

大模型的训练方法、装置、设备及存储介质

5 本申请要求于 2024 年 03 月 05 日提交中国专利局、申请号为 202410252065.9、申请名称为“大模型的训练方法、装置、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

10 本申请涉及大数据技术领域，具体涉及一种大模型的训练方法、装置、设备及存储介质。

背景技术

15 随着大模型技术与大数据的进一步发展，重视数据隐私和安全已经成为了世界性的趋势，如何在保障数据和模型隐私的同时使得双方协作完成大模型训练与推理是一个焦点。
目前，一个训练完成的大模型通常含有十亿、百亿级别的参数量，需要算力以及私有数据对其进行微调。中小型企业通常具有本地化数据，可以用来微调大模型，模型持有方、数据持有方希望合作出一个大模型，但均不愿意泄露各自的内容给彼此，但是中小型企业的本地客户端的算力又存在局限。

20 因此，相关技术中，在对大模型训练过程中，存在数据安全性不高、算力利用不充分的技术问题。

发明内容

本申请提供一种大模型的训练方法、装置、设备及存储介质，从而解决在对大模型训练过程中，存在数据安全性不高、算力利用不充分的技术问题的问题。

25 第一方面，本申请提供一种大模型的训练方法，所述大模型包括头部适配器、模型主体和尾部适配器；其中所述头部适配器包括第一云端部分和第一本地部分，其中所述尾部适配器包括第二云端部分和第二本地部分；在训练所述大模型的过程中所述第一云端部分、所述模型主体和所述第二云端部分部署于云端，所述第一本地部分和所述第二本地部分部署于本地客户端；

30 所述方法包括：

本地客户端获取训练数据，并对所述训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据；

本地客户端将所述第一分片训练数据发送至云端；

云端将所述第一分片训练数据输入至所述第一云端部分进行运算，本地客户端将所述

第二分片训练数据输入至所述第一本地部分进行运算，以输出第一特征向量并发送至云端；

云端将所述第一特征向量输入至所述模型主体，以输出第二特征向量；

云端将所述第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量，并将第二分片特征向量发送至客户端；

- 5 云端将所述第一分片特征向量输入至所述第二云端部分进行运算，本地客户端将所述第二分片特征向量输入至所述第二本地部分进行运算，以输出所述训练数据的预测结果；

云端和本地客户端根据预设数量的训练数据的预测结果，迭代调整所述大模型至模型收敛。

- 本申请在训练大模型的过程中，将大模型拆分为三个部分，中间的模型主体 M0 以及
- 10 头部适配器 Adaptor Header 和尾部适配器 Adaptor Tail 两个 Adaptor 部分，头部适配器包括第一云端部分 AH1 和第一本地部分 AH2，尾部适配器包括第二云端部分 AT1 和第二本地部分 AT2，每个适配器分为两部分，一半位于云端 Cloud，一半位于本地客户端 Client，对大模型的训练数据为来自 Client 的行业私有化数据，将其分片进入后，分别进入云端和本地客户端进行第一次联合训练，经过第一次训练后，得到特征向量形式的第一特征向量，
- 15 再输入至云端进行全部训练数据的明文训练，云端的输出再分片处理分别在云端和本地客户端进行第二次联合训练，得到训练数据的预测结果，根据训练数据的预测结果可得到训练好的大模型。基于上述训练方法，在大模型与小模型的协作推理运算过程中模型与数据均不出本地，云端仅能获知分片后的部分数据，以及难以还原出原始输入的特征向量，保护了隐私，提高了数据的安全性，且仅将头尾两个 Adaptor 分散到云端和客户端，利用加密的客户端数据做 Adaptor 部分的模型训练，模型主体结构依然位于云端，利用 Cloud 算力进行训练与推理，保证了算力性能，同时有效利用 Adaptor 微调机制，Adaptor 参数量仅
- 20 占整个原始大模型百分之几或千分之几，能达到与全模型全参数微调几乎相同效果的特性，实现了在安全、充分利用算力的前提下对大模型的有效训练。

- 可选地，所述云端将所述第一分片训练数据输入至所述第一云端部分进行运算，本地
- 25 客户端将所述第二分片训练数据输入至所述第一本地部分进行运算，以输出第一特征向量并发送至云端，包括：

- 云端将所述第一分片训练数据输入至所述第一云端部分进行运算，本地客户端将所述第二分片训练数据输入至所述第一本地部分进行运算，且所述第一云端部分和所述第一本地部分进行运算的过程中通过多方安全计算的加密运算方式，经过所述第一云端部分和所述
- 30 所述第一本地部分的运算，以输出第一特征向量并发送至云端。

其中，本申请可基于多方安全计算共享运算参数，从而实现第一云端部分和第一本地

部分基于秘密分享协议的联合训练，通过多方安全计算将原始数据转化为难以还原的特征向量，进一步地提高了大模型训练的安全性及可行性。

可选地，所述云端将所述第一分片特征向量输入至所述第二云端部分进行运算，本地客户端将所述第二分片特征向量输入至所述第二本地部分进行运算，以输出所述训练数据的预测结果，包括：

云端将所述第一分片特征向量输入至所述第二云端部分进行运算，本地客户端将所述第二分片特征向量输入至所述第二本地部分进行运算，且所述第二云端部分和所述第二本地部分进行运算的过程中通过多方安全计算的加密运算方式，经过所述第二云端部分和所述第二本地部分的运算，以输出所述训练数据的预测结果。

其中，本申请对尾部适配器的运算也基于多方安全计算共享运算参数，从而实现第二云端部分和第二本地部分基于秘密分享协议的联合训练，进一步地提高了大模型训练的安全性及可行性。

可选地，所述多方安全计算包括 ABY 混合协议或秘密共享协议。

可选地，所述云端和本地客户端根据预设数量的训练数据的预测结果，迭代调整所述大模型至模型收敛，包括：

云端根据所述预测结果与所述训练数据对应的预设标签得到损失参数；云端根据所述损失参数调整所述大模型的第一云端部分和所述第二云端部分的模型参数；本地客户端根据所述损失参数，调整所述大模型的第一本地部分和所述第二本地部分的模型参数；本地客户端若确定所述大模型的预测误差大于预设阈值，则重新执行“本地客户端获取训练数据”的步骤以继续训练所述大模型，若确定所述大模型的预测误差小于或等于预设阈值，则停止迭代训练。

可选地，所述云端根据所述预测结果与所述训练数据对应的预设标签得到损失参数，包括：

云端将所述预测结果与所述预设标签输入至损失函数中，以输出所述损失参数。

可选地，所述损失函数的计算公式为：

$$L = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic})$$

其中，L 表示损失函数，M 表示类别数量，N 表示样本数量， y_{ic} 是符号函数，表示第 i 个样本的属于第 c 个类别的概率， p_{ic} 表示模型预测第 i 个样本属于第 c 个类别的概率，i 和 N 是任意正整数。

这里，本申请得到训练数据的预测结果后，再与预测标签计算损失，反馈整个模型网

络进行模型收敛计算，从而调整大模型的模型参数，以调整大模型直至满足要求，实现了对大模型的训练。

可选地，在所述云端根据所述预测结果与所述训练数据对应的预设标签得到损失参数，并根据所述损失参数调整所述大模型的第一云端部分和所述第二云端部分的模型参数之前，还包括：

本地客户端根据预设加密算法，对所述训练数据对应的预设标签进行加密处理，以得到加密标签；本地客户端将所述加密标签发送至所述云端。

这里，为了进一步保证在大模型训练过程中的数据安全，本申请本地客户端将训练数据对应的预设标签发送至云端时，首先对预设标签进行加密处理，云端无法获取原始的预设标签，进一步地提高了大模型训练的安全性。

