

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2019年6月20日 (20.06.2019)



(10) 国际公布号
WO 2019/114618 A1

- (51) 国际专利分类号:
G06N 3/08 (2006.01)
- (21) 国际申请号: PCT/CN2018/119725
- (22) 国际申请日: 2018年12月7日 (07.12.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201711319390.9 2017年12月12日 (12.12.2017) CN
- (71) 申请人: 杭州海康威视数字技术股份有限公司 (HANGZHOU HIKVISION DIGITAL TECHNOLOGY CO., LTD.) [CN/CN]; 中国浙江省杭州市滨江区阡陌路555号, Zhejiang 310051 (CN)。
- (72) 发明人: 谢迪 (XIE, Di); 中国浙江省杭州市滨江区阡陌路555号, Zhejiang 310051 (CN)。 浦世亮 (PU, Shiliang); 中国浙江省杭州市滨江区阡陌路555号, Zhejiang 310051 (CN)。
- (74) 代理人: 北京柏杉松知识产权代理事务所 (普通合伙) (PATENTSINO IP FIRM); 中国北京市西城区北三环中路27号商厦大厦413室, Beijing 100029 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,

(54) Title: DEEP NEURAL NETWORK TRAINING METHOD AND APPARATUS, AND COMPUTER DEVICE

(54) 发明名称: 一种深度神经网络训练方法、装置及计算机设备

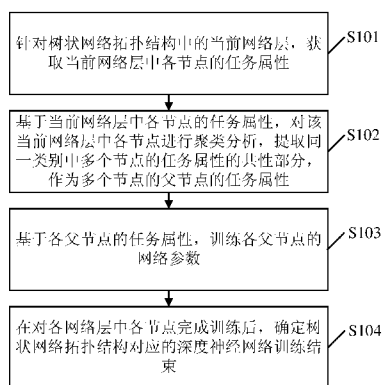


图1

- S101 FOR A CURRENT NETWORK LAYER IN A TREE NETWORK TOPOLOGY STRUCTURE, ACQUIRE A TASK ATTRIBUTE OF EACH NODE IN THE CURRENT NETWORK LAYER
- S102 BASED ON THE TASK ATTRIBUTE OF EACH NODE IN THE CURRENT NETWORK LAYER, PERFORM CLUSTERING ANALYSIS ON EACH NODE IN THE CURRENT NETWORK LAYER TO EXTRACT A COMMON PART OF THE TASK ATTRIBUTES OF MULTIPLE NODES IN THE SAME CATEGORY AND TAKE SAME AS A TASK ATTRIBUTE OF A PARENT NODE OF THE MULTIPLE NODES
- S103 BASED ON THE TASK ATTRIBUTE OF EACH PARENT NODE, TRAIN A NETWORK PARAMETER OF EACH PARENT NODE
- S104 AFTER THE TRAINING OF EACH NODE IN EACH NETWORK LAYER IS COMPLETED, DETERMINE THAT THE TRAINING OF A DEEP NEURAL NETWORK CORRESPONDING TO THE TREE NETWORK TOPOLOGY STRUCTURE HAS ENDED

(57) Abstract: A deep neural network training method and apparatus, and a computer device. The deep neural network training method comprises: for a current network layer in a tree network topology structure, acquiring a task attribute of each node in the current network layer (S101); based on the task attribute of each node in the current network layer, performing clustering analysis on each node in the current network layer to extract a common part of the task attributes of multiple nodes in the same category and take same as a task attribute of a parent node of the multiple nodes (S102); based on the task attribute of each parent node, training a network parameter of each parent node (S103); and after the training of each node in each network layer is completed, determining that the training of a deep neural network corresponding to the tree network topology structure has ended (S104). By means of the method, the computational efficiency of deep learning can be improved.

(57) 摘要: 一种深度神经网络训练方法、装置及计算机设备, 其中, 深度神经网络训练方法包括: 针对树状网络拓扑结构中的当前网络层, 获取当前网络层中各节点的任务属性(S101); 基于当前网络层中各节点的任务属性, 对当前网络层中各节点进行聚类分析, 提取同一类别中多个节点的任务属性的共性部分, 作为多个节点的父节点的任务属性(S102); 基于各父节点的任务属性, 训练各父节点的网络参数(S103); 在对各网络层中各节点完成训练后, 确定树状网络拓扑结构对应的深度神经网络训练结束(S104)。通过本方法可以提高深度学习的运算效率。

PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

- (84)** 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

一种深度神经网络训练方法、装置及计算机设备

本申请要求于 2017 年 12 月 12 日提交中国专利局、申请号为 201711319390.9 发明名称为“一种深度神经网络训练方法、装置及计算机设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及机器学习技术领域，特别是涉及一种深度神经网络训练方法、装置及计算机设备。

背景技术

10 深度学习是人工智能神经网络基础上发展而来的一种机器学习方法，深度神经网络作为深度学习的主要模型，通过模仿人脑的机制来解释数据，是一种通过建立和模拟人脑进行分析学习的智能模型，其在语音识别、图像分类、人脸识别、自然语言处理、广告投放等应用领域已被广泛应用。

15 目前，大多数深度学习只针对单个任务，例如，对目标的属性进行检测、对目标的状态进行估计等。针对于复杂的场景，往往需要实现多个任务，通常使用的方法是，利用多个神经网络分别针对各任务进行运算，然后再将运算结果进行合并，这个过程非常消耗时间，并且由于每一个神经网络中存在高度的冗余性，导致深度学习的运算效率较低。

