



(12) 发明专利申请

(10) 申请公布号 CN 113065843 A

(43) 申请公布日 2021.07.02

(21) 申请号 202110275915.3

(22) 申请日 2021.03.15

(71) 申请人 腾讯科技(深圳)有限公司

地址 518057 广东省深圳市南山区高新区
科技中一路腾讯大厦35层

(72) 发明人 晏赛 骆颖民 赵丽丽 杨蕾
刘才良 陈保贺 王玢 张肇勇
文书凡

(74) 专利代理机构 广州三环专利商标代理有限
公司 44202

代理人 熊永强 贾允

(51) Int. Cl.

G06Q 10/10 (2012.01)

G06K 9/62 (2006.01)

G06N 20/00 (2019.01)

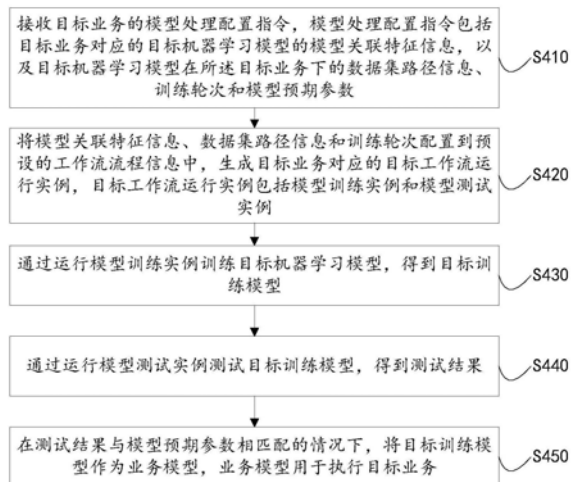
权利要求书3页 说明书17页 附图10页

(54) 发明名称

一种模型处理方法、装置、电子设备和存储
介质

(57) 摘要

本申请公开了一种模型处理方法、装置、电
子设备和存储介质,所述方法包括:接收业务模
型的模型处理配置指令,将模型处理配置指令中
的内容配置到预设的工作流程信息中,生成业
务模型对应的目标工作流运行实例,目标工作流
运行实例包括模型训练实例和模型测试实例。运
行模型训练实例,以训练目标机器学习模型,得
到目标训练模型。运行模型测试实例,以测试目
标训练模型,得到测试结果。在测试结果与模型
预期参数相匹配的情况下,将目标训练模型作为
业务模型。该方法可以自动完成模型训练、测试
等的流程,提高了模型处理的效率,简化了模型
处理的操作,从而节省了人工成本和运维成本。



1. 一种模型处理方法,其特征在于,所述方法包括:

接收目标业务的模型处理配置指令,所述模型处理配置指令包括所述目标业务对应的目标机器学习模型的模型关联特征信息,以及所述目标机器学习模型在所述目标业务下的数据集路径信息、训练轮次和模型预期参数;

将所述模型关联特征信息、所述数据集路径信息和所述训练轮次配置到预设的 workflows 信息中,生成所述目标业务对应的目标 workflow 运行实例,所述目标 workflow 运行实例包括模型训练实例和模型测试实例;

通过运行所述模型训练实例训练所述目标机器学习模型,得到目标训练模型;

通过运行所述模型测试实例测试所述目标训练模型,得到测试结果;

在所述测试结果与所述模型预期参数相匹配的情况下,将所述目标训练模型作为业务模型,所述业务模型用于执行所述目标业务。

2. 根据权利要求1所述的模型处理方法,其特征在于,所述数据集路径信息包括训练数据集路径信息,所述目标机器学习模型的数目有多个,所述通过运行所述模型训练实例训练所述目标机器学习模型,得到目标训练模型包括:

根据所述训练数据集路径信息,获取训练数据集;

从所述训练数据集中获取当前训练轮次对应的训练数据集;

基于当前训练轮次对应的训练数据集,分别对上一训练轮次对应的目标训练后模型进行训练,得到当前训练轮次对应的训练后模型;其中,在当前训练轮次为第一训练轮次时,所述上一训练轮次对应的目标训练后模型为多个目标机器学习模型;

对当前训练轮次对应的训练后模型,按照每个训练后模型对应的训练效果数据由大至小依次排序,得到当前训练轮次对应的训练后模型序列,所述训练效果数据表征所述训练后模型与所述目标业务的匹配程度;

