(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 111465946 B
(45)授权公告日 2024.05.28

(21)申请号 201880070077.2
(22)申请日 2018.10.26
(65)同一申请的已公布的文献号
申请公布号 CN 111465946 A
(43)申请公布日 2020.07.28
(30)优先权数据
62/578,356 2017.10.27 US
(85)PCT国际申请进入国家阶段日
2020.04.27
(86)PCT国际申请的申请数据
PCT/EP2018/079401 2018.10.26
(87)PCT国际申请的公布数据
WO2019/081705 EN 2019.05.02
(73)专利权人 渊慧科技有限公司
地址 英国伦敦
(72)发明人 C.T.费尔南多 K.西蒙扬
K.卡乌克科格鲁 刘寒骁
O.温亚尔斯
(74)专利代理机构 北京市柳沈律师事务所 11105
专利代理师 金玉洁
(51)Int.Cl.
G06N 3/082(2023.01)
G06N 3/045(2023.01)
(56)对比文件
CN 106471526 A,2017.03.01
CN 105719001 A,2016.06.29
Esteban Real 等."Large-Scale Evolution of Image Classifiers".《arXiv:1703.01041v2 [cs.NE]》.2017,第1-18页.
Masanori Suganuma 等."A Genetic Programming Approach to Designing Convolutional Neural Network Architectures".《arXiv:1704.00764v2 [cs.NE]》.2017,第1-9页.
审查员 马金驹

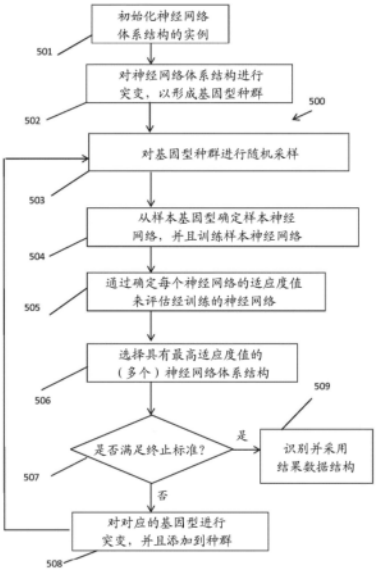
权利要求书4页 说明书14页 附图4页

(54)发明名称

使用分层表示进行神经网络体系结构搜索

(57)摘要

一种用于自动确定神经网络体系结构的计算机实现的方法将神经网络体系结构表示为在多个层级中定义有向非循环分层图形集合的数据结构。每个图形都有输入、输出以及输入和输出之间的多个节点。在每一层级,对应的节点集合通过有向边成对连接,有向边指示对一个节点的输出执行的操作,以生成对另一个节点的输入。每个层级与对应的操作集合相关联。在最低层级,与每个边相关联的操作是从原语操作集合中选择的。该方法包括重复生成新的样本神经网络体系结构,并评估它们的适应度。修改是通过选择层级、在该层级选择两个节点、以及根据与分层结构的更低层级相关联的操作来修改、移除或添加这些节点之间的边来执行的。



CN 111465946 B

1.一种用于自动确定神经网络体系结构的计算机实现的方法,包括:

生成将神经网络体系结构表示为定义分层图形集合的数据结构的数据,所述分层图形集合包括图形的一系列分层层级,每个图形具有输入、输出、输入和输出之间的多个节点以及一个或多个边,每个边连接两个相应节点,其中,节点对应于由所述神经网络体系结构定义的神经网络内的特征图,每个边将所述边的输入节点连接到所述边的输出节点并且对应于对所述边的输入节点的特征图执行的、以向所述边的输出节点提供特征图的操作,使得每个图形执行操作,所述一系列分层层级包括最低分层层级和多个附加分层层级,生成所述数据包括定义所述分层图形集合,并且定义所述分层图形集合包括:

定义包括一个或多个最低层级图形的最低分层层级,其中,与所述一个或多个最低层级图形的边对应的操作是从原语神经网络操作集合中选择的,并且

递归地定义附加分层层级,其中,每个相应附加分层层级包括一个或多个图形的相应集合,其中,递归地定义每个相应附加分层层级包括,对于所述相应附加分层层级中的一个或多个图形中的每个图形,从所述一系列分层层级中在所述相应附加分层层级之前的分层层级中的图形中选择较低层级图形的集合,并且通过装配从在所述相应附加分层层级之前的分层层级中的图形中选择的较低层级图形来生成表示所述图形的数据;

基于定义分层图形集合的数据结构初始化两个或更多个样本神经网络体系结构的种群,其中,所述种群中的每个相应样本神经网络体系结构通过修改由所述分层图形集合的边执行的操作中的一个或多个操作来初始化;

生成具有所述种群中的样本神经网络体系结构的样本神经网络;

训练所述样本神经网络以接收传感器数据并生成指示所述传感器数据与多个类别之一相关联的对应输出;或者训练所述样本神经网络以处理包括字序列的输入数据序列;或者训练所述样本神经网络以生成识别要由机器人执行的动作的对应输出;

通过确定所述样本神经网络中的每个样本神经网络的适应度值来评估所述样本神经网络;以及

根据所确定的适应度值来选择所述种群中的样本神经网络体系结构中的一个或多个样本神经网络体系结构,以确定神经网络体系结构。

2.根据权利要求1所述的方法,其中,对于每个相应附加分层层级,与所述相应附加分层层级中的一个或多个图形的边对应的操作是从所述一系列分层层级中的所述相应附加分层层级之前的分层层级中的一个或多个图形所执行的操作集合中选择的。

3.根据权利要求1所述的方法,其中,所述原语神经网络操作集合包括恒等操作,其中当在特征图上执行所述恒等操作时,使所述特征图不变。

4.根据权利要求1所述的方法,其中,所述原语神经网络操作集合包括至少一个卷积操作,其中当在特征图上执行所述卷积操作时,所述特征图的分辨率不变。

5.根据权利要求1所述的方法,其中,所述原语神经网络操作集合包括随后是批量归一化操作的至少一个卷积操作。

6.根据权利要求1所述的方法,其中,所述原语神经网络操作集合包括无连接操作,所述无连接操作定义了由所述操作对应的边链接的节点之间不具有直接连接。

7.根据权利要求1所述的方法,其中,所述原语神经网络操作集合包括当在特征图上执行时使所述特征图不变的恒等操作,并且所述初始化包括将由所述分层图形集合的边执行

的操作中的至少一些操作初始化为恒等操作。

8.根据权利要求1所述的方法,其中,所述选择包括将来自所述种群的样本神经网络体系结构的适应度值彼此进行比较。

9.根据权利要求1所述的方法,其中,确定至少两个样本神经网络体系结构还包括对根据所确定的适应度值选择的样本神经网络体系结构进行突变,并重复生成、训练、评估和选择以演化样本神经网络体系结构。

10.根据权利要求9所述的方法,其中,所述突变包括选择分层层级之一、选择所选择的层级中的图形、选择所选择的图形中的前置节点和后置节点、以及用另一操作替换对应于连接所选择的节点的边的操作。

11.根据权利要求1所述的方法,其中,所述选择还包括:

提供多个评估工作者,所述多个评估工作者中的每个评估工作者被配置为评估样本神经网络体系结构;

当每个工作者变得可用时,将样本神经网络体系结构分配给工作者用于评估,其中每个工作者执行具有样本神经网络体系结构的样本神经网络的生成、训练和评估,并且当完成时,被分配另外的样本神经网络体系结构;

将评估结果添加到由评估工作者共享的数据存储中;以及

使用数据存储中的评估来控制选择。

12.根据权利要求1所述的方法,还包括根据所确定的神经网络体系结构构建神经网络。

13.根据权利要求12所述的方法,还包括在用于训练和/或推断的神经网络系统中使用所述神经网络。

14.根据权利要求12所述的方法,还包括使所述神经网络在神经网络系统中可用于经由API进行训练和/或推断。

15.根据权利要求13所述的方法,其中,所述神经网络系统包括所述神经网络的多个实例。

16.一种自动确定神经网络体系结构的系统,包括一个或多个计算机和存储指令的一个或多个存储设备,所述指令当被一个或多个计算机执行时,使所述一个或多个计算机执行用于自动确定神经网络体系结构的操作,所述操作包括:

生成将神经网络体系结构表示为定义分层图形集合的数据结构的数据,所述分层图形集合包括图形的一系列分层层级,每个图形具有输入、输出、输入和输出之间的多个节点以及一个或多个边,每个边连接两个相应节点,其中,节点对应于由所述神经网络体系结构定义的神经网络内的特征图,每个边将所述边的输入节点连接到所述边的输出节点并且对应于对所述边的输入节点的特征图执行的、以向所述边的输出节点提供特征图的操作,使得每个图形执行操作,所述一系列分层层级包括最低分层层级和多个附加分层层级,生成所述数据包括定义所述分层图形集合,并且定义所述分层图形集合包括:

定义包括一个或多个最低层级图形的最低分层层级,其中,与所述一个或多个最低层级图形的边对应的操作是从原语神经网络操作集合中选择的,以及

递归地定义附加分层层级,其中,每个相应附加分层层级包括一个或多个图形的相应集合,其中,递归地定义每个相应附加分层层级包括,对于所述相应附加分层层级中的一个

或多个图形中的每个图形,从所述一系列分层层级中在所述相应附加分层层级之前的分层层级中的图形中选择较低层级图形的集合,并且通过装配从在所述相应附加分层层级之前的分层层级中的图形中选择的较低层级图形来生成表示所述图形的数据;

基于定义分层图形集合的数据结构初始化两个或更多个样本神经网络体系结构的种群,其中,所述种群中的每个相应样本神经网络体系结构通过修改由所述分层图形集合的边执行的操作中的一个或多个操作来初始化;

生成具有所述种群中的样本神经网络体系结构的样本神经网络;

训练所述样本神经网络以接收传感器数据并生成指示所述传感器数据与多个类别之一相关联的对应输出;或者训练所述样本神经网络以处理包括字序列的输入数据序列;或者训练所述样本神经网络以生成标识要由机器人执行的动作的对应输出;

通过确定所述样本神经网络中的每个样本神经网络的适应度值来评估所述样本神经网络;以及

根据所确定的适应度值来选择所述种群中的样本神经网络体系结构中的一个或多个样本神经网络体系结构,以确定神经网络体系结构。

17.根据权利要求16所述的系统,其中,对于每个相应附加分层层级,与所述相应附加分层层级中的一个或多个图形的边对应的操作是从所述一系列分层层级中的所述相应附加分层层级之前的分层层级中的一个或多个图形所执行的操作集合中选择的。

18.根据权利要求16所述的系统,其中,确定至少两个样本神经网络体系结构包括通过以下来初始化一个样本神经网络体系结构或样本神经网络体系结构种群:

通过至少定义由所述分层图形集合的边执行的操作,初始化定义所述分层图形集合的数据结构的至少一个实例,以及

通过修改由所述分层图形集合的边执行的操作来对数据结构的所述至少一个实例进行突变。

19.一个或多个存储指令的非暂时性计算机可读存储介质,所述指令当被一个或多个计算机执行时,使得所述一个或多个计算机执行用于自动确定神经网络体系结构的操作,所述操作包括:

生成将神经网络体系结构表示为定义分层图形集合的数据结构的数据,所述分层图形集合包括图形的一系列分层层级,每个图形具有输入、输出、输入和输出之间的多个节点以及一个或多个边,每个边连接两个相应节点,其中,节点对应于由所述神经网络体系结构定义的神经网络内的特征图,每个边将所述边的输入节点连接到所述边的输出节点并且对应于对所述边的输入节点的特征图执行的、以向所述边的输出节点提供特征图的操作,使得每个图形执行操作,所述一系列分层层级包括最低分层层级和多个附加分层层级,生成所述数据包括定义所述分层图形集合,并且定义所述分层图形集合包括:

定义包括一个或多个最低层级图形的最低分层层级,其中,与所述一个或多个最低层级图形的边对应的操作是从原语神经网络操作集合中选择的,以及

递归地定义附加分层层级,其中,每个相应附加分层层级包括一个或多个图形的相应集合,其中,递归地定义每个相应附加分层层级包括,对于所述相应附加分层层级中的一个或多个图形中的每个图形,从所述一系列分层层级中在所述相应附加分层层级之前的分层层级中的图形中选择较低层级图形的集合,并且通过装配从在所述相应附加分层层级之前

的分层层级中的图形中选择的较低层级图形来生成表示所述图形的数据;

基于定义分层图形集合的数据结构初始化两个或更多个样本神经网络体系结构的种群,其中,所述种群中的每个相应样本神经网络体系结构通过修改由所述分层图形集合的边执行的操作中的一个或多个操作来初始化;

生成具有所述种群中的样本神经网络体系结构的样本神经网络;

训练所述样本神经网络以接收传感器数据并生成指示所述传感器数据与多个类别之一相关联的对应输出;或者训练所述样本神经网络以处理包括字序列的输入数据序列;或者训练所述样本神经网络以生成标识要由机器人执行的动作的对应输出;

通过确定所述样本神经网络中的每个样本神经网络的适应度值来评估所述样本神经网络;以及

根据所确定的适应度值来选择所述种群中的样本神经网络体系结构中的一个或多个样本神经网络体系结构,以确定神经网络体系结构。

使用分层表示进行神经网络体系结构搜索

技术领域

[0001] 本说明书涉及用于自动确定神经网络体系结构的系统和方法。

背景技术

[0002] 神经网络是采用一层或多层非线性单元来预测接收到的输入的输出的机器学习模型。一些神经网络除了输出层之外还包括一个或多个隐藏层。每个隐藏层的输出被用作网络中下一层(即下一隐藏层或输出层)的输入。网络的每一层根据相应参数集的当前值从接收的输入生成输出。

[0003] 一些神经网络是循环神经网络(recurrent neural network)。循环神经网络是接收输入序列并从输入序列生成输出序列的神经网络。特别地,循环神经网络可以使用来自前一时间步(time step)的网络的内部状态中的一些或全部内部状态来计算当前时间步的输出。循环神经网络的示例是长短期记忆(long short term memory,LSTM)神经网络,其中LSTM神经网络包括一个或多个LSTM记忆块。每个LSTM记忆块可以包括一个或多个单元,每个单元包括输入门、遗忘门和输出门,其允许该单元存储该单元的先前状态,例如用于生成当前激活或提供给LSTM神经网络的其他组件。

发明内容

[0004] 本说明书描述了一种确定神经网络体系结构的系统和方法,其被实现为一个或多个位置中的一个或多个计算机上的一个或多个计算机程序。神经网络体系结构例如可以是卷积和/或循环神经网络的体系结构。

[0005] 包括对应于所确定的神经网络体系结构的神经网络的神经网络系统可以被配置为接收任何种类的数字数据输入,并基于该输入生成任何种类的得分、分类或回归输出。例如,它可以是被配置为对输入数据(例如由传感器捕获并表征真实世界的数据项)进行分类的神经网络系统。例如,数据项可以是图像(诸如由相机捕获的图像)或者声音信号(诸如由麦克风捕获的声音信号)。神经网络系统可以通过基于由传感器数据的实例(例如,传感器数据(诸如由相机捕获的图像)的真实世界实例;但是原则上可以模拟传感器数据的实例中的一些或所有)以及指示传感器数据的实例与预定类别集合中的一个相关联的对应标签组成的训练集通过监督学习来训练。可以通过将由传感器捕获的真实世界的传感器数据输入到神经网络系统中,并且获得对应的标签作为神经网络系统的输出,以指示输入的传感器数据与类别之一相关联,来使用经训练的神经网络系统。

[0006] 可替换地,神经网络系统可以被配置为处理输入数据序列(诸如字或数据项序列),或者它可以是生成神经网络系统(generative neural network system)的一部分,或者它可以是强化学习系统的一部分,在这种情况下,它可以生成标识要由智能体(agent)(可以例如是机器人)执行的动作的输出。

[0007] 总的来说,本公开提出了一种用于自动确定神经网络体系结构的计算机实现的方法,该神经网络体系结构以定义分层的有向非循环图形(directed acyclic graph)集合的

数据结构表示任何给定的神经网络体系结构。每个图形具有输入、输出以及输入和输出之间的多个节点。

[0008] 如上所述,图形集合是分层的,在多个层级中的每一个层级定义。这些层级可以与对应的节点子集相关联。每个层级的节点子集包括下一更高层级的节点子集、以及(多个)附加节点。分层表示在每一层级包括对应的节点子集的相应节点对之间的有向边(directed edge)。边(edge)指示对节点之一的输出执行操作以生成到另一个节点的输入。该表示的某个层级的两个节点之间的边可以在下一更低层级由一个或多个附加节点(即与所述下一更低层级相关联但不是与所述某个层级相关联的节点子集的一部分的节点子集的节点)代替表示,并且由所述某个层级的两个节点之间的边和/或这两个节点和附加节点之间的边和/或附加节点中的两个附加节点之间的边来表示。

[0009] 每个层级可以与对应的操作集合相关联。除最低层级之外的给定层级的任何节点对之间的操作可以是与更低层级相关联的操作子集之一。在倒数第二层级上,与每个边相关联的操作可以从构成分层结构的最低层级的原语操作集合中选择。

[0010] 该方法包括通过修改现有的(多个)神经网络体系结构来重复生成新的样本神经网络体系结构,并评估它们的适应度。通过选择现有神经网络体系结构的层级,选择该层级处的两个节点,以及在这些节点之间修改、移除或添加操作来执行修改。修改或添加操作是使用与分层结构的更低层级相关联的操作来执行的。