发明内容

20 本申请实施例的目的在于提供一种深度神经网络训练方法、装置及计算机设备，以提高深度学习的运算效率。具体技术方案如下：

第一方面，本申请实施例提供了一种深度神经网络训练方法，所述方法包括：

25 针对树状网络拓扑结构中的当前网络层，获取所述当前网络层中各节点的任务属性，其中，所述树状网络拓扑结构中的各节点为对应于不同任务的神经网络，所述树状网络拓扑结构中的叶节点为针对指定任务已完成训练的神经网络；

基于所述当前网络层中各节点的任务属性，对所述当前网络层中各节点

进行聚类分析，提取同一类别中多个节点的任务属性的共性部分，作为所述多个节点的父节点的任务属性；

基于各父节点的任务属性，训练各父节点的网络参数；

在对各网络层中各节点完成训练后，确定所述树状网络拓扑结构对应的
5 深度神经网络训练结束。

可选的，所述基于所述当前网络层中各节点的任务属性，对所述当前网络层中各节点进行聚类分析，提取同一类别中多个节点的任务属性的共性部分，作为所述多个节点的父节点的任务属性，包括：

10 根据所述当前网络层中各节点的任务属性，通过预设相似性度量算法，生成对应于各节点的任务属性的相似性度量矩阵；

根据所述相似性度量矩阵，将相似性大于预设阈值的多个节点确定为同一类别；

提取同一类别中多个节点的任务属性的共性部分，作为所述多个节点的父节点的任务属性。

15 可选的，所述基于各父节点的任务属性，训练各父节点的网络参数，包括：

针对任一父节点，基于该父节点的任务属性，将该父节点的各子节点的输出特征作为该父节点的输入，训练该父节点的网络参数。

20 可选的，所述基于各父节点的任务属性，训练各父节点的网络参数，包括：

针对任一父节点，基于该父节点的任务属性，利用具有特征信号控制机制的预设结构生成该父节点；获取并根据该父节点各子节点的输出特征中与所述任务属性相关的各输出特征的加权组合，训练该父节点的网络参数。

25 可选的，所述在对各网络层中各节点完成训练后，确定所述树状网络拓扑结构对应的深度神经网络训练结束，包括：

从所述叶节点所处的网络层开始，按照从底层至顶层的训练顺序，依次

训练各网络层中节点的网络参数，完成各节点的训练；

在顶层的各节点训练完毕后，确定所述树状网络拓扑结构对应的深度神经网络训练结束。

5 第二方面，本申请实施例提供了一种深度神经网络训练装置，所述装置包括：

获取模块，用于针对树状网络拓扑结构中的当前网络层，获取所述当前网络层中各节点的任务属性，其中，所述树状网络拓扑结构中的各节点为对应于不同任务的神经网络，所述树状网络拓扑结构中的叶节点为针对指定任务已完成训练的神经网络；

10 聚类模块，用于基于所述当前网络层中各节点的任务属性，对所述当前网络层中各节点进行聚类分析，提取同一类别中多个节点的任务属性的共性部分，作为所述多个节点的父节点的任务属性；

训练模块，用于基于各父节点的任务属性，训练各父节点的网络参数；

15 确定模块，用于在对各网络层中各节点完成训练后，确定所述树状网络拓扑结构对应的深度神经网络训练结束。

可选的，所述聚类模块，具体用于：

根据所述当前网络层中各节点的任务属性，通过预设相似性度量算法，生成对应于各节点的任务属性的相似性度量矩阵；

20 根据所述相似性度量矩阵，将相似性大于预设阈值的多个节点确定为同一类别；

提取同一类别中多个节点的任务属性的共性部分，作为所述多个节点的父节点的任务属性。

可选的，所述训练模块，具体用于：

25 针对任一父节点，基于该父节点的任务属性，将该父节点的各子节点的输出特征作为该父节点的输入，训练该父节点的网络参数。

可选的，所述训练模块，具体用于：

-4-

针对任一父节点，基于该父节点的任务属性，利用具有特征信号控制机制的预设结构生成该父节点；获取并根据该父节点的各子节点的输出特征中与所述任务属性相关的各输出特征的加权组合，训练该父节点的网络参数。

可选的，所述确定模块，具体用于：

- 5 从所述叶节点所处的网络层开始，按照从底层至顶层的训练顺序，依次训练各网络层中节点的网络参数，完成各节点的训练；

在顶层的各节点训练完毕后，确定所述树状网络拓扑结构对应的深度神经网络。

- 10 第三方面，本申请实施例提供了一种计算机可读存储介质，用于存储可执行代码，所述可执行代码用于在运行时执行：本申请实施例第一方面所提供的深度神经网络训练方法。

