



(12)发明专利申请

(10)申请公布号 CN 106462797 A

(43)申请公布日 2017.02.22

(21)申请号 201580024851.2

(74)专利代理机构 上海专利商标事务所有限公
司 31100

(22)申请日 2015.04.20

代理人 袁逸

(30)优先权数据

61/992,168 2014.05.12 US

14/483,054 2014.09.10 US

(51)Int.Cl.

G06N 3/02(2006.01)

G06N 99/00(2010.01)

(85)PCT国际申请进入国家阶段日
2016.11.11

(86)PCT国际申请的申请数据

PCT/US2015/026621 2015.04.20

(87)PCT国际申请的公布数据

W02015/175156 EN 2015.11.19

(71)申请人 高通股份有限公司

地址 美国加利福尼亚州

(72)发明人 C·M·维任斯基 D·J·朱利安
A·莎拉

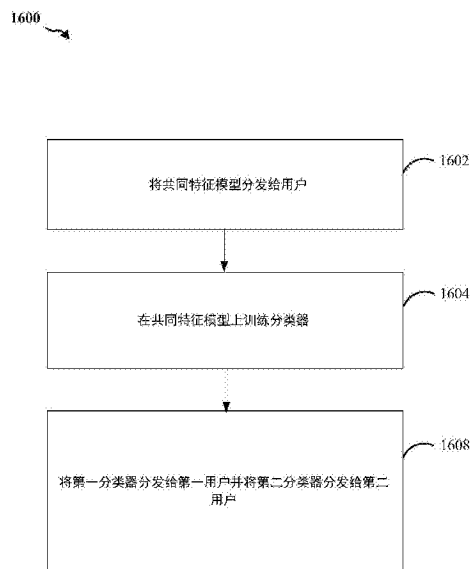
权利要求书2页 说明书26页 附图16页

(54)发明名称

共同特征上的定制分类器

(57)摘要

一种生成分类器模型的方法包括将共同特征模型分发至两个或更多个用户。在该共同特征模型之上训练多个分类器。该方法还包括将该多个分类器中的第一分类器分发至第一用户并将该多个分类器中的第二分类器分发至第二用户。



1. 一种生成分类器模型的方法,包括:
将共同特征模型分发给多个用户;
在所述共同特征模型上训练多个分类器;以及
将所述多个分类器中的第一分类器分发给所述多个用户中的第一用户并将所述多个分类器中的第二分类器分发给所述多个用户中的第二用户。
2. 如权利要求1所述的方法,其特征在于,所述多个分类器中的至少一者是在从至少一个实体获得的带标签数据集上训练的。
3. 如权利要求2所述的方法,其特征在于,所述至少一个实体包括所述多个用户中的指定用户。
4. 如权利要求2所述的方法,其特征在于,进一步包括为在从所述至少一个实体获得的所述带标签数据上训练的每个分类器提供度量。
5. 如权利要求4所述的方法,其特征在于,所述度量包括关于分类准确度或带标签图像的数目的信息。
6. 如权利要求1所述的方法,其特征在于,进一步包括:
接收从远程设备上的数据计算出的至少一个特征;
至少部分地基于所述至少一个特征来确定用于对所述远程设备上的数据进行分类的至少一个分类器;以及
将至少一个分类器的指示分发给所述远程设备。
7. 如权利要求1所述的方法,其特征在于,进一步包括:
接收从远程设备上的数据计算出的特征;
至少部分地基于接收到的特征来计算分类;以及
向所述远程设备传送所述分类。
8. 如权利要求1所述的方法,其特征在于,进一步包括:
组合所述第一分类器和所述第二分类器以生成被配置成对与一个或多个带标签数据集相关联的类进行分类的组合式分类器;以及
将所述组合式分类器分发给所述多个用户中的至少一者。
9. 如权利要求1所述的方法,其特征在于,所述多个分类器中的至少一者是至少部分地基于来自至少一个实体的反馈来训练的。
10. 一种用于生成分类器模型的装置,包括:
存储器;以及
耦合至所述存储器的至少一个处理器,所述至少一个处理器被配置成:
将共同特征模型分发给多个用户;
在所述共同特征模型上训练多个分类器;以及
将所述多个分类器中的第一分类器分发给所述多个用户中的第一用户并将所述多个分类器中的第二分类器分发给所述多个用户中的第二用户。
11. 如权利要求10所述的装置,其特征在于,所述至少一个处理器被进一步配置成在从至少一个实体获得的带标签数据集上训练所述多个分类器中的至少一者。
12. 如权利要求11所述的装置,其特征在于,所述至少一个实体包括所述多个用户中的指定用户。

13. 如权利要求11所述的装置,其特征在于,所述至少一个处理器被进一步配置成为在从所述至少一个实体获得的所述带标签数据上训练的每个分类器提供度量。

14. 如权利要求13所述的装置,其特征在于,所述度量包括关于分类准确度或带标签图像的数目的信息。

15. 如权利要求10所述的装置,其特征在于,所述至少一个处理器被进一步配置成:

接收从远程设备上的数据计算出的至少一个特征;

至少部分地基于所述至少一个特征来确定用于对所述远程设备上的数据进行分类的至少一个分类器;以及

将至少一个分类器的指示分发给所述远程设备。

16. 如权利要求10所述的装置,其特征在于,所述至少一个处理器被进一步配置成:

接收从远程设备上的数据计算出的特征;

至少部分地基于接收到的特征来计算分类;以及

向所述远程设备传送所述分类。

17. 如权利要求10所述的装置,其特征在于,所述至少一个处理器被进一步配置成:

组合所述第一分类器和所述第二分类器以生成被配置成对与一个或多个带标签数据集相关联的类进行分类的组合式分类器;以及

将所述组合式分类器分发给所述多个用户中的至少一者。

18. 如权利要求10所述的装置,其特征在于,所述至少一个处理器被进一步配置成至少部分地基于来自至少一个实体的反馈来训练所述多个分类器中的至少一者。

19. 一种用于生成分类器模型的装备,包括:

用于将共同特征模型分发给多个用户的装置;

用于在所述共同特征模型上训练多个分类器的装置;以及

用于将所述多个分类器中的第一分类器分发给所述多个用户中的第一用户并将所述多个分类器中的第二分类器分发给所述多个用户中的第二用户的装置。

20. 一种用于生成分类器模型的计算机程序产品,包括:

其上编码有程序代码的非瞬态计算机可读介质,所述程序代码包括:

用于将共同特征模型分发给多个用户的程序代码;

用于在所述共同特征模型上训练多个分类器的程序代码;以及

用于将所述多个分类器中的第一分类器分发给所述多个用户中的第一用户并将所述多个分类器中的第二分类器分发给所述多个用户中的第二用户的程序代码。

共同特征上的定制分类器

[0001] 相关申请的交叉引用

[0002] 本申请要求于2014年5月12日提交且题为“CUSTOMIZED CLASSIFIER OVER COMMON FEATURES (共同特征上的定制分类器)”的美国临时专利申请No. 61/992,168的权益,其公开内容通过援引全部明确纳入于此。

[0003] 背景

技术领域

[0004] 本公开的某些方面一般涉及神经系统设计,并且尤其涉及用于生成共同特征集上的定制分类器的系统和方法。

[0005] 背景

[0006] 可包括一群互连的人工神经元(即,神经元模型)的人工神经网络是一种计算设备或者表示将由计算设备执行的方法。人工神经网络可具有生物学神经网络中的对应的结构和/或功能。然而,人工神经网络可为其中传统计算技术是麻烦的、不切实际的、或不胜任的某些应用提供创新且有用的计算技术。由于人工神经网络能从观察中推断出功能,因此这样的网络在因任务或数据的复杂度使得通过常规技术来设计该功能较为麻烦的应用中是特别有用的。

[0007] 概述

[0008] 在本公开的一方面,公开了一种生成分类器模型的方法。该方法包括将共同特征模型分发至两个或更多个用户。该方法还包括在该共同特征模型之上训练多个分类器。该方法还包括将该多个分类器中的第一分类器分发至该两个或更多个用户中的第一用户,并且将该多个分类器中的第二分类器分发至该两个或更多个用户中的第二用户。

[0009] 在本公开的另一方面,公开了一种用于生成分类器模型的装置。该装置包括存储器以及耦合至该存储器的一个或多个处理器。(诸)处理器被配置成将共同特征模型分发至两个或更多个用户。(诸)处理器还被配置成在该共同特征模型之上训练多个分类器。这些处理器被进一步配置成将该多个分类器中的第一分类器分发至该两个或更多个用户中的第一用户,并且将该多个分类器中的第二分类器分发至该两个或更多个用户中的第二用户。

[0010] 在本公开的又一方面,公开了一种用于生成分类器模型的装备。该方法包括用于将共同特征模型分发至两个或更多个用户的装置。该方法还包括用于在该共同特征模型之上训练多个分类器的装置。该方法还包括用于将该多个分类器中的第一分类器分发至该两个或更多个用户中的第一用户,并且将该多个分类器中的第二分类器分发至该两个或更多个用户中的第二用户的装置。

[0011] 在本公开的又一方面,公开了一种用于生成分类器模型的计算机程序产品。该计算机程序产品包括其上编码有程序代码的非瞬态计算机可读介质。该程序代码包括用于将共同特征模型分发至两个或更多个用户的程序代码。该程序代码还包括用于在该共同特征模型之上训练多个分类器的程序代码。该程序代码还包括用于将该多个分类器中的第一分

类器分发至该两个或更多用户中的第一用户,并且将该多个分类器中的第二分类器分发至该两个或更多用户中的第二用户的程序代码。

[0012] 这已较宽泛地勾勒出本公开的特征和技术优势以便下面的详细描述可以被更好地理解。本公开的附加特征和优点将在下文描述。本领域技术人员应该领会,本公开可容易被用作修改或设计用于实施与本公开相同的目的的其他结构的基础。本领域技术人员还应认识到,这样的等效构造并不脱离所附权利要求中所阐述的本公开的教导。被认为是本公开的特性的新颖特征在其组织和操作方法两方面连同进一步的目的是和优点在结合附图来考虑以下描述时将被更好地理解。然而,要清楚理解的是,提供每一幅附图均仅用于解说和描述目的,且无意作为对本公开的限定的定义。

[0013] 附图简述

[0014] 在结合附图理解下面阐述的详细描述时,本公开的特征、本质和优点将变得更加明显,在附图中,相同附图标记始终作相应标识。

[0015] 图1解说了根据本公开的某些方面的示例神经元网络。

[0016] 图2解说了根据本公开的某些方面的计算网络(神经系统或神经网络)的处理单元(神经元)的示例。

[0017] 图3解说了根据本公开的某些方面的尖峰定时依赖可塑性(STDP)曲线的示例。

[0018] 图4解说了根据本公开的某些方面的用于定义神经元模型的行为的正态相和负态相的示例。

[0019] 图5解说了根据本公开的某些方面的使用通用处理器来设计神经网络的示例实现。

[0020] 图6解说了根据本公开的某些方面的设计其中存储器可以与个体分布式处理单元对接的神经网络的示例实现。

[0021] 图7解说了根据本公开的某些方面的基于分布式存储器和分布式处理单元来设计神经网络的示例实现。

[0022] 图8解说了根据本公开的某些方面的神经网络的示例实现。

[0023] 图9是解说根据本公开的各方面的用于学习模型的示例性数据流的框图。

[0024] 图10是解说根据本公开的各方面的分类器的示例性架构的框图。

[0025] 图11是解说根据本公开的各方面的用于学习模型的示例性数据流的框图。

[0026] 图12是解说根据本公开的各方面的用于生成分类器的示例性数据流的流程图。

[0027] 图13解说了根据本公开的各方面的用于学习模型的方法。

[0028] 图14解说了根据本公开的各方面的用于学习模型的方法。

[0029] 图15解说了根据本公开的各方面的用于生成分类器模型的方法。

[0030] 图16解说了根据本公开的各方面的用于生成分类器模型的方法。

[0031] 图17解说了根据本公开的各方面的用于生成分类器模型的方法。

[0032] 详细描述

[0033] 以下结合附图阐述的详细描述旨在作为各种配置的描述,而无意表示可实践本文中所描述的概念的仅有的配置。本详细描述包括具体细节以便提供对各种概念的透彻理解。然而,对于本领域技术人员将显而易见的是,没有这些具体细节也可实践这些概念。在一些实例中,以框图形式示出众所周知的结构和组件以避免湮没此类概念。

[0034] 基于本教导,本领域技术人员应领会,本公开的范围旨在覆盖本公开的任何方面,不论其是与本公开的任何其他方面相独立地还是组合地实现的。例如,可以使用所阐述的任何数目的方面来实现装置或实践方法。另外,本公开的范围旨在覆盖使用作为所阐述的本公开的各个方面的补充或者与之不同的其他结构、功能性、或者结构及功能性来实践的此类装置或方法。应当理解,所披露的本公开的任何方面可由权利要求的一个或多个元素来实施。

[0035] 措辞“示例性”在本文中用于表示“用作示例、实例或解说”。本文中描述为“示例性”的任何方面不必被解释为优于或胜过其他方面。

[0036] 尽管本文描述了特定方面,但这些方面的众多变体和置换落在本公开的范围之内。虽然提到了优选方面的一些益处和优点,但本公开的范围并非旨在被限定于特定益处、用途或目标。相反,本公开的各方面旨在能宽泛地应用于不同的技术、系统配置、网络和协议,其中一些作为示例在附图以及以下对优选方面的描述中解说。详细描述和附图仅仅解说本公开而非限定本公开,本公开的范围由所附权利要求及其等效技术方案来定义。

[0037] 示例神经系统、训练及操作

[0038] 图1解说了根据本公开的某些方面的具有多级神经元的示例人工神经系统100。神经系统100可具有神经元级102,该神经元级102通过突触连接网络104(即,前馈连接)来连接到另一神经元级106。为简单起见,图1中仅解说了两级神经元,尽管神经系统中可存在更少或更多级神经元。应注意,一些神经元可通过侧向连接来连接至同层中的其他神经元。此外,一些神经元可通过反馈连接来后向连接至先前层中的神经元。

[0039] 如图1所解说的,级102中的每一个神经元可以接收可由前级的神经元(未在图1中示出)生成的输入信号108。信号108可表示级102的神经元的输入电流。该电流可在神经元膜上累积以对膜电位进行充电。当膜电位达到其阈值时,该神经元可激发并生成输出尖峰,该输出尖峰将被传递到下一级神经元(例如,级106)。在一些建模办法中,神经元可以连续地向下一级神经元传递信号。该信号通常是膜电位的函数。此类行为可在硬件和/或软件(包括模拟和数字实现,诸如以下所述那些实现)中进行仿真或模拟。

[0040] 在生物学神经元中,在神经元激发时生成的输出尖峰被称为动作电位。该电信号是相对迅速、瞬态的神经脉冲,其具有约为100mV的振幅和约为1ms的历时。在具有一系列连通的神经元(例如,尖峰从图1中的一级神经元传递至另一级神经元)的神经系统的特定实施例中,每个动作电位都具有基本上相同的振幅和历时,并且因此该信号中的信息可仅由尖峰的频率和数目、或尖峰的时间来表示,而不由振幅来表示。动作电位所携带的信息可由尖峰、发放了尖峰的神经元、以及该尖峰相对于一个或数个其他尖峰的时间来确定。尖峰的重要性可由向各神经元之间的连接所应用的权重来确定,如以下所解释的。

[0041] 尖峰从一级神经元向另一级神经元的传递可通过突触连接(或简称“突触”)网络104来达成,如图1中所解说的。相对于突触104,级102的神经元可被视为突触前神经元,而级106的神经元可被视为突触后神经元。突触104可接收来自级102的神经元的输出信号(即,尖峰),并根据可调节突触权重 $w_1^{(i,i+1)}$ 、.....、 $w_P^{(i,i+1)}$ 来按比例缩放那些信号,其中P是级102的神经元与级106的神经元之间的突触连接的总数,并且i是神经元级的指示符。在图1的示例中,i表示神经元级102并且i+1表示神经元级106。此外,经按比例缩放的信号可

被组合以作为级106中每个神经元的输入信号。级106中的每个神经元可基于对应的组合输入信号来生成输出尖峰110。可使用另一突触连接网络(图1中未示出)将这些输出尖峰110传递到另一级神经元。

[0042] 生物学突触可以仲裁突触后神经元中的兴奋性或抑制性(超极化)动作,并且还可用于放大神经元信号。兴奋性信号使膜电位去极化(即,相对于静息电位增大膜电位)。如果在某个时间段内接收到足够的兴奋性信号以使膜电位去极化到高于阈值,则在突触后神经元中发生动作电位。相反,抑制性信号一般使膜电位超极化(即,降低膜电位)。抑制性信号如果足够强则可抵消掉兴奋性信号之和并阻止膜电位到达阈值。除了抵消掉突触兴奋以外,突触抑制还可对自发活跃神经元施加强力的控制。自发活跃神经元是指在没有进一步输入的情况下(例如,由于其动态或反馈而)发放尖峰的神经元。通过压制这些神经元中的动作电位的自发生成,突触抑制可对神经元中的激发模式进行定形,这一般被称为雕刻。取决于期望的行为,各种突触104可充当兴奋性或抑制性突触的任何组合。

[0043] 神经系统100可由通用处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现场可编程门阵列(FPGA)或其他可编程逻辑器件(PLD)、分立的门或晶体管逻辑、分立的硬件组件、由处理器执行的软件模块、或其任何组合来仿真。神经系统100可用在大范围的应用中,诸如图像和模式识别、机器学习、电机控制、及类似应用等。神经系统100中的每一神经元可被实现为神经元电路。被充电至发起输出尖峰的阈值的神经元膜可被实现为例如对流经其的电流进行积分的电容器。

[0044] 在一方面,电容器作为神经元电路的电流积分器件可被除去,并且可使用较小的忆阻器元件来替代它。这种办法可应用于神经元电路中,以及其中大容量电容器被用作电流积分器的各种其他应用中。另外,每个突触104可基于忆阻器元件来实现,其中突触权重变化可与忆阻器电阻的变化有关。使用纳米特征尺寸的忆阻器,可显著地减小神经元电路和突触的面积,这可使得实现大规模神经系统硬件实现更为切实可行。

[0045] 对神经系统100进行仿真的神经处理器的功能性可取决于突触连接的权重,这些权重可控制神经元之间的连接的强度。突触权重可存储在非易失性存储器中以在掉电之后保留该处理器的功能性。在一方面,突触权重存储器可实现在与主神经处理器芯片分开的外部芯片上。突触权重存储器可与神经处理器芯片分开地封装成可更换的存储卡。这可向神经处理器提供多种多样的功能性,其中特定功能性可基于当前附连至神经处理器的存储卡中所存储的突触权重。

