

(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
2020 年 7 月 16 日 (16.07.2020)



(10) 国际公布号
WO 2020/143303 A1

(51) 国际专利分类号:
G06N 3/08 (2006.01)

(21) 国际申请号: PCT/CN2019/117310

(22) 国际申请日: 2019 年 11 月 12 日 (12.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910023779.1 2019 年 1 月 10 日 (10.01.2019) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 金戈(JIN, Ge); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 徐亮(XU, Liang); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市精英专利事务所(SHENZHEN TALENT PATENT SERVICE); 中国广东省深圳市福田区深南中路 6009 号绿景广场 B 栋 20 层 B, Guangdong 518000 (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,

(54) Title: METHOD AND DEVICE FOR TRAINING DEEP LEARNING MODEL, COMPUTER APPARATUS, AND STORAGE MEDIUM

(54) 发明名称: 深度学习模型训练方法、装置、计算机设备及存储介质

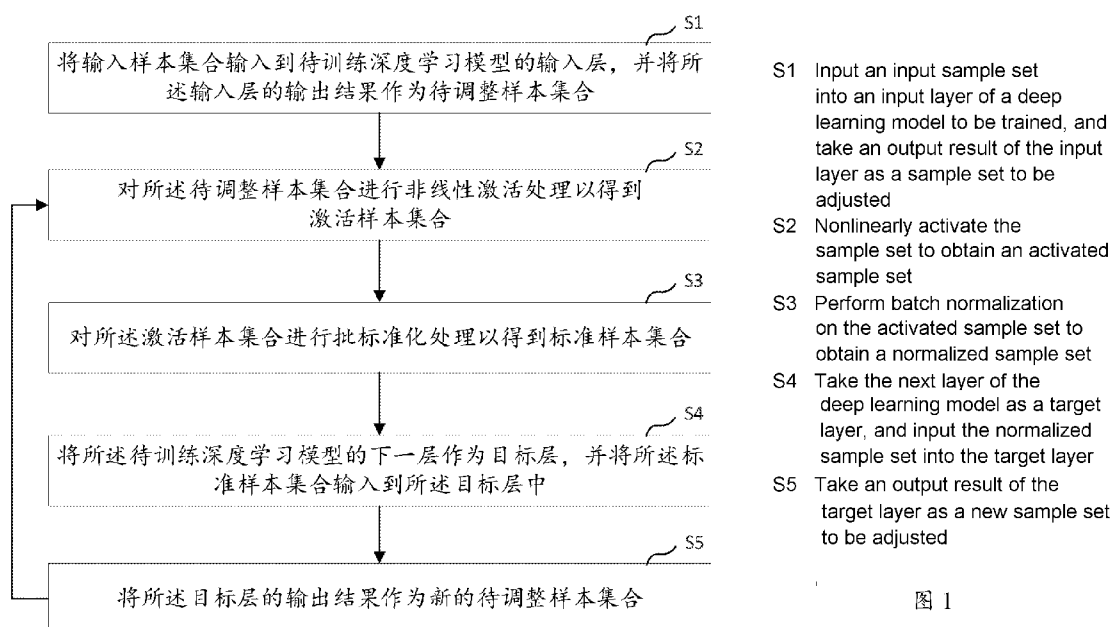


图 1

(57) Abstract: Disclosed in embodiments of the present invention are a method and device for training a deep learning model, a computer apparatus, and a storage medium. The method relates to artificial intelligence technology, and comprises: inputting an input sample set into an input layer of a deep learning model to be trained, and taking an output result of the input layer as a sample set to be adjusted; nonlinearly activating the sample set to obtain an activated sample set; performing batch normalization on the activated sample set to obtain a normalized sample set; taking the next layer of the deep learning model as a target layer, and inputting the normalized

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, 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))。

sample set into the target layer; and taking an output result of the target layer as a new sample set to be adjusted, and continuing to nonlinearly activate the sample set to obtain an activated sample set.

(57) 摘要: 本申请实施例公开了一种深度学习模型训练方法、装置、计算机设备及存储介质。其中, 该方法属于人工智能技术, 该方法包括: 将输入样本集合输入到待训练深度学习模型的输入层, 并将输入层的输出结果作为待调整样本集合; 对待调整样本集合进行非线性激活处理以得到激活样本集合; 对激活样本集合进行批标准化处理以得到标准样本集合; 将待训练深度学习模型的下一层作为目标层, 并将标准样本集合输入到所述目标层中; 将目标层的输出结果作为新的待调整样本集合, 并继续对待调整样本集合进行非线性激活处理以得到激活样本集合。

发明名称：深度学习模型训练方法、装置、计算机设备及存储介质

本申请要求于2019年1月10日提交中国专利局、申请号为201910023779.1、申请名称为“深度学习模型训练方法、装置、计算机设备及存储介质”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

本申请涉及人工智能技术领域，尤其涉及一种深度学习模型训练方法、装置、计算机设备及存储介质。

背景技术

深度学习是机器学习研究中的一个新的领域，其动机在于建立、模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据，例如图像，声音和文本等数据。

深度学习模型，例如，卷积神经网络(Convolutional Neural Network, CNN)需经大量数据训练后才能实际使用。在深度学习模型的训练过程中，大多选择用批标准化(Batch Normalization, BN)的方法对深度学习模型的各层进行处理，使得网络在每一层传递的过程中样本的差异性有所降低，然而现有的处理方法对下一层网络的控制不够，导致深度学习模型的训练效果不佳。

发明内容

本申请实施例提供了一种深度学习模型训练方法、装置、计算机设备及存储介质，旨在提高深度学习模型的训练效果。

第一方面，本申请实施例提供了一种深度学习模型训练方法，其包括：

将输入样本集合输入到待训练深度学习模型的输入层，并将所述输入层的输出结果作为待调整样本集合；

对所述待调整样本集合进行非线性激活处理以得到激活样本集合；

对所述激活样本集合进行批标准化处理以得到标准样本集合；

将所述待训练深度学习模型的下一层作为目标层，并将所述标准样本集合输入到所述目标层中；

将所述目标层的输出结果作为新的待调整样本集合，并返回所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合的步骤。

第二方面，本申请实施例还提供了一种深度学习模型训练装置，其包括：

第一输入单元，用于将输入样本集合输入到待训练深度学习模型的输入层，并将所述输入层的输出结果作为待调整样本集合；

第一激活单元，用于对所述待调整样本集合进行非线性激活处理以得到激活样本集合；

第一批标准化单元，用户对所述激活样本集合进行批标准化处理以得到标准样本集合；

第二输入单元，用于将所述待训练深度学习模型的下一层作为目标层，并将所述标准样本集合输入到所述目标层中；

通知单元，用于将所述目标层的输出结果作为新的待调整样本集合，并通知激活单元返回所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合的步骤。