第四方面，本申请实施例提供了一种应用程序，用于在运行时执行：本申请实施例第一方面所提供的深度神经网络训练方法。

- 15 第五方面，本申请实施例提供了一种计算机设备，包括处理器和计算机可读存储介质，其中，

所述计算机可读存储介质，用于存放可执行代码；

所述处理器，用于执行所述计算机可读存储介质上所存放的可执行代码时执行：本申请实施例第一方面所提供的深度神经网络训练方法。

- 20 综上所述，本申请实施例提供的方案中，通过构建树状网络拓扑结构，获取该树状网络拓扑结构中已完成训练的当前网络层中各节点的任务属性，基于任务属性，对当前网络层中各节点进行聚类分析，并将同一类别中多个节点的任务属性的共性部分作为父节点的任务属性，这样就可以根据各父节点的任务属性训练各父节点的网络参数，在对各网络层中各节点完成训练后，可以确定树状网络拓扑结构对应的深度神经网络训练完毕，训练后的深度神经网络可以实现多个任务。由于树状网络拓扑结构中的叶节点为针对指定任务已训练好的神经网络，通过对网络层中各节点的任务属性的共性部分进行
25 提取，可以对指定任务对应的神经网络进行复用，通过一层一层的树状网络

拓扑结构的运算，即可以利用一个完整的深度神经网络实现多个指定任务，并且基于任务属性对节点进行聚类，构建属于同一类别的节点的父节点，该父节点可以实现子节点的共性任务，因此可以有效减小神经网络间的冗余，进而提高深度学习的运算效率。

5 附图说明

为了更清楚地说明本申请实施例和现有技术的技术方案，下面对实施例和现有技术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申请的一些实施例，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，还可以根据这些附图获得其他的附图。

- 10 图 1 为本申请实施例的深度神经网络训练方法的流程示意图；
图 2 为本申请实施例的深度神经网络训练装置的结构示意图；
图 3 为本申请实施例的计算机设备的结构示意图。

具体实施方式

- 15 为使本申请的目的、技术方案、及优点更加清楚明白，以下参照附图并举实施例，对本申请进一步详细说明。显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

下面通过具体实施例，对本申请进行详细的说明。

- 20 为了提高深度学习的运算效率，本申请实施例提供了一种深度神经网络训练方法、装置及计算机设备。

下面，首先对本申请实施例所提供的深度神经网络训练方法进行介绍。

- 25 本申请实施例所提供的一种深度神经网络训练方法的执行主体可以为实现多个指定任务的计算机设备，执行主体中至少包括具有数据处理能力的核心处理芯片。实现本申请实施例所提供的一种深度神经网络训练方法的方式可以为设置于执行主体中的软件、硬件电路和逻辑电路中的至少一种方式。

如图1所示，为本申请实施例所提供的一种深度神经网络训练方法，该深

度神经网络训练方法可以包括如下步骤：

S101，针对树状网络拓扑结构中的当前网络层，获取当前网络层中各节点的任务属性。

5 针对多个指定任务，利用预设训练算法，例如反向传播算法、正向传播算法、梯度下降训练算法等，可以得到针对各指定任务的神经网络，每个神经网络可以实现一个具体的任务，例如目标属性检测、目标状态估计等，这些任务之间往往存在一定的相似性，因此实现这些任务的各神经网络中也存在冗余，也就是，已训练好的神经网络之间存在一些相同的网络模型用于实现相同的任务。基于上述考虑，可以基于用于执行不同指定任务的神经网络，
10 设计一个完整的深度神经网络，实现各指定任务。该深度神经网络可以为树状网络拓扑结构，其中每个节点可以为对应于不同任务的神经网络。该树状网络拓扑结构中一共存在三种类型的节点：叶节点、根节点及中间节点。由于最终目的是实现各指定任务，因此，叶节点为针对指定任务已完成训练的神经网络，可以将叶节点视为编码器。中间节点和根节点为基于下一层的节点，
15 向上一层一层训练出来的神经网络。

S102，基于当前网络层中各节点的任务属性，对该当前网络层中各节点进行聚类分析，提取同一类别中多个节点的任务属性的共性部分，作为多个节点的父节点的任务属性。

20 由于树状网络拓扑结构中的一个网络层中，各节点所对应的任务之间具有一定的相似性，例如，识别目标的性别和识别目标的年龄时，由于这两个指定任务都需要首先检测出人体或者人脸区域，以此为基础进行性别分类与年龄估计，则可以将这两个指定任务对应的两个节点划分为同一类别，并且提取这两个指定任务的共性部分，即对人体或者人脸区域的检测，则可以将对人体或者人脸区域的检测作为上述两个节点的父节点的任务属性。这样，
25 在通过深度神经网络进行运算时，可以先执行对人体或者人脸区域进行检测的任务，在检测出人体或者人脸区域后，再进行目标的性别识别和目标的年龄估计的任务。

聚类分析是将数据分类到不同的类别的过程，因此，同一类别中的对象有很大的相似性。聚类分析的方式有多种，例如系统聚类法、分解法、加入

法、动态聚类法等，这里不做限定。在本实施例中，聚类分析就是将任务属性相近的节点划分为同一类别，因此可以先根据任务属性的相似性进行任务的聚类，将同一类别的任务属性对应的节点设置为拥有公共的父节点。为了提高聚类分析的效率，可以利用相似性度量进行聚类，因此，可选的，对网络层中各节点进行聚类得到父节点的方式可以包括如下步骤：