[0011] 更具体地表达上述一般概念,根据一个创新方面,一种用于自动确定神经网络体系结构的计算机实现的方法。该方法将神经网络体系结构表示为定义的分层图形(具体地,分层有向非循环图形)集合的数据结构。每个图形具有输入、输出以及输入和输出之间的多个节点。每个节点(或者至少每个除了分层结构的输入节点之外的节点)对应于(输出)由神经网络体系结构定义的神经网络内的特征图(feature map)。该特征图可以包括在由该体系结构表示的神经网络中嵌入从神经网络的输入向量导出的特征。两个节点由(有向)边连接,其中该边将该边的输入节点连接到该边的输出节点。每个边对应于对该边的输入节点的特征图执行的操作,以向该边的输出节点提供特征图。因此,每个图形执行操作(之后描述为模体(motif))。

[0012] 分层图形集合包括图形的多个分层层级,具体地包括一个或多个最低层级图形的最低分层层级、以及包括一个或多个下一层级图形的一个或多个下一更高分层层级。从原语神经网络操作中选择对应于一个或多个最低层级图形的边的操作(模体)(在之后的描述中,最低层级图形在层级2,并且原语操作本身在层级1)。对应于一个或多个下一层级图形的边的操作是从由一个或多个最低层级图形执行的操作集合中选择的。

[0013] 该方法可以包括确定至少两个样本神经网络体系结构,具体地通过至少定义由表示样本神经网络体系结构的相应的分层图形集合中的边执行的操作。这些样本体系结构可以用于确定用于构建神经网络的神经网络体系结构。这可能涉及使用遗传算法,在这种情况下,样本神经网络体系结构可能随着时间的推移而演化。可替换地,它可以包括随机搜索算法,在这种情况下,可以顺序地评估样本体系结构,直到识别出满足要求的体系结构。可替换地,可以采用其他技术。

[0014] 因此,该方法还可以包括生成具有样本神经网络体系结构的样本神经网络,训练样本神经网络,通过确定样本神经网络中的每一个的适应度值来评估样本神经网络,然后

根据确定的适应度值选择一个或多个样本神经网络体系结构以确定神经网络体系结构。可以使用适应度测试来计算适应度值,该适应度测试例如可以是对包括样本神经网络并被训练以执行计算测试的神经网络系统执行该任务的能力的测试。

[0015] 在一些实施方式中,多个图形分层层级包括一系列下一更高的分层层级。对应于分层层级中的每个分层层级的边的操作是从操作集合(即前一层级的模体)中选择的,其中该操作集合是由该系列中的至少一个前一(更低)层级的一个或多个图形执行的。这些操作可以但不一定只来自紧邻的前一层级,例如,层级处的操作也可以包括一个或多个原语操作,诸如下面描述的无连接/无边操作。

[0016] 原语神经网络操作集合可以包括例如一个或多个卷积操作,其中该一个或多个卷积操作中的一些或全部可以是步长1的卷积操作。优选地,这种卷积操作被配置为使得当在特征图上被执行时,它使特征图的输入分辨率不变。这可以促进在实现方法时不同层级处的操作的链。出于类似的原因,卷积操作之后可以是批量归一化(batch normalization)操作。可以提供例如具有不同的核大小、和/或可分离/不可分离类型的多个卷积操作。

[0017] 原语神经网络操作集合可以附加地或可替换地包括恒等操作,即当在特征图上执行时使特征图基本不变的操作。这可能是有利的,因为通过改变两个节点是否彼此等价,可以允许节点被有效地添加到图形中/从图形中移除。

[0018] 原语神经网络操作集合可以附加地或可替换地包括无连接操作,该无连接操作定义了由该操作所对应的边链接的节点之间不具有直接连接。这实际上是无边操作,因为无连接操作有效地定义了节点之间不具有边。也就是说,节点中的一个节点的输出不会作为输入传递给该节点中的另一个节点。

[0019] 原语神经网络操作集合可以附加地或可替换地包括最大池化操作、平均池化操作和诸如LSTM或GRU(gated recurrent unit,门控循环单元)操作的循环操作中的一个或多个。

[0020] 确定至少两个样本神经网络体系结构可以包括初始化样本神经网络体系结构的种群(population)。这可以包括通过定义至少由分层图形集合的边执行的操作(在一些实施方式中,对恒等操作)来初始化定义分层图形集合的数据结构的至少一个实例。然后,这些操作可以通过随机修改由分层图形集合的边执行的操作来突变(mutate)。这种初始化对于包括随机搜索的各种搜索过程是有用的,在这种情况下,对于每个搜索迭代,可以只生成/初始化一个样本体系结构。

[0021] 对于遗传算法,可以采用更大种群的样本神经网络体系结构。然后,该选择可以包括将来自种群的样本神经网络体系结构的适应度值彼此进行比较。这可以通过任何一种用于遗传算法的基于适应度/奖励的选择机制来实现,但是已经发现锦标赛选择(tournament selection)工作得特别好。优选地,锦标赛选择被用于相当大比例(例如等于或大于1%、3%或5%)的种群。因此,确定样本神经网络体系结构还可以包括对根据所确定的适应度值选择的样本神经网络体系结构进行突变,并重复生成、训练、评估和选择以演化样本神经网络体系结构。

[0022] 该突变可以包括选择分层层级之一、选择所选择的层级中的图形(模体)、以及选择所选择的图形中的前置节点(predecessor node)和后置节点(successor node)。该选择中的一些或所有选择可以是随机的。可以选择前置节点和后置节点,使得在突变之前的神

经网络中,后置节点的输出没有传递(直接或经由其他节点传递)到前置节点。也就是说,可以选择前置节点和后置节点来与现有的边/操作保持一致,以确保突变不会改变图形的非循环性质;一般地,在有向非循环图形中具有固有的拓扑排序。

[0023] 该突变还可以包括用另一操作(可以是随机选择的)替换对应于连接所选择的节点的边的操作。选择与原始操作相同的操作可能是允许的,但不一定是允许的。

[0024] 在一些实施方式中,该选择还包括提供多个评估工作者,这些评估工作者可以被实现为软件智能体,每个被配置为评估样本神经网络体系结构。当每个工作者变得可用时,样本神经网络体系结构可以分配给工作者用于评估。例如,工作者可以从定义评估体系结构的数据队列中挑选样本。然后,每个工作者可以执行对具有相应体系结构的样本神经网络的生成、训练和评估。当完成时,可以异步地给工作者分配另外的样本神经网络体系结构进行评估。工作者生成的评估结果可以被添加到共享数据存储中。控制器软件还可以访问共享数据存储,以便根据它们的评估结果来控制对体系结构的选择。

[0025] 许多不同的技术可以用于评估适应度值(即神经网络的性能),包括诸如网络在执行诸如分类任务的计算任务时的准确性、网络内的损失或成本函数或强化学习任务中获得的得分/奖励等。在确定误差或平均误差的情况下,适应度值可以是该误差的负值。样本神经网络的适应度值可以可选地被确定为在执行计算任务时包括样本神经网络的一个或多个实例的(更大的)神经网络系统的性能的度量。

[0026] 该方法还可以包括如果任何进一步的设计工作是必要的则根据所确定的神经网络体系结构设计神经网络、和/或用该体系结构构建神经网络。该构建可以是自动的。它可以包括将所确定的神经网络体系结构的一个或多个实例合并到(更大的)神经网络系统中。该方法还可以包括使用神经网络进行训练和/或推断;或者使神经网络可经由API(application programming interface,应用编程接口)使用(例如,用于训练和/或推断)。因此,该方法/系统的实施方式的一个用例涉及用户为机器学习处理提供数据;然后,该方法/系统可以用于生成(具体地,演化)神经网络体系结构;然后,具有该结构的神经网络可以供用户使用。

[0027] 在相关方面,自动确定神经网络体系结构的系统包括存储数据结构的数据结构存储器。每个数据结构都被配置为将神经网络体系结构表示为分层图形集合。每个图形可以具有输入、输出以及输入和输出之间的多个节点。节点可以对应于由神经网络体系结构定义的神经网络内的特征图,例如在输出该特征图的意义上。节点的输入可以是作为其他节点的输出的一个或多个特征图。每个节点对可以由一个边连接。每个边可以将该边的输入节点连接到该边的输出节点,并且可以对应于在该边的输入节点的特征图上执行的操作,以向该边的输出节点提供特征图,使得每个图形执行操作。

