

用于训练学习模型的方法、装置和计算设备

技术领域

[01]本说明书的实施例涉及机器学习领域，具体地，涉及用于训练学习模型的方法、装置和计算设备。

5 背景技术

[02]深度学习源于对人工神经网络的研究，是近年来机器学习研究的一个新领域，并且在各个领域展现了广阔的应用前景。目前，在大多数深度学习方法中，深度学习模型（也可以被称为深度神经网络）通常是以批量学习方式训练的。在这种方式中，一般需要在学习任务开始前准备好训练数据集（也就是离线训练（Offline Training））。

10 [03]然而，在目前的很多应用中，数据可能是以流的形式先后到达的，那么批量学习方式对于这样的应用而言可能不太实用。

发明内容

[04]考虑到现有技术的上述问题，本说明书的实施例提供了用于训练学习模型的方法、装置和计算设备。

15 [05]一方面，本说明书的实施例提供了一种用于训练学习模型的方法，包括：接收当前流式样本数据；基于所述当前流式样本数据对当前深度学习模型进行训练，其中，浅层学习模型的参数被用作所述当前深度学习模型的初始化参数，所述浅层学习模型是基于与所述当前流式样本数据具有关联性的历史样本数据训练得到的。

[06]另一方面，本说明书的实施例提供了一种用于训练学习模型的装置，包括：接收单元，用于接收当前流式样本数据；训练单元，用于基于所述当前流式样本数据对当前深度学习模型进行训练，其中，浅层学习模型的参数被用作所述当前深度学习模型的初始化参数，所述浅层学习模型是基于与所述当前流式样本数据具有关联性的历史样本数据训练得到的。

20

[07]另一方面，本说明书的实施例提供了一种计算设备，包括：至少一个处理器；与所述至少一个处理器进行通信的存储器，其上存储有可执行指令，所述可执行指令在被所述至少一个处理器执行时使得所述至少一个处理器实现上述方法。

25

[08]可见,在该技术方案中,在基于当前流式样本数据对当前深度学习模型进行训练时,通过将训练好的浅层学习模型的参数用作当前深度学习模型的初始化参数,能够加快深度学习模型的收敛速度,从而能够高效地完成模型训练过程,而且也有利于提升深度学习模型的性能。

5 附图说明

[09]下文将以明确易懂的方式,通过对优选实施例的说明并结合附图来对本发明上述特性、技术特征、优点及其实现方式予以进一步说明,其中:

[10]图 1 是根据一个实施例的用于训练学习模型的方法的示意性流程图。

[11]图 2A 是根据一个实施例的用于浅层学习模型训练的示例性过程。

10 [12]图 2B 是根据一个实施例的用于深度学习模型训练的示例性过程。

[13]图 3 是根据一个实施例的用于训练学习模型的装置的示意性框图。

[14]图 4 是根据一个实施例的用于训练学习模型的计算设备的硬件结构图。

具体实施方式

15 [15]现在将参考各实施例讨论本文描述的主题。应当理解的是,讨论这些实施例仅是为了使得本领域技术人员能够更好地理解并且实现本文描述的主题,并非是对权利要求书中所阐述的保护范围、适用性或者示例的限制。可以在不脱离权利要求书的保护范围的情况下,对所讨论的元素的功能和排列进行改变。各个实施例可以根据需要,省略、替换或者添加各种过程或组件。

20 [16]目前,通常以批量学习方式训练深度学习模型。在批量学习方式中,一般需要在学习任务开始前准备好训练数据集,然后利用这些训练数据集来训练深度学习模型(即,离线训练)。然而,在一些应用中,数据可能是以流的形式先后到达的,那么批量学习方式可能不太适合这样的应用。另外,在批量学习方式中,提前准备好的训练数据集可能会占用较大的存储空间,那么对于某些存储空间受限的场景而言也是不太适合的。