第三方面，本申请实施例还提供了一种计算机设备，包括存储器以及与所述存储器相连的处理器；所述存储器用于存储计算机程序；所述处理器用于运行所述存储器中存储的计算机程序，以执行如下步骤：

将输入样本集合输入到待训练深度学习模型的输入层，并将所述输入层的输出结果作为待调整样本集合；

对所述待调整样本集合进行非线性激活处理以得到激活样本集合；

对所述激活样本集合进行批标准化处理以得到标准样本集合；

将所述待训练深度学习模型的下一层作为目标层，并将所述标准样本集合输入到所述目标层中；

将所述目标层的输出结果作为新的待调整样本集合，并返回所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合的步骤。

第四方面，本申请实施例还提供一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时使所述处理器执行以下步骤：

将输入样本集合输入到待训练深度学习模型的输入层，并将所述输入层的输出结果作为待调整样本集合；

对所述待调整样本集合进行非线性激活处理以得到激活样本集合；

对所述激活样本集合进行批标准化处理以得到标准样本集合；

将所述待训练深度学习模型的下一层作为目标层，并将所述标准样本集合输入到所述目标层中；

将所述目标层的输出结果作为新的待调整样本集合，并返回所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合的步骤。

附图说明

为了更清楚地说明本申请实施例技术方案，下面将对实施例描述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。

图 1 为本申请实施例提供的一种深度学习模型训练方法的流程示意图；

图 2 为本申请实施例提供的一种深度学习模型训练方法的子流程示意图；

图 3 为本申请实施例提供的一种深度学习模型训练方法的子流程示意图；

图 4 为本申请另一实施例提供的一种深度学习模型训练方法的流程示意图；

图 5 为本申请实施例提供的一种深度学习模型训练方法的子流程示意图；

图 6 为本申请实施例提供的一种深度学习模型训练装置的示意性框图；

图 7 为本申请实施例提供的一种深度学习模型训练装置的第一激活单元单元的示意性框图；

图 8 为本申请实施例提供的一种深度学习模型训练装置的第一批标准化单元的示意性框图；

图 9 为本申请实施例提供的一种深度学习模型训练装置的第一批标准化单元的获取单元的示意性框图；

图 10 为本申请另一实施例提供的一种深度学习模型训练装置的示意性框图；

图 11 为本申请另一实施例提供的一种深度学习模型训练装置的第一分词单元的示意性框图；以及

图 12 为本申请实施例提供的一种计算机设备的示意性框图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例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做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。

应当理解，当在本说明书和所附权利要求书中使用时，术语“包括”和“包含”指示所描述特征、整体、步骤、操作、元素和/或组件的存在，但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

还应当理解，在此本申请说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本申请。如在本申请说明书和所附权利要求书中所使用的那样，除非上下文清楚地指明其它情况，否则单数形式的“一”、“一个”及“该”意在包括复数形式。

还应当进一步理解，在本申请说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合，并且包括这些组合。

如在本说明书和所附权利要求书中所使用的那样，术语“如果”可以依据上下文被解释为“当... 时”或“一旦”或“响应于确定”或“响应于检测到”。类似地，短语“如果确定”或“如果检测到[所描述条件或事件]”可以依据上下文被解释为意指“一旦确定”或“响应于确定”或“一旦检测到[所描述条件或事件]”或“响应于检测到[所描述条件或事件]”。

请参阅图 1，图 1 是本申请实施例提供的一种深度学习模型训练方法的流程示意图。如图所示，该方法包括以下步骤 S1-S5：

S1，将输入样本集合输入到待训练深度学习模型的输入层，并将所述输入层的输出结果作为待调整样本集合。

在本申请实施例中，通过输入样本集合来对待训练深度学习模型进行训练。待训练深度学习模型包括输入层、多个隐藏层以及输出层。

具体实施中，将输入样本集合输入到待训练深度学习模型的输入层，以对待训练深度学习模型的输入层进行训练。

在本申请实施例中，输入层输出结果时，将输入层的输出结果作为待调整样本集合，并对待调整样本集合进行调整后再输入到待训练深度学习模型的下

一层中。

S2, 对所述待调整样本集合进行非线性激活处理以得到激活样本集合。

具体实施中, 对所述待调整样本集合进行非线性激活处理以得到激活样本集合。通过对所述待调整样本集合进行非线性激活处理能够提高待训练深度学习模型的非线性因素, 提高待训练深度学习模型的表现力。

在一实施例中, 以上步骤 S2 具体包括如下步骤:

通过预设的非线性激活函数对所述待调整样本集合进行非线性激活处理以得到激活样本集合。

需要说明的是, 在本申请中, 非线性激活函数包括: Sigmoid 函数、Tanh 函数以及 ReLU (Rectified Linear Unit, 修正线性单元) 函数, 本申请对此不作具体限定。

S3, 对所述激活样本集合进行批标准化处理以得到标准样本集合。

具体实施中, 通过对所述激活样本集合进行批标准化处理以得到标准样本集合。批标准化处理能够降低样本在深度学习模型各层传递过程中的差异性, 从而提高了模型的训练效果。

在本申请实施例中, 先对待调整样本集合进行非线性激活处理得到激活样本集合后, 再对激活样本集合进行批标准化处理以得到标准样本集合。通过将批标准化处理调整到非线性激活处理之后 (非线性激活处理会提高样本的差异性), 能够获得对于下一层网络更好的控制, 从而提高了对待训练深度学习模型的训练效果。

在一实施例中, 参见图 2, 以上步骤 S3 包括如下步骤 S31-S32:

S31, 获取所述激活样本集合中各样本的均值以及方差。

具体实施中, 对激活样本集合进行批标准化处理需要用到激活样本集合中各样本的均值以及方差, 为此首先计算激活样本集合中各样本的均值以及方差。

在一实施例中, 参见图 3, 以上步骤 S31 具体包括如下步骤 S311-S312:

S311, 通过以下公式 $\mu = \frac{\sum_{i=1}^m x_i}{m}$ 计算激活样本集合中各样本的均值 μ 。

具体实施中, 通过以下公式 $\mu = \frac{\sum_{i=1}^m x_i}{m}$ 计算激活样本集合中各样本的均值 μ , 其中, i 为样本的序号, m 为样本的数量, x_i 为样本的值。

S312, 通过以下公式 $\sigma^2 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \mu)^2$ 计算激活样本集合中各样本的方差 σ 。