[0028] 分层图形集合可以包括图形的多个分层层级、包括一个或多个最低层级图形的最低分层层级、和包括一个或多个下一层级图形的下一更高的分层层级。对应于一个或多个最低层级图形的边的操作可以从原语神经网络操作集合中选择。对应于一个或多个下一层级图形的边的操作可以从由一个或多个最低层级图形执行的操作集合中选择。该系统可以包括样本模块,其中样本模块用于通过至少定义由表示样本神经网络体系结构的相应的分层图形集合中的边执行的操作来定义至少两个样本神经网络体系结构。该系统还可以包括评估模块,其中评估模块用于:生成具有样本神经网络体系结构的样本神经网络;训练样本

神经网络;通过确定样本神经网络中的每一个的适应度值来评估样本神经网络;和/或根据确定的适应度值选择样本神经网络体系结构中的一个或多个样本神经网络体系结构。

[0029] 本说明书中描述的主题可以在特定实施例中实现,以便实现以下优点中的一个或多个。

[0030] 上述系统和方法的实施例可以生成在计算任务(诸如处理传感器数据以生成分类数据或生成控制数据的输出)方面具有与人类设计的神经网络一样好或更好的性能以及具有更短的时间空间的性能的神经网络体系结构。因此,该方法的实施例导致改进的计算机系统,对于给定水平的计算训练资源(例如,处理器时间),该计算机系统可以被训练为以更高的效率执行计算任务,诸如传感器数据分类。

[0031] 此外,已经发现,与用于执行体系结构搜索的已知算法相比,所描述的系统和方法的实施例执行神经网络体系结构搜索,在计算资源上有很大的节省(例如,具有给定处理速率的处理器单元使用的处理时间减少)。在各种实验中,与已知技术相比,观察到速度增加了6倍至200倍以上,从而产生执行图像分类任务达到基本相同的精度水平或者在某些情况下更好的神经网络体系结构。

[0032] 由所述方法生成的神经网络可以在硬件中实现为新颖的计算机系统,该计算机系统包括实现神经网络的节点和边的相应功能的多个处理器单元,处理器单元通过由方法设计的物理信号路径(电气和/或光学电路)连接。因此,神经网络具有不同于任何现有计算机系统的硬件结构。可替换地,该方法可以由模拟新颖硬件体系结构的处理器单元和信号路径的一个或多个处理器实现。

附图说明

[0033] 为了举例,现在将仅参考以下附图来描述所公开的方法的示例,附图中:

[0034] 图1示出了可以在根据本公开的方法中使用的原语操作;

[0035] 由图2(a)和2(b)构成的图2示出了在根据本公开的方法中使用图1的原语操作形成层级2模体;

[0036] 由图3(a)和3(b)构成的图3示出了另外两个层级2模体;

[0037] 由图4(a)和4(b)构成的图4示出了在图2的方法中使用图2和3的层级2模体形成层级3模体;

[0038] 图5是示出根据本公开的用于确定神经网络体系结构并使用所确定的神经网络体系结构来执行计算任务的方法的步骤的流程图;和

[0039] 图6示意性地示出了用于执行图1的方法的系统的结构。

具体实施方式

[0040] 下面参照图5描述根据本公开的用于确定(生成)神经网络体系结构的方法的示例。在此之前,我们描述由该方法生成的神经网络。

[0041] 每个神经网络具有神经体系结构,神经体系结构可以表示为由通过有向边成对连接的多个节点构成的有向非循环图形。这些节点包括至少一个输入节点和至少一个输出节点。为简单起见,我们将考虑只有单个输入节点(“单源”)和单个输出节点(“单汇聚(sink)”)的情况,使得神经网络将源处的输入转换为汇聚处的输出。该网络是非循环的,即

不存在从任何给定节点开始、沿着有向边(沿边的方向)延伸并返回该给定节点的路径。

[0042] 图形的每个节点可以基于其输入生成特征图。特征图可以是向量的数组(例如,二维数组,但是在该示例的变型中,该数组可以具有任何其他维度)。每个向量具有一个或多个分量;典型地,特征图的每个向量具有相同(多个)数量的分量。

[0043] 从节点中的第一节点延伸到该节点中的第二节点的每个有向边与原语操作集合中相应的一个原语操作相关联,所述相应的一个原语操作在第一节点中变换特征图以产生该边将其传递到第二节点的特征图。

[0044] 形式上,神经网络由多个节点(用变量i标记)构成,所述多个节点中的每个节点从其输入形成特征图x_i。节点通过有向边成对连接。从第一节点i到第二节点j的有向边对特征图x_i执行操作,以生成第二节点的输入。神经网络体系结构由表示(G,o)定义,表示(G,o)由两个元素构成:

[0045] 1.可用操作池$o=\{o_1,o_2,...\}$。

[0046] 2.指定操作的神经网络图的邻接矩阵G,其中$G_{ij}=k$意味着该图包含节点i和j,具有对应于第k个操作o_k的从节点j到节点i的有向边。

[0047] 该体系结构(也可以替代地被表示为arch)是通过根据邻接矩阵G的装配操作o来获得的:

[0048] arch=assemble(G,o)。(1)

[0049] 单个输入节点可由i=1指定。节点数为|G|,输出节点为i=|G|。对其他节点进行编号,使得节点j的集合具有小于i的标签(网络是非循环的,这保证了这种拓扑排序是可能的),其中从该节点j的集合中可以通过沿着由边中的一个边在由该边指定的方向上移动而到达给定的节点i。该节点集合被称为节点i的直接前置节点(direct predecessor node)。因此,节点i的特征图x_i是从其直接前置节点的特征图x_j中按照拓扑排序获得的:

[0050] $$x_i = merge\ [\{o_{G_{ij}}(x_j)\}_{j<i}],\ i = 2... |G| \quad (2)$$

[0051] 这里,merge(合并)是将多个特征图合并成一个特征图的操作。在一个示例中,这可以被实现为深度上级联(depthwise concatenation)(即,x_i是向量$\{o_{G_{ij}}(x_j)\}_{j<i}$的级联)。执行merge操作的替代方式是通过元素上(element-wise)的加法。然而,这不太灵活,因为它要求传入的特征图包含相同数量的通道(组件)。在任何情况下,这种合并操作可以由被执行1×1卷积的有向边连接的两个节点的链执行,其中两个节点中的每个节点通过级联执行合并操作。

[0052] 分层神经网络体系结构表示的关键思想在于,神经网络体系结构是根据由多个层级构成的分层结构来构建的。除了最低层级之外的每个层级与节点子集相关联。最高层级可以例如仅与单个输入节点和单个输出节点相关联。每个更低层级与下一更高层级的节点子集加上一个或多个附加节点相关联。倒数第二层级(“层级2”)与神经网络体系结构的所有节点相关联。在除最低层级之外的每个层级处,我们可以通过与该层相关联的相应节点对之间的有向边来表示与该层相关联的两个节点之间的数据流。因此,对于层级2,如上所述,该边中的每个边表示可用操作池(称为可用操作)中相应的一个,而在每个更高层级,边表示该层级的两个节点之间(典型地,经由一个或多个更低层级的(多个)节点)的数据流。

[0053] 在分层结构的每个层级处定义一个或多个模体,其中,除了最低层级以外,给定层

级的每个模体由更低层级的模体构成。也就是说,在构建更高层级的模体期间,更低层级的模体被重新用作构建块(操作)。具体地,除了第一层以外,给定层级的每个模体可以由紧邻的更低层级的模体构成。

[0054] 考虑L层级的分层结构,其中第l层级包含M_l个模体。最高层级l=L只包含对应于整个体系结构的单个模体,并且最低层级l=1是原语操作集合。我们递归地将层级l中的第m模体$o_m^{(l)}$定义为更低层级模体的组合$\boldsymbol{o}^{(l-1)} = \left\{o_1^{(l-1)}, o_2^{(l-1)}, \ldots\right\}$。使用这种方式,网络结构矩阵G可以被分解成矩阵集合$\left\{G_m^{(l)}\right\}$,其中矩阵集合$\left\{G_m^{(l)}\right\}$中的每个矩阵根据其网络结构对应于模体之一(即层级l的第m模体),这样我们可以将层级l的第m模体写成:

[0055] $$o_m^{(l)} = assemble\left(G_m^{(l)}, \boldsymbol{o}^{(l-1)}\right) \quad \forall l \in 2, \ldots, L$$

[0056] 数据结构$\left(\left\{\left\{G_m^{(l)}\right\}_{m=1}^{M_l}\right\}_{l=1}^{L}, \boldsymbol{o}^{(1)}\right)$提供了分层神经网络体系结构表示,其中该数据结构通过所有层级的模体的网络结构和底层级原语集合确定。

[0057] 我们现在考虑分层结构中层级l=1的原语操作。在一个示例中,有六个这样的原语操作(l=1,M_l=6),如下所示:

[0058] • C通道的1×1卷积(convolution)

[0059] • C通道的3×3深度上卷积(depthwise convolution)

[0060] • C通道的3×3可分离卷积(separable convolution)

