



(12)发明专利申请

(10)申请公布号 CN 109328362 A

(43)申请公布日 2019.02.12

(21)申请号 201680085917.3

科拉伊·卡武克曲奥卢

(22)申请日 2016.12.30

拉亚·泰·哈德塞尔

(30)优先权数据

拉兹万·帕什卡努

62/339,719 2016.05.20 US

詹姆斯·柯克帕特里克

(85)PCT国际申请进入国家阶段日

赫伯特·约瑟夫·索耶

2018.11.20

(74)专利代理机构 中原信达知识产权代理有限公司 11219

(86)PCT国际申请的申请数据

代理人 李宝泉 周亚荣

PCT/US2016/069589 2016.12.30

(51)Int.Cl.

(87)PCT国际申请的公布数据

G06N 3/04(2006.01)

W02017/200597 EN 2017.11.23

(71)申请人 渊慧科技有限公司

权利要求书3页 说明书11页 附图3页

地址 英国伦敦

(72)发明人 尼尔·查尔斯·拉比诺维茨

纪尧姆·德雅尔丹

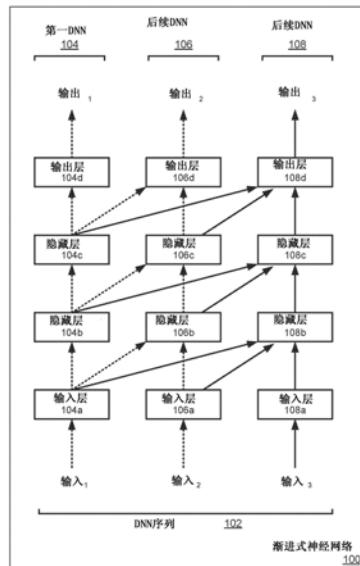
安德烈-亚历山德鲁·鲁苏

(54)发明名称

渐进式神经网络

(57)摘要

用于执行机器学习任务序列的方法和系统。一种系统包括深度神经网络(DNN)序列，所述深度神经网络(DNN)序列包括：与第一机器学习任务对应的第一DNN，其中，所述第一DNN包括第一多个索引层，并且所述第一多个索引层中的每个层被配置为接收相应的层输入并且处理所述层输入以生成相应的层输出；以及与一个或者多个相应的机器学习任务对应的一个或者多个后续DNN，其中，每个后续DNN包括相应的多个索引层，并且相应的多个索引层中具有大于1的索引的每个层从所述相应的后续DNN的先前层和相应的先前DNN的一个或者多个先前层接收输入，其中，先前层是其索引比当前索引小1的层。



1. 一种由一个或者多个计算机实现的神经网络系统,所述神经网络系统包括深度神经网络DNN序列,

其中,已经对所述DNN序列中的每个DNN进行了训练以执行相应的机器学习任务,以及其中,所述DNN序列包括:

与第一机器学习任务对应的第一DNN,其中, (i) 所述第一DNN包括第一多个索引层,并且 (ii) 所述第一多个索引层中的每个层被配置为接收相应的层输入并且处理所述层输入以生成相应的层输出;以及

与一个或者多个相应的机器学习任务对应的一个或者多个后续DNN,其中, (i) 每个后续DNN包括相应的多个索引层,并且 (ii) 相应的多个索引层中具有大于1的索引的每个层从 (i) 所述相应的后续DNN的先前层和 (ii) 相应的先前DNN的一个或者多个先前层接收输入,其中,先前层是其索引比当前索引小1的层。

2. 根据权利要求1所述的系统,其中,相应的多个索引层中具有等于1的索引的每个层接收相应的后续DNN输入。

3. 根据权利要求1或者2中的任一项所述的系统,其中, (i) 所述第一多个索引层包括L个隐藏层,并且 (ii) 每个相应的多个索引层s包括M_s个隐藏层。

4. 根据权利要求3所述的系统,其中,对于每个s,L不等于M_s。

5. 根据权利要求3所述的系统,其中,对于一个或者多个s,L等于M_s。

6. 根据权利要求3至5中的任一项所述的系统,其中,所述第一多个索引层中的每个层包括隐藏激活 $h_i^{(1)} \in \mathbb{R}^{n_i}$, 其中, n_i表示在层i≤L处的隐藏单元的数量。

7. 根据权利要求6所述的系统,其中,相应的多个索引层s中的每个层包括隐藏激活 $h_j^{(s)} \in \mathbb{R}^{n_j}$, 其中, n_j表示在层j≤M_s处的隐藏单元的数量。

8. 根据权利要求7所述的系统,其中,机器学习任务的序列包括k+1个机器学习任务,以及其中,第k个后续DNN的激活由以下等式给出:

$$h_i^{(k)} = \sigma \left(W_i^{(k)} h_{i-1}^{(k)} + \sum_{j < k} U_{ij}^{(k)} h_{i-1}^{(j)} \right)$$

其中, $h_i^{(k)}$ 表示第k个后续DNN的激活, $W_i^{(k)} \in \mathbb{R}^{n_1 \times n_{i-1}}$ 表示第k个后续DNN的层i的权重矩阵, $U_{ij}^{(k)} \in \mathbb{R}^{n_1 \times n_j}$ 表示从第k个后续DNN的层i到第j个后续DNN的层i-1的横向连接,并且 σ 表示元素形式的非线性。

9. 根据权利要求7所述的系统,其中,机器学习任务的序列包括k个机器学习任务,以及其中,第k个后续DNN的激活由以下等式给出:

$$h_i^{(k)} = \sigma \left(W_i^{(k)} h_{i-1}^{(k)} + U_{ij}^{(k)} \sigma \left(V_{ij}^{(k)} \alpha_{i-1}^{(<k)} h_{i-1}^{(<k)} \right) \right)$$

其中, $h_i^{(k)}$ 表示第k个后续DNN的激活, $W_i^{(k)} \in \mathbb{R}^{n_1 \times n_{i-1}}$ 表示第k个后续DNN的层i的权重矩阵, $U_{ij}^{(k)} \in \mathbb{R}^{n_1 \times n_j}$ 表示从第k个后续DNN的层i到第j个后续DNN的层i-1的横向连接, σ 表

示元素形式的非线性, $V_{ij}^{(k)} \in \mathbb{R}^{n_{i-1} \times n_{i-1}^{(<k)}}$ 表示投影矩阵, 并且 $\alpha_{i-1}^{(<k)}$ 是学习标量。

10. 根据权利要求1至9中的任一项所述的系统, 其中, 机器学习任务的序列包括独立的机器学习任务。

11. 根据权利要求1至10中的任一项所述的系统, 其中, 机器学习任务的序列包括 (i) 对抗机器学习任务、(ii) 分类任务、(iii) 机器人学习任务或者 (iv) 生成建模任务中的一个或者多个。

12. 一种用于按顺序训练深度神经网络DNN序列以执行机器学习任务序列的计算机实现的方法, 所述序列中的每个DNN与相应的机器学习任务对应, 并且所述方法包括:

对于所述序列中的第一机器学习任务:

