

使用待交换数据来执行机器学习的方法、设备和系统

5

技术领域

本公开的示例性实施例总体说来涉及人工智能的机器学习领域，具体说来，涉及一种使用待交换数据来执行机器学习的方法、设备和系统。

背景技术

10

随着大数据、云计算和人工智能等技术的发展，机器学习被广泛地用来从海量数据中挖掘其中隐藏的有用信息。

15

为了应用机器学习，通常需要从给定的训练数据集中学习出一个由特征及其参数组成的模型函数，当新数据到来时，可以针对该新数据应用这个模型。为了更好地学习或应用模型，往往需要来自各个方面的数据参与到模型的例如训练、测试或预估等过程中，这些数据可从相应的数据提供方购买或以其它方式获得。例如，银行在进行诸如获客、反欺诈等业务时，常常需要结合各种额外数据来执行机器学习，作为示例，所述额外数据可包括：移动互联网行为数据（诸如手机号码、通讯录数据、手机型号、生产厂商、硬件信息、经常使用的 APP、社交分享的内容等等）、移动设备通信数据（诸如手机号码、通讯录数据和通话记录）、移动运营商数据（诸如手机号码、上网浏览行为和 APP 使用行为）。

20

25

实践中，为了保证数据安全性和/或机器学习效果，会借助第三方通过利用各个数据提供方的数据来提供机器学习服务。相应地，各个数据提供方可将分别用相同密钥进行加密的数据提供给所述第三方，使得第三方能够在不获取数据明文的情况下完成数据的拼接，并基于拼接结果来进行机器学习。然而，应注意的是，在进行上述加密数据的交换时，很容易由于第三方与某一数据提供方之间的串通而泄露用户的隐私信息或其它不适于披露的信息，并且，交换的数据很容易在未经授权的情况下被再次使用或出售，难以从技术上保障数据的合法使用。例如，当互联网应用方面的数据提供方将自己的数据提供给第三方，以结合银行的数据来执行机器学习时，会担心自己用户的隐私被无端泄露，也会担心数据不经授权而被再次使用或出售。另一方面，银行同样会担心数据内容的泄露和/或数据的未授权使用。

30

发明内容

35

根据本公开的一个示例性实施例，提供了一种使用待交换数据来执行机

器学习的设备, 包括: 初次加密数据接收单元, 被配置为分别从第一数据提供方接收第一初次加密结果数据和从第二数据提供方接收第二初次加密结果数据, 其中, 第一初次加密结果数据由第一数据提供方通过使用第一加密函数对第一待交换数据进行加密而得, 第二初次加密结果数据由第二数据提供方通过使用第二加密函数对第二待交换数据进行加密而得, 其中, 第一待交换数据与第二待交换数据至少部分对应; 初次加密数据发送单元, 被配置为分别将第一初次加密结果数据发送到第二数据提供方和将第二初次加密结果数据发送到第一数据提供方; 再次加密数据接收单元, 被配置为分别从第一数据提供方接收第二再次加密结果数据和从第二数据提供方接收第一再次加密结果数据, 其中, 第一再次加密结果数据由第二数据提供方通过使用第二加密函数对第一初次加密结果数据进行加密而得, 第二再次加密结果数据由第一数据提供方通过使用第一加密函数对第二初次加密结果数据进行加密而得; 以及机器学习执行单元, 被配置为通过将第一再次加密结果数据与第二再次加密结果数据进行拼接来获取机器学习样本, 并基于机器学习样本来执行机器学习。

根据本公开的另一示例性实施例, 提供一种使用待交换数据来执行机器学习的方法, 包括: 分别从第一数据提供方接收第一初次加密结果数据和从第二数据提供方接收第二初次加密结果数据, 其中, 第一初次加密结果数据由第一数据提供方通过使用第一加密函数对第一待交换数据进行加密而得, 第二初次加密结果数据由第二数据提供方通过使用第二加密函数对第二待交换数据进行加密而得, 其中, 第一待交换数据与第二待交换数据至少部分对应; 分别将第一初次加密结果数据发送到第二数据提供方和将第二初次加密结果数据发送到第一数据提供方; 分别从第一数据提供方接收第二再次加密结果数据和从第二数据提供方接收第一再次加密结果数据, 其中, 第一再次加密结果数据由第二数据提供方通过使用第二加密函数对第一初次加密结果数据进行加密而得, 第二再次加密结果数据由第一数据提供方通过使用第一加密函数对第二初次加密结果数据进行加密而得; 以及通过将第一再次加密结果数据与第二再次加密结果数据进行拼接来获取机器学习样本, 并基于机器学习样本来执行机器学习。

