(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 109840591 B
(45)授权公告日 2021.08.03

(21)申请号 201711227185.X
(22)申请日 2017.11.29
(65)同一申请的已公布的文献号
申请公布号 CN 109840591 A
(43)申请公布日 2019.06.04
(73)专利权人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼
(72)发明人 陈普 廖乔勃
(51)Int.Cl.
G06N 3/08(2006.01)
(56)对比文件
CN 106204780 A,2016.12.07
CN 106204780 A,2016.12.07
CN 105575389 A,2016.05.11
CN 103389719 A,2013.11.13
CN 107195186 A,2017.09.22
CN 107124276 A,2017.09.01
CN 106856508 A,2017.06.16
US 2017154113 A1,2017.06.01
US 8706659 B1,2014.04.22
CN 106502889 A,2017.03.15
审查员 王青

权利要求书3页 说明书12页 附图5页

(54)发明名称

模型训练系统、方法和存储介质

(57)摘要

本发明提供了一种模型训练系统、方法和存储介质，涉及机器学习领域。该模型训练系统，包括云数据存储平台和云模型训练平台；云数据存储平台用于存储训练数据，以及用于接收训练数据调用请求，根据训练数据调用请求，将与数据调用指令对应的训练数据导出至云模型训练平台；云模型训练平台用于接收模型训练创建指令，获取待训练模型，以及用于生成并向云数据存储平台发送训练数据调用请求，以及用于利用从云数据存储平台导出的训练数据，训练待训练模型，得到训练成果模型。利用本发明的技术方案能够降低训练数据发生泄露的风险。

CN 109840591 B

1.一种提供云服务的模型训练系统,其特征在于,包括云数据存储平台、云模型训练平台和云模型存储平台;

所述云数据存储平台,用于存储训练数据,所述训练数据由数据提供者上传至所述云数据存储平台;

所述云模型存储平台,用于存储待训练模型;

所述云模型训练平台,用于接收用户输入的模型训练创建指令,根据所述模型训练创建指令从所述云模型存储平台获取所述待训练模型,调用所述云数据存储平台存储的所述训练数据,根据所述训练数据训练所述待训练模型,得到训练成果模型,其中,所述待训练模型由模型提供者或所述用户上传至所述云模型存储平台。

2.根据权利要求1所述的模型训练系统,其特征在于,所述模型训练系统还包括鉴权中心,

所述鉴权中心,用于接收所述用户输入的鉴权许可请求,所述鉴权许可请求用于确定所述训练数据的权限。

3.根据权利要求1所述的模型训练系统,其特征在于,

所述云数据存储平台,用于接收所述数据提供者提供的所述训练数据的标签,所述训练数据的标签用于表征所述训练数据的内容。

4.根据权利要求1所述的模型训练系统,其特征在于,所述系统还包括检索数据平台,

所述检索数据平台,用于获取所述训练数据的信息,所述训练数据的信息包括所述训练数据的数据所有者信息、所述训练数据的数据上传日期中的任意一种或全部。

5.根据权利要求4所述的模型训练系统,其特征在于,

所述检索数据平台,用于根据所述训练数据的标签和所述训练数据的信息中的至少一个建立所述训练数据的数据索引表;

所述检索数据平台,还用于接收所述用户输入的包括检索关键词的检索指令,根据所述检索关键词在所述数据索引表中进行查找以生成检索结果,所述检索结果包括与所述检索关键词相关的训练数据的信息或与所述检索关键词相关的训练数据的标签中的至少一个。

6.根据权利要求5所述的模型训练系统,其特征在于,

所述检索数据平台,还用于将所述检索结果发送给用户终端以显示给所述用户;

所述检索数据平台,还用于接收所述用户终端发送的针对所述检索结果的数据选取指令,所述数据选取指令用于指示所述数据检索平台从所述检索结果中确定所述训练数据。

7.根据权利要求1所述的模型训练系统,其特征在于,

所述云模型训练平台,用于得到所述训练成果模型后,将所述训练成果模型发送至所述云模型存储平台。

8.根据权利要求1至7任一项所述的系统,其特征在于,所述模型训练系统还包括数据稽查系统,

所述数据稽查系统,用于判定所述数据提供者上传的所述训练数据的有效性。

9.根据权利要求1至7任一项所述的系统,其特征在于,

所述云数据存储平台设置有访问接口,所述访问接口用于接收所述数据提供者上传的所述训练数据。

10.根据权利要求1至7任一项所述的系统，其特征在于，还包括模型推理平台；

所述模型推理平台，用于调用所述训练成果模型，将待处理数据导入所述训练成果模型进行模型推理。

11.根据权利要求1至7任一项所述的系统，其特征在于，所述训练数据设置有数据路由，所述云模型训练平台根据所述数据路由调用所述训练数据。

12.根据权利要求11所述的系统，其特征在于，所述数据路由包括所述训练数据的统一资源定位符路径。

13.一种模型训练方法，其特征在于，应用于提供云服务的模型训练系统，所述模型训练系统包括云数据存储平台、云模型训练平台和云模型存储平台，所述方法包括：

所述云数据存储平台存储训练数据，所述训练数据由数据提供者上传至所述云数据存储平台；

所述云模型存储平台存储待训练模型；

所述云模型训练平台接收用户输入的模型训练创建指令，根据所述模型训练创建指令从所述云模型存储平台获取所述待训练模型，调用所述云数据存储平台存储的所述训练数据，根据所述训练数据训练所述待训练模型，得到训练成果模型，其中，所述待训练模型由模型提供者或所述用户上传至所述云模型存储平台。

14.根据权利要求13所述的模型训练方法，其特征在于，所述模型训练系统还包括鉴权中心，所述方法还包括：

所述鉴权中心接收所述用户输入的鉴权许可请求，所述鉴权许可请求用于确定所述训练数据的权限。

15.根据权利要求13所述的模型训练方法，其特征在于，所述方法还包括：

所述云数据存储平台接收所述数据提供者提供的所述训练数据的标签，所述训练数据的标签用于表征所述训练数据的内容。

16.根据权利要求13所述的模型训练方法，其特征在于，所述模型训练系统还包括检索数据平台，所述方法还包括：

所述检索数据平台获取所述训练数据的信息，所述训练数据的信息包括所述训练数据的数据所有者信息、所述训练数据的数据上传日期中的任意一种或全部。

17.根据权利要求16所述的模型训练方法，其特征在于，所述方法还包括：

所述检索数据平台根据所述训练数据的标签和所述训练数据的信息中的至少一个建立所述训练数据的数据索引表；

所述检索数据平台接收所述用户输入的包括检索关键词的检索指令，根据所述检索关键词在所述数据索引表中进行查找以生成检索结果，所述检索结果包括与所述检索关键词相关的训练数据的信息或与所述检索关键词相关的训练数据的标签中的至少一个。

18.根据权利要求17所述的模型训练方法，其特征在于，所述方法还包括：

所述检索数据平台将所述检索结果发送给用户终端以显示给所述用户；

所述检索数据平台接收所述用户终端发送的针对所述检索结果的数据选取指令，所述数据选取指令用于指示所述数据检索平台从所述检索结果中确定所述训练数据。

19.根据权利要求13所述的模型训练方法，其特征在于，所述方法还包括

所述云模型训练平台在得到所述训练成果模型后，将所述训练成果模型发送至所述云

模型存储平台。

20.根据权利要求13至19任一项所述的模型训练方法,其特征在于,所述模型训练系统还包括数据稽查系统,所述方法还包括:

所述数据稽查系统判定所述数据提供者上传的所述训练数据的有效性。

21.根据权利要求13至19任一项所述的模型训练方法,其特征在于,

所述云数据存储平台设置有访问接口,所述访问接口用于接收所述数据提供者上传的所述训练数据。

22.根据权利要求13至19任一项所述的模型训练方法,其特征在于,还包括模型推理平台;所述方法还包括:

所述模型推理平台调用所述训练成果模型,将待处理数据导入所述训练成果模型进行模型推理。

23.根据权利要求13至19任一项所述的模型训练方法,其特征在于,所述训练数据设置有数据路由,所述云模型训练平台根据所述数据路由调用所述训练数据。

24.根据权利要求23所述的模型训练方法,其特征在于,所述数据路由包括所述训练数据的统一资源定位符路径。

25.一种存储介质,其特征在于,所述存储介质上存储有程序,所述程序被处理器执行时实现如权利要求13至24中任意一项所述的模型训练方法。

模型训练系统、方法和存储介质

技术领域

[0001] 本发明涉及机器学习领域，尤其涉及一种模型训练系统、方法和存储介质。

背景技术

[0002] 深度学习广泛应用于人工智能和计算机视觉等领域。深度学习需要进行模型训练，在模型训练过程中模型开发者需要设计好特定模型，利用数据集进行多次迭代训练，从而得到符合期望要求的深度学习模型。其中，数据集是决定训练出的模型的稳定性和精确度是否符合期望要求的关键。数据集可由数据提供者提供。