[0061] • 3x 3最大池化(max-pooling)

[0062] • 3x 3平均池化(average-pooling)

[0063] • 恒等(identity)

[0064] 六种可能的这样的原语操作中的三种在图1中示出。这些原语操作中的每一个在卷积神经网络领域中都是众所周知的,并且可以总结如下。

[0065] 由$\sigma_1^{(1)}$表示的第一个原语操作是1x1卷积。这是一种操作,在接收到相等长度的第一向量的二维数组时,形成第二向量的相同数量的第二二维数组,其中:每个第二向量是第一向量中的对应的一个第一向量的函数(通常是线性函数);使用相同的函数从对应的第一向量生成每个第二向量。

[0066] 由$\sigma_2^{(1)}$表示的第二个原语操作是3×3深度上卷积。这是另一种操作,在接收到相等长度的第一向量的二维数组时,形成第二向量的第二二维数组,但是在这种情况下,第二向量中的每个第二向量被形成为第一向量的二维数组的对应的3×3块(patch)中的第一向量的函数(通常是线性函数)。这些块可以是重叠的。同样,每个第二向量使用相同的函数生成,除了可选择地二维数组的边处的第二向量可以由对较少数量的第一向量进行操作的修改函数生成。该卷积是深度上的,其中第二向量的分量(仅)由第一向量的块的对应分量形成。

[0067] 注意,六个原语操作中的另一个(图1中未示出)是可分离卷积,其中卷积核是可分离的,作为两个向量的外积。可分离的卷积可以等价于深度上的卷积(随后是1×1卷积)。

[0068] 由$\sigma_3^{(1)}$表示的另一种可能的原语操作是3×3最大池化操作,这是另一种操作,其中该操作在接收到相等长度的第一向量的二维数组时,形成第二向量的第二二维数组,并且每个第二向量由第一向量的二维数组的对应的3×3块形成,但是在这种情况下,每个第二向量的每个分量是该块中第一向量的对应分量的最大值。

[0069] 类似地,六个可能的原语操作中的另一个(图1中未示出)是3×3平均池化操作,其中3×3平均池化操作与3×3最大池化操作的不同之处在于,每个第二向量由第一向量的数组的对应3×3块形成,使得每个第二向量的每个分量是该块的第一向量的对应分量的平均。

[0070] 在一个示例中,所有原语是步长1,卷积后的特征图被填充以保持其空间分辨率。所有卷积操作之后可以是批量归一化(即增加/减少第一值,并且乘以第二值,以提供零均值和单位方差)、以及ReLU激活(即设置为(i)它们的值和(ii)零中的更高者)。每个卷积操作的输出通道数量被固定为常数C(例如通过填充)。我们注意到具有更大感受野(receptive field)和更多通道的卷积可以被表达为由这样的原语构成的模体。实际上,通过将3×3卷积堆叠成链状结构可以获得大的感受野,并且通过深度上级联来合并多个卷积的输出可以获得具有更多通道的更宽的卷积。

[0071] 恒等操作是一种原语操作,其中当它与连接节点i和节点j的有向边相关联时,将特征图x_i输出到节点j。

[0072] 我们还引入了一个特殊的“无(none)”操作,它表示节点i和j之间没有边。它被添加到每个层级的操作池中。

[0073] 图2(a)示出了使用图1的原语操作定义的可能的层级2网络结构$G_1^{(2)}$的示例,并且图2(b)示出了这些原语是如何基于$G_1^{(2)}$被装配到对应的层级2模体$\sigma_1^{(2)}$的。图3(a)和3(b)示出了另外两个层级2模体$\sigma_2^{(2)}$和$\sigma_3^{(2)}$。

[0074] 图4(a)示出了层级3网络结构$G_1^{(3)}$,并且图4(b)示出了图2(b)和图3的原语是如何基于$G_1^{(3)}$被装配以形成层级3模体$\sigma_1^{(3)}$的。

[0075] 我们现在参照图5解释根据本公开的计算机实现的方法500(其是示例)。该方法通过将神经网络体系结构表示(G,o)的实例视为基因型(genotypes),对神经网络体系结构采用演化策略。

[0076] 步骤501和502是初始化基因型种群。在501中,创建至少一个初始基因型。初始基因型是表示具有多个层级L(大于1,并且优选大于2)的分层神经网络的神经网络体系结构表示。例如,该(或每个)初始基因型可以被创建为“不重要(trivial)”基因型,这是表示神经网络的神经网络表示,其中每个层级的唯一模体是恒等映射链。因此,在初始基因型中使用的唯一原语操作是恒等原语。

[0077] 在步骤502中,通过应用多个随机突变使(多个)初始基因型多样化,以形成神经网络表示的种群。例如,从(多个)初始基因型中的单一的一个初始基因型开始,通过对(多个)初始基因型应用相应的多个突变,可以形成多个基因型(神经网络体系结构表示)。注意,这与一些已知技术相反,其中在这些已知技术中,所有的初始基因型是不重要的(trivial)网

络。多样化步骤502提供了具有重要的(non-trivial)体系结构的搜索空间的改进的初始覆盖,并且有助于避免由手工初始化例程引入的偏差。事实上,根据下面讨论的适应度测试,甚至在步骤502中生成的基因型所表示的神经网络中的一些神经网络也可以相当好地执行。

[0078] 步骤502的突变可以使用下面的动作空间来执行。单个突变由以下一系列动作构成:

[0079] 1.在2到L的范围内选择l的值。这相当于对非原语层级(“目标层级”)进行采样,其中非原语层级不是原语层级(l=1)。

[0080] 2.在目标层级中采样模体m(“目标模体”)。

[0081] 3.对目标模体中的随机节点i(“后置节点”)进行采样,其中该目标模体中的随机节点i不是该模体的输入节点。

[0082] 4.对目标模体中的随机节点j(前置节点)进行采样,其中该目标模体中的随机节点j不是节点i并且不是节点i的后置节点(即既不是直接后置节点也不是间接后置节点,即不是通过跟随多个有向边可以到达节点i的节点)。

[0083] 5.用随机采样的操作$o_{k'}^{(l-1)}$替换从j到i的当前操作$o_k^{(l-1)}$。

[0084] 在由两个层级构成的平面基因型的情况下(其中两个层级中的一个层级是固定原语的层级),第一步被省略并且l被设置为2。突变可以总结为:

[0085] $$\left[G_m^{(l)}\right]_{ij} = k' \qquad (4)$$

[0086] 其中(l,m,i,j,k')是从它们各自的域中随机采样的。注意的是,突变过程强大到足以对目标模体进行各种修改,诸如:

[0087] 1.添加新边:如果$o_k^{(l-1)} = none$且$o_{k'}^{(l-1)} \neq none$

[0088] 2.改变现有边:如果$o_k^{(l-1)} \neq none$且$o_{k'}^{(l-1)} \neq none$且$o_k^{(l-1)} \neq o_{k'}^{(l-1)}$

[0089] 3.移除现有边:如果$o_k^{(l-1)} \neq none$并且如果$o_{k'}^{(l-1)} = none$

[0090] 步骤503-508实现基于锦标赛选择的演化搜索算法。从随机基因型的初始种群开始,锦标赛选择提供了从种群中挑选至少一个有希望的基因型并将其突变的后代放回种群中的机制。通过重复此过程,种群的质量随着时间的推移而不断精确。当满足终止标准(例如,经过了固定的时间量)时在整个种群当中具有最高适应度值(例如,通过从头开始在固定数量的步骤训练包括神经网络的神经网络系统而获得的验证准确性)的基因型被选择作为最终输出(确定的神经网络体系结构)。

[0091] 具体地,在步骤503中,从当前基因型种群(例如种群的5%)中随机选择种群中多个基因型的样本(即数量K,其至少为2,优选大于2,但小于种群中基因型的数量)。

[0092] 在步骤504中,每个样本基因型(数据结构的实例)被装配以形成对应的样本神经网络,并且样本神经网络被训练。每个样本神经网络可以用随机权重(或者以变化的相等权重)初始化。因此,504不依赖于权重继承,尽管将其合并到步骤504中是简单的。该训练可以使用固定数量的训练步骤来训练每个网络。

[0093] 在步骤505中,使用适应度测试来评估经训练的网络的适应度,以导出每个样本神

经网络的适应度值。每个样本神经网络的适应度值存储在存储器中。

[0094] 如果第一次没有执行步骤504和505(见下文),使得给定的样本神经网络已经被训练和评估,则对于该样本神经网络可以省略步骤504,并且在步骤505中,可以从存储器中检索评估结果。