具体实施中, 通过公式 $\sigma^2 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \mu)^2$ 计算激活样本集合中各样本的方差 σ 。

其中, i 为样本的序号, m 为样本的数量, x_i 为样本的值。

S32, 根据预设的批标准化公式以及所述激活样本集合中各样本的均值以及方差对所述激活样本集合进行批处理化处理。

具体实施中, 在获取了激活样本集合中各样本的均值以及方差后, 根据预设的批标准化公式以及所述激活样本集合中各样本的均值以及方差对所述激活样本集合进行批处理化处理。

在本申请实施例中, 批标准化公式为 $y_i = \gamma \frac{wx_i - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta$, 其中, 在以上公式中, i 为样本的序号, x_i 为激活样本集合中的样本的数值, y_i 为相应的标准样本集合中的样本的数值, μ 为激活样本集合中各样本的均值, σ 为激活样本集合中各样本的方差, m 为激活样本集合中样本的数量, w 、 γ 、 β 以及 ϵ 为待训练深度学习模型的参数, 以上随机初始化生成, 然后训练过程中通过迭代来更新这些参数。

S4, 将所述待训练深度学习模型的下一层作为目标层, 并将所述标准样本集合输入到所述目标层中。

具体实施中, 将所述待训练深度学习模型的下一层作为目标层, 并将所述标准样本集合输入到所述目标层中, 以对所述目标层进行训练。

S5, 将所述目标层的输出结果作为新的待调整样本集合, 并返回所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合的步骤。

本申请的方案中, 将目标层的输出结果作为新的待调整样本集合, 并且返回所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合的步骤, 然后再对激活样本集合进行批标准化处理以得到标准样本集合, 之后再待训练深度学习模型的下一层作为目标层, 并将所述标准样本集合输入到所述目标层中, 以对所述目标层进行训练。以此类推直到所述待训练深度学习模型输出结果。

通过以上方法, 可实现对待训练深度学习模型的上一层的输出结果先后进行非线性激活处理以及批标准化处理后, 再输入到待训练深度学习模型的下一

层中，直到待训练深度学习模型的输出层时，直接输出结果。

本申请实施例中，通过将批标准化处理调整到非线性激活处理之后（非线性激活处理会提高样本的差异性），从而批标准化处理直接作用于深度学习模型的下一层结构，能够获得对于下一层结构更好的控制，提高了对待训练深度学习模型的训练效果。

图4是本申请另一实施例提供的一种深度学习模型训练方法的流程示意图。如图4所示，本实施例的深度学习模型训练方法包括步骤S41-S47。其中步骤S43-S47与上述实施例中的步骤S1-S5类似，在此不再赘述。下面详细说明本实施例中所增加的步骤S41-S42。

S41，对训练文本进行分词处理以得到分词样本集合，其中，所述分词样本集合为对所述训练文本进行分词后得到的样本组成的集合。

在本实施例中，训练文本为预存在终端中的文本，可直接调用获取。

具体实施中，分词指的是将一个汉字序列切分成一个个单独的词。分词就是将连续的字序列按照一定的规范重新组合成词序列的过程。分词是文本处理中的一个基础步骤。

通过对训练文本进行分词处理以得到分词样本集合，其中，分词样本集合为对所述训练文本进行分词后得到的样本（词语）组成的集合。

在一实施例中，在一实施例中，参见图5，以上步骤S41具体包括如下步骤S411-S412：

S411，通过预设的分词工具对训练文本进行分词处理以得到初始分词样本集合。

具体实施中，常用的分词工具为结巴分词工具。本实施例中，采用结巴分词工具对训练文本进行分词处理以得到初始分词样本集合。结巴分词工具适用于对中文文本的分词，其对中文文本分词的准确性极高，从而可提高本方案准确性。

或者，在其他实施例中，可采用其他分词工具来对训练文本进行分词处理，本申请对此不作具体限定。

S412，将所述初始分词样本集合中的停止词去除以得到所述分词样本集合。

具体实施中，将所述初始分词样本集合中的停止词去除以得到分词样本集合。需要说明的是，停止词（stop word），常为介词、副词或连词等。例如，"

在"、"里面"、"也"、"的"、"它"、"为"等都为停止词。

S42, 通过预设词向量工具对所述分词样本集合中的样本进行词向量训练以得到所述输入样本集合, 其中, 所述输入样本集合为所述分词样本集合中的样本的词向量组成的集合。

具体实施中, 采用 word2vec 作为词向量工具, word2vec 是一种自然语言处理工具, 其作用就是将自然语言中的字词转为计算机可以理解的词向量。

传统的词向量容易受维数灾难的困扰, 且任意两个词之间都是孤立的, 不能体现词和词之间的关系, 因此本实施例采用 word2vec 来得到词向量, 其可通过计算向量之间的距离来体现词与词之间的相似性, 使得训练结果更加准确。

本实施例中, 通过 word2vec 对分词样本集合中的样本进行词向量训练以得到各样本的词向量。将分词样本集合中个样本的词向量组合得到输入样本集合。

或者, 在其他实施例中, 可采用其他词向量工具对所述分词样本集合中的样本进行词向量训练, 本申请对此不作具体限定。

图 6 是本申请实施例提供的一种深度学习模型训练装置 60 的示意性框图。如图 6 所示, 对应于以上深度学习模型训练方法, 本申请还提供一种深度学习模型训练装置 60。该深度学习模型训练装置 60 包括用于执行上述深度学习模型训练方法的单元, 该装置可以被配置于台式电脑、平板电脑、手提电脑、等终端中。具体地, 请参阅图 6, 该深度学习模型训练装置 60 包括第一输入单元 61、第一激活单元 62、第一批标准化单元 63、第二输入单元 64 以及通知单元 65。

第一输入单元 61, 用于将输入样本集合输入到待训练深度学习模型的输入层, 并将所述输入层的输出结果作为待调整样本集合; 第一激活单元 62, 用于对所述待调整样本集合进行非线性激活处理以得到激活样本集合; 第一批标准化单元 63, 用户对所述激活样本集合进行批标准化处理以得到标准样本集合; 第二输入单元 64, 用于将所述待训练深度学习模型的下一层作为目标层, 并将所述标准样本集合输入到所述目标层中; 通知单元 65, 用于将所述目标层的输出结果作为新的待调整样本集合, 并通知激活单元返回所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合的步骤。

在一实施例中, 参见图 7, 第一激活单元 62 包括第二激活单元 621。第二激活单元 621, 用于通过预设的非线性激活函数对所述待调整样本集合进行非线性激活处理以得到激活样本集合。