根据本公开的另一示例性实施例, 提供一种执行机器学习的系统, 包括: 第一数据提供方, 被配置为通过使用第一加密函数对第一待交换数据进行加密而得到第一初次加密结果数据; 第二数据提供方, 被配置为通过使用第二加密函数对第二待交换数据进行加密而得到第二初次加密结果数据, 其中, 第一待交换数据与第二待交换数据至少部分对应; 机器学习执行设备, 被配置为分别从第一数据提供方接收第一初次加密结果数据和从第二数据提供方接收第二初次加密结果数据, 并分别将第一初次加密结果数据发送到第二数

据提供方和将第二初次加密结果数据发送到第一数据提供方，其中，第一数据提供方通过使用第一加密函数对第二初次加密结果数据进行加密而得到第二再次加密结果数据，第二数据提供方通过使用第二加密函数对第一初次加密结果数据进行加密而得到第一再次加密结果数据，并且，机器学习执行设备分别从第一数据提供方接收第二再次加密结果数据和从第二数据提供方接收第一再次加密结果数据，并通过将第一再次加密结果数据与第二再次加密结果数据进行拼接来获取机器学习样本，以基于机器学习样本来执行机器学习。

根据本公开的另一示例性实施例，提供一种执行机器学习的方法，包括：

10 由第一数据提供方通过使用第一加密函数对第一待交换数据进行加密而得到第一初次加密结果数据；由第二数据提供方通过使用第二加密函数对第二待交换数据进行加密而得到第二初次加密结果数据，其中，第一待交换数据与第二待交换数据至少部分对应；由机器学习执行设备分别从第一数据提供方接收第一初次加密结果数据和从第二数据提供方接收第二初次加密结果数据，

15 并分别将第一初次加密结果数据发送到第二数据提供方和将第二初次加密结果数据发送到第一数据提供方；由第一数据提供方通过使用第一加密函数对第二初次加密结果数据进行加密而得到第二再次加密结果数据；由第二数据提供方通过使用第二加密函数对第一初次加密结果数据进行加密而得到第一再次加密结果数据；由机器学习执行设备分别从第一数据提供方接收第二

20 再次加密结果数据和从第二数据提供方接收第一再次加密结果数据；以及由机器学习执行设备通过将第一再次加密结果数据与第二再次加密结果数据进行拼接来获取机器学习样本，以基于机器学习样本来执行机器学习。

根据本公开的另一示例性实施例，提供一种使用待交换数据来执行机器学习的计算机可读存储介质，其中，在所述计算机可读存储介质上记录有用于执行如上所述的任一方法的计算机程序。

根据本公开的另一示例性实施，提供一种使用待交换数据来执行机器学习的计算装置，包括存储部件和处理器，其中，存储部件中存储有计算机可执行指令集合，当所述计算机可执行指令集合被所述处理器执行时，执行如上所述的任一方法。

30 根据本公开示例性实施例的使用待交换数据来执行机器学习的方法、设备和系统，可安全可靠地使用外部数据来提供机器学习服务，不仅保证了数据的内容不被泄露，而且可防止数据在未经授权的情况下被再次使用。

附图说明

35 从下面结合附图对本公开示例性实施例的详细描述中，本公开示例性实施例的这些和/或其它方面和优点将变得更加清楚并更容易理解，其中：

图 1 示出根据本公开示例性实施例的使用待交换数据来执行机器学习的设备的框图；

图 2 示出根据本公开示例性实施例的使用待交换数据来执行机器学习的方法的流程图；以及

5 图 3 示出根据本公开示例性实施例的执行机器学习的系统使用数据提供方的数据来执行机器学习的示意图。

具体实施方式

为了使本领域技术人员更好地理解本公开的示例性实施例，下面结合附图和具体实施方式对本公开的示例性实施例作进一步详细说明。在此需要说明的是，在本公开中出现的“并且/或者”、“和/或”均表示包含三种并列的情况。例如“包括 A 和/或 B”表示如下三种并列的情况：(1) 包括 A；(2) 包括 B；(3) 包括 A 和 B。又例如“执行步骤一并且/或者步骤二”表示如下三种并列的情况：(1) 执行步骤一；(2) 执行步骤二；(3) 执行步骤一和步骤二。

图 1 示出根据本公开示例性实施例的使用待交换数据来执行机器学习的设备的框图。

这里，作为示例，执行机器学习的设备可相对独立地存在于各个数据提供方之外，而仅作为提供机器学习服务的第三方。相应地，所述设备可利用来自各个数据提供方的待交换数据（或进一步结合自身数据）来执行机器学习模型的训练、测试或应用，从而向外提供针对某个预测目标的机器学习模型和/或相应的预测结果，或者，所述设备可直接应用相应的机器学习预测结果，例如，基于机器学习预测结果来完成诸如获客等业务。

