1->Edge->Node 2)可以用于实现不同群(层)神经元之间通过突触的顺向/前馈连接;

[0158] 2,如果该Edge的上、下游容器是同一个Node,它们之间的拓扑关系(例如,Node 1->Edge->Node 1)可用于实现同一群(层)神经元之间通过突触的互相连接,也可以用于实现神经元通过自突触(autapse)连接回自身的反馈连接;

[0159] 3,如果该Edge的上、下游容器是来自于不同Layer的Node,它们之间的拓扑关系可以用于实现不同层神经元之间通过突触的跨层连接。

[0160] 当一些Edge分别用于表征一至多个突触,以及一些Node分别用于表征一至多个神

神经元,它们构成的拓扑关系如Node 1→Edge 1→Node N→Edge N→Node 1,那么:

[0161] 1,如果各个Node分属不同的Layer,则该拓扑关系可以用于实现不同层神经元之间通过顺馈连接、跨层连接和反馈连接形成的神经环路;

[0162] 2,如果各个Node属于同一个Layer,则该拓扑关系可以用于实现由一至多个(或一至多群)不同神经元构成的反馈回路。

[0163] 在前述例子中,Edge中的突触可以通过访问上、下游容器中的神经元以获取其激发时序信息,结合自身参数(如权重)进行运算并将运算结果传播至上、下游容器中的神经元,可以实现神经脉冲通过突触在神经元间的传导,以及Hebbian、Anti-Hebbian、STDP等长、短时可塑性机制;Node中的神经元可根据接受到的(通过神经递质或神经调制传递的)信息发生功能改变或塑型(神经元可塑性的一种)。

[0164] 当一个Edge用于表征一至多个突触,并且其对应的上、下游容器中至少有一个是表征一至多个突触的Edge时,它们之间的拓扑关系(例如,Edge 1→Edge 2→Edge 3)可以用于实现突触和突触间的直接连接关系与直接信息交换。

[0165] 上述参数数据库用于存储网络的各项参数数据(包括初始化参数和运行时参数)。参数数据库可选为二进制文件或文本文件。文本文件可采用CSV文件格式或以其它字符分隔数据的文件格式。每个容器都可以有一个或多个对应的参数数据库。如可以使NodeParam、EdgeParam或Param容纳的参数由一至多个参数数据库存储,也可以让多个NodeParam、EdgeParam或Param共享一至多个参数数据库以存储相同的参数。用户可将网络中各个容器的参数数据库置于模型文件路径中相应的子文件夹。

[0166] 本系统支持的建模设计方式可以是任何层级(或尺度)的模型分解为数据(data)和操作(operation)两部分。如前所述,数据可由NodeParam、EdgeParam或Param容纳,并由对应的参数数据库存储。操作是可以访问和更新前述数据的可执行程序(如函数及包含函数的类)。

[0167] 例如,神经元的建模可以用传统神经元模型,将其ReLU激活函数设计为操作,并将其阈值参数设计为数据;神经元的建模也可以用脉冲神经元模型,将其leaky integrate-and-fire model的函数设计为操作,并将其参数设计为数据。

[0168] 再如,用户可以通过定义一至多种操作,使同一个Node中的各个神经元(无需通过Edge)直接互相访问和更新彼此的数据,以实现信息的快速交换,可用于模拟生物脑神经系统中的电突触。

[0169] 又如,用户建模时可以通过定义一至多种操作,使同一个Edge中的各个突触直接互相访问和更新彼此的数据,以实现信息的快速交换,用于模拟生物脑神经系统中同一个神经元的树突上多个突触间互相交换信息并进行逻辑运算的情况,包括分流抑制(shunting inhibition)等机制。

[0170] 综上所述,由于本系统提供了灵活而统一的建模方式,而且提供的多层级树状结构的网络描述方式支持生物脑神经系统全尺度的建模和灵活的网络拓扑结构,从而将建模尺度和建模丰富度有机统一,并将各个尺度的全部模型融合为统一的神经网络进行运算。再加上数据以张量的形式表征和存储,这也使本系统既可以支持脉冲神经网络,也可以兼容传统神经网络(深度学习)以及其它以张量为主要数据表征方式的算法。