在一实施例中，参见图 8，第一批标准化单元 63 包括获取单元 631 以及第二批标准化单元 632。

获取单元 631，用于获取所述激活样本集合中各样本的均值以及方差；第二批标准化单元 632，用于根据预设的批标准化公式以及所述激活样本集合中各样本的均值以及方差对所述激活样本集合进行批处理化处理。

在一实施例中，参见图 9，获取单元 631 包括第一计算单元 6311 以及第二计算单元 6312。

第一计算单元 6211，用于通过以下公式 $\mu = \frac{\sum_{i=1}^m x_i}{m}$ 计算激活样本集合中各样本的均值 μ ；第二计算单元 6312，用于通过以下公式 $\sigma^2 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \mu)^2$ 计算激活样本集合中各样本的方差 σ ；其中， i 为样本的序号， m 为样本的数量， x_i 为样本的值。

图 10 是本申请另一实施例提供的一种深度学习模型训练装置 60 的示意性框图。如图 10 所示，本实施例的深度学习模型训练装置 60 是上述实施例的基础上增加了第一分词单元 66 以及训练单元 67。

第一分词单元 66，用于对训练文本进行分词处理以得到分词样本集合，所述分词样本集合为对所述训练文本进行分词后得到的样本组成的集合；训练单元 67，用于通过预设词向量工具对所述分词样本集合中的样本进行词向量训练以得到所述输入样本集合，所述输入样本集合为所述分词样本集合中的样本的词向量组成的集合。

在一实施例中，参见图 11，第一分词单元 66 包括第二分词单元 661 以及去除单元 662。

第二分词单元 661，用于通过预设的分词工具对训练文本进行分词处理以得到初始分词样本集合；去除单元 662，用于将所述初始分词样本集合中的停止词去除以得到所述分词样本集合。

需要说明的是，所属领域的技术人员可以清楚地了解到，上述深度学习模型训练装置 60 和各单元的具体实现过程，可以参考前述方法实施例中的相应描述，为了描述的方便和简洁，在此不再赘述。

上述深度学习模型训练装置 60 可以实现为一种计算机程序的形式，该计算机程序可以在如图 12 所示的计算机设备上运行。

请参阅图 12，图 12 是本申请实施例提供的一种计算机设备的示意性框图。该计算机设备 500 是终端，其中，终端可以是智能手机、平板电脑、笔记本电脑、台式电脑、个人数字助理和穿戴式设备等具有通信功能的电子设备。

参阅图 12，该计算机设备 500 包括通过系统总线 501 连接的处理器 502、存储器和网络接口 505，其中，存储器可以包括非易失性存储介质 503 和内存储器 504。

该非易失性存储介质 503 可存储操作系统 5031 和计算机程序 5032。该计算机程序 5032 被执行时，可使得处理器 502 执行一种深度学习模型训练方法。

该处理器 502 用于提供计算和控制能力，以支撑整个计算机设备 500 的运行。

该内存储器 504 为非易失性存储介质 503 中的计算机程序 5032 的运行提供环境，该计算机程序 5032 被处理器 502 执行时，可使得处理器 502 执行一种深度学习模型训练方法。

该网络接口 505 用于与其它设备进行网络通信。本领域技术人员可以理解，图 12 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备 500 的限定，具体的计算机设备 500 可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

其中，所述处理器 502 用于运行存储在存储器中的计算机程序 5032，以实现本申请的深度学习模型训练方法。

应当理解，在本申请实施例中，处理器 502 可以是中央处理单元 (Central Processing Unit, CPU)，该处理器 502 还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现成可编程门阵列 (Field-Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。其中，通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

本领域普通技术人员可以理解的是实现上述实施例的方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成。该计算机程序可存储于一存储介质中，该存储介质为计算机可读存储介质。该计算机程序被该计算机系统中的至少一个处理器执行，以实现上述方法的实施例的流程步骤。

因此，本申请还提供一种存储介质。该存储介质可以为计算机可读存储介质。该存储介质存储有计算机程序。该计算机程序被处理器执行时使处理器执行本申请的深度学习模型训练方法。

所述存储介质为实体的、非瞬时性的存储介质，例如可以是U盘、移动硬盘、只读存储器（Read-Only Memory, ROM）、磁碟或者光盘等各种可以存储程序代码的实体存储介质。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

以上所述，仅为本申请的具体实施方式，但本申请明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

权 利 要 求 书

1.一种深度学习模型训练方法，包括：

将输入样本集合输入到待训练深度学习模型的输入层，并将所述输入层的输出结果作为待调整样本集合；

对所述待调整样本集合进行非线性激活处理以得到激活样本集合；

对所述激活样本集合进行批标准化处理以得到标准样本集合；

将所述待训练深度学习模型的下一层作为目标层，并将所述标准样本集合输入到所述目标层中；

将所述目标层的输出结果作为新的待调整样本集合，并返回所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合的步骤。

2.根据权利要求1所述的方法，其中，所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合，包括：

通过预设的非线性激活函数对所述待调整样本集合进行非线性激活处理以得到激活样本集合。

3.根据权利要求2所述的方法，其中，所述非线性激活函数包括 Sigmoid 函数、Tanh 函数以及 ReLU 函数。

4.根据权利要求1所述的方法，其中，所述对所述激活样本集合进行批标准化处理以得到标准样本集合，包括：

获取所述激活样本集合中各样本的均值以及方差；

根据预设的批标准化公式以及所述激活样本集合中各样本的均值以及方差对所述激活样本集合进行批处理化处理。

5.根据权利要求4所述的方法，其中，所述获取所述激活样本集合中各样本的均值以及方差，包括：

通过以下公式 $\mu = \frac{\sum_{i=1}^m x_i}{m}$ 计算激活样本集合中各样本的均值 μ ；

通过以下公式 $\sigma^2 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \mu)^2$ 计算激活样本集合中各样本的方差 σ ；