训练所述序列中与所述第一机器学习任务对应的第一DNN以执行所述第一机器学习任务, 其中, (i) 所述第一DNN包括第一多个索引层, 并且 (ii) 所述第一多个索引层中的每个层被配置为接收相应的层输入并且处理所述层输入以生成相应的层输出;

对于所述序列中的每个后续机器学习任务:

训练与所述机器学习任务对应的后续DNN以执行所述机器学习任务, 其中, (i) 所述后续DNN包括后续的多个索引层, 并且 (ii) 所述后续的多个索引层中具有大于1的索引的每个层从 (i) 所述后续DNN的先前层和 (ii) 相应的先前DNN的一个或者多个先前层接收输入, 其中, 先前层是其索引比当前索引小1的层。

13. 根据权利要求12所述的方法, 其中, 相应的多个索引层中具有等于1的索引的每个DNN层接收相应的后续DNN输入。

14. 根据权利要求12或者13中的任一项所述的方法, 其中, 所述机器学习任务序列包括独立的机器学习任务。

15. 根据权利要求12至14中的任一项所述的方法, 其中, 所述机器学习任务序列包括 (i) 对抗机器学习任务、(ii) 分类任务、(iii) 机器人学习任务或者 (iv) 生成建模任务中的一个或者多个。

16. 根据权利要求12至15中的任一项所述的方法, 其中, (i) 所述第一DNN包括一个或者多个相应的DNN参数, 并且 (ii) 每个后续DNN包括一个或者多个相应的后续DNN参数, 以及

其中, 训练每个后续DNN包括: 将先前DNN的先前DNN参数设置为常量值。

17. 根据权利要求12至16中的任一项所述的方法, 其中, 训练每个后续DNN进一步包括: 使用机器学习训练技术来调整所述相应的后续DNN参数的值。

18. 一种用于使用深度神经网络DNN序列来处理输入的计算机实现的方法, 其中, 已经对所述DNN序列中的每个DNN进行了训练以执行相应的机器学习任务, 所述DNN序列包括:

与第一机器学习任务对应的第一DNN, 其中, (i) 所述第一DNN包括第一多个索引层, 并且 (ii) 所述第一多个索引层中的每个层被配置为接收相应的层输入并且处理所述层输入以生成相应的层输出; 以及

与一个或者多个相应的机器学习任务对应的一个或者多个后续DNN, 其中, (i) 每个后续DNN包括相应的多个索引层, 并且 (ii) 相应的多个索引层中具有大于1的索引的每个层从 (i) 所述相应的后续DNN的先前层和 (ii) 相应的先前DNN的一个或者多个先前层接收输入, 其中, 先前层是其索引比当前索引小1的层;

所述方法包括：

接收作为与所述DNN序列中的最后的后续DNN对应的机器学习任务的一部分的输入；以及

使用所述序列中的所述最后的后续DNN来处理所述输入以生成所述机器学习任务的最后的后续DNN输出。

19. 根据权利要求18所述的方法，其中，相应的多个索引层中具有等于1的索引的每个层接收相应的后续DNN输入。

20. 根据权利要求18或者19中的任一项所述的方法，其中，与所述DNN序列中的所述最后的后续DNN对应的所述机器学习任务包括 (i) 对抗机器学习任务、(ii) 分类任务、(iii) 机器人学习任务或者 (iv) 生成建模任务。

21. 一种计算机可读指令，所述计算机可读指令在由至少一个计算机执行时使所述至少一个计算机引起执行根据权利要求12至20中的任一项所述的方法。

22. 一种包括一个或者多个计算机和存储指令的一个或者多个存储装置的系统，所述指令在由所述一个或者多个计算机执行时使所述一个或者多个计算机执行根据权利要求12至20中的任一项所述的方法的操作。

23. 一种计算机可读指令，所述计算机可读指令在由至少一个计算机执行时使所述至少一个计算机实现根据权利要求1至11中的任一项所述的系统。

渐进式神经网络

背景技术

[0001] 本说明书涉及神经网络。

[0002] 神经网络是采用一个或者多个非线性单元层来预测接收到的输入的输出的机器学习模型。除了输出层之外,一些神经网络还包括一个或者多个隐藏层。每个隐藏层的输出用作网络中下一层(即,下一隐藏层或者输出层)的输入。网络的每个层根据相应参数集合的当前值来从接收到的输入生成输出。

发明内容

[0003] 本说明书描述了被实施为位于一个或者多个位置的一个或者多个计算机上的计算机程序的系统可以按顺序学习多个机器学习任务的方式,其中,保留了任务特征,使得新任务可以受益于所有先前学过的特征。

[0004] 通常,本说明书中所描述的主题的一个创新方面可以体现为一种由一个或者多个计算机实现的神经网络系统,该神经网络系统包括深度神经网络(DNN)序列,其中,已经对DNN序列中的每个DNN进行了训练以执行相应的机器学习任务,以及其中,DNN序列包括:与第一机器学习任务对应的第一DNN,其中,(i)第一DNN包括第一多个索引层,并且(ii)第一多个索引层中的每个层被配置为接收相应的层输入并且处理该层输入以生成相应的层输出;以及与一个或者多个相应的机器学习任务对应的一个或者多个后续DNN,其中,(i)每个后续DNN包括相应的多个索引层,并且(ii)相应的多个索引层中具有大于1的索引的每个层从(i)相应的后续DNN的先前层和(ii)相应的先前DNN的一个或者多个先前层接收输入,其中,先前层是其索引比当前索引小1的层。

[0005] 单独地或者组合地,前述以及其它实施例可以分别可选地包括以下特征中的一个或者多个。在一些实施方式中,相应的多个索引层中具有等于1的索引的每个层接收相应的后续DNN输入。

[0006] 在一些实施方式中,(i)第一多个索引层包括L个隐藏层,并且(ii)每个相应的多个索引层s包括Ms个隐藏层。

[0007] 在一些实施方式中,对于每个s,L不等于Ms。

[0008] 在一些实施方式中,对于一个或者多个s,L等于Ms。

[0009] 在一些实施方式中,第一多个索引层中的每个层包括隐藏激活 $h_i^{(1)} \in \mathbb{R}^{n_i}$,其中, n_i 表示在层 $i \leq L$ 处的隐藏单元的数量。

[0010] 在一些实施方式中,相应的多个索引层s中的每个层包括隐藏激活 $h_j^{(s)} \in \mathbb{R}^{n_j}$,其中, n_j 表示在层 $j \leq M_s$ 处的隐藏单元的数量。

[0011] 在一些实施方式中,机器学习任务序列包括k+1个机器学习任务,以及其中,第k个后续DNN的激活由以下等式给出:

$$[0012] \quad h_i^{(k)} = \sigma \left(W_i^{(k)} h_{i-1}^{(k)} + \sum_{j < k} U_{ij}^{(k)} h_{i-1}^{(j)} \right)$$

[0013] 其中, $h_i^{(k)}$ 表示第 k 个后续DNN的激活, $W_i^{(k)} \in \mathbb{R}^{n_1 \times n_{i-1}}$ 表示第 k 个后续DNN的层 i 的权重矩阵, $U_{ij}^{(k)} \in \mathbb{R}^{n_1 \times n_j}$ 表示从第 k 个后续DNN的层 i 到第 j 个后续DNN的层 $i-1$ 的横向连接, 并且 σ 表示元素形式的非线性。