25 [17]本说明书的实施例提出了一种基于流式样本数据来训练学习模型的技术方案。在本说明书中,流式样本数据通常可以包括由数据样本源持续生成的样本数据,例如 Web 应用程序生成的日志文件、网购数据、游戏玩家活动数据、社交网站信息数据等。流式样本数据也可以称为实时样本数据,其时间跨度通常在数百毫秒到数秒之间。基于流式

样本数据进行模型训练通常也可以认为是在线学习。

[18]具体而言，在本说明书的技术方案中，可以接收当前流式样本数据。然后，基于当前流式样本数据对当前深度学习模型进行训练。其中，浅层学习模型的参数可以被用作当前深度学习模型的初始化参数，而浅层学习模型可以是基于与当前流式样本数据具有关联性的历史样本数据训练得到的。

[19]可见，在该技术方案中，在基于当前流式样本数据对当前深度学习模型进行训练时，通过将训练好的浅层学习模型的参数用作当前深度学习模型的初始化参数，能够加快深度学习模型的收敛速度，从而能够高效地完成模型训练过程，而且也有利于提升深度学习模型的性能。

[20]此外，在该技术方案中，基于流式样本数据来对深度学习模型进行训练，可以不需要提前准备好训练数据集，从而能够有效地节省存储空间。

[21]在某些情况下，数据可能会表现出概念偏移，即数据的统计特性随着时间的推移以不可预见的方式变化的现象。那么，通过基于流式样本数据来对深度学习模型进行训练，能够随着数据的变化来及时地调整深度学习模型，从而能够提升深度学习模型的预测效果，并且也具有具有良好的可扩展性。

[22]在本说明书中，浅层学习模型的结构可以比深度学习模型的结构简单。例如，浅层学习模型可以不具有隐藏层，或者其具有的隐藏层的层数比深度学习模型的隐藏层的层数少。

[23]例如，浅层学习模型可以是逻辑回归模型（Logistic Regression，LR），而深度学习模型可以具有一层或多层的隐藏层。在一些情况下，为了便于实现，可以在初始构建深度学习模型时，将其构建为具有一层的隐藏层。

[24]下面将结合具体实施例来详细地描述上述技术方案。

[25]图 1 是根据一个实施例的用于训练学习模型的方法的示意性流程图。

[26]如图 1 所示，在步骤 102 中，可以接收当前流式样本数据。

[27]在步骤 104 中，可以基于当前流式样本数据来对当前深度学习模型进行训练。

[28]浅层学习模型的参数可以被用作当前深度学习模型的初始化参数，而浅层学习模型可以是基于与当前流式样本数据具有关联性的历史样本数据训练得到的。

[29]例如，历史样本数据与当前流式样本数据之间的关联性可以表示历史样本数据与当

前流式样本数据具有一个或多个相同的样本特征。

[30]在一个实施例中,上述历史样本数据可以是当前流式样本数据之前的历史流式样本数据。例如,历史流式样本数据可能包括当前流式样本数据之前的一批或多批流式样本数据。在这种情况下,浅层学习模型可以是基于历史流式样本数据在线训练得到的。

- 5 [31]在一个实施例中,上述历史样本数据可以是与当前流式样本数据具有关联性的离线样本数据。例如,离线样本数据可以是与当前流式样本数据具有一个或多个相同的样本特征的数据。在这种情况下,浅层学习模型可以是基于离线样本数据离线训练得到的。

[32]可见,在本说明书中,浅层学习模型可以通过在线方式或离线方式来训练得到,这样能够适应不同的应用需求。

- 10 [33]在一个实施例中,在将浅层学习模型的参数用作当前深度学习模型的初始化参数时,可以将浅层学习模型的各层的参数用作当前深度学习模型的对应各层的初始化参数。比如,可以采用一一映射的方式将浅层学习模型的各层的参数映射到当前深度学习模型的对应各层上。这样,不仅有利于加快深度学习模型的收敛速度,缩短训练时间,而且有利于提升深度学习模型的性能。此外,当前深度学习模型的其余各层的参数可以采用随
15 机化的方式来初始化。

[34]在一个实施例中,在步骤 104 中,可以利用在线梯度下降(Online Gradient Descent)等本领域中各种适用的方法来对当前深度学习模型进行训练。