参照图 1，执行机器学习的设备可包括初次加密数据接收单元 100、初次加密数据发送单元 200、再次加密数据接收单元 300 和机器学习执行单元 400。这些单元可以是执行相应计算机程序步骤的虚拟单元，也可以是具有实体结构的物理单元，例如，运行有相应程序步骤的处理单元或在处理单元控制下执行操作以实现相应功能的模块。作为示例，这些单元之间可共享至少一部分共同的组件（例如，接口），甚至，某些虚拟单元的功能可合并单个实体中，例如，由单个实体来完成初次加密结果数据和/或再次加密结果数据的接收/发送处理。

具体说来，初次加密数据接收单元 100 被配置为分别从第一数据提供方接收第一初次加密结果数据和从第二数据提供方接收第二初次加密结果数据，其中，第一初次加密结果数据由第一数据提供方通过使用第一加密函数对第一待交换数据进行加密而得，第二初次加密结果数据由第二数据提供方通过使用第二加密函数对第二待交换数据进行加密而得，其中，第一待交换数据

与第二待交换数据至少部分对应。

这里，作为示例，初次加密数据接收单元 100 可借助网络（例如，云服务网络）从第一数据提供方和第二数据提供方分别接收它们各自生成的初次加密结果数据；或者，初次加密数据接收单元 100 可通过直接或经由中间设备与各个数据方连接而接收各个初次加密结果数据。这里，每个数据提供方均具有各自的数据资源，这些数据之间的至少一部分具有对应性，例如，这些数据提供方可分别具有关于特定用户的银行数据、移动运营商数据、互联网数据、资产数据、信用数据等。相应地，第一数据提供方和第二数据提供方可分别针对至少部分对应的第一待交换数据和第二待交换数据进行初次加密。这里，第一数据提供方可使用第一加密函数对第一待交换数据进行初次加密，而第二数据提供方可使用第二加密函数对第二待交换数据进行初次加密，作为示例，第一加密函数和第二加密函数分别是第一数据提供方和第二数据提供方私有的不为其他方所知的可交换函数。

初次加密数据发送单元 200 被配置为分别将第一初次加密结果数据发送到第二数据提供方和将第二初次加密结果数据发送到第一数据提供方。

这里，初次加密数据发送单元 200 可将由初次加密数据接收单元 100 所接收的初次加密结果数据以互换的方式分别发送到各个数据提供方。作为示例，初次加密数据发送单元 200 可按照与接收初次加密结果数据的相同路径反向地发送初次加密结果数据，在这种情况下，初次加密数据发送单元 200 可与初次加密数据接收单元 100 集成在单个实体中，该实体被配置为针对不同的传输对象和传输方向进行操作。

再次加密数据接收单元 300 分别从第一数据提供方接收第二再次加密结果数据和从第二数据提供方接收第一再次加密结果数据，其中，第一再次加密结果数据由第二数据提供方通过使用第二加密函数对第一初次加密结果数据进行加密而得，第二再次加密结果数据由第一数据提供方通过使用第一加密函数对第二初次加密结果数据进行加密而得。

这里，第一数据提供方在接收到由初次加密数据发送单元 200 发送的第二初次加密结果数据之后，使用其私有的第一加密函数对所述第二初次加密结果数据进行再次加密，而第二数据提供方在接收到由初次加密数据发送单元 200 发送的第一初次加密结果数据之后，使用其私有的第二加密函数对所述第一初次加密结果数据进行再次加密。通过上述方式，第一数据提供方可获得第二再次加密结果数据，第二数据提供方可获得第一再次加密结果数据。

相应地，再次加密数据接收单元 300 可分别从各个数据提供方接收其产生的再次加密结果数据。作为示例，再次加密数据接收单元 300 可按照与接收初次加密结果数据的相同路径来接收再次加密结果数据，在这种情况下，再次加密数据接收单元 300 可与初次加密数据接收单元 100 集成在单个实体

中，该实体被配置为针对不同的接收对象进行操作。此外，作为示例，初次加密数据接收单元 100、初次加密数据发送单元 200 和再次加密数据接收单元 300 三者可集成在单个实体（例如，收发器）中，该实体被配置为针对不同的传输对象和传输方向来执行相应的数据发送/接收处理。