第一步，根据当前网络层中各节点的任务属性，通过预设相似性度量算法，生成对应于各节点的任务属性的相似性度量矩阵。

第二步，根据相似性度量矩阵，将相似性大于预设阈值的多个节点确定为同一类别。

10 第三步，提取同一类别中多个节点的任务属性的共性部分，作为多个节点的父节点的任务属性。

预设相似性度量算法可以为欧氏距离、曼哈顿距离、切比雪夫距离、闵可夫斯基距离、标准化欧氏距离、马氏距离、夹角余弦、汉明距离等，这里不做具体的限定，通过这些算法可以生成任务属性与任务属性之间的相似性度量矩阵，根据该相似性度量矩阵，如果多个任务属性的相似性大于预设阈值，则说明多个任务属性较为相近，则可以将实现这些任务属性的节点确定为同一类别，并且这些任务属性的共性部分可以只运算一次，在基于对任务属性的共性部分的执行后，再执行各任务属性的特性部分，这样就可以有效减少各神经网络之间的冗余，提高运行效率。

20 S103，基于各父节点的任务属性，训练各父节点的网络参数。

在确定各父节点的任务属性后，即可以确定各父节点需要实现的任务，可以使用实现该任务的传统的神经网络模型，但是各父节点的网络参数需要进行训练才可以确定，也就是使得各父节点的输出可以满足作为其子节点的神经网络的任务的要求，因此，可以通过如下步骤训练各父节点的网络参数：

25 针对任一父节点，基于该父节点的任务属性，将该父节点各子节点的输出特征作为该父节点的输入，训练该父节点的网络参数。

网络参数的训练过程可以是通过将各父节点的子节点的输出特征作为该父节点的输入，然后通过对网络参数进行不断的调整使得各子节点的输出特

征满足指定任务。针对树状网络拓扑结构的叶节点，由于神经网络的倒数第一层往往为瞬时函数，而倒数第二层为特征层，即倒数第二层的输出为编码的特征，因此将倒数第二层的输出作为上一层级的节点的输入。

5 由于不同特征之间会有相互影响，如果输入父节点的特征权值相同，在经过父节点和子节点的运算后，得到的结果可能与原始的任务结果有较大的差别，因此，父节点可以由任意具有信号控制机制的结构构成，例如LSTM（Long Short-Term Memory，长短期记忆）网络机制、注意力Attention机制、GRU（Gated Recurrent Unit，门控循环单元）网络机制等，通过对父节点的输入进行加权组合，训练父节点的网络参数，并在训练的过程中调整权值。即，
10 还可以通过如下步骤训练各父节点的网络参数：

针对任一父节点，基于该父节点的任务属性，利用具有特征信号控制机制的预设结构生成该父节点；获取并根据该父节点的各子节点的输出特征中与所述任务属性相关的各输出特征的加权组合，训练该父节点的网络参数。

15 父节点可以由任意具有信号控制机制的结构构成的本质就是针对特定的任务属性，自动选择相关的特征信号、消除与任务属性不相关的特征信号。并且通过加权组合，可以给予节点中的有效部分分配较大的权值，给无效部分分配较小的权值，这样，就可以在任务结果中突显期望达到的任务效果，同时屏蔽无效部分，从而可以提高指定任务上的性能。

20 S104，在对各网络层中各节点完成训练后，确定树状网络拓扑结构对应的深度神经网络训练结束。

基于上述过程可以得到树状网络拓扑结构中一个网络层的各节点，通过对各网络层中各节点进行训练，训练结束后即可确定树状网络拓扑结构对应的深度神经网络。

可选的，确定深度神经网络训练结束的步骤，具体可以包括：

25 从叶节点所处的网络层开始，按照从底层至顶层的训练顺序，依次训练各网络层中节点的网络参数，完成各节点的训练；

在顶层的各节点训练完毕后，确定树状网络拓扑结构对应的深度神经网络训练结束。

5 执行主体可以周期性的检测是否有新的神经网络，如果新添加了一个指定任务对应的神经网络，则可以按照上述过程，将该神经网络作为树状网络拓扑结构的叶节点，一步步自底至顶训练，直至根节点。为了保证更新的一致性，同一网络层中每一个节点的训练过程是相互独立的，不会受到其他节点的干扰。

10 应用本实施例，通过构建树状网络拓扑结构，获取该树状网络拓扑结构中已完成训练的当前网络层中各节点的任务属性，基于任务属性，对当前网络层中各节点进行聚类分析，并将同一类别中多个节点的任务属性的共性部分作为父节点的任务属性，这样就可以根据各父节点的任务属性训练各父节点的网络参数，在对各网络层中各节点完成训练后，可以确定树状网络拓扑结构对应的深度神经网络训练完毕，训练后的深度神经网络可以实现多个任务。由于树状网络拓扑结构中的叶节点为针对指定任务已训练好的神经网络，通过对网络层中各节点的任务属性的共性部分进行提取，可以对指定任务对应的神经网络进行复用，通过一层一层的树状网络拓扑结构的运算，即可以
15 利用一个完整的深度神经网络实现多个指定任务，并且通过基于任务属对节点进行聚类，构建属于同一类别的节点的父节点，该父节点可以实现子节点的共性任务，因此可以有效减小神经网络间的冗余，进而提高深度学习的运算效率。

20 相应于上述方法实施例，本申请实施例提供了一种深度神经网络训练装置，如图2所示，该深度神经网络训练装置可以包括：

获取模块210，用于针对树状网络拓扑结构中的当前网络层，获取所述当前网络层中各节点的任务属性，其中，所述树状网络拓扑结构中的各节点为对应于不同任务的神经网络，所述树状网络拓扑结构中的叶节点为针对指定
25 任务已完成训练的神经网络；

聚类模块220，用于基于所述当前网络层中各节点的任务属性，对所述当前网络层中各节点进行聚类分析，提取同一类别中多个节点的任务属性的共性部分，作为所述多个节点的父节点的任务属性；

