



(12)发明专利申请

(10)申请公布号 CN 106548190 A

(43)申请公布日 2017. 03. 29

(21)申请号 201610711793.7

(22)申请日 2016.08.23

(30)优先权数据

10-2015-0132679 2015.09.18 KR

(71)申请人 三星电子株式会社

地址 韩国京畿道水原市

(72)发明人 姜晓娥 金荷映

(74)专利代理机构 北京铭硕知识产权代理有限公司

公司 11286

代理人 闫红玉 张川绪

(51)Int. Cl.

G06K 9/62(2006.01)

G06N 3/08(2006.01)

G09B 9/00(2006.01)

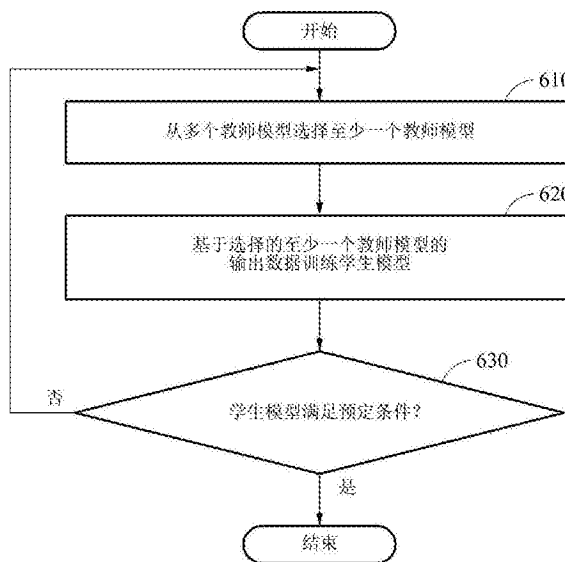
权利要求书2页 说明书12页 附图6页

(54)发明名称

模型训练方法和设备以及数据识别方法

(57)摘要

本发明公开一种模型训练方法和设备以及数据识别方法。这里公开一种模型训练方法、数据识别方法和模型训练设备。模型训练方法包括：从多个教师模型选择教师模型；在学生模型处接收输入数据；基于选择的教师模型的输出数据训练学生模型，输出数据对应于输入数据。



1. 一种模型训练方法,包括:  
从多个教师模型选择教师模型;  
在学生模型处接收输入数据;  
基于选择的教师模型的输出数据训练学生模型,所述输出数据对应于所述输入数据。
2. 如权利要求1所述的模型训练方法,其中,选择的步骤包括:基于所述多个教师模型的精度,从所述多个教师模型选择教师模型。
3. 如权利要求1所述的模型训练方法,还包括:从所述多个教师模型输出数据,  
其中,选择的步骤包括:基于所述多个教师模型的输出数据之间的相关性,从所述多个教师模型选择教师模型,输出数据对应于输入数据。
4. 如权利要求3所述的模型训练方法,其中,选择的步骤包括:响应于确定所述选择的教师模型的输出数据与另一选择的教师模型的输出数据之间的相关性小于阈值,从所述多个教师模型选择另一教师模型。
5. 如权利要求1所述的模型训练方法,其中,训练的步骤还包括:基于学生模型的输出数据训练学生模型。
6. 如权利要求1所述的模型训练方法,其中,迭代执行选择的步骤和训练的步骤,直到学生模型满足预定条件为止。
7. 如权利要求1所述的模型训练方法,其中,训练的步骤包括:基于学生模型的输出数据与所述选择的教师模型的第一输出数据之间的第一损耗,以及从学生模型的隐藏层得到的分类器层的输出数据与所述选择的教师模型的第二输出数据之间的第二损耗,来训练学生模型,第一输出数据对应于输入数据,  
所述方法还包括:  
使用不同的方法确定第一损耗和第二损耗。
8. 如权利要求7所述的模型训练方法,其中,确定第一损耗和第二损耗的步骤包括:基于从所述选择的教师模型输出的第一输出数据和从不同的选择的教师模型输出的第二输出数据,确定第一损耗和第二损耗。
9. 如权利要求7所述的模型训练方法,其中,确定第一损耗和第二损耗的步骤包括:通过向第一输出数据和第二输出数据施加不同的权重,来确定第一损耗和第二损耗。
10. 如权利要求7所述的模型训练方法,还包括:将用于分类器层的初始权重设置为用于所述多个教师模型中具有与输入到分类器层的数据的大小最接近的大小的教师模型的初始权重,并将具有与输入到分类器层的数据的大小最接近的大小的教师模型选择为所述选择的教师模型。
11. 如权利要求1所述的模型训练方法,其中,训练的步骤包括:基于所述选择的教师模型的输出数据和与输入数据对应的正确答案数据,来训练学生模型。
12. 如权利要求1所述的模型训练方法,其中,所述多个教师模型具有不同的初始权重、不同的神经网络结构、不同的超参数或包括不同的集成。
13. 如权利要求1所述的模型训练方法,还包括:基于输入到选择的一个教师模型的数据的大小,确定学生模型的结构。
14. 一种数据识别方法,包括:  
接收将被识别的目标数据;

使用训练的模型识别目标数据，

其中，基于教师模型的输出数据训练所述模型，输出数据对应于由所述模型接收的输入数据，从多个教师模型选择教师模型。

15. 如权利要求14所述的数据识别方法，还包括：基于所述多个教师模型的精度或所述多个教师模型的输出数据之间的相关性选择教师模型，输出数据对应于输入数据。

16. 如权利要求14所述的数据识别方法，还包括：基于所述模型的输出数据与选择的教师模型的第一输出数据之间的第一损耗，以及从所述模型的隐藏层得到的分类器层的输出数据与选择的教师模型的第二输出数据之间的第二损耗，来训练所述模型，第一输出数据对应于由所述模型接收的输入数据，

使用不同的方法确定第一损耗和第二损耗。

17. 一种模型训练设备，包括：

处理器，被配置为训练学生模型；

存储器，被配置为存储训练的学生模型，

其中，处理器被配置为从多个教师模型选择教师模型，并基于选择的教师模型的输出数据来训练学生模型，输出数据对应于由学生模型接收的输入数据。

18. 如权利要求17所述的模型训练设备，其中，处理器被配置为基于所述多个教师模型的精度或所述多个教师模型的输出数据之间的相关性，从所述多个教师模型选择教师模型，输出数据对应于输入数据。

19. 如权利要求17所述的模型训练设备，其中，处理器被配置为基于学生模型的输出数据与选择的教师模型的第一输出数据之间的第一损耗，以及从学生模型的隐藏层得到的分类器层的输出数据与选择的教师模型的第二输出数据之间的第二损耗，来训练学生模型，第一输出数据对应于输入数据，

处理器使用不同的方法确定第一损耗和第二损耗。

20. 一种模型训练器，包括：

处理器，被配置为从包括三个或更多的教师模型的多个教师模型选择两个教师模型，使用选择的两个教师模型训练学生模型，并输出训练的学生模型；

存储器，被配置为存储训练的学生模型，

其中，处理器被配置为基于与预定阈值相关的两个教师模型的精度，选择两个教师模型。

21. 如权利要求20所述的模型训练器，其中，处理器被配置为按精度的递减顺序选择具有高于所述预定阈值的精度的两个教师模型。

## 模型训练方法和设备以及数据识别方法

[0001] 本申请要求于2015年9月18日提交到韩国知识产权局的第10-2015-0132679号韩国专利申请的优先权,所述韩国专利申请的公开通过引用全部包含于此。

### 技术领域

[0002] 与这里公开的示例性实施例一致的方法和设备涉及一种模型训练方法和设备以及数据识别方法。

### 背景技术

[0003] 最近,为了将输入模式分类为属于预定组的成员,研究者正在积极地研究将由人执行的有效和精确的模式识别技术应用于计算机的方法。一个这样的研究领域聚焦于通过数学表达式对人类的生物神经细胞的特征进行建模的人工神经网络。为了将输入模式分类为预定组,神经网络采用模拟人类大脑的学习能力的算法。通过该算法,神经网络可生成输入模式与输出模式之间的映射。也可将生成这样映射的能力称为神经网络的“学习能力”。此外,神经网络可具有基于学习的结果生成针对将被用于学习的输入模式的相对准确的输出的泛化能力。

[0004] 此外,正在进行使人工神经网络小型化的同时最小化识别速率的降低的研究。

### 发明内容

[0005] 示例性实施例可至少解决以上提到的问题和/或缺点以及以上未描述的其他缺点。此外,示例性实施例无需克服上述的缺点,并且示例性实施例可不需要克服上述问题中的任何一个问题。

[0006] 根据示例性实施例的一方面,提供一种模型训练方法,包括:从多个教师模型选择教师模型;在学生模型接收输入数据;基于选择的教师模型的输出数据训练学生模型,输出数据对应于输入数据。