其中， i 为样本的序号， m 为样本的数量， x_i 为样本的值。

6.根据权利要求1所述的方法，其中，在所述将输入样本集合输入到待训练深度学习模型的输入层，并将所述输入层的输出结果作为待调整样本集合之前，所述方法还包括：

对训练文本进行分词处理以得到分词样本集合，所述分词样本集合为对所述训练文本进行分词后得到的样本组成的集合；

通过预设词向量工具对所述分词样本集合中的样本进行词向量训练以得到所述输入样本集合，所述输入样本集合为所述分词样本集合中的样本的词向量组成的集合。

7.根据权利要求6所述的方法，其中，所述对训练文本进行分词处理以得到分词样本集合，包括：

通过预设的分词工具对训练文本进行分词处理以得到初始分词样本集合；

将所述初始分词样本集合中的停止词去除以得到所述分词样本集合。

8.根据权利要求6所述的方法，其中，所述词向量工具为 word2vec。

9.根据权利要求7所述的方法，其中，所述分词工具为结巴分词工具。

10.一种深度学习模型训练装置，包括：

第一输入单元，用于将输入样本集合输入到待训练深度学习模型的输入层，并将所述输入层的输出结果作为待调整样本集合；

第一激活单元，用于对所述待调整样本集合进行非线性激活处理以得到激活样本集合；

第一批标准化单元，用户对所述激活样本集合进行批标准化处理以得到标准样本集合；

第二输入单元，用于将所述待训练深度学习模型的下一层作为目标层，并将所述标准样本集合输入到所述目标层中；

通知单元，用于将所述目标层的输出结果作为新的待调整样本集合，并通知激活单元返回所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合的步骤。

11.一种计算机设备，包括存储器以及与所述存储器相连的处理器；其中，所述存储器用于存储计算机程序；所述处理器用于运行所述存储器中存储的计算机程序，以执行如下步骤：

将输入样本集合输入到待训练深度学习模型的输入层，并将所述输入层的输出结果作为待调整样本集合；

对所述待调整样本集合进行非线性激活处理以得到激活样本集合；

对所述激活样本集合进行批标准化处理以得到标准样本集合；

将所述待训练深度学习模型的下一层作为目标层，并将所述标准样本集合输入到所述目标层中；

将所述目标层的输出结果作为新的待调整样本集合，并返回所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合的步骤。

12.根据权利要求 11 所述的计算机设备，其中，所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合的步骤包括：

通过预设的非线性激活函数对所述待调整样本集合进行非线性激活处理以得到激活样本集合。

13.根据权利要求 12 所述的计算机设备，其中，所述非线性激活函数包括 Sigmoid 函数、Tanh 函数以及 ReLU 函数。

14.根据权利要求 11 所述的计算机设备，其中，所述对所述激活样本集合进行批标准化处理以得到标准样本集合的步骤包括：

获取所述激活样本集合中各样本的均值以及方差；

根据预设的批标准化公式以及所述激活样本集合中各样本的均值以及方差对所述激活样本集合进行批处理化处理。

15.根据权利要求 14 所述的计算机设备，其中，所述获取所述激活样本集合中各样本的均值以及方差的步骤包括：

通过以下公式 $\mu = \frac{\sum_{i=1}^m x_i}{m}$ 计算激活样本集合中各样本的均值 μ ；

通过以下公式 $\sigma^2 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \mu)^2$ 计算激活样本集合中各样本的方差 σ ；

其中， i 为样本的序号， m 为样本的数量， x_i 为样本的值。

16.根据权利要求 11 所述的计算机设备，其中，在所述将输入样本集合输入到待训练深度学习模型的输入层，并将所述输入层的输出结果作为待调整样本集合之前，所述处理器还执行如下步骤：

对训练文本进行分词处理以得到分词样本集合，所述分词样本集合为对所述训练文本进行分词后得到的样本组成的集合；

通过预设词向量工具对所述分词样本集合中的样本进行词向量训练以得到所述输入样本集合，所述输入样本集合为所述分词样本集合中的样本的词向量组成的集合。

17.根据权利要求 16 所述的计算机设备，其中，所述对训练文本进行分词处

理以得到分词样本集合的步骤包括：

通过预设的分词工具对训练文本进行分词处理以得到初始分词样本集合；
将所述初始分词样本集合中的停止词去除以得到所述分词样本集合。

18.根据权利要求 16 所述的计算机设备,其中,所述词向量工具为 word2vec。

19.根据权利要求 17 所述的计算机设备,其中,所述分词工具为结巴分词工具。

20.一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,其中,所述计算机程序被处理器执行时使所述处理器执行以下步骤:

将输入样本集合输入到待训练深度学习模型的输入层,并将所述输入层的输出结果作为待调整样本集合;

对所述待调整样本集合进行非线性激活处理以得到激活样本集合;

对所述激活样本集合进行批标准化处理以得到标准样本集合;

将所述待训练深度学习模型的下一层作为目标层,并将所述标准样本集合输入到所述目标层中;