[0046] 图2解说了根据本公开的某些方面的计算网络(例如,神经系统或神经网络)的处理单元(例如,神经元或神经元电路)202的示例性示图200。例如,神经元202可对应于来自图1的级102和106的任何神经元。神经元202可接收多个输入信号 204_1-204_N ,这些输入信号可以是该神经系统外部的信号、或是由同一神经系统的其他神经元所生成的信号、或这两者。输入信号可以是电流、电导、电压、实数值的和/或复数值的。输入信号可包括具有定点或浮点表示的数值。可通过突触连接将这些输入信号递送到神经元202,突触连接根据可调节突触权重 $206_1-206_N (W_1-W_N)$ 对这些信号进行按比例缩放,其中N可以是神经元202的输入连接总数。

[0047] 神经元202可组合这些经按比例缩放的输入信号,并且使用组合的经按比例缩放的输入来生成输出信号208(即,信号Y)。输出信号208可以是电流、电导、电压、实数值的和/

或复数值的。输出信号可以是具有定点或浮点表示的数值。随后该输出信号208可作为输入信号传递至同一神经系统的其他神经元、或作为输入信号传递至同一神经元202、或作为该神经系统的输出来传递。

[0048] 处理单元(神经元)202可由电路来仿真,并且其输入和输出连接可由具有突触电路的电连接来仿真。处理单元202及其输入和输出连接也可由软件代码来仿真。处理单元202还可由电路来仿真,而其输入和输出连接可由软件代码来仿真。在一方面,计算网络中的处理单元202可以是模拟电路。在另一方面,处理单元202可以是数字电路。在又一方面,处理单元202可以是具有模拟和数字组件两者的混合信号电路。计算网络可包括任何前述形式的处理单元。使用这样的处理单元的计算网络(神经系统或神经网络)可用在大范围的应用中,诸如图像和模式识别、机器学习、电机控制、及类似应用等。

[0049] 在神经网络的训练过程期间,突触权重(例如,来自图1的权重 $w_1^{(i,j+1)}$ 、.....、 $w_P^{(i,j+1)}$ 和/或来自图2的权重206₁-206_N)可用随机值来初始化并根据学习规则而被增大或减小。本领域技术人员将领会,学习规则的示例包括但不限于尖峰定时依赖可塑性(STDP)学习规则、Hebb规则、Oja规则、Bienenstock-Copper-Munro (BCM) 规则等。在一些方面,这些权重可稳定或收敛至两个值(即,权重的双峰分布)之一。该效应可被用于减少每个突触权重的位数、提高从/向存储突触权重的存储器读取和写入的速度、以及降低突触存储器的功率和/或处理器消耗。

[0050] 突触类型

[0051] 在神经网络的硬件和软件模型中,突触相关功能的处理可基于突触类型。突触类型可以是非可塑突触(权重和延迟没有改变)、可塑突触(权重可改变)、结构化延迟可塑突触(权重和延迟可改变)、全可塑突触(权重、延迟和连通性可改变)、以及基于此的变型(例如,延迟可改变,但在权重或连通性方面没有改变)。多种类型的优点在于处理可以被细分。例如,非可塑突触不会使用待执行的可塑性功能(或等待此类功能完成)。类似地,延迟和权重可塑性可被细分成可一起或分开地、顺序地或并行地运作的操作。不同类型的突触对于适用的每一种不同的可塑性类型可具有不同的查找表或公式以及参数。因此,这些方法将针对该突触的类型来访问相关的表、公式或参数。

[0052] 还进一步牵涉到以下事实:尖峰定时依赖型结构化可塑性可独立于突触可塑性地来执行。结构化可塑性即使在权重幅值没有改变的情况下(例如,如果权重已达最小或最大值、或者其由于某种其他原因而不被改变)也可被执行,因为结构化可塑性(即,延迟改变的量)可以是pre-post(前-后)尖峰时间差的直接函数。替换地,结构化可塑性可被设为权重变化量的函数或者可基于与权重或权重变化的界限有关的条件来设置。例如,突触延迟可仅在权重变化发生时或者在权重到达0的情况下才改变,但在这些权重为最大值时则不改变。然而,具有独立函数以使得这些过程能被并行化从而减少存储器访问的次数和交叠可能是有利的。

[0053] 突触可塑性的确定

[0054] 神经元可塑性(或简称“可塑性”)是大脑中的神经元和神经网络响应于新的信息、感官刺激、发展、损坏、或机能障碍而改变其突触连接和行为的能力。可塑性对于生物学中的学习和记忆、以及对于计算神经元科学和神经网络是重要的。已经研究了各种形式的可塑性,诸如突触可塑性(例如,根据Hebbian理论)、尖峰定时依赖可塑性(STDP)、非突触可塑

性、活跃性依赖可塑性、结构化可塑性和自稳态可塑性。

[0055] STDP是调节神经元之间的突触连接的强度的学习过程。连接强度是基于特定神经元的输出与收到输入尖峰(即,动作电位)的相对定时来调节的。在STDP过程下,如果至某个神经元的输入尖峰平均而言倾向于紧挨在该神经元的输出尖峰之前发生,则可发生长期增强(LTP)。于是使得该特定输入在一定程度上更强。另一方面,如果输入尖峰平均而言倾向于紧接在输出尖峰之后发生,则可发生长期抑压(LTD)。于是使得该特定输入在一定程度上更弱,并由此得名“尖峰定时依赖可塑性”。因此,使得可能是突触后神经元兴奋原因的输入甚至在将来作出贡献的可能性更大,而使得不是突触后尖峰的原因的输入在将来作出贡献的可能性更小。该过程继续,直至初始连接集合的子集保留,而所有其他连接的影响减小至无关紧要的水平。

[0056] 由于神经元可在其许多输入都在一短时段内发生(即,累积性足以引起输出)时产生输出尖峰,因此通常保留下来的输入子集包括倾向于在时间上相关的那些输入。另外,由于在输出尖峰之前发生的输入被加强,因此提供对相关性的最早充分累积性指示的那些输入将最终变成至该神经元的最后输入。

[0057] STDP学习规则可因变于突触前神经元的尖峰时间 t_{pre} 与突触后神经元的尖峰时间 t_{post} 之间的时间差(即, $t = t_{post} - t_{pre}$)来有效地适配将该突触前神经元连接到该突触后神经元的突触的突触权重。STDP的典型公式化是若该时间差为正(突触前神经元在突触后神经元之前激发)则增大突触权重(即,增强该突触),以及若该时间差为负(突触后神经元在突触前神经元之前激发)则减小突触权重(即,抑压该突触)。

[0058] 在STDP过程中,突触权重随时间推移的改变可通常使用指数式衰退来达成,如由下式给出的:

$$[0059] \quad \Delta w(t) = \begin{cases} a_+ e^{-t/k_+} + \mu, & t > 0 \\ a_- e^{t/k_-}, & t < 0 \end{cases}, \quad (1)$$

[0060] 其中 k_+ 和 $k_- \tau_{\text{sign}}(\Delta t) \tau_{\text{sign}}(\Delta t)$ 分别是针对正和负时间差的时间常数, a_+ 和 a_- 是对应的比例缩放幅值,以及 μ 是可应用于正时间差和/或负时间差的偏移。

[0061] 图3解说了根据STDP,突触权重作为突触前(presynaptic)和突触后(postsynaptic)尖峰的相对定时的函数而改变的示例性示图300。如果突触前神经元在突触后神经元之前激发,则对应的突触权重可被增大,如曲线图300的部分302中所解说的。该权重增大可被称为该突触的LTP。从曲线图部分302可观察到,LTP的量可因变于突触前和突触后尖峰时间之差而大致呈指数式地下降。相反的激发次序可减小突触权重,如曲线图300的部分304中所解说的,从而导致该突触的LTD。

[0062] 如图3中的曲线图300中所解说的,可向STDP曲线图的LTP(因果性)部分302应用负偏移 μ 。 x 轴的交越点306($y=0$)可被配置成与最大时间滞后重合以考虑到来自层 $i-1$ 的各因果性输入的相关性。在基于帧的输入(即,呈特定历时的包括尖峰或脉冲的帧的形式的输入)的情形中,可计算偏移值 μ 以反映帧边界。该帧中的第一输入尖峰(脉冲)可被视为要么如直接由突触后电位所建模地随时间衰退,要么在对神经状态的影响的意义上随时间衰退。如果该帧中的第二输入尖峰(脉冲)被视为与特定时间帧相关或有关,则该帧之前和之后的有关时间可通过使STDP曲线的一个或多个部分偏移以使得这些有关时间中的值可以

不同(例如,对于大于一个帧为负,而对于小于一个帧为正)来在该时间帧边界处被分开并在可塑性意义上被不同地对待。例如,负偏移 μ 可被设为偏移LTP以使得曲线实际上在大于帧时间的pre-post时间处变得低于零并且它由此为LTD而非LTP的一部分。

[0063] 神经元模型及操作

[0064] 存在一些用于设计有用的尖峰发放神经元模型的一般原理。良好的神经元模型在以下两个计算态相(regime)方面可具有丰富的潜在行为:重合性检测和功能计算。此外,良好的神经元模型应当具有允许时间编码的两个要素:输入的抵达时间影响输出时间,以及重合性检测能具有窄时间窗。最后,为了在计算上是有吸引力的,良好的神经元模型在连续时间上可具有闭合形式解,并且具有稳定的行为,包括在靠近吸引子和鞍点之处。换言之,有用的神经元模型是可实践且可被用于建模丰富的、现实的且生物学一致的行为并且可被用于对神经电路进行工程设计和反向工程两者的神经元模型。

[0065] 神经元模型可取决于事件,诸如输入抵达、输出尖峰或其他事件,无论这些事件是内部的还是外部的。为了达成丰富的行为库,能展现复杂行为的状态机可能是期望的。如果事件本身的发生在撇开输入贡献(若有)的情况下能影响状态机并约束该事件之后的动态,则该系统的将来状态并非仅是状态和输入的函数,而是状态、事件和输入的函数。

[0066] 在一方面,神经元n可被建模为尖峰带漏泄积分激发神经元,其膜电压 $v_n(t)$ 由以下动态来支配:

$$[0067] \quad \frac{dv_n(t)}{dt} = \alpha v_n(t) + \beta \sum_m w_{m,n} y_m(t - \Delta t_{m,n}) \quad (2)$$

[0068] 其中 α 和 β 是参数, $w_{m,n}$ 是将突触前神经元m连接至突触后神经元n的突触的突触权重,以及 $y_m(t)$ 是神经元m的尖峰发放输出,其可根据 $\Delta t_{m,n}$ 被延迟达树突或轴突延迟才抵达神经元n的胞体。

[0069] 应注意,从建立了对突触后神经元的充分输入的时间直至该突触后神经元实际上激发的时间存在延迟。在动态尖峰发放神经元模型(诸如Izhikevich简单模型)中,如果在去极化阈值 v_t 与峰值尖峰电压 v_{peak} 之间有差量,则可引发时间延迟。例如,在该简单模型中,神经元胞体动态可由关于电压和恢复的微分方程对来支配,即:

$$[0070] \quad \frac{dv}{dt} = (k(v - v_t)(v - v_r) - u + I) / C \quad (3)$$

$$[0071] \quad \frac{du}{dt} = a(b(v - v_r) - u) \quad (4)$$

[0072] 其中 v 是膜电位, u 是膜恢复变量, k 是描述膜电位 v 的时间尺度的参数, a 是描述恢复变量 u 的时间尺度的参数, b 是描述恢复变量 u 对膜电位 v 的阈下波动的敏感度的参数, v_r 是膜静息电位, I 是突触电流,以及 C 是膜的电容。根据该模型,神经元被定义为在 $v > v_{peak}$ 时发放尖峰。

[0073] Hunzinger Cold模型

[0074] Hunzinger Cold神经元模型是能再现丰富多样的各种神经行为的最小双态相尖峰发放线性动态模型。该模型的一维或二维线性动态可具有两个态相,其中时间常数(以及耦合)可取决于态相。在阈下态相中,时间常数(按照惯例为负)表示漏泄通道动态,其一般

作用于以生物学一致的线性方式使细胞返回到静息。阈上态相中的时间常数(按照惯例为正)反映抗漏泄通道动态,其一般驱动细胞发放尖峰,而同时在尖峰生成中引发等待时间。

[0075] 如图4中所解说的,该模型400的动态可被划分成两个(或更多个)态相。这些态相可被称为负态相402(也可互换地称为带漏泄积分激发(LIF)态相,勿与LIF神经元模型混淆)以及正态相404(也可互换地称为抗漏泄积分激发(ALIF)态相,勿与ALIF神经元模型混淆)。在负态相402中,状态在将来事件的时间趋向于静息(v_-)。在该负态相中,该模型一般展现出时间输入检测性质及其他阈下行为。在正态相404中,状态趋向于尖峰发放事件(v_s)。在该正态相中,该模型展现出计算性质,诸如取决于后续输入事件而引发发放尖峰的等待时间。在事件方面对动态进行公式化以及将动态分成这两个态相是该模型的基础特性。

[0076] 线性双态相二维动态(对于状态 v 和 u)可按照惯例定义为:

$$[0077] \quad \tau_p \frac{dv}{dt} = v + q_p \quad (5)$$

$$[0078] \quad -\tau_u \frac{du}{dt} = u + r \quad (6)$$

[0079] 其中 q_p 和 r 是用于耦合的线性变换变量。

[0080] 符号 p 在本文中用于标示动态态相,在讨论或表达具体态相的关系时,按照惯例对于负态相和正态相分别用符号“-”或“+”来替换符号 p 。

[0081] 模型状态由膜电位(电压) v 和恢复电流 u 来定义。在基本形式中,态相在本质上是由模型状态来决定的。该精确和通用的定义存在一些细微却重要的方面,但目前考虑该模型在电压 v 高于阈值(v_+)的情况下处于正态相404中,否则处于负态相402中。

[0082] 态相相关时间常数包括负态相时间常数 τ_- 和正态相时间常数 τ_+ 。恢复电流时间常数 τ_u 通常是与态相无关的。出于方便起见,负态相时间常数 τ_- 通常被指定为反映衰退的负量,从而用于电压演变的相同表达式可用于正态相,在正态相中指数和 τ_+ 将一般为正,正如 τ_u 那样。

[0083] 这两个状态元素的动态可在发生事件之际通过使状态偏离其零倾线(null-cline)的变换来耦合,其中变换变量为:

$$[0084] \quad q_p = -\tau_p \beta u - v_p \quad (7)$$

$$[0085] \quad r = \delta (v + \epsilon) \quad (8)$$

[0086] 其中 δ 、 ϵ 、 β 和 v_- 、 v_+ 是参数。 v_p 的两个值是这两个态相的参考电压的基数。参数 v_- 是负态相的基电压,并且膜电位在负态相中一般将朝向 v_- 衰退。参数 v_+ 是正态相的基电压,并且膜电位在正态相中一般将趋向于背离 v_+ 。

[0087] v 和 u 的零倾线分别由变换变量 q_p 和 r 的负数给出。参数 δ 是控制 u 零倾线的斜率的比例缩放因子。参数 ϵ 通常被设为等于 $-v_-$ 。参数 β 是控制这两个态相中的 v 零倾线的斜率的电阻值。 τ_p 时间常数参数不仅控制指数式衰退,还单独地控制每个态相中的零倾线斜率。

[0088] 该模型可被定义为在电压 v 达到值 v_s 时发放尖峰。随后,状态可在发生复位事件(其可以与尖峰事件完全相同)之际被复位:

$$[0089] \quad v = \hat{v}_- \quad (9)$$

[0090] $u = u + \Delta u$ (10)

[0091] 其中 \hat{v}_+ 和 Δu 是参数。复位电压 \hat{v}_+ 通常被设为 v_- 。

[0092] 依照瞬时耦合的原理, 闭合形式解不仅对于状态是可能的 (且具有单个指数项), 而且对于到达特定状态的时间也是可能的。闭合形式状态解为:

$$[0093] \quad v(t + \Delta t) = (v(t) + q_\rho) e^{\frac{\Delta t}{\tau_\rho}} - q_\rho \quad (11)$$

$$[0094] \quad u(t + \Delta t) = (u(t) + r) e^{\frac{\Delta t}{\tau_u}} - r \quad (12)$$

[0095] 因此, 模型状态可仅在发生事件之际被更新, 诸如在输入 (突触前尖峰) 或输出 (突触后尖峰) 之际被更新。还可在任何特定时间 (无论是否有输入或输出) 执行操作。

[0096] 而且, 依照瞬时耦合原理, 突触后尖峰的时间可被预计, 因此到达特定状态的时间可提前被确定而无需迭代技术或数值方法 (例如, 欧拉数值方法)。给定了先前电压状态 v_0 , 直至到达电压状态 v_f 之前的时间延迟由下式给出:

$$[0097] \quad \Delta t = \tau_\rho \log \frac{v_f + q_\rho}{v_0 + q_\rho} \quad (13)$$

[0098] 如果尖峰被定义为发生在电压状态 v 到达 v_s 的时间, 则从电压处于给定状态 v 的时间起测量的直至发生尖峰前的时间量或即相对延迟的闭合形式解为:

$$[0099] \quad \Delta t_s = \begin{cases} \tau_+ \log \frac{v_s + q_+}{v + q_+} & \text{if } v > \hat{v}_+ \\ \infty & \text{otherwise} \end{cases} \quad (14)$$

[0100] 其中 \hat{v}_+ 通常被设为参数 v_+ , 但其他变型可以是可能的。

[0101] 模型动态的以上定义取决于该模型是在正态相还是负态相中。如所提及的, 耦合和态相 ρ 可基于事件来计算。出于状态传播的目的, 态相和耦合 (变换) 变量可基于在上一 (先前) 事件的时间的状态来定义。出于随后预计尖峰输出时间的目的, 态相和耦合变量可基于在下一 (当前) 事件的时间的状态来定义。

[0102] 存在对该 Cold 模型、以及在时间上执行模拟、仿真、或建模的若干可能实现。这包括例如事件-更新、步阶-事件更新、以及步阶-更新模式。事件更新是其中基于事件或“事件更新” (在特定时刻) 来更新状态的更新。步阶更新是以间隔 (例如, 1ms) 来更新模型的更新。这不一定利用迭代方法或数值方法。通过仅在事件发生于步阶处或步阶间的情况下才更新模型或即通过“步阶-事件更新”, 基于事件的实现以有限的时间分辨率在基于步阶的模拟器中实现也是可能的。”

[0103] 尽管本公开已经描述了尖峰神经元、神经元模型和网络的各种示例, 但本公开不限于此。相反, 非尖峰神经元和网络也可用于实现本公开的一些方面。

[0104] 分布式模型学习以及共同特征上的定制分类器

[0105] 本公开的各方面涉及用于在各设备也使用模型的同时以分布式松协调方式继续学习该模型的过程。在一示例性方面, 深度神经网络 (DNN) 可被用于移动设备上的图像中的对象识别, 其中诸移动设备将信息发回中央服务器以继续完善该模型。为了解释方便, 示例