[0014] 在一些实施方式中, 机器学习任务序列包括 k 个机器学习任务, 以及其中, 第 k 个后续DNN的激活由以下等式给出:

$$[0015] \quad h_i^{(k)} = \sigma \left(W_i^{(k)} h_{i-1}^{(k)} + U_{ij}^{(k)} \sigma \left(V_{ij}^{(k)} \alpha_{i-1}^{(<k)} h_{i-1}^{(<k)} \right) \right)$$

[0016] 其中, $h_i^{(k)}$ 表示第 k 个后续DNN的激活, $W_i^{(k)} \in \mathbb{R}^{n_1 \times n_{i-1}}$ 表示第 k 个后续DNN的层 i 的权重矩阵, $U_{ij}^{(k)} \in \mathbb{R}^{n_1 \times n_j}$ 表示从第 k 个后续DNN的层 i 到第 j 个后续DNN的层 $i-1$ 的横向连接, σ 表示元素形式的非线性, $V_{ij}^{(k)} \in \mathbb{R}^{n_{i-1} \times n_{i-1}^{(<k)}}$ 表示投影矩阵, 并且 $\alpha_{i-1}^{(<k)}$ 是学习标量。

[0017] 在一些实施方式中, 机器学习任务序列包括独立的机器学习任务。

[0018] 在一些实施方式中, 机器学习任务序列包括 (i) 对抗机器学习任务、(ii) 分类任务、(iii) 机器人学习任务或者 (iv) 生成建模任务中的一个或者多个。

[0019] 通常, 本说明书中所描述的主题的另一创新方面可以体现为一种用于按顺序训练深度神经网络 (DNN) 序列以执行机器学习任务序列的方法, 序列中的每个DNN与相应的机器学习任务对应, 并且该方法包括: 对于序列中的第一机器学习任务: 训练序列中与第一机器学习任务对应的第一DNN以执行第一机器学习任务, 其中, (i) 第一DNN包括第一多个索引层, 并且 (ii) 第一多个索引层中的每个层被配置为接收相应的层输入并且处理该层输入以生成相应的层输出; 对于序列中的每个后续机器学习任务: 训练与机器学习任务对应的后续DNN以执行机器学习任务, 其中, (i) 后续DNN包括后续的多个索引层, 并且 (ii) 后续的多个索引层中具有大于1的索引的每个层从 (i) 后续DNN的先前层和 (ii) 相应的先前DNN的一个或者多个先前层接收输入, 其中, 先前层是其索引比当前索引小1的层。

[0020] 该方面的其它实施例包括记录在一个或者多个计算机存储装置上的对应计算机系统、设备、和计算机程序, 每个计算机系统、设备、和计算机程序配置为执行方法的动作。一个或者多个计算机的系统可以配置为借助于安装在系统上的软件、固件、硬件、或者其任何组合来执行特定操作或者动作, 该软件、固件、硬件、或者其任何组合在运行时可以使系统执行动作。一个或者多个计算机程序可以配置为借助于包括指令来执行特定操作或者动作, 该指令在由数据处理设备执行时使设备执行动作。

[0021] 单独地或者组合地, 前述以及其它实施例可以分别可选地包括以下特征中的一个或者多个。在一些实施方式中, 相应的多个索引层中具有等于1的索引的每个DNN层接收相应的后续DNN输入。

[0022] 在一些实施方式中, 机器学习任务序列包括独立的机器学习任务。

[0023] 在一些实施方式中, 机器学习任务序列包括 (i) 对抗机器学习任务、(ii) 分类任

务、(iii) 机器人学习任务、或者(iv) 生成建模任务中的一个或者多个。

[0024] 在一些实施方式中, (i) 第一DNN包括一个或者多个相应的DNN参数, 并且(ii) 每个后续DNN包括一个或者多个相应的后续DNN参数, 并且训练每个后续DNN包括: 将先前DNN的先前DNN参数设置为常量值。

[0025] 在一些实施方式中, 训练每个后续DNN进一步包括: 通过使用机器学习训练技术来调整相应的后续DNN参数的值。

[0026] 通常, 本说明书中所描述的主题的另一创新方面可以体现为一种用于通过使用深度神经网络(DNN)序列来处理输入的方法, 其中, 已经对DNN序列中的每个DNN进行了训练以执行相应的机器学习任务, DNN序列包括: 与第一机器学习任务对应的第一DNN, 其中, (i) 第一DNN包括第一多个索引层, 并且(ii) 第一多个索引层中的每个层被配置为接收相应的层输入并且处理该层输入以生成相应的层输出; 以及与一个或者多个相应的机器学习任务对应的一个或者多个后续DNN, 其中, (i) 每个后续DNN包括相应的多个索引层, 并且(ii) 相应的多个索引层中具有大于1的索引的每个层从(i) 相应的后续DNN的先前层和(ii) 相应的先前DNN的一个或者多个先前层接收输入, 其中, 先前层是其索引比当前索引小1的层; 该方法包括: 接收输入作为与DNN序列中的最后的后续DNN对应的机器学习任务的一部分; 以及通过使用序列中的最后的后续DNN来处理输入以生成机器学习任务的最后的后续DNN输出。

[0027] 该方面的其它实施例包括记录在一个或者多个计算机存储装置上的对应计算机系统、设备、和计算机程序, 每个计算机系统、设备、和计算机程序配置为执行方法的动作。一个或者多个计算机的系统可以配置为借助于安装在系统上的软件、固件、硬件、或者其任何组合来执行特定操作或者动作, 该软件、固件、硬件、或者其任何组合在运行时可以使系统执行动作。一个或者多个计算机程序可以配置为借助于包括指令来执行特定操作或者动作, 该指令在由数据处理设备执行时使设备执行动作。

[0028] 单独地或者组合地, 前述以及其它实施例可以分别可选地包括以下特征中的一个或者多个。在一些实施方式中, 相应的多个索引层中具有等于1的索引的每个层接收相应的后续DNN输入。

[0029] 在一些实施方式中, 与DNN序列中的最后的后续DNN对应的机器学习任务包括(i) 对抗机器学习任务、(ii) 分类任务、(iii) 机器人学习任务或者(iv) 生成建模任务。

[0030] 可以在特定实施例中实施本说明书中所描述的主题以实现以下优点中的一个或者多个。

[0031] 为了在神经网络中实施迁移学习, 系统通常执行微调方法。例如, 可以在源域(在该源域中, 数据通常是充足的)上预先训练神经网络, 可以使神经网络的输出层适应目标域, 并且可以通过使用机器学习技术(反向传播等技术)来对神经网络进行微调。在一些设置中, 例如, 当执行单个任务时, 这种微调方法可能就足够了。然而, 在其它设置中, 例如, 当跨多个任务实施迁移学习时, 这种方法会具有若干缺点并且可能不够。例如, 虽然微调方法可以允许恢复目标域中的专家性能, 但是这是丢弃先前学习过的与先前任务对应的神经网络参数的破坏性过程。为了克服这些问题, 可以在进行微调之前复制每个神经网络以明确地记住所有先前任务。