可选地，所述云端根据所述预测结果与所述训练数据对应的预设标签得到损失参数，并根据所述损失参数调整所述大模型的第一云端部分和所述第二云端部分的模型参数，包括：

云端根据所述预设加密算法，对所述预测结果进行加密处理，以得到加密预设结果；云端根据所述加密预测结果与所述加密标签得到损失参数，并根据所述损失参数调整所述大模型的第一云端部分和所述第二云端部分的模型参数。

可选地，所述预设加密算法为同态加密算法。

这里，云端根据本地客户端对预设标签的预设加密算法对训练数据的预测结果进行加密处理，从而能够将得到的加密预测结果与加密标签进行运算处理，得到损失参数，进而在安全可靠的前提下对模型参数进行调整。

可选地，所述对所述训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据，包括：

对所述训练数据进行随机分片处理，以得到第一分片训练数据和第二分片训练数据。

其中，本申请在对训练数据进行分片处理时，可采用随机分片的方式，通过此方式，云端无法还原出原始的训练数据，进而充分保护了本地客户端的数据安全。

可选地，所述对所述训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据，包括：

根据第一预设秘密分片算法，对所述训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据。

其中，本申请在对训练数据进行分片处理时，可采用第一预设秘密分片算法的方式，通过此方式，云端无法还原出原始的训练数据，进而充分保护了本地客户端的数据安全。

可选地，所述云端将所述第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量，包括：

云端将所述第二特征向量进行随机分片处理，以得到第一分片特征向量和第二分片特征向量。

5 可选地，所述云端将所述第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量，包括：

云端根据第二预设秘密分片算法，将所述第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量。

10 这里，本申请云端可通过随机或预设算法对第二特征向量进行分片处理，对数据进行进一步地拆分，以实现安全可靠的云端与本地客户端的联合训练，进一步地提高了大模型训练的数据安全及数据隐私。

可选地，所述训练好的大模型的头部适配器和尾部适配器部署于本地客户端，模型主体部署于云端。

15 其中，本申请训练好的大模型的头部适配器和尾部适配器部署于本地，模型主体部署于云端，在数据的推理预测阶段也需要双方合作才能完成，任何一方都无法独立完成推理，本地客户端的数据权益得到保障，云端的模型以及算力权益得到保障，提高了大模型预测的安全性以及算力资源的利用率。

可选地，上述方法还包括：

20 本地客户端获取一条待预测数据；本地客户端将所述待预测数据输入至所述头部适配器，以输出第一预测向量；本地客户端将所述第一预测向量发送至云端；云端将所述第一预测向量输入至所述模型主体，以输出第二预测向量；云端将所述第二预测向量发送至本地客户端；本地客户端将所述第二预测向量输入至所述尾部适配器，以输出所述待预测数据的预测结果。

25 这里，本申请能够根据训练好的大模型进行推理预测，对于在本地客户端的待预测数据，首先通过本地客户端的头部适配器将待预测数据转化为相对于云端加密的第一预测向量，再运用云端的算力，对待预测数据进行充分推理预测，再通过本地客户端的尾部适配器输出待预测数据的预测结果，实现了安全、高效的数据推理预测。

可选地，所述第一预测向量的数据格式为加密格式。

其中，第一预测向量的加密格式进一步地保证了本地客户端数据的安全性。

30 第二方面，本申请提供了一种大模型的训练装置，所述大模型包括头部适配器、模型主体和尾部适配器；其中所述头部适配器包括第一云端部分和第一本地部分，其中所述尾

部适配器包括第二云端部分和第二本地部分；在训练所述大模型的过程中所述第一云端部分、所述模型主体和所述第二云端部分部署于云端，所述第一本地部分和所述第二本地部分部署于本地客户端；

所述装置包括：

- 5 第一处理模块，用于本地客户端获取训练数据，并对所述训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据；

第一发送模块，用于本地客户端将所述第一分片训练数据发送至云端；

- 第二处理模块，用于云端将所述第一分片训练数据输入至所述第一云端部分进行运算，本地客户端将所述第二分片训练数据输入至所述第一本地部分进行运算，以输出第一特征
10 向量并发送至云端；

第一输入模块，用于云端将所述第一特征向量输入至所述模型主体，以输出第二特征向量；

第三处理模块，用于云端将所述第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量，并将第二分片特征向量发送至客户端；

- 15 第四处理模块，用于云端将所述第一分片特征向量输入至所述第二云端部分进行运算，本地客户端将所述第二分片特征向量输入至所述第二本地部分进行运算，以输出所述训练数据的预测结果；

第五处理模块，用于云端和本地客户端根据预设数量的训练数据的预测结果，迭代调整所述大模型至模型收敛。

- 20 可选地，所述第二处理模块具体用于：

云端将所述第一分片训练数据输入至所述第一云端部分进行运算，本地客户端将所述第二分片训练数据输入至所述第一本地部分进行运算，且所述第一云端部分和所述第一本地部分进行运算的过程中通过多方安全计算的加密运算方式，经过所述第一云端部分和所述第一本地部分的运算，以输出第一特征向量并发送至云端。

- 25 可选地，所述第四处理模块具体用于：

云端将所述第一分片特征向量输入至所述第二云端部分进行运算，本地客户端将所述第二分片特征向量输入至所述第二本地部分进行运算，且所述第二云端部分和所述第二本地部分进行运算的过程中通过多方安全计算的加密运算方式，经过所述第二云端部分和所述第二本地部分的运算，以输出所述训练数据的预测结果。

- 30 可选地，所述多方安全计算包括 ABY 混合协议或秘密共享协议。

可选地，所述第五处理模块具体用于；

云端根据所述预测结果与所述训练数据对应的预设标签得到损失参数；云端根据所述损失参数调整所述大模型的第一云端部分和所述第二云端部分的模型参数；本地客户端根据所述损失参数，调整所述大模型的第一本地部分和所述第二本地部分的模型参数；本地客户端若确定所述大模型的预测误差大于预设阈值，则重新执行“本地客户端获取训练数据”的步骤以继续训练所述大模型，若确定所述大模型的预测误差小于或等于预设阈值，则停止迭代训练。

可选地，所述第五处理模块还具体用于：云端将所述预测结果与所述预设标签输入至损失函数中，以输出所述损失参数。

可选地，所述损失函数的计算公式为：

$$L = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic})$$

其中， L 表示损失函数， M 表示类别数量， N 表示样本数量， y_{ic} 是符号函数，表示第 i 个样本的属于第 c 个类别的概率， p_{ic} 表示模型预测第 i 个样本属于第 c 个类别的概率， i 和 N 是任意正整数。

可选地，在所述第五处理模块用于云端根据所述预测结果与所述训练数据对应的预设标签得到损失参数，并根据所述损失参数调整所述大模型的第一云端部分和所述第二云端部分的模型参数之前，上述装置还包括加密处理模块，用于：

本地客户端根据预设加密算法，对所述训练数据对应的预设标签进行加密处理，以得到加密标签；本地客户端将所述加密标签发送至所述云端。

可选地，所述第五处理模块具体用于：

云端根据所述预设加密算法，对所述预测结果进行加密处理，以得到加密预设结果；云端根据所述加密预测结果与所述加密标签得到损失参数，并根据所述损失参数调整所述大模型的第一云端部分和所述第二云端部分的模型参数。

可选地，所述预设加密算法为同态加密算法。

可选地，所述第一处理模块具体用于：

对所述训练数据进行随机分片处理，以得到第一分片训练数据和第二分片训练数据。

可选地，所述第一处理模块具体用于：

根据第一预设秘密分片算法，对所述训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据。

可选地，所述第三处理模块具体用于：

云端将所述第二特征向量进行随机分片处理，以得到第一分片特征向量和第二分片特

征向量。

可选地，所述第三处理模块具体用于：

云端根据第二预设秘密分片算法，将所述第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量。

- 5 可选地，所述训练好的大模型的头部适配器和尾部适配器部署于本地客户端，模型主体部署于云端。

可选地，上述装置还包括推理预测模块，用于：

- 本地客户端获取一条待预测数据；本地客户端将所述待预测数据输入至所述头部适配器，以输出第一预测向量；本地客户端将所述第一预测向量发送至云端；云端将所述第一
10 预测向量输入至所述模型主体，以输出第二预测向量；云端将所述第二预测向量发送至本地客户端；本地客户端将所述第二预测向量输入至所述尾部适配器，以输出所述待预测数据的预测结果。