性数据流和其它描述被应用于图像和对象识别。然而,本公开不限于此,而是替换地或补充地可以利用任何感官模态。

[0106] 本公开的其它方面涉及生成分类器模型。分类器模型可以在共同特征集上定制。在一个示例性方面,中央服务器可被配置成从一个或多个用户设备接收带标签示例的素材库。用户设备可包括个人计算机(PC)、电视机、视频游戏系统、移动设备,诸如膝上型计算机、平板PC、智能电话、或其它便携式电子设备。

[0107] 服务器可以配置有与数据集有关的统计特征集。在一些方面,数据集例如可对应于特定感官模态(图像、声音、取向、位置等)。服务器可基于接收到的带标签示例素材库以及统计特征集来生成分类器。

[0108] 图5解说了根据本公开的某些方面的使用通用处理器502来进行(之前提及的)学习模型、生成分类器模型、和/或更新分类器集合的示例实现500。与计算网络(神经网络)相关联的变量(神经信号)、突触权重、系统参数,延迟,频槽信息,参数更新,离群者信息,模型更新,特征信息,示例和/或标签信息可被存储在存储器块504中,而在通用处理器502处执行的指令可从程序存储器506中加载。在本公开的一方面,被加载到通用处理器502中的指令可包括用于以下操作的代码:从一个或多个用户接收模型更新,基于先前模型和模型更新来计算经更新模型,和/或基于经更新模型来向该一个或多个用户传送与经更新模型有关的数据。

[0109] 在本公开的另一方面,被加载到通用处理器502中的指令可包括用于以下操作的代码:基于共享式推断模型来从服务器接收数据、基于接收到的数据来生成包括一个或多个模型参数的模型、基于该模型来计算推断、基于该推断来计算一个或多个模型参数更新和/或基于模型参数更新来向服务器传送数据。

[0110] 在本公开的又一方面,被加载到通用处理器502中的指令可包括用于以下操作的代码:将第一分类器集合应用于第一数据集和/或基于第一分类器集合的输出或者应用第一分类器集合的性能量度中的一者或多者来向远程设备请求分类器更新。

[0111] 在本公开的又一方面,被加载到通用处理器502中的指令可包括用于以下操作的代码:将共同特征模型分发至多个用户、在共同特征模型之上训练多个分类器,和/或将该多个分类器中的第一分类器分发至该多个用户中的第一用户并将该多个分类器中的第二分类器分发至该多个用户中的第二用户。

[0112] 图6解说了根据本公开的某些方面的前述学习模型和/或生成分类器模型的示例实现600,其中存储器602可以经由互连网络604与计算网络(神经网络)的个体(分布式)处理单元(神经处理器)606对接。与计算网络(神经网络)相关联的变量(神经信号)、突触权重、系统参数,延迟,频率槽信息,参数更新、离群者信息、模型更新、特征信息、示例和/或标签信息可被存储在存储器602中,并且可从存储器602经由互连网络604的连接被加载到每个处理单元(神经处理器)606中。在本公开的一方面,处理单元606可被配置成从一个或多个用户接收模型更新、基于先前模型和模型更新来计算经更新模型,和/或基于经更新模型来向该一个或多个用户传送与经更新模型有关的数据。

[0113] 在本公开的另一方面,处理单元606可被配置成基于共享式推断模型来从服务器接收数据、基于接收到的数据来生成包括一个或多个模型参数的模型、基于该模型来计算推断、基于该推断来计算一个或多个模型参数更新,和/或基于模型参数更新来向服务器传

送数据。

[0114] 在本公开的又一方面,处理单元606可被配置成将第一分类器集合应用于第一数据集和/或基于第一分类器集合的输出或者应用第一分类器集合的性能量度中的一者或者来向远程设备请求分类器更新。

[0115] 在本公开的又一方面,处理单元606可被配置成将共同特征模型分发至多个用户、在该共同特征模型之上训练多个分类器,和/或将该多个分类器中的第一分类器分发至该多个用户中的第一用户并将该多个分类器中的第二分类器分发至该多个用户中的第二用户。

[0116] 图7解说了前述学习模型和/或生成分类器模型的示例实现700。如图7中所解说的,一个存储器组702可与计算网络(神经网络)的一个处理单元704直接对接。每一个存储器组702可存储与对应的处理单元(神经处理器)704相关联的变量(神经信号)、突触权重、和/或系统参数,延迟,频率槽信息,参数更新,离群者信息,模型更新,特征信息,示例和/或标签信息。在本公开的一方面,处理单元704可被配置成从一个或多个用户接收模型更新、基于先前模型和模型更新来计算经更新模型,和/或基于经更新模型来向该一个或多个用户传送与经更新模型有关的数据。

[0117] 在本公开的另一方面,处理单元704可被配置成基于共享式推断模型来从服务器接收数据、基于接收到的数据来生成包括一个或多个模型参数的模型、基于该模型来计算推断、基于该推断来计算一个或多个模型参数更新、和/或基于(诸)模型参数更新来向服务器传送数据。

[0118] 在本公开的又一方面,处理单元704可被配置成将第一分类器集合应用于第一数据集和/或基于第一分类器集合的输出或者应用第一分类器集合的性能量度中的一者或者来向远程设备请求分类器更新。

[0119] 在本公开的又一方面,处理单元704可被配置成将共同特征模型分发至多个用户、在该共同特征模型之上训练多个分类器,和/或将该多个分类器中的第一分类器分发至该多个用户中的第一用户并将该多个分类器中的第二分类器分发至该多个用户中的第二用户。

[0120] 图8解说了根据本公开的某些方面的神经网络800的示例实现。如图8中所解说的,神经网络800可具有多个局部处理单元802,它们可执行本文所描述的方法的各种操作。每个局部处理单元802可包括存储该神经网络的参数的局部状态存储器804和局部参数存储器806。另外,局部处理单元802可具有用于存储局部模型程序的局部(神经元)模型程序(LMP)存储器808、用于存储局部学习程序的局部学习程序(LLP)存储器810、以及局部连接存储器812。此外,如图8中所解说的,每个局部处理单元802可与用于提供针对该局部处理单元的各局部存储器的配置的配置处理器单元814对接,并且与提供各局部处理单元802之间的路由的路由连接处理单元816对接。

[0121] 在一种配置中,神经元模型被配置成从一个或多个用户接收模型更新、基于先前模型和模型更新来计算经更新模型,和/或基于经更新模型来向该一个或多个用户传送与经更新模型有关的数据。该神经元模型包括接收装置、计算装置和传送装置。在一个方面,该接收装置、计算装置和/或传送装置可以是被配置成执行所叙述的功能的通用处理器502、程序存储器506、存储器块504、存储器602、互连网络604、处理单元606、处理单元704、

局部处理单元802、和或路由连接处理单元816。

[0122] 在另一配置中,神经元模型被配置成基于共享式推断模型来从服务器接收数据、基于该模型来计算推断、基于该推断来计算一个或多个模型参数更新,和/或基于模型参数更新来向服务器传送数据。该神经元模型包括接收装置、计算装置和传送装置。在一个方面,该接收装置、用于计算推断的装置、用于计算模型参数更新的装置和/或传送装置可以是配置成执行所叙述的功能的通用处理器502、程序存储器506、存储器块504、存储器602、互连网络604、处理单元606、处理单元704、局部处理单元802、和或路由连接处理单元816。

[0123] 在又一配置中,神经元模型被配置成将第一分类器集合应用于第一数据集和/或基于第一分类器集合的输出或者应用第一分类器集合的性能量度中的一者或多者来向远程设备请求分类器更新。神经元模型包括应用装置和请求装置。在一个方面,该应用装置和/或请求装置可以是配置成执行所叙述的功能的通用处理器502、程序存储器506、存储器块504、存储器602、互连网络604、处理单元606、处理单元704、局部处理单元802、和/或路由连接处理单元816。

[0124] 在又一配置中,神经元模型被配置成将共同特征模型分发至各用户、在共同特征模型之上训练分类器,和/或将各分类器中的第一分类器分发至第一用户并将第二分类器分发至第二用户。神经元模型包括用于分发共同特征模型的装置、训练装置、以及用于将该多个分类器中的第一分类器分发至第一用户并将第二分类器分发至该多个用户中的第二用户的装置(“用于分发分类器的装置”)。在一个方面,用于分发共同特征模型的装置、训练装置和/或用于分发分类器的装置可以是配置成执行所叙述的功能的通用处理器502、程序存储器506、存储器块504、存储器602、互连网络604、处理单元606、处理单元704、局部处理单元802、和/或路由连接处理单元816。

[0125] 在另一配置中,神经元模型被配置成将共同特征映射图集应用于来自第一指定用户的带标签示例的第一素材库以学习第一分类器模型、将该共同特征映射图集应用于来自第二指定用户的带标签示例的第二素材库以学习第二分类器模型,和/或分发该分类器模型。神经元模型包括用于将共同特征映射图集应用于来自第一指定用户的带标签示例的第一素材库以学习第一分类器模型的装置、用于将共同特征映射图集应用于来自第二指定用户的带标签示例的第二素材库以学习第二分类器模型的装置以及分发装置。在一方面,用于将共同特征映射图集应用于来自第一指定用户的带标签示例的第一素材库以学习第一分类器模型的装置、用于将共同特征映射图集应用于来自第二指定用户的带标签示例的第二素材库以学习第二分类器模型的装置、和/或分发装置可以是配置成执行所叙述的功能的通用处理器502、程序存储器506、存储器块504、存储器602、互连网络604、处理单元606、处理单元704、局部处理单元802、和或路由连接处理单元816。

[0126] 在另一配置中,前述装置可以是配置成执行由前述装置所叙述的功能的任何模块或任何装置。

[0127] 根据本公开的某些方面,每个局部处理单元802可被配置成基于神经网络的一个或多个期望功能性特征来确定神经网络的参数,以及随着所确定的参数被进一步适配、调谐和更新来使这一个或多个功能性特征朝着期望的功能性特征发展。

[0128] 图9是解说根据本公开的各方面的用于学习模型的示例性数据流900的框图。参照

图9,在框902,可训练神经网络以学习具有初始权重W0的模型。在一些方面,该神经网络可被训练以学习用于对训练图像集进行对象识别的模型。该神经网络可例如包括深度神经网络(DNN)。DNN是具有多个隐藏层的神经网络。

[0129] 在框904,可将初始模型权重(也被称为“模型”)W0推送出或分发到各用户(例如,移动设备(诸如智能电话)或其它设备)或其它实体。在一些方面,该模型可被广泛地分发(例如,1亿或十亿个设备的数量级)。

[0130] 在框906和908,每一移动设备可使用模型W0来执行特定任务。例如,在一些方面,模型W0可提供对移动设备上的数据的分类。例如,模型W0可标识和/或标示设备用户的图片中的对象。在一些方面,对象可使用模型W0来自动标识或标示。另外,每一移动设备可以在图片被拍摄时或者在一些情形中在图片被预览时学习模型参数更新,移动设备i还可计算并累积模型参数更新 $\Delta W_{0,i}$ 。在一些方面,设备i只可使用分布式模型W0的参数(例如,权重)来进行推断,并且可以不在本地应用其更新。

[0131] 参数更新 $\Delta W_{0,i}$ 可以按数种方式来计算。例如,参数更新 $\Delta W_{0,i}$ 可通过向用户提示标签并使用后向传播或者在该时段内以该模型中的一层为目标并基于例如自动编码器目标函数计算该层的权重梯度来被计算。当然,也可使用其它类型的目标函数。例如,在一些方面,也可使用稀疏自动编码器、收缩自动编码器、去噪自动编码器目标函数等。此类目标函数可以用正则化惩罚来使重构最小化。参数更新也可使用无监督式苏醒-休眠过程或其它更新技术来计算。

[0132] 在框910中,移动设备(例如,智能电话)可经由中央服务器/中枢来发送其模型权重更新 $\Delta W_{0,i}$ 以供收集。在一些方面,模型权重更新可以在周期性的基础上(诸如每天、每周或每月)被发送到中央服务器。当然,这仅是示例性的并且不是限定性的。例如,在一些方面,移动设备可响应于来自服务器的请求(例如,服务器可轮询更新)来发回更新。在另一示例中,移动设备可响应于服务器请求或者与周期性的所调度更新相组合地发送更新。在又一示例中,更新可基于训练示例的累积来发回(例如,自从上一次提供的更新以来拍摄图片的时间或者自从上一次提供的更新以来的图片数)。

[0133] 在框912,中央服务器/中枢可进而基于从移动设备接收到的模型权重更新 $\Delta W_{0,i}$ 来计算新模型W1。

[0134] 在一些方面,在框914可经由验证过程来验证新模型。在框916,可将新模型W1推送出或分发到诸移动设备用户。在框918和920,每一移动设备可使用模型W1来执行特定任务。此后,该过程可被重复以进一步更新该模型。

[0135] 计算模型更新

[0136] 经更新模型可用各种方式来计算。例如,在一些方面,经更新模型可被如下计算:

$$[0137] \quad W_{k+1} = W_k + \eta \left(\frac{1}{n} \sum_i \Delta W_{k,i} \right) \quad (15)$$

[0138] 其中n是用户更新的数目,而 η 是学习参数。

[0139] 在一些方面,权重更新可被预归一化。例如,权重更新可通过在发回权重更新之前藉以学习的图片数来预归一化(除以该图片数)。这可提供模型权重的直接平均。

[0140] 在一些方面,更新也可被加权。在一个示例中,权重更新可因变于 p_i (用于计算 $\Delta W_{0,i}$ 的图像数)来被加权。如此,来自拍摄数百张图片的用户的权重更新可具有比来自只拍

摄例如一张图片的用户的权重更新更大的影响。相应地,在具有该修改的情况下,经更新模型可被计算为:

$$[0141] \quad W_{k+1} = W_k + \eta \left(\sum_i p_i \Delta W_{k,i} \right) / \left(\sum_i p_i \right) \quad (16)$$

[0142] 在加权更新的情形中,保护免遭用户的过度加权可以是合乎需要的。即,针对例如拍摄不感兴趣的图片(例如,白墙的众多图片)的用户、来自个体用户的过度呈现、以及尝试故意使模型降级的攻击者进行防护可以是合乎需要的。一种办法将会是在运行权重更新之前将图片数 p_i 封顶或限制为 $1 \leq p_i \leq p_{\max}$ 或等价地 $p_i \leftarrow \min(p_i, p_{\max})$ 。在该情形中,可使用来自多个用户(例如,所有用户或其片段(诸如同龄群))的合计的或大量的权重更新来平均化并针对来自攻击者的权重更新进行防护。此外,具有大元素值的权重更新 $\Delta W_{k,i}$ 可被过滤掉或归一化。

[0143] 模型验证

[0144] 因为推送出或分发到用户的模型除了学习新模型更新之外还可支持活跃推断,所以验证经更新模型可以是有用的。例如,在一些方面,模型性能可被验证以确保新学习到的权重不使推断性能过度降级。另一方面,当推断性能被过度降级时,可发起纠正性动作。

[0145] 在一些方面,可以在验证数据集上测量经更新模型性能。在一个示例中,经更新模型性能可通过计算对象识别的准确度或F得分来衡量。在该示例中,经更新模型可以只在验证性能不降低达超过预定量(例如,定义的百分比或固定差量)的情况下才被分发或推送出。如果性能的确降低达超过目标量,则可实现纠正性措施。例如,在一些方面,可以在一时段(例如,该轮)内忽略模型更新,可以向用户发送通知(例如以便重置其 Δ (增量)权重和/或使用当前模型或先前模型)。

[0146] 在一些方面,如下所述的离群者检测器可标识例如要移除/忽略的用户权重的子集。经更新模型然后可基于其余权重来被重新计算。经更新模型还可经受重新测试和验证过程。如果该模型仍不满足目标度量,则可使用附加或更多的限制性离群者过滤器。

[0147] 在一些方面,可使用梯度方向上的线搜索。例如,这可通过计算具有不同学习速率的若干潜在经更新模型以及使用具有最佳验证性能的模型、具有满足目标验证性能阈值的最大学习速率的模型、或根据验证性能结果所选择的模型来进行。

[0148] 新的或经更新的模型还可包括将先前模型用于推断以及将新模型用于计算权重更新的指示。这可允许学习朝将降低模型性能而不影响推断性能的方向探索数个步骤。

[0149] 在其他方面,用户设备可以维护两个模型(例如, W_0 和 W_1)。例如,用户设备可维护一个模型(例如, W_0)以及来自服务器的对应于模型更新的 Δ (例如, W_1)。由于这两个模型可能在欧氏距离上接近,因此设备可以用比在存储两个分开的模型中使用的存储器更少的存储器来维护这两个模型。在该办法中,在数个步骤之后,如果模型性能提升到新的更佳性能点,则以要使用该新模型的指示来推送出该模型。如果模型性能没有改善,则可采用一个或多个纠正性动作(例如,以上描述的纠正性动作)。

[0150] 此外,可执行对验证数据集的健全性检查以确保验证数据集未过时(例如,错过正驱动特征学习并且恰适地更新验证数据集的新对象(诸如新电话、汽车等))。

[0151] 离群者检测

[0152] 在一些配置中,离群者检测器过滤器可任选地被包括以检测具有反复的权重更新

(这指示问题或潜在攻击)的个体权重更新或用户/设备。离群者过滤器可对照权重更新群集来测试这些权重更新。例如,可计算关于给定权重值的更新分布。如果权重更新超过目标个数的标准偏差,则其可被标识为离群者。

[0153] 类似地,当使用如上所引述的线搜索时,梯度向量方向应当指向局部最小值。如果用户梯度和群集平均梯度的内积低于阈值,则其可被标记为离群者。如果梯度量值或梯度元素超过该群集的数个标准偏差,则其也可被标记为离群者。可使用其他统计测试。此外,可在有或没有作为离群者被包括的在测梯度的情况下计算群集和分布。

[0154] 如果给定梯度更新被记录为离群者,则其对于当前这轮权重更新可被略去,或者可在更新中被给予较小加权。如果用户被标志为反复地提供离群者权重更新,则可标志这些更新以供进一步调查或者可将该用户永久地标志为可能有意或无意地攻击模型准确性的用户。在一些方面,被标志用户可以按贡献比非标志用户少的方式来添加它们的权重更新。例如,经更新模型可被计算为:

$$[0155] \quad W_{k+1} = W_k + \eta \left(\frac{1}{n_{\text{non-flagged}}} \sum_{i \in \text{non-flagged}} \Delta W_{ki} + \frac{\gamma}{n_{\text{flagged}}} \sum_{i \in \text{flagged}} \Delta W_{ki} \right) \quad (17)$$