从当前训练轮次对应的训练后模型序列中,由前往后确定当前训练轮次对应的目标训练后模型;

若当前训练轮次对应的目标训练后模型不满足预设训练条件,则返回执行所述从所述训练数据集中获取当前训练轮次对应的训练数据集及后续步骤,直至当前训练轮次对应的目标训练后模型满足预设训练条件时,从当前训练轮次对应的目标训练后模型中确定目标训练模型。

3. 根据权利要求2所述的模型处理方法,其特征在于,所述从所述训练数据集中获取当前训练轮次对应的训练数据集包括:

基于预设采样条件,确定当前训练轮次对应的采样数量,其中,第一个训练轮次对应的采样数量低于其他训练轮次的采样数量;

基于所述当前训练轮次对应的采样数量,从所述训练数据集中获取所述采样数量个训练数据,得到当前训练轮次对应的当前训练数据集。

4. 根据权利要求2所述的模型处理方法,其特征在于,所述从当前训练轮次对应的训练后模型序列中,由前往后确定当前训练轮次对应的目标训练后模型包括:

基于预设筛选条件,确定当前训练轮次对应的筛选数量,其中,第一个训练轮次对应的筛选数量高于其他训练轮次的筛选数量;

根据所述当前训练轮次对应的筛选数量,从所述训练后模型序列中由前到后确定多个

训练后模型。

5. 根据权利要求1所述的模型处理方法,其特征在于,所述接收业务模型的模型处理配置指令之前,所述方法还包括:

接收初始数据集;

将所述初始数据集配置到预设的数据解析流程信息中,生成所述初始数据集对应的数据集解析实例;

通过运行所述数据集解析实例解析所述初始数据集,得到数据解析结果;

将所述数据解析结果发送到云端进行存储。

6. 根据权利要求1所述的模型处理方法,其特征在于,所述方法还包括:

接收机器学习模型;

生成所述机器学习模型的初始应用状态信息,所述初始应用状态信息表征所述机器学习模型不可应用;

获取所述机器学习模型的训练数据路径信息;

将所述训练数据路径信息配置到预设的工作流实例中,生成模型验证实例;

通过运行所述模型验证实例训练所述机器学习模型,得到所述机器学习模型对应的验证数据;

在所述验证数据满足预设应用条件时,将所述机器学习模型的初始应用状态信息更新为可应用。

7. 根据权利要求6所述的模型处理方法,其特征在于,所述接收业务模型的模型处理配置指令包括:

显示模型处理配置界面;

响应于所述模型处理配置界面中的模型配置确认指令,从所述初始应用状态信息更新为可应用的机器学习模型中获取所述目标业务对应的目标机器学习模型的模型关联特征信息,并获取所述目标机器学习模型在所述目标业务下的数据集路径信息、训练轮次和模型预期参数;

基于所述模型关联特征信息、所述数据集路径信息、所述训练轮次和所述模型预期参数,生成所述模型处理配置指令。

8. 一种模型处理装置,其特征在于,所述装置包括:配置信息获取模块、 workflow 运行实例生成模块、模型训练模块、模型测试模块和业务模型确定模块;

所述配置信息获取模块用于接收目标业务的模型处理配置指令,所述模型处理配置指令包括所述目标业务对应的目标机器学习模型的模型关联特征信息,以及所述目标机器学习模型在所述目标业务下的数据集路径信息、训练轮次和模型预期参数;

所述 workflow 运行实例生成模块用于将所述模型关联特征信息、所述数据集路径信息和所述训练轮次配置到预设的工作流流程信息中,生成所述目标业务对应的目标 workflow 运行实例,所述目标 workflow 运行实例包括模型训练实例和模型测试实例;

所述模型训练模块用于通过运行所述模型训练实例训练所述目标机器学习模型,得到目标训练模型;

所述模型测试模块用于通过运行所述模型测试实例测试所述目标训练模型,得到测试结果;

所述业务模型确定模块用于在所述测试结果与所述模型预期参数相匹配的情况下,将所述目标训练模型作为业务模型,所述业务模型用于执行所述目标业务。

9. 一种电子设备,其特征在于,所述电子设备包括处理器和存储器,所述存储器中存储有至少一条指令或至少一段程序,所述至少一条指令或所述至少一段程序由所述处理器加载并执行以实现如权利要求1-7任一项所述的一种模型处理方法。