训练模块230, 用于基于各父节点的任务属性, 训练各父节点的网络参数;

确定模块240, 用于在对各网络层中各节点完成训练后, 确定所述树状网络拓扑结构对应的深度神经网络训练结束。

可选的, 所述聚类模块220, 具体可以用于:

- 5 根据所述当前网络层中各节点的任务属性, 通过预设相似性度量算法, 生成对应于各节点的任务属性的相似性度量矩阵;

根据所述相似性度量矩阵, 将相似性大于预设阈值的多个节点确定为同一类别;

- 10 提取同一类别中多个节点的任务属性的共性部分, 作为所述多个节点的父节点的任务属性。

可选的, 所述训练模块230, 具体可以用于:

针对任一父节点, 基于该父节点的任务属性, 将该父节点的各子节点的输出特征作为该父节点的输入, 训练该父节点的网络参数。

可选的, 所述训练模块230, 具体可以用于:

- 15 针对任一父节点, 基于该父节点的任务属性, 利用具有特征信号控制机制的预设结构生成该父节点; 获取并根据该父节点各子节点的输出特征中与所述任务属性相关的各输出特征的加权组合, 训练该父节点的网络参数。

可选的, 所述确定模块240, 具体可以用于:

- 20 从所述叶节点所处的网络层开始, 按照从底层至顶层的训练顺序, 依次训练各网络层中节点的网络参数, 完成各节点的训练;

在顶层的各节点训练完毕后, 确定所述树状网络拓扑结构对应的深度神经网络训练结束。

- 25 应用本实施例, 通过构建树状网络拓扑结构, 获取该树状网络拓扑结构中已完成训练的当前网络层中各节点的任务属性, 基于任务属性, 对当前网络层中各节点进行聚类分析, 并将同一类别中多个节点的任务属性的共性部分作为父节点的任务属性, 这样就可以根据各父节点的任务属性训练各父节

点的网络参数，在对各网络层中各节点完成训练后，可以确定树状网络拓扑结构对应的深度神经网络训练完毕，训练后的深度神经网络可以实现多个任务。由于树状网络拓扑结构中的叶节点为针对指定任务已训练好的神经网络，通过对网络层中各节点的任务属性的共性部分进行提取，可以对指定任务对应的神经网络进行复用，通过一层一层的树状网络拓扑结构的运算，即可以利用一个完整的深度神经网络实现多个指定任务，并且通过基于任务属对节点进行聚类，构建属于同一类别的节点的父节点，该父节点可以实现子节点的共性任务，因此可以有效减小神经网络间的冗余，进而提高深度学习的运算效率。

10

另外，相应于上述实施例所提供的深度神经网络训练方法，本申请实施例提供了一种计算机可读存储介质，用于存储可执行代码，所述可执行代码用于在运行时执行：上述实施例所提供的深度神经网络训练方法。

本实施例中，计算机可读存储介质存储有在运行时执行本申请实施例所提供的深度神经网络训练方法的可执行代码，因此能够实现：通过构建树状网络拓扑结构，获取该树状网络拓扑结构中已完成训练的当前网络层中各节点的任务属性，基于任务属性，对当前网络层中各节点进行聚类分析，并将同一类别中多个节点的任务属性的共性部分作为父节点的任务属性，这样就可以根据各父节点的任务属性训练各父节点的网络参数，在对各网络层中各节点完成训练后，可以确定树状网络拓扑结构对应的深度神经网络训练完毕，训练后的深度神经网络可以实现多个任务。由于树状网络拓扑结构中的叶节点为针对指定任务已训练好的神经网络，通过对网络层中各节点的任务属性的共性部分进行提取，可以对指定任务对应的神经网络进行复用，通过一层一层的树状网络拓扑结构的运算，即可以利用一个完整的深度神经网络实现多个指定任务，并且通过基于任务属对节点进行聚类，构建属于同一类别的节点的父节点，该父节点可以实现子节点的共性任务，因此可以有效减小神经网络间的冗余，进而提高深度学习的运算效率。

25

另外，相应于上述实施例所提供的深度神经网络训练方法，本申请实施

例提供了一种应用程序，用于在运行时执行：上述实施例所提供的深度神经网络训练方法。

本实施例中，应用程序在运行时执行本申请实施例所提供的深度神经网络训练方法，因此能够实现：通过构建树状网络拓扑结构，获取该树状网络拓扑结构中已完成训练的当前网络层中各节点的任务属性，基于任务属性，
5 对当前网络层中各节点进行聚类分析，并将同一类别中多个节点的任务属性的共性部分作为父节点的任务属性，这样就可以根据各父节点的任务属性训练各父节点的网络参数，在对各网络层中各节点完成训练后，可以确定树状网络拓扑结构对应的深度神经网络训练完毕，训练后的深度神经网络可以实现多个任务。由于树状网络拓扑结构中的叶节点为针对指定任务已训练好的神经网络，通过对网络层中各节点的任务属性的共性部分进行提取，可以对指定任务对应的神经网络进行复用，通过一层一层的树状网络拓扑结构的运算，即可以利用一个完整的深度神经网络实现多个指定任务，并且通过基于任务属对节点进行聚类，构建属于同一类别的节点的父节点，该父节点可以
10 实现子节点的共性任务，因此可以有效减小神经网络间的冗余，进而提高深度学习的运算效率。