可选地，所述第一预测向量的数据格式为加密格式。

第三方面，本申请提供一种大模型的训练设备，包括：至少一个处理器和存储器；

- 15 所述存储器存储计算机执行指令；

所述至少一个处理器执行所述存储器存储的计算机执行指令，使得所述至少一个处理器执行如上第一方面以及第一方面各种可能的设计所述的大模型的训练方法。

- 第四方面，本发明提供一种计算机可读存储介质，所述计算机可读存储介质中存储有计算机执行指令，当处理器执行所述计算机执行指令时，实现如上第一方面以及第一方面
20 各种可能的设计所述的大模型的训练方法。

第五方面，本发明提供一种计算机程序产品，包括计算机程序，所述计算机程序被处理器执行时，实现如上第一方面以及第一方面各种可能的设计所述的大模型的训练方法。

- 本申请提供的大模型的训练方法、装置、设备及存储介质，其中该方法对大模型的训练数据为来自 Client 的行业私有化数据，将其分片进入后，分别进入云端和本地客户端进行第一次联合训练，经过第一次训练后，得到特征向量形式的第一特征向量，再输入至云端进行全部训练数据的明文训练，云端的输出再分片处理分别在云端和本地客户端进行第二次联合训练，得到训练数据的预测结果，根据训练数据的预测结果可得到训练好的大模型。基于上述训练方法，在大模型与小模型的协作推理运算过程中模型与数据均不出本地，云端仅能获知分片后的部分数据，以及难以还原出原始输入的特征向量，保护了隐私，提高了数据的安全性，且仅将头尾两个 Adaptor 分散到云端和客户端，利用加密的客户端数据做 Adaptor 部分的模型训练，模型主体结构依然位于云端，利用 Cloud 算力进行训练与
25
30

推理，保证了算力性能，同时有效利用 Adaptor 微调机制，Adaptor 参数量仅占整个原始大模型百分之几或千分之几，能达到与全模型全参数微调几乎相同效果的特性，实现了在安全、充分利用算力的前提下对大模型的有效训练。

5 附图说明

此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本申请的实施例，并与说明书一起用于解释本申请的原理。

- 图 1 为本申请实施例提供的一种云端与本地客户端联合模型结构示意图；
- 图 2 为本申请实施例提供的一种基于 Adaptor 方式的大模型结构示意图；
- 10 图 3 为本申请实施例提供的一种大模型的训练过程模型结构示意图；
- 图 4 为本申请实施例提供的一种大模型的训练方法的流程示意图；
- 图 5 为本申请实施例提供的一种 ABY 协议的基本运算逻辑的示意图；
- 图 6 为本申请实施例提供的一种参数训练过程示意图；
- 图 7 为本申请实施例提供的一种训练好的大模型的结构示意图；
- 15 图 8 为本申请实施例提供的另一种大模型的训练方法的流程示意图；
- 图 9 为本申请实施例提供的一种大模型的训练设备的结构示意图。

通过上述附图，已示出本申请明确的实施例，后文中将有更详细的描述。这些附图和文字描述并不是为了通过任何方式限制本申请构思的范围，而是通过参考特定实施例为本领域技术人员说明本申请的概念。

20

具体实施方式

- 这里将详细地对示例性实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本申请相一致的所有实施方式。相反，它们仅是与如所附权利要求书中所详述的、本申请的一些方面相一致的装置和方法的例子。
- 25

需要说明的是，本公开所涉及的用户信息（包括但不限于用户设备信息、用户个人信息等）和数据（包括但不限于用于分析的数据、存储的数据、展示的数据等），均为经用户授权或者经过各方充分授权的信息和数据，并且相关数据的收集、使用和处理需要遵守相关法律法规和标准，并提供有相应的操作入口，供用户选择授权或者拒绝。

30

一个通用的预训练完成的大模型通常含有十亿、百亿级别的参数量，需要算力以及私有数据对其进行微调。中小型企业通常具有本地化数据，可以用来微调大模型，从而获得

定制化的大模型。然而，现状或难点是：模型持有方、数据持有方希望合作出一个大模型，但均不愿意泄露各自的内容（模型以及数据）给彼此。

目前，有以下三种方式训练大模型：方式一，提出了基于分布式协作的隐私计算方法及系统，该技术方案主要模块包括软件开发工具包 SDK、控制器和多个隐私计算一体机，5 控制器部署在边缘云中，多个隐私计算一体机部署在私有云中。在协作进行隐私计算任务的流程中，首先由 SDK 通过控制器向 n 个隐私计算一体机申请使用数据集，数据要素持有者可以通过隐私计算一体机向控制器进行数据集登记注册，内容包括但不限于：数据的特征、数据来源、数据数量、数据使用权报价、数据的隐私规则等。SDK 接收到多个隐私计算一体机数据后，开发人员可以在 SDK 中开发隐私计算应用，开发完毕后将该隐私计算应用发送给控制器，控制器对隐私计算应用进行分解，将得到的计算任务分别发送给 n 10 个隐私计算一体机，每个隐私计算一体机根据自身存储的数据集对计算任务进行计算，通过控制器将得到的计算结果发送给 SDK。方式二，提出了一种联邦学习方法和系统、第一隐私计算平台和第二隐私计算平台，该技术方案实现了在不共享受保护的隐私数据且自有数据不传递到外部的前提下，通过加密的方式交换模型相关信息从而实现协同优化联邦学习模型，且两个隐私计算平台遵循地协议不同。具体地，与第一数据方基于第一数据方持有的第一参数密文执行第一交互操作，得到第一待聚合数据，第一参数密文为第一数据方基于其自身的数据集在本地进行训练并按照第一加密协议加密得到，第一待聚合数据符合第二隐私计算平台支持的第二加密协议。将持有的第一待聚合数据通过路由节点转发至第二隐私计算平台，以使第二隐私计算平台利用第一待聚合数据和第二待聚合数据进行第一 15 聚合计算，得到聚合参数；其中，第二待聚合数据根据第二隐私计算平台侧的各第二数据方的第二参数密文计算得到，第二待聚合数据符合第二加密协议；接收路由节点转发的第二隐私计算平台返回的聚合参数，利用聚合参数执行下一轮迭代计算。方式三，提出了一种敏感数据隐私保护的天际线查询方法及系统，该技术方案将敏感数据信息通过加法秘密共享上传到两个不合谋的第一服务器和第二服务器，两个服务器通过联合隐私计算生成天际线图，且各自拥有对应份额图；在用户端将用户信息通过加法秘密共享上传，各服务器在各自的份额图上联合地通过私有信息检索方法检索用户所期望查询的相似信息份额，并将相似信息份额发送至用户端；用户端恢复完整的查询信息，以进行辅助决策。

然而，上述方式均存在缺点：方式一提出的技术方案存在一定的不足，主要在于数据的特征、数据来源、数据数量等要素在隐私计算任务协作过程中会出本地私有云，这给数据安全带来风险。此外就是算力层面，最终的隐私计算任务由 SDK 通过控制器发送回 n 20 个隐私计算一体机本地执行，对于联邦场景中参与隐私计算的分散参与方而言，算力不充