[0003] 现阶段，用户可在数据提供商处购买下载数据权限。下载数据权限通过后，用户可将数据下载至本地保存。当需要进行模型训练时，将下载至本地保存的数据拷贝到模型训练系统中，实现模型训练。但是，下载至本地保存的数据发生泄漏的风险较大。

发明内容

[0004] 本申请提供了一种模型训练系统、方法和存储介质，能够降低训练数据发生泄露的风险。

[0005] 第一方面，本申请提供了一种模型训练系统，包括云数据存储平台和云模型训练平台；云数据存储平台用于存储训练数据，以及用于接收训练数据调用请求，根据训练数据调用请求，将与数据调用指令对应的训练数据导出至云模型训练平台；云模型训练平台用于接收模型训练创建指令，获取待训练模型，以及用于生成并向云数据存储平台发送训练数据调用请求，以及用于利用从云数据存储平台导出的训练数据，训练待训练模型，得到训练成果模型。

[0006] 根据第一方面，在第一方面的第一种可能中，模型训练系统还包括检索数据平台和鉴权中心；云数据存储平台包括权限网关；检索数据平台用于根据数据提供者提供的训练数据，建立数据索引表，以及用于接收检索指令，根据检索指令在数据索引表中进行数据检索，并生成检索结果，以及用于接收用户终端针对检索结果的数据选取指令，根据数据选取指令向鉴权中心发起鉴权许可请求，鉴权许可请求包括训练数据的数据标识；鉴权中心用于接收鉴权许可请求，根据鉴权许可请求创建数据标识的数据令牌，并将数据令牌下发给权限网关和用户终端；云模型训练平台还用于向权限网关发送训练数据调用请求，训练数据调用请求包括鉴权中心下发至用户终端的数据令牌；权限网关用于建立第一对应关系，第一对应关系为数据标识与数据令牌一一对应的关系，以及用于接收训练数据调用请求，根据训练数据调用请求中的数据令牌，在第一对应关系中查找目标数据标识，目标数据标识为与训练数据调用请求中的数据令牌对应的数据标识，以及用于将目标数据标识对应的训练数据导出至云模型训练平台。

[0007] 根据第一方面，在第一方面的第二种可能中，模型训练系统还包括检索数据平台和鉴权中心；云数据存储平台包括权限网关和至少一个数据存储服务器；检索数据平台用于根据数据提供者提供的训练数据，建立数据索引表，以及接收检索指令，根据检索指令在

数据索引表中进行数据检索，并生成检索结果，以及用于接收用户终端针对检索结果的数据选取指令，根据数据选取指令向鉴权中心发起鉴权许可请求，鉴权许可请求包括训练数据的数据标识；鉴权中心用于接收鉴权许可请求，根据鉴权许可请求创建数据标识的数据令牌，并将数据令牌下发给权限网关和用户终端；云模型训练平台还用于向权限网关发送训练数据调用请求，训练数据调用请求包括鉴权中心下发至用户终端的数据令牌；权限网关用于建立第二对应关系，第二对应关系为数据令牌与数据路由的对应关系，数据路由包括训练数据的统一资源定位符路径，以及用于接收训练数据调用请求，根据训练数据调用请求中的数据令牌，在第二对应关系中查找目标数据路由，目标数据路由为与训练数据调用请求中的数据令牌对应的数据路由，以及用于访问目标数据存储服务器，以将目标数据存储服务器中目标数据路由指示的训练数据导出至云模型训练平台，目标数据存储服务器为与目标数据路由对应的数据存储服务器。

[0008] 根据第一方面的第二种可能，在第一方面的第三种可能中，模型训练系统还包括访问路由器，权限网关通过访问路由器中预定的标准访问接口从目标数据存储服务器中导出目标数据路由指示的训练数据。

[0009] 根据第一方面的第一种可能或第二种可能，在第一方面的第四种可能中，权限网关还用于获取更新判断参数，判断更新判断参数是否满足更新条件，以及用于若判定更新判断参数满足更新条件，向鉴权中心发送更新请求，以及用于与鉴权中心同步更新数据令牌；鉴权中心还用于接收更新请求，根据更新请求更新数据令牌。

[0010] 根据第一方面的第四种可能，在第一方面的第五种可能中，更新判断参数包括对鉴权许可请求的拒绝次数；权限网关还用于监测鉴权中心对鉴权许可请求的处理过程，以及用于若监测到鉴权中心对鉴权许可请求的拒绝次数超出更新条件中的拒绝次数更新阈值，则向鉴权中心发送更新请求。

[0011] 根据第一方面的第五种可能，在第一方面的第六种可能中，更新判断参数包括训练数据的调用次数；权限网关还用于获取一段时长内的训练数据的调用次数，以及用于若在一段时长内，同一训练数据的调用次数超出更新条件中的调用次数更新阈值，则向鉴权中心发送更新请求。

[0012] 根据第一方面，在第一方面的第七种可能中，云模型训练平台还用于训练得到训练成果模型后，销毁云模型训练平台内训练训练成果模型所利用的训练数据和待训练模型。

[0013] 根据第一方面，在第一方面的第八种可能中，模型训练系统还包括数据稽查系统；数据稽查系统用于对数据提供者上传的训练数据进行有效性认证，拒绝将有效性认证失败的训练数据存入云数据存储平台。

[0014] 根据第一方面，在第一方面的第九种可能中，模型训练系统还包括云模型存储平台；云模型存储平台用于提供待训练模型，以及保存训练成果模型。

[0015] 根据第一方面的第九种可能，在第一方面的第十种可能中，模型训练系统还包括镜像平台和模型推理平台；镜像平台用于存储模型推理运行环境；模型推理平台用于接收推理请求，推理请求包括待处理数据，以及从镜像平台加载模型推理运行环境，以及从云模型存储平台调用训练成果模型，将待处理数据导入训练成果模型进行模型推理。

[0016] 第二方面，本申请提供了一种模型训练方法，包括：云模型训练平台接收模型训练

创建指令，获取待训练模型；云模型训练平台生成并向云数据存储平台发出训练数据调用请求，以调用云数据存储平台中存储的训练数据；云数据存储平台接收训练数据调用请求，将与训练数据调用请求对应的训练数据导出至云模型训练平台；云模型训练平台利用从云数据存储平台导出的训练数据，训练待训练模型，得到训练成果模型。

[0017] 根据第二方面，在第二方面的第一种可能中，上述模型训练方法还包括：检索数据平台根据数据提供者提供的训练数据，建立数据索引表；检索数据平台接收检索指令，根据检索指令在数据索引表中进行数据检索，并生成检索结果；检索数据平台接收用户终端的数据选取指令，根据数据选取指令向鉴权中心发起鉴权许可请求，鉴权许可请求包括训练数据的数据标识；鉴权中心接收鉴权许可请求，根据鉴权许可请求创建数据标识的数据令牌，并将数据令牌下发给权限网关和用户终端；权限网关根据下发得到的数据令牌，建立第一对应关系，第一对应关系为数据标识与数据令牌一一对应的关系。

[0018] 根据第二方面的第一种可能，在第二方面的第二种可能中，云模型训练平台生成并向云数据存储平台发送训练数据调用请求，包括：云模型训练平台生成并向权限网关发送训练数据调用请求，训练数据调用请求包括鉴权中心下发至用户终端的数据令牌；云数据存储平台接收训练数据调用请求，将与训练数据调用请求对应的训练数据导出至云模型训练平台，包括：云数据存储平台中的权限网关接收训练数据调用请求，根据训练数据调用请求中的数据令牌，在第一对应关系中查找目标数据标识，并将目标数据标识对应的训练数据导出至云模型训练平台，目标数据标识为与训练数据调用请求中的数据令牌对应的数据标识。