本申请实施例还提供了一种计算机设备，如图3所示，包括处理器301和计算机可读存储介质302，其中，

20 计算机可读存储介质302，用于存放可执行代码；

处理器301，用于执行计算机可读存储介质302上所存放的可执行代码时执行：上述实施例所提供的深度神经网络训练方法。

计算机可读存储介质302与处理器301之间可以通过有线连接或者无线连接的方式进行数据传输，并且计算机设备可以通过有线通信接口或者无线通信接口与其他的设备进行通信。
25

上述计算机可读存储介质可以包括RAM（Random Access Memory，随机存取存储器），也可以包括NVM（Non-Volatile Memory，非易失性存储器），例如至少一个磁盘存储器。可选的，计算机可读存储介质还可以是至少一个

位于远离上述处理器的存储装置。

上述处理器可以是通用处理器，包括CPU（Central Processing Unit，中央处理器）、NP（Network Processor，网络处理器）等；还可以是DSP（Digital Signal Processing，数字信号处理器）、ASIC（Application Specific Integrated Circuit，专用集成电路）、FPGA（Field-Programmable Gate Array，现场可编程门阵列）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。

本实施例中，该计算机设备的处理器通过读取计算机可读存储介质中存储的可执行代码，并通过运行该可执行代码，能够实现：通过构建树状网络拓扑结构，获取该树状网络拓扑结构中已完成训练的当前网络层中各节点的任务属性，基于任务属性，对当前网络层中各节点进行聚类分析，并将同一类别中多个节点的任务属性的共性部分作为父节点的任务属性，这样就可以根据各父节点的任务属性训练各父节点的网络参数，在对各网络层中各节点完成训练后，可以确定树状网络拓扑结构对应的深度神经网络训练完毕，训练后的深度神经网络可以实现多个任务。由于树状网络拓扑结构中的叶节点为针对指定任务已训练好的神经网络，通过对网络层中各节点的任务属性的共性部分进行提取，可以对指定任务对应的神经网络进行复用，通过一层一层的树状网络拓扑结构的运算，即可以利用一个完整的深度神经网络实现多个指定任务，并且通过基于任务属对节点进行聚类，构建属于同一类别的节点的父节点，该父节点可以实现子节点的共性任务，因此可以有效减小神经网络间的冗余，进而提高深度学习的运算效率。

对于计算机可读存储介质、应用程序以及计算机设备实施例而言，由于其所涉及的方法内容基本相似于前述的方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要

素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

本说明书中的各个实施例均采用相关的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于装置、计算机可读存储介质、应用程序以及计算机设备实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

以上所述仅为本申请的较佳实施例而已，并不用以限制本申请，凡在本申请的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本申请保护的范围之内。

权 利 要 求

1、一种深度神经网络训练方法，其特征在于，所述方法包括：

针对树状网络拓扑结构中的当前网络层，获取所述当前网络层中各节点的任务属性，其中，所述树状网络拓扑结构中的各节点为对应于不同任务的神经网络，所述树状网络拓扑结构中的叶节点为针对指定任务已完成训练的神经网络；

基于所述当前网络层中各节点的任务属性，对所述当前网络层中各节点进行聚类分析，提取同一类别中多个节点的任务属性的共性部分，作为所述多个节点的父节点的任务属性；

10 基于各父节点的任务属性，训练各父节点的网络参数；

在对各网络层中各节点完成训练后，确定所述树状网络拓扑结构对应的深度神经网络训练结束。

2、根据权利要求1所述的方法，其特征在于，所述基于所述当前网络层中各节点的任务属性，对所述当前网络层中各节点进行聚类分析，提取同一类别中多个节点的任务属性的共性部分，作为所述多个节点的父节点的任务属性，包括：

根据所述当前网络层中各节点的任务属性，通过预设相似性度量算法，生成对应于各节点的任务属性的相似性度量矩阵；

20 根据所述相似性度量矩阵，将相似性大于预设阈值的多个节点确定为同一类别；

提取同一类别中多个节点的任务属性的共性部分，作为所述多个节点的父节点的任务属性。

3、根据权利要求1所述的方法，其特征在于，所述基于各父节点的任务属性，训练各父节点的网络参数，包括：

25 针对任一父节点，基于该父节点的任务属性，将该父节点的各子节点的输出特征作为该父节点的输入，训练该父节点的网络参数。

4、根据权利要求1所述的方法，其特征在于，所述基于各父节点的任务

属性，训练各父节点的网络参数，包括：

针对任一父节点，基于该父节点的任务属性，利用具有特征信号控制机制的预设结构生成该父节点；获取并根据该父节点的各子节点的输出特征中与所述任务属性相关的各输出特征的加权组合，训练该父节点的网络参数。

- 5 5、根据权利要求1所述的方法，其特征在于，所述在对各网络层中各节点完成训练后，确定所述树状网络拓扑结构对应的深度神经网络训练结束，包括：

从所述叶节点所处的网络层开始，按照从底层至顶层的训练顺序，依次训练各网络层中节点的网络参数，完成各节点的训练；

- 10 在顶层的各节点训练完毕后，确定所述树状网络拓扑结构对应的深度神经网络训练结束。

6、一种深度神经网络训练装置，其特征在于，所述装置包括：

- 15 获取模块，用于针对树状网络拓扑结构中的当前网络层，获取所述当前网络层中各节点的任务属性，其中，所述树状网络拓扑结构中的各节点为对应于不同任务的神经网络，所述树状网络拓扑结构中的叶节点为针对指定任务已完成训练的神经网络；