10. 一种计算机可读存储介质,其特征在于,所述存储介质包括处理器和存储器,所述存储器中存储有至少一条指令或至少一段程序,所述至少一条指令或所述至少一段程序由所述处理器加载并执行以实现如权利要求1-7任一项所述的一种模型处理方法。

一种模型处理方法、装置、电子设备和存储介质

技术领域

[0001] 本申请涉及机器学习技术领域,尤其涉及一种模型处理方法、装置、电子设备和存储介质。

背景技术

[0002] 机器学习平台作为模型处理工具,可以快速定位到合适的机器学习算法,为机器学习算法分配相应的资源,即可基于该机器学习算法进行模型训练。

[0003] 现有技术中,机器学习平台通常是面向算法工程师的,导致机器学习平台的使用门槛较高,非算法工程师难以进行模型训练,提高了模型处理的操作难度,从而降低了模型处理的效率。

发明内容

[0004] 本申请提供了一种模型处理方法、装置、电子设备和存储介质,得到提高了模型处理的效率,简化了模型处理操作的技术效果。

[0005] 一方面,本申请提供了一种模型处理方法,所述方法包括:

[0006] 接收目标业务的模型处理配置指令,所述模型处理配置指令包括所述目标业务对应的目标机器学习模型的模型关联特征信息,以及所述目标机器学习模型在所述目标业务下的数据集路径信息、训练轮次和模型预期参数;

[0007] 将所述模型关联特征信息、所述数据集路径信息和所述训练轮次配置到预设的工作流流程信息中,生成所述目标业务对应的目标 workflow 运行实例,所述目标 workflow 运行实例包括模型训练实例和模型测试实例;

[0008] 通过运行所述模型训练实例训练所述目标机器学习模型,得到目标训练模型;

[0009] 通过运行所述模型测试实例测试所述目标训练模型,得到测试结果;

[0010] 在所述测试结果与所述模型预期参数相匹配的情况下,将所述目标训练模型作为业务模型,所述业务模型用于执行所述目标业务。

[0011] 另一方面提供了一种模型处理装置,

[0012] 所述装置包括:配置信息获取模块、workflow 运行实例生成模块、模型训练模块、模型测试模块和业务模型确定模块;

[0013] 所述配置信息获取模块用于接收目标业务的模型处理配置指令,所述模型处理配置指令包括所述目标业务对应的目标机器学习模型的模型关联特征信息,以及所述目标机器学习模型在所述目标业务下的数据集路径信息、训练轮次和模型预期参数;

[0014] 所述 workflow 运行实例生成模块用于将所述模型关联特征信息、所述数据集路径信息和所述训练轮次配置到预设的工作流流程信息中,生成所述目标业务对应的目标 workflow 运行实例,所述目标 workflow 运行实例包括模型训练实例和模型测试实例;

[0015] 所述模型训练模块用于通过运行所述模型训练实例训练所述目标机器学习模型,得到目标训练模型;

[0016] 所述模型测试模块用于通过运行所述模型测试实例测试所述目标训练模型,得到测试结果;

[0017] 所述业务模型确定模块用于在所述测试结果与所述模型预期参数相匹配的情况下,将所述目标训练模型作为业务模型,所述业务模型用于执行所述目标业务。

[0018] 另一方面提供了一种电子设备,所述电子设备包括处理器和存储器,所述存储器中存储有至少一条指令或至少一段程序,所述至少一条指令或所述至少一段程序由所述处理器加载并执行以实现如上述所述的一种模型处理方法。

[0019] 另一方面提供了一种计算机可读存储介质,所述存储介质包括处理器和存储器,所述存储器中存储有至少一条指令或至少一段程序,所述至少一条指令或所述至少一段程序由所述处理器加载并执行以实现如上述所述的一种模型处理方法。

[0020] 另一方面还提供了一种计算机程序产品,该计算机程序产品包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述所述的一种模型处理方法。

[0021] 本申请提供了一种模型处理方法、装置、电子设备和存储介质,所述方法包括:接收业务模型的模型处理配置指令,将模型处理配置指令中的内容配置到预设的工作流流程信息中,生成业务模型对应的目标工作流运行实例,目标工作流运行实例包括模型训练实例和模型测试实例。运行模型训练实例,以训练目标机器学习模型,得到目标训练模型。运行模型测试实例,以测试目标训练模型,得到测试结果。在测试结果与模型预期参数相匹配的情况下,将目标训练模型作为业务模型。该方法可以自动完成模型训练、测试等的流程,提高了模型处理的效率,简化了模型处理的操作,从而节省了人工成本和运维成本。