[35]在一个实施例中,在对当前深度学习模型训练结束而得到训练后的深度学习模型之后,可以将训练后的深度学习模型的性能和当前深度学习模型的性能进行比较。

- 20 [36]例如,用于评估性能的指标可以包括曲线下面积(Area Under Curve, AUC)、准确率、覆盖率、F1 Score 等各种适用的指标。本说明书对此不作限定。

[37]如果训练后的深度学习模型与当前深度学习模型相比性能得到提升,比如,训练后的深度学习模型的 AUC 高于当前深度学习模型的 AUC,则可以将训练后的深度学习模型作为最新深度学习模型。

- 25 [38]如果训练后的深度学习模型与当前深度学习模型相比性能未得到提升,比如,训练后的深度学习模型的 AUC 低于当前深度学习模型的 AUC,或者这二者的 AUC 基本上接近,则可以考虑对当前深度学习模型进行更新。

[39]例如,可以增加当前深度学习模型的隐藏层的层数,得到增加层数的深度学习模型。

比如,可以将当前深度学习模型的隐藏层的层数增加一层或多层,这可以根据实际需求等来决定。

[40]之后,可以基于当前流式样本数据对增加层数后的深度学习模型进行训练,得到新深度学习模型。

- 5 [41]在一个实施例中,浅层学习模型的参数可以被用作增加层数后的深度学习模型的初始化参数。这样,可以有助于加快增加层数后的深度学习模型的收敛速度,从而高效地得到新深度学习模型,而且也有利于提升新深度学习模型的性能。

[42]之后,可以基于新深度学习模型与当前深度学习模型的性能比较结果,来确定最新深度学习模型。

- 10 [43]例如,如果新深度学习模型与当前深度学习模型相比性能得到提升,则可以将新深度学习模型作为最新深度学习模型。如果新深度学习模型与当前深度学习模型相比性能未得到提升,则可以将当前深度学习模型作为最新深度学习模型。

[44]可见,通过上述方式,能够有效地选择当前最优的深度学习模型作为最新深度学习模型。

- 15 [45]在一个实施例中,可以直接将最新深度学习模型作为要应用的最新学习模型。

[46]或者,在另一实施例中,可以将最新深度学习模型和浅层学习模型进行加权,得到要应用的最新学习模型。

[47]例如,最新深度学习模型和浅层学习模型可以分别具有相应的权重。它们的权重可以是预先定义的,也可以是用户根据实际需求来定义的。比如,最新深度学习模型的权重可以为 70%,而浅层学习模型的权重可以为 30%。或者,最新深度学习模型的权重可以为 100%,而浅层学习模型的权重可以为 0。由此,可以适用于不同的应用场景。

- 20

[48]下面将结合具体例子来描述本说明书中的技术方案。应理解的是,这些例子仅是为了帮助本领域技术人员更好地理解各个实施例,而非限制其范围。

[49]图 2A 是根据一个实施例的用于浅层学习模型训练的示例性过程。

- 25 [50]如图 2A 所示,在步骤 202A 中,可以获取历史样本数据。

[51]在步骤 204A 中,可以基于历史样本数据进行浅层学习模型训练,以得到浅层学习模型。

[52]例如,历史样本数据可以是上述当前流式样本之前的一批或多批流式样本数据。在

这种情况下，可以在每次接收到一批流式样本数据时进行浅层学习模型训练。这样，可能需要一批或多批流式样本数据来在线地完成浅层学习模型训练，从而得到浅层学习模型。

5 [53]再例如，历史样本数据可以是与上述当前流式样本具有关联性的离线样本数据。在这种情况下，可以基于离线样本数据，以离线的方式进行浅层学习模型训练，从而得到浅层学习模型。

[54]可以理解的是，图 2A 示出的过程可以发生在深度学习模型训练过程（例如，图 2B 所示的过程）之前，以便将浅层学习模型的参数用作深度学习模型的初始化参数。