[0171] 上述操作管理器7用于管理所有可在本系统上运行的操作。操作可以是能够运行

在通用CPU、ARM、DSP、GPU或其它处理器的程序(包括代码段、函数和类)。所有的操作构成了操作库。操作管理器7提供了编程接口用于根据操作的编号或名称查询和调取指定操作。用户可在模型描述模块中为各个容器指定需要执行的操作,运行时由调度器调度执行相应的操作。这样保证了该系统具有一定跨硬件平台通用性,可以运行在通用CPU、GPU、ARM、DSP等硬件平台上。

[0172] 上述配置描述模块1用于描述当前网络运行环境的配置参数。诸如系统可用的内存池大小、以何种方式运行(单次、多次,连续运行)、读写硬盘数据的频率上限和下限、启动突触和神经元裁剪与新生过程的条件等等。

[0173] 上述配置管理器6用于读取上述配置描述模块1,以获取系统配置参数,并提供编程接口给其它组件调用。

[0174] 上述网络模型对象由网络构建器9构造并驻留在内存中。它表征了整个网络,包括所有容器、拓扑关系和参数数据,是调度器调度运行的对象。

[0175] 上述规则管理器12用于读取模型描述模块3中用户声明的规则,并在调度器11调度网络模型对象的运算时对这些规则进行解释。用户可以在模型描述模块3中为各个容器指定一至多条规则。全部规则构成了一个规则库。规则管理器12对规则库中的规则按照一定优先级进行排序,当作用于一个容器的多条规则互相之间发生冲突时,只有最高优先级的规则被执行。当一个容器没有指定任何规则时,规则管理器12采用默认规则执行。

[0176] 规则库中的规则包括(但不限于):遍历规则、内存使用规则、数据I/O规则、突触与神经元的裁剪与新生规则等等。遍历规则可以用于指导调度器按照一定时间间隔或遍历周期重复遍历或跳过遍历网络的全部或特定容器,以将计算资源集中在需要计算密集的子网络,提高数据利用效率;内存使用规则可以用于指导调度器合理安排主内存和协处理器内存的使用;数据I/O规则可以用于指导调度器调度数据在主内存和协处理器内存间,以及内存与硬盘间交换的频次,节省I/O资源,提高整体运算效率。

[0177] 上述数据管理器8包括一至多种解码器和编码器。解码器用于读取、解析用户指定格式的数据文件并将其内容转换为可供计算机计算的数据类型。编码器用于将内存中的数据按照用户指定格式进行序列化以便写回硬盘。数据文件的文件类型可以是二进制文件,也可以是文本文件(采用Unicode或ASCII格式)。用户可以在数据管理器8中添加自定义的解码器和编码器,以读写自定义格式的文件。

[0178] 上述网络构建器9读取模型描述模块3,解析网络的拓扑结构,并通过数据管理器8读取数据文件,在内存中构建网络模型对象。

[0179] 上述网络管理器10提供了构建网络模型对象的编程接口,该接口通过调用网络构建器9以构建网络模型对象。网络管理器9还提供了访问、遍历和操作网络模型对象的编程接口,包括支持通过编号或名称来查询和更新任意容器、神经元、突触、参数等。支持的遍历顺序包括(但不限于):

[0180] 1,深度优先遍历;

[0181] 2,广度优先遍历;

[0182] 3,按照模型描述模块中指定的规则进行遍历。

[0183] 遍历的实现方式可以包括(但不限于):

[0184] 1,循环式遍历;

[0185] 2,递归式遍历。

[0186] 上述调度器11用于分配硬件资源和调度运算过程,保证运算效率最优。调度器11管理了一个主内存池和一至多个设备内存池,以合理分配网络模型对象在主内存和各个设备内存中的使用情况。主内存池用于管理主内存的使用;每个协处理器(可以是ARM,GPU,DSP,ASIC)各有一个对应的设备内存池,用于管理相应设备内存的使用。其容量上限和下限由用户通过配置描述模块1指定。

[0187] 上述调度器11管理了一个线程池,用于动态安排子线程参与多线程运算,以合理安排主计算单元(可以是CPU、ARM等)和协处理器(可以是ARM、GPU、DSP等)以及I/O设备(硬盘、摄像头、音频输入、控制输出等)的运算负载。