分这一点通常是一个棘手的问题，如何有效利用云端大厂的充足算力高效地协作执行隐私计算任务是有待研究的。方式二主要在于仅聚焦于协议转换本身。尽管两方私有数据可以不出本地，对于计算效率、全量参数传递带来的较大计算成本未考虑到。方式三主要在于决策服务提供商需要将自身拥有的全量历史数据以加法秘密分享的形式上传至两服务器，这就已经造成了数据出域，无法实现“数据可用不可见”。此外，全量历史数据上传至两台服务器，伴随着海量用户数据不断产生，会对通信链路造成极大压力。由两台服务器进行网格式的信息检索，检索出用户所期望查询的相似信息后，再将各自的相似信息份额发送至用户端，用户端恢复完整的相似信息以便于用户进行查询并提供辅助决策，这种方式虽然实现了查询用户与数据隔离，有一定程度的隐私保障，但它本质上是一种传统的数据库查询，并不贴近具体的垂直领域对大模型（Large Language Model）处理海量数据、实现模型训练与推理的实际应用诉求。

综合上述可知，相关技术存在对大模型训练过程中，数据安全性不高、算力利用不充分的技术问题。

为了解决上述问题，本申请实施例提供一种大模型的训练方法、装置、设备及介质，该方法云端和本地客户端双方协作共建定制化大模型，实现隐私保护和权益保护，隐私包括云端的模型以及客户端的私有数据，在合作的过程中隐私不暴露给对方，所形成的模型在使用过程中仍然能体现各自的权益。

可选地，本申请实施例可实现以下功能：

（1）通过秘密分享保护数据与模型隐私：在大模型与小模型的协作推理运算过程中使用各类秘密分享协议，模型与数据均不出本地。且能够使用异构互通算子。安全算子强调的是基于密码学构造的、采用安全计算协议实现的基础计算工具，例如基于多方的矩阵乘法、欧氏距离、大小比较计算等。

（2）合理利用算力保障整体性能：仅将头尾两个适配器 Adaptor 分散到云端（Cloud）和本地客户端（Client）做 Adaptor 部分的模型训练。模型大部分结构依然位于云端，利用 Cloud 算力进行训练与推理，保证性能。

（3）有效利用 Adaptor 微调机制：Adaptor 一般为原始大模型中的部分有效层，如参数的低秩矩阵，或者为另外嵌入到大模型每一层中的占比极小的参数层。Adaptor 参数量仅占整个原始大模型百分之几或千分之几，能达到与全模型全参数微调几乎相同效果的特性。利用这一特性将模型极小部分（也即 Adaptor）的部分参数分散到 Client 侧，与 Cloud 侧共同完成基于隐私保护/秘密分享协议的安全训练/推理。

可选地，图 1 为本申请实施例提供的一种云端与本地客户端联合模型结构示意图，如

图 1 所示，这里的本地客户端也称为客户端，本申请实施例能够通过云端和本地客户端对原始大模型的联合训练，得到训练好的大模型。

可选地，图 2 为本申请实施例提供的一种基于 Adaptor 方式的大模型结构示意图，如图 2 所示，模型结构：模型由 3 部分构成，中间的模型 M0 以及头尾两个 Adaptor 部分，
5 包括头部适配器（Adaptor Header，AH）和尾部适配器（Adaptor Tail，AT）。Adaptor 为原始大模型中头部和尾部的一小部分有效层，通常仅占百分之几甚至千分之几的参数量，专门用于对原始模型进行微调。微调方式：将训练数据 Data 喂给大模型进行微调训练，在训练过程中，只有 Adaptor 部分的参数进行更新，M0 部分的参数始终保持不变。方法特色：这种微调方式可以仅微调原始大模型百分之几甚至千分之几的参数量，可以达到和微调全
10 部原始模型几乎一致的模型推理准确率。

由图 2 可知，模型在预训练时，大模型 $M=AH+M0+AT$ ，微调后即训练完成后，训练好的大模型 $M^*=AH^*+M0+AT^*$ ，只有 Adaptor 部分的参数进行更新，M0 部分的参数始终保持不变。可以理解的是，这里的“+”不表示数学计算的“+”，表示模型 M0 以及头尾两个 Adaptor 部分的结合叠加。

15 可以理解的是，本申请实施例示意的结构并不构成对大模型的训练系统架构的具体限定。在本申请另一些可行的实施方式中，上述架构可以包括比图示更多或更少的部件，或者组合某些部件，或者拆分某些部件，或者不同的部件布置，具体可根据实际应用场景确定，在此不做限制。图 1 和图 2 所示的部件可以以硬件，软件，或软件与硬件的组合实现。

在具体实现过程中，上述云端以及本地客户端中包括处理器或者处理设备。

20 应理解，上述处理设备可以通过处理器读取存储器中的指令并执行指令的方式实现，也可以通过芯片电路实现。

另外，本申请实施例描述的网络架构以及业务场景是为了更加清楚的说明本申请实施例的技术方案，并不构成对于本申请实施例提供的技术方案的限定，本领域普通技术人员可知，随着网络架构的演变和新业务场景的出现，本申请实施例提供的技术方案对于类似
25 的技术问题，同样适用。

下面结合具体的实施例对本申请的技术方案进行详细的说明：

可选地，图 3 为本申请实施例提供的一种大模型的训练过程模型结构示意图，如图 3 所示，训练过程的大模型包括头部适配器（第一云端部分和第一本地部分组成头部适配器）、模型主体 M0 和尾部适配器（第二云端部分和第二本地部分组成尾部适配器）；其中头部
30 适配器包括第一云端部分 AH1 和第一本地部分 AH2，其中尾部适配器包括第二云端部分 AT1 和第二本地部分 AT2；在训练大模型的过程中第一云端部分、模型主体和第二云端部

分部署于云端，第一本地部分和第二本地部分部署于本地客户端。

在训练的过程中，数据 Data 拆分为两部分 D1 和 D2。

其中，Adaptor 一般为原始大模型中的部分有效层，如参数的低秩矩阵，或者为另外嵌入到大模型每一层中的占比极小的参数层。Adaptor 分为 Adaptor Header 和 Adaptor Tail 两部分，分别记为 AH、AT。他们以分散的形式各自的一部分位于 Cloud 端，另一部分位于 Client 端。大模型其他部分参数仍然位于 Cloud 端。

如图 3 所示，云端部分和本地部分可以通过多方计算 (Multi-party Computation, MPC) 算子运算实现联合训练。

如图 3 所示，协作训练过程如下：

10 步骤一：原始训练数据为来自 Client 的行业私有化数据，将其分片进入下一步基于秘密共享计算的训练过程中。

通过步骤一，Client 输入数据得以保护。

步骤二： Adaptor Header 分为两半，一半位于 Cloud，一半位于 Client，双方进行第一次基于秘密分享协议的联合训练。

15 通过步骤二，Adaptor 部分参数分散持有。

步骤三：经过 Adaptor Header 的训练输出，为明文形式。

此时中间层的输出数据已经是特征向量，很难还原出原始输入。

步骤四：在 Cloud 一侧进行全部训练数据的明文训练。

Cloud 端运算 M0 部分可以使得大模型不泄露，同时充分利用 Cloud 端算力。

20 步骤五：M0 的输出再次秘密分享分为两半，Adaptor Tail 也分为两半，一半位于 Cloud，一半位于 Client，双方进行第二次基于秘密分享协议的联合训练。

Adaptor 部分参数继续分散持有。

可选地，图 4 为本申请实施例提供的一种大模型的训练方法的流程示意图，用于对图 3 中的大模型进行训练。本申请实施例的执行主体可以为图 1 中的云端与本地客户端联合模型，具体执行主体可以根据实际应用场景确定。如图 4 所示，该方法包括如下步骤：