- 5 机器学习执行单元 400 被配置为通过将第一再次加密结果数据与第二再次加密结果数据进行拼接来获取机器学习样本，并基于机器学习样本来执行机器学习。

具体说来，机器学习执行单元 400 可首先基于第一再次加密结果数据和第二再次加密数据来生成机器学习样本。这里，作为示例，机器学习执行单元 400 除了基于各个数据提供方的待交换数据之间的对应性（例如，标识信息）将第一再次加密结果数据和第二再次加密结果数据两者进行拼接之外，还可进一步拼接其他对应的数据（例如，执行机器学习的设备自身所具有的数据）。如上所述，各个数据提供方的待交换数据描述了对象在某些方面的属性或针对某个预测目标的标记（label），相应地，机器学习执行单元 400 可针对各个标识信息分别生成包括相应的属性信息和/或标记信息的拼接数据记录，并可进一步通过对这些拼接数据记录执行诸如特征抽取等特征处理而得到相应的机器学习样本。作为示例，机器学习执行单元 400 可在获取机器学习的训练样本之后，批量地使用所述训练样本来训练机器学习模型，可选地，在训练机器学习模型的过程中，还可进一步获取用来衡量模型训练成果的测试样本以测试训练的模型。作为另一示例，在得到机器学习模型（例如，训练出机器学习模型）之后，机器学习执行单元 400 可获取将对其进行预估的预测样本，以利用机器学习模型针对预测样本给出关于预测目标的预测结果，可选地，在得到预测结果之后，机器学习执行单元 400 可进一步应用这样的预测结果，例如，基于预测结果来执行诸如获客等业务。

- 25 如上所述，机器学习执行单元 400 可执行机器学习模型的训练、测试和/或预测，从而向外提供机器学习模型和/或预测结果，并可选地进一步应用预测结果。

由此可见，图 1 所示的执行机器学习的设备可利用外部数据来提供机器学习服务，不仅确保了各个数据提供方的数据内容的安全性，而且避免了数据在未经授权的情况下被使用。

图 2 示出根据本公开示例性实施例的使用待交换数据来执行机器学习的方法的流程图。作为示例，图 2 所示的方法可由图 1 所示的设备来执行，也可由其他计算装置来执行，例如，可通过运行相应的计算机程序来执行所述方法。

- 35 参照图 2，在步骤 S100 中，分别从第一数据提供方接收第一初次加密结果数据和从第二数据提供方接收第二初次加密结果数据，其中，第一初次加

密结果数据由第一数据提供方通过使用第一加密函数对第一待交换数据进行加密而得，第二初次加密结果数据由第二数据提供方通过使用第二加密函数对第二待交换数据进行加密而得，其中，第一待交换数据与第二待交换数据至少部分对应。

5 这里，第一数据提供方和第二数据提供方分别具有将被交换到第三方以进行机器学习的一部分数据。而且，第一数据提供方所具有的第一待交换数据与第二数据提供方所具有的第二待交换数据之间至少部分地具有对应性，即，两者所针对的至少一部分对象具有一致性。这里，第一待交换数据和第二待交换数据均可具有一条或多条数据记录，每条数据记录可具有各自的标识信息，借助该标识信息可在不同的待交换数据集之间完成至少一部分具有相同标识信息的数据记录的拼接。此外，来自不同数据提供方的数据记录可承载着对象在某些方面的属性或针对预测目标的标记。作为示例，第一待交换数据之中的每条第一待交换数据记录可至少包括标识信息和属性信息，第二待交换数据之中的每条第二待交换数据记录可至少包括标识信息和关于机器学习目标的标记信息。此外，第二待交换数据也可进一步包括一些属性信息。在这种情况下，第二数据提供方会希望借助第一数据提供方的属性信息更好地挖掘出关于机器学习目标的规律。

10 相应地，可从第一数据提供方接收第一初次加密结果数据，并从第二数据提供方接收第二初次加密结果数据。这里，可同时或按照任何顺序异步地接收第一初次加密结果数据和第二初次加密结果数据。具体说来，第一初次加密结果数据由第一数据提供方通过使用第一加密函数对第一待交换数据进行加密而得，第二初次加密结果数据由第二数据提供方通过使用第二加密函数对第二待交换数据进行加密而得。这里，作为示例，第一加密函数为第一数据提供方的私有函数，第二加密函数为第二数据提供方的私有函数，并且，
25 第一加密函数和第二加密函数之间构成单向可交换私有函数。可选地，第一加密函数可以是具有第一私有素数的第一幂函数，第二加密函数可以是具有第二私有素数的第二幂函数，从而进一步确保加密结果无法被破解。

30 接下来，在步骤 S200 中，分别将第一初次加密结果数据发送到第二数据提供方和将第二初次加密结果数据发送到第一数据提供方。这里，可在接收到第一初次加密结果数据之后，将接收的第一初次加密结果数据发送到第二数据提供方，并且，在接收到第二初次加密结果数据之后，将接收的第二初次加密结果数据发送到第一数据提供方。应注意，本公开的示例性实施例对于转发初次加密结果数据至另一方的时机和顺序不进行任何限制。

35 然后，在步骤 S300 中，分别从第一数据提供方接收第二再次加密结果数据和从第二数据提供方接收第一再次加密结果数据，其中，第一再次加密结果数据由第二数据提供方通过使用第二加密函数对第一初次加密结果数据进