[0007] 选择的步骤可包括:基于所述多个教师模型的精度,从所述多个教师模型选择教师模型。

[0008] 所述模型训练方法还可包括:从所述多个教师模型输出数据,选择的步骤可包括:基于所述多个教师模型的输出数据之间的相关性,从所述多个教师模型选择教师模型,输出数据对应于输入数据。

[0009] 选择的步骤可包括:响应于确定所述选择的教师模型的输出数据与另一选择的教师模型的输出数据之间的相关性小于阈值,从所述多个教师模型选择另一教师模型。

[0010] 训练的步骤还可包括:基于学生模型的输出数据训练学生模型。

[0011] 可迭代执行选择的步骤和训练的步骤,直到学生模型满足预定条件为止。

[0012] 训练的步骤可包括基于学生模型的输出数据与所述选择的教师模型的第一输出数据之间的第一损耗,以及从学生模型的隐藏层得到的分类器层的输出数据与所述选择的教师模型的第二输出数据之间的第二损耗,来训练学生模型,第一输出数据对应于输入数

据,所述方法还可包括:使用不同的方法确定第一损耗和第二损耗。

[0013] 确定第一损耗和第二损耗的步骤可包括:基于从所述选择的教师模型输出的第一输出数据和从不同的选择的教师模型输出的第二输出数据,来确定第一损耗和第二损耗。

[0014] 确定第一损耗和第二损耗的步骤可包括:通过向第一输出数据和第二输出数据施加不同的权重,来确定第一损耗和第二损耗。

[0015] 所述模型训练方法还可包括:将用于分类器层的初始权重设置为用于多个教师模型中具有与输入到分类器层的数据的大小最接近的大小的教师模型的初始权重,并将具有与输入到分类器层的数据的大小最接近的大小的教师模型选择为所述选择的教师模型。

[0016] 训练的步骤可包括:基于所述选择的教师模型的输出数据和与输入数据对应的正确答案数据,来训练学生模型。

[0017] 多个教师模型可具有不同的初始权重、不同的神经网络结构、不同的超参数或包括不同的集成。

[0018] 所述模型训练方法还可包括:基于输入到选择的一个教师模型的数据的大小,确定学生模型的结构。

[0019] 根据另一个示例性实施例的一方面,提供一种数据识别方法,包括:接收将被识别的目标数据;使用训练的模型识别目标数据,其中,基于教师模型的输出数据训练所述模型,输出数据对应于由所述模型接收的输入数据,从多个教师模型选择教师模型。

[0020] 所述数据识别方法还可包括:基于所述多个教师模型的精度或所述多个教师模型的输出数据之间的相关性选择教师模型,输出数据对应于输入数据。

[0021] 所述数据识别方法还可包括:基于所述模型的输出数据与选择的模型的第一输出数据之间的第一损耗,以及从所述模型的隐藏层得到的分类器层的输出数据与选择的教师模型的第二输出数据之间的第二损耗,来训练所述模型,第一输出数据对应于由所述模型接收的输入数据,使用不同的方法确定第一损耗和第二损耗。

[0022] 根据另一个示例性实施例的一方面,提供一种非暂时性计算机可读记录介质,当运行时,使计算机执行模型训练方法,模型训练方法包括:从多个教师模型选择教师模型;在学生模型处接收输入数据;并基于选择的教师模型的输出数据训练学生模型,输出数据对应于输入数据。

[0023] 根据另一个示例性实施例的一方面,提供一种模型训练设备,包括:处理器,被配置为训练学生模型;存储器,被配置为存储训练的学生模型,其中,处理器被配置为从多个教师模型选择教师模型,并基于选择的教师模型的输出数据来训练学生模型,输出数据对应于由学生模型接收的输入数据。

[0024] 处理器可被配置为基于所述多个教师模型的精度或所述多个教师模型的输出数据之间的相关性,从所述多个教师模型选择教师模型,输出数据对应于输入数据。

[0025] 处理器可被配置为基于学生模型的输出数据与选择的教师模型的第一输出数据之间的第一损耗,以及从学生模型的隐藏层得到的分类器层的输出数据与选择的教师模型的第二输出数据之间的第二损耗,来训练学生模型,第一输出数据对应于输入数据,而且处理器可使用不同的方法确定第一损耗和第二损耗。

## 附图说明

[0026] 通过下面结合附图对特定示例性实施例进行的详细描述,示例性实施例的上面和其他方面将会变得清楚和更容易理解,其中:

[0027] 图1是示出根据示例性实施例的教师模型和学生模型的示图;

[0028] 图2、图3和图4是示出根据示例性实施例的从多个教师模型选择至少一个教师模型并训练学生模型的处理的示图;

[0029] 图5是示出根据示例性实施例的使用学生模型的分层器的训练处理的示图;

[0030] 图6是示出根据示例性实施例的模型训练方法的流程图;

[0031] 图7是示出根据示例性实施例的数据识别方法的流程图;

[0032] 图8是示出根据示例性实施例的模型训练设备的框图;

[0033] 图9是示出根据示例性实施例的数据识别设备的框图。

### 具体实施方式

[0034] 现在将对示例性实施例进行详细描述,所述示例性实施例的示例被示出在附图中,其中,相同的参考标号始终表示相同的元件。

[0035] 以下,将参照附图对示例性实施例进行详细地描述。下面具体结构或功能性描述是示例性的并仅描述示例性实施例的特定特征,而且示例性实施例的范围不被限制到在本说明书中提供的描述。可由本领域的普通技术人员对示例性实施例进行各种改变和修改。在这里可不描述公知的功能或配置。

[0036] 图1是示出根据示例性实施例的教师模型和学生模型的示图。

[0037] 参照图1,示出教师模型110和学生模型120。教师模型110和学生模型120是被训练以针对预定输入输出预定输出的模型,并可包括例如神经网络。神经网络指使用通过边彼此连接的大量人工神经元模拟生物系统的计算能力的识别模型。然而,应该理解,教师模型110和学生模型120不限于神经网络,并也可在其他类型的网络和设备中实现。

[0038] 神经网络使用通过将生物神经元的功能进行简化来配置的人工神经元,而且人工神经元可通过具有连接权重的边来彼此连接。连接权重(神经网络的参数)是边的预定值,并且可被称为连接强度。神经网络可通过人工神经元执行人类大脑的认知功能或学习处理。也可将人工神经元称为节点。

[0039] 神经网络可包括多个层。例如,神经网络可包括输入层、隐藏层和输出层。输入层可接收将被用于执行训练的输入并将输入传输到隐藏层,输出层可基于从隐藏层的节点接收的信号产生神经网络的输出。隐藏层可布置在输入层与输出层之间。隐藏层可将从输入层接收的训练数据改变为可容易预测的值。包括在输入层和隐藏层中的节点可通过具有连接权重的边彼此连接,并且包括在隐藏层和输出层中的节点也可通过具有连接权重的边彼此连接。输入层、隐藏层和输出层可分别包括多个节点。

[0040] 神经网络可包括多个隐藏层。可将包括多个隐藏层的神经网络称为深度神经网络。可将训练深度神经网络称为深度学习。可将包括在隐藏层中的节点称为隐藏节点。在深度神经网络中设置的隐藏层的数量不限于任何特定数量。

[0041] 可通过监督学习训练神经网络。监督学习指这样的方法:将输入数据和与输入数据对应的输出数据提供给神经网络,并更新边的连接权重从而可输出与输入数据对应的输出数据。例如,模型训练设备可通过德尔塔定律和误差反向传播学习来更新人工神经元中

的边的连接权重。

[0042] 误差反向传播学习指这样的方法：通过正演计算估计针对提供的输入数据的损耗，并更新连接权重以在沿从输出层向隐藏层和输入层的反向方向传播估计的损耗的处理中降低损耗。可按输入层、隐藏层和输出层的顺序来执行神经网络的处理。然而，在误差反向传播学习中，可按输出层、隐藏层和输入层的顺序来更新连接权重。以下，根据示例性实施例，训练神经网络指训练神经网络的参数。此外，训练的神经网络指应用了训练的参数的神经网络。

[0043] 教师模型110和学生模型120可以是配置为识别相同目标的不同大小的神经网络。然而，应该理解，不要求教师模型110和学生模型120是不同的大小。

[0044] 教师模型110是基于从将被识别的目标数据提取的足够大的数量的特征来相对高准确性地识别目标数据的模型。教师模型110可以是大小大于学生模型120的神经网络。例如，与学生模型120相比，教师模型110可包括更大数量的隐藏层、更大数量的节点或他们的组合。

[0045] 学生模型120可以是大小小于教师模型110的神经网络。由于相对小的大小，所以学生模型120可具有比教师模型110更高的识别速率。可使用教师模型110训练学生模型120以针对输入数据提供教师模型110的输出数据。例如，教师模型110的输出数据可以从教师模型110输出的逻辑值、概率值或从教师模型110的隐藏层得到的分类器层的输出值。因此，可获得在输出与从教师模型110输出的值相同的值时具有比教师模型110高的识别速率的学生模型120。可将前述处理称为模型压缩。模型压缩是一种基于教师模型110的输出数据来训练学生模型120来代替基于与真实标签对应的正确答案数据来训练学生模型120的方案。