[0032] 在一些情况下, 蒸馏方法可以提供多任务学习的潜在解决方案。然而, 蒸馏方法假设训练数据可用于所有任务——这种假设可能并不总是成立的。

[0033] 如本说明书中描述的渐进式神经网络系统提供了一种模型架构,通过构造,该体系架构支持跨机器学习任务序列的迁移学习。虽然实施微调方法的系统仅在进行初始化时结合先前学习过的知识,但是渐进式神经网络系统的结构允许将来自先前训练过的模型的先前学习过的知识集成在当前模型的每一层。通过按照这种方式来组合先前学习过的知识,渐进式神经网络系统具有更复杂的组合性。与其它神经网络系统相比较,这种组合性可以使渐进式神经网络系统能够实现加速学习。

[0034] 另外,渐进式神经网络系统自然地累积经验并且不受设计遗忘的影响,这使得它们理想地适用于处理连续的或者终身的学习任务。

[0035] 如本说明书中描述的渐进式神经网络系统不假设被训练的机器学习任务是相关的或者相似的。例如,实施标准的预训练和微调方法的其它神经网络系统通常假设它们被训练的机器学习任务是重叠的。相反,渐进式神经网络系统不做这种假设,并且甚至可以应用于机器学习任务是正交的或者甚至是对抗的设置。具体地,渐进式神经网络系统为每个新任务分配新的深度神经网络,该深度神经网络的权重被随机初始化。这种深度神经网络可以通过先前深度神经网络之间的横向连接来重新使用、修改或者忽略先前学习过的特征。与在训练新的深度神经网络时冻结先前学习过的参数相结合,横向连接确保机器学习任务之间没有干扰并且防止发生灾难性遗忘。

[0036] 在附图和下面的描述中阐述了本说明书的主题的一个或者多个实施例的细节。本主题的其它特征、方面、和优点将通过描述、附图以及权利要求书而变得显而易见。

附图说明

- [0037] 图1示出了示例渐进式神经网络系统。
- [0038] 图2是训练渐进式神经网络系统以执行机器学习任务序列的示例过程的流程图。
- [0039] 图3是通过使用渐进式神经网络来处理输入的示例过程的流程图。
- [0040] 各个附图中的类似附图标记和标号指示类似的元件。

具体实施方式

[0041] 图1示出了示例渐进式神经网络系统100。系统100是被实施为位于一个或者多个位置的一个或者多个计算机上的计算机程序的系统的示例,在该一个或者多个计算机中,可以实施下面描述的系统、组件、和技术。

[0042] 渐进式神经网络系统100可以按顺序学习多个机器学习任务,其中,保留了任务特征,使得新任务可以受益于所有先前学习过的特征,并且使得在每个机器学习任务上,可以对最终的神经网络系统进行评估。

[0043] 在一些实施方式中,多个机器学习任务包括一个或者多个独立的机器学习任务,例如,其关系彼此正交的机器学习任务。在一些实施方式中,机器学习任务中的一个或者多个包括对抗机器学习任务。在一些实施方式中,机器学习任务序列中的机器学习任务可以包括不同的分类任务,诸如,图像处理任务、语音识别任务、自然语言处理任务、或者光学字符识别任务。例如,包括图像处理任务的机器学习任务序列可以包括一个或者多个不同的图像识别任务,其中,每个图像识别任务需要识别图像中的不同物体或者图案。作为另一示例,包括语音识别任务的机器学习任务序列可以包括一个或者多个热门词检测任务,其中,

每个任务需要识别不同的热门词或者热门词序列。

[0044] 在一些实施方式中,机器学习任务序列中的机器学习任务可以包括机器人学习任务。例如,机器学习任务序列中的第一机器学习任务可以包括感觉运动学习任务,诸如,移动。机器学习任务序列中的第二机器学习任务可以包括第二感觉运动学习任务,诸如,抓握。机器学习任务序列中的后续机器学习任务可以包括其它感觉运动学习任务(诸如,活动物体分类)或者交互式学习任务(诸如,与人类同伴一起联合操纵物体),或者语言学习任务,诸如,人类语言的情境化意义。

[0045] 在一些实施方式中,机器学习任务序列中的机器学习任务可以包括生成建模任务,例如,密度估计任务。

[0046] 渐进式神经网络系统100包括深度神经网络(DNN)序列102。例如,如在图1中示出的,DNN序列可以包括第一DNN 104和一个或者多个后续DNN。为了方便起见,图1示出了两个后续DNN,例如,后续DNN 106和108。然而,在一些情况下,渐进式神经网络系统100可以包括更多的或者更少的后续DNN。

[0047] 包括在DNN序列中的DNN的类型取决于DNN被训练执行的机器学习任务的类型。例如,在机器学习任务序列包括语言建模任务的情况下,DNN序列102可以包括一个或者多个递归神经网络,例如,长短期记忆神经网络。作为另一示例,在机器学习任务序列包括图像分类任务或者声学建模任务的情况下,DNN序列可以包括一个或者多个卷积神经网络。下面参照图2描述DNN序列可以被训练执行的其它示例机器学习任务。

[0048] DNN序列102中的第一DNN 104与机器学习任务序列中的第一机器学习任务对应。即,DNN序列102中的第一DNN 104是被配置为例如通过对适当的训练数据进行训练来执行第一机器学习任务的DNN。第一DNN 104包括多个神经网络层,例如,输入层104a、隐藏层104b和104c以及输出层104d。为了方便起见,将第一DNN 104图示为具有四个神经网络层,然而,在一些情况下,第一DNN 104可以包括更多的或者更少的神经网络层。例如,通常,第一DNN 104可以包括L个索引层,例如,层 $i=1, \dots, L$ 。层 $i=1$ 可以对应于第一DNN 104的底部输入层,例如,输入层104a,其被配置为接收第一DNN输入,例如,输入1。层 $i=L$ 可以对应于第一DNN 104的顶部输出层,例如,输出层104d,其被配置为输出第一DNN输出,例如,输出1。另外,为了方便起见,将DNN序列中的DNN图示为具有堆叠神经网络层的列。然而,在一些实施方式中,DNN序列中的DNN可以是包括其它类型的神经网络层(例如,长短期记忆递归神经网络层)的列。

[0049] 第一DNN 104的多个神经网络层中的每个神经网络层可以包括一个或者多个隐藏单元,并且可以被配置为接收相应的层输入(例如,从先前神经网络层)并且处理该层输入以生成相应的层输出。如上所述,在一些实施方式中,多个层可以包括L个索引神经网络层。

L个层中的每个层根据神经网络层的参数 $\Theta_i^{(1)}$ 来生成相应的层输出。神经网络层参数

$\Theta_i^{(1)}$ 是可以例如通过使用标准的反向传播技术来训练的参数。下面参照图2更详细地描述训练渐进式神经网络系统中的第一DNN。

[0050] DNN序列102中的后续DNN与机器学习任务序列中的后续机器学习任务对应。即,DNN序列中的每个后续DNN是可以被配置为例如通过对适当的训练数据进行训练来执行后续机器学习任务的DNN。每个后续DNN(例如,后续DNN 106和108)包括多个神经网络层。例