[0188] 上述调度器11管理了一个节点数据输入缓存器、一个节点数据输出缓存器、一个边数据输入缓存器、一个边数据输出缓存器,用于缓存需要读写硬盘或I/O设备的数据。它们优先采用环形队列的数据结构。各个缓存器的容量、读写硬盘或I/O设备的频次上限和下限,读写硬盘或I/O设备的吞吐量上限和下限,由用户通过配置描述模块1指定。调度器11根据处理器、硬盘和I/O设备的负载,适时地安排硬盘、I/O设备读写以避免I/O阻塞。

[0189] 由于使用了调度器11合理分配处理器及协处理器、内存、硬盘、I/O设备等硬件资源的使用,使本系统适合在硬件资源(如内存)相对有限的嵌入式设备上高效运行。

[0190] 本系统提供了按照一定触发条件和执行规则自动执行突触和神经元的裁剪与新生的功能。用户可以在配置描述模块1中指定启动裁剪或新生过程的触发条件,并在模型描述模块3中指定裁剪或新生过程的执行规则。执行规则可以作用于整个网络模型对象,也可以作用于子网络或特定容器。裁剪或新生过程由调度器11调度执行,可以在网络运行状态时执行,也可在网络暂停状态时执行。

[0191] 上述启动裁剪或新生过程的触发条件可以包括(但不限于)下面的一种或多种:

[0192] 1,用户命令,即用户通过键盘或鼠标及其它方式输入命令给本系统,本系统接收到命令后即刻或稍后执行裁剪或新生过程;

[0193] 2,持续执行,即只要网络模型或其子区域达到符合裁剪或新生过程的规则就执行裁剪或新生过程;

[0194] 3,间隔执行,即系统每间隔一定时间或遍历周期,适时地自动启动裁剪或新生过程。

[0195] 该裁剪过程的执行规则,分为突触裁剪规则和神经元裁剪规则。突触裁剪规则可以包括(但不限于)下面的一种或多种:

[0196] 1,某个突触的参数和指定参照突触集合中所有突触参数的统计量达到一定数值关系(如某突触的权重小于指定边中全部突触权重平均值的1%),则该突触应被裁剪;

[0197] 2,某个突触的参数和指定阈值达到一定数值关系(如某突触的权重小于10.0),则该突触应被裁剪;

[0198] 3,某个突触超过一定时间或遍历周期没有被触发,则该突触应被裁剪;

[0199] 4,某个突触被其它运算过程标记为可以被裁剪,则该突触应被裁剪。

[0200] 神经元裁剪规则可以包括(但不限于)下面的一种或多种:

[0201] 1,某个神经元没有输入的突触,则该神经元应被裁剪;

[0202] 2,某个神经元没有输出的突触,则该神经元应被裁剪;

[0203] 3,某个神经元没有输入及输出的突触,则该神经元应被裁剪;

[0204] 4,某个神经元的参数和指定参照神经元集合中所有神经元参数的统计量达到一定数值关系(如某神经元的阈值大于指定节点中全部神经元阈值最大值),则该神经元应被裁剪;

[0205] 5,某个神经元的参数和指定阈值达到一定数值关系(如某神经元的阈值大于1000.0),则该神经元应被裁剪;

[0206] 6,某个神经元超过一定时间或遍历周期没有发放,则该神经元应被裁剪;

[0207] 7,某个神经元被其它运算过程标记为可以被裁剪,则该神经元应被裁剪。

[0208] 该新生过程的执行规则,分为神经元新生规则和突触新生规则。神经元新生规则可以包括(但不限于)下面的一种或多种:

[0209] 1,某个节点容器的现存神经元数量和该节点容器的总容量达到一定比例或数值关系,则以其总容量的一定比例或数量新生神经元;

[0210] 2,某个节点容器按照一定速率,即每间隔一定时间或遍历周期以其总容量的一定比例或数量新生神经元;

[0211] 3,某个节点容器被其它运算过程标记为需要新生神经元,并按照一定速率(即每间隔一定时间或遍历周期以其总容量的一定比例或数量)新生神经元。

[0212] 突触新生规则可以包括(但不限于)下面的一种或多种:

[0213] 1,某个边容器的现存突触数量和该边容器的总容量达到一定比例或数值关系,则以其总容量的一定比例或数量新生突触;

[0214] 2,某个边容器按照一定速率(即每间隔一定时间或遍历周期以其总容量的一定比例或数量)新生突触;

[0215] 3,某个边容器被其它运算过程标记为需要新生突触,并按照一定速率(即每间隔一定时间或遍历周期以其总容量的一定比例或数量)新生突触;

[0216] 4,某个节点容器存在没有输入或输出突触的神经元,则在对应的各个边容器中分别为其新生输入突触或输出突触。

[0217] 上述调度器负责调度突触和神经元裁剪与新生的执行。当启动突触和神经元裁剪或新生的条件被触发时,调度器从其所管理的线程池中分配一至多个子线程,各自负责网络模型对象中的部分区域或特定容器。子线程会遍历所辖区域中的各个容器并按照指定的规则执行神经元裁剪或新生过程。

[0218] 由于本系统提供了按照一定条件和规则自动执行突触和神经元裁剪与新生的功能,并且提供了多种灵活的启动裁剪和新生过程的触发条件与执行过程的规则,免去了神经网络开发者需要自行编写突触和神经元裁剪与新生程序的负担,提高了开发的灵活性和高效性。突触与神经元的裁剪过程可以和新生过程交替配合使用,可以优化神经网络的编码效率、大幅压缩神经网络的尺寸和所需存储空间、节省了内存并提高了运算效率,使本系统适合在硬件资源有限的嵌入式设备上运行。本系统通过支持突触与神经元的裁剪和新生,有利于模拟生物脑神经系统中丰富的机制(如海马体的突触和神经元的凋亡与新生现象),能够更好地支持类脑智能和认知计算。

[0219] 上述日志管理器5用于记录系统运行时生成的日志,日志用于提示用户系统的工作状态与异常,以便于调试和维护。日志由一系列字符串和时间戳组成,可以在命令行环境

中显示或者保存在文件中用文本浏览器显示。日志管理器由日志记录编程接口和日志管理服务构成。日志记录编程接口由用户在程序中调用,并把日志数据传给日志管理服务。日志管理服务由独立的线程运行以避免阻塞网络运算,它将接收到的日志数据按照时间戳统一排序并缓存在内存中,当缓存的数据量多达一定程度时按照一定顺序保存至硬盘并清理缓存。

[0220] 上述运行监测模块13用于接收和响应用户输入并管理整个系统的运行状态。它采用状态机的设计,包含有默认状态、网络构建状态、网络运行状态、网络暂停状态。它拥有一个消息队列用于接收并缓存用户输入的命令,并有一个独立线程用于及时响应队列中的命令,以使状态机在不同的状态间切换。用户可通过键盘、鼠标、编程接口或其它方式输入命令。命令包括(但不限于):构建网络命令、开始运行命令、暂停运行命令、结束运行命令、突触与神经元裁剪命令以及突触与神经元新生命命令。

[0221] 下面结合附图3对系统的运行原理做简单的描述:

[0222] S1:系统启动并初始化运行环境;

[0223] S2:运行环境进入默认状态;

[0224] S3:通过配置管理其读取配置描述模块获取配置参数;

[0225] S4:等待命令输入;

[0226] S5:判断是否接收到构建网络命令,直至判断结果为是,进入下一步;

[0227] S6:当接收到构建网络命令时,运行环境切换至网络构建状态;

[0228] S7:初始化网络管理器和规则管理器;

[0229] S8:通过网络构建器读取模型描述模块并构建网络模型对象,通过数据管理器读取参数数据库;

[0230] S9:等待命令输入;

[0231] S10:判断是否收到开始运行命令,当判断结果为否时,返回步骤S9再次等待命令输入,当判断结果为是时,进入下一步;

[0232] S11:运行环境进入网络运行状态;

[0233] S12:调度执行;

[0234] S13:判断是否收到暂停运行命令,如果判断结果为是,则进入步骤S14,如果判断结果为否,则进入步骤S17;

[0235] S14:运行环境进入网络暂停状态;

[0236] S15:等待命令输入;

[0237] S16:判断是否收到开始运行命令,如果判断结果为是,则返回步骤S11,如果判断结果为否,返回步骤S15;