聚类模块，用于基于所述当前网络层中各节点的任务属性，对所述当前网络层中各节点进行聚类分析，提取同一类别中多个节点的任务属性的共性部分，作为所述多个节点的父节点的任务属性；

- 20 训练模块，用于基于各父节点的任务属性，训练各父节点的网络参数；

确定模块，用于在对各网络层中各节点完成训练后，确定所述树状网络拓扑结构对应的深度神经网络训练结束。

7、根据权利要求6所述的装置，其特征在于，所述聚类模块，具体用于：

- 25 根据所述当前网络层中各节点的任务属性，通过预设相似性度量算法，生成对应于各节点的任务属性的相似性度量矩阵；

根据所述相似性度量矩阵，将相似性大于预设阈值的多个节点确定为同

一类别；

提取同一类别中多个节点的任务属性的共性部分，作为所述多个节点的父节点的任务属性。

8、根据权利要求6所述的装置，其特征在于，所述训练模块，具体用于：

5 针对任一父节点，基于该父节点的任务属性，将该父节点的各子节点的输出特征作为该父节点的输入，训练该父节点的网络参数。

9、根据权利要求6所述的装置，其特征在于，所述训练模块，具体用于：

10 针对任一父节点，基于该父节点的任务属性，利用具有特征信号控制机制的预设结构生成该父节点；获取并根据该父节点各子节点的输出特征中与所述任务属性相关的各输出特征的加权组合，训练该父节点的网络参数。