如,后续DNN 106包括输入层106a、隐藏层106b和106c以及输出层106d。同样,后续DNN 108包括输入层108a、隐藏层108b和108c以及输出层108d。为了方便起见,将后续DNN 106和108图示为分别具有四个神经网络层。然而,在一些情况下,后续DNN中的每个后续DNN可以包括更多的或者更少的神经网络层。例如,通常,后续DNN可以包括M个索引神经网络层,例如,i=1、…、M。层i=1可以对应于后续DNN的底部神经网络层,例如,输入层106a或者108a,其被配置为接收后续DNN输入,例如,输入2或者输入3。层i=M可以对应于后续DNN的顶部神经网络层,例如,输出层106d或者108d,其被配置为输出后续DNN输出,例如,输出2或者输出3。

[0051] 后续DNN中具有大于1的索引的每个神经网络层被配置为从(i)后续DNN的先前层和(ii)相应的先前DNN的一个或者多个先前层接收输入,其中,先前层是其索引比当前索引小1的层。例如,后续DNN 106的神经网络层106b被配置为从(i)DNN 106的神经网络层106a和(ii)DNN 104的神经网络层104a接收输入。同样,后续DNN 106的神经网络层106c被配置为从(i)DNN 106的神经网络层106b和(ii)DNN 104的神经网络层104b接收输入。

[0052] 作为一般示例,后续DNN的神经网络层1可以从DNN的神经网络层1-1接收输入,并且从相应的先前DNN的一个或者多个先前神经网络层1-1接收输入。例如,DNN序列中的第三DNN的神经网络层1可以从第三DNN的神经网络层1-1接收输入,并且从第一和第二DNN的神经网络层1-1接收两个输入。具有等于1的索引的神经网络层接收相应的DNN输入,例如,输入2和输入3。不需要DNN之间一对一对,例如,可以将后续DNN中的层连接至先前DNN的每个其它层。作为另一示例,可以根本不使用先前DNN中的一些DNN,例如,先前DNN的一个或者多个层可以仅包括先前DNN内的层之间的连接。

[0053] 在一些实施方式中,每个后续DNN可以包括Ms个神经网络层,其中,Ms个层中的每个层i包括隐藏激活 $h_i^{(s)} \in \mathbb{R}^{n_i}$,其中,n_i表示在层i≤Ms处的隐藏单元的数量。在一些情况下,第一DNN中的层数L可以不等于每个后续DNN中的层数,例如,对于每个s,L可以不等于Ms。在其它情况下,第一DNN中的层数L可以等于一个或者多个后续DNN中的层数,例如,对于一个或者多个s,L可以等于Ms。在一些情况下,后续DNN中的层数可以发生变化,例如,对于每个s,Ms不相同。

[0054] 如上所述,可以训练渐进式神经网络系统100以执行机器学习任务序列。DNN序列102中的第一DNN 104与机器学习任务序列中的第一机器学习任务对应,即,DNN序列102中的第一DNN 104是可以被配置为例如通过对适当的训练数据进行训练来执行第一机器学习任务的DNN。当接收到或者选择了机器学习任务序列中的第二机器学习任务时,将后续DNN(例如,DNN 106)实例化。每当接收到或者选择了机器学习任务序列中的后续机器学习任务时,将对应的后续DNN实例化。按照这种方式,渐进式神经网络系统100中的DNN的总数可以等于机器学习任务序列中的机器学习任务的数量。

[0055] 在一些实施方式中,机器学习任务序列可以包括K+1个机器学习任务。在这些情况下,第k个后续DNN的第i层的隐藏激活可以由以下等式给出:

$$[0056] h_i^{(k)} = \sigma \left(W_i^{(k)} h_{i-1}^{(k)} + \sum_{j < k} U_{ij}^{(k)} h_{i-1}^{(j)} \right) \quad (1)$$

[0057] 其中, $h_i^{(k)}$ 表示第k个后续DNN的激活, $W_i^{(k)} \in \mathbb{R}^{n_1 \times n_{i-1}}$ 表示第k个后续DNN的层i的

权重或者参数矩阵, $U_{ij}^{(k)} \in \mathbb{R}^{n_1 \times n_j}$ 表示从第k个后续DNN的层i到第k-1个后续DNN的层i-1的横向连接, 并且 σ 表示元素形式的非线性。在一些实施方式中, $\sigma(x) = \max(0, x)$ 。通常, 可以向不同的DNN层应用不同的激活——不要求在每个DNN层应用相同的变换。图1图示了上面描述的系统, 其中, $K=3$ 。

[0058] 在一些实施方式中, 可以利用至其它DNN层的非线性横向连接来增强DNN层。换句话说, 来自其它DNN层的连接不需要只表示来自其它DNN的信息的简单读数。相反, 连接可以表示信息的学习过的、可能非线性的变换。在一些情况下, 这种非线性横向连接可以改进初始调节并且执行降维。

[0059] 例如, 上面参照等式(1)描述了基本的横向连接。可以用其第一层是在 n_{i-1} 维子空间上的投影的单个隐藏层多层感知器(MLP)来代替这种基本的横向连接。通过用非线性横向连接来代替基本的横向连接, 源自横向连接的参数的数量可以与用于第一DNN的参数的量值相同。

[0060] 例如, 机器学习任务的序列可以包括 $K+1$ 个机器学习任务, 并且第k个后续DNN的第i层的激活可以由以下等式给出:

$$[0061] h_i^{(k)} = \sigma \left(W_i^{(k)} h_{i-1}^{(k)} + U_{ij}^{(k)} \sigma \left(V_{ij}^{(k)} \alpha_{i-1}^{(<k)} h_{i-1}^{(<k)} \right) \right)$$

[0062] 其中, $h_i^{(k)}$ 表示第k个后续DNN的激活, $W_i^{(k)} \in \mathbb{R}^{n_1 \times n_{i-1}}$ 表示第k个后续DNN的层i的权重或者参数矩阵, $U_{ij}^{(k)} \in \mathbb{R}^{n_1 \times n_j}$ 表示从第k个后续DNN的层i到第j个后续DNN的层i-1的横向连接, σ 表示元素形式的非线性, $V_{ij}^{(k)} \in \mathbb{R}^{n_{i-1} \times n_{i-1}^{(<k)}}$ 表示投影矩阵, 并且 $\alpha_{i-1}^{(<k)}$ 是例如, 被初始化为较小的随机值以确保 $h_{i-1}^{(<k)}$ 和 $h_i^{(k)}$ 的尺度在进行初始化时具有可比性的学习标量。在第k个后续DNN是全连接层的情况下, 可以通过包括附加线性层来代替横向连接, 与输入的数量相比较可能具有不同数量的输出。在第k个后续DNN是卷积神经网络的情况下, 可以通过在图像高度和宽度上执行 1×1 卷积并且使用密集的全连接输入和输出连接来代替横向连接。在一些情况下, 在表示被连接并且被馈送至当前DNN之前, 可以在这种层的顶部添加额外的非线性。