将所述目标层的输出结果作为新的待调整样本集合,并返回所述对所述待调整样本集合进行非线性激活处理以得到激活样本集合的步骤。

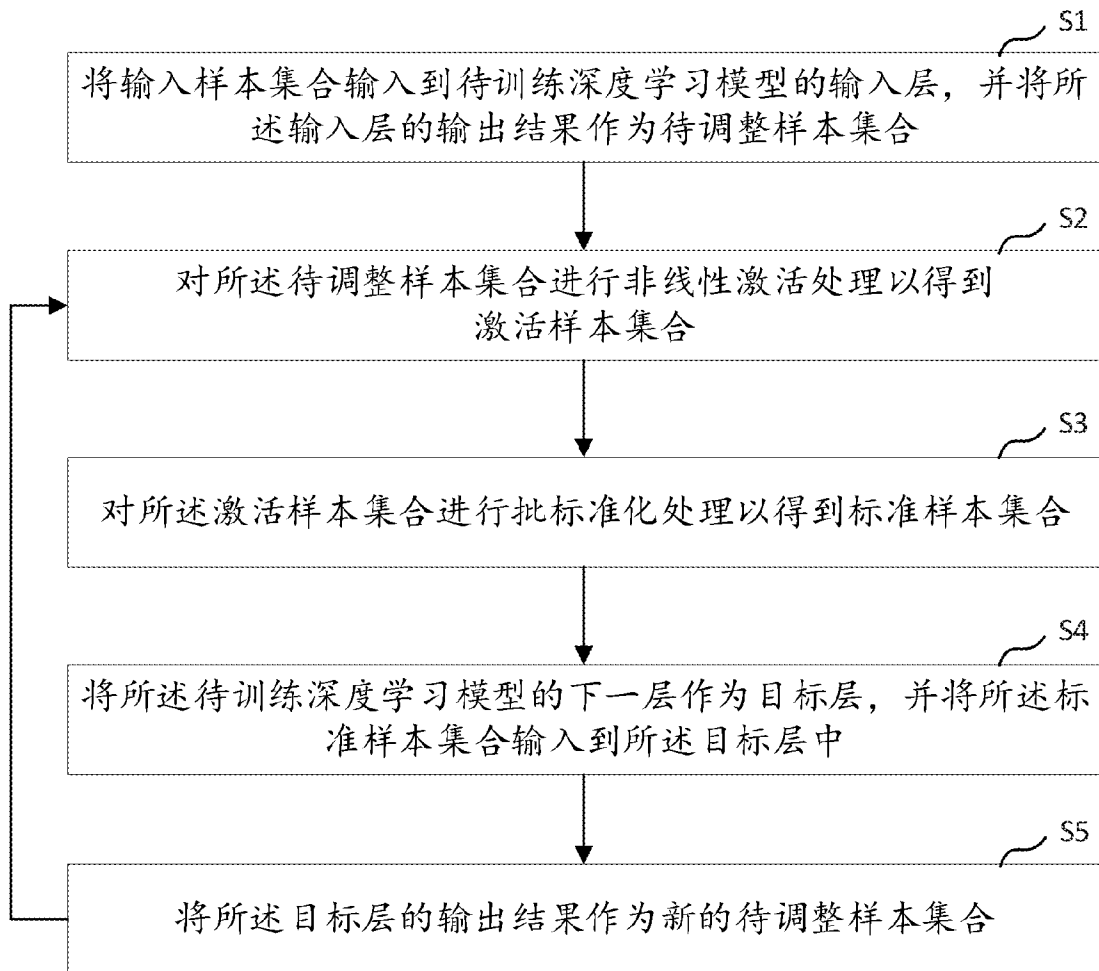


图 1

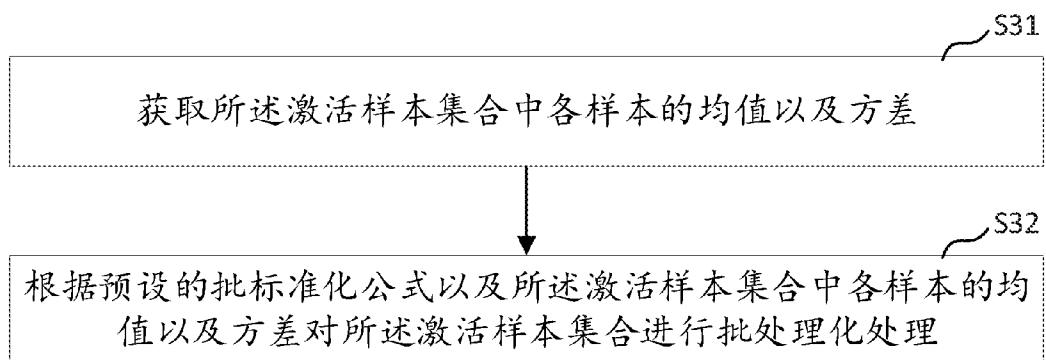


图 2

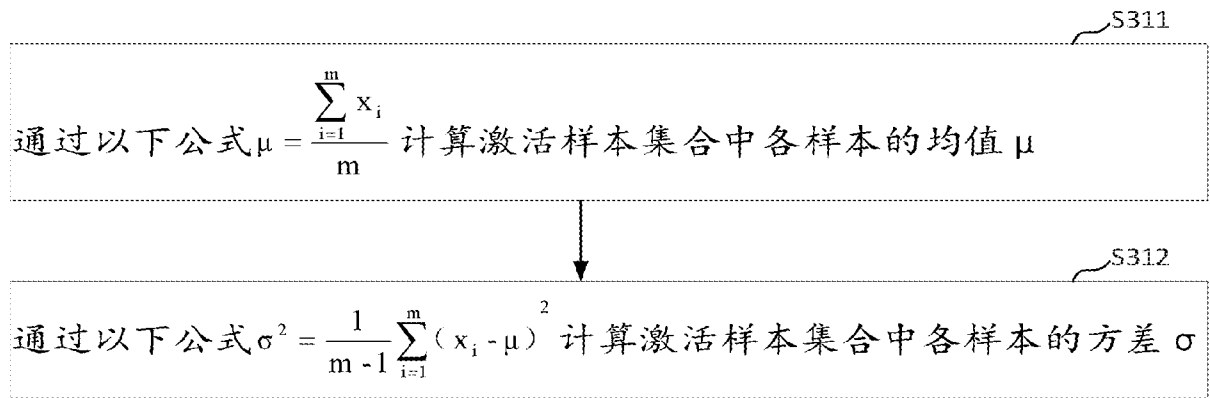


图 3

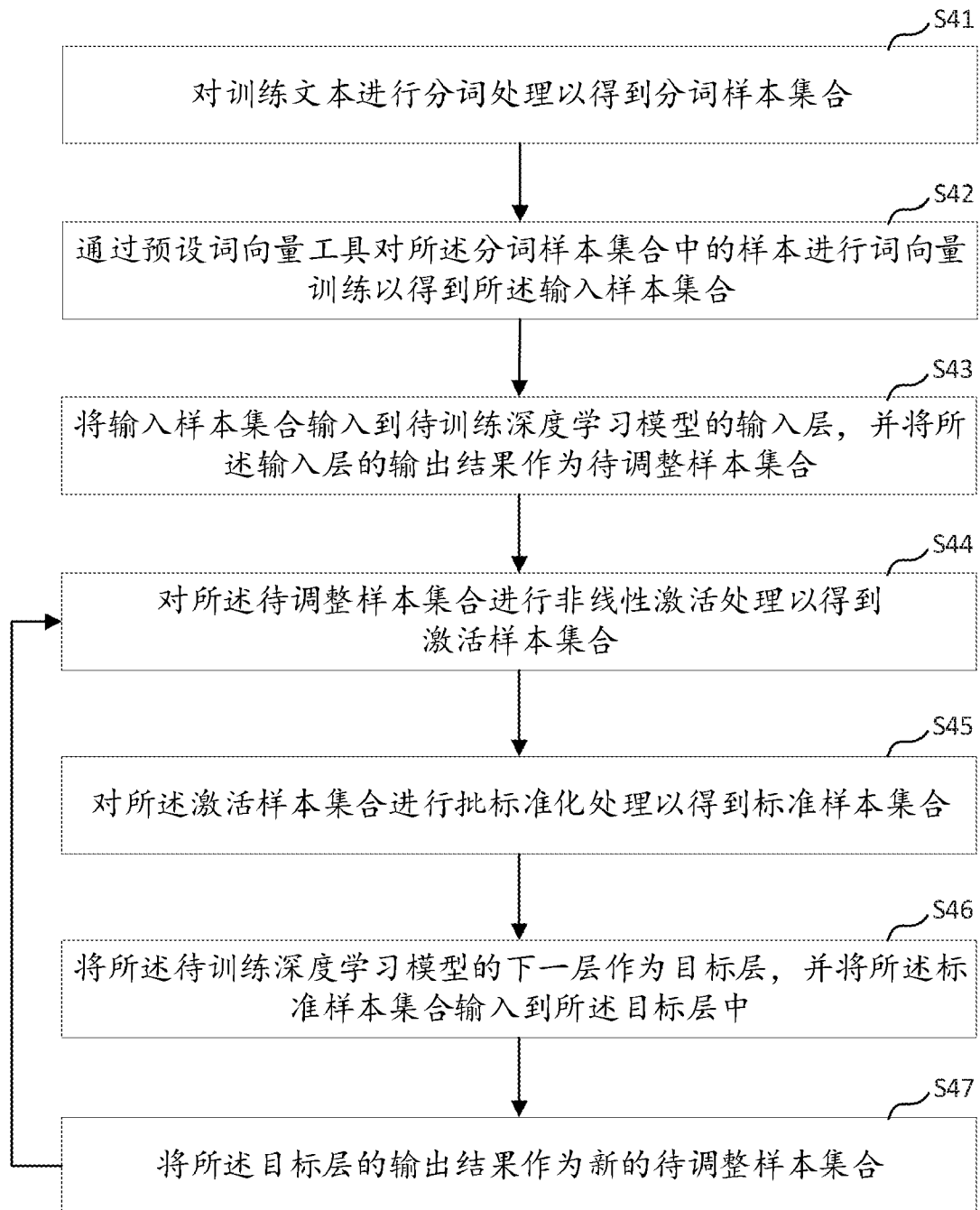


图 4

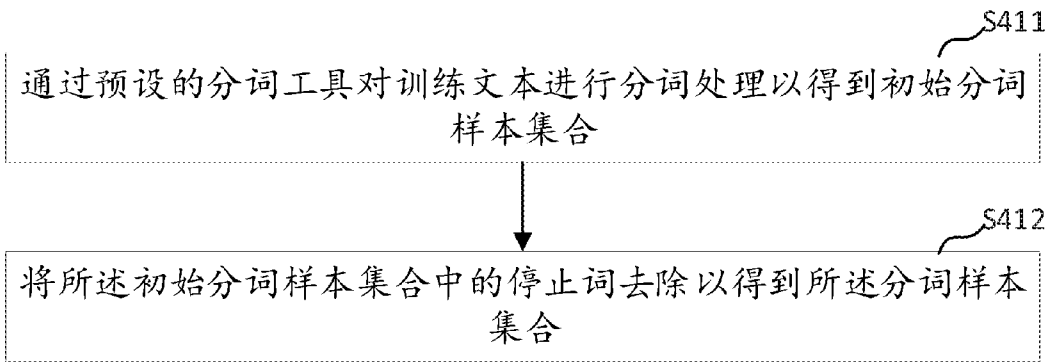


图 5

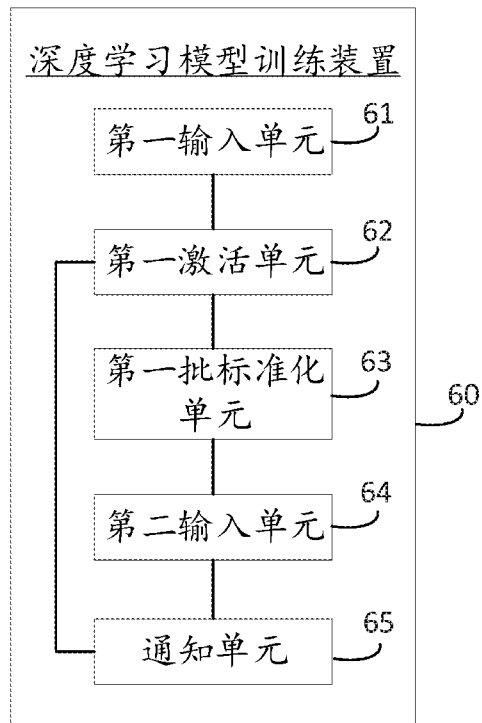


图 6

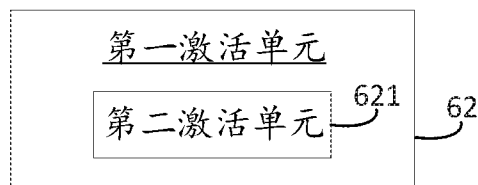


图 7

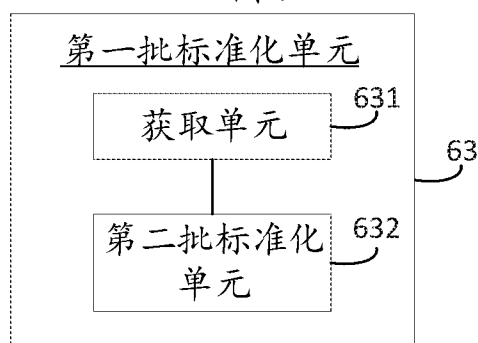


图 8

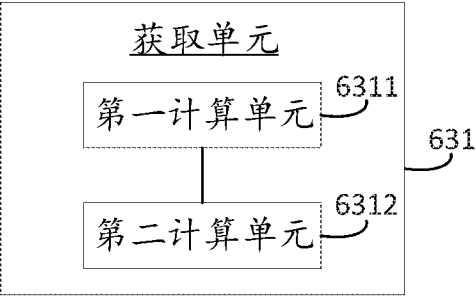


图 9

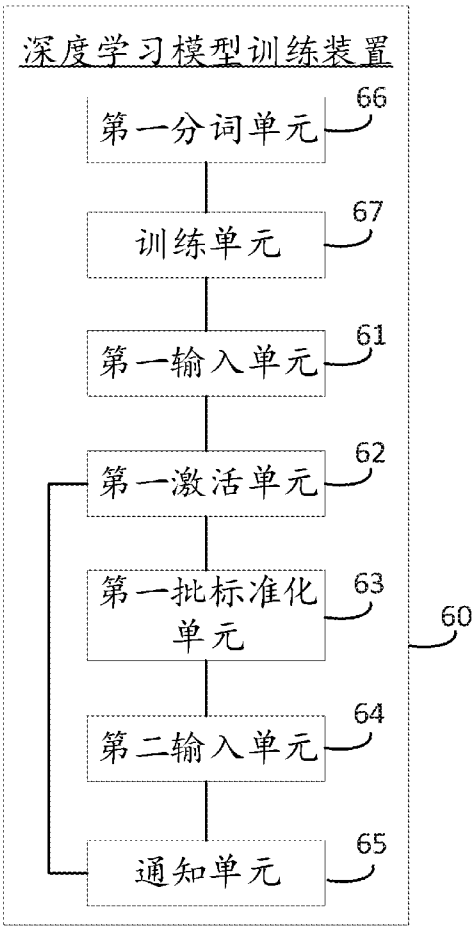


图 10

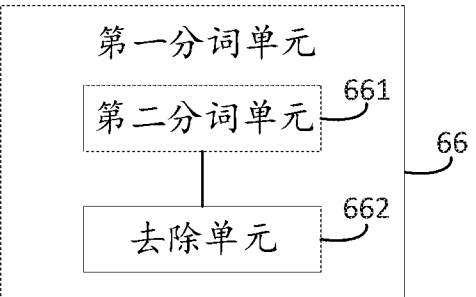


图 11

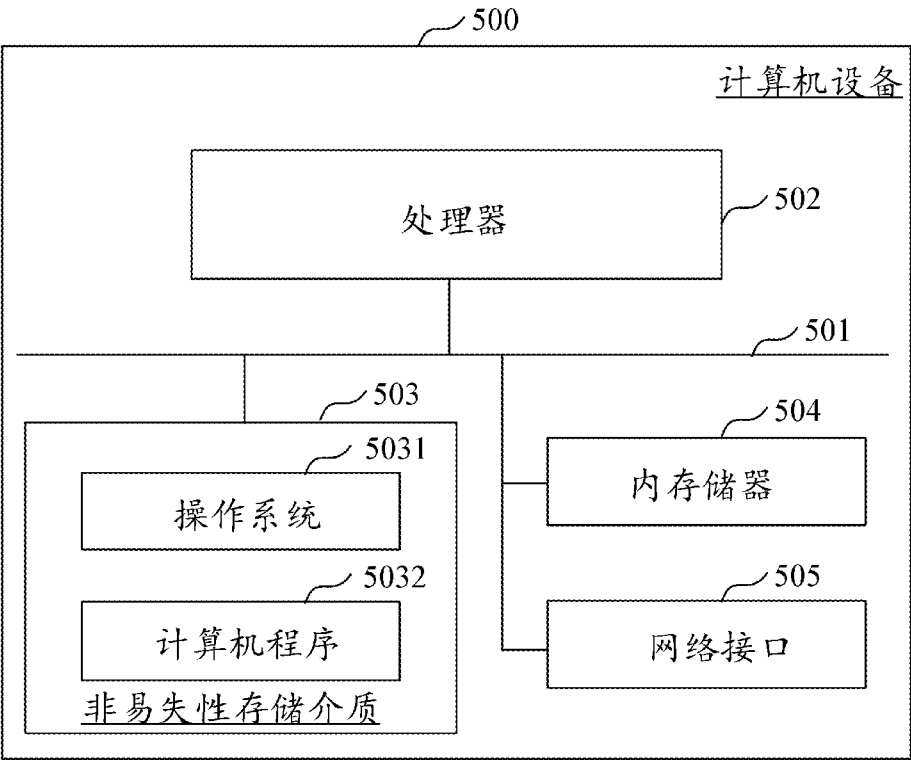


图 12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/117310

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)

CNABS, CNTXT, CNKI, VEN, 深度学习, 神经网络, 模型, 训练, 样本, 集合, 激活, 标准, 层, 迭代, 调整, 更新, 递归, 收敛, deep learning, neural network, model, train+, sample, set, activat+, standard, layer, iteration, adjust+, updat+, recurs+, converg+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109886402 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 14 June 2019 (2019-06-14) description, paragraphs 4-163	1-20
X	CN 108334943 A (ZHEJIANG UNIVERSITY) 27 July 2018 (2018-07-27) description, paragraphs 5-33	1-20
A	CN 108898218 A (ALIBABA GROUP HOLDING LIMITED) 27 November 2018 (2018-11-27) entire document	1-20