[0156] 其中 $\gamma < 1$ 以提供被标志用户的较小贡献。替换地,这些权重可从更新中被排除(例如, $\gamma = 0$)。

[0157] 相反,离群者检测器可以基于更新颖的图像、和/或包含更多新颖特征图像来辅助确定更新。离群者检测器还可以辅助标识提供此类新颖图像的用户。此外,当新颖图像和/或用户被标识出时,离群者标识器可以对这些图像、用户和/或特征进行减权(unweight)。

[0158] 架构更新

[0159] 经更新模型可包括架构更新。由于模型可在学习继续进行的同时被用于推断任务,因此基于初始训练数据和设备特性来制订初始模型能力的大小以使得这些训练数据不被过拟合可能是有益的。然而,随分布式学习的进展,性能可能变得受限模型容量。与此同时,随模型的学习,诸最低层可以开始收敛至它们所能学习的东西(例如,第一层中的Gabor型边缘检测器)。这些层中的附加训练可具有有限潜力。为了处置这点,模型可以生长和/或收缩。

[0160] 在一种办法中,可监视计算和验证阶段期间的模型性能以确定对于当前模型容量是否存在收敛。收敛度量包括追踪关于验证集的模型性能以确定该性能是否已饱和达目标数目个历元,或者以查看权重更新的特征,诸如权重更新的量值、大于阈值的数个元素的稀疏性、和/或梯度方向的相干性。训练历元数和/或训练示例数还可被用于确定模型生长和/或收缩是否将是有益的。

[0161] 如果指示了模型生长,则可例如增加一个或多个现有层中的神经元数目、或者可向该模型添加一个或多个附加层。新的神经元和/或层能以若干方式中的一种来添加。

[0162] 在一个办法中,此被添加的架构(例如,神经元和/或层)可被添加以使得不具有即时影响。例如,可用零(0)值来配置层中新节点的权重和/或可用形成恒等函数的权重集来配置新层。随后,后续学习历元将开始学习经完善的权重。

[0163] 在另一办法中,可按上述办法来添加新层,并且随后可实现对中央数据训练集的训练(监督式和无监督式)以获得更佳的初始权重集。

[0164] 尽管以上办法允许生长模型,但是可使用相同指示来缩减模型的可塑/可学习部

分。例如,可以冻结底层以使得仅针对诸较高层来计算和传送权重更新。

[0165] 分类器特征学习者拆分

[0166] 图10是解说根据本公开的各方面的分类器1000的示例性架构的框图。构造分类器的一个挑战是如何添加或扣减分类标签、或甚至是在不用从零开始学习的情况下构建更复杂的推断引擎。一个观察是诸如深度神经网络之类的模型可在较低层中可被视为学习特征并且在较高层中被视为关于这些特征的推断引擎。此外,特征学习可以从大量训练示例中受益最多,而推断引擎可在其正使用高质量特征的情况下用少得多的示例来学习。使用此观察,分布式学习能够学习关于特征的模型更新(诸如权重)并且每次从零开始在顶层中使用中央数据集来重新训练分类器,如图10所示。

[0167] 作为示例,在图10的分类器的示例性架构1000中,设备提供输入数据1002(例如,可拍摄照片或提供其他感官输入数据)。可基于输入数据1002来计算模型权重更新以提供分布式的习得特征1004作为DNN的特征层。这些设备可随后将这些权重更新(周期性地或以其他方式)发送给集中习得推断引擎1006,并且可基于这些权重更新来更新特征模型,如先前所描述的。此后,可固定这些特征模型权重并且可使用集中标记式数据集1008在这些特征之上训练新的图像分类器。结果所得的模型随后可被分发或推送出给这些设备以用于改进的对象识别能力和进一步特征模型学习。

[0168] 作为对此的扩展,集中学习式推断引擎1006可利用不断习得的特征来添加、扣减、组合对象标签、扩展至标记多个对象、或者提供其他改进。添加、扣减、或组合对象标签可通过恰适地修改用于在特征之上学习推断引擎的数据集(诸如修改现有图像的标签和/或添加/移除图像)来进行。

[0169] 类似地,由于推断引擎1006被构建在分布式习得特征1004之上,因此可使用新的推断过程、架构、或办法。例如,新推断过程可包括标记修改,诸如为每个图像提供多个标签而非为每个图像提供单个标签。在另一示例中,推断引擎架构可通过从人工神经网络(ANN)办法切换至尖峰神经网络办法、支持向量机(SVM)办法、或其他办法来被修改。

[0170] 另外,通过集中地学习推断引擎,可通过使用不同训练数据集在相同特征集上训练不同推断引擎来学习不同推断引擎以用于不同的使用情形、设备、或应用。

[0171] 作为又一扩展,作为冻结特征权重且仅训练推断引擎的替代,特征权重还可从学习一个或多个训练数据集以计算结果所得的模型来被进一步完善,该结果所得的模型可被分发给用户和/或设备。相反,推断引擎可按与这些特征相同的分布式方式来被习得。

[0172] 在一个配置中,分布式学习可以在很大程度上是无监督式的,其中在用户纠正来自推断引擎的一个或多个模型标签时偶尔有监督式输入。在此配置中,可使用不止一个学习过程来计算模型权重更新。同样,在此配置中,具有特征/推断学习概念性拆分的分布式学习可使用用户标签来在本地更新推断引擎,从而用户更快地看到模型改进。

[0173] 另外,在一些方面,出于隐私、选择退出、和/或带宽的目的,用户图像可以不被提供给中央服务器。在此情形中,图像可与本地标签一起被高速缓存在用户设备中,从而当新模型在该用户设备处被接收时推断引擎可通过基于本地存储的图像和标签更新权重来被自动完善。这可允许用户具有保留基于标签纠正的更新的推断引擎而同时该模型继续以分布式方式学习。

[0174] 图11是解说根据本公开的各方面的用于学习模型的示例性数据流1100的框图。参

照图11,在框1102,可在中央数据集上训练初始模型 W_0 。在框1104,可将初始模型 W_0 推送出或分发给例如用户1和用户2。当然,这仅仅是示例性的,并且该模型可被分发给任何数目的用户、用户群、或其他实体。

[0175] 初始模型 W_0 可被用于在每个用户设备处的识别和学习模型(1106,1108)。然而,所习得的权重更新可被异步地提供给中央服务器。即,每个用户可以异步地向中央服务器发送权重更新(1106和1108)。当中央服务器接收到单个模型更新(例如,在框1110从用户1或者在框1116从用户2接收)、或者可以是在一时间段上来自用户的子集的模型更新集时,该服务器可计算新模型并将其分发或推送出给这些用户(1112、1118)。

[0176] 在一些方面,这些新模型更新可仅被发送给提供更新的用户子集。即,这些更新可以有差别地分发。例如,在框1110,中央服务器收集来自用户1的权重更新并且进而计算新模型 W_1 。在框1112,中央服务器仅向用户1推送出或分发 W_1 。更新可被表达为:

$$[0177] \quad W_1 = W_0 + \eta \left(\frac{1}{n} \sum_i \Delta W_{k,i} \right) \quad (18)$$

[0178] 其中求和是针对该子集中的该一个或数个用户。在框1114,新模型 W_1 可被用于用户1设备处的识别和学习。

[0179] 在框1108,中央服务器可以从另一用户(例如,从用户2)或用户子集接收模型更新集合并计算另一新模型更新(1116)。该模型更新可如下被应用于正在发展中的模型:

$$[0180] \quad W_2 = W_1 + \eta \left(\frac{1}{n} \sum_i \Delta W_{k,i} \right) \quad (19)$$

[0181] ,即使这些更新是在较老的模型(诸如 W_0)上计算出的。在框1118,可将新模型 W_2 分发至另一用户(例如,用户2)或用户群。在框1120,可将新模型 W_2 用于用户2设备处的识别和学习。

[0182] 在一些方面,当接收到模型更新(例如, W_1 或 W_2)时,这些模型可被标记或以其他方式配置有关于它们基于哪个模型的指示。如果更新是在从初始模型(例如, W_0)开始的阈值数量的更新后被接收到的,则这些更新可由于陈旧而被丢弃。

[0183] 在异步更新情形中,异步更新次数可基于若干因素。在一个示例中,可针对不同的用户子集在不同时间规划更新以帮助对服务器和网络资源进行负载平衡。在第二示例中,更新可以在满足某一设备本地度量(诸如计算出目标数量的局部模型更新(例如目标数量的图片))或者从设备回到中央服务器的高带宽信道的突然可用后被发送。

[0184] 在一些方面,所习得的权重更新可按在每个学习计算(例如,图片)之后应用这些更新的在线方式、或者使用在目标数目的学习计算(例如,图片)之后应用这些更新的小批次过程来被本地应用。例如,用户可发送回自从中央服务器接收到上一个模型起的总累积权重更新。该办法可允许用户冒着中间性能由于推断是在未经验证的模型上执行而降级的风险来局部地改进其模型并更快地探索模型空间中的更多地方。

[0185] 性能降级的风险可通过在本地维护两个模型(例如,一个用于报告推断值,一个用于学习模型空间中的更多地方)来降低。当然,所维护的模型数仅仅是示例性的,并且可以根据资源可用性来维护任何数目的模型。这可通过维护模型 W 以及发展中的更新 ΔW 并使用 W 来进行推断且使用 $W + \eta \Delta W$ 来进行模型学习来完成。

[0186] 在这些模型探索情形中,中央服务器可以如先前描述的方法中那样应用模型更

新。在一些方面,中央服务器可对照验证集来测试不同的更新以确定哪个更新提供更好的模型更新。

[0187] 传送模型参数

[0188] 与分布式模型学习相关联的一个挑战是高容量模型的模型大小可能相对较大,因此推送出这些模型并收回习得模型参数的简单办法可消耗大量带宽。另外,对于中央服务器,从大量设备(例如,数亿到数十亿个设备)接收模型更新可产生要维护的非常大的流。存在可被用于减少带宽和设备处的存储器利用的若干方法。

[0189] 从设备到中央服务器

[0190] 第一办法是对每一用户(设备)发送到中央服务器的 ΔW 进行子采样。如果模型具有大量(例如,数百万或数十亿)权重参数,则 ΔW 向量具有那么多的元素。每一用户(设备)可发送元素子集(例如,随机的元素子集),而不是使该数百万或数十亿个设备中的每一者均向中央服务器发送完整的权重向量。因为每一 ΔW 元素通常被计算成使误差函数最小化,所以单就每一元素更新来看其应在良好方向上。因为存在大量设备,所以如果所有用户均发送其所有更新而不是使用合适的统计平均,则带宽可能未被高效利用。在一方面,服务器可发送关于在模型被推送出到用户(设备)时要发回的参数的数目的参数 n_p 。本地用户设备可在 ΔW 向量中随机选择要发送给中央服务器的相应数目的元素位置。如此,在每一学习更新上,本地设备可以只计算用于计算目标 ΔW 元素的中间值。此外,本地用户设备然后可以随时间只跟踪 ΔW 的 n_p 个元素。该设备可以在适当时间将这 n_p 个元素发送到中央服务器。

[0191] 在下一迭代期间,可实现其它选项。例如,在一种配置中,设备可保持相同的元素位置集合或者可重新生成新的随机元素位置。另外,由中央服务器推送出的参数 n_p 的值可随时间变化,例如以便计及增加的设备数、改变的模型大小、增加的带宽、以及其它因素。

[0192] 在另一配置中,中央服务器可以从各设备接收全部或一些参数并且可以对用于更新模型的 ΔW 进行子采样。可以这样做以便控制中央服务器中针对每一更新执行的计算量。在一些方面,中央服务器还可使用来自接收到的所有更新的 ΔW 的随机子集。在其它方面,中央服务器可丢弃接收到的更新中的某一些。

[0193] 从中央服务器到设备

[0194] 将模型更新分发或推送出到设备也可消耗大量带宽。存在可实现以减少带宽的若干办法。

[0195] 一种办法是向用户广播或多播权重更新,因为模型对于所有用户而言是共同的。例如,在一种配置中,权重更新可以在开销信道(诸如蜂窝短消息服务(SMS)信道、广播信道或开销位置)中被分发。

[0196] 另一种办法是使用基于先前模型权重值的算法来压缩模型权重更新。例如,对于模型 W_{k+1} ,中央服务器可计算 $W_{k+1} - W_k$ 并且然后对结果所得向量使用标准压缩过程,这应具有小值以发送模型更新。当用户(设备)接收到更新时,该设备可解压该更新并将其添加到先前模型。替换地,霍夫曼压缩可基于所估计的概率 $p(W_{k+1} | W_k)$ 来被使用。

[0197] 另外,如果权重更新中存在相关性(诸如可出现在使用动量的DNN中),则服务器可计算权重更新的二重差分:

[0198] $(W_{k+1} - W_k) - (W_k - W_{k-1})$ (20)

[0199] 二重差分可被压缩并发送以用于模型更新。在某一方面,基于概率的压缩可被使

用 $p(W_{k+1} | W_k, W_{k-1})$ 。

[0200] 在又一种办法中,中央服务器可指示供用户设备针对当前迭代来聚焦(例如,更新)的给定模型层或权重集。在该情形中,服务器可指示权重集被作为目标以用于当前模型更新迭代。设备可跟踪只与目标权重集有关的权重更新。类似于上文,诸设备还可选择该目标权重集的随机子集。诸设备可以在迭代结束时向中央服务器发送其模型权重更新。服务器进而可针对该权重集计算出经更新模型并只发出这些经更新权重以用于下一次模型更新。在一些方面,随时间迭代可将不同的权重集作为学习目标。

[0201] 作为上述办法的扩展,中央服务器在一些方面可指示不同的用户子集将不同的层或权重子集作为目标以用于迭代。中央服务器还可使用验证检查阶段以检查哪层或哪个模型子集对性能具有最大影响并且只推送那些更新。

[0202] 共同特征上的定制分类器

[0203] 本公开的各方面进一步涉及共同特征函数上的定制分类器。假设用户想要通过外观来标识野外的蘑菇,且用户希望为此能力付费。一个或多个蘑菇专家以使这些专家能为其工作而受到奖励的方式将其知识转告给该用户和其它蘑菇采集者的高效方式是什么?除了蘑菇之外,可受益于专家标记的对象类的其它示例包括汽车、动物、美术作品、医疗诊断图像等。

[0204] 根据本公开的各方面,公开了利用机器学习能力的分类器。实体集合(例如,指定的用户或专家)可以向中央服务器或“模型商店”提供带标签示例的素材库。该中央服务器还可包括与特定感官模态(或模态组合)有关的统计特征集。这些特征可以按无监督方式被学习。服务器可使用习得特征和专家提供的带标签示例集合两者来计算分类器。服务器可将计算出的分类器的参数分发至将允许用户计算他们遇到的各种对象的类的设备。

[0205] 被消耗以存储分类器参数的存储器通常将比完整的训练素材库小许多数量级。该客户端-服务器架构还可允许在两个训练素材库的级联上训练单个分类器的可能性,从而赋予用户合并来自相异或交叠领域的多个专家的分类知识的能力。

[0206] 可以在服务器侧和用户设备侧两者上提供基础设施以确保设备拥有适当的特征集、存储分类器参数的能力、以及部署这些参数以实现正确分类的能力。相应地,在本公开的一方面,公开了用于在潜在可能非常大数量的用户之间共享一个或多个指定用户或专家的分类专长的过程。一个或多个用户可能希望使用移动设备来收集感官数据并且在一些情形中将这数据分类成有意义的标签(例如,用相机观看蘑菇以标识蘑菇类型)。可以是带标签数据素材库形式的“专长”可被提供给一个或多个服务器,该一个或多个服务器可将带标签数据素材库与先前习得特征集相组合以便在这些特征之上计算出分类器。服务器然后将分类器分发至任何感兴趣用户的设备。可以存在比服务器多得多的用户。

[0207] 图12是解说根据本公开的各方面的用于生成分类器的示例性数据流1200的流程图。参照图12,在框1102,服务器可将特征集F分发至一个或多个用户。

[0208] 在一些方面,用户可各自具有用于从输入数据中计算相同的特征的机制。确保所有用户共享相同的输入特征的一种途径将会是使服务器将这些特征连同版本号一起推送出或分发至所有用户。例如,具有一层或多层的人工神经网络(ANN)可计算特征,并因此服务器可通过将该ANN的连接权重发送到所有用户来执行该数据流的这一部分。这些特征还可使用分布式无监督学习来被共同学习。

[0209] 对向一组用户提供分类器感兴趣的实体、专家或其它指定用户可以向服务器提供带标签数据。该带标签数据素材库可保持对服务器可用以使其能够在输入特征已经改变的情况下重新训练分类器。在框1204,例如,专家可以向服务器发送包括数据D和标签L的带标签示例的素材库E。例如,带标签示例素材库可以是图像集(以及针对每一图像的唯一性标签集)的形式。

[0210] 在框1206,服务器可构建或学习分类器,该分类器学习示例素材库E中的每一带标签资料 D_i 及其相应的专家提供的标签 L_i 之间的映射。存在可能分类器和学习方法的许多选择。例如,在一方面,ANN可被用于分类器且训练可使用后向传播来进行。

[0211] 在一些方面,还可构造组合多个专家的训练集中包含的知识的分类器(例如,可标识图像中的蘑菇和蝴蝶两者的分类器)。该组合式分类器可以例如通过使用两个训练素材库的混合来执行训练来构造。

[0212] 在附加方面,分类器参数可包括架构参数(例如,在ANN的情形中是层中的单元数)。这在所建议或指示的给定素材库使用具有较高容量并因此在给定层中具有较多单元的分类器的复杂情况下可以是有用的。

[0213] 在一方面,训练可涉及在固定特征 $F(d)$ 上训练分类器,或者可涉及通过例如后向传播遍历各特征层以及诸顶上的分类层来微调特征。在另一方面,来自固定特征 $F(d)$ 和/或经更新的 $F(d)$ 的权重 Δ (增量)可被发出到用户设备。在另一方面,训练可包括在相同的共享特征 $F(d)$ 之上训练针对两个专家训练集的两个单独分类器。

[0214] 另外,分类器还可被以阶级方式配置或组织以使得分类器具有顶层或一般层以及更具体的分类器。例如,顶层分类器可将图像分类为汽车,而更具体的分类器可分类汽车类型(例如,轿车、运动型多功能车、跑车等)。还可提供特定分类器的更多层。例如,一个特定层可将图像分类为1958Ferrari GT California Spyder。

[0215] 在一些方面,分类器可被应用于数据集并被配置成执行顶级分类。基于该顶级分类,分类器可以向服务器请求一个或多个附加分类器。在接收到该一个或多个附加分类器之际,可执行对数据集的更具体的分类。

[0216] 在框1208,用户可经由服务器(例如,“模型商店”)从中央选项集中选择大量知识或分类器。用户可以向服务器指示关于它想要下载哪个分类器的选择。这可例如采取显示可供下载的所有分类器的在线商店的形式。除了分类器之外,该商店还可给予用户下载顶级特征以计算分类器的选项(例如,在蘑菇分类器的情形中,用户可首先下载针对自然图像的低级视觉特征集)。用户还可指定将要作为组合式分类器来被下载的多个分类器。