[0019] 根据第二方面，在第二方面的第三种可能中，上述模型训练方法还包括：检索数据平台根据数据提供者提供的训练数据，建立数据索引表；检索数据平台接收检索指令，根据检索指令在数据索引表中进行数据检索，并生成并发送检索结果；检索数据平台接收用户终端针对检索结果的数据选取指令，根据数据选取指令向鉴权中心发起鉴权许可请求，鉴权许可请求包括训练数据的数据标识；鉴权中心接收鉴权许可请求，根据鉴权许可请求创建数据标识的数据令牌，并将数据令牌下发给权限网关和用户终端；权限网关根据下发得到的数据令牌，建立第二对应关系，第二对应关系为数据令牌与数据路由的对应关系，数据路由包括训练数据的统一资源定位符路径。

[0020] 根据第二方面的第三种可能，在第二方面的第四种可能中，云模型训练平台生成并向云数据存储平台发送训练数据调用请求，包括：云模型训练平台生成并向权限网关发送训练数据调用请求，训练数据调用请求包括鉴权中心下发至用户终端的数据令牌；云数据存储平台接收训练数据调用请求，将与训练数据调用请求对应的训练数据导出至云模型训练平台，包括：云数据存储平台中的权限网关接收训练数据调用请求，根据训练数据调用请求中的数据令牌，在第二对应关系中查找目标数据路由，目标数据路由为与训练数据调用请求中的数据令牌对应的数据路由；权限网关访问目标数据存储服务器，以将目标数据存储服务器中目标数据路由指示的训练数据导出至云模型训练平台，目标数据存储服务器为与目标数据路由对应的数据存储服务器。

[0021] 根据第二方面或第二方面的第一种可能至第四种可能中的任意一种可能，在第二方面的第五种可能中，上述模型训练方法还包括：权限网关获取更新判断参数，判断更新判断参数是否满足更新条件；若判定更新判断参数满足更新条件，权限网关向鉴权中心发送

更新请求;鉴权中心接收更新请求,根据更新请求更新数据令牌;权限网关与鉴权中心同步更新数据令牌。

[0022] 根据第二方面的第五种可能,在第二方面的第六种可能中,更新判断参数包括对鉴权许可请求的拒绝次数;权限网关获取更新判断参数,判断更新判断参数是否满足更新条件,包括:权限网关监测鉴权中心对鉴权许可请求的处理过程,并获取鉴权中心对鉴权许可请求的拒绝次数,并判断鉴权中心对鉴权许可请求的拒绝次数是否超出更新条件中的拒绝次数更新阈值;若判定更新判断参数满足更新条件,权限网关向鉴权中心发送更新请求,包括:若监测到鉴权中心对鉴权许可请求的拒绝次数超出更新条件中的拒绝次数更新阈值,则向鉴权中心发送更新请求。

[0023] 根据第二方面的第五种可能,在第二方面的第七种可能中,更新判断参数包括训练数据的调用次数;权限网关获取更新判断参数,判断更新判断参数是否满足更新条件,包括:权限网关获取一段时长内的训练数据的调用次数,判断在一段时长内,同一训练数据的调用次数是否超出更新条件中的调用次数更新阈值;若判定更新判断参数满足更新条件,权限网关向鉴权中心发送更新请求,包括:若在一段时长内,同一训练数据的调用次数超出更新条件中的调用次数更新阈值,则向鉴权中心发送更新请求。

[0024] 根据第二方面,在第二方面的第八种可能中,在云模型训练平台利用从云数据存储平台导出的训练数据,训练待训练模型,得到训练成果模型之后,还包括:云模型训练平台销毁云模型训练平台内训练训练成果模型所利用的训练数据和待训练模型。

[0025] 根据第二方面,在第二方面的第九种可能中,上述模型训练方法还包括:数据稽查系统对数据提供者上传的训练数据进行有效性认证;数据稽查系统拒绝将有效性认证失败的训练数据存入云数据存储平台。

[0026] 根据第二方面,在第二方面的第十种可能中,在云模型训练平台利用从云数据存储平台导出的训练数据,训练待训练模型,得到训练成果模型之后,还包括:云模型存储平台保存训练成果模型。

[0027] 根据第二方面的第十种可能,在第二方面的第十一种可能中,上述模型训练方法还包括:模型推理平台接收推理请求,推理请求包括待处理数据;模型推理平台从镜像平台加载模型推理运行环境,并从云模型存储平台调用训练成果模型,将待处理数据导入训练成果模型进行模型推理。

[0028] 第三方面,本申请提供了一种存储介质,存储介质上存储有程序,程序被处理器执行时实现上述技术方案中的模型训练方法。

[0029] 本申请提供了一种模型训练系统、方法和存储介质,可应用于深度学习场景中。模型训练系统可包括云数据存储平台和云模型训练平台。云数据存储平台存储训练数据。云模型训练平台接收用户的模型训练创建指令,触发执行模型训练。云模型训练平台通过向云数据存储平台发送训练数据调用请求,调用云数据存储平台存储的训练数据。云模型训练平台利用获取的待训练模型和从云数据存储平台导出的训练数据进行模型训练。在本申请中,云数据存储平台和云模型训练平台相互独立,将训练数据的存储与模型训练两种功能分离。云数据存储平台和云模型训练平台均以云系统为基础实现,模型训练过程在云系统中进行,进行模型训练的用户无法将训练数据下载至本地,训练数据存在于云数据存储平台和正在进行模型训练的云模型训练平台。也就是说,训练数据不会从本地的用户侧泄

露,从而降低了训练数据发生泄露的风险。

附图说明

[0030] 图1为本发明实施例的模型训练系统的应用场景示意图;

[0031] 图2为本发明一实施例中一种模型训练系统的结构示意图;

[0032] 图3为本发明另一实施例中一种模型训练系统的结构示意图;

[0033] 图4为本发明又一实施例中一种模型训练系统的结构示意图;

[0034] 图5为本发明一实施例中一种模型训练方法的流程图;

[0035] 图6为本发明一实施例中一种模型训练方法的一种具体实现方式的流程图;

[0036] 图7为本发明一实施例中一种模型训练方法的另一种具体实现方式的流程图。

具体实施方式

[0037] 本发明实施例提供一种模型训练系统、方法和存储介质,可应用于深度学习(Deep Learning)的场景中,可实现对深度学习模型的训练,也可实现对深度学习模型的应用,比如,利用训练处的深度学习模型进行推理。本发明实施例的模型训练系统可在云端完成模型训练、模型推理等功能。图1为本发明实施例的模型训练系统的应用场景示意图。如图1所示,模型训练系统可在云服务系统上运行,云服务系统可由云系统以及向外提供访问接口的系统集群网关构成。用户可通过用户终端使用账号及密码通过网络连接到云系统。云系统包括多个内部网络互通的服务器。模型训练系统可通过数据模型仓库实现训练数据和训练模型的存储和提供。模型训练系统可通过深度学习数据库实现模型训练系统与用户的人机交互,可通过鉴权服务系统完成用户与模型训练系统的各项权利的鉴权,可通过训练推理系统完成模型的训练和推理。

[0038] 图2为本发明一实施例中一种模型训练系统的结构示意图。如图2所示,模型训练系统包括云数据存储平台11和云模型训练平台12。

[0039] 云数据存储平台11用于存储训练数据,以及用于接收训练数据调用请求,根据训练数据调用请求,将与数据调用指令对应的训练数据导出至云模型训练平台12。

[0040] 训练数据为用于对训练模型所需的数据,云数据存储平台11可存储多个训练数据,训练数据可视为由多条数据形成的数据集。训练数据可包括图像、视频、音频等,在此并不限定。云数据存储平台11在存储训练数据时,可为训练数据分配数据标识,数据标识用于标识训练数据,可作为查找数据存储位置的标识符。在一个示例中,为了区分不同的训练数据,训练数据的数据标识具有唯一性,也就是说,不同的训练数据的数据标识不同。

[0041] 云数据存储平台11可接收数据提供者上传的训练数据。示例性地,数据提供者可利用客户端通过超文本传输协议(HyperText Transfer Protocol,HTTP)连接到云系统的后端,从而与云数据存储平台11进行信息交互。在一个示例中,云数据存储平台11可向数据提供者提供上传训练数据的标准协议,标准协议中可包括数据格式、压缩格式以及数据类型等。云数据存储平台11可对数据提供者上传的训练数据进行检测,若确定数据提供者上传的训练数据不符合标准协议,则云数据存储平台11可拒绝存储不符合标准协议的训练数据。