[0046] 可将多个教师模型110用于训练学生模型120。可从多个教师模型110选择至少一个教师模型并可使用选择的至少一个教师模型来训练学生模型120。可迭代执行从多个教师模型110选择至少一个教师模型并训练学生模型120的处理，直到学生模型120满足预定条件为止。在该示例中，可在每次执行训练处理时重新选择被选择用来训练学生模型120的至少一个教师模型。例如，可选择一个或多个教师模型来用于训练学生模型120。

[0047] 下面将参照图2、图3和图4来描述从多个教师模型110选择一个或多个教师模型并训练学生模型120的处理。

[0048] 图2、图3和图4是示出根据示例性实施例的从多个教师模型选择至少一个教师模型并训练学生模型的处理的示图。

[0049] 参照图2，示出从多个教师模型选择一个教师模型来训练学生模型220的处理。可由模型训练设备执行从多个教师模型选择一个教师模型的处理。

[0050] 根据示例性实施例，模型训练设备是被配置为训练用于数据识别的神经网络的装置，并可使用被专门配置为执行与训练用于数据识别的神经网络相关的特定功能的单个处理器或多个处理器来实现。模型训练设备也可使用包括在不同装置中的多个模块来实现。在该示例中，例如，多个模块可通过网络彼此连接。

[0051] 多个教师模型是将被用于训练学生模型220的预先训练的模型，并可具有各种结构和各种精度。多个教师模型可包括第一教师模型至第N教师模型（例如，第一教师模型205、第二教师模型210、第三教师模型215、第N教师模型225）。

[0052] 如在下面详细地描述,多个教师模型可具有不同的初始权重,具有不同的神经网络结构,具有不同的超参数,或包括不同的集成(ensemble,系综)。

[0053] 多个教师模型可具有不同的初始权重。初始权重是神经网络中的边的连接权重的初始值。初始权重可极大地影响误差反向传播神经网络训练中的学习速率和收敛速率。可使用诸如例如随机初始化和预先训练的各种方法将被施加到多个教师模型的初始权重设置为不同的值。

[0054] 多个教师模型可具有不同的神经网络结构。例如,多个教师模型可具有包括不同数量的隐藏层、不同数量的滤波器和不同大小的核,连同不同形状、大小以及网络配置的各种神经网络结构。

[0055] 多个教师模型可包括不同的超参数。超参数可指诸如例如学习速率和动量的训练相关参数。

[0056] 多个教师模型可包括不同的集成。教师模型可包括单个神经网络,或多个神经网络的集成。当教师模型包括多个神经网络的集成时,相应教师模型的集成可与另一个教师模型的集成不同。

[0057] 模型训练设备可基于多个教师模型的精度从多个教师模型选择一个教师模型。例如,模型训练设备可从多个教师模型选择具有最高精度的教师模型。模型训练设备可随机或根据预定的顺序选择具有大于或等于阈值的精度的教师模型中的一个教师模型。

[0058] 模型训练设备可使用选择的教师模型210的输出数据作为学生模型220的标签来训练学生模型220。模型训练设备可基于指示针对输入数据的学生模型220的输出数据与选择的教师模型210的输出数据之间的差的损耗,来训练学生模型220。模型训练设备可计算损耗,并基于随机梯度下降(SGD)来训练学生模型220以降低损耗。然而,示例性实施例不限于基于SGD降低损耗,也可使用其他技术(例如,其他梯度下降方法)来代替SGD或与SGD组合。

[0059] 模型训练设备可沿从输出层向隐藏层和输入层的反向方向传播损耗,并更新连接权重以降低损耗。可将反方向的误差传播称为逆推法。

[0060] 模型训练设备可定义用于测量当前设置的连接权重的最优性的目标函数,基于目标函数的结果持续地改变连接权重,并迭代地执行训练。例如,目标函数可以是由神经网络使用以计算针对输入数据的实际输出数据与期望输出的值(例如,选择的教师模型210的输出数据)之间的损耗的损耗函数。模型训练设备可更新连接权重以降低损耗函数的值。

[0061] 如等式1所示,模型训练设备可计算损耗。

[0062] [等式1]

[0063]  $Loss(\theta) = f(P_{Teacher(i)}, P_{Student})$

[0064] 在等式1中, $Loss(\theta)$ 表示损耗函数, $\theta$ 表示将由学生模型220学习的参数。 $P_{Teacher(i)}$ 表示选择的第*i*教师模型的输出数据, $P_{Student}$ 表示针对输入数据的学生模型220的输出数据, $f$ 表示交叉熵、softmax函数或者选择的第*i*教师模型的输出数据 $P_{Teacher(i)}$ 与学生模型220的输出数据 $P_{Student}$ 之间的欧几里得距离。

[0065] 图2示出从第一教师模型至第*N*教师模型选择第二教师模型210并基于表示选择的第二教师模型的输出数据的 $P_{Teacher(2)}$ 来训练学生模型220的示例。

[0066] 在另一示例中,模型训练设备还可基于正确答案数据 $T_r$ 来训练学生模型220。如等

式2所示,模型训练设备还可基于正确答案数据 $T_r$ 来计算损耗。

[0067] [公式2]

$$[0068] \text{Loss}(\theta) = \alpha f(P_{\text{Teacher}(i)}, P_{\text{Student}}) + \beta g(T_r, P_{\text{Student}})$$

[0069] 在公式2中, $g$ 表示交叉熵,softmax函数或者正确答案数据 $T_r$ 与学生模型220的输出数据 $P_{\text{Student}}$ 之间的欧几里得距离。 $\alpha$ 是表示施加到选择的第 $i$ 教师模型的输出数据 $P_{\text{Teacher}(i)}$ 的权重的常数, $\beta$ 是表示施加到正确答案数据 $T_r$ 的权重的常数。

[0070] 模型训练设备可通过调节 $\alpha$ 和 $\beta$ 的值,来确定选择的第 $i$ 教师模型的输出数据 $P_{\text{Teacher}(i)}$ 和正确答案数据 $T_r$ 在训练学生模型220的处理中的影响。例如,在 $\alpha$ 的值被设置为大于 $\beta$ 的值的条件下,相对于正确答案数据 $T_r$ ,模型训练设备可更加关注选择的第 $i$ 教师模型的输出数据 $P_{\text{Teacher}(i)}$ ,来训练学生模型220。

[0071] 参照图3,示出从多个教师模型选择至少两个教师模型来训练学生模型320的处理。可由模型训练设备执行从多个教师模型选择至少两个教师模型的处理。多个教师模型是将被用于训练学生模型320的预先训练的模型,并可具有各种结构和各种精度。

[0072] 模型训练设备可基于多个教师模型的精度从多个教师模型选择至少两个教师模型。模型训练设备可从多个教师模型选择具有大于或等于阈值的精度的教师模型。模型训练设备可基于多个教师模型的精度排列多个教师模型,并按精度的递减顺序或另一顺序选择预定数量的教师模型(例如, $k$ 个教师模型)。

[0073] 模型训练设备可基于针对输入数据的多个教师模型的输出数据之间的相关性,从多个教师模型选择至少两个教师模型。当选择的至少两个教师模型310的输出数据之间的相关性小于阈值时,模型训练设备可从包括第一教师模型305至第 $N$ 教师模型325的多个教师模型选择包括第二教师模型和第三教师模型的至少两个教师模型310。例如,当选择的教师模型310的输出数据之间的相关性小于第二阈值时,模型训练设备可从具有大于或等于第一阈值的精度的教师模型选择至少两个教师模型310。在另一示例中,模型训练设备可按与具有最高精度的教师模型的输出数据的相关性的递升的顺序,从具有大于或等于阈值的精度的教师模型选择具有最高精度的教师模型和预定数量的教师模型(例如, $k$ 个教师模型)。

[0074] 模型训练设备可从用户接收从多个教师模型试探性地选择至少两个教师模型的输入。

[0075] 图3示出从第一教师模型305至第 $N$ 教师模型325选择第二教师模型和第三教师模型,并基于表示选择的第二教师模型的输出数据的 $P_{\text{Teacher}(2)}$ 和表示选择的第三教师模型的输出数据的 $P_{\text{Teacher}(3)}$ 来训练学生模型320的示例。

[0076] 模型训练设备还可基于正确答案数据 $T_r$ 训练学生模型320。在从多个教师模型选择至少两个教师模型,并还基于正确答案数据 $T_r$ 训练学生模型320的情况下,模型训练设备可如等式3所示计算损耗。