[0095] 在一个实施方式中,对于给定的样本基因型,步骤504和505中的训练和评估可以包括使用从样本基因型生成的神经网络作为组件来形成第一神经网络系统。从样本基因型生成的样本神经网络可以被称为第一神经网络系统的神经网络部分("单元(cell)")。例如,第一神经网络系统可以根据多个神经网络部分(所述多个神经网络部分例如以前馈方式顺序排列)的第一模板(例如,预定体系结构)生成。模板的一个或多个神经网络部分可以被实现为从基因型生成的神经网络。模板的其他神经网络部分中的一个或多个可以以一些其他方式实现,例如实现为一个或多个层(诸如卷积层、池化层和/或softmax层)的预定集合。

[0096] 在这种情况下,在步骤504中,第一神经网络系统然后可以被训练(例如,基于训练集通过标准神经网络训练算法(诸如监督学习)训练)以执行计算任务,包括训练从样本基因型生成的(多个)神经网络部分。在步骤505中,适应度测试然后可以是经训练的第一神经网络系统在执行计算任务时成功的度量。

[0097] 在步骤506中,识别(赢得锦标赛)并选择在步骤505中评估的具有最高适应度值的神经网络。在一个变型中,在这个阶段可以识别和选择多于一个的神经网络,例如预定数量的具有最高适应度值的神经网络。

[0098] 在步骤507中,确定是否满足终止标准(例如,步骤集合503-506的已经执行了至少预定次数)。如果该确定是肯定的,方法500转到步骤509(下面讨论)。

[0099] 可替换地,如果步骤507中的确定是否定的,则在步骤508中,通过以上关于步骤502描述的过程,对在步骤506中选择的(多个)神经网络从其中被装配的(多个)样本基因型进行突变。被突变的基因型(数据结构的附加实例)被添加到种群中。该过程然后返回到步骤503。因此,步骤集合503-508构成了典型地执行多次(例如至少40次)的循环。

[0100] 注意,选择压力(selection pressure)是由锦标赛大小控制的,在我们的示例中,锦标赛大小设置为种群大小的5%。在方法500中,不从种群中移除基因型,允许其随时间生长,从而保持体系结构多样性。

[0101] 方法500可以由异步分布式系统实现,该系统由负责执行随基因型的演化(步骤501-502和506-508)的单个控制器和N_w个工作者的集合构成。控制器和工作者都可以访问记录基因型种群及其适应度的共享表格存储器M、以及包含具有未知适应度的基因型(该基因型应该被评估)的数据队列Q。

[0102] 工作者负责他们的评估(步骤503、504和505)。每当存在可用的基因型,工作者就会从Q中挑选一个未经评估的基因型,将其装配成体系结构,进行训练和验证,然后在M中记录验证准确性(适应度)。

[0103] 此外,控制器通过控制当前空闲(例如,因为Q为空而不能开始执行步骤504和505)的任何工作者来执行步骤506和508。具体地,每当有工作者可用时,控制器将从M中执行基因型的锦标赛选择,随后对所选择的基因型进行突变并将其插入到Q中以进行适应度评估。

[0104] 注意,不要求同步,在体系结构演化期间,所有工作者都可以被完全占用。

[0105] 在步骤509中,基因型种群中的基因型之一被选择作为“所确定的神经网络体系结构”,并且这可替换地用于后续编程任务,其中后续编程任务可以是在步骤504和505中使用的编程任务。所确定的神经网络体系结构可以是从其中来生成神经网络的基因型,其中该神经网络在执行步骤505的所有时间中达到最高的适应度值。

[0106] 所确定的神经网络体系结构可以用于形成新的神经网络(通过对它执行合并操作),然后训练它(例如,用于新的任务)。

[0107] 在步骤509的一个可能实施方式中,对应于所确定的神经网络体系结构的神经网络被用作第二神经网络系统的一个或多个组件(“单元(cell)”)。第二神经网络系统可以使用多个神经网络部分(所述多个神经网络部分例如以前馈方式顺序排列)的第二模板(例如预定体系结构)形成。神经网络部分中的一个或多个可以被实现为从所确定的神经网络体系结构生成的神经网络。可替换地或附加地,一个或多个其他神经网络部分可以以一些其他方式实现,例如实现为一个或多个层(诸如卷积层、池化层和/或softmax层)的预定集合。

[0108] 第二神经网络系统然后可以被训练以执行计算任务,例如包括训练从所确定的神经网络体系结构生成的(多个)神经网络部分。计算任务可以与步骤504和505中使用的计算任务相同。

[0109] 优选地,第一模板和第二模板是不同的。第一模板优选定义比第二模板更小的神经网络系统。特别地,第二模板可以包括更多的神经网络部分(例如,第二神经网络系统可以包括比第一神经网络系统包括样本神经网络的实例更多的使用所确定的神经网络体系结构形成的神经网络的实例)。使用由步骤509的第二模板指定的更小的神经网络系统来执行步骤504和505,意味着必须执行多次以生成所确定的神经网络体系结构的步骤504和505可以用比只需要执行一次的步骤509的计算成本更少的计算成本来执行。

[0110] 注意,在方法500的变型中,省略了步骤503、507和508。相反,在步骤502中生成的所有基因型种群在步骤504中用于生成相应的经训练的神经网络,并且在步骤505中对所有经训练的神经网络进行评估。在步骤506,选择具有最高适应度值的经训练的神经网络作为所确定的神经网络体系结构,然后该方法进行到步骤509。方法500的这种变型的优点是它可以在整个种群上并行运行,大大减少了搜索时间。

[0111] 图6示出了用于执行方法500的系统600。它包括用于存储包括数据结构(基因型)的种群的数据的数据结构存储器601。它还包括可以执行方法500的步骤501、502和508的适应度模块602、以及可以执行方法500的步骤503-507的评估模块603。它还包括用于执行步骤509的确定的神经网络体系结构采用模块604。在步骤503-508的循环期间,工作者可以被动态地分配到适应度模块602或评估模块603。这允许并行性的优点被有效地实现。

[0112] 在使用标准学习问题的实验中,如上所述的方法和系统能够生成第二神经网络系统,该第二神经网络与传统的神经网络训练技术相比具有减少了97%以上的训练时间,同时生成了高质量的结果。因为与神经网络体系结构搜索技术相关的计算处理时间以前是巨大的,所以这是很重要的。

[0113] 具体地,在使用CIFAR-10训练集、并且使用40,000个训练图像和10,000个验证图像的图像分类任务中,使用200个P100 GPU(graphics processing unit,图形处理单元)通过方法500执行7000个步骤的演化搜索花费1.5天。使用前面段落中讨论的方法500的变型,使用200个P100 GPU搜索200个体系结构花费一个小时。相比之下,已知技术使用250个GPU

需要11天,使用450个GPU需要4天。已经发现,使用方法500的变型在一个小时内生成的神经网络系统基本上和这些已知技术一样执行图像分类任务(即,具有几乎相同的分类误差),尽管已经使用少于1%的处理器时间生成。使用方法500生成的神经网络系统以比使用方法500的变型生成的神经网络系统更低的分类误差执行分类任务。

[0114] 在本文档中,一个或多个计算机构成的系统被配置为执行特定的操作或动作意味着该系统已经在其上安装了软件、固件、硬件或它们的组合,这些软件、固件、硬件或它们的组合在操作中使得该系统执行这些操作或动作。对于被配置为执行特定操作或动作的一个或多个计算机程序,意味着所述一个或多个程序包括当由数据处理装置执行时使该装置执行这些操作或动作的指令。

[0115] 本说明书中描述的主题和功能操作的实施例可以在数字电子电路中、在有形体现的计算机软件或固件中、在包括本说明书中公开的结构及其结构等价物的计算机硬件中、或者在它们中的一个或多个的组合中实现。本说明书中描述的主题的实施例可以被实现为一个或多个计算机程序,即编码在有形的非暂时性程序载体上的计算机程序指令的一个或多个模块,用于由数据处理装置执行或控制数据处理装置的操作。可替换地或附加地,程序指令可以被编码在人工生成的传播信号上(例如,机器生成的电、光或电磁信号),其中该人工生成的传播信号被生成以编码信息,用于传输到合适的接收器设备以由数据处理装置执行。计算机存储介质可以是机器可读存储设备、机器可读存储基底、随机或串行存取存储器设备或者它们中的一个或多个的组合。然而,计算机存储介质不是传播信号。

[0116] 术语“数据处理装置”包括用于处理数据的所有种类的装置、设备和机器,例如包括可编程处理器、计算机或多个处理器或计算机。该装置可以包括专用逻辑电路,例如FPGA(field programmable gate array,现场可编程门阵列)或ASIC(application specific integrated circuit,专用集成电路)。除了硬件之外,该装置还可以包括为所讨论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、操作系统或它们中的一个或多个的组合的代码。