[0217] 此外,用户可指定分类器阶级中的所需层。例如,用户可能想要用于对水果(诸如苹果)进行分类的通用分类器,或者可以在各种类型的苹果(例如,Granny Smith、Pink Lady、Fuji、Gala等)之间进行进一步区分的更具体的分类器。

[0218] 在框1210,服务器可以按描述分类器C的参数的形式向用户提供所请求的知识。一旦用户已指定要构建和/或下载的一个或多个分类器,服务器就可将该分类器的参数分发或推送给用户设备。在基于ANN的分类器的情形中,这些参数例如可以包括连接权重和偏置项。

[0219] 在一些实例中,服务器可自动向用户推送某个分类器或分类器阶级中的层的参数。这可例如基于经由用户提供的感官信息(例如,用户具有众多的跑车图像——可提供更

具体的分类器以使该用户能够对捕捉到的跑车图像进行进一步分类)。

[0220] 在框1212,用户可收集数据d(例如,用她的智能电话拍摄蘑菇的图片)。d中的特征F(d)可使用例如先前提提供的特征集F来在本地计算。分类器C可被应用于这些特征以计算所估计的专家标签以用于未知刺激(例如,蘑菇类型)。

[0221] 一旦装备有特征集F和所下载的分类器C,用户就可收集数据集d(例如,蘑菇图像),提取其特征F(d),并将这些信息馈送到分类器以获得分类C(F(d))。分类器在这些特征上的输出可表示专家对该观察的类的意见,该意见与专家先前提提供的带标签素材库E一致。根据本公开的各方面,可以相对快速地计算出许多分类器(例如,相对较浅的ANN)。这意味着分类可以在捕获数据之际立即发生并且可作为数据捕获过程的一部分被立即呈现给用户。例如,用户的智能电话相机取景器可以实时地在图像本身上显示蘑菇的估计类型。

[0222] 替换地,如果分类器是复杂的,则对用户的数据d的分类可通过首先在该设备上计算特征F(d)并将这些特征发送到服务器来回头在服务器上执行。服务器然后可计算分类C(F(d))并将结果发回用户设备。

[0223] 用户反馈

[0224] 当用户能够对其设备上的数据进行分类时,他们可任选地希望提供与该系统有关的反馈。这一反馈例如可采取以下形式:

[0225] 类型1:新标签,如果用户相信分类器已生成针对给定输入的不正确标签并知道正确标签应是什么;

[0226] 类型2:“错误标签”,如果用户相信分类器已针对给定输入生成了不正确标签,但不知道正确标签应是什么;或者

[0227] 类型3:要加载不同分类器的请求,如果基于分类器的初始结果,用户想要对相同的数据应用更专用的分类器。

[0228] 反馈可被提供给用户设备、服务器、专家或指定用户、或用户群或其它实体。在一些方面,类型1反馈可被用于构建私有分类器。例如,在用户能提供附加带标签示例的场合,可以从专家提供的分类器推导出私有分类器。类型2反馈可以被隔离地或优选地与来自其它用户的反馈相组合地使用以便通过提供带否定标签的示例来重新训练该分类器。

[0229] 类型3反馈可被用于构建对象类至其它分类器之间的关联的数据库。例如,使用用于宽泛对象分类的分类器的某人可能拍摄了苹果,接收标签“苹果”,并且然后切换至更具体的苹果分类器以确定苹果的具体品种。该动作能以反馈的形式被捕捉到,以使得提供标签“苹果”的其它分类器也可自动向用户提供切换至相同的具体苹果分类器的选项。通过累积这一类型3反馈,系统可将多个分类器排序或组织成分类器阶级并允许以上下文相关方式切换至更具体的分类器。要从一个分类器切换至更具体的分类器的决定可被自动化并且基于例如用户在特定对象上停留多久、或者单个图像中呈现了类的多少实例(如果图像中有许多“苹果”,则例如更具体的苹果分类器可能是有用的)。

[0230] 模型商店

[0231] 这些分类器和专家模型的前端清算场所可以是模型商店。模型商店可允许某些用户(例如,专家)上传其带标签数据集并为使用其数据集构建的分类器设置价格。模型商店还可允许用户购买具有上述后端过程的模型以用于训练模型并重用共享特征的效率。

[0232] 模型商店中的定价可允许对每一专家的带标签数据集的一次性定价或者可允许

组合定价。组合定价的示例可包括给定特征集上的第一分类器的更高价格以及用相同的特征集构建的后续分类器的降低的价格。替换地，组合定价可包括第一蘑菇分类器的给定价格以及来自其它专家的后续蘑菇分类器的折扣定价。

[0233] 后端可计算添加附加专家标记数据的增量式改进的某些联合性能得分以帮助确定增量价格。模型商店还可显示用以帮助用户选择要购买哪些专家数据集的度量，诸如分类准确度、带标签图像数等。

[0234] 模型商店还可允许用户上传例如来自在他们的手机上捕获的一些蘑菇图像的特征F(d)以评估哪个蘑菇分类器最适合其数据。在来自用户的样本图像上达成最高分类准确度的模型将会是要购买的那个模型

[0235] 其它示例性用例

[0236] 在一些方面，用户可以从“应用商店”或者与共同特征集相组合地工作的其它应用销路来购买特定分类器。

[0237] 在一方面，水果和蔬菜的粗略分类器可标识将用户设备感测到的对象标识为苹果。此外，通过停留或悬停在对象上方，分类器可加载更具体的苹果分类器（例如，对照相同的共同特征函数训练的分类器）以告知用户他们正看着Granny Smith苹果。在一些方面，分类器可标识可以对对象进行进一步分类的一个或多个其它分类器。

[0238] 在另一方面，知道很多关于火车的事情的用户可购买专家的火车分类器并用其自己的知识来对其进行扩增。

[0239] 在又一方面，到达沃斯旅行的戴着谷歌眼镜的用户可创建将两个专家分类器——名人的脸和芝士——合并成一个组合式标记器以用于抬头显示的定制分类器。

[0240] 在又一方面，危险位置中的配备有相机或其它传感器的一群机器人可使用其相机（或其它传感器）以及无监督学习来发现用于表示其环境中的纹理（砂砾、草地、泥地、碎石）的良好视觉特征。另外，使用加速度计和里程表，一些侦查机器人可基于该机器人是否能在这种地形上前进来将标签“可通过”和“不可通过”指派给不同的纹理。机器人还可在这些特征之上学习自定义分类器。自定义分类器然后可以与这群机器人中的其余机器人共享。

[0241] 尽管本公开的各方面已经描述了尖峰发放神经元和尖峰发放神经元模型，但这仅仅是示例性的并且也可使用非尖峰发放神经元和神经元模型。此外，本文公开的概念和技术可用于尖峰发放和非尖峰发放分布式学习两者。

[0242] 图13解说了根据本公开的各方面的用于学习模型的方法1300。在框1302中，该过程从一个或多个用户接收一个或多个模型更新。在框1304中，该过程基于先前模型和模型更新来计算经更新模型。此外，在框1306中，该过程基于经更新模型来将与经更新模型的子集有关的数据传送到一个或多个用户。

[0243] 在一些方面，经更新模型可基于性能度量和/或模型容量来验证。

[0244] 在一些方面，经更新模型可基于在对模型更新的比较性分析的基础上检测离群者来计算。

[0245] 在一些方面，经更新模型可包括模型架构和/或学习速率的变化。该架构和/或学习速率基于对照验证数据的模型性能和/或权重更新的稀疏性来确定。

[0246] 在一些方面，该子集可以只包括该模型的新训练的层。在一些方面，该子集可包括模型的随机子集。

[0247] 图14解说了根据本公开的各方面的用于学习模型的方法1400。在框1402中,该过程基于共享式推断模型从服务器接收数据。在框1404中,该过程基于接收到的数据来生成包括一个或多个模型参数的模型。在框1406中,该过程基于模型来计算推断。在框1408中,该过程基于推断来计算一个或多个模型参数更新。此外,在框1410中,该过程基于模型参数更新来向服务器传送数据。

[0248] 在一些方面,该过程还包括使用本地高速缓存的训练示例来训练分类器。

[0249] 在一些方面,数据可基于当前模型更新与先前模型更新之间的差异来传送。例如,该差异可以在动量模型中被压缩或使用。

[0250] 在一些方面,计算模型参数更新和/或基于模型参数更新来传送数据包括选择要计算和/或发送的模型参数的随机子集。

[0251] 图15解说了根据本公开的各方面的用于更新分类器集合的方法1500。在框1502中,该过程将第一分类器集合应用于第一数据集。该数据可包括传感器数据或存储在用户设备上的其它数据。此外,在框1504中,该过程基于第一分类器集合的输出和/或应用第一分类器集合的性能量度来向远程设备请求分类器更新。

[0252] 在一些方面,该请求可基于上下文信息。上下文信息例如可包括用户输入信息、给定时间段(例如,天、周、月等)内的观察数、位置、活动、加速度计、剩余电池寿命(例如,如果电池寿命低,则可指示低复杂性分类器)。在附加方面,该请求可基于计算负载。例如,在计算负载高(例如,高于预定阈值)的场合,可指示较低复杂性分类器。另一方面,在计算负载低(例如,低于预定阈值)的场合,可使用更复杂的分类器。

[0253] 在一些方面,性能量度可包括分类器的准确度或置信度、多个分类器的契合性的指示、分类器计算的速度等。

[0254] 图16解说了根据本公开的各方面的用于生成分类器模型的方法1600。在框1602中,该过程将共同特征模型分发至各用户。在框1604中,该过程在共同特征模型之上训练分类器。此外,在框1606中,该过程将第一分类器分发至第一用户并将第二分类器分发至第二用户。

[0255] 在一些方面,可以在从实体获得的带标签数据集上训练一个或多个分类器。实体可包括用户、某一指定用户或其它实体。可以为在从实体获得的带标签数据集上训练的每一分类器提供度量。该度量例如可包括关于分类准确度或带标签图像的数目的信息。

[0256] 在一些方面,该过程接收从远程设备上的数据中计算出的一个或多个特征。另外,该过程基于该一个或多个特征来确定用于对远程设备上的数据进行分类的一个或多个分类器。进而,该过程将该一个或多个分类器的指示分发至远程设备。

[0257] 在一些方面,该过程接收从远程设备上的数据计算出的特征。该过程还基于接收到的特征来计算分类。此外,该过程将该分类传送到该远程设备。

[0258] 在一些方面,该过程组合第一分类器和第二分类器以生成组合式分类器。该组合式分类器可被配置成对与一个或多个带标签数据集相关联的类进行分类。该过程还将该组合式分类器分发至一个或多个用户。

[0259] 图17解说了根据本公开的各方面的用于生成分类器模型的方法1700。在框1702中,该过程将共同特征映射图集应用于来自第一指定用户的带标签示例的第一素材库以学习第一分类器模型。在框1704中,该过程将该共同特征映射图集应用于来自第二指定用户

的带标签示例的第二素材库以学习第二分类器模型。此外,在框1706中,该过程将包括第一分类器模型和第二分类器模型的分类器模型分发至一个或多个用户。

[0260] 在一些方面,可生成组合式分类器。该组合式分类器可基于带标签示例的素材库、带标签示例的附加素材库并使用第一共同特征函数集和第二共同特征函数集来生成。如此,组合式分类器可被配置成对与带标签示例的一个或多个素材库以及附加素材库相关联的类进行分类。

[0261] 以上所描述的方法的各种操作可由能够执行相应功能的任何合适的装置来执行。这些装置可包括各种硬件和/或软件组件和/或模块,包括但不限于电路、专用集成电路(ASIC)、或处理器。一般而言,在附图中有解说的操作的场合,那些操作可具有带相似编号的相应配对装置加功能组件。

[0262] 如本文所使用的,术语“确定”涵盖各种各样的动作。例如,“确定”可包括演算、计算、处理、推导、研究、查找(例如,在表、数据库或其他数据结构中查找)、探知及诸如此类。另外,“确定”可包括接收(例如接收信息)、访问(例如访问存储器中的数据)、及类似动作。而且,“确定”可包括解析、选择、选取、确立及类似动作。

[0263] 如本文所使用的,引述一系列项目中的“至少一个”的短语是指这些项目的任何组合,包括单个成员。作为示例,“a、b或c中的至少一个”旨在涵盖:a、b、c、a-b、a-c、b-c、以及a-b-c。

[0264] 结合本公开所描述的各种解说性逻辑框、模块、以及电路可用设计成执行本文所描述功能的通用处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现场可编程门阵列信号(FPGA)或其他可编程逻辑器件(PLD)、分立的门或晶体管逻辑、分立的硬件组件或其任何组合来实现或执行。通用处理器可以是微处理器,但在替换方案中,处理器可以是任何市售的处理器、控制器、微控制器、或状态机。处理器还可以被实现为计算设备的组合,例如DSP与微处理器的组合、多个微处理器、与DSP核心协同的一个或多个微处理器、或任何其它此类配置。

[0265] 结合本公开所描述的方法或算法的步骤可直接在硬件中、在由处理器执行的软件模块中、或在这两者的组合中体现。软件模块可驻留在本领域所知的任何形式的存储介质中。可使用的存储介质的一些示例包括随机存取存储器(RAM)、只读存储器(ROM)、闪存、可擦除可编程只读存储器(EPROM)、电可擦除可编程只读存储器(EEPROM)、寄存器、硬盘、可移动盘、CD-ROM,等等。软件模块可包括单条指令、或许多条指令,且可分布在若干不同的代码段上,分布在不同的程序间以及跨多个存储介质分布。存储介质可被耦合到处理器以使得该处理器能从/向该存储介质读写信息。在替换方案中,存储介质可以被整合到处理器。

[0266] 本文所公开的方法包括用于实现所描述的方法的一个或多个步骤或动作。这些方法步骤和/或动作可以彼此互换而不会脱离权利要求的范围。换言之,除非指定了步骤或动作的特定次序,否则具体步骤和/或动作的次序和/或使用可以改动而不会脱离权利要求的范围。

[0267] 所描述的功能可在硬件、软件、固件或其任何组合中实现。如果以硬件实现,则示例硬件配置可包括设备中的处理系统。处理系统可以用总线架构来实现。取决于处理系统的具体应用和整体设计约束,总线可包括任何数目的互连总线和桥接器。总线可将包括处理器、机器可读介质、以及总线接口的各种电路链接在一起。总线接口可用于尤其将网络适

配器等经由总线连接至处理系统。网络适配器可用于实现信号处理功能。对于某些方面,用户接口(例如,按键板、显示器、鼠标、操纵杆,等等)也可以被连接到总线。总线还可以链接各种其他电路,诸如定时源、外围设备、稳压器、功率管理电路以及类似电路,它们在本领域中是众所周知的,因此将不再进一步描述。

[0268] 处理器可负责管理总线和一般处理,包括执行存储在机器可读介质上的软件。处理器可用一个或多个通用和/或专用处理器来实现。示例包括微处理器、微控制器、DSP处理器、以及其他能执行软件的电路系统。软件应当被宽泛地解释成意指指令、数据、或其任何组合,无论是被称作软件、固件、中间件、微代码、硬件描述语言、或其他。作为示例,机器可读介质可包括随机存取存储器(RAM)、闪存、只读存储器(ROM)、可编程只读存储器(PROM)、可擦式可编程只读存储器(EPROM)、电可擦式可编程只读存储器(EEPROM)、寄存器、磁盘、光盘、硬驱动器、或者任何其他合适的存储介质、或其任何组合。机器可读介质可被实施在计算机程序产品中。该计算机程序产品可以包括包装材料。

[0269] 在硬件实现中,机器可读介质可以是处理系统中与处理器分开的一部分。然而,如本领域技术人员将容易领会的,机器可读介质或其任何部分可在处理系统外部。作为示例,机器可读介质可包括传输线、由数据调制的载波、和/或与设备分开的计算机产品,所有这些都可由处理器通过总线接口来访问。替换地或补充地,机器可读介质或其任何部分可被集成到处理器中,诸如高速缓存和/或通用寄存器文件可能就是这种情形。虽然所讨论的各种组件可被描述为具有特定位置,诸如局部组件,但它们也可按各种方式来配置,诸如某些组件被配置成分布式计算系统的一部分。

[0270] 处理系统可以被配置为通用处理系统,该通用处理系统具有一个或多个提供处理器功能性的微处理器、以及提供机器可读介质中的至少一部分的外部存储器,它们都通过外部总线架构与其他支持电路系统链接在一起。替换地,该处理系统可以包括一个或多个神经元形态处理器以用于实现本文所述的神经元模型和神经系统模型。作为另一替换方案,处理系统可以用带有集成在单块芯片中的处理器、总线接口、用户接口、支持电路系统、和至少一部分机器可读介质的专用集成电路(ASIC)来实现,或者用一个或多个现场可编程门阵列(FPGA)、可编程逻辑器件(PLD)、控制器、状态机、门控逻辑、分立硬件组件、或者任何其他合适的电路系统、或者能执行本公开通篇所描述的各种功能性的电路的任何组合来实现。取决于具体应用和加诸于整体系统上的总设计约束,本领域技术人员将认识到如何最佳地实现关于处理系统所描述的功能性。

[0271] 机器可读介质可包括数个软件模块。这些软件模块包括当由处理器执行时使处理系统执行各种功能的指令。这些软件模块可包括传送模块和接收模块。每个软件模块可以驻留在单个存储设备中或者跨多个存储设备分布。作为示例,当触发事件发生时,可以从硬驱动器中将软件模块加载到RAM中。在软件模块执行期间,处理器可以将一些指令加载到高速缓存中以提高访问速度。随后可将一个或多个高速缓存行加载到通用寄存器文件中以供处理器执行。在以下述及软件模块的功能性时,将理解此类功能性是在处理器执行来自该软件模块的指令时由该处理器来实现的。

[0272] 如果以软件实现,则各功能可作为一条或多条指令或代码存储在计算机可读介质上或藉其进行传送。计算机可读介质包括计算机存储介质和通信介质两者,这些介质包括促成计算机程序从一地向另一地转移的任何介质。存储介质可以是能被计算机访问的任何

可用介质。作为示例而非限定,此类计算机可读介质可包括RAM、ROM、EEPROM、CD-ROM或其他光盘存储、磁盘存储或其他磁存储设备、或能用于携带或存储指令或数据结构形式的期望程序代码且能被计算机访问的任何其他介质。另外,任何连接也被正当地称为计算机可读介质。例如,如果软件是使用同轴电缆、光纤电缆、双绞线、数字订户线(DSL)、或无线技术(诸如红外(IR)、无线电、以及微波)从web网站、服务器、或其他远程源传送而来,则该同轴电缆、光纤电缆、双绞线、DSL或无线技术(诸如红外、无线电、以及微波)就被包括在介质的定义之中。如本文中所使用的盘(disk)和碟(disc)包括压缩碟(CD)、激光碟、光碟、数字多用碟(DVD)、软盘、和蓝光[®]碟,其中盘(disk)常常磁性地再现数据,而碟(disc)用激光来光学地再现数据。因此,在一些方面,计算机可读介质可包括非瞬态计算机可读介质(例如,有形介质)。另外,对于其他方面,计算机可读介质可包括瞬态计算机可读介质(例如,信号)。上述的组合应当也被包括在计算机可读介质的范围内。