附图说明

[0022] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅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。

[0023] 图1为本申请实施例提供的一种模型处理方法的应用场景示意图;

[0024] 图2为本申请实施例提供的区块链进行分布式存储的示意图;

[0025] 图3为本申请实施例提供的区块链结构的示意图;

[0026] 图4为本申请实施例提供的一种模型处理方法的流程图;

[0027] 图5为本申请实施例提供的一种模型处理方法中接收业务模型的模型处理配置指令的流程图;

[0028] 图6为本申请实施例提供的一种模型处理方法目标业务的模型配置界面的示意图;

[0029] 图7为本申请实施例提供的一种模型处理方法目标业务的模型处理配置界面的示意图;

[0030] 图8为本申请实施例提供的一种模型处理方法中目标工作流运行实例的管理流程的示意图;

- [0031] 图9为本申请实施例提供一种模型处理方法中 workflow 运维页面的示意图；
- [0032] 图10为本申请实施例提供一种模型处理方法对多个目标机器学习模型进行训练的方法流程图；
- [0033] 图11为本申请实施例提供一种模型处理方法中对初始数据集进行解析的方法流程图；
- [0034] 图12为本申请实施例提供一种模型处理方法中初始数据集的上传界面的示意图；
- [0035] 图13为本申请实施例提供一种模型处理方法中在资源管理平台对初始数据集进行数据解析的流程示意图；
- [0036] 图14为本申请实施例提供一种模型处理方法中展示初始数据集解析结果的界面示意图；
- [0037] 图15为本申请实施例提供一种模型处理方法中对机器学习模型进行验证的方法流程图；
- [0038] 图16为本申请实施例提供一种模型处理方法中机器学习模型的状态标注界面的示意图；
- [0039] 图17为本申请实施例提供一种模型处理方法中通用模块和业务模块的结构示意图；
- [0040] 图18为本申请实施例提供一种模型处理装置的结构示意图；
- [0041] 图19为本申请实施例提供一种用于实现本申请实施例所提供的方法的设备的硬件结构示意图。

具体实施方式

[0042] 为使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请作进一步地详细描述。显然，所描述的实施例仅仅是本申请的一部分实施例，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的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。

[0043] 在本申请的描述中，需要理解的是，术语“第一”、“第二”仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个该特征。而且，术语“第一”、“第二”等适用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施。

[0044] 请参见图1，其显示了本申请实施例提供一种模型处理方法的应用场景示意图，该应用场景包括客户端110、云端120和业务服务器130，客户端110响应于用户的输入的信息，接收业务模型的模型处理配置指令，并根据模型处理配置指令和预设的工作流流程信息，生成目标业务对应的目标 workflow 运行实例，该目标 workflow 运行实例包括模型训练实例和模型测试实例。客户端110对目标机器学习模型进行处理，客户端110运行模型训练实例，从云端120处获取训练数据集，以训练目标机器学习模型，得到目标训练模型。客户端110运行模型测试实例，从云端120处获取测试数据集，以测试目标训练模型，得到测试结果。客户

端110在测试结果与模型预期参数匹配的情况下,将目标训练模型作为业务模型发送到业务服务器130,执行目标业务。

[0045] 在本申请实施例中,客户端110包括智能手机、台式电脑、平板电脑、笔记本电脑、数字助理、智能可穿戴设备等类型的实体设备,也可以包括运行于实体设备中的软体,例如应用程序等。本申请实施例中实体设备上运行的操作系统可以包括但不限于安卓系统、IOS系统、linux、Unix、windows等。客户端110可以运行模型训练实例和模型测试实例,执行模型训练步骤和模型测试步骤。客户端110包括UI (User Interface,用户界面)层,客户端110通过UI层对外提供业务模型的显示,以及模型处理配置界面,另外,基于API (Application Programming Interface,应用程序接口)将向云端120请求数据集。

[0046] 客户端中应用机器学习平台,可以自动执行机器学习的过程,对目标机器学习模型进行处理。机器学习平台通常只提供算法训练所需计算资源,算法研究人员需手动构建自己的算法训练环境,并且评估算法和上线等操作也需要手动配置。因此机器学习平台通常要求用户对算法模型有深刻的理解,并且需要能熟练的用编程语言实现,导致非算法工程师难以在机器学习平台上对模型进行训练。