行加密而得，第二再次加密结果数据由第一数据提供方通过使用第一加密函数对第二初次加密结果数据进行加密而得。

这里，第一数据提供方在接收到第二初次加密结果数据之后，通过使用其所拥有的第一加密函数对第二初次加密结果数据进行再次加密而得到第二再次加密结果数据，而第二数据提供方在接收到第一初次加密结果数据之后，通过使用其所拥有的第二加密函数对第一初次加密结果数据进行再次加密而得到第一再次加密结果数据。

相应地，在该步骤中，可从第一数据提供方接收第二再次加密结果数据，并从第二数据提供方接收第一再次加密结果数据。这里，可同时或按照任何顺序异步地接收第一再次加密结果数据和第二再次加密结果数据。

在步骤 S400 中，通过将第一再次加密结果数据与第二再次加密结果数据进行拼接来获取机器学习样本，并基于机器学习样本来执行机器学习。

这里，由于第一再次加密结果数据所依据的第一待交换数据和第二再次加密结果数据所依据的第二待交换数据之间至少部分具有对应性，因此，可通过将第一再次加密结果数据和第二再次加密结果数据进行拼接来得到扩展了属性信息的拼接数据记录，作为示例，该拼接数据记录还可额外包括其他的信息（例如，来自执行机器学习的设备自身所持有的数据记录中的属性信息等）。在获得机器学习样本之后，可执行相应的机器学习处理，例如，基于机器学习训练样本来训练机器学习模型；基于机器学习测试样本来控制模型训练的进程；基于机器学习预测样本来应用机器学习模型执行预估服务。此外，在该步骤中，还可直接应用机器学习模型的预测结果，例如，在获客业务中，针对预测出的潜在客户进行推广活动等。也就是说，在该步骤中，机器学习样本可以是机器学习训练样本、机器学习测试样本或机器学习预测样本，相应地，可基于机器学习样本来训练机器学习模型、测试机器学习模型或使用机器学习模型来进行预测。

可以看出，在根据本公开示例性实施例的使用待交换数据来执行机器学习的方法中，数据提供方全程仅使用了自己的私密函数来执行加密，而该私密函数对于其他方而言均为秘密。并且，机器学习服务的提供方仅能够接触到加密后的结果数据，且不同数据提供方的加密函数彼此独立且互相隐秘。在这种情况下，基于外部数据来执行机器学习则能够确保数据的安全性，并限制了未授权使用数据的情况。

图 3 示出根据本公开示例性实施例的执行机器学习的系统使用数据提供方的数据来执行机器学习的示意图。

参照图 3，根据本公开示例性实施例的执行机器学习的系统可包括第一数据提供方、第二机器学习提供方和机器学习执行设备。在图 3 所示的流程中，“第一数据提供方”具体是第一数据提供方的数据提供设备，“第二数据

提供方”具体是第二数据提供方的数据提供设备。

在图 3 所示的系统中，第一数据提供方和第二数据提供方均具有各自的待交换数据。这里，“交换”是指为了扩展地进行数据挖掘而采取的共享行为，包括但不限于将数据从提供方发送到获取方的过程。这里，提供方是指待交换数据的提供者，可以是待交换数据的直接或间接来源；获取方是指期望获取待交换数据以基于获取的各方数据来执行机器学习的服务提供者。

在下面的描述中，为便于理解，可将下面情况作为应用示例而非限制性描述：第一数据提供方为互联网数据提供商，其拥有的数据描述了用户的网络浏览行为，而第二数据提供方为银行，其拥有的数据描述了用户是否成为银行客户的获客结果（即，标记），作为示例，银行的数据还可进一步包括用户的其他属性。应理解，获客业务仅仅作为示例，而非用于限制本公开的示例性实施里，实际上，本公开的示例性实施可应用于任何基于多方数据来执行机器学习的情况，例如，反欺诈、推荐等业务中。

具体说来，第一数据提供方被配置为通过使用第一加密函数对第一待交换数据进行加密而得到第一初次加密结果数据。具体说来，在步骤 S11 中，针对第一待交换数据 DATA1，第一数据提供方可使用其私有的加密函数 $h(x)$ 对其进行加密，以得到第一初次加密结果数据 $h(DATA1)$ 。作为示例，假设 DATA1 中的任何一条数据记录 X_n (n 为自然数) 可包括标识信息 kn 和至少一个属性信息 $fn1$ 、 $fn2$ 、 $fn3$... fnm (其中， m 为大于或等于 1 的整数)，相应地， $h(X_n) = h(kn) h(fn1) h(fn2) h(fn3) \dots h(fnm)$ 。作为示例， $h(x) = a * x \% p$ ，或者， $h(x) = x^a \% p$ ，其中， a 为第一数据提供方私有的大素数，而 p 为共享的大素数。