[0042] 云数据存储平台11中可设置一备份区域,该备份区域可用于对训练数据进行备

份，避免数据出现意外，如数据误操作等导致无法恢复的情况。

[0043] 训练数据调用请求是云模型训练平台12生成并发送的，根据训练数据调用请求可得知云模型训练平台12请求调用的训练数据。在一个示例中，训练数据调用请求可包括数据标识。云数据存储平台11接收训练数据调用请求，可查找训练数据调用请求需要调用的训练数据，并将请求调用的训练数据导出至云模型训练平台12，以供云模型训练平台12利用导出的训练数据进行模型训练。

[0044] 云模型训练平台12用于接收模型训练创建指令，获取待训练模型，以及用于生成并向云数据存储平台11发送训练数据调用请求，以及用于利用从云数据存储平台11导出的训练数据，训练待训练模型，得到训练成果模型。

[0045] 其中，云模型训练平台12可获取用户或模型提供者上传的待训练模型，也可从云系统中的模型数据库中获取待训练模型。

[0046] 在一个示例中，示例性地，用户可利用用户终端20通过超文本传输协议连接到云系统的后端，从而与云模型训练平台12进行信息交互。用户可通过用户终端20向云模型训练平台12发送模型训练创建指令，以触发云模型训练平台12创建模型训练任务。云模型训练平台12可利用待训练模型和训练数据进行模型训练。示例性的，模型训练可指将训练数据导入待训练模型进行多次迭代训练，从而得到经训练后的模型即训练成果模型。

[0047] 需要说明的是用户终端20的使用者可包括用户、数据提供者或模型提供者。

[0048] 本发明实施例中的云数据存储平台11可视为图1中数据模型仓库的一部分。本发明实施例中的云模型训练平台12可视为图1中训练推理系统的一部分。

[0049] 在本发明实施例中，云数据存储平台11和云模型训练平台12相互独立，将训练数据的存储与模型训练两种功能分离。云数据存储平台11和云模型训练平台12均以云系统为基础实现，模型训练过程在云系统中进行，进行模型训练的用户无法将训练数据下载至本地，训练数据存在于云数据存储平台11和正在进行模型训练的云模型训练平台12。也就是说，训练数据不会从本地的用户侧泄露，从而降低了训练数据发生泄露的风险。

[0050] 图3为本发明另一实施例中一种模型训练系统的结构示意图。图3与图2的不同之处在于，图2中的云数据存储平台11还包括图3中的权限网关111；图3所示的模型训练系统还可包括检索数据平台13、鉴权中心14、数据稽查系统15、云模型存储平台16、镜像平台17和模型推理平台18。

[0051] 检索数据平台13用于根据数据提供者提供的训练数据，建立数据索引表。用户可通过检索数据平台13对云数据存储平台11中存储的训练数据进行搜索查询。

[0052] 在一个示例中，在数据提供者上传训练数据后，检索数据平台13可对训练数据进行分析处理，得到训练数据的数据集大小、数据集规模、数据所有者信息、数据上传日期等数据基本信息，便于用户了解训练数据的基本信息。

[0053] 在一个示例中，云数据存储平台11还可要求数据提供者在上传训练数据时，提供训练数据的标签，训练数据的标签可表征训练数据的特征。具体的，训练数据的标签可以为训练数据表征的内容的关键词。比如，数据提供者在上传训练数据时，为训练数据标记的标签为“车牌”和“小型车”。检索数据平台13在建立数据索引表的过程中，也可将训练数据的标签添加入数据检索表，以便于用户在检索训练数据时，利用训练数据的特征进行检索。

[0054] 检索数据平台13用于接收检索指令，根据检索指令在数据索引表中进行数据检

索，并生成检索结果。具体的，检索指令中可包括一个或多个检索关键词，可根据检索关键词在数据索引表中的训练数据的标签中进行查找。检索结果可包括与检索指令中的检索关键词相关的训练数据的信息，比如训练数据的名称、编号、关键词以及训练数据中的部分数据示例等。在一个示例中，检索结果可包括按照与检索关键词的相关程度的大小依次排列的训练数据的信息，使用户能够更直观地得到与检索关键字最相关的训练数据。在另一个示例中，也可在根据检索关键词检索到的训练数据的信息中随机筛选固定数据的训练数据的信息提供给用户。比如，每次检索生成的检索结果包括十条训练数据的信息。检索数据平台13可将检索结果发送给用户终端20，用户终端20可显示检索结果。

[0055] 用户接收到检索结果后，还可通过用户终端20针对检索结果发出数据选取指令。数据检索平台接收用户终端20针对检索结果的数据选取指令，根据数据选取指令向鉴权中心14发起鉴权许可请求。数据选取指令可用于指示选取检索结果中的一项或多项训练数据的信息，从而确定模型训练需要的训练数据。

[0056] 确定模型训练需要的训练数据后，向鉴权中心14发起鉴权许可请求，鉴权许可请求可包括训练数据的数据标识，向鉴权中心14请求训练数据的调用权限。

[0057] 本发明实施例中的检索数据平台13可视为图1中的深度学习数据库的至少一部分。

[0058] 鉴权中心14用于接收鉴权许可请求，根据鉴权许可请求创建数据标识的数据令牌，并将数据令牌下发给权限网关111和用户终端20。

[0059] 鉴权许可请求用于请求训练数据的调用权限。鉴权中心14可决定是否同意检索数据平台13发送来的鉴权许可请求。示例性的，鉴权许可请求可包括针对训练数据的付费信息，若付费信息表明用户对针对训练数据付费成功，鉴权中心14可同意鉴权许可请求，并创建数据标识的数据令牌。鉴权中心14同意鉴权许可请求后，还可生成并保存数据鉴权信息，数据鉴权信息可包括用户标识和数据标识。示例性的，数据鉴权信息可具有有效时长，即在有效时长内，若用户再次请求同样的训练数据时，鉴权许可请求可直接被鉴权中心14同意通过，不需要进行审核。有效时长可根据工作场景和工作需求设定，在此并不限定。比如，有效时长可为一年或永久。

[0060] 数据令牌(即数据Token)可标识某个操作中的训练数据，作为数据调用的一种安全凭证使用。比如，数据令牌标识后续过程中数据调用操作中的训练数据。在一个示例中，数据令牌可实现为安全插件。鉴权中心14将创建的数据令牌下发给用户终端20，以使得用户终端20可利用数据令牌通过权限网关111从云数据存储平台11导出与数据令牌对应的训练数据。同时，鉴权中心14也将创建的数据令牌保存在鉴权中心14。

[0061] 云模型训练平台12还用于向权限网关111发送训练数据调用请求，训练数据调用请求包括鉴权中心14下发至用户终端20的数据令牌。

[0062] 比如，用户终端20在请求训练数据时，可将数据令牌添加入模型训练创建指令，云模型训练平台12可解析模型训练创建指令，得到下发至用户终端20的数据令牌，并将下发至用户终端20的数据令牌添加入训练数据调用请求中。云模型训练平台12通过训练数据调用请求中的数据令牌从云数据存储平台11调用与数据令牌对应的训练数据。

[0063] 在一种实现方式中，云数据存储平台11具体可实现为第三方公用服务器。第三方公用服务器不属于数据提供者、模型提供者和用户，是一个公用的用于存储训练数据且能

够导出训练数据的服务器。调用训练数据可利用数据令牌与数据标识的对应关系进行授权调用。

[0064] 权限网关111用于建立第一对应关系，第一对应关系为数据标识与数据令牌的对应关系。数据标识与数据令牌一一对应，数据令牌也具有唯一性，也就是说，不同的数据标识对应不同的数据令牌。权限网关111在接收到训练数据调用请求时，根据训练数据调用请求中的数据令牌，在第一对应关系中查找目标数据标识，目标数据标识为与训练数据调用请求中的数据令牌对应的数据标识，并将目标数据标识对应的训练数据导出至云模型训练平台12。

[0065] 当云数据存储平台11接收到训练数据调用请求后，权限网关111会对比训练数据调用请求中的数据令牌是否与权限网关111中存储的数据令牌；若训练数据调用请求中的数据令牌能够与权限网关111中存储的数据令牌匹配，则允许调用训练数据，并将与训练数据调用请求中的数据令牌对应的训练数据导出。

[0066] 为了保障模型训练过程中的数据安全，避免训练数据被越权使用，可根据实际情况对数据令牌进行更新。权限网关111可用于获取更新判断参数，判断更新判断参数是否满足更新条件。若判定更新判断参数满足更新条件，权限网关111向鉴权中心14发送更新请求，以及用于与鉴权中心14同步更新数据令牌。鉴权中心14接收更新请求，根据更新请求更新数据令牌。