[55]图 2B 是根据一个实施例的用于深度学习模型训练的示例性过程。

10 [56]如图 2B 所示，在步骤 202B 中，可以接收当前流式样本数据。

[57]在步骤 204B 中，可以基于当前流式样本数据对当前深度学习模型进行训练，以得到训练后的深度学习模型。

15 [58]例如，在进行训练时，可以首先对当前深度学习模型进行初始化。比如，可以获取浅层学习模型的各层的参数，然后将浅层学习模型的各层的参数作为当前深度学习模型的对应各层的初始化参数。可以以随机化的方式来初始化当前深度学习模型的其余各层的参数。

[59]在步骤 206B 中，可以确定训练后的深度学习模型与当前深度学习模型相比性能是否得到提升。

20 [60]如果得到提升，则在步骤 208B 中，可以将训练后的深度学习模型作为最新深度学习模型。

[61]如果没有得到提升，则在步骤 210B 中，可以增加当前深度学习模型的隐藏层的层数。比如，可以将其增加一层或多层。

[62]在步骤 212B 中，可以对增加层数后的深度学习模型进行训练，从而得到新深度学习模型。

25 [63]例如，在对增加层数后的深度学习模型进行训练时，可以将浅层学习模型的各层的参数用作增加层数后的深度学习模型的对应各层的初始化参数，而增加层数后的深度学习模型的其余各层的参数可以通过随机化的方式来初始化。

[64]在步骤 214B 中，可以确定新深度学习模型与当前深度学习模型相比性能是否得到

提升。

[65]如果得到提升，则在步骤 216B 中，可以将新深度学习模型作为最新深度学习模型。

[66]如果没有得到提升，则在步骤 218B 中，可以将当前深度学习模型作为最新深度学习模型。

5 [67]应当理解的是，图 2A 和图 2B 示出的各个步骤的次序仅是示例性的。根据实际应用或不同的设计逻辑，这些步骤的次序可以相应地改变。

[68]图 3 是根据一个实施例的用于训练学习模型的装置的示意性框图。

[69]如图 3 所示，装置 300 可以包括接收单元 302 和训练单元 304。

10 [70]接收单元 302 可以接收当前流式样本数据。训练单元 304 可以基于所述当前流式样本数据对当前深度学习模型进行训练。其中，浅层学习模型的参数可以被用作当前深度学习模型的初始化参数，浅层学习模型可以是基于与当前流式样本数据具有关联性的历史样本数据训练得到的。

[71]可见，在该技术方案中，通过将训练好的浅层学习模型的参数用作当前深度学习模型的初始化参数，这样在基于当前流式样本数据对当前深度学习模型进行训练时，能够
15 加快深度学习模型的收敛速度，从而能够高效地完成模型训练过程，而且也有利于提升深度学习模型的性能。

[72]在一个实施例中，历史样本数据可以是当前流式样本数据之前的历史流式样本数据。那么，浅层学习模型可以是基于历史流式样本数据在线训练得到的。

[73]在一个实施例中，历史样本数据可以是离线样本数据。那么，浅层学习模型可以是
20 基于离线样本数据离线训练得到的。

[74]在一个实施例中，装置 300 还可以包括评估单元 306。

[75]在训练单元 304 对当前深度学习模型训练结束，得到训练后的深度学习模型之后，评估单元 306 可以对训练后的深度学习模型的性能进行评估。

[76]如果训练后的深度学习模型与当前深度学习模型相比性能得到提升，则评估单元
25 306 可以将训练后的深度学习模型作为最新深度学习模型。

[77]如果训练后的深度学习模型与当前深度学习模型相比性能未得到提升，则训练单元 304 可以增加当前深度学习模型的隐藏层的层数，得到增加层数后的深度学习模型。训练单元 304 可以基于当前流式样本数据对增加层数后的深度学习模型进行训练，以得到