[0077] [等式3]

$$[0078] \text{Loss}(\theta) = \alpha f(P_{\text{Teacher}(i)}, P_{\text{Student}}) + \beta g(P_{\text{Teacher}(j)}, P_{\text{Student}})$$

$$[0079] + \gamma h(T_r, P_{\text{Student}})$$

[0080] 在等式3中, $P_{\text{Teacher}(j)}$ 表示选择的第 $j$ 教师模型的输出数据, $g$ 表示交叉熵、softmax函数或者选择的第 $j$ 教师模型的输出数据 $P_{\text{Teacher}(j)}$ 与学生模型320的输出数据 $P_{\text{Student}}$ 之间的

欧几里得距离。 $h$ 表示交叉熵、softmax函数或者正确答案数据 $T_r$ 与学生模型320的输出数据 $P_{Student}$ 之间的欧几里得距离。 $\beta$ 是表示施加到选择的第 $j$ 教师模型的输出数据 $P_{Teacher(j)}$ 的权重的常数, $\gamma$ 是表示施加到正确答案数据 $T_r$ 的权重的常数。模型训练设备可通过调节 $\alpha$ 、 $\beta$ 和 $\gamma$ 的值,来确定选择的第 $i$ 教师模型的输出数据 $P_{Teacher(i)}$ 、选择的第 $j$ 教师模型的输出数据 $P_{Teacher(j)}$ 和正确答案数据 $T_r$ 的影响。

[0081] 模型训练设备可训练学生模型320以降低通过等式3计算的损耗。

[0082] 参照图4,示出还基于学生模型420的输出数据训练学生模型420的处理。模型训练设备可基于从包括第一教师模型405至第 $N$ 教师模型425的多个教师模型选择的至少一个教师模型410的输出数据和学生模型420的输出数据来计算损耗,并训练学生模型420以降低损耗。

[0083] 在某些情况下,基于选择的至少一个教师模型410训练的学生模型420可具有比选择的至少一个教师模型410更高的精度。因此,通过还基于具有相对高的精度的学生模型420的输出数据训练学生模型420,模型训练设备可增加学生模型420的学习速率或精度。

[0084] 模型训练设备可测量学生模型420的精度,并将学生模型420的精度与多个教师模型的精度进行比较,从而确定学生模型420的输出数据是否将被用于训练。

[0085] 例如,当学生模型420的精度高于多个教师模型的精度中的最高精度时,模型训练设备可确定学生模型420的输出数据将被用于训练。在另一示例中,模型训练设备可通过将学生模型420的精度与多个教师模型的精度的统计值进行比较,来确定学生模型420的输出数据是否将被用于训练。例如,统计值可以是精度的平均值或精度的前 $k\%$ 的值。此外,根据各种示例性实施例的设计,可使用各种标准来确定学生模型420的输出数据是否将被用于训练。

[0086] 图4示出基于选择的第二教师模型的输出数据 $P_{Teacher(2)}$ 、选择的第三教师模型的输出数据的 $P_{Teacher(3)}$ 和学生模型420的输出数据 $P_{Student}$ 训练学生模型420的示例。

[0087] 图5是示出根据示例性实施例的使用学生模型的分层器的训练处理的示图。

[0088] 参照图5,学生模型500可包括输入层510、隐藏层520、输出层530和分类器层540。可基于输入数据550和输出数据560训练学生模型500。在该示例中,输出数据560可对应于针对输入数据550从至少一个教师模型输出的数据,所述至少一个教师模型从多个教师模型被选择。例如,可根据上面有关图2至图4描述的技术选择所述至少一个教师模型。

[0089] 虽然参照图2至图4描述了使用输出层训练学生模型的处理,但是根据示例性实施例的模型训练设备可使用输出层和从隐藏层得到的分类器层来训练学生模型。

[0090] 在隐藏层520中,可通过预定水平对神经网络的隐藏节点进行分类。隐藏层520可布置在输入层510与输出层530之间。例如,隐藏层520可以是卷积神经网络(NCC)中的完全连接层或卷积滤波器,或由预定功能或特征分类的各种类型的层或滤波器。隐藏层520可包括第一隐藏层520-1、第二隐藏层520-2、第 $i$ 隐藏层520- $i$ 至第 $n$ 隐藏层520- $n$ 。

[0091] 分类器层540可以从隐藏层520派生出的层。与输出层530类似,分类器层540可通过分析从派生出分类器层540的隐藏层520接收的值,来提供与预定元素对应的输出数据。分类器层可包括第一分类器层540-1至第 $j$ 分类器层540- $j$ 。以下,为了便于描述,将描述使用第 $j$ 分类器层540- $j$ 训练学生模型500的处理。然而,描述也可适用于剩余的分类器层。

[0092] 模型训练设备还可基于第 $j$ 分类器层540- $j$ 的输出数据与选择的教师模型的输出

数据560之间的损耗来训练学生模型500。可使用等式1至等式3来计算第j分类器层540-j的输出数据与选择的教师模型的输出数据560之间的损耗。

[0093] 可通过反向传播将计算的损耗反向传播到第i隐藏层520-i。第i隐藏层520-i可基于从得到的第j分类器层540-j接收的损耗和从上面的第i+1隐藏层接收的损耗更新连接权重,将两个损耗相加,并将两个损耗的和传输到下面的第i-1隐藏层。

[0094] 当额外使用第j分类器层540-j来训练学生模型500时,可使用不同的方法计算输出层530的损耗和第j分类器层540-j的损耗。以下,将输出层530的损耗称为第一损耗,并将第j分类器层540-j的损耗称为第二损耗。此外,也可将将被用于计算第一损耗的选择的教师模型的输出数据称为第一输出数据,也可将将被用于计算第二损耗的选择的教师模型的输出数据称为第二输出数据。

[0095] 模型训练设备可使用不同的方法计算第一损耗和第二损耗。模型训练设备可设置不同的选择的教师模型以输出第一输出数据和第二输出数据。例如,模型训练设备可基于从选择的第一教师模型输出的第一输出数据来计算第一损耗,并基于从选择的第三教师模型输出的第二输出数据来计算第二损耗。

[0096] 虽然第一输出数据和第二输出数据被从相同的选择的教师模型输出,但是模型训练设备可确定将被施加到第一输出数据和第二输出数据的不同的权重。例如,在选择第一教师模型和第三教师模型而且将计算第一损耗和第二损耗的情况下,模型训练设备可通过在计算第一损耗时向第一教师模型的输出数据设置更大权重以及在计算第二损耗时向第三教师模型的输出数据设置更大权重,来计算第一损耗和第二损耗。此外,模型训练设备可通过向输出层530和分类器层540中更靠近输出层530的层针对正确答案数据设置更大的权重,来计算损耗。应该理解,示例性实施例不限于以上示例性技术,可使用很多不同的设置权重的方法来计算第一损耗和第二损耗。

[0097] 模型训练设备可确定将被施加到输出层530和第j分类器层540-j的不同的初始权重。模型训练设备可将第j分类器层540-j的初始权重设置为具有与输入到第j分类器层540-j的数据的大小最相似的大小的教师模型的初始权重,所述教师模型从选择的至少一个教师模型被选择。类似地,模型训练设备可将输出层530的初始权重设置为具有与输入到输出层530的数据的大小最相似的大小的教师模型的初始权重,所述教师模型从选择的至少一个教师模型被选择。例如,当输入到第j分类器层540-j的例如输入特征映射的数据的大小是128时,模型训练设备可将具有最相似的输入大小的教师模型的初始权重设置为第j分类器层540-j的初始权重,所述教师模型从选择的至少一个教师模型被选择。

[0098] 例如,可如等式4所示计算第j分类器层540-j的损耗。

[0099] [等式4]

[0100]  $Loss_j(\theta) = \alpha f_j(P_{Student\_j}, P_{Teacher(i)})$

[0101]  $+ \beta g_j(P_{Student\_j}, P_{Teacher(1)}) + \gamma h_j(P_{Student\_j}, T_r)$

[0102] 在等式4中, $Loss_j$ 表示计算第j分类器层540-j的损耗的损耗函数, $P_{Teacher(1)}$ 表示选择的第1教师模型的输出数据, $P_{Student\_j}$ 表示针对输入数据550的第j分类器层540-j的输出数据。 $f_j$ 表示熵、softmax函数或者第j分类器层540-j的输出数据 $P_{Student\_j}$ 与选择的第i教师模型的输出数据 $P_{Teacher(i)}$ 之间的欧几里得距离, $g_j$ 表示熵、softmax函数或者第j分类器层540-j的输出数据 $P_{Student\_j}$ 与选择的第1教师模型的输出数据 $P_{Teacher(1)}$ 之间的欧几里得距