第二数据提供方被配置为通过使用第二加密函数对第二待交换数据进行加密而得到第二初次加密结果数据。具体说来，在步骤 S21 中，针对第二待交换数据 DATA2，第二数据提供方可使用其私有的加密函数 $g(x)$ (这里， $g(x)$ 与 $h(x)$ 构成单向可交换私有函数) 对其进行加密，以得到第二初次加密结果数据 $g(DATA2)$ 。作为示例，假设 DATA2 中的任何一条数据记录 Y_j (j 为自然数) 可包括标识信息 kj 和关于预测目标的标记信息 lj ，相应地， $g(Y_j) = g(kj) g(lj)$ 。作为示例， $g(x) = b * x \% p$ ，或者， $g(x) = x^b \% p$ ，其中， b 为第二数据提供方私有的大素数，而 p 为共享的大素数。这里，应注意，第二数据提供方所拥有的第二待交换数据记录除了标识信息和标记信息之外，也可包括其他的属性信息。

机器学习执行设备被配置为分别从第一数据提供方接收第一初次加密结果数据和从第二数据提供方接收第二初次加密结果数据，并分别将第一初次加密结果数据发送到第二数据提供方和将第二初次加密结果数据发送到第一数据提供方。具体说来，在步骤 S12 中，机器学习执行设备接收从第一数据

提供方发送的第一初次加密结果数据 h (DATA1), 并且, 在步骤 S22 中, 机器学习执行设备接收从第二数据提供方发送的第二初次加密结果数据 g (DATA2)。此后, 机器学习执行设备在步骤 S31 中将从第二数据提供方接收到的第二初次加密结果数据 g (DATA2) 发送到第一数据提供方, 并在步骤
5 S32 中将从第一数据提供方接收到的第一初次加密结果数据 h (DATA1) 发送到第二数据提供方。

接下来, 在步骤 S13 中, 第一数据提供方通过使用第一加密函数 $h(x)$ 对第二初次加密结果数据 g (DATA2) 进行加密而得到第二再次加密结果数据 $h(g(\text{DATA2}))$, 相应地, 在步骤 S23 中, 第二数据提供方通过使用第二
10 加密函数 $g(x)$ 对第一初次加密结果数据 h (DATA1) 进行加密而得到第一再次加密结果数据 $g(h(\text{DATA1}))$ 。

接着, 在步骤 S33 中, 机器学习执行设备接收从第一数据提供方发送的第二再次加密结果数据 $h(g(\text{DATA2}))$, 并且, 在步骤 S34 中, 机器学习执行设备接收从第二数据提供方发送的第一再次加密结果数据 $g(h(\text{DATA1}))$ 。

15 这里, 应注意, 本公开的示例性实施例不限制数据传输的路径, 例如, 可借由云服务, 在诸如公有云或私有云的网络部署下进行数据传输, 也可通过设备直接互联或经由中间媒介的互联来完成数据的传输。此外, 上述步骤的时间顺序也不受图 3 所示顺序的限制, 例如, 第一数据提供方和第二数据提供方执行加密的时间顺序不受限制, 而机器学习执行设备也可同时或异步地
20 与第一数据提供方和第二数据提供方之间传输数据。

最后, 在步骤 S35 中, 机器学习执行设备通过将第一再次加密结果数据 $g(h(\text{DATA1}))$ 与第二再次加密结果数据 $h(g(\text{DATA2}))$ 进行拼接来获取机器学习样本, 以基于机器学习样本来执行机器学习。这里, 作为示例, 可通过经过加密的标识信息来完成数据之间的拼接, 也就是说, 内容相同的标识
25 信息加密结果可代表对应的数据记录, 机器学习执行设备可将这样对应的数据记录进行拼接, 以得到具有额外的属性信息和/或标记信息的拼接数据记录。可选地, 通过针对这样的拼接数据记录执行特征抽取等特征处理, 可得到相应的机器学习样本, 从而进一步进行机器学习模型的训练、测试或预估。

应理解, 图 1 和图 3 中所示出的设备可被分别配置为执行特定功能的软件、硬件、固件或上述项的任意组合。例如, 这些设备及其组成部分可对应于专用的集成电路, 也可对应于纯粹的软件代码, 还可对应于软件与硬件相结合的单元或模块。

基于图 1-3 所公开的内容, 本公开的实施例中还提供一种数据提供设备, 包括至少一个计算装置和至少一个存储指令的存储装置, 其中, 所述指令在被所述至少一个计算装置运行时, 促使所述至少一个计算装置执行以下
35 步骤: 使用第一加密函数对第一待交换数据进行加密而得到第一初次加密结