[0067] 更新判断参数可包括对鉴权许可请求的拒绝次数、训练数据的调用次数、数据令牌的存在时长等参数中的一项或多项。

[0068] 比如，更新判断参数包括对鉴权许可请求的拒绝次数。权限网关111可监测鉴权中心14对鉴权许可请求的处理过程，从而得到鉴权中心14对鉴权许可请求的拒绝次数。若权限网关111监测到鉴权中心14对鉴权许可请求的拒绝次数超出更新条件中的拒绝次数更新阈值，则向鉴权中心14发送更新请求。

[0069] 拒绝次数更新阈值可根据工作场景和工作需求设定，在此并不限定。鉴权中心14删除原数据令牌，并生成新的数据令牌，并将新的数据令牌下发给客户终端和权限网关111，以使得权限网关111可以与鉴权中心14同步更新数据令牌。数据令牌在鉴权中心14和权限网关111中更新时，需要停止训练数据调用请求的执行，待鉴权中心14和权限网关111中的数据令牌更新完毕后，再执行训练数据调用请求。在数据令牌更新完毕后，若训练数据调用请求中包含的仍然是原数据令牌，训练数据调用请求中的原数据令牌失效，则无法调用训练数据。

[0070] 又比如，更新判断参数包括训练数据的调用次数。权限网关111可获取一段时长内的训练数据的调用次数。若在一段时长内，权限网关111确定同一训练数据的调用次数超出更新条件中的调用次数更新阈值，则向鉴权中心14发送更新请求。统计训练数据的一段时长和调用次数更新阈值可根据工作场景和工作需求设定，在此并不限定。

[0071] 还比如，更新判断参数包括数据令牌的存在时长。权限网关111可设置数据令牌的更新周期时长，并记录数据令牌的存在时长。若权限网关111确定数据令牌的存在时长达到更新周期时长，则向鉴权中心14发送更新请求。数据令牌的更新周期时长可根据工作场景和工作需求设定，在此并不限定。

[0072] 需要说明的是，更新判断参数和更新条件并不限于上述举例。权限网关111也可接

收用户的更新策略配置指令，根据更新策略配置指令设置更新判断参数和更新条件。

[0073] 云模型存储平台16用于提供待训练模型，以及保存训练成果模型。云模型存储平台16中存储的模型可以是模型提供者上传的模型，也可以是云模型训练平台12训练得到的训练成果模型。

[0074] 在一个示例中，上述云模型训练平台12在训练得到训练成果模型后，可将训练成果模型发送至云模型存储平台16保存，并销毁云模型训练平台12内训练训练成果模型所利用的训练数据和待训练模型，还可将云模型训练平台12内的训练成果模型销毁，以防止遗留在云模型训练平台12的训练数据和模型即待训练模型和训练成果模型泄露。

[0075] 在一个示例中，数据稽查系统15先于云数据存储平台11接收到数据提供者上传的训练数据。数据稽查系统15用于对数据提供者上传的训练数据进行有效性认证，拒绝将有效性认证失败的训练数据存入云数据存储平台11。比如，若数据提供者上传的训练数据与云数据存储平台11存储的训练数据重复，或者数据提供者上传的数据的数据格式不符合云数据存储平台11的标准协议，则数据稽查系统15判定数据提供者上传的训练数据无效，即上传的训练数据有效性认证失败。若数据稽查系统15判定数据提供者上传的训练数据有效，则可通过检索数据平台13向云数据存储平台11发送存储指令，以使得云数据存储平台11将数据提供者上传的训练数据持久存储。

[0076] 需要说明的是，对数据提供者上传的训练数据进行有效性认证的方式并不限于上述方式。数据稽查系统15可保证模型训练系统中所使用的训练数据的真实有效性。

[0077] 镜像平台17用于存储模型推理运行环境。具体的，模型推理运行环境可包括系统环境和训练成果模型对应的运行框架环境。

[0078] 模型推理平台18可接收推理请求，推理请求包括待处理数据。推理请求可由用户终端20发送。示例性的，用户终端20可通过应用程序编程接口(Application Programming Interface，API)向模型推理平台18发送推理请求。模型推理平台18接收推理请求后，从镜像平台17加载模型推理运行环境，并从云模型存储平台16调用训练成果模型，将待处理数据导入训练成果模型进行模型推理。

[0079] 本发明实施例中的数据检索平台可视为图1中深度学习数据库中的至少一部分。本发明实施例中的鉴权中心14可视为图1中鉴权服务系统中的至少一部分。本发明实施例中的模型推理平台18可视为图1中训练推理系统中的一部分。

[0080] 图4为本发明又一实施例中一种模型训练系统的结构示意图。图4所示的模型训练系统与图3所示的模型训练系统的不同之处在于，云数据存储平台11可实现为数据提供者的至少一个私有服务器。

[0081] 在云数据存储平台11包括权限网关111和至少一个数据存储服务器112即私有服务器的条件下，调用训练数据可利用数据令牌与数据路由的对应关系进行授权调用。

[0082] 数据路由可包括训练数据的统一资源定位符(Uniform Resource Locator，URL)路径，还可包括数据访问方法和从云数据存储平台11导出训练数据的标准。数据提供者在上传训练数据的同时也可上传训练数据对应的数据路由至检索数据平台13。

[0083] 检索数据平台13也可对数据路由进行合法性检测，若确定数据路由不合法，则拒绝存储数据路由。比如，检索数据平台13确定数据路由无法访问或数据路由的格式不符合模型训练系统中预设的标准，则拒绝存储数据路由。示例性的，检索数据平台13可向权限网

关111和鉴权中心14发送拒绝指令，以使得权限网关111和鉴权中心14均拒绝存储路由数据。

[0084] 权限网关111可建立第二对应关系，第二对应关系为数据令牌与数据路由的对应关系。示例性的，第二对应关系可实现为数据路由表。训练数据具有对应的数据路由，训练数据与数据令牌一一对应，数据令牌与数据路由也一一对应。在检索数据平台13建立数据索引表时，可将对应的数据路由保存在权限网关111中。

[0085] 权限网关111接收训练数据调用请求后，根据训练数据调用请求中的数据令牌，在第二对应关系中查找目标数据路由。目标数据路由为与训练数据调用请求中的数据令牌对应的数据路由。权限网关111可根据与数据令牌对应的数据路由，访问目标数据存储服务器112，以将目标数据存储服务器112中目标数据路由指示的训练数据导出至云模型训练平台12。目标数据存储服务器112为与目标数据路由对应的数据存储服务器112。

[0086] 为了保证数据存储服务器112即私有服务器中的训练数据的安全性，可建立安全加密远程访问。在一个实例中，模型训练系统还可包括访问路由器。权限网关111通过访问路由器中预定的标准访问接口从目标数据存储服务器112中导出目标数据路由指示的训练数据。比如，标准访问接口为restful访问接口，并可将restful访问接口的路径作为数据路由。

[0087] 在一个示例中，为了进一步保证数据存储服务器112中的训练数据的安全性。权限网关111可随机选取数据令牌，并验证数据令牌的合法性。若权限网关111确定数据令牌非法，则可更新数据路由表，即更新第二对应关系，具体可实现为更新第二对应关系中的数据令牌。

[0088] 图5为本发明一实施例中一种模型训练方法的流程图。该模型训练方法可适用于上述实施例中的模型训练系统。如图5所示，模型训练方法可包括步骤S201和步骤S204。

[0089] 在步骤S201中，云模型训练平台接收模型训练创建指令，获取待训练模型；

[0090] 在步骤S202中，云模型训练平台生成并向云数据存储平台发出训练数据调用请求，以调用云数据存储平台中存储的训练数据；

[0091] 在步骤S203中，云数据存储平台接收训练数据调用请求，将与训练数据调用请求对应的训练数据导出至云模型训练平台；

[0092] 在步骤S204中，云模型训练平台利用从云数据存储平台导出的训练数据，训练待训练模型，得到训练成果模型。

[0093] 上述步骤S201至步骤S204的说明可参见上述实施例中的云模型训练平台和云数据存储平台的相关说明。

[0094] 在本发明实施例中，云数据存储平台和云模型训练平台相互独立，将训练数据的存储与模型训练两种功能分离。云数据存储平台和云模型训练平台均以云系统为基础实现，模型训练过程在云系统中进行，进行模型训练的用户无法将训练数据下载至本地，训练数据存在于云数据存储平台和正在进行模型训练的云模型训练平台。也就是说，训练数据不会从本地的用户侧泄露，从而降低了训练数据发生泄露的风险。