10、根据权利要求6所述的装置，其特征在于，所述确定模块，具体用于：

从所述叶节点所处的网络层开始，按照从底层至顶层的训练顺序，依次训练各网络层中节点的网络参数，完成各节点的训练；

15 在顶层的各节点训练完毕后，确定所述树状网络拓扑结构对应的深度神经网络训练结束。

11、一种计算机可读存储介质，其特征在于，用于存储可执行代码，所述可执行代码用于在运行时执行：权利要求 1-5 任一项所述的深度神经网络训练方法。

20 12、一种应用程序，其特征在于，用于在运行时执行：权利要求1-5任一项所述的深度神经网络训练方法。

13、一种计算机设备，其特征在于，包括处理器和计算机可读存储介质，其中，

所述计算机可读存储介质，用于存放可执行代码；

25 所述处理器，用于执行所述计算机可读存储介质上所存放的可执行代码时执行：权利要求 1-5 任一项所述的深度神经网络训练方法。

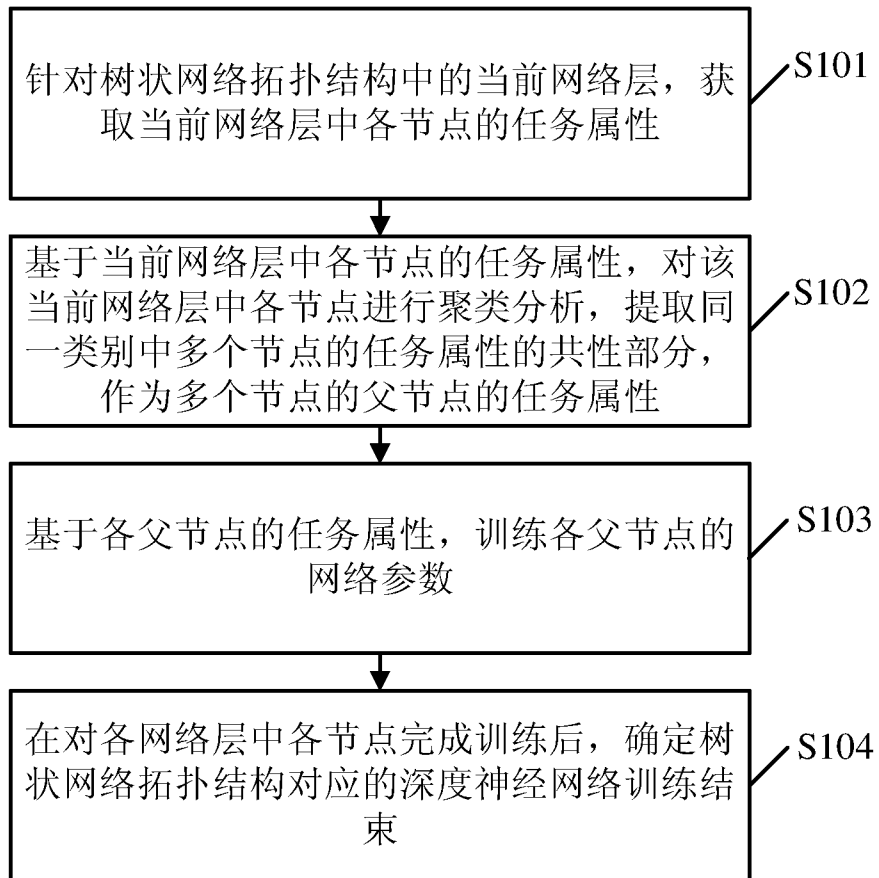


图1

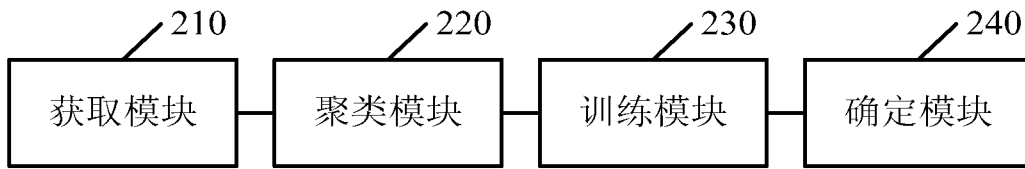


图2

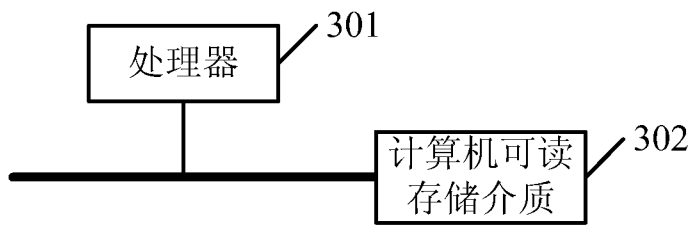


图3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/119725

A. CLASSIFICATION OF SUBJECT MATTER

G06N 3/08(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI, IEEE: 神经网络, 深度学习, 深度神经网络, 卷积神经网络, 树状, 分支, 拓扑, 叶, 节点神经网络, 多任务, neural networks, deep learning, deep neural networks, DNN, tree, leaf, multi-task, multitask learning

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 106295803 A (SUZHOU INSTITUTE FOR ADVANCED STUDY, USTC ET AL.) 04 January 2017 (2017-01-04) description, paragraphs 0006-0012, and figure 1	1-13
A	CN 107368294 A (ALIBABA GROUP HOLDING LIMITED) 21 November 2017 (2017-11-21) entire document	1-13
A	CN 107424612 A (BEIJING SOGOU TECHNOLOGY DEVELOPMENT CO., LTD.) 01 December 2017 (2017-12-01) entire document	1-13
A	US 9564123 B1 (SOUNDHOUND, INC.) 07 February 2017 (2017-02-07) entire document	1-13

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

20 February 2019

Date of mailing of the international search report

06 March 2019

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/119725

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	106295803	A	04 January 2017	None	
CN	107368294	A	21 November 2017	None	
CN	107424612	A	01 December 2017	None	
US	9564123	B1	07 February 2017	None	

国际检索报告

国际申请号

PCT/CN2018/119725

<p>A. 主题的分类</p> <p>G06N 3/08(2006.01) i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																	
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G06N</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CNPAT, WPI, EPODOC, CNKI, IEEE: 神经网络, 深度学习, 深度神经网络, 卷积神经网络, 树状, 分支, 拓扑, 叶, 节点神经网络, 多任务, neural networks, deep learning, deep neural networks, DNN, tree, leaf, multi-task, multitask learning</p>																	
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 106295803 A (中国科学技术大学苏州研究院 等) 2017年 1月 4日 (2017 - 01 - 04) 说明书第0006--0012段, 附图1</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>CN 107368294 A (阿里巴巴集团控股有限公司) 2017年 11月 21日 (2017 - 11 - 21) 全文</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>CN 107424612 A (北京搜狗科技发展有限公司) 2017年 12月 1日 (2017 - 12 - 01) 全文</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>US 9564123 B1 (SOUNDHOUND, INC.) 2017年 2月 7日 (2017 - 02 - 07) 全文</td> <td>1-13</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 106295803 A (中国科学技术大学苏州研究院 等) 2017年 1月 4日 (2017 - 01 - 04) 说明书第0006--0012段, 附图1	1-13	A	CN 107368294 A (阿里巴巴集团控股有限公司) 2017年 11月 21日 (2017 - 11 - 21) 全文	1-13	A	CN 107424612 A (北京搜狗科技发展有限公司) 2017年 12月 1日 (2017 - 12 - 01) 全文	1-13	A	US 9564123 B1 (SOUNDHOUND, INC.) 2017年 2月 7日 (2017 - 02 - 07) 全文	1-13
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
A	CN 106295803 A (中国科学技术大学苏州研究院 等) 2017年 1月 4日 (2017 - 01 - 04) 说明书第0006--0012段, 附图1	1-13															
A	CN 107368294 A (阿里巴巴集团控股有限公司) 2017年 11月 21日 (2017 - 11 - 21) 全文	1-13															
A	CN 107424612 A (北京搜狗科技发展有限公司) 2017年 12月 1日 (2017 - 12 - 01) 全文	1-13															
A	US 9564123 B1 (SOUNDHOUND, INC.) 2017年 2月 7日 (2017 - 02 - 07) 全文	1-13															
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																	
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件</p> <p>“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</p> <p>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</p> <p>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</p> <p>“&” 同族专利的文件</p>																	
国际检索实际完成的日期	2019年 2月 20日	国际检索报告邮寄日期															
		2019年 3月 6日															
ISA/CN的名称和邮寄地址	中国知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	受权官员															
		杨栋															
传真号 (86-10)62019451		电话号码 86-(10)-53961344															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/119725

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	106295803	A	2017年 1月 4日	无	
CN	107368294	A	2017年 11月 21日	无	
CN	107424612	A	2017年 12月 1日	无	
US	9564123	B1	2017年 2月 7日	无	

表 PCT/ISA/210 (同族专利附件) (2015年1月)