离,  $h_j$ 表示熵、softmax函数或者第j分类器层540-j的输出数据 $P_{Student\_j}$ 与正确答案数据 $T_r$ 之间的欧几里得距离。 $\alpha$ 、 $\beta$ 和 $\gamma$ 分别是表示将被施加到选择的第i教师模型的输出数据 $P_{Teacher(i)}$ 的权重的常数、被施加到选择的第1教师模型的输出数据 $P_{Teacher(1)}$ 的权重的常数和被施加到正确答案数据 $T_r$ 的权重的常数。

[0103] 以上描述了使用输出层530和第j分类器层540-j训练学生模型500的处理。然而,上述描述也可适用于包括在分类器层540中的剩余分类器层。

[0104] 具体地,模型训练设备可使用不同的方法来计算多个分类器层的损耗。例如,模型训练设备可基于不同的选择的教师模型的输出数据来计算多个分类器层的损耗。此外,模型训练设备可通过设置将被施加到选择的教师模型的输出数据的不同权重来计算多个分类器层的损耗。此外,模型训练设备可通过向多个分类器层施加不同的初始权重来计算多个分类器层的损耗。

[0105] 图6是示出根据示例性实施例的模型训练方法的流程图。

[0106] 可由包括在模型训练设备中的处理器执行模型训练方法。可将处理器专门配置为执行在图6中示出的操作。

[0107] 参照图6,在操作610中,模型训练设备从多个教师模型选择至少一个教师模型。多个教师模型可具有不同的初始权重,可具有不同的神经网络结构,可具有不同的超参数,或可包括不同的集成。

[0108] 模型训练设备可基于多个教师模型的精度从多个教师模型选择至少一个教师模型。例如,模型训练设备可从多个教师模型选择具有最高精度的教师模型,或从多个教师模型选择具有大于或等于阈值的精度的教师模型。此外,模型训练设备可基于多个教师模型的精度来排列多个教师模型,并按精度的递减顺序或根据另一预定顺序选择预定数量的教师模型(例如,k个教师模型)。

[0109] 模型训练设备可基于针对输入数据的多个教师模型的输出数据之间的相关性,从多个教师模型选择至少两个教师模型。此外,当选择的教师模型的输出数据之间的相关性小于阈值时,模型训练设备可从多个教师模型选择至少两个教师模型。

[0110] 模型训练设备可从用户接收从多个教师模型试探性地选择至少一个教师模型的输入。然而,应该理解,示例性实施例不限于试探性地选择至少一个教师模型的输入,也可根据示例性实施例使用其他类型的选择。

[0111] 在操作620中,模型训练设备可基于针对输入数据的选择的教师模型的输出数据来训练学生模型。模型训练设备可基于针对输入数据的学生模型的输出数据与选择的教师模型的输出数据之间的损耗来训练学生模型。

[0112] 模型训练设备还可基于学生模型的输出数据来训练学生模型。此外,模型训练设备还可基于与输入数据对应的正确答案数据来训练学生模型。

[0113] 模型训练设备可基于针对输入数据的学生模型的输出数据与选择的至少一个教师模型的第一输出数据之间的第一损耗,和从学生模型的隐藏层得到的分类器层的输出数据与选择的至少一个教师模型的第二输出数据之间的第二损耗,来训练学生模型。

[0114] 根据示例性实施例,可使用不同的方法确定第一损耗和第二损耗。可基于从不同的选择的教师模型输出的第一输出数据和第二输出数据来确定第一损耗和第二损耗,或通过向第一输出数据和第二输出数据施加不同的权重来确定第一损耗和第二损耗。可将分类

器层的初始权重设置为具有与输入到分类器层的数据的大小最相似的大小的教师模型的初始权重,教师模型从选择的至少一个教师模型被选择。

[0115] 模型训练设备可基于输入到选择的教师模型的数据的大小来确定学生模型的结构。模型训练设备可基于输入到选择的教师模型的数据的大小与输入到学生模型的数据的大小之间的差,来改变学生模型的结构。例如,当输入到选择的教师模型的数据的大小对应于64像素×64像素,并且输入到学生模型的数据的大小对应于32像素×32像素时,可改变学生模型的结构,以使学生模型可具有与选择的教师模型的感受域相同或相似的感受域(例如,刺激将触发神经元的放电(firing)的感觉空间的特定区域)。

[0116] 模型训练设备可基于输入到选择的教师模型的数据的大小,改变包括在学生模型中的输入层的节点的数量、隐藏层的数量和隐藏层的节点的数量。

[0117] 在操作630中,模型训练设备可确定学生模型是否满足预定条件。

[0118] 例如,在操作630中,模型训练设备可确定学生模型的精度是否大于或等于阈值。当学生模型的精度小于阈值时,模型训练设备可重新执行操作610和620。每当重新执行操作610时,可基于预定标准变化地选择多个教师模型中的至少一个教师模型。例如,当重新执行操作610时,可选择除在先前操作620中被用于训练学生模型的教师模型之外的剩余教师模型中的至少一个教师模型。通过上述处理,可有效地防止将学生模型过拟合到单个教师模型。此外,模型训练设备可基于训练处理(例如,学生模型的精度),控制将被施加到选择的教师模型的输出数据和正确答案数据的权重。例如,当训练学生模型时,相对于选择的教师模型的输出数据,模型训练设备可针对正确答案数据设置更大的权重。相反地,当学生模型的精度大于或等于阈值时,模型训练设备可终止训练学生模型。

[0119] 作为另一示例,在操作630中,模型训练设备可确定学生模型的训练计数是否满足预设迭代计数。当学生模型的训练计数不满足预设迭代计数时,模型训练设备可重新执行操作610和操作620。在该示例中,每当重新执行操作610时,可基于预定标准变化地选择多个教师模型中的至少一个教师模型。相反地,当学生模型的训练计数满足预设迭代计数时,模型训练设备可终止训练学生模型。

[0120] 图7是示出根据示例性实施例的数据识别方法的流程图。

[0121] 可由包括在数据识别设备中的处理器执行数据识别方法。可将处理器专门配置为执行该操作。

[0122] 参照图7,在操作710中,数据识别设备接收将被识别的目标数据。目标数据是将通过预先训练的模型识别的数据,并可包括例如图像数据、视频数据、语音数据、时间序列数据、传感器数据或他们的各种组合。

[0123] 在操作720中,数据识别设备使用预先训练的模型识别目标数据。模型表示能够从目标数据检测对象的神经网络或对对象进行分类或分簇的神经网络。根据示例性实施例的模型也可能执行除对目标进行检测、分类或分簇之外的其他功能。

[0124] 可基于针对输入数据的至少一个教师模型的输出数据来训练模型,至少一个教师模型从多个教师模型被选择。例如,可基于针对输入数据的选择的至少一个教师模型的输出数据与学生模型的输出数据之间的损耗来训练模型。

[0125] 在该示例中,基于多个教师模型的精度或基于针对输入数据的多个教师模型的输出数据之间的相关性,来选择至少一个教师模型。

[0126] 此外,可基于以下项来训练模型:(i)针对输入数据的模型的输出数据与选择的至少一个教师模型的第一输出数据之间的第一损耗,(ii)从模型的隐藏层得到的分类器层的输出数据与选择的至少一个教师模型的第二输出数据之间的第二损耗。在该示例中,可使用不同的方法确定第一损耗和第二损耗,但不限于此,可选地,可使用相同的方法确定第一损耗和第二损耗。

[0127] 参照图1至图6提供的描述可适用于训练模型以识别目标数据的处理,因此为了简明,将省略重复的描述。

[0128] 图8是示出根据示例性实施例的模型训练设备的框图。

[0129] 参照图8,模型训练设备800包括处理器810和存储器820。模型训练设备800是训练用于数据识别的神经网络的设备,并可使用单个处理器或多个处理器来实现。

[0130] 处理器810可从多个教师模型选择至少一个教师模型,并基于针对输入数据的选择的至少一个教师模型的输出数据来训练学生模型。例如,处理器810可基于针对输入数据的学生模型的输出数据与选择的至少一个教师模型的输出数据之间的损耗来训练学生模型。可将处理器810实现为很多不同类型的硬件处理器,并可被专门配置为除了其他类型的功能以外还执行上述选择至少一个教师模型和训练学生模型的功能。

[0131] 处理器810可基于多个教师模型的精度或基于针对输入数据的多个教师模型的输出数据之间的相关性,从多个教师模型选择至少一个教师模型。

[0132] 处理器810可基于以下项来训练学生模型:(i)针对输入数据的学生模型的输出数据与选择的至少一个教师模型的第一输出数据之间的第一损耗,(ii)从学生模型的隐藏层得到的分类器层的输出数据与选择的至少一个教师模型的第二输出数据之间的第二损耗。根据示例性实施例,可使用不同的方法确定第一损耗和第二损耗。