A	CN 108734193 A (HEFEI LINTU INFORMATION TECHNOLOGY CO., LTD.) 02 November 2018 (2018-11-02) entire document	1-20
A	CN 108959265 A (SHENZHEN NIUDINGFENG TECHNOLOGY CO., LTD.) 07 December 2018 (2018-12-07) entire document	1-20
A	KR 20010095960 A (HYUNDAI IND DEV & CONSTR CO., LTD) 07 November 2001 (2001-11-07) entire document	1-20



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

05 February 2020

Date of mailing of the international search report

18 February 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/117310

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	109886402	A	14 June 2019	None	
CN	108334943	A	27 July 2018	None	
CN	108898218	A	27 November 2018	None	
CN	108734193	A	02 November 2018	None	
CN	108959265	A	07 December 2018	None	
KR	20010095960	A	07 November 2001	None	

国际检索报告

国际申请号

PCT/CN2019/117310

A. 主题的分类 G06N 3/08 (2006.01) i 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		
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06N 包含在检索领域中的除最低限度文献以外的检索文献 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)) CNABS, CNTXT, CNKI, VEN, 深度学习, 神经网络, 模型, 训练, 样本, 集合, 激活, 标准, 层, 迭代, 调整, 更新, 递归, 收敛, deep learning, neural network, model, train+, sample, set, activat+, standard, layer, iteration, adjust+, updat+, recurs+, converg+		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 109886402 A (平安科技深圳有限公司) 2019年 6月 14日 (2019 - 06 - 14) 说明书第4-163段	1-20