[0095] 图6为本发明一实施例中一种模型训练方法的一种具体实现方式的流程图。如图6所示，模型训练方法可包括步骤S301至步骤S315。

[0096] 在步骤301中，数据稽查系统对数据提供者上传的训练数据进行有效性认证。

[0097] 在步骤302中，数据稽查系统拒绝将有效性认证失败的训练数据存入云数据存储平台。

[0098] 在步骤303中，检索数据平台根据数据提供者提供的训练数据，建立数据索引表。

[0099] 在步骤304中，检索数据平台接收检索指令，根据检索指令在数据索引表中进行数据检索，并生成检索结果。

[0100] 在步骤305中，检索数据平台接收用户终端的数据选取指令，根据数据选取指令向鉴权中心发起鉴权许可请求。

[0101] 其中，鉴权许可请求包括训练数据的数据标识。

[0102] 在步骤306中，鉴权中心接收鉴权许可请求，根据鉴权许可请求创建数据标识的数据令牌，并将数据令牌下发给云数据存储平台中的权限网关和用户终端。

[0103] 在步骤307中，云数据存储平台中的权限网关根据下发得到的数据令牌，建立第一对应关系。

[0104] 其中，第一对应关系为数据标识与数据令牌的对应关系。

[0105] 在步骤308中，云模型训练平台接收模型训练创建指令，获取待训练模型。

[0106] 在步骤309中，云模型训练平台生成并向云数据存储平台中的权限网关发送训练数据调用请求，以调用云数据存储平台中存储的训练数据。

[0107] 其中，训练数据调用请求包括鉴权中心下发至用户终端的数据令牌。

[0108] 在步骤310中，云数据存储平台中的权限网关接收训练数据调用请求，根据训练数据调用请求中的数据令牌，在第一对应关系中查找目标数据标识，并将目标数据标识对应的训练数据导出至云模型训练平台。

[0109] 其中，目标数据标识为与训练数据调用请求中的数据令牌对应的数据标识。

[0110] 在步骤311中，云模型训练平台利用从云数据存储平台导出的训练数据，训练待训练模型，得到训练成果模型。

[0111] 在步骤312中，云模型存储平台保存训练成果模型。

[0112] 在步骤313中，云模型训练平台销毁云模型训练平台内训练训练成果模型所利用的训练数据和待训练模型。

[0113] 在步骤314中，模型推理平台接收推理请求，推理请求包括待处理数据。

[0114] 在步骤315中，模型推理平台从镜像平台加载模型推理运行环境，并从云模型存储平台调用训练成果模型，将待处理数据导入训练成果模型进行模型推理。

[0115] 图7为本发明一实施例中一种模型训练方法的另一种具体实现方式的流程图。图7与图6的不同之处在于，图6中的步骤S307可替换为图7中的步骤S316；图6中的步骤S310可替换为图7中的步骤S317和步骤S318。

[0116] 在步骤S316中，云数据存储平台中的权限网关根据下发得到的数据令牌，建立第二对应关系。

[0117] 其中，第二对应关系为数据令牌与数据路由的对应关系。数据路由包括训练数据的统一资源定位符路径。

[0118] 在步骤S317中，云数据存储平台中的权限网关接收训练数据调用请求，根据训练数据调用请求中的数据令牌，在第二对应关系中查找目标数据路由。

[0119] 其中，目标数据路由为与训练数据调用请求中的数据令牌对应的数据路由。

[0120] 在步骤S318中，云数据存储平台中的权限网关访问目标数据存储服务器，以将目标数据存储服务器中目标数据路由指示的训练数据导出至云模型训练平台。

[0121] 其中，目标数据存储服务器为与目标数据路由对应的数据存储服务器。

[0122] 在一个示例中，还可以根据具体场景对数据令牌进行更新，从而保证训练数据的安全。权限网关获取更新判断参数，判断更新判断参数是否满足更新条件。若判定更新判断参数满足更新条件，权限网关向鉴权中心发送更新请求。鉴权中心接收更新请求，根据更新请求更新数据令牌。权限网关与鉴权中心同步更新数据令牌。

[0123] 示例性的，更新判断参数包括对鉴权许可请求的拒绝次数。数据令牌更新过程可具体为：权限网关监测鉴权中心对鉴权许可请求的处理过程，并获取鉴权中心对鉴权许可请求的拒绝次数，并判断鉴权中心对鉴权许可请求的拒绝次数是否超出更新条件中的拒绝次数更新阈值；若监测到鉴权中心对鉴权许可请求的拒绝次数超出更新条件中的拒绝次数更新阈值，则向鉴权中心发送更新请求。

[0124] 示例性的，更新判断参数包括训练数据的调用次数。数据令牌更新过程可具体为：权限网关获取一段时长内的训练数据的调用次数，判断在一段时长内，同一训练数据的调用次数是否超出更新条件中的调用次数更新阈值；若在一段时长内，同一训练数据的调用次数超出更新条件中的调用次数更新阈值，则向鉴权中心发送更新请求。

[0125] 上述方法实施例中各步骤的说明内容可参照上述系统实施例中的相关说明。

[0126] 本发明实施例还可提供一种存储介质，该存储介质上存储有程序，程序被处理器执行时实现上述实施例中的模型训练方法。

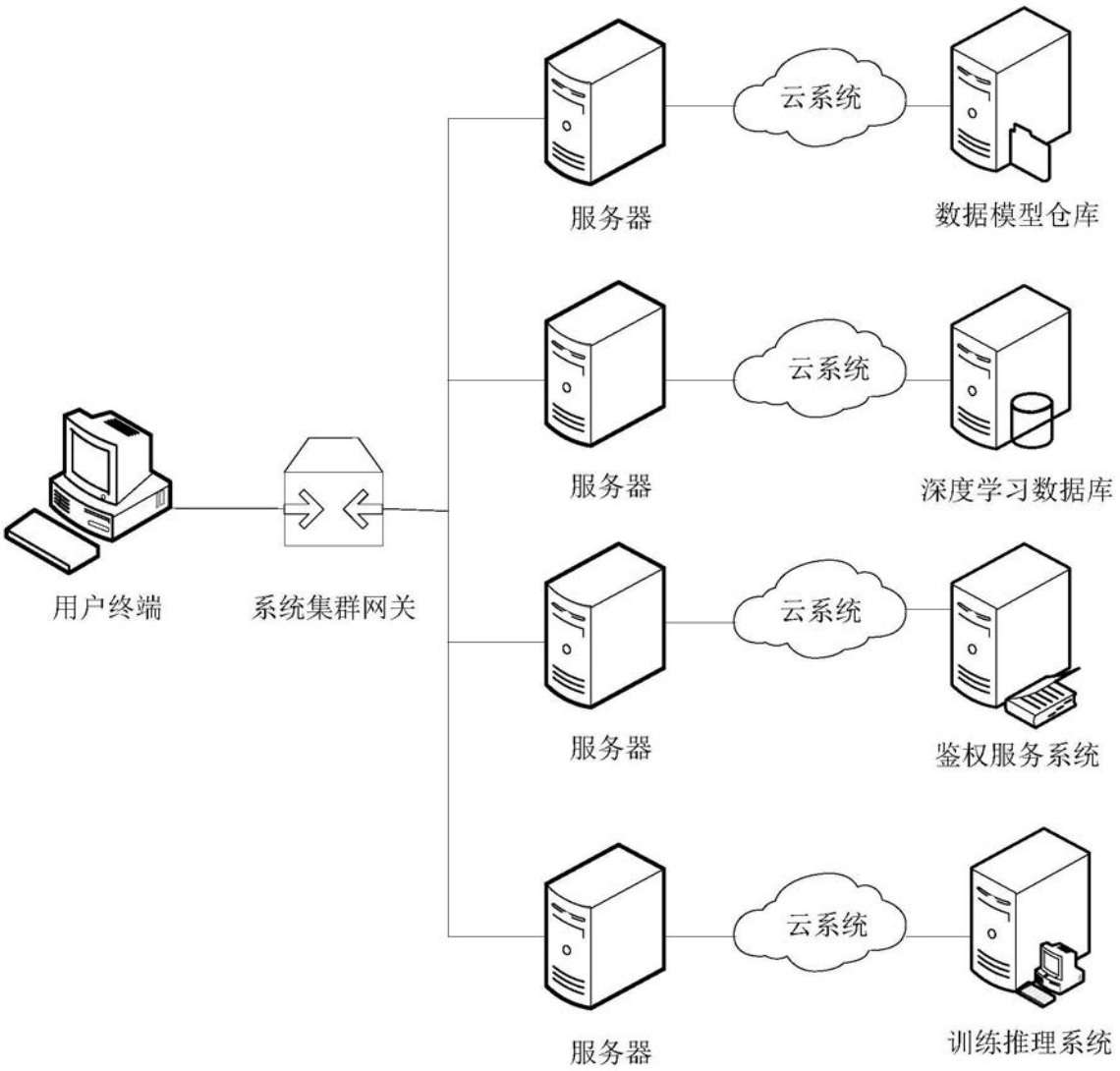
云系统
服务器
数据模型仓库
云系统
服务器
深度学习数据库
用户终端
系统集群网关
云系统
服务器
鉴权服务系统
云系统
服务器
训练推理系统