25 S401：本地客户端获取训练数据，并对训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据。

可选地，这里可以获取任意一个训练数据，对着一个训练数据进行分片处理，得到第一分片训练数据和第二分片训练数据。

30 可选地，对训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据，包括：

对训练数据进行随机分片处理，以得到第一分片训练数据和第二分片训练数据。

其中，本申请实施例在对训练数据进行分片处理时，可采用随机分片的方式，通过此方式，云端无法还原出原始的训练数据，进而充分保护了本地客户端的数据安全。

可选地，对训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据，

5 包括：

根据第一预设秘密分片算法，对训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据。

可以理解的是，第一预设秘密分片算法可以根据实际情况确定，本申请实施例对此不作具体限制。

10 其中，本申请实施例在对训练数据进行分片处理时，可采用第一预设秘密分片算法的方式，通过此方式，云端无法还原出原始的训练数据，进而充分保护了本地客户端的数据安全。

S402：本地客户端将第一分片训练数据发送至云端。

15 S403：云端将第一分片训练数据输入至第一云端部分进行运算，本地客户端将第二分片训练数据输入至第一本地部分进行运算，以输出第一特征向量并发送至云端。

可选地，云端将第一分片训练数据输入至第一云端部分进行运算，本地客户端将第二分片训练数据输入至第一本地部分进行运算，以输出第一特征向量并发送至云端，包括：

20 云端将第一分片训练数据输入至第一云端部分进行运算，本地客户端将第二分片训练数据输入至第一本地部分进行运算，且第一云端部分和第一本地部分进行运算的过程中通过多方安全计算的加密运算方式，经过第一云端部分和第一本地部分的运算，以输出第一特征向量并发送至云端。

其中，本申请实施例可基于多方安全计算共享运算参数，从而实现第一云端部分和第一本地部分基于秘密分享协议的联合训练，通过多方安全计算将原始数据转化为难以还原的特征向量，进一步地提高了大模型训练的安全性及可行性。

25 可选地，多方安全计算包括 ABY 混合协议或秘密共享（Secret Sharing，SS）协议。ABY 协议是一种两方专用的安全计算框架，它由三个协议混合组成，其中 A 代表算数秘密共享（Arithmetic sharing）、B 代表布尔秘密共享（Boolean sharing），Y 代表姚氏混淆电路（Yao's garbled circuits）。

30 其中，客户端的训练数据进行秘密分片，分散到 Cloud 端和 Client 端，同时基于秘密分享协议（如 ABY，ss 等协议）训练 AH 部分的网络参数。

以 ABY 协议为例，具体地对训练过程进行说明：

ABY 协议的基本运算逻辑见图 5。例如，Client 拥有某行业微调数据集：

“如果物价水平（即 CPI）上升，那么同样数量的货币购买同样的商品和服务的能力就会下降，即货币的购买力降低。反之，如果物价水平下降，那么 同样数量的货币购买同样的商品和服务的能力就会提高，即货币的购买力提高。”

5 此时，数据集 token 的结果为[如果，物价，水平，（，即， C，P，I，），上升，那么，同样，数量，的，货币，购买，商品，和，服务，能力，就，会，下降，购买力，降低，反之，提高]， N=27 是所有单词的列表。接下来对单词表进行编码和向量嵌入，以部分 tokens 嵌入成 768 维向量为例，图 6 为本申请实施例提供的一种参数训练过程示意图。如图 6 所示，在本例中我们得到的输入数据为 27*768 的矩阵，记为输入数据矩阵 D。令
10 AH、AT 为 m*n 维矩阵（m 和 n 为任意正整数），在分片中我们分别讲 $m/2 * n$ 维矩阵放在 Cloud 端和 Client 端。接下来按照 aby 协议切割矩阵 D，并生成双方的随机三元组，aby 秘密分享协议通过三元组来盲化两边的计算数据，双方计算过程中无法逆推得到对方的数据。

S404：云端将第一特征向量输入至模型主体，以输出第二特征向量。

15 其中，第一特征向量为秘密共享形式数据明文。
得到解密后的秘密共享形式数据明文，准备进入 Cloud 端大模型其他部分参数进行训练：经过 AH 部分的运算后，我们得到原始输入数据的分片相乘形式 $D1*D2$ ，该秘密全部放置到 Cloud 端，进入固定部分 M0 进行训练。

数据在 Cloud 端大模型其他部分参数进行训练，充分利用云端算力，得到秘密共享形式数据明文输出，准备再次分片进入 AT 参与再一次的秘密共享训练。
20

S405：云端将第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量，并将第二分片特征向量发送至客户端。

可选地，云端将第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量，包括：

25 云端将第二特征向量进行随机分片处理，以得到第一分片特征向量和第二分片特征向量。

可选地，云端将第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量，包括：

30 云端根据第二预设秘密分片算法，将第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量。

这里，本申请实施例云端可通过随机或预设算法对第二特征向量进行分片处理，对数

据进行进一步地拆分，以实现安全可靠的云端与本地客户端的联合训练，进一步地提高了大模型训练的数据安全及数据隐私。

S406：云端将第一分片特征向量输入至第二云端部分进行运算，本地客户端将第二分片特征向量输入至第二本地部分进行运算，以输出训练数据的预测结果。

5 可选地，数据分散到 Cloud 端和 Client 端，同时基于秘密分享协议（如 aby, ss 等协议）训练 AT 部分的网络参数，AT 部分计算逻辑与步骤 S403 中相同。

可选地，云端将第一分片特征向量输入至第二云端部分进行运算，本地客户端将第二分片特征向量输入至第二本地部分进行运算，以输出训练数据的预测结果，包括：

10 云端将第一分片特征向量输入至第二云端部分进行运算，本地客户端将第二分片特征向量输入至第二本地部分进行运算，且第二云端部分和第二本地部分进行运算的过程中通过多方安全计算的加密运算方式，经过第二云端部分和第二本地部分的运算，以输出训练数据的预测结果。

15 其中，本申请实施例对尾部适配器的运算也基于多方安全计算共享运算参数，从而实现第二云端部分和第二本地部分基于秘密分享协议的联合训练，进一步地提高了大模型训练的安全性及可行性。

S407：云端和本地客户端根据预设数量的训练数据的预测结果，迭代调整所述大模型至模型收敛。

可选地，云端和本地客户端根据预设数量的训练数据的预测结果，迭代调整所述大模型至模型收敛，包括：

20 云端根据预测结果与训练数据对应的预设标签得到损失参数；云端根据损失参数调整大模型的第一云端部分和第二云端部分的模型参数；本地客户端根据损失参数，调整大模型的第一本地部分和第二本地部分的模型参数；本地客户端若确定大模型的预测误差大于预设阈值，则重新执行“本地客户端获取训练数据”的步骤以继续训练大模型，若确定大模型的预测误差小于或等于预设阈值，则停止迭代训练。

25 可以理解的是，这里的预设阈值可以根据实际情况确定，本申请实施例对此不作具体限制。

可选地，得到训练输出，再与标签计算损失，反馈整个模型网络进行模型收敛计算。具体地，对于训练数据 token 的掩码进行预测，并计算损失函数，这里取一般的多分类交叉熵损失函数进行说明。AT 的输出为经过分类层计算的 logits 值，它将与真实标签值计算损失。基于上述步骤执行模型迭代训练，直至模型收敛到一定精度。

30 可选地，云端根据预测结果与训练数据对应的预设标签得到损失参数，包括：

云端将预测结果与预设标签输入至损失函数中，以输出损失参数。

可选地，损失函数的计算公式为：

$$L = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic})$$

其中，L 表示损失函数，M 表示类别数量，N 表示样本数量， y_{ic} 是符号函数（取 0 或 1，如果样本 i 的真是类别等于 c 则取 1，否则取 0），表示第 i 个样本的属于第 c 个类别的概率， p_{ic} 表示模型预测第 i 个样本属于第 c 个类别的概率，i 和 N 是任意正整数。