[0047] 机器学习(Machine Learning,ML)是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域。机器学习和深度学习通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、式教学习等技术。

[0048] 在本申请实施例中,云端120可以包括一个独立运行的服务器,或者分布式服务器,或者由多个服务器组成的服务器集群。云端120可以为区块链结构的分布式存储系统。如图2所示,区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。以分布式系统为区块链系统为例,参见图2,图2是本发明实施例提供的分布式系统200应用于区块链系统的一个可选的结构示意图,由多个节点(接入网络中的任意形式的计算设备,如服务器、用户终端)和客户端形成,节点之间形成组成的点对点(P2P,Peer To Peer)网络,P2P协议是一个运行在传输控制协议(TCP,Transmission Control Protocol)协议之上的应用层协议。在分布式系统中,任何机器如服务器、终端都可以加入而成为节点,节点包括硬件层、中间层、操作系统层和应用层。

[0049] 参见图2示出的区块链系统中各节点的功能,涉及的功能包括:

[0050] 1) 路由,节点具有的基本功能,用于支持节点之间的通信。

[0051] 节点除具有路由功能外,还可以具有以下功能:

[0052] 2) 应用,用于部署在区块链中,根据实际业务需求而实现特定业务,记录实现功能相关的数据形成记录数据,在记录数据中携带数字签名以表示任务数据的来源,将记录数据发送到区块链系统中的其他节点,供其他节点在验证记录数据来源以及完整性成功时,将记录数据添加到临时区块中。

[0053] 例如,应用实现的业务包括:

[0054] 2.1) 钱包,用于提供进行电子货币的交易的功能,包括发起交易(即,将当前交易的交易记录发送给区块链系统中的其他节点,其他节点验证成功后,作为承认交易有效的

响应,将交易的记录数据存入区块链的临时区块中;当然,钱包还支持查询电子货币地址中剩余的电子货币;

[0055] 2.2) 共享账本,用于提供账目数据的存储、查询和修改等操作的功能,将对账目数据的操作的记录数据发送到区块链系统中的其他节点,其他节点验证有效后,作为承认账目数据有效的响应,将记录数据存入临时区块中,还可以向发起操作的节点发送确认。

[0056] 2.3) 智能合约,计算机化的协议,可以执行某个合约的条款,通过部署在共享账本上的用于在满足一定条件时而执行的代码实现,根据实际的业务需求代码用于完成自动化的交易,例如查询买家所购买商品的物流状态,在买家签收货物后将买家的电子货币转移到商户的地址;当然,智能合约不仅限于执行用于交易的合约,还可以执行对接收的信息进行处理的合约。

[0057] 3) 区块链,包括一系列按照产生的先后时间顺序相互接续的区块(Block),新区块一旦加入到区块链中就不会再被移除,区块中记录了区块链系统中节点提交的记录数据。

[0058] 参见图3,图3是本发明实施例提供的区块结构(Block Structure)一个可选的示意图,每个区块中包括本区块存储交易记录的哈希值(本区块的哈希值)、以及前一区块的哈希值,各区块通过哈希值连接形成区块链。另外,区块中还可以包括有区块生成时的时间戳等信息。区块链(Blockchain),本质上是一个去中心化的数据库,是一串使用密码学方法相关联产生的数据块,每一个数据块中包含了相关的信息,用于验证其信息的有效性(防伪)和生成下一个区块。

[0059] 请参见图4,其显示了一种模型处理方法,可应用于客户端侧,该方法包括:

[0060] S410.接收目标业务的模型处理配置指令,模型处理配置指令包括目标业务对应的目标机器学习模型的模型关联特征信息,以及目标机器学习模型在目标业务下的数据集路径信息、训练轮次和模型预期参数;

[0061] 进一步地,客户端可以接收目标业务的模型处理配置指令,模型处理配置指令表征用户配置的对模型进行处理的指令信息,基于模型处理配置指令,客户端可以自动执行模型训练、模型测试等模型处理的方法。

[0062] 进一步地,模型关联特征信息表征目标机器学习模型的属性信息。数据集路径信息表征存储数据集的地址信息,该地址信息可以位于云端,训练轮次表征在目标业务下对目标机器学习模型进行训练的次数,模型预期参数表征对应应用于目标业务的模型的预期应用条件,在经过训练并测试后的目标机器学习模型达到模型预期参数时,可以将该模型应用于对应的目标业务。