[0238] S17:判断是否达到指定停止条件(包括接收到结束运行命令等),如果判定结果为否,返回步骤S12,如果判定结果为是,结束运行。

[0239] 当系统初始化时,上述状态机在默认状态,启动消息队列,开始接收用户输入;当接收到构建网络命令时,状态机切换至网络构建状态,并构建网络模型对象;当接收到开始运行命令时,状态机切换至网络运行状态,并执行网络运算;当接收到暂停运行命令时,状态机切换至网络暂停状态,并暂停网络运算;当接收到结束运行命令时,状态机保存网络数据至硬盘,系统结束并退出。在状态机处于网络运行状态或网络暂停状态时,如果消息队列

中有突触与神经元裁剪命令,则通过调度器启动突触与神经元裁剪过程;如果消息队列中有突触与神经元新生命命令,则通过调度器启动突触与神经元新生过程。由于本系统使用了运行监测模块来管理系统的工作状态,可以在应用环境不需要网络运算时将系统切换至网络暂停状态,以节省功耗,使本系统适用于嵌入式系统。

[0240] 上述图像化显示模块4用于读取网络数据,并将其显示给用户,便于开发、监测和调试。图像化显示模块4可以直接读取网络模型对象在内存中的数据,也可以读取保存在硬盘的数据。图像化显示模块4采用独立线程以避免阻塞网络运算,因此它可以在网络调度运行中实时显示,也可以在网络调度运行结束后显示。

[0241] 本说明书中各个实施例采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似部分互相参见即可。对于实施例公开的装置而言,由于其与实施例公开的方法相对应,所以描述的比较简单,相关之处参见方法部分说明即可。

[0242] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下,在其它实施例中实现。因此,本发明将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