这里，本申请实施例得到训练数据的预测结果后，再与预测标签计算损失，反馈整个模型网络进行模型收敛计算，从而调整大模型的模型参数，以调整大模型直至满足要求，实现了对大模型的训练。

10 可选地，在云端根据预测结果与训练数据对应的预设标签得到损失参数，并根据损失参数调整大模型的第一云端部分和第二云端部分的模型参数之前，还包括：

本地客户端根据预设加密算法，对训练数据对应的预设标签进行加密处理，以得到加密标签；本地客户端将加密标签发送至云端。

15 这里，为了进一步保证在大模型训练过程中的数据安全，本申请实施例本地客户端将训练数据对应的预设标签发送至云端时，首先对预设标签进行加密处理，云端无法获取原始的预设标签，进一步地提高了大模型训练的安全性。

可选地，云端根据预测结果与训练数据对应的预设标签得到损失参数，并根据损失参数调整大模型的第一云端部分和第二云端部分的模型参数，包括：

20 云端根据预设加密算法，对预测结果进行加密处理，以得到加密预设结果；云端根据加密预测结果与加密标签得到损失参数，并根据损失参数调整大模型的第一云端部分和第二云端部分的模型参数。

可选地，预设加密算法为同态加密算法。

25 这里，云端根据本地客户端对预设标签的预设加密算法对训练数据的预测结果进行加密处理，从而能够将得到的加密预测结果与加密标签进行运算处理，得到损失参数，进而在安全可靠的前提下对模型参数进行调整。

可选地，将上述步骤中训练完毕的 AH、AT 合并，并全部置于 Client 端，得到最终训练好的大模型。训练好的大模型的头部适配器和尾部适配器部署于本地客户端，模型主体部署于云端。

30 示范性地，图 7 为本申请实施例提供的一种训练好的大模型的结构示意图，通过图 7 中的训练好的大模型，能够实现数据推理预测。

本申请实施例在训练大模型的过程中，将大模型拆分为三个部分，中间的模型主体 M0 以及头部适配器 Adaptor Header 和尾部适配器 Adaptor Tail 两个 Adaptor 部分，头部适配器包括第一云端部分 AH1 和第一本地部分 AH2，尾部适配器包括第二云端部分 AT1 和第二本地部分 AT2，每个适配器分为两部分，一半位于云端 Cloud，一半位于本地客户端 Client，对大模型的训练数据为来自 Client 的行业私有化数据，将其分片进入后，分别进入云端和本地客户端进行第一次联合训练，经过第一次训练后，得到特征向量形式的第一特征向量，再输入至云端进行全部训练数据的明文训练，云端的输出再分片处理分别在云端和本地客户端进行第二次联合训练，得到训练数据的预测结果，根据训练数据的预测结果可得到训练好的大模型。基于上述训练方法，在大模型与小模型的协作推理运算过程中模型与数据均不出本地，云端仅能获知分片后的部分数据，以及难以还原出原始输入的特征向量，保护了隐私，提高了数据的安全性，且仅将头尾两个 Adaptor 分散到云端和客户端，利用加密的客户端数据做 Adaptor 部分的模型训练，模型主体结构依然位于云端，利用 Cloud 算力进行训练与推理，保证了算力性能，同时有效利用 Adaptor 微调机制，Adaptor 参数量仅占整个原始大模型百分之几或千分之几，能达到与全模型全参数微调几乎相同效果的特性，实现了在安全、充分利用算力的前提下对大模型的有效训练。

可选地，本申请实施例能够通过训练好的大模型实现数据的推理训练，相应地，图 8 为本申请实施例提供的另一种大模型的训练方法的流程示意图，如图 8 所示，该方法包括：

- S801：本地客户端获取训练数据，并对训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据。
- S802：本地客户端将第一分片训练数据发送至云端。
- S803：云端将第一分片训练数据输入至第一云端部分进行运算，本地客户端将第二分片训练数据输入至第一本地部分进行运算，以输出第一特征向量并发送至云端。
- S804：云端将第一特征向量输入至模型主体，以输出第二特征向量。
- S805：云端将第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量，并将第二分片特征向量发送至客户端。
- S806：云端将第一分片特征向量输入至第二云端部分进行运算，本地客户端将第二分片特征向量输入至第二本地部分进行运算，以输出训练数据的预测结果。
- S807：云端和本地客户端根据预设数量的训练数据的预测结果，迭代调整所述大模型至模型收敛。
- 其中，步骤 S801-S807 的实现方式与步骤 S401-S407 的实现方式相同，在此不作赘述。
- S808：本地客户端获取一条待预测数据。

S809: 本地客户端将待预测数据输入至头部适配器, 以输出第一预测向量。

第一预测向量的数据格式为加密格式。

其中, 第一预测向量针对于云端为加密格式, 云端无法得到具体的数据, 提高了数据推理预测的安全性。

5 首先, 推理数据经过 Client 端 AH 部分, 输出为原始明文推理数据的某种打乱形式, 无法反推, 从而保障了推理数据隐私, 接下来该形式的推理数据被传递到 Cloud 端大模型剩余参数部分参与推理。

S810: 本地客户端将第一预测向量发送至云端。

S811: 云端将第一预测向量输入至模型主体, 以输出第二预测向量。

10 上一步中的数据在模型剩余参数部分进行推理计算, 得到计算输出, 即第二预测向量。

S812: 云端将第二预测向量发送至本地客户端。

S813: 本地客户端将第二预测向量输入至尾部适配器, 以输出待预测数据的预测结果。

上一步的计算输出传递到 Client 端, 进入 AT 部分进行推理计算, 得到模型推理结果输出。

15 本申请实施例实现了隐私保护, 合作前的大模型以及训练数据, 合作后的增量模型以及推理数据都没有暴露给对方, 也实现了权益保护, 双方协作形成联合大模型, 推理阶段也需要双方合作才能完成, 任何一方都无法独立完成推理。Client 方的数据权益得到保障, Cloud 方的模型以及算力权益得到保障。

20 本申请实施例能够根据训练好的大模型进行推理预测, 对于在本地客户端的待预测数据, 首先通过本地客户端的头部适配器将待预测数据转化为相对于云端加密的第一预测向量, 再运用云端的算力, 对待预测数据进行充分推理预测, 再通过本地客户端的尾部适配器输出待预测数据的预测结果, 实现了安全、高效的数据推理预测。

25 本申请实施例还提供的一种大模型的训练装置, 本申请实施例的装置包括: 第一处理模块、第一发送模块、第二处理模块、第一输入模块、第三处理模块、第四处理模块和第五处理模块。这里的大模型的训练装置可以是上述云端或者本地客户端的处理装置、处理器本身, 或者是实现处理器的功能的芯片或者集成电路。这里需要说明的是, 第一处理模块、第一发送模块、第二处理模块、第一输入模块、第三处理模块、第四处理模块和第五处理模块的划分只是一种逻辑功能的划分, 物理上两者可以是集成的, 也可以是独立的。

30 其中, 第一处理模块, 用于本地客户端获取训练数据, 并对训练数据进行分片处理, 以得到第一分片训练数据和第二分片训练数据;

第一发送模块, 用于本地客户端将第一分片训练数据发送至云端;