[0273] 因此,某些方面可包括用于执行本文中给出的操作的计算机程序产品。例如,此类计算机程序产品可包括其上存储(和/或编码)有指令的计算机可读介质,这些指令能由一个或多个处理器执行以执行本文中所描述的操作。对于某些方面,计算机程序产品可包括包装材料。

[0274] 此外,应当领会,用于执行本文中所描述的方法和技术的模块和/或其它恰适装置能由用户终端和/或基站在适用的场合下载和/或以其他方式获得。例如,此类设备能被耦合至服务器以促成用于执行本文中所描述的方法的装置的转移。替换地,本文所述的各种方法能经由存储装置(例如,RAM、ROM、诸如压缩碟(CD)或软盘等物理存储介质等)来提供,以使得一旦将该存储装置耦合至或提供给用户终端和/或基站,该设备就能获得各种方法。此外,可利用适于向设备提供本文所描述的方法和技术的任何其他合适的技术。

[0275] 将理解,权利要求并不被限定于以上所解说的精确配置和组件。可在以上所描述的方法和装置的布局、操作和细节上作出各种改动、更换和变形而不会脱离权利要求的范围。

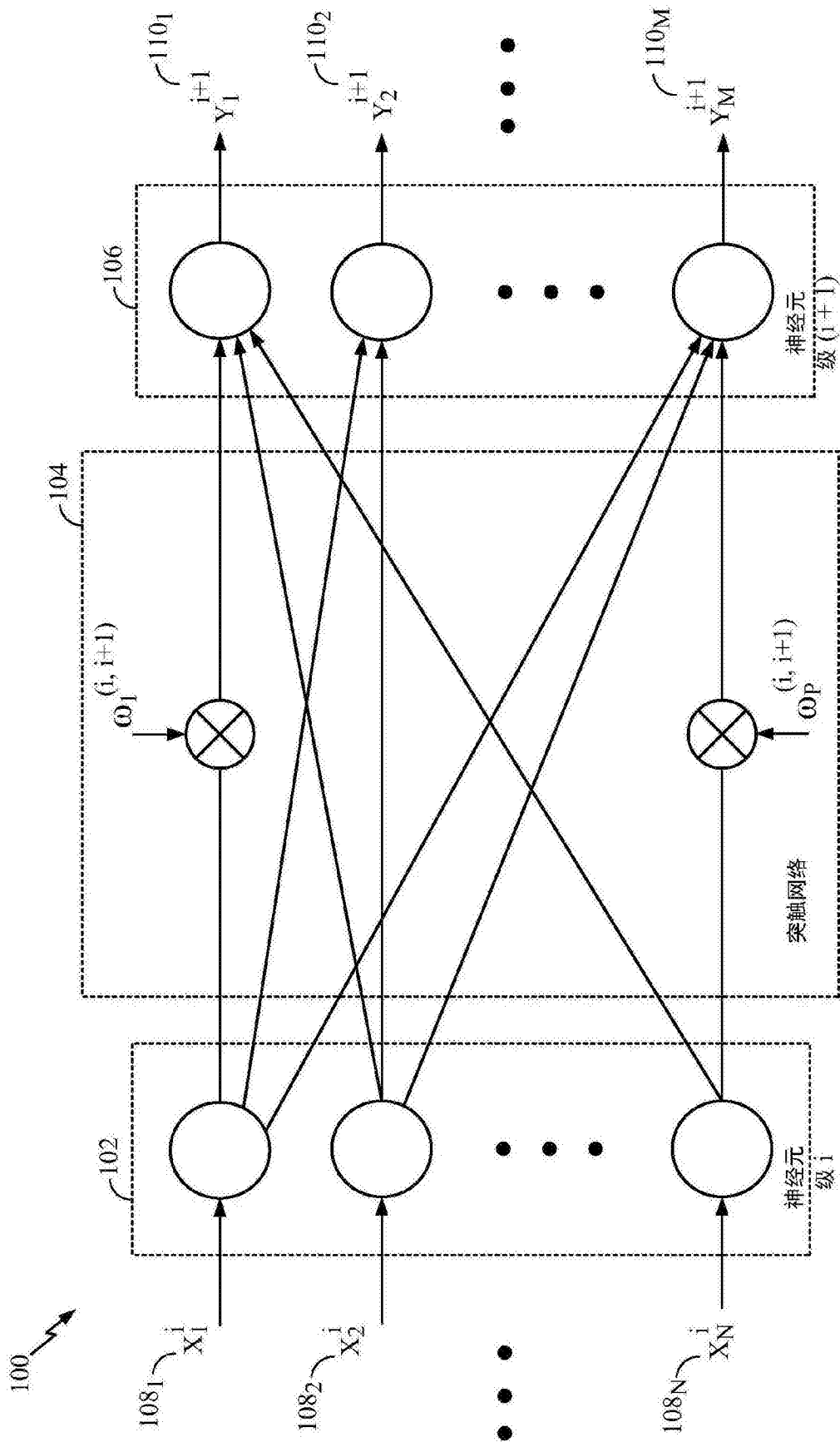


图1

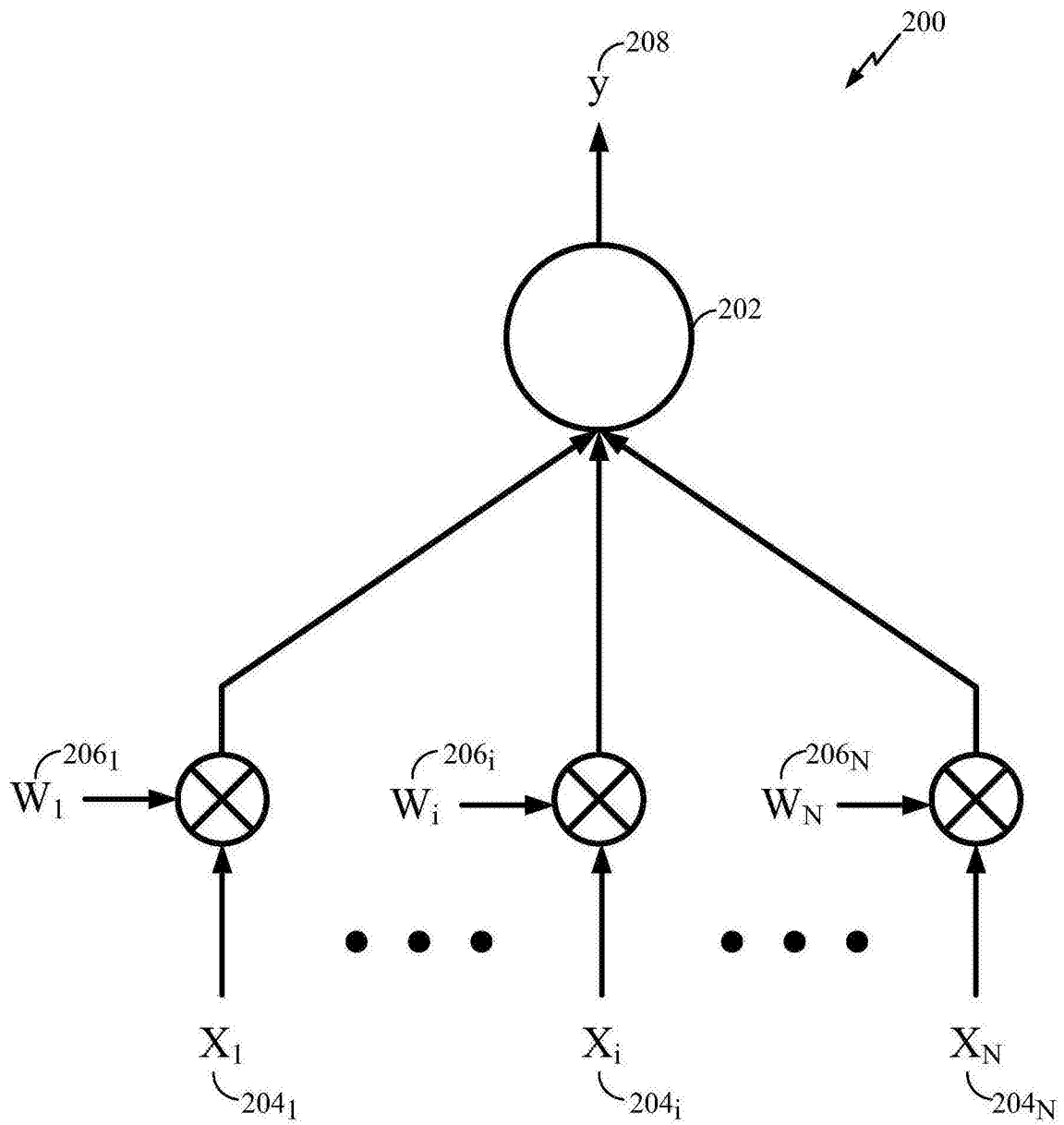


图2

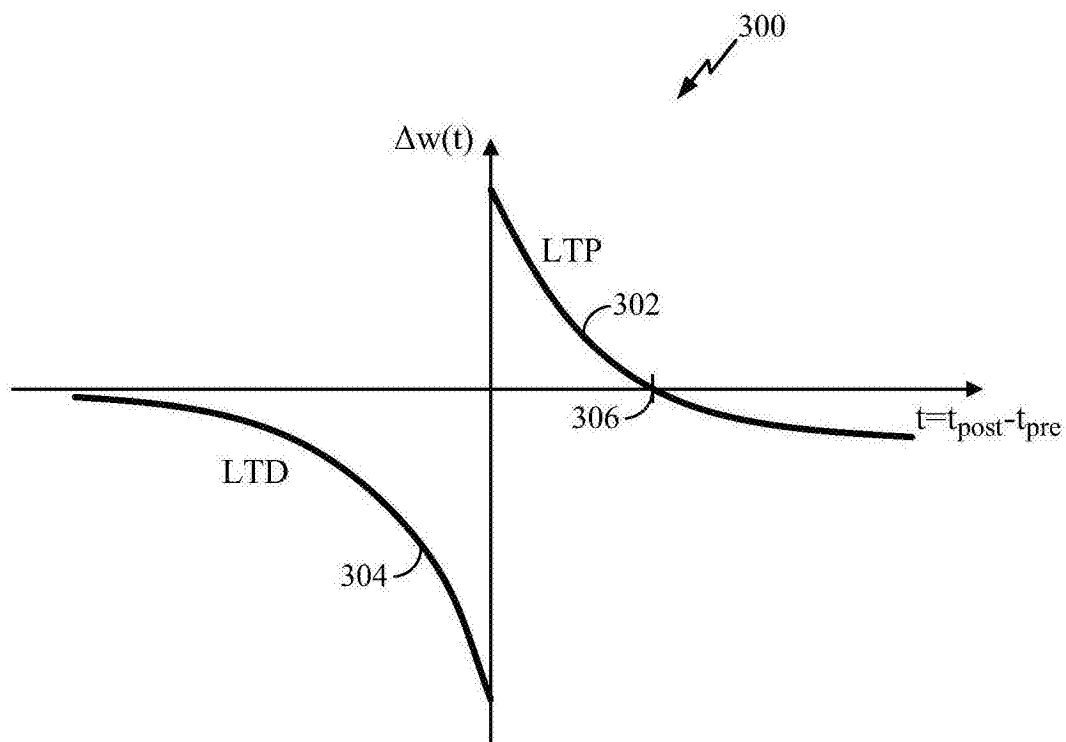


图3

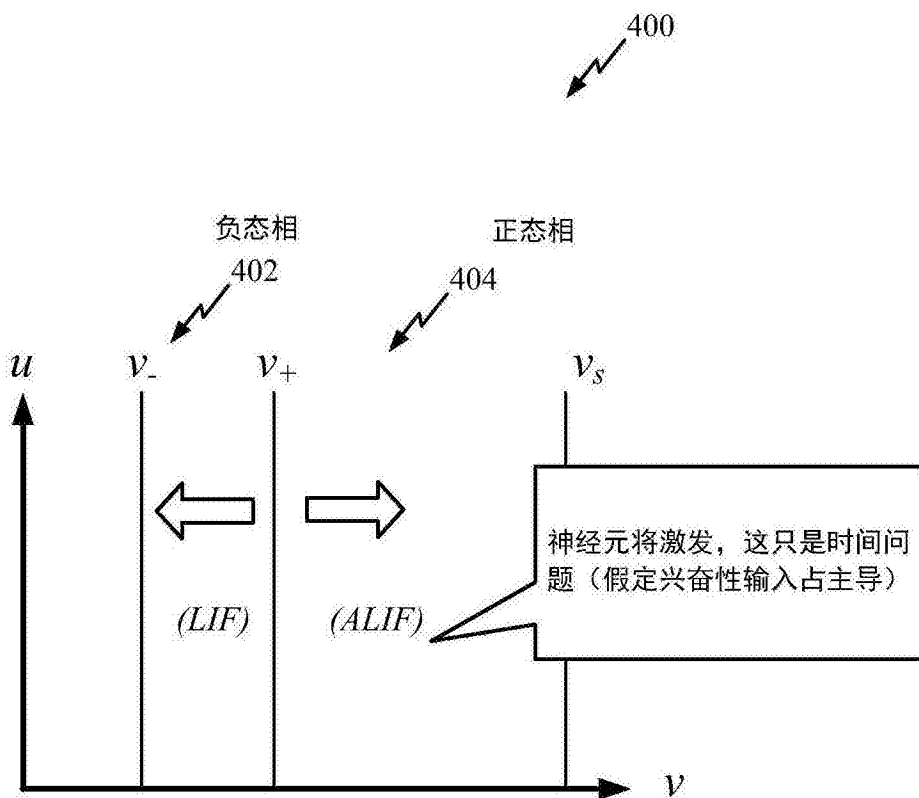


图4

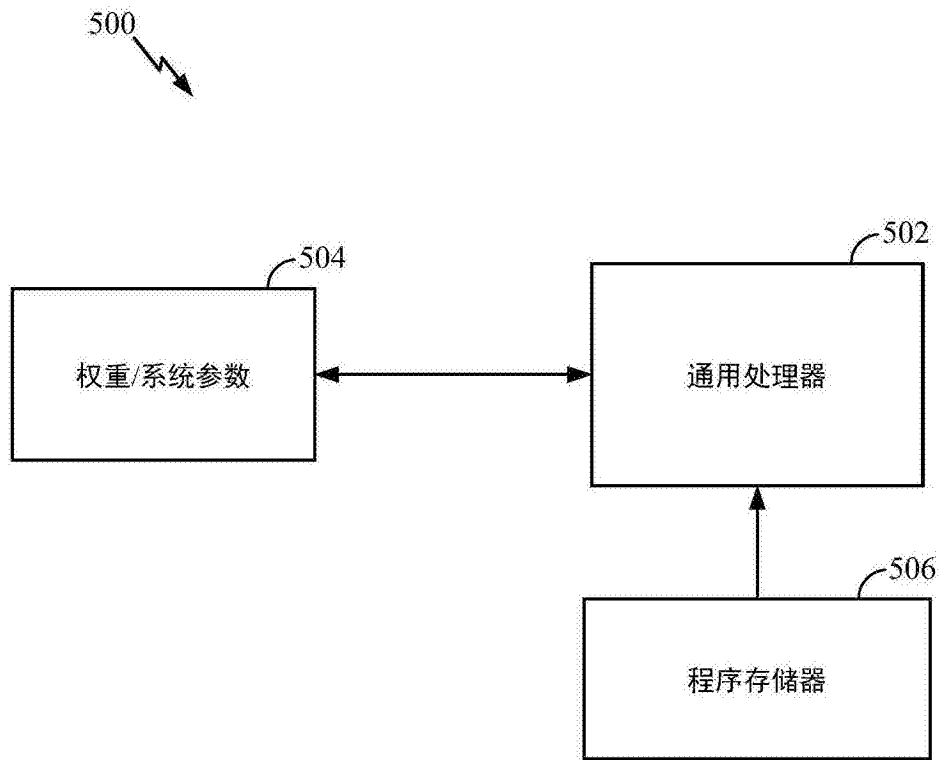


图5

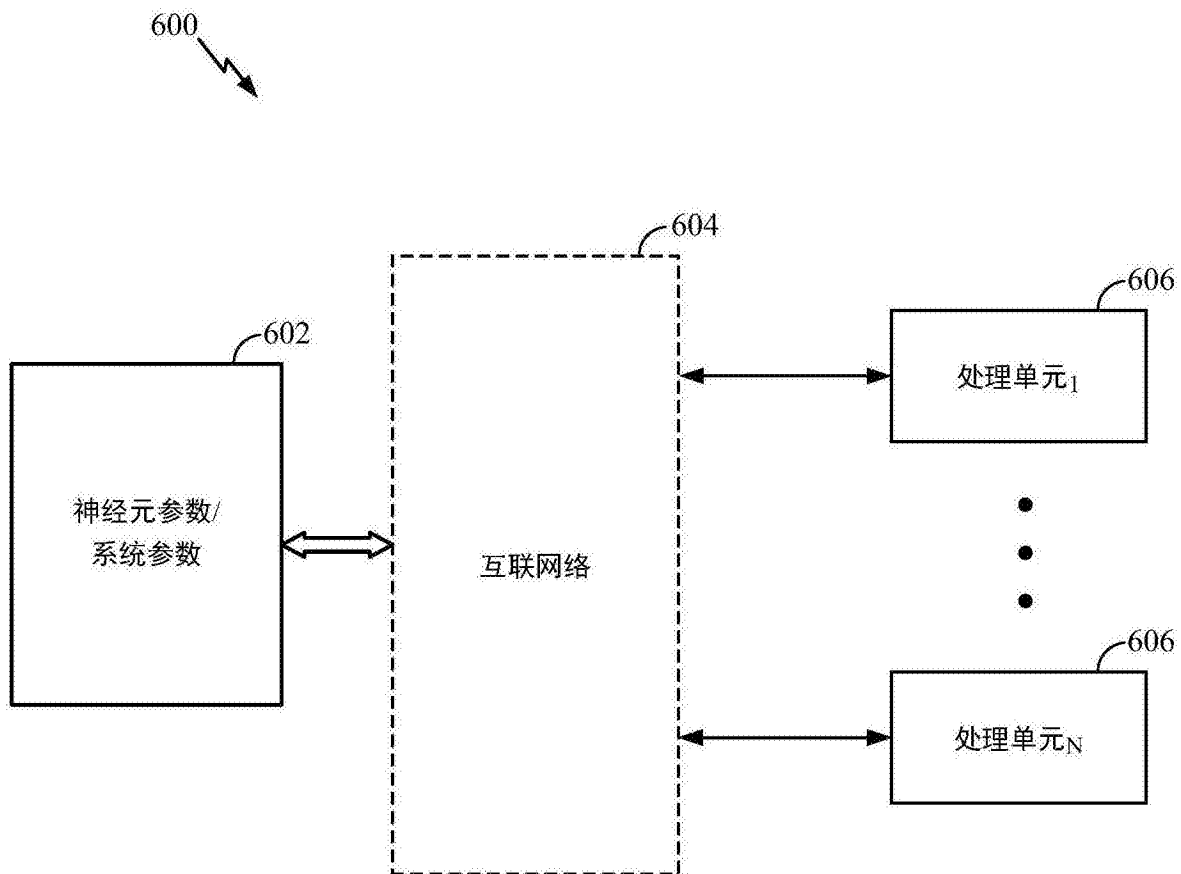


图6

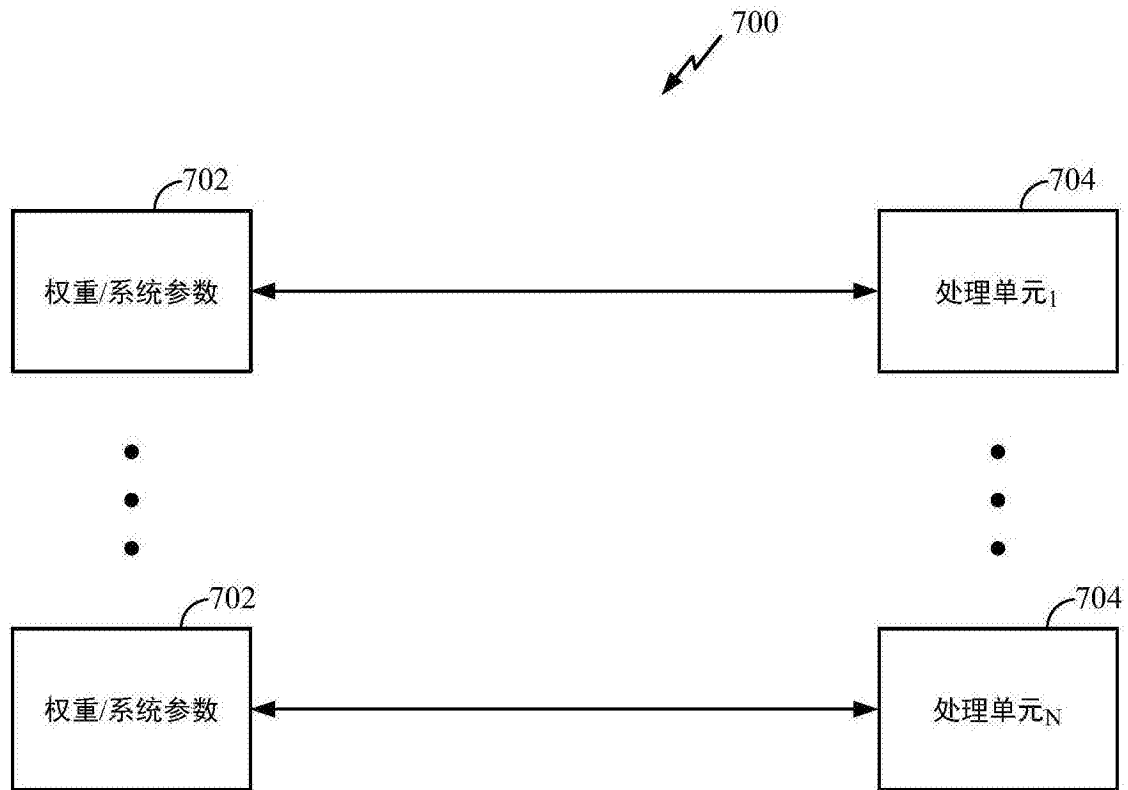


图7

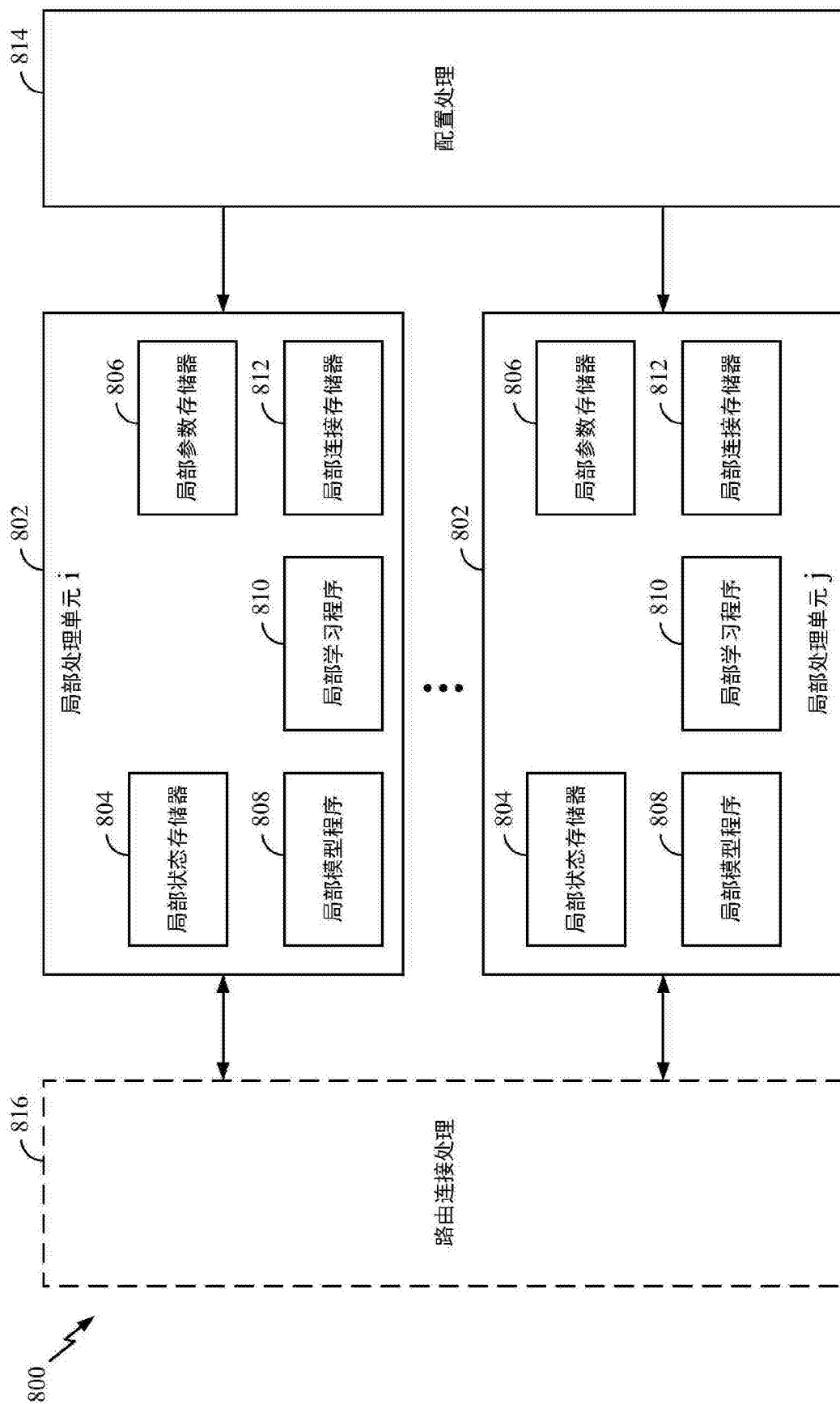


图8