[0063] 在一些实施方式中, $K+1$ 个机器学习任务可以是独立的机器学习任务, 即, 没有对任务的类型或者 K 个任务之间的关系进行假设。例如, $K+1$ 个机器学习任务可以是正交的或者对抗的。下面分别参照图2和图3描述训练渐进式神经网络系统以执行 $K+1$ 个机器学习任务以及通过使用渐进式神经网络来处理例如与第 $K+1$ 个机器学习任务对应的输入。

[0064] 图2是训练渐进式神经网络系统以执行机器学习任务序列的示例过程200的流程图。为了方便起见, 将过程200描述为由位于一个或者多个位置的一个或者多个计算机的系统执行。例如, 包括深度神经网络(DNN)序列的、根据本说明书被适当编程的系统可以执行过程200, 其中, 序列中的每个DNN与相应的机器学习任务对应。例如, 图1的根据本说明书被适当编程的渐进式神经网络系统100可以执行过程200。

[0065] 对于序列中的第一机器学习任务, 系统训练序列中与第一机器学习任务对应的第一DNN以执行第一机器学习任务(步骤202)。第一DNN可以包括第一数量的索引层, 例如, L 个隐藏层, 其中, 第一数量的索引层中的每个层被配置为接收相应的层输入并且处理该层输

入以根据一个或者多个相应的第一DNN参数来生成相应的层输出。

[0066] 在一些实施方式中,第一数量的索引层中的每个层可以包括隐藏激活 $h_i^{(1)} \in \mathbb{R}^{n_i}$,其中, n_i 表示在层 $i \leq L$ 处的隐藏单元的数量。

[0067] 对于序列中的每个后续机器学习任务,系统训练与机器学习任务对应的后续DNN以执行机器学习任务(步骤204)。系统不需要执行与用于训练先前DNN(例如,第一DNN)的学习算法相同的学习算法来训练后续DNN。换句话说,系统可以向DNN序列中的每个DNN应用不同的训练算法。例如,DNN序列中的一个或者多个DNN可以针对图像分类被训练,而DNN序列中的一个或者多个其它DNN可以针对生成建模任务被训练,例如,可以是变分自编码器。

[0068] 后续DNN可以包括随后的多个索引层,其中,随后的多个索引层中具有大于1的索引的每个层从(i)后续DNN的先前层(其索引比当前索引小1的层)和(ii)相应的先前DNN的一个或者多个先前层接收输入。在一些实施方式中,随后的多个索引层中具有等于1的索引的每个层接收相应的后续DNN输入。随后的多个索引层中的每个层被配置为接收相应的层输入并且处理该层输入以根据一个或者多个相应的后续DNN参数来生成相应的层输出。

[0069] 在一些实施方式中,训练每个后续DNN可以包括:将先前DNN的先前DNN参数设置为常量值,例如,设置为先前训练的值;以及通过使用机器学习训练技术(例如,对梯度进行反向传播)来调整相应的后续DNN参数的值。在一些实施方式中,在进行训练之前将与后续机器学习任务对应的每个后续DNN随机初始化。在其它实施方式中,可以将与后续机器学习任务对应的每个后续DNN初始化,使得后续DNN最初产生与先前DNN相同的输出。对每个后续DNN的初始化影响上面的等式(1)在输出层实例化的方式。

[0070] 图3是通过使用渐进式神经网络来处理输入的示例过程300的流程图。为了方便起见,将过程300描述为由位于一个或者多个位置的一个或者多个计算机的系统执行。例如,包括深度神经网络(DNN)序列的、根据本说明书被适当编程的系统(例如,图1的渐进式神经网络100)可以执行过程300。

[0071] 系统接收作为与DNN序列中的最后的后续DNN对应的机器学习任务的一部分的输入(步骤302)。例如,输入可以是作为机器学习任务的与机器学习任务序列中的第 $K+1$ 个机器学习任务对应的一部分的输入。第 $K+1$ 个机器学习任务可以独立于机器学习任务序列中的其它机器学习任务,即,第 $K+1$ 个机器学习任务可以与机器学习任务序列中的其它机器学习任务中的一个或者多个正交或者对抗。

[0072] 系统通过使用序列中的最后的后续DNN来处理输入以生成机器学习任务的最后的后续DNN输出(步骤304)。最后的后续DNN包括若干索引层,并且(i)若干索引层中具有大于1的索引的每个层从(i)最后的后续DNN的先前层和(ii)DNN序列中每个先前DNN的一个或者多个先前层接收输入,其中,先前层是其索引比当前索引小1的层。相应的多个索引层中具有等于1的索引的每个层接收相应的后续DNN输入。

[0073] 由于系统的如上面参照图1描述的结构,系统通过使用从机器学习任务序列中的先前机器学习任务学习到的特征来处理接收到的输入。即,系统不仅通过使用序列中的最后的后续DNN,而且还通过使用最后的后续DNN的至少一个层连接至其的每个DNN来处理接收到的输入。即使与接收到的输入对应的机器学习任务与由先前DNN执行的先前机器学习任务无关,该结构也适用。

[0074] 在一些实施方式中,接收到的输入可以是分类机器学习任务的一部分,诸如,图像

处理任务、语音识别任务、自然语言处理任务、或者光学字符识别任务。在这些情况下,生成的输出可以包括对接收到的输入进行分类的输出。

[0075] 可以在数字电子电路系统中、有形体现的计算机软件或者固件中、计算机硬件中(包括本说明书所公开的结构及其结构等效物)、或者它们中的一个或者多个的组合中实施本说明书中描述的主题的实施例和功能操作。可以将本说明书中描述的主题的实施例实施为一个或者多个计算机程序,即,编码在有形非暂时性程序载体上以由数据处理设备执行或者控制数据处理设备的操作的计算机程序指令的一个或者多个模块。可替代地或者另外,程序指令可以编码在人工生成的传播信号上,例如,机器生成的电气、光学、或者电磁信号,生成该信号是为了对信息进行编码以便传输至合适的接收器设备供数据处理设备执行。计算机存储介质可以是机器可读存储装置、机器可读存储基板、随机或者串行存取存储器装置、或者它们中的一个或者多个的组合。然而,计算机存储介质不是传播信号。

[0076] 术语“数据处理设备”囊括了用于处理数据的所有种类的设备、装置、和机器,包括:例如,可编程处理器、计算机、或者多个处理器或者计算机。该设备可以包括专用逻辑电路系统,例如,FPGA(现场可编程门阵列)或者ASIC(专用集成电路)。除了硬件之外,该设备还可以包括为探讨中的计算机程序创建执行环境的代码,例如,构成处理器固件、协议栈、数据库管理系统、操作系统、或者它们中的一个或者多个的组合的代码。