[0133] 输入数据是将被用于训练学生模型的训练数据,并可包括例如图像数据、音频数据(例如,语音数据、音乐数据等)或他们的各种组合。

[0134] 存储器820可存储由处理器810训练的学生模型。

[0135] 图9是示出根据示例性实施例的数据识别设备的框图。

[0136] 参照图9,数据识别设备900包括接收器910和处理器920。数据识别设备900是能够通过预先训练的模型来识别接收的目标数据的设备,并可被设置在诸如例如智能电话、平板计算机、膝上型计算机、台式计算机、电视机、可穿戴装置、安全系统和智能家居系统的各种计算装置和/或系统中。

[0137] 接收器910可接收将被识别的目标数据。

[0138] 处理器920可使用预先训练的模型来识别目标数据。可基于针对输入数据的至少一个教师模型的输出数据来训练模型,至少一个教师模型从多个教师模型被选择。例如,可基于针对输入数据的学生模型的输出数据与选择的至少一个教师模型的输出数据之间的损耗来训练模型。

[0139] 示例性实施例可从多个教师模型选择至少一个教师模型,并使用选择的至少一个教师模型来训练学生模型,从而有效地增加学生模型的学习速率和精度。

[0140] 此外,示例性实施例可基于预定准则从多个教师模型变化地选择至少一个教师模型,从而防止将学生模型过拟合到预定教师模型。

[0141] 示例性实施例也可使用不同的方法计算学生模型的输出层的损耗和学生模型的

分类器层的损耗,从而即使学生模型是深度神经网络,也能有效地更新神经网络中的连接权重。此外,根据特定示范性实施例,可通过从多个教师模型随机地选择至少一个教师模型,来迭代执行训练学生模型的处理,直到学生模型的精度满足预定条件为止。此外,根据特定示范性实施例,可选择多个教师模型并将多个教师模型用于重新训练单个学生模型。

[0142] 可使用硬件组件和软件组件来实现在此所描述的单元和/或模块。例如,硬件组件可包括麦克风、放大器、带通滤波器、音频数字转换器和处理装置。可使用被配置为通过执行算数、逻辑和输入/输出操作来运行和/或执行程序代码的一个或多个硬件装置来实现处理装置。处理装置可包括处理器、控制器和算术逻辑单元、数字信号处理器、微型计算机、现场可编程阵列、可编程逻辑单元、微型处理器或能够以限定的方式响应并执行指令的任何其它装置。处理装置可运行操作系统(OS)和运行在OS上的一个或多个软件应用。处理装置也可响应于软件的执行来访问、存储、操控、处理和创建数据。为了简单起见,以单数的形式描述上面的处理装置的描述,然而,本领域技术人员将理解:处理装置可包括多个处理元件并还可包括多种类型的处理元件。例如,处理装置可包括多个处理器或一个处理器以及一个控制器。此外,不同的处理配置(诸如,并行处理器)是可行的。

[0143] 根据示范性实施例实施的软件可包括用于独立地或共同地指示和/或配置处理装置按需要进行操作的计算机程序、一段代码、指令或它们的一些组合,从而将处理装置转换为专用处理器。可在任何类型的机器、组件、物理的或虚拟的设备、计算机存储介质或装置中,或可以以能够将指令或数据提供给处理装置或由处理装置解释的传播信号波来永久地或临时地实施软件和数据。软件也可被分布在联网的计算机系统上,使得软件按照分布方式被存储并被执行。软件和数据可由一个或多个非暂时性计算机可读记录介质进行存储。

[0144] 根据上述示范性实施例的方法可被记录在包括程序指令以执行上述示范性实施例的各种操作的非暂时性计算机可读介质中。所述介质还可包括单独的或与所述程序指令结合的数据文件、数据结构等。在所述介质上记录的程序指令可以是针对携带示范性实施例的功能的目的而专门设计和构建的那些程序指令,或者它们可以是对计算机软件领域的技术人员公知和可用的程序指令。非暂时性计算机可读介质的示例包括磁介质(诸如,硬盘、软盘和磁带)、光学介质(诸如,CD-ROM盘、DVD和/或蓝光盘)、磁光介质(诸如,光盘)和专门被配置为存储并执行程序指令的硬件装置(诸如,只读存储器(ROM)、随机存取存储器(RAM)、闪存(例如,USB闪存驱动器、存储卡、记忆棒)等)。程序指令的示例包括机器代码(诸如,由编译器产生的机器代码)和包含可由计算机使用解释器来执行的更高级代码的文件两者。上述装置可被配置为作为一个或多个软件模块以便执行上述示范性实施例的操作,或者反之亦然。

[0145] 以上已描述了多个示范性实施例。然而,应理解,可对这些示范性实施例进行各种修改。例如,如果以不同的顺序执行描述的技术和/或如果以不同的方式组合和/或通过其他组件或他们的等价物替代或补充描述的系统、结构、装置或电路中的组件,则可实现合适的结果。因此,很多不同类型的实施方式落入权利要求的范围内。