果数据，将第一初次加密结果数据发送到机器学习执行设备，从机器学习执行设备接收第二初次加密结果数据，通过使用第一加密函数对第二初次加密结果数据进行加密而得到第二再次加密结果数据，将第二再次加密结果数据发送给机器学习执行设备；或者，使用第二加密函数对第二待交换数据进行加密而得到第二初次加密结果数据，将第二初次加密结果数据发送到机器学习执行设备，从机器学习执行设备接收第一初次加密结果数据，通过使用第二加密函数对第一初次加密结果数据进行加密而得到第一再次加密结果数据，将第一再次加密结果数据发送给机器学习执行设备。

本公开实施例提供的所述数据提供设备中，可选地，第一待交换数据之中的每条第一待交换数据记录至少包括标识信息和属性信息；第二待交换数据之中的每条第二待交换数据记录至少包括标识信息和关于机器学习目标的标记信息。

本公开实施例提供的所述数据提供设备中，可选地，第一加密函数为第一数据提供方的私有函数，第二加密函数为第二数据提供方的私有函数，并且，第一加密函数和第二加密函数之间构成单向可交换私有函数。

本公开实施例提供的所述数据提供设备中，可选地，第一加密函数为具有第一私有大素数的第一幂函数，第二加密函数为具有第二私有大素数的第二幂函数。

基于图 1-3 所公开的内容，本公开的实施例中还提供一种由计算装置执行的数据提供方法，包括：使用第一加密函数对第一待交换数据进行加密而得到第一初次加密结果数据，将第一初次加密结果数据发送到机器学习执行设备，从机器学习执行设备接收第二初次加密结果数据，使用第一加密函数对第二初次加密结果数据进行加密而得到第二再次加密结果数据，将第二再次加密结果数据发送给机器学习执行设备；或者，使用第二加密函数对第二待交换数据进行加密而得到第二初次加密结果数据，将第二初次加密结果数据发送到机器学习执行设备，从机器学习执行设备接收第一初次加密结果数据，使用第二加密函数对第一初次加密结果数据进行加密而得到第一再次加密结果数据，将第一再次加密结果数据发送给机器学习执行设备。

本公开实施例提供的所述数据提供方法中，可选地，第一待交换数据之中的每条第一待交换数据记录至少包括标识信息和属性信息；第二待交换数据之中的每条第二待交换数据记录至少包括标识信息和关于机器学习目标的标记信息。

本公开实施例提供的所述数据提供方法中，可选地，第一加密函数为第一数据提供方的私有函数，第二加密函数为第二数据提供方的私有函数，并且，第一加密函数和第二加密函数之间构成单向可交换私有函数。

本公开实施例提供的所述数据提供方法中，可选地，第一加密函数为具

有第一私有素数的第一幂函数，第二加密函数为具有第二私有素数的第二幂函数。

以上参照图 1 到图 3 描述了根据本公开示例性实施例的使用待交换数据来执行机器学习的设备、方法和系统。应理解，上述方法可通过记录在计算机可读存储介质上的程序来实现，相应地，根据本公开的示例性实施例，可提供一种使用待交换数据来执行机器学习的计算机可读存储介质，其中，在所述计算机可读存储介质上记录有用于执行以下方法步骤的计算机程序：(A) 分别从第一数据提供方接收第一初次加密结果数据和从第二数据提供方接收第二初次加密结果数据，其中，第一初次加密结果数据由第一数据提供方通过使用第一加密函数对第一待交换数据进行加密而得，第二初次加密结果数据由第二数据提供方通过使用第二加密函数对第二待交换数据进行加密而得，其中，第一待交换数据与第二待交换数据至少部分对应；(B) 分别将第一初次加密结果数据发送到第二数据提供方和将第二初次加密结果数据发送到第一数据提供方；(C) 分别从第一数据提供方接收第二再次加密结果数据和从第二数据提供方接收第一再次加密结果数据，其中，第一再次加密结果数据由第二数据提供方通过使用第二加密函数对第一初次加密结果数据进行加密而得，第二再次加密结果数据由第一数据提供方通过使用第一加密函数对第二初次加密结果数据进行加密而得；以及 (D) 通过将第一再次加密结果数据与第二再次加密结果数据进行拼接来获取机器学习样本，并基于机器学习样本来执行机器学习。

上述计算机可读存储介质中的计算机程序可在诸如客户端、主机、代理装置、服务器等计算机设备中部署的环境中运行，应注意，所述计算机程序还可用于执行除了上述步骤以外的附加步骤或者在执行上述步骤时执行更为具体的处理，这些附加步骤和进一步处理的内容已经参照图 1 到图 3 进行了描述，这里为了避免重复将不再进行赘述。

此外，本公开的示例性实施例还可以实现为计算装置，该计算装置包括存储部件和处理器，存储部件中存储有计算机可执行指令集合，当所述计算机可执行指令集合被所述处理器执行时，执行使用待交换数据来执行机器学习的方法。