[0077] 可以用任何形式的程序语言(包括:编译语言或者解释语言、或者陈述性语言或者程序语言)来编写计算机程序(也可以被称为或者描述为程序、软件、软件应用、模块、软件模块、脚本、或者代码),并且可以按照任何形式(包括:作为独立式程序或者模块、组件、子例程、或者适合用于计算环境的其它单元)来部署计算机程序。计算机程序可以但并非必须与文件系统中的文件对应。可以将程序存储在保持其它程序或者数据(例如,存储在标记语言文档中的一个或者多个脚本)的文件的一部分中,或者存储在专用于所探讨中的程序的单个文件中,或者存储在多个协作文件(例如,存储一个或者多个模块、子程序、或者部分代码的文件)中。可以将计算机程序部署为在一个计算机上执行或者在位于一个站点处或者分布在多个站点中并且通过通信网络互相连接的多个计算机上执行。

[0078] 如在本说明书中使用的,“引擎”或者“软件引擎”是指提供与输入不同的输出的软件实现的输入/输出系统。引擎可以是编码的功能块,诸如,库、平台、软件开发工具包(“SDK”)、或者对象。可以将每个引擎实施在包括一个或者多个处理器和计算机可读介质的任何适当类型的计算装置上,例如,服务器、移动电话、平板计算机、笔记本计算机、音乐播放器、电子书阅读器、膝上型或者台式计算机、PDA、智能电话、或者其它固定的或者便携式装置。此外,可以在同一计算装置上或者不同的计算装置上实施引擎中的两个或者更多个。

[0079] 可以通过一个或者多个可编程计算机来执行本说明书中描述的过程和逻辑流,该一个或者多个可编程计算机执行一个或者多个计算机程序以通过操作输入数据并且生成输出来执行功能。还可以通过专用逻辑电路系统(例如,FPGA(现场可编程门阵列))或者ASIC(专用集成电路)来执行处理和逻辑流,并且还可以将是被实施为逻辑电路系统(例如,FPGA(现场可编程门阵列))或者ASIC(专用集成电路)。

[0080] 适合于执行计算机程序的计算机包括,例如,可以基于通用或者专用微处理器或者两者或者任何其它种类的中央处理单元。一般而言,中央处理单元将接收来自只读存储器或者随机存取存储器或者两者的指令和数据。计算机的必要元件是:用于进行或者执行

指令的中央处理单元、和用于存储指令和数据的一个或者多个存储器装置。一般而言，计算机还将包括用于存储数据的一个或者多个海量存储装置（例如，磁盘、磁光盘、或者光盘），或者计算机可以操作地耦合以接收来自该海量存储装置的数据或者将数据传输至该海量存储装置或者进行两者。然而，计算机不需要具有这种装置。此外，计算机可以嵌入在另一装置中，例如，移动电话、个人数字助理（PDA）、移动音频或者视频播放器、游戏机、全球定位系统（GPS）接收器、或者便携式存储装置（例如，通用串行总线（USB）闪存驱动），仅举数例。

[0081] 适合于存储计算机程序指令和数据的计算机可读介质包括所有形式的非易失性存储器、介质和存储器装置，包括：例如，半导体存储器装置（例如，EPROM、EEPROM、和闪速存储器装置）、磁盘（例如，内部硬盘或者可移动盘）、磁光盘、CD-ROM盘和DVD-ROM盘。处理器和存储器可以由专用逻辑电路系统补充或者可以并入该专用逻辑电路系统中。

[0082] 为了提供与用户的交互，可以在计算机上实施本说明书中描述的主题的实施例，该计算机具有：用于向用户显示信息的显示装置，例如，CRT（阴极射线管）或者LCD（液晶显示器）监视器；以及键盘和指向装置，例如，鼠标或者轨迹球，用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互；例如，提供给用户的反馈可以是任何形式的传感反馈，例如，视觉反馈、听觉反馈或者触觉反馈；并且可以用任何形式（包括声输入、语音输入或者触觉输入）来接收来自用户的输入。另外，计算机可以通过将文档发送至用于所使用的装置并且接收来自该装置的文档（例如，通过响应于从网络浏览器接收的请求来将网页发送至在用户的用户装置上的网络浏览器）来与用户进行交互。

[0083] 可以将本说明书中描述的主题的实施例实施在包括后台组件的计算系统（例如，作为数据服务器）、或者包括中间件组件的计算系统（例如，应用服务器）、或者包括前端组件的计算系统（例如，具有图形用户界面或者网络浏览器的用户计算机，用户可以通过该图形用户界面或者该网络浏览器来与本发明中所描述的主题的实施方式交互）、或者包括一个或者多个这种后台组件、中间件组件或者前端组件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信（例如，通信网络）来将系统的组件相互连接。通信网络的示例包括：局域网（“LAN”）和广域网（“WAN”），例如，互联网。

[0084] 计算系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。

[0085] 虽然本说明书包含了许多具体实施细节，但是不应该将这些细节视为对任何发明或者可能被要求的内容的范围的限制，而是作为可以针对特定发明的特定实施例的特征的描述。在本说明书中在单独实施例的背景下描述的某些特征还可以组合地实施在单个实施例中。相反，在单个实施例的背景下描述的各种特征也可以单独地或者按照任何合适的子组合实施在多个实施例中。此外，虽然上文可能将特征描述为以某些组合来起作用并且最初甚至同样地对该特征进行了要求，但是在一些情况下，可以从组合中删除来自所要求的组合的一个或者多个特征。并且所要求的组合可以指向子组合或者子组合的变化。

[0086] 同样，虽然在附图中按照特定顺序示出了操作，但是不应该将其理解为需要按照所述的特定顺序或者按照相继的顺序来进行这种操作，或者需要进行所有图示的操作以实现期望的结果。在某些情况下，多任务处理和并行处理可以是有利的。此外，不应该将在上

述实施例中的各种系统模块和组件的分离理解为在所有实施例中需要这种分离，并且应该理解，所描述的程序组件和系统通常可以一起集成在单个软件产品中或者封装到多个软件产品中。

[0087] 已经描述了本主题的特定实施例。其它实施例在以下权利要求书的范围内。例如，可以按照不同的顺序来执行权利要求书中阐述的动作，并且仍然实现期望的结果。作为一个示例，在附图中描绘的过程不一定要求示出的特定顺序或者相继顺序，以实现期望的结果。在某些实施方式中，多任务处理和并行处理可以是有利的。