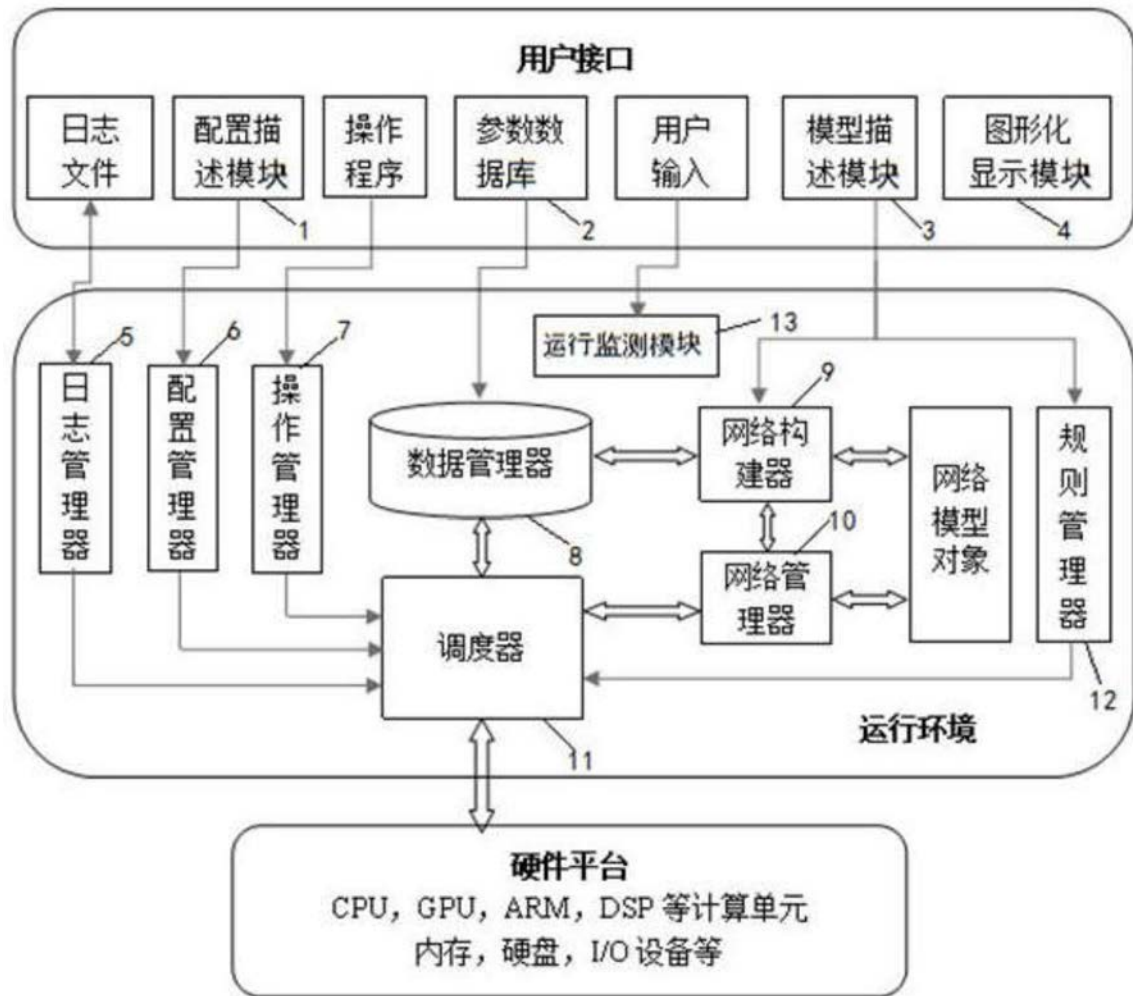


图1

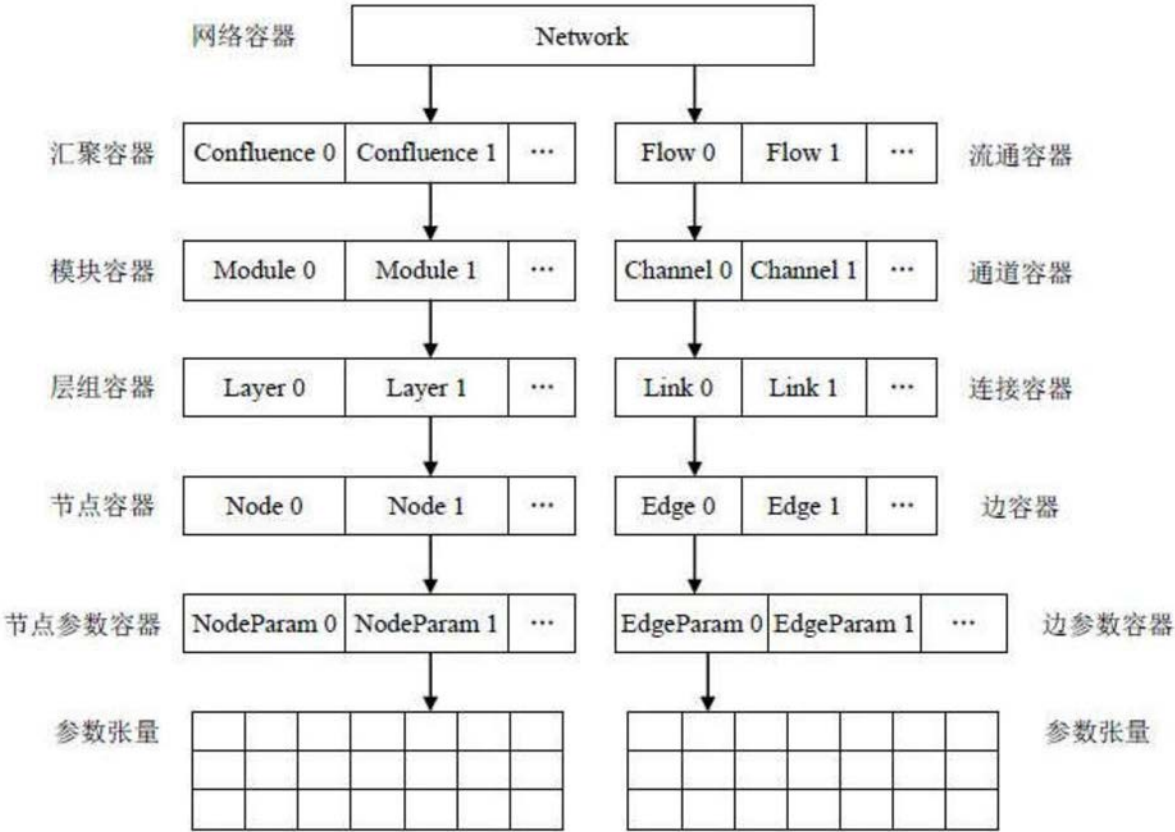