X	CN 108334943 A (浙江大学) 2018年 7月 27日 (2018 - 07 - 27) 说明书第5-33段	1-20
A	CN 108898218 A (阿里巴巴集团控股有限公司) 2018年 11月 27日 (2018 - 11 - 27) 全文	1-20
A	CN 108734193 A (合肥麟图信息科技有限公司) 2018年 11月 2日 (2018 - 11 - 02) 全文	1-20
A	CN 108959265 A (深圳市牛鼎丰科技有限公司) 2018年 12月 7日 (2018 - 12 - 07) 全文	1-20
A	KR 20010095960 A (HYUNDAI IND DEV & CONSTR CO., LTD) 2001年 11月 7日 (2001 - 11 - 07) 全文	1-20
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “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) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “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 “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 “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 “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 “&” 同族专利的文件		
国际检索实际完成的日期 2020年 2月 5日		国际检索报告邮寄日期 2020年 2月 18日
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 史江峰 电话号码 62089926

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/117310

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	109886402	A	2019年 6月 14日	无	
CN	108334943	A	2018年 7月 27日	无	
CN	108898218	A	2018年 11月 27日	无	
CN	108734193	A	2018年 11月 2日	无	
CN	108959265	A	2018年 12月 7日	无	
KR	20010095960	A	2001年 11月 7日	无	