具体说来，所述计算装置可以部署在服务器或客户端中，也可以部署在分布式网络环境中的节点装置上。此外，所述计算装置可以是 PC 计算机、平板装置、个人数字助理、智能手机、web 应用或其他能够执行上述指令集合的装置。

这里，所述计算装置并非必须是单个的计算装置，还可以是任何能够单独或联合执行上述指令（或指令集）的装置或电路的集合体。计算装置还可以是集成控制系统或系统管理器的一部分，或者可被配置为与本地或远程（例

如, 经由无线传输) 以接口互联的便携式电子装置。

在所述计算装置中, 处理器可包括中央处理器(CPU)、图形处理器(GPU)、可编程逻辑装置、专用处理器系统、微控制器或微处理器。作为示例而非限制, 处理器还可包括模拟处理器、数字处理器、微处理器、多核处理器、处理器阵列、网络处理器等。

根据本公开示例性实施例的使用待交换数据来执行机器学习的方法中所描述的某些操作可通过软件方式来实现, 某些操作可通过硬件方式来实现, 此外, 还可通过软硬件结合的方式来实现这些操作。

处理器可运行存储在存储部件之一中的指令或代码, 其中, 所述存储部件还可以存储数据。指令和数据还可经由网络接口装置而通过网络被发送和接收, 其中, 所述网络接口装置可采用任何已知的传输协议。

存储部件可与处理器集成为一体, 例如, 将RAM或闪存布置在集成电路微处理器等之内。此外, 存储部件可包括独立的装置, 诸如, 外部盘驱动、存储阵列或任何数据库系统可使用的其他存储装置。存储部件和处理器可在操作上进行耦合, 或者可例如通过I/O端口、网络连接等互相通信, 使得处理器能够读取存储在存储部件中的文件。

此外, 所述计算装置还可包括视频显示器(诸如, 液晶显示器)和用户交互接口(诸如, 键盘、鼠标、触摸输入装置等)。计算装置的所有组件可经由总线和/或网络而彼此连接。

本公开的实施例还提供一种包括至少一个计算装置和至少一个存储指令的存储装置的使用待交换数据来执行机器学习的设备, 其中, 所述指令在被所述至少一个计算装置运行时, 促使所述至少一个计算装置执行本公开任一实施例所描述的方法的步骤。例如执行以下步骤: 从第一数据提供方接收第一初次加密结果数据和从第二数据提供方接收第二初次加密结果数据; 将第一初次加密结果数据发送到第二数据提供方和将第二初次加密结果数据发送到第一数据提供方; 从第一数据提供方接收第二再次加密结果数据和从第二数据提供方接收第一再次加密结果数据; 以及通过将第一再次加密结果数据与第二再次加密结果数据进行拼接来获取机器学习样本, 并基于机器学习样本来执行机器学习。

根据本公开示例性实施例的使用待交换数据来执行机器学习的方法所涉及的操作可被描述为各种互联或耦合的功能块或功能示图。然而, 这些功能块或功能示图可被均等地集成为单个的逻辑装置或按照非确切的边界进行操作。

具体说来, 如上所述, 根据本公开示例性实施例的使用待交换数据来执行机器学习的计算装置可包括存储部件和处理器, 其中, 存储部件中存储有计算机可执行指令集合, 当所述计算机可执行指令集合被所述处理器执行时,

执行下述步骤：从第一数据提供方接收第一初次加密结果数据和从第二数据提供方接收第二初次加密结果数据，其中，第一初次加密结果数据由第一数据提供方通过使用第一加密函数对第一待交换数据进行加密而得，第二初次加密结果数据由第二数据提供方通过使用第二加密函数对第二待交换数据进行加密而得，其中，第一待交换数据与第二待交换数据至少部分对应；将第一初次加密结果数据发送到第二数据提供方和将第二初次加密结果数据发送到第一数据提供方；从第一数据提供方接收第二再次加密结果数据和从第二数据提供方接收第一再次加密结果数据，其中，第一再次加密结果数据由第二数据提供方通过使用第二加密函数对第一初次加密结果数据进行加密而得，第二再次加密结果数据由第一数据提供方通过使用第一加密函数对第二初次加密结果数据进行加密而得；以及通过将第一再次加密结果数据与第二再次加密结果数据进行拼接来获取机器学习样本，并基于机器学习样本来执行机器学习。

应注意，以上已经结合图 1 到图 3 描述了根据本公开示例性实施例的使用待交换数据来执行机器学习的各处理细节，这里将不再赘述计算装置执行各步骤时的处理细节。

以上已经描述了本公开的各示例性实施例，应理解，上述描述仅是示例性的，并非穷尽性的，并且本公开也不限于所披露的各示例性实施例。在不偏离本公开的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。因此，本公开的保护范围应该以权利要求的范围为准。