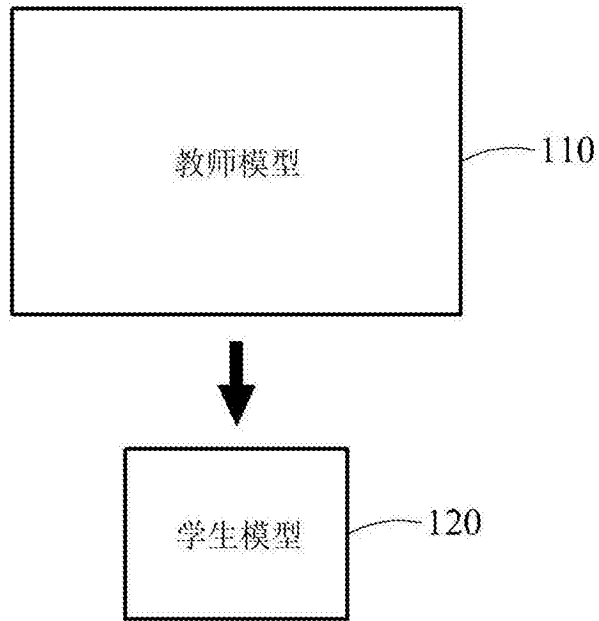


图1

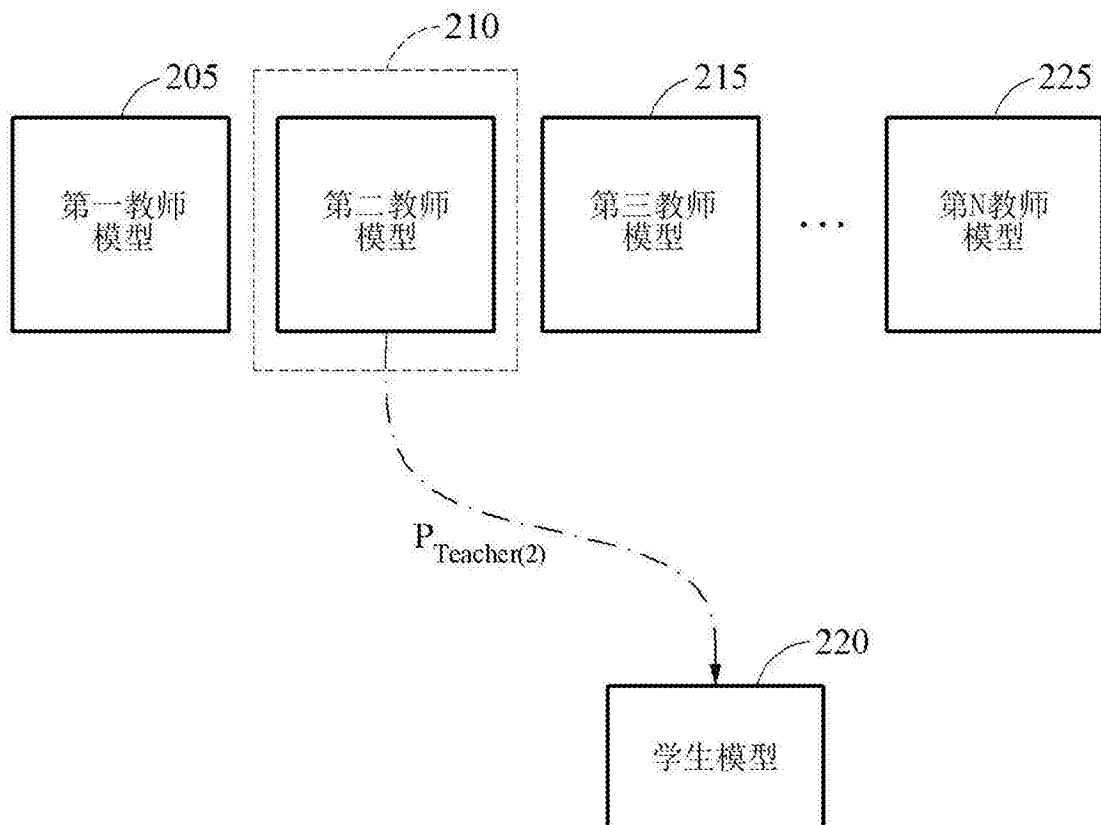


图2

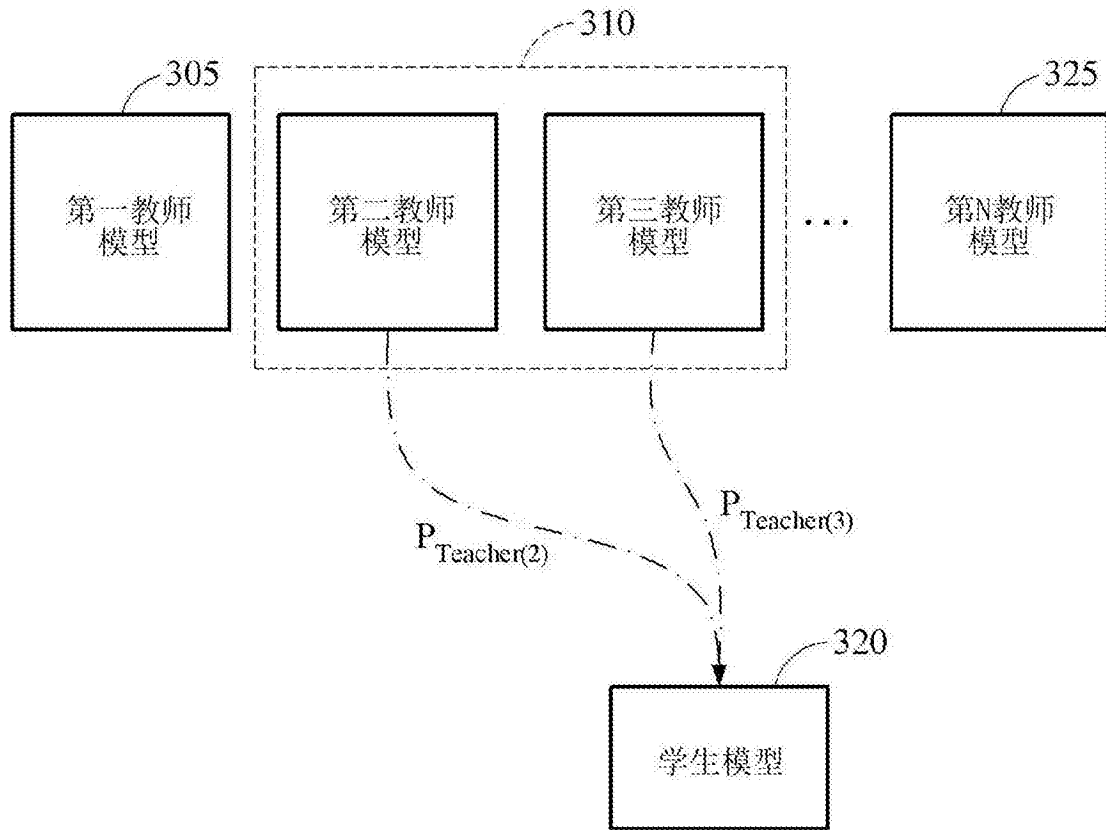


图3

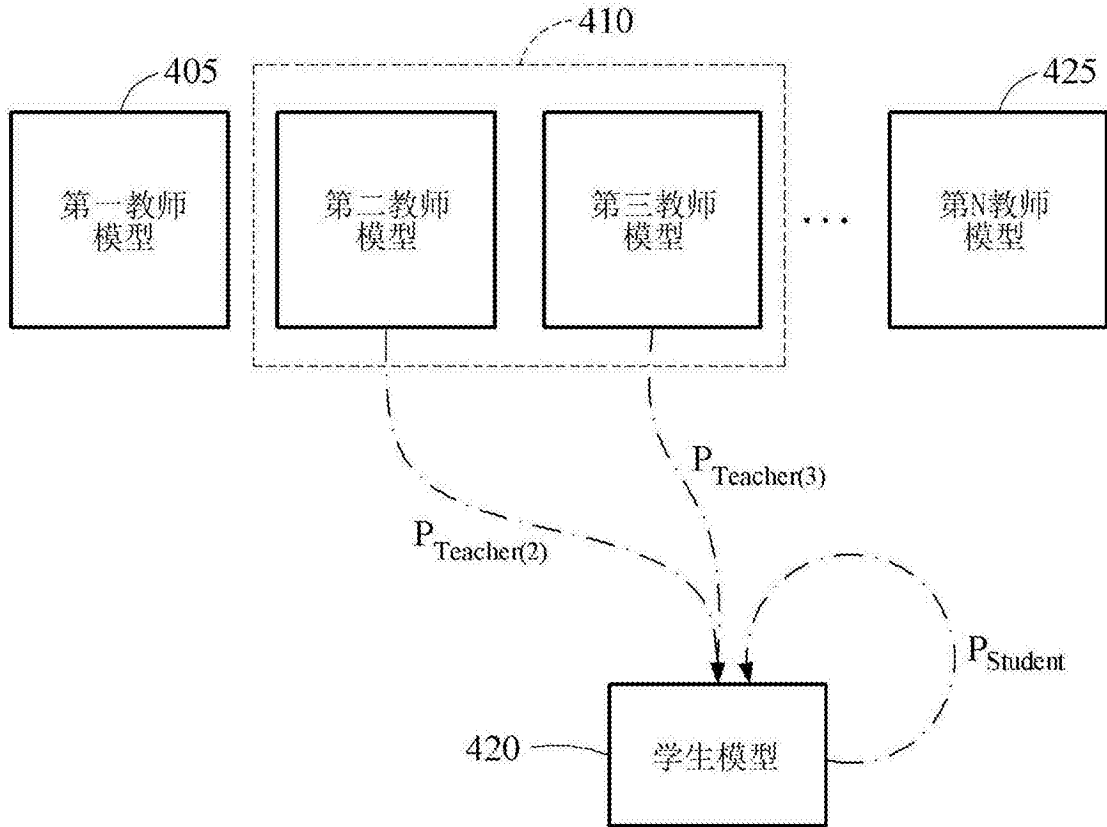


图4

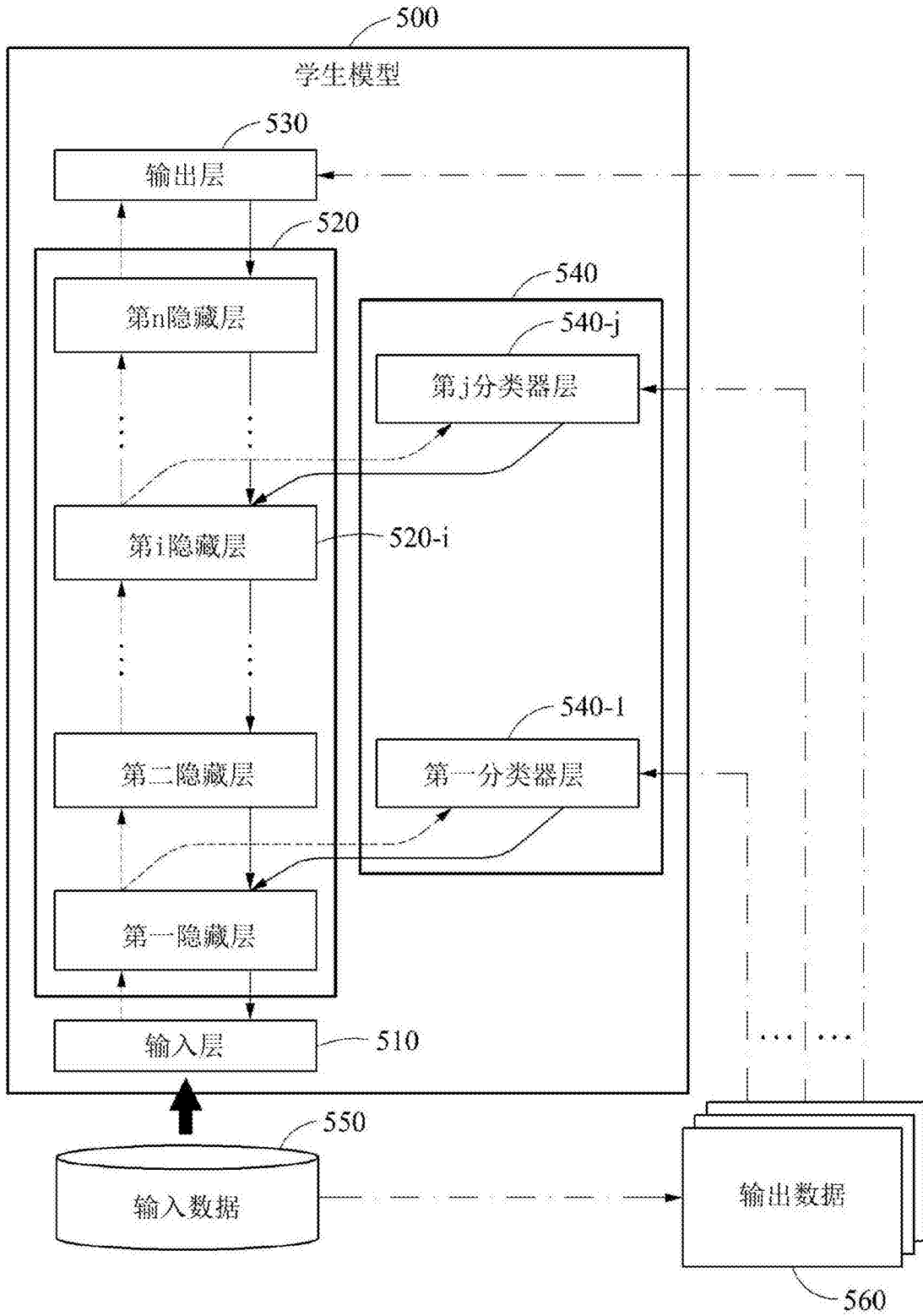