[0117] 计算机程序(也可以被称为或描述为程序、软件、软件应用、模块、软件模块、脚本或代码)可以以任何形式的编程语言编写,包括编译或解释语言、或声明性或过程性语言,并且它可以以任何形式部署,包括作为独立程序或作为模块、组件、子例程或适合在计算环境中使用的其他单元。计算机程序可以(但不是必须)对应于文件系统中的文件。程序可以存储在保存其他程序或数据的文件的一部分中,例如存储在标记语言文档中的一个或多个脚本,存储在专用于所讨论的程序的单个文件中,或者存储在多个协调文件(例如存储一个或多个模块、子程序或部分代码的文件)中。计算机程序可以被部署为在一个计算机或位于一个站点或分布在多个站点并通过通信网络互连的多个计算机上执行。

[0118] 在本说明书中,“引擎”或“软件引擎”是指软件实现的输入/输出系统,其提供不同于输入的输出。引擎可以是编码的功能块,诸如库、平台、软件开发工具包(“SDK”)或对象。每个引擎可以在任何适当类型的计算设备(例如服务器、移动电话、平板电脑、笔记本电脑、音乐播放器、电子书阅读器、膝上型或台式计算机、PDA、智能电话或包括一个或多个处理器和计算机可读介质的其他固定或便携式设备)上实现。另外,两个或更多个引擎可以在相同的计算设备上实现,或者在不同的计算设备上实现。

[0119] 本说明书中描述的过程和逻辑流可以由执行一个或多个计算机程序的一个或多

个可编程计算机来执行,以通过对输入数据进行操作并生成输出来执行功能。过程和逻辑流也可以由专用逻辑电路来执行,并且装置也可以被实现为专用逻辑电路,例如,FPGA(现场可编程门阵列)或ASIC(专用集成电路)。例如,过程和逻辑流可以由图形处理单元(GPU)来执行,并且装置也可以被实现为图形处理单元(GPU)。

[0120] 举例来说,适于执行计算机程序的计算机可以基于通用或专用微处理器或两者、或者任何其他类型的中央处理单元。通常,中央处理单元将从只读存储器或随机存取存储器或两者接收指令和数据。计算机的基本元件是用于执行或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多个存储设备。通常,计算机还将包括用于存储数据的一个或多个大容量存储设备,例如磁盘、磁光盘或光盘,或者可操作地耦合到一个或多个大容量存储设备,以从一个或多个大容量存储设备接收数据或向一个或多个大容量存储设备传送数据,或者两者都包括。然而,计算机不是必须具有这样的设备。此外,计算机可以嵌入到另一个设备(例如,移动电话、个人数字助理(PDA)、移动音频或视频播放器、游戏控制台、全球定位系统(Global Positioning System,GPS)接收器或便携式存储设备(例如通用串行总线(universal serial bus,USB)闪存驱动器),仅举几个示例)中。

[0121] 适用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、介质和存储设备,包括例如:半导体存储设备,例如,EPROM、EEPROM和闪存存储器设备;磁盘,例如内部硬盘或可移动磁盘;磁光盘;和CD ROM和DVD-ROM。处理器和存储器可以由专用逻辑电路来补充或并入专用逻辑电路中。

[0122] 为了提供与用户的交互,本说明书中描述的主题的实施例可以在计算机上实现,该计算机具有用于向用户显示信息的显示设备(例如CRT(cathode ray tube,阴极射线管)或LCD(liquid crystal display,液晶显示器))以及用户可以通过其向计算机提供输入的键盘和指示设备(例如,鼠标或轨迹球)。也可以使用其他类型的设备来提供与用户的交互;例如,提供给用户的反馈可以是任何形式的感觉反馈,例如视觉反馈、听觉反馈或触觉反馈;并且来自用户的输入可以以任何形式(包括声音、语音或触觉输入)接收。此外,计算机可以通过向用户使用的设备发送文档和从用户使用的设备接收文档来与用户交互;例如通过响应于从网络浏览器接收到的请求,将网页发送到用户客户端设备上的网络浏览器。

[0123] 本说明书中描述的主题的实施例可以在计算系统中实现,该计算系统包括后端组件(例如作为数据服务器),或者包括中间件组件(例如应用服务器),或者包括前端组件(例如具有图形用户界面或网络浏览器的客户端计算机,用户可以通过该图形用户界面或网络浏览器与本说明书中描述的主题的实施方式进行交互),或者包括一个或多个这样的后端组件、中间件组件或前端组件的任意组合。系统的组件可以通过任何形式或介质的数字数据通信(例如通信网络)来互连。通信网络的示例包括局域网(“LAN”)和广域网(“WAN”)(例如互联网)。

[0124] 计算系统可以包括客户端和服务器。客户端和服务器通常彼此远离,并且通常通过通信网络进行交互。客户端和服务器的关系是由运行在各自计算机上的计算机程序生成的,并且相互之间具有客户端-服务器关系。

[0125] 虽然本说明书包含许多具体的实施细节,但这些不应被解释为对任何发明或所要求保护的范围的限制,而是对特定发明的特定实施例所特有的特征的描述。本说明书中在单独实施例的上下文中描述的某些特征也可以在单个实施例中组合实现。相反,在单个实

施例的上下文中描述的各种特征也可以在多个实施例中单独或以任何合适的子组合来实现。此外,尽管特征可以在上面被描述为在某些组合中起作用,甚至最初也是这样要求保护的,但是在某些情况下,来自所要求保护的组合的一个或多个特征可以从该组合中删除,并且所要求保护的组合可以针对子组合或子组合的变型。

[0126] 类似地,尽管在附图中以特定的顺序描述了操作,但是这不应该被理解为要求以所示的特定顺序或顺序执行这些操作、或者执行所有示出的操作以获得期望的结果。在某些情况下,多任务和并行处理可以是有利的。此外,上述实施例中的各种系统模块和组件的分离不应被理解为在所有实施例中都需要这样的分离,并且应当理解,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。

[0127] 已经描述了主题的特定实施例。其他实施例在所附权利要求的范围内。例如,权利要求中列举的动作可以以不同的顺序执行,并且仍然获得期望的结果。作为一个示例,附图中描述的过程不一定需要所示的特定顺序或顺序以获得期望的结果。在某些实施方式中,多任务和并行处理可以是有利的。

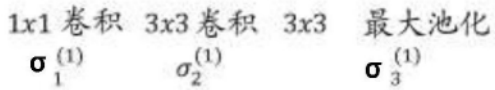

1x1 卷积
3x3 卷积
3x3
最大池化
σ (1) 1
σ (1) 2
σ (1) 3

图1

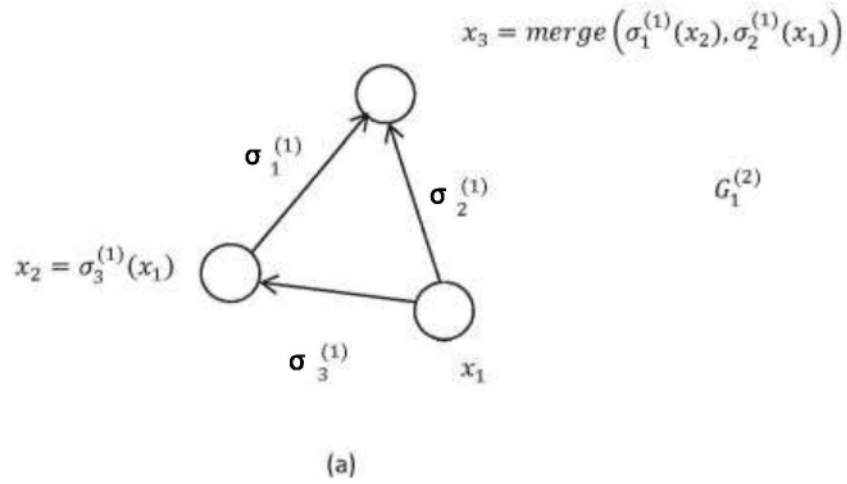

$x_3 = merge\left(\sigma_1^{(1)}(x_2), \sigma_2^{(1)}(x_1)\right)$
σ (1) 1
σ (1) 2
$G_1^{(2)}$
$x_2 = \sigma_3^{(1)}(x_1)$
σ (1) 3
x_1
(a)
1x1
3x3
σ (2) 1
池化
(b)