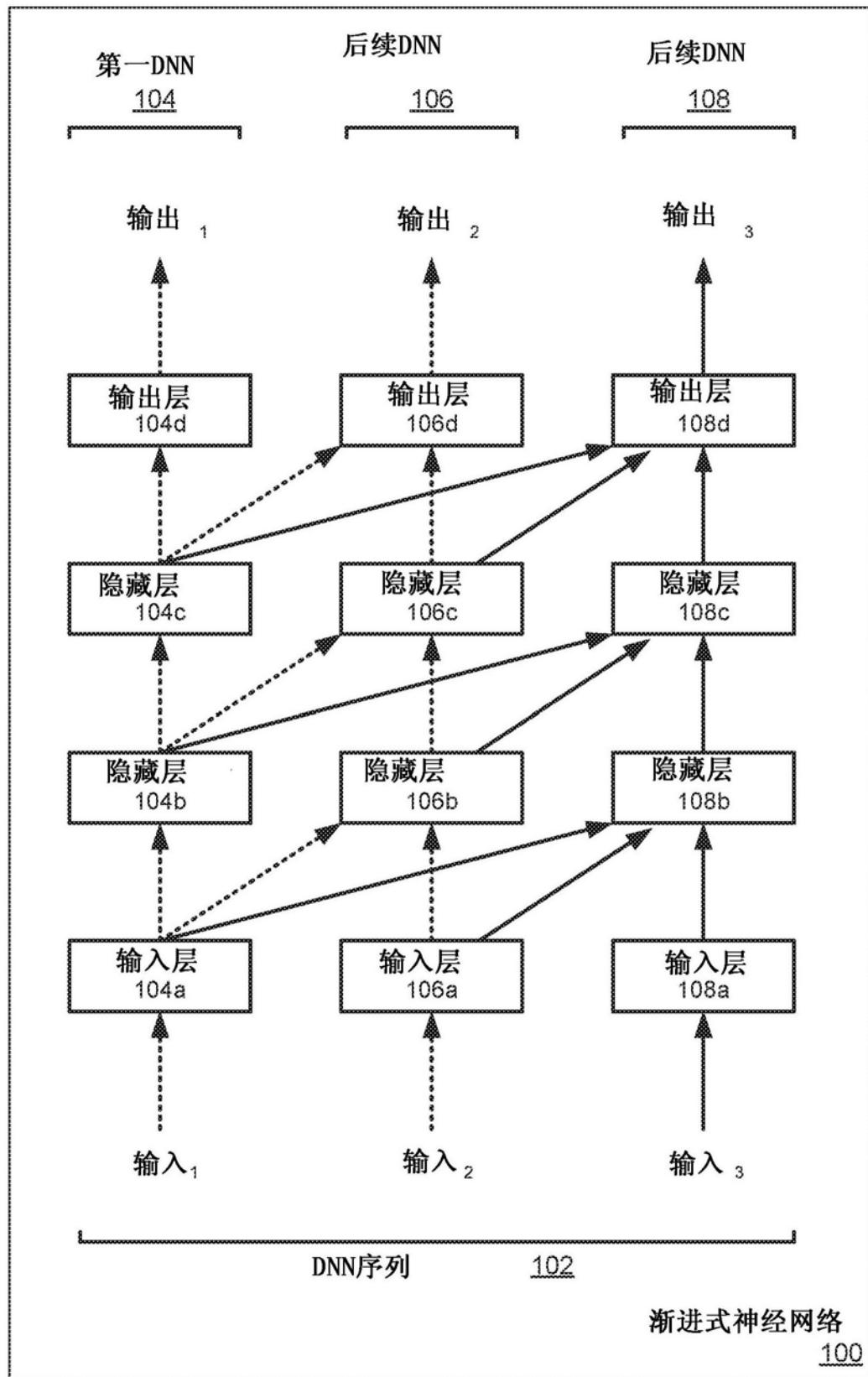


图1

200 ↘

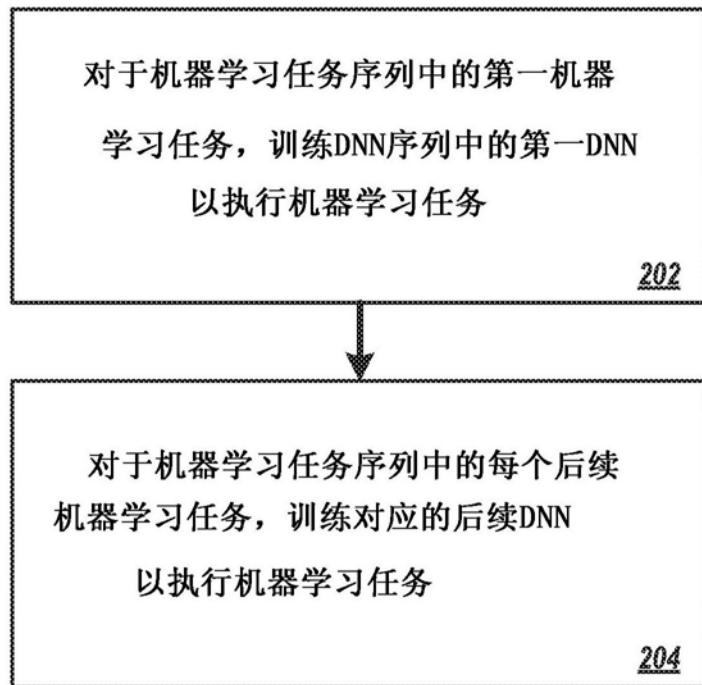


图2

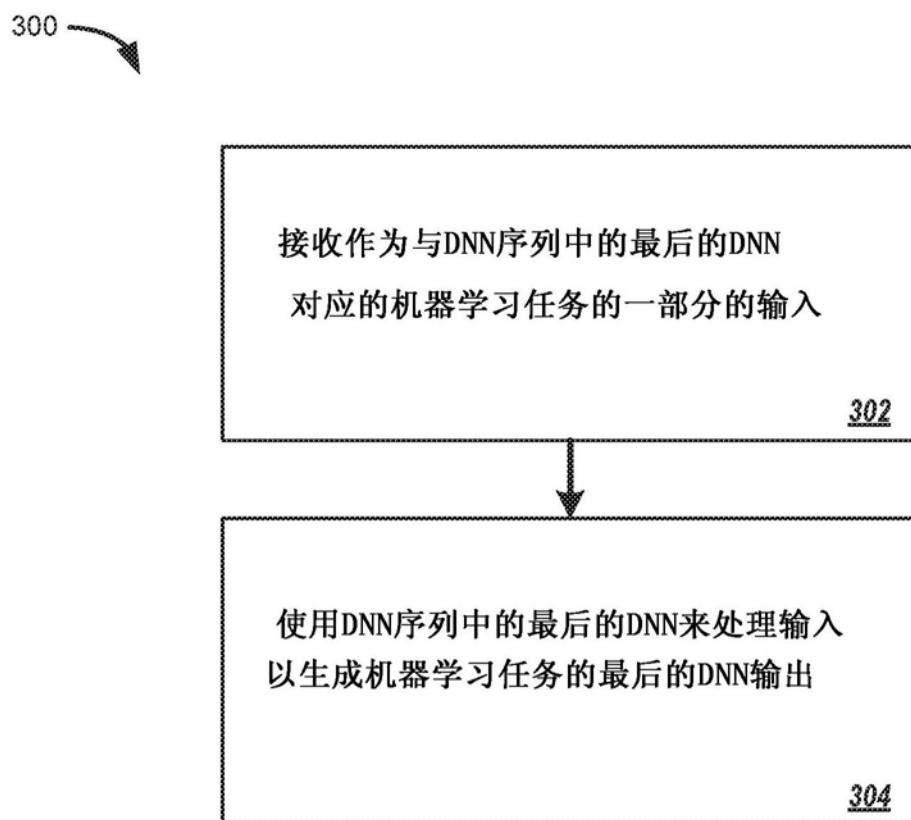


图3