[0063] 在一个具体的实施例中,请参见图6,如图6所示为目标机器学习模型的模型配置界面,在模型配置界面可以输入模型配置指令,模型配置指令可以包括已选模板、训练方式、业务类型、应用组、模型名称、模型描述、维护人等信息,检测到用户输入的模型配置指令,可以对目标机器学习模型进行配置。例如,选择视频分类的模型作为已选模板,设置训练方式为分类,业务类型为视频业务,应用组和维护人可以填写相关工作人员的信息。模型描述可以为对目标机器学习模型的描述信息。在完成模型配置后,可以配置对目标机器学习模型的模型处理配置指令,基于模型处理配置指令,对目标机器学习模型进行处理。

[0064] 进一步地,请参见图5,接收目标业务的模型处理配置指令包括:

[0065] S510.显示模型处理配置界面;

[0066] S520. 响应于模型处理配置界面中的模型配置确认指令, 从初始应用状态信息更新为可应用的机器学习模型中获取目标业务对应的目标机器学习模型的模型关联特征信息, 并获取目标机器学习模型在目标业务下的数据集路径信息、训练轮次和模型预期参数;

[0067] S530. 基于模型关联特征信息、数据集路径信息、训练轮次和模型预期参数, 生成模型处理配置指令。

[0068] 在一个具体的实施例中, 请参见图7, 如图7所示为模型处理配置界面。在模型处理配置界面可以进行模型关联特征信息、数据集路径信息、模型预期参数、服务器标识信息等的配置。其中, 模型关联特征信息可以包括语料特征、任务类型和目标机器学习模型, 目标机器学习模型为与语料特征相关的模型信息, 例如语料特征为图像特征, 则相关的模型信息为用于处理图像信息的模型。数据集路径信息可以包括训练数据路径信息和测试数据路径信息, 训练数据路径信息为获取训练数据集的路径信息, 测试数据路径信息为获取测试数据集的路径信息。模型预期参数可以包括预期的精确率、预期的召回率和预期的综合函数F1, F1为精确率和召回率的调和均值。

[0069] 响应于模型配置确认指令, 获取模型配置区域输入的信息、数据集配置区域输入的信息、训练轮次配置区域输入的信息和模型预期参数配置区域输入的信息, 基于模型配置区域输入的信息、数据集配置区域输入的信息、训练轮次配置区域输入的信息和模型预期参数配置区域输入的信息, 客户端生成模型处理配置指令。客户端可以将模型处理配置指令与预设的工作流程信息组合, 得到目标 workflow 运行实例。在参数配置完成后, 客户端运行目标 workflow 运行实例时, 客户端可以在 workflow 运维页面展示模型训练、测试或上线的实时状态和参数配置详情, 用户可以随时查看该实时状态和参数配置详情。

[0070] 用户在模型处理配置界面进行配置后, 客户端就可以自动执行模型训练和测试的步骤, 简化了用户操作。

[0071] S420. 将模型关联特征信息、数据集路径信息和训练轮次配置到预设的工作流程信息中, 生成目标业务对应的目标 workflow 运行实例, 目标 workflow 运行实例包括模型训练实例和模型测试实例;

[0072] 进一步地, 如图8所示, 为目标 workflow 运行实例的管理流程的示意图。在客户端中预设有训练→测试→上线的工作流程信息, 将模型关联特征信息、数据集路径信息和训练轮次增加到该 workflow 流程信息, 生成与该业务模型对应的目标 workflow 运行实例。客户端通过运行目标 workflow 运行实例, 自动执行模型训练的步骤、模型测试的步骤和模型上线的步骤。

[0073] 其中, 目标 workflow 运行实例中包括模型训练实例和模型测试实例, 客户端运行模型训练实例, 可以执行训练步骤, 对目标机器学习模型进行训练, 得到目标训练模型。客户端运行模型测试实例, 可以执行测试步骤, 对目标训练模型进行测试, 得到测试结果。模型上线的步骤为将测试结果与模型预期参数进行比对, 当测试结果满足模型预期参数时, 将目标训练模型作为业务模型, 自动执行上线步骤, 将业务模型发送到上线服务器进行上线处理。