图1

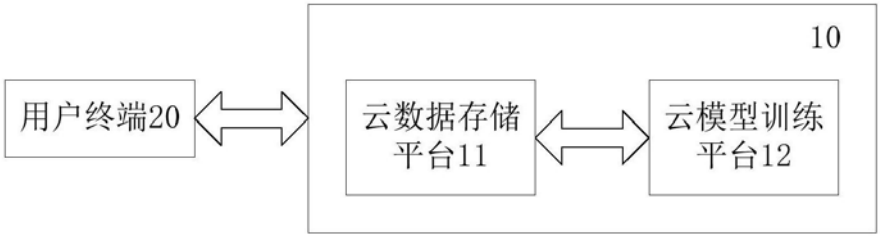
10
用户终端20
云数据存储
平台11
云模型训练
平台12

图2

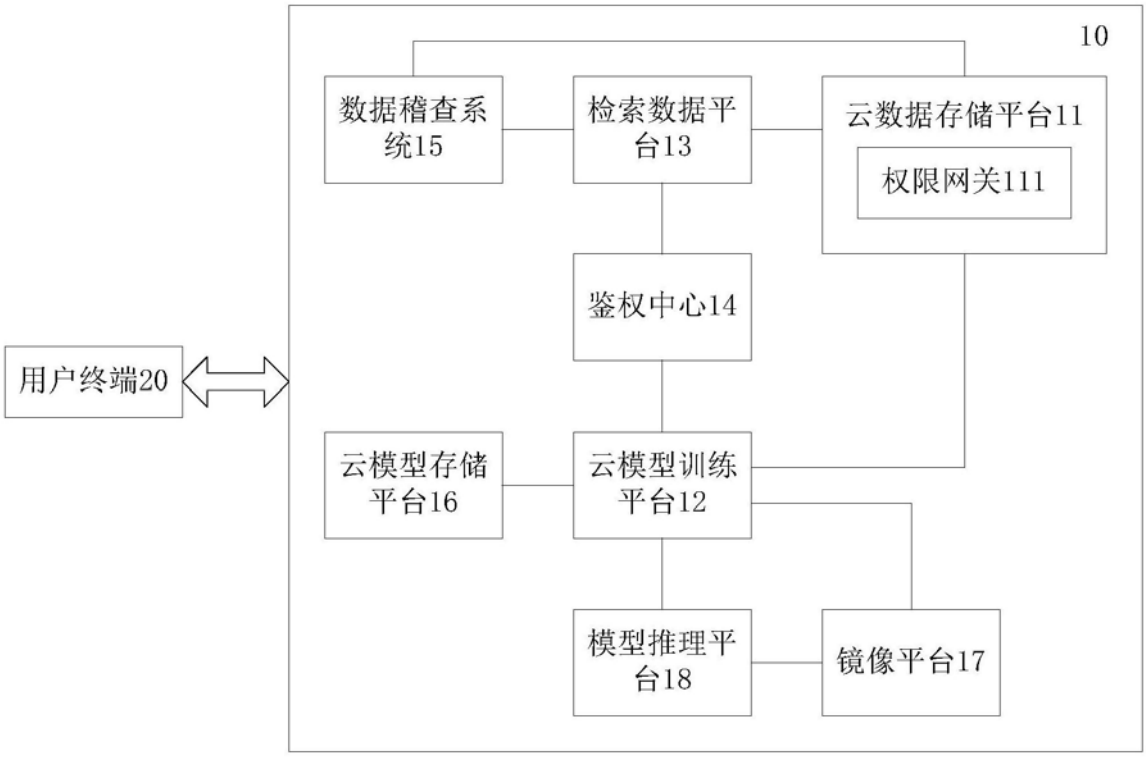
10
数据稽查系统15
检索数据平台13
云数据存储平台11
权限网关111
鉴权中心14
用户终端20
云模型存储平台16
云模型训练平台12
模型推理平台18
镜像平台17

图3

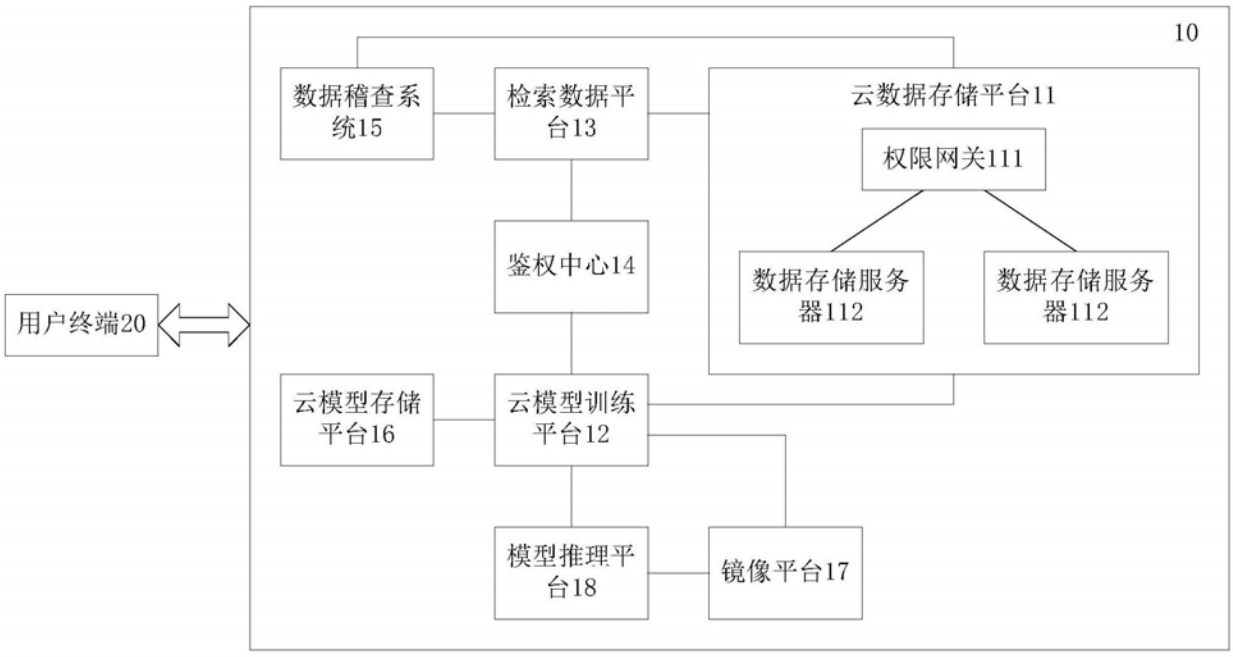
10
数据稽查系统15
检索数据平台13
云数据存储平台11
权限网关111
鉴权中心14
数据存储服务器112
数据存储服务器112
用户终端20
云模型存储平台16
云模型训练平台12
模型推理平台18
镜像平台17