图2

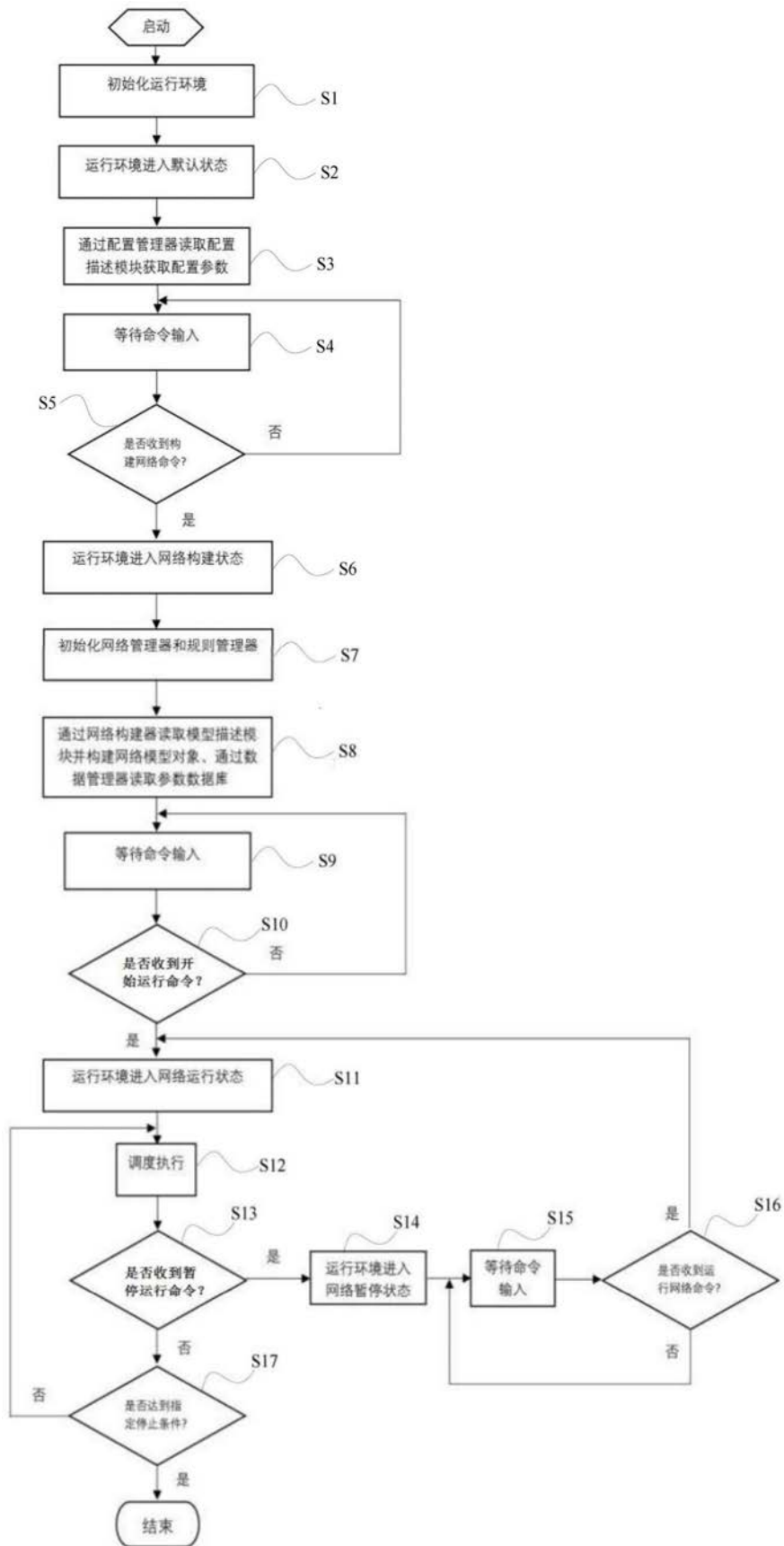


图3