[0074] workflow 运维页面如图9所示。 workflow 运维页面中还包括训练结果对应的页面、测试结果对应的页面和上线结果对应的页面。在训练过程结束后, 响应于用户输入的指令, 客户端可以跳转到相应的页面并显示训练结果的详细信息。在测试过程结束后, 响应于用户输

入的指令,客户端可以跳转到相应的页面并显示测试结果的详细信息。在上线过程结束后,响应于用户输入的指令,客户端可以跳转到相应的页面并显示上线结果的详细信息。

[0075] 在一个具体的实施例中,客户端可以利用资源分配平台分配的虚拟机资源运行目标 workflow 运行实例,在虚拟机上运行目标 workflow 运行实例时,可以定时将运行中的目标 workflow 运行实例的状态和数据同步到后端服务器。

[0076] 利用 workflow 的方式,可以为算法模型提供一站式自动化管理,从而降低模型训练和评估的门槛,提高模型处理的效率。

[0077] S430.通过运行模型训练实例训练目标机器学习模型,得到目标训练模型;

[0078] 进一步地,通过运行模型训练实例,可以对多个目标机器学习模型进行多轮训练,从训练后模型中筛选目标训练后模型,基于目标训练后模型确定与目标业务的匹配的目标训练模型。

[0079] 在一个具体的实施例中,在客户端上运行模型训练实例,或者将模型训练实例发送到资源分配平台,资源分配平台接收资源分配平台为模型训练实例分配的虚拟机资源,基于虚拟机资源,运行模型训练实例,对目标机器学习模型进行训练,得到目标训练模型。

[0080] 进一步地,请参见图10,模型训练实例包括训练数据路径信息和训练参数,目标机器学习模型的数目有多个,通过运行模型训练实例,训练目标机器学习模型,得到目标训练模型包括:

[0081] S1010.根据训练数据路径信息,获取训练数据集;

[0082] S1020.从训练数据集中获取当前训练轮次对应的训练数据集;

[0083] S1030.基于当前训练轮次对应的训练数据集,分别对上一训练轮次对应的目标训练后模型进行训练,得到当前训练轮次对应的训练后模型;其中,在当前训练轮次为第一训练轮次时,上一训练轮次对应的目标训练后模型为多个目标机器学习模型;

[0084] S1040.对当前训练轮次对应的训练后模型,按照每个训练后模型对应的训练效果数据由大至小依次排序,得到当前训练轮次对应的训练后模型序列;

[0085] S1050.从当前训练轮次对应的训练后模型序列中,由前往后确定当前训练轮次对应的目标训练后模型;

[0086] S1060.若当前训练轮次对应的目标训练后模型不满足预设训练条件,则返回执行从训练数据集中获取当前训练轮次对应的训练数据集及后续步骤,直至当前训练轮次对应的目标训练后模型满足预设训练条件时,从当前训练轮次对应的目标训练后模型中确定目标训练模型。

[0087] 在一个具体的实施例中,客户端根据训练数据路径信息,从云端中获取训练数据集。在只有一个目标机器学习模型需要训练时,客户端根据训练数据集和训练参数,对该目标机器学习模型进行自动训练,得到目标训练模型。在用户选择了多个目标机器学习模型的情况下,客户端需要获取每个目标机器学习模型对应的训练后模型,并从这些训练后模型中选择目标训练模型。

[0088] 客户端在对多个目标机器学习模型进行训练时,可以对训练数据集进行采样,得到当前训练数据集。客户端将目标机器学习模型作为当前待训练模型。客户端根据当前训练数据集和训练参数,对多个目标机器学习模型进行训练,得到多个训练后模型。客户端将训练后模型根据训练效果数据由大到小排序,训练效果数据表征训练后模型与目标业务的

匹配程度,也就是说训练效果数据越大的训练后模型越适用于执行目标业务。在获取训练效果数据时,客户端可以计算每个训练后模型的损失数据,通过训练后模型的损失数据表征模型的训练效果,损失数据越小则代表模型的训练效果越好。根据训练后模型的输出结果和训练数据集中的标注信息间的差值,可以计算得到训练后模型的损失数据。在一个具体的实施例中,损失数据可以是交叉熵损失数据。

[0089] 按照损失数据的大小,客户端从小到大对多个训练后模型进行排序,得到训练后模型序列。客户端对训练后模型序列进行筛选,得到目标训练后模型,目标训练后模型可以包括多个训练后模型。若当前训练轮次对应的目标训练后模型不满足预设训练条件,则客户端返回执行从训练数据集中获取当前训练轮次对应的训练数据集及后续步骤,直至当前训练轮次对应的目标训练后模型满足预设训练条件时,从当前训练轮次对应的目标训练后模型中确定目标训练模型。