- 第二处理模块，用于云端将第一分片训练数据输入至第一云端部分进行运算，本地客户端将第二分片训练数据输入至第一本地部分进行运算，以输出第一特征向量并发送至云端；
- 第一输入模块，用于云端将第一特征向量输入至模型主体，以输出第二特征向量；
- 5 第三处理模块，用于云端将第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量，并将第二分片特征向量发送至客户端；
- 第四处理模块，用于云端将第一分片特征向量输入至第二云端部分进行运算，本地客户端将第二分片特征向量输入至第二本地部分进行运算，以输出训练数据的预测结果；
- 第五处理模块，用于云端和本地客户端根据预设数量的训练数据的预测结果，迭代调
- 10 整所述大模型至模型收敛。
- 可选地，第二处理模块具体用于：
- 云端将第一分片训练数据输入至第一云端部分进行运算，本地客户端将第二分片训练数据输入至第一本地部分进行运算，且第一云端部分和第一本地部分进行运算的过程中通过多方安全计算的加密运算方式，经过第一云端部分和第一本地部分的运算，以输出第一
- 15 特征向量并发送至云端。
- 可选地，第四处理模块具体用于：
- 云端将第一分片特征向量输入至第二云端部分进行运算，本地客户端将第二分片特征向量输入至第二本地部分进行运算，且第二云端部分和第二本地部分进行运算的过程中通过多方安全计算的加密运算方式，经过第二云端部分和第二本地部分的运算，以输出训练
- 20 数据的预测结果。
- 可选地，多方安全计算包括 ABY 混合协议或秘密共享协议。
- 可选地，第五处理模块具体用于：
- 云端根据预测结果与训练数据对应的预设标签得到损失参数；云端根据损失参数调整大模型的第一云端部分和第二云端部分的模型参数；本地客户端根据损失参数，调整大模
- 25 型的第一本地部分和第二本地部分的模型参数；本地客户端若确定大模型的预测误差大于预设阈值，则重新执行“本地客户端获取训练数据”的步骤以继续训练大模型，若确定大模型的预测误差小于或等于预设阈值，则停止迭代训练。
- 可选地，第五处理模块还具体用于：云端将预测结果与预设标签输入至损失函数中，以输出损失参数。
- 30 可选地，损失函数的计算公式为：

$$L = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic})$$

其中， L 表示损失函数， M 表示类别数量， N 表示样本数量， y_{ic} 是符号函数，表示第 i 个样本的属于第 c 个类别的概率， p_{ic} 表示模型预测第 i 个样本属于第 c 个类别的概率， i 和 N 是任意正整数。

- 5 可选地，在第五处理模块用于云端根据预测结果与训练数据对应的预设标签得到损失参数，并根据损失参数调整大模型的第一云端部分和第二云端部分的模型参数之前，上述装置还包括加密处理模块，用于：

本地客户端根据预设加密算法，对训练数据对应的预设标签进行加密处理，以得到加密标签；本地客户端将加密标签发送至云端。

- 10 可选地，第五处理模块具体用于：

云端根据预设加密算法，对预测结果进行加密处理，以得到加密预设结果；云端根据加密预测结果与加密标签得到损失参数，并根据损失参数调整大模型的第一云端部分和第二云端部分的模型参数。

可选地，预设加密算法为同态加密算法。

- 15 可选地，第一处理模块具体用于：

对训练数据进行随机分片处理，以得到第一分片训练数据和第二分片训练数据。

可选地，第一处理模块具体用于：

根据第一预设秘密分片算法，对训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据。

- 20 可选地，第三处理模块具体用于：

云端将第二特征向量进行随机分片处理，以得到第一分片特征向量和第二分片特征向量。

可选地，第三处理模块具体用于：

云端根据第二预设秘密分片算法，将第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量。

- 25 可选地，训练好的大模型的头部适配器和尾部适配器部署于本地客户端，模型主体部署于云端。

可选地，上述装置还包括推理预测模块，用于：

- 30 本地客户端获取一条待预测数据；本地客户端将待预测数据输入至头部适配器，以输出第一预测向量；本地客户端将第一预测向量发送至云端；云端将第一预测向量输入至模

型主体，以输出第二预测向量；云端将第二预测向量发送至本地客户端；本地客户端将第二预测向量输入至尾部适配器，以输出待预测数据的预测结果。

可选地，第一预测向量的数据格式为加密格式。

参考图 9，其示出了适于用来实现本公开实施例的大模型的训练设备的结构示意图，

- 5 该大模型的训练设备 900 可以为终端设备或服务器。其中，终端设备可以包括但不限于诸如移动电话、笔记本电脑、数字广播接收器、个人数字助理（Personal Digital Assistant，简称 PDA）、平板电脑（Portable Android Device，简称 PAD）、便携式多媒体播放器（Portable Media Player，简称 PMP）、车载终端（例如车载导航终端）等等的移动终端以及诸如数字 TV、台式计算机等等的固定终端。图 9 示出的大模型的训练设备仅仅是一个示例，不
10 应对本公开实施例的功能和使用范围带来任何限制。

- 如图 9 所示，大模型的训练设备 900 可以包括处理装置（例如中央处理器、图形处理器等）901，其可以根据存储在只读存储器（Read Only Memory，简称 ROM）902 中的程序或者从存储装置 908 加载到随机访问存储器（Random Access Memory，简称 RAM）903 中的程序而执行各种适当的动作和处理。在 RAM 903 中，还存储有大模型的训练设备 900
15 操作所需的各种程序和数据。处理装置 901、ROM 902 以及 RAM 903 通过总线 904 彼此相连。输入至/输出（I/O）接口 905 也连接至总线 904。

- 通常，以下装置可以连接至 I/O 接口 905：包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入至装置 906；包括例如液晶显示器（Liquid Crystal Display，简称 LCD）、扬声器、振动器等输出装置 907；包括例如磁带、硬盘等的存储装置 908；以及通信装置 909。通信装置 909 可以允许大模型的训练设备 900 与其他设备进行无线或有线通信以交换数据。虽然图 9 示出了具有各种装置的大模型的训练设备 900，但是应理解的是，并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。

- 特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件
25 程序。例如，本公开的实施例包括一种计算机程序产品，其包括承载在计算机可读介质上的计算机程序，该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信装置 909 从网络上被下载和安装，或者从存储装置 908 被安装，或者从 ROM 902 被安装。在该计算机程序被处理装置 901 执行时，执行本公开实施例的方法中限定的上述功能。

- 30 需要说明的是，本公开上述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限

于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。

计算机可读存储介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、

- 5 光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开中，计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读信号
- 10 号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读信号介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：电线、光缆、RF（射频）等等，或者上述的任意合适的组合。

- 上述计算机可读介质可以是上述大模型的训练设备中所包含的；也可以是单独存在，
- 15 而未装配入该大模型的训练设备中。

上述计算机可读介质承载有一个或者多个程序，当上述一个或者多个程序被该大模型的训练设备执行时，使得该大模型的训练设备执行上述实施例所示的方法。

- 可以以一种或多种程序设计语言或其组合来编写用于执行本公开的操作的计算机程序代码，上述程序设计语言包括面向对象的程序设计语言——诸如 Java、Smalltalk、C++，
- 20 还包括常规的过程式程序设计语言——诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网（Local Area Network，简称 LAN）或广域网（Wide Area Network，简称 WAN）——连接到用户
- 25 计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

- 附图中的流程图和框图，图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分，该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。
- 30

也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本公开实施例中所涉及到的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。其中，单元的名称在某种情况下并不构成对该单元本身的限定，例如，第一获取单元还可以被描述为“获取至少两个网际协议地址的单元”。

本文中以上描述的功能可以至少部分地由一个或多个硬件逻辑部件来执行。例如，非限制性地，可以使用的示范类型的硬件逻辑部件包括：现场可编程门阵列（FPGA）、专用集成电路（ASIC）、专用标准产品（ASSP）、片上系统（SOC）、复杂可编程逻辑设备（CPLD）等等。

在本公开的上下文中，机器可读介质可以是有形的介质，其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦除可编程只读存储器（EPROM 或快闪存储器）、光纤、便捷式紧凑盘只读存储器（CD-ROM）、光学储存设备、磁储存设备、或上述内容的任何合适组合。

本申请实施例的大模型的训练设备，可以用于执行本申请上述各方法实施例中的技术方案，其实现原理和技术效果类似，此处不再赘述。

本申请实施例还提供一种计算机可读存储介质，该计算机可读存储介质中存储有计算机执行指令，计算机执行指令被处理器执行时用于实现上述任一项的大模型的训练方法。

本申请实施例还提供一种计算机程序产品，包括计算机程序，计算机程序被处理器执行时，用于实现上述任一项的大模型的训练方法。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各