新深度学习模型。之后，评估单元 306 可以基于新深度学习模型与当前深度学习模型的性能比较结果，确定最新深度学习模型。

[78]在一个实施例中，浅层学习模型的参数可以被用作增加层数后的深度学习模型的初始化参数。

5 [79]在一个实施例中，如果新深度学习模型与当前深度学习模型相比性能得到提升，则评估单元 306 可以将新深度学习模型作为最新深度学习模型。

[80]如果新深度学习模型与当前深度学习模型相比性能未得到提升，则评估单元 306 可以将当前深度学习模型作为最新深度学习模型。

10 [81]在一个实施例中，装置 300 还可以包括加权单元 308。加权单元 308 可以将最新深度学习模型和浅层学习模型进行加权，以得到最新学习模型。

[82]装置 300 的各个单元可以执行图 1 至 2B 的方法实施例中的相应步骤，因此，为了描述的简洁，装置 300 的各个单元的具体操作和功能此处不再赘述。

15 [83]上述装置 300 可以采用硬件实现，也可以采用软件实现，或者可以通过软硬件的组合来实现。例如，装置 300 在采用软件实现时，其可以通过其所在设备的处理器将存储器（比如非易失性存储器）中对应的可执行指令读取到内存中运行来形成。

20 [84]图 4 是根据一个实施例的用于训练学习模型的计算设备的硬件结构图。如图 4 所示，计算设备 400 可以包括至少一个处理器 402、存储器 404、内存 406 和通信接口 408，并且至少一个处理器 402、存储器 404、内存 406 和通信接口 408 经由总线 410 连接在一起。至少一个处理器 402 执行在存储器 404 中存储或编码的至少一个可执行指令（即，上述以软件形式实现的元素）。

[85]在一个实施例中，在存储器 404 中存储的可执行指令在被至少一个处理器 402 执行时，使得计算设备实现以上结合图 1-2B 描述的各种过程。

[86]计算设备 400 可以采用本领域任何适用的形式来实现，例如，其包括但不限于台式计算机、膝上型计算机、智能电话、平板计算机、消费电子设备、可穿戴智能设备等等。

25 [87]本说明书的实施例还提供了一种机器可读存储介质。该机器可读存储介质可以存储有可执行指令，可执行指令在被机器执行时使得机器实现上面参照图 1-2B 描述的方法实施例的具体过程。

[88]例如，机器可读存储介质可以包括但不限于随机存取存储器（Random Access

Memory, RAM)、只读存储器(Read-Only Memory, ROM)、电可擦除可编程只读存储器(Electrically-Erasable Programmable Read-Only Memory, EEPROM)、静态随机存取存储器(Static Random Access Memory, SRAM)、硬盘、闪存等等。

[89]应当理解的是,本说明书中的各个实施例均采用递进的方式来描述,各个实施例之间相同或相似的部分相互参见即可,每个实施例重点说明的都是与其它实施例的不同之处。例如,对于上述关于装置的实施例、关于计算设备的实施例以及关于机器可读存储介质的实施例而言,由于它们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。

[90]上文对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。10 在一些情况下,在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外,在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中,多任务处理和并行处理也是可以的或者可能是有利的。

[91]应当理解的是,对于本领域普通技术人员来说,对本说明书中的实施例进行的各种15 修改将是显而易见的,并且可以在不脱离权利要求书的保护范围的情况下,将本文所定义的一般性原理应用于其它变型。

权利要求书

1、一种用于训练学习模型的方法，包括：

接收当前流式样本数据；

5 基于所述当前流式样本数据对当前深度学习模型进行训练，其中，浅层学习模型
参数被用作所述当前深度学习模型的初始化参数，所述浅层学习模型是基于与所述当前
流式样本数据具有关联性的历史样本数据训练得到的。

2、根据权利要求 1 所述的方法，其中，所述历史样本数据是所述当前流式样本数
据之前的历史流式样本数据，所述浅层学习模型是基于所述历史流式样本数据在线训练
得到的。