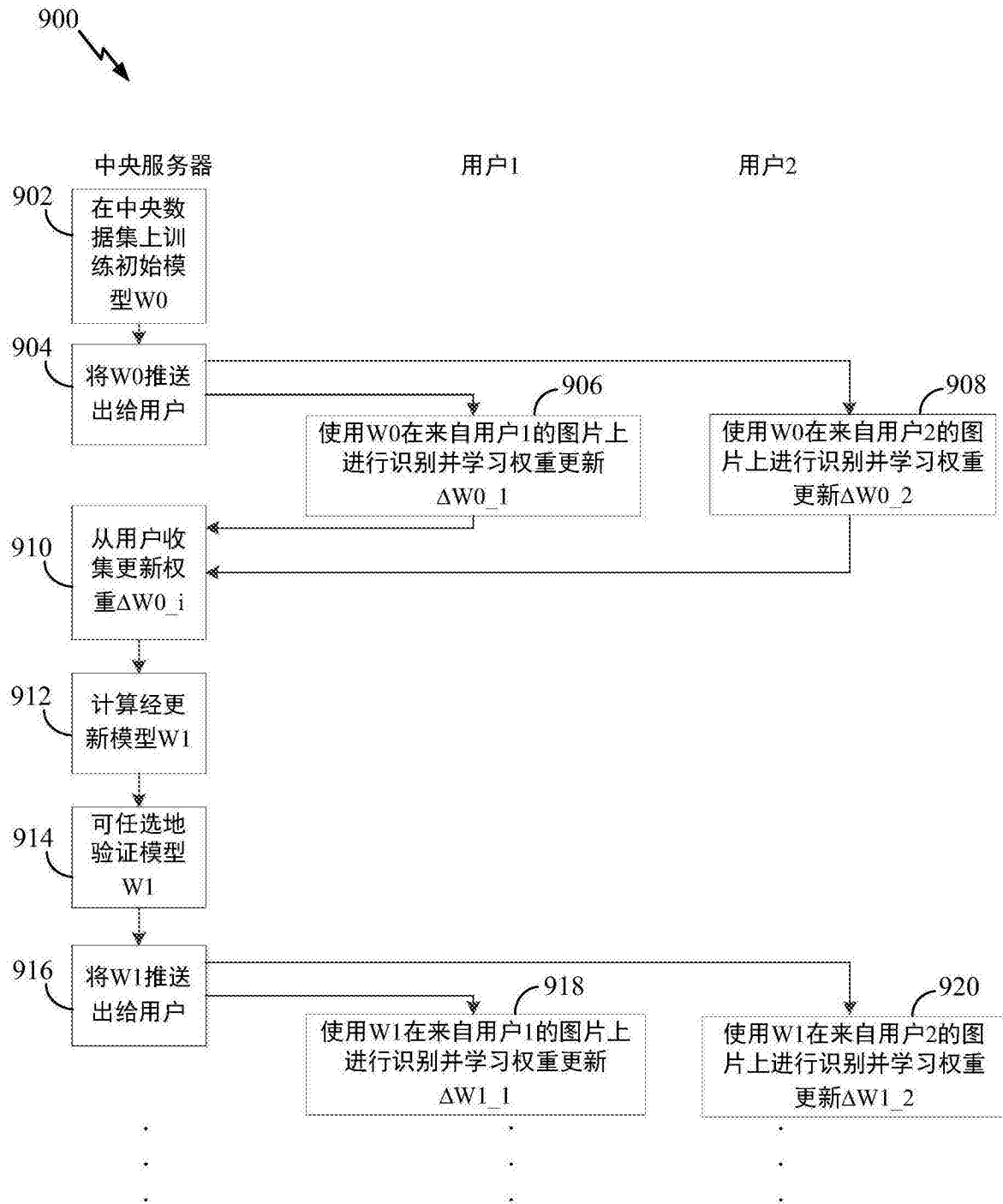


图9

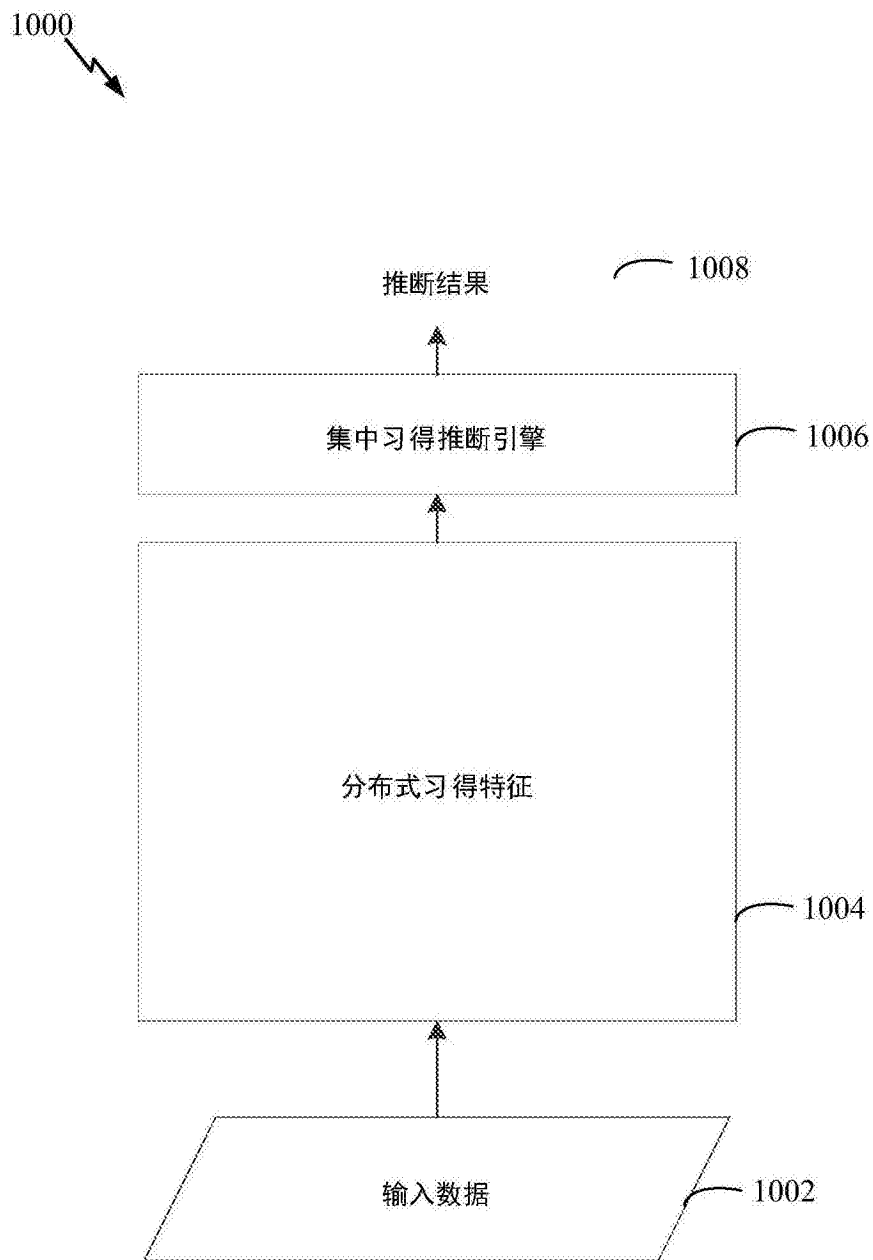


图10

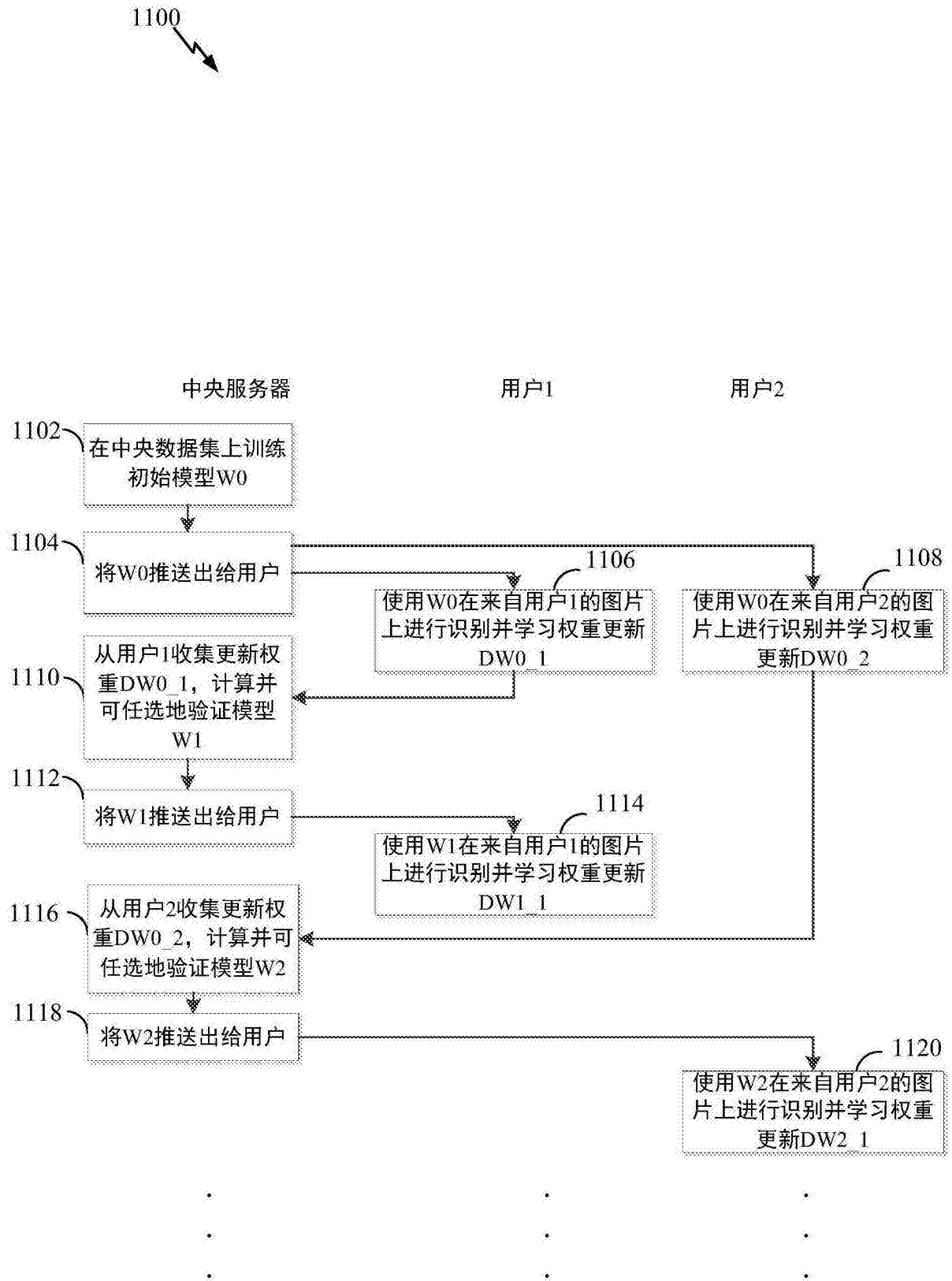


图11

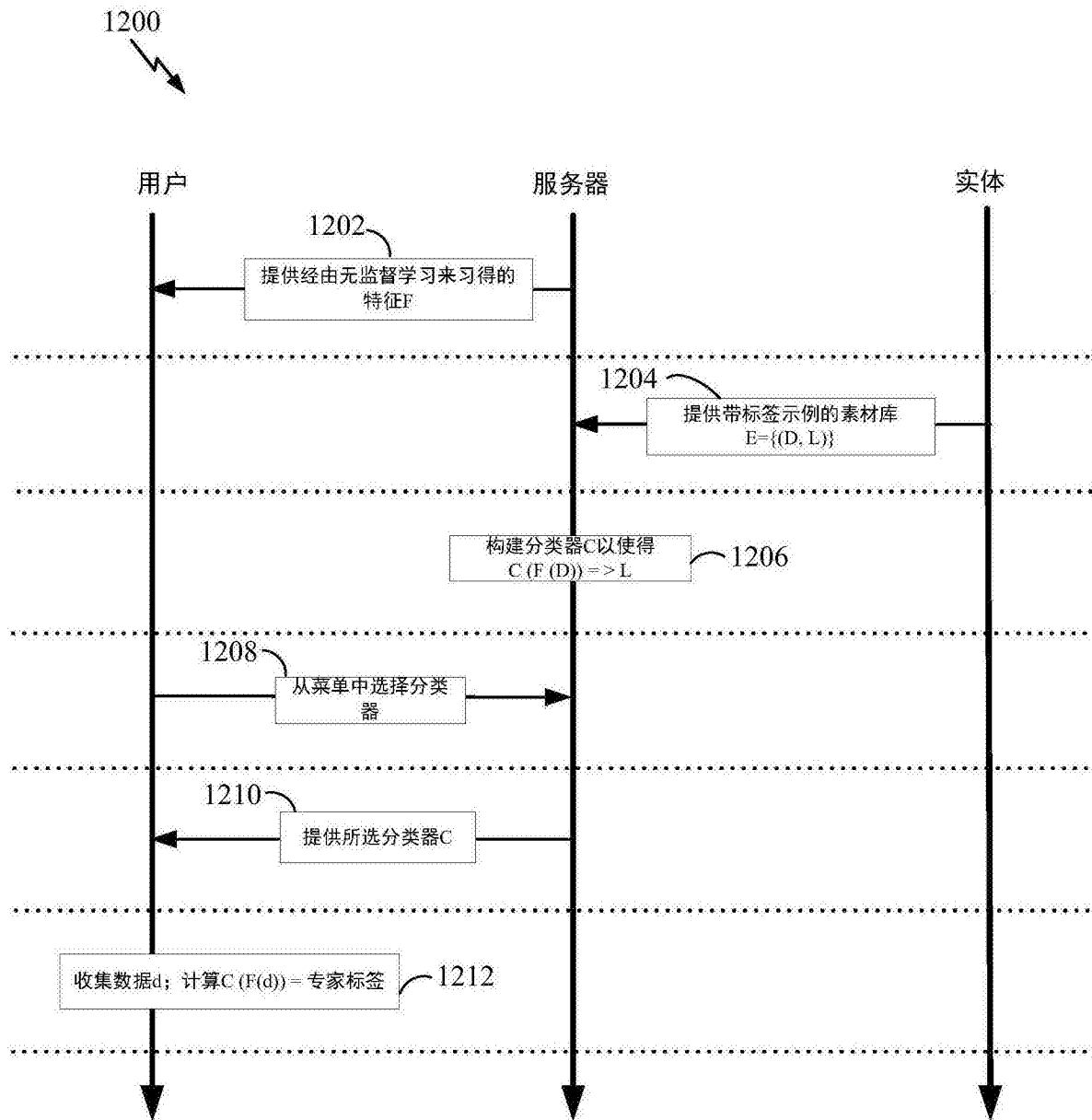


图12

1300

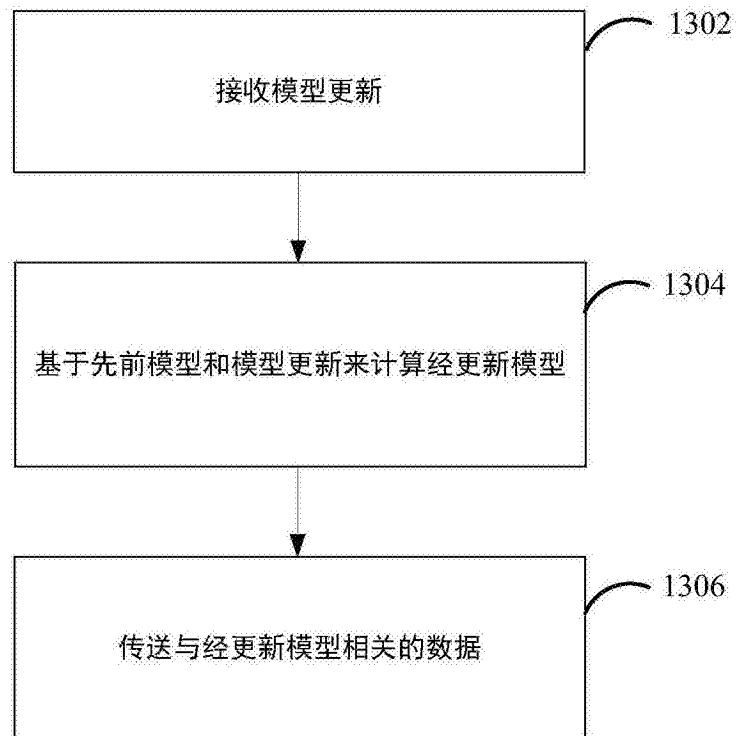


图13

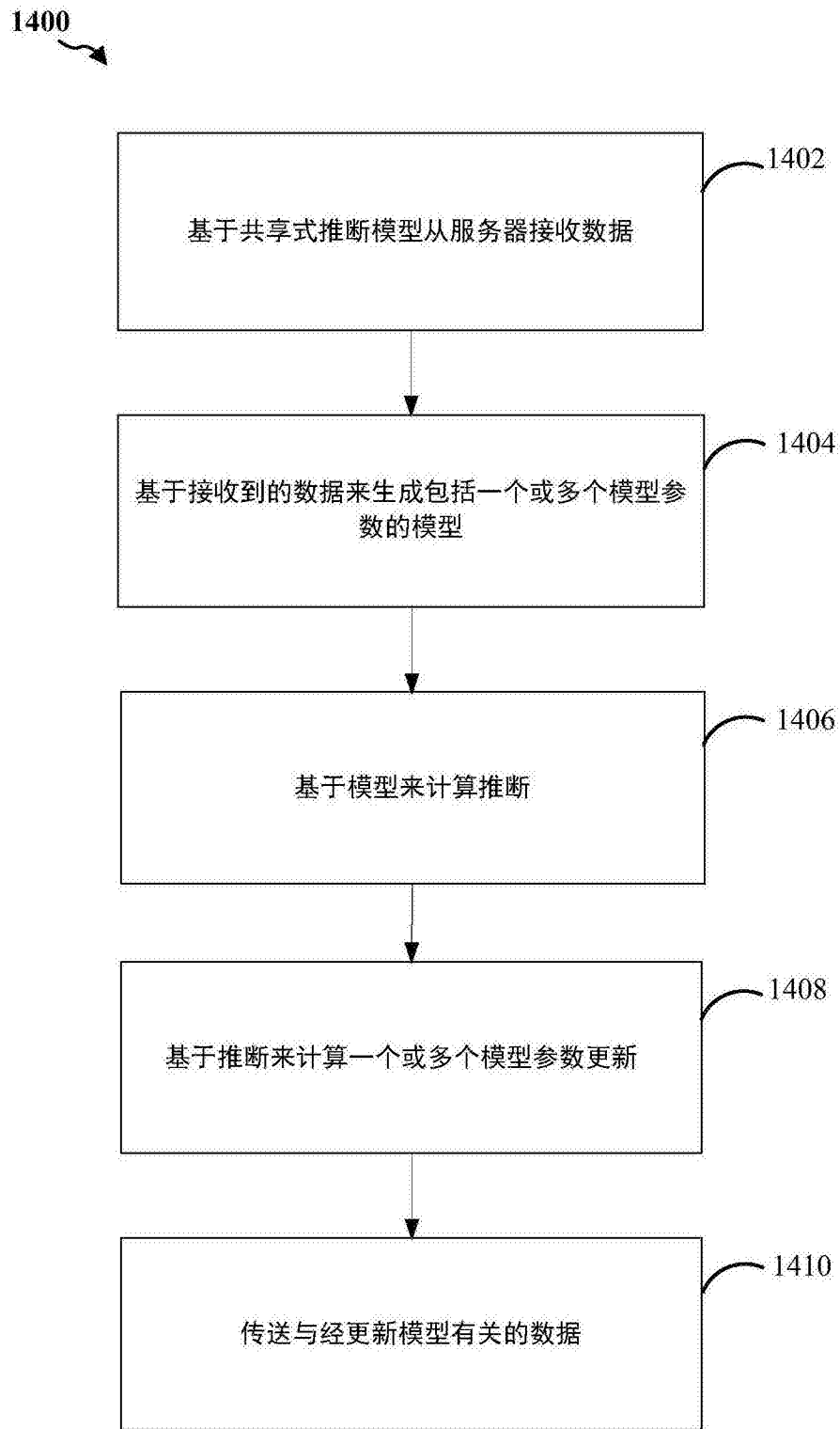


图14

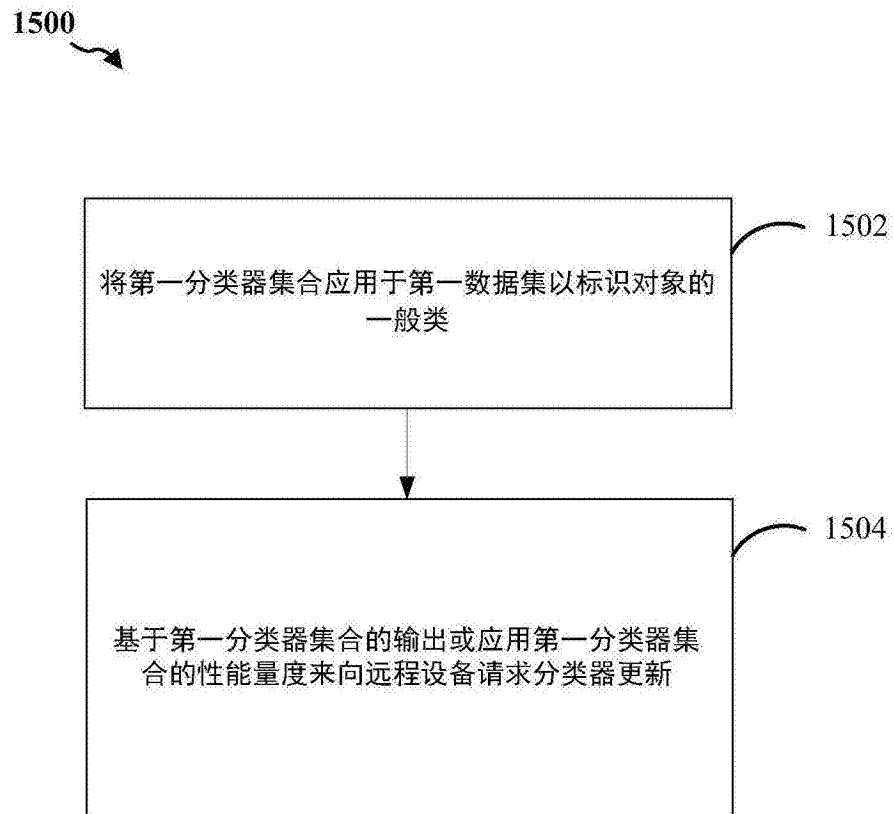


图15

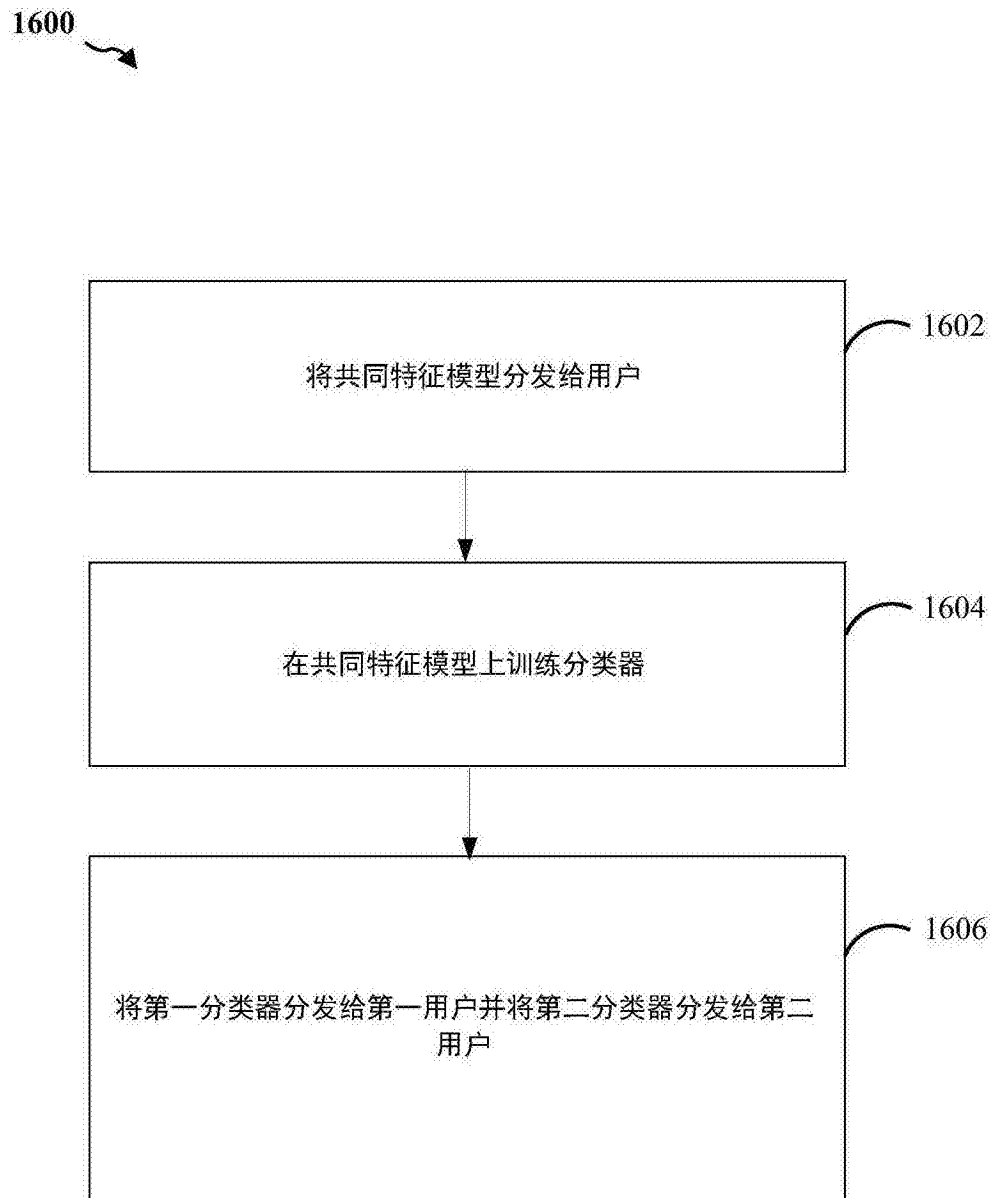


图16

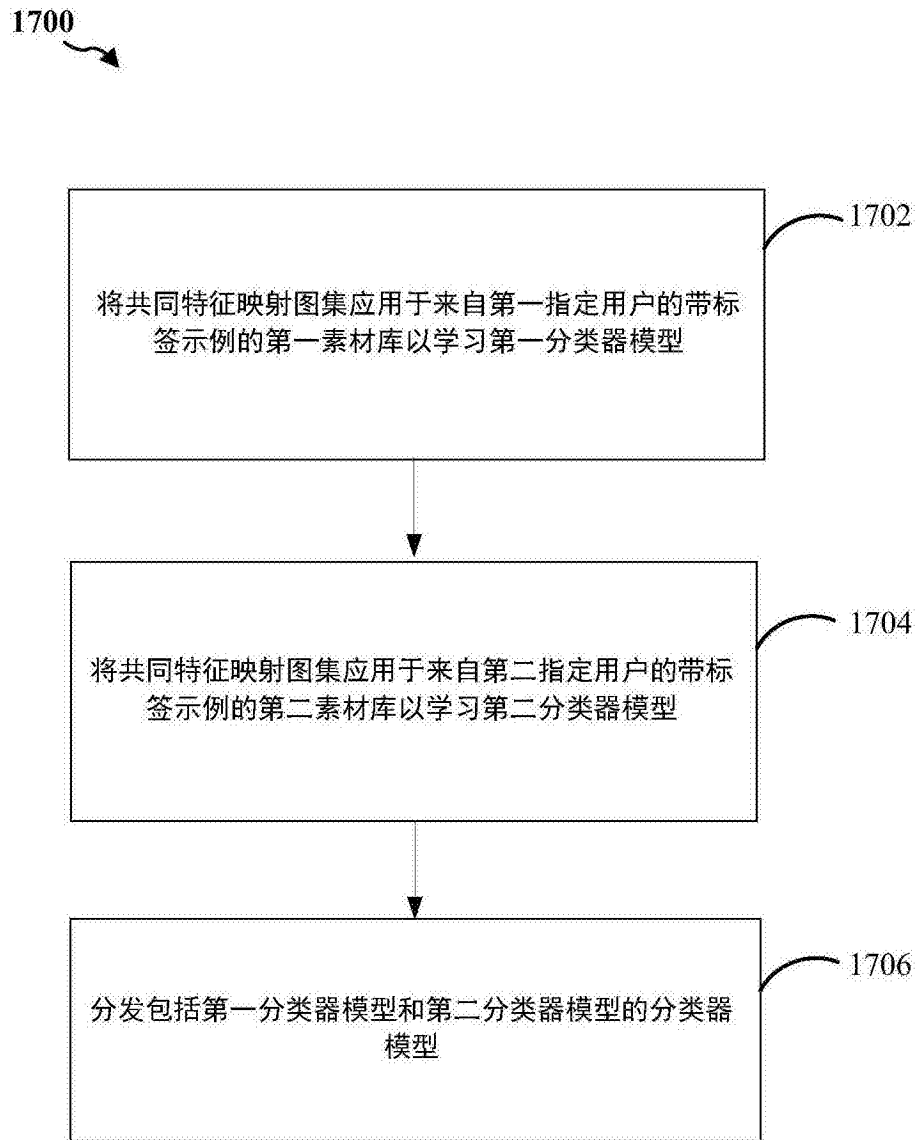


图17