图2

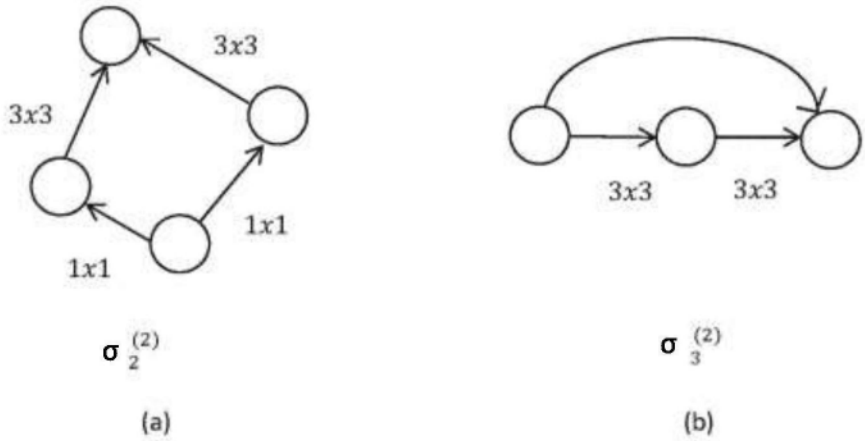


图3

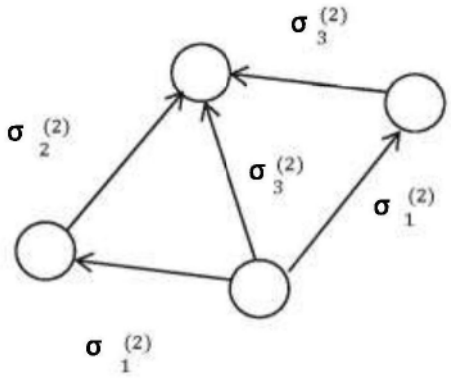


图4(a)

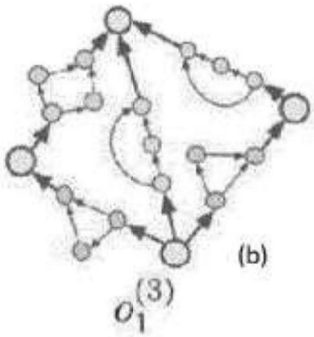


图4(b)

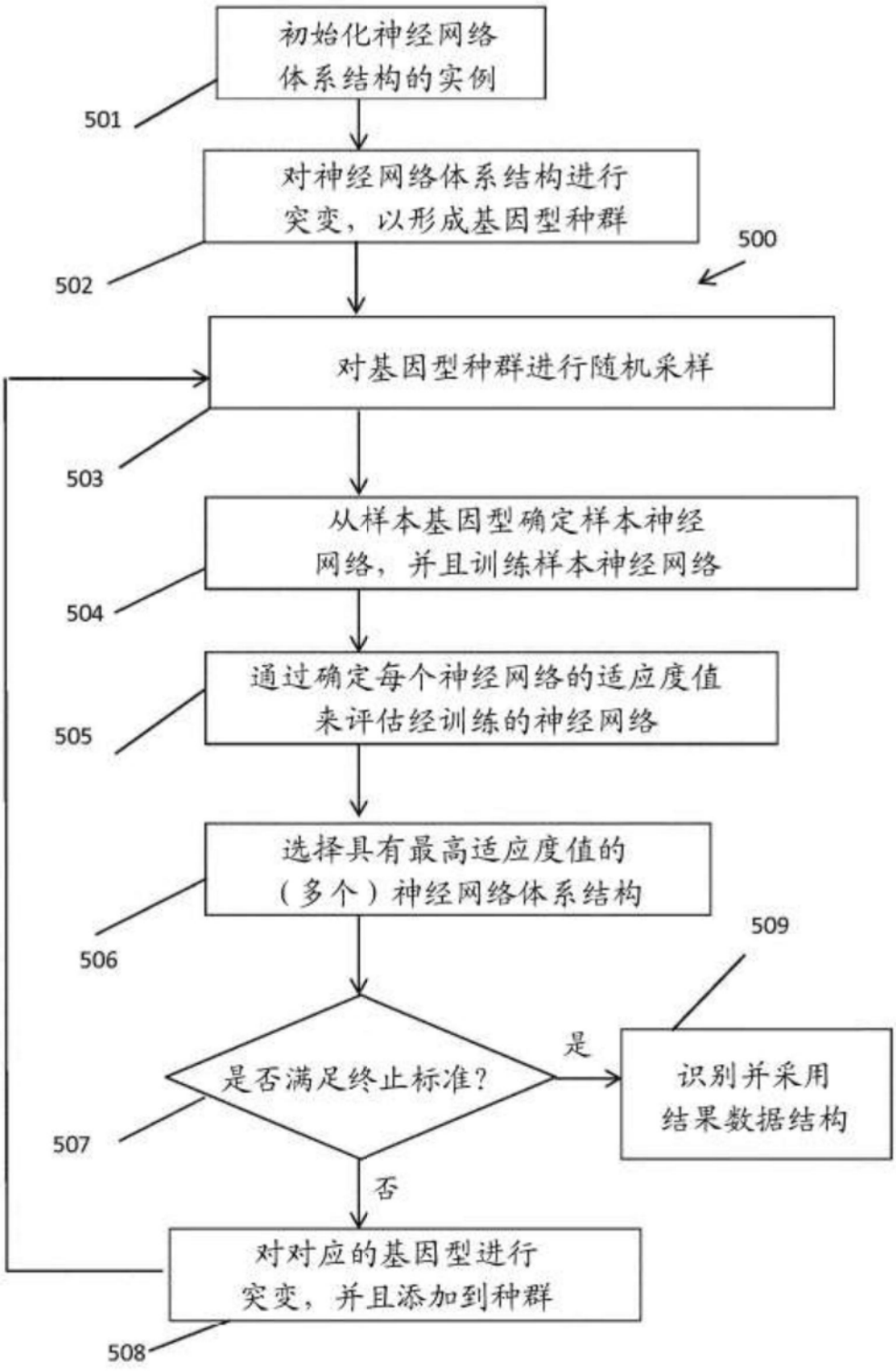

初始化神经网络
体系结构的实例
501
对神经网络体系结构进行
突变，以形成基因型种群
502
500
对基因型种群进行随机采样
503
从样本基因型确定样本神经
网络，并且训练样本神经网络
504
通过确定每个神经网络的适应度值
来评估经训练的神经网络
505
选择具有最高适应度值的
（多个）神经网络体系结构
506
509
是否满足终止标准？
507
是
识别并采用
结果数据结构
否
对对应的基因型进行
突变，并且添加到种群
508

图5

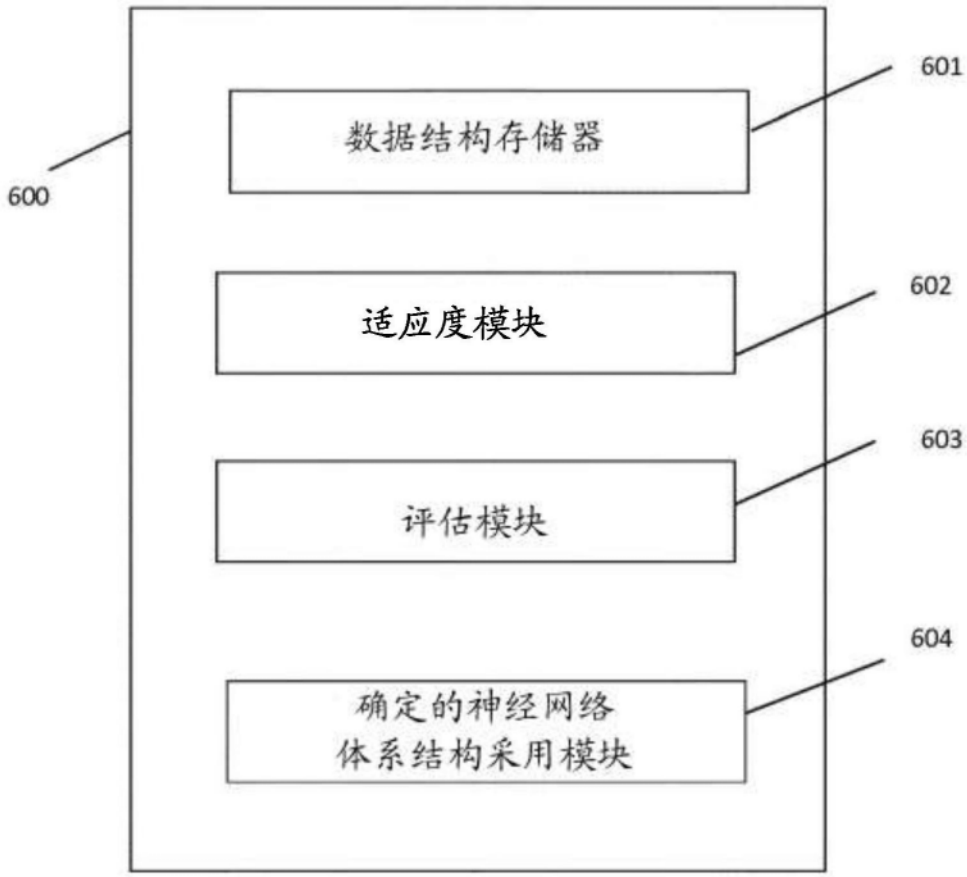


图6