图5

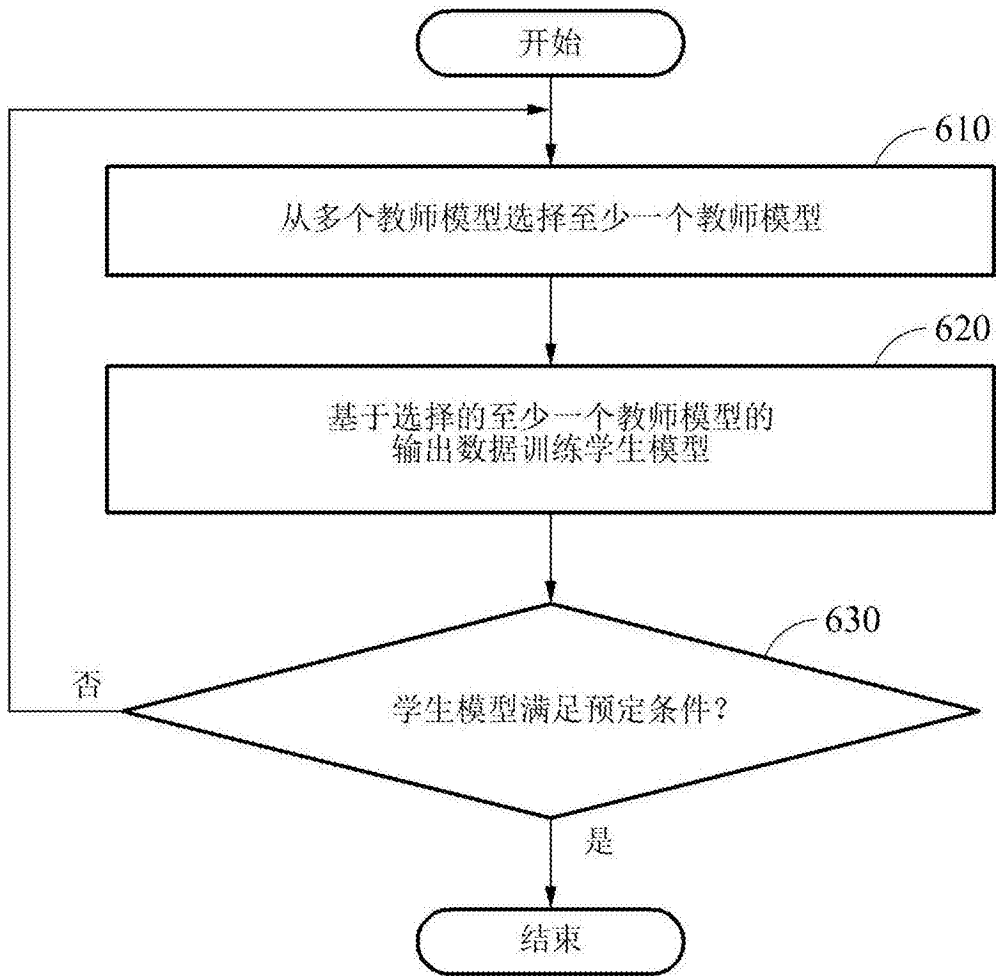


图6

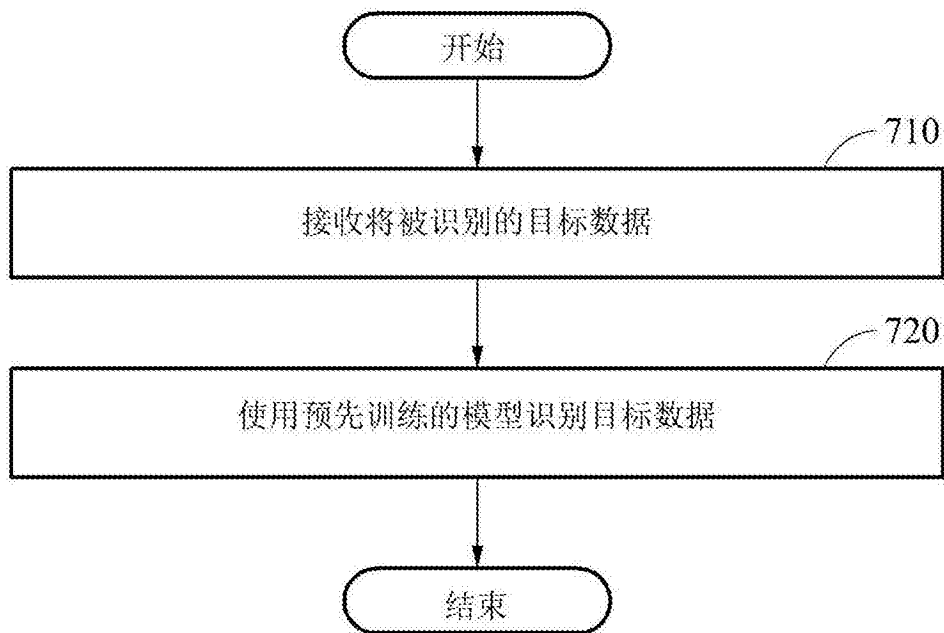


图7

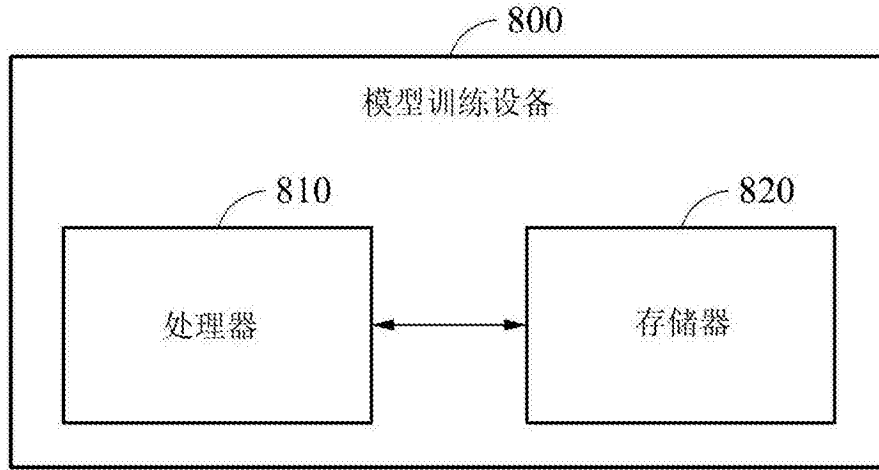


图8

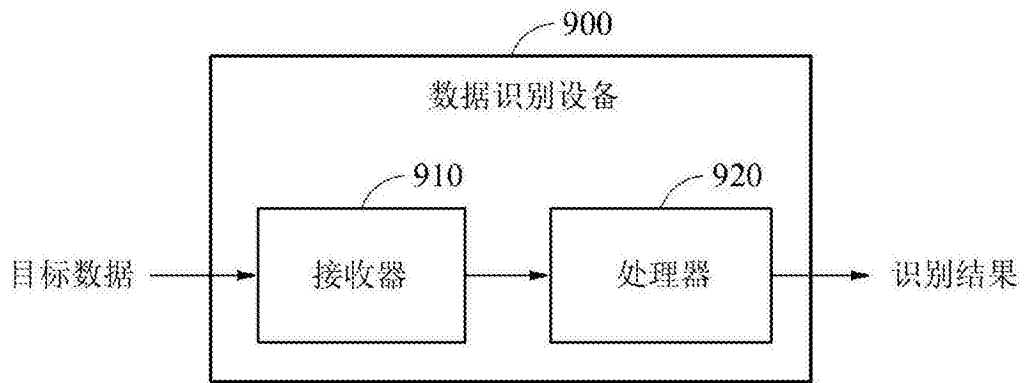


图9