个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

- 5 本领域技术人员在考虑说明书及实践这里公开的发明后，将容易想到本申请的其它实施方案。本申请旨在涵盖本申请的任何变型、用途或者适应性变化，这些变型、用途或者适应性变化遵循本申请的一般性原理并包括本申请未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的，本申请的真正范围和精神由下面的权利要求书指出。

应当理解的是，本申请并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围进行各种修改和改变。本申请的范围仅由所附的权利要求书来限制。

权 利 要 求 书

1、一种大模型的处理方法，其特征在于，所述大模型包括头部适配器、模型主体和尾部适配器；其中所述头部适配器包括第一云端部分和第一本地部分，其中所述尾部适配器包括第二云端部分和第二本地部分；在训练所述大模型的过程中所述第一云端部分、所述模型主体和所述第二云端部分部署于云端，所述第一本地部分和所述第二本地部分部署于本地客户端；

所述方法包括：

本地客户端获取训练数据，并对所述训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据；

本地客户端将所述第一分片训练数据发送至云端；

云端将所述第一分片训练数据输入至所述第一云端部分进行运算，本地客户端将所述第二分片训练数据输入至所述第一本地部分进行运算，以输出第一特征向量并发送至云端；

云端将所述第一特征向量输入至所述模型主体，以输出第二特征向量；

云端将所述第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量，并将第二分片特征向量发送至客户端；

云端将所述第一分片特征向量输入至所述第二云端部分进行运算，本地客户端将所述第二分片特征向量输入至所述第二本地部分进行运算，以输出所述训练数据的预测结果；

云端和本地客户端根据预设数量的训练数据的预测结果，迭代调整所述大模型至模型收敛。

20

2、根据权利要求 1 所述的方法，其特征在于，所述云端将所述第一分片训练数据输入至所述第一云端部分进行运算，本地客户端将所述第二分片训练数据输入至所述第一本地部分进行运算，以输出第一特征向量并发送至云端，包括：

云端将所述第一分片训练数据输入至所述第一云端部分进行运算，本地客户端将所述第二分片训练数据输入至所述第一本地部分进行运算，且所述第一云端部分和所述第一本地部分进行运算的过程中通过多方安全计算的加密运算方式，经过所述第一云端部分和所述第一本地部分的运算，以输出第一特征向量并发送至云端。

25

3、根据权利要求 1 所述的方法，其特征在于，所述云端将所述第一分片特征向量输入至所述第二云端部分进行运算，本地客户端将所述第二分片特征向量输入至所述第二本

30

地部分进行运算，以输出所述训练数据的预测结果，包括：

- 5 云端将所述第一分片特征向量输入至所述第二云端部分进行运算，本地客户端将所述第二分片特征向量输入至所述第二本地部分进行运算，且所述第二云端部分和所述第二本地部分进行运算的过程中通过多方安全计算的加密运算方式，经过所述第二云端部分和所述第二本地部分的运算，以输出所述训练数据的预测结果。

4、根据权利要求 2 或 3 所述的方法，其特征在于，所述多方安全计算为 ABY 混合协议或秘密共享协议。

- 10 5、根据权利要求 1 所述的方法，其特征在于，所述云端和本地客户端根据预设数量的训练数据的预测结果，迭代调整所述大模型至模型收敛，包括：

云端根据所述预测结果与所述训练数据对应的预设标签得到损失参数；

云端根据所述损失参数调整所述大模型的第一云端部分和所述第二云端部分的模型参数；

- 15 本地客户端根据所述损失参数，调整所述大模型的第一本地部分和所述第二本地部分的模型参数；

本地客户端若确定所述大模型的预测误差大于预设阈值，则重新执行“本地客户端获取训练数据”的步骤以继续训练所述大模型，若确定所述大模型的预测误差小于或等于预设阈值，则停止迭代训练。

20

- 6、根据权利要求 5 所述的方法，其特征在于，所述云端根据所述预测结果与所述训练数据对应的预设标签得到损失参数，包括：

云端将所述预测结果与所述预设标签输入至损失函数中，以输出所述损失参数。

- 25 7、根据权利要求 6 所述的方法，其特征在于，所述损失函数的计算公式为：

$$L = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic})$$

其中，L 表示损失函数，M 表示类别数量，N 表示样本数量， y_{ic} 是符号函数，表示第 i 个样本的属于第 c 个类别的概率， p_{ic} 表示模型预测第 i 个样本属于第 c 个类别的概率，i 和 N 是任意正整数。

30

- 8、根据权利要求 5 所述的方法，其特征在于，在所述云端根据所述预测结果与所

述训练数据对应的预设标签得到损失参数，并根据所述损失参数调整所述大模型的第一云端部分和所述第二云端部分的模型参数之前，还包括：

本地客户端根据预设加密算法，对所述训练数据对应的预设标签进行加密处理，以得到加密标签；

5 本地客户端将所述加密标签发送至所述云端。

9、根据权利要求 8 所述的方法，其特征在于，所述云端根据所述预测结果与所述训练数据对应的预设标签得到损失参数，并根据所述损失参数调整所述大模型的第一云端部分和所述第二云端部分的模型参数，包括：

10 云端根据所述预设加密算法，对所述预测结果进行加密处理，以得到加密预设结果；

云端根据所述加密预测结果与所述加密标签得到损失参数，并根据所述损失参数调整所述大模型的第一云端部分和所述第二云端部分的模型参数。

15 10、根据权利要求 8 或 9 所述的方法，其特征在于，所述预设加密算法为同态加密算法。

11、根据权利要求 1 至 3 任一项所述的方法，其特征在于，所述对所述训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据，包括：

20 对所述训练数据进行随机分片处理，以得到第一分片训练数据和第二分片训练数据。

12、根据权利要求 1 至 3 任一项所述的方法，其特征在于，所述对所述训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据，包括：

25 根据第一预设秘密分片算法，对所述训练数据进行分片处理，以得到第一分片训练数据和第二分片训练数据。

13、根据权利要求 1 至 3 任一项所述的方法，其特征在于，所述云端将所述第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量，包括：

30 云端将所述第二特征向量进行随机分片处理，以得到第一分片特征向量和第二分片特征向量。

14、根据权利要求 1 至 3 任一项所述的方法，其特征在于，所述云端将所述第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量，包括：

35 云端根据第二预设秘密分片算法，将所述第二特征向量进行分片处理，以得到第一分片特征向量和第二分片特征向量。

15、根据权利要求 1 至 3 任一项所述的方法，其特征在于，所述训练好的大模型的头部适配器和尾部适配器部署于本地客户端，模型主体部署于云端。

40 16、根据权利要求 15 所述的方法，其特征在于，还包括：

本地客户端获取一条待预测数据；

本地客户端将所述待预测数据输入至所述头部适配器，以输出第一预测向量；

本地客户端将所述第一预测向量发送至云端；

云端将所述第一预测向量输入至所述模型主体，以输出第二预测向量；
云端将所述第二预测向量发送至本地客户端；
本地客户端将所述第二预测向量输入至所述尾部适配器，以输出所述待预测数据的预测结果。

5

17、根据权利要求 16 所述的方法，其特征在于，所述第一预测向量的数据格式为加密格式。

10 18、一种大模型的训练设备，其特征在于，包括：处理器，以及与所述处理器通信连接的存储器；

所述存储器存储计算机执行指令；

所述处理器执行所述存储器存储的计算机执行指令，以实现如权利要求 1 至 17 中任一项所述的方法。

15 19、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质中存储有计算机执行指令，所述计算机执行指令被处理器执行时用于实现如权利要求 1 至 17 中任一项所述的方法。