图4

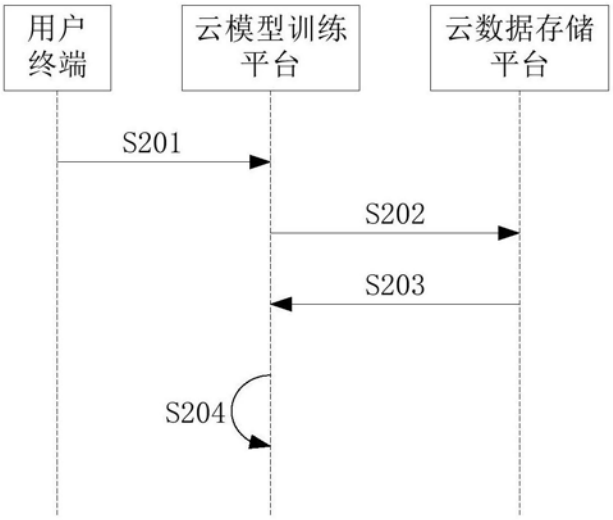

图5

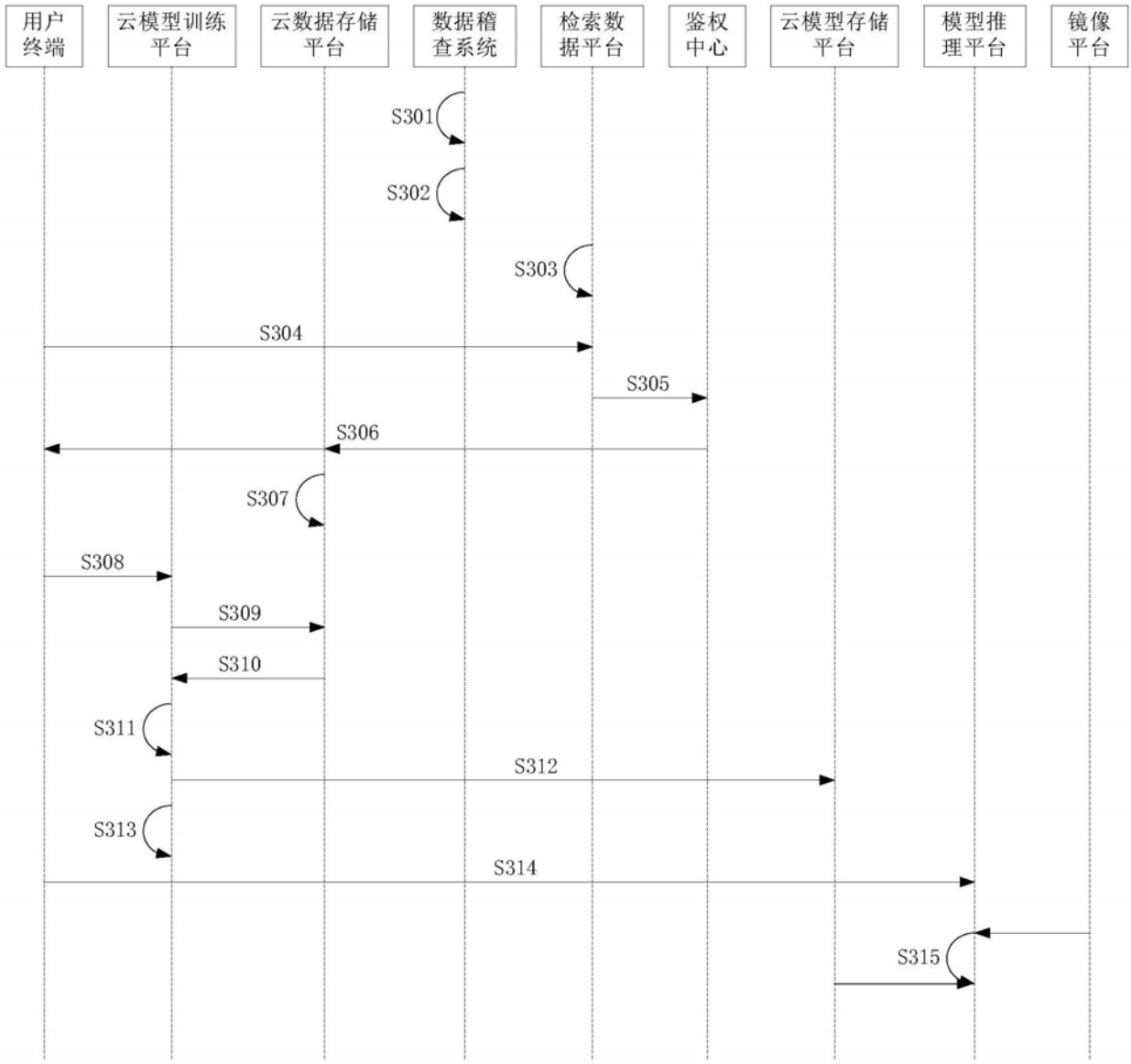
用户终端
云模型训练平台
云数据存储平台
数据稽查系统
检索数据平台
鉴权中心
云模型存储平台
模型推理平台
镜像平台
S301
S302
S303
S304
S305
S306
S307
S308
S309
S310
S311
S312
S313
S314
S315

图6

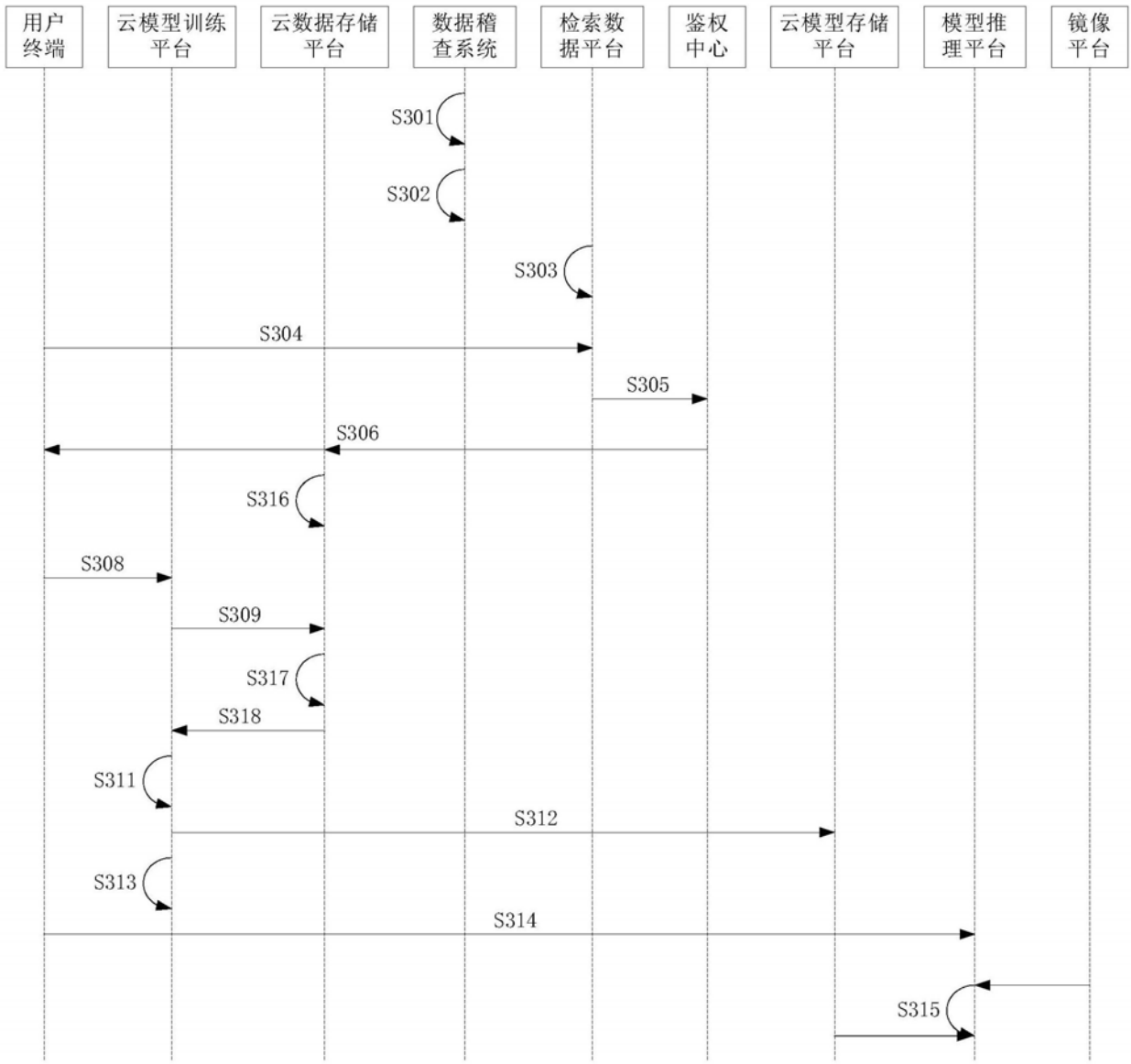
用户终端
云模型训练平台
云数据存储平台
数据稽查系统
检索数据平台
鉴权中心
云模型存储平台
模型推理平台
镜像平台
S301
S302
S303
S304
S305
S306
S316
S308
S309
S317
S318
S311
S312
S313
S314
S315

图7