10 3、根据权利要求 1 所述的方法，其中，所述历史样本数据是离线样本数据，所述
浅层学习模型是基于所述离线样本数据离线训练得到的。

4、根据权利要求 1 至 3 中任一项所述的方法，还包括：

在对所述当前深度学习模型训练结束，得到训练后的深度学习模型之后：

15 如果所述训练后的深度学习模型与所述当前深度学习模型相比性能得到提升，则将
所述训练后的深度学习模型作为最新深度学习模型；

如果所述训练后的深度学习模型与所述当前深度学习模型相比性能未得到提升，则
进行以下操作：

增加所述当前深度学习模型的隐藏层的层数，得到增加层数后的深度学习模型；

20 基于所述当前流式样本数据对所述增加层数后的深度学习模型进行训练，以得到新
深度学习模型；

基于所述新深度学习模型与所述当前深度学习模型的性能比较结果，确定最新深度
学习模型。

5、根据权利要求 4 所述的方法，其中，所述浅层学习模型的参数被用作所述增加
层数后的深度学习模型的初始化参数。

25 6、根据权利要求 4 或 5 所述的方法，其中，所述基于所述新深度学习模型与所述
训练后的深度学习模型的性能比较结果，确定最新深度学习模型，包括：

如果所述新深度学习模型与所述当前深度学习模型相比性能得到提升，则将所述新
深度学习模型作为最新深度学习模型；

30 如果所述新深度学习模型与所述当前深度学习模型相比性能未得到提升，则将所述
当前深度学习模型作为最新深度学习模型。

7、根据权利要求 4 至 6 中任一项所述的方法，还包括：

将所述最新深度学习模型和所述浅层学习模型进行加权，以得到最新学习模型。

8、一种用于训练学习模型的装置，包括：

接收单元，用于接收当前流式样本数据；

5 训练单元，用于基于所述当前流式样本数据对当前深度学习模型进行训练，其中，浅层学习模型的参数被用作所述当前深度学习模型的初始化参数，所述浅层学习模型是基于与所述当前流式样本数据具有关联性的历史样本数据训练得到的。

9、根据权利要求 8 所述的装置，其中，所述历史样本数据是所述当前流式样本数据之前的历史流式样本数据，所述浅层学习模型是基于所述历史流式样本数据在线训练得到的。

10 10、根据权利要求 8 所述的装置，其中，所述历史样本数据是离线样本数据，所述浅层学习模型是基于所述离线样本数据离线训练得到的。

11、根据权利要求 8 至 10 中任一项所述的装置，还包括评估单元，

其中，在所述训练单元对所述当前深度学习模型训练结束，得到训练后的深度学习模型之后：

15 如果所述训练后的深度学习模型与所述当前深度学习模型相比性能得到提升，则所述评估单元用于将所述训练后的深度学习模型作为最新深度学习模型；

如果所述训练后的深度学习模型与所述当前深度学习模型相比性能未得到提升，则：

20 所述训练单元还用于增加所述当前深度学习模型的隐藏层的层数，得到增加层数后的深度学习模型，并且基于所述当前流式样本数据对所述增加层数后的深度学习模型进行训练，以得到新深度学习模型；

所述评估单元用于基于所述新深度学习模型与所述当前深度学习模型的性能比较结果，确定最新深度学习模型。

12、根据权利要求 11 所述的装置，其中，所述浅层学习模型的参数被用作所述增加层数后的深度学习模型的初始化参数。

25 13、根据权利要求 11 或 12 所述的装置，其中，所述评估单元具体用于：

如果所述新深度学习模型与所述当前深度学习模型相比性能得到提升，则将所述新深度学习模型作为最新深度学习模型；

如果所述新深度学习模型与所述当前深度学习模型相比性能未得到提升，则将所述当前深度学习模型作为最新深度学习模型。

30 14、根据权利要求 11 至 13 中任一项所述的装置，还包括：

加权单元，用于将所述最新深度学习模型和所述浅层学习模型进行加权，以得到最

新学习模型。

15、一种计算设备，包括：

至少一个处理器；

与所述至少一个处理器进行通信的存储器，其上存储有可执行指令，所述可执行指

- 5 令在被所述至少一个处理器执行时使得所述至少一个处理器实现根据权利要求 1 至 7 中任一项所述的方法。