[0090] 其中,在一些实施例中,预设训练条件可以为目标训练后模型为最后一个训练轮次输出的目标训练后模型或目标训练后模型中只存在一个训练后模型。在目标训练后模型中只存在一个训练后模型时,当前训练轮次对应的目标训练后模型满足预设训练条件,此时训练过程结束。客户端可以确定目标训练后模型中存在的这个训练后模型为目标训练模型。在另一些实施例中,预设训练条件也可以为模型处理配置指令所指示的训练轮次耗尽。在目标训练后模型为最后一个训练轮次输出的目标训练后模型的情况下,在客户端完成了最后一个训练轮次,输出了最后一个训练轮次得到的目标训练后模型时,若目标训练后模型中仍然存在多个训练后模型,但由于训练轮次已经耗尽,故当前训练轮次对应的目标训练后模型满足预设训练条件,此时训练过程结束。客户端可以从多个训练后模型中选择目标训练模型,例如可以选择训练效果数据最大的训练后模型作为目标训练模型,如果同时存在两个或两个以上的训练后模型的训练效果数据均为最大,则可以将这些训练后模型均作为目标训练模型。

[0091] 进一步地,从训练数据集中获取当前训练轮次对应的训练数据集包括:

[0092] 基于预设采样条件,确定当前训练轮次对应的采样数量,其中,第一个训练轮次对应的采样数量低于其他训练轮次的采样数量;

[0093] 基于当前训练轮次对应的采样数量,从训练数据集中获取采样数量个训练数据,得到当前训练轮次对应的当前训练数据集。

[0094] 在一些具体的实施例中,在每一个训练轮次中,根据预设的采样条件,可以确定当前训练轮次对应的采样数量,根据当前训练轮次对应的采样数量,对训练数据集进行采样,得到当前训练轮次对应的当前训练数据集。预设的采样条件为第一个训练轮次对应的采样数量低于其他训练轮次的采样数量。

[0095] 在一种实施方式中,预设的采样条件可以为除第一个训练轮次以外其他训练轮次对应的当前训练数据集中训练数据的数量相同,其他训练轮次对应的当前训练数据集中训练数据的数量均大于第一个训练轮次对应的当前训练数据集中训练数据的数量。

[0096] 在另一种实施方式中,预设的采样条件也可以为每一个训练轮次对应的当前训练数据集中训练数据的数量大于上一个训练轮次对应的当前训练数据集中训练数据的数量,也就是当前训练数据集中训练数据的数量随训练轮次的增加而增加。

[0097] 在一个具体的实施例中,用户选择了N个目标机器学习模型,并选择了数据集M作

为训练数据集,预设进行x轮训练,采样参数设置为y,采样参数为第一个训练轮次中训练数据集的采样数量对应的参数。

[0098] 在当前训练轮次为第一个训练轮次时,客户端可以根据采样参数y,对数据集M进行采样,得到当前训练数据集m。

[0099] 在当前训练轮次为非第一个训练轮次时,客户端可以根据下述公式(1)更新当前训练轮次对应的采样数量,该公式为:

$$[0100] \quad y_1 = y^{\frac{x-b}{x}} \dots\dots\dots(1)$$

[0101] 其中,b表示当前已经训练的次数。b的最大值为x-1。

[0102] 例如,用户选择了N个目标机器学习模型,并选择了数据集M作为训练数据集,预设进行2轮训练,采样参数设置为0.1。在第一个训练轮次时,从数据集M中采样10%的数据,对N个目标机器学习模型进行训练。

[0103] 利用公式(2)计算第二个训练轮次时对数据集进行采样的采样参数可得:

$$[0104] \quad y_1 = y^{\frac{x-b}{x}} = 0.1^{\frac{1}{2}} = 31.62\% \dots\dots\dots(2)$$

[0105] 在第二个训练轮次时,从数据集M中采样31.62%的数据,对第一个训练轮次得到的目标训练后模型进行训练。在第二个训练轮次结束后,从第二个训练轮次后的模型中选择损失数据最小的模型作为目标训练模型。

[0106] 在每一个训练轮次时更新训练数据集的采样数量,逐次增加当前训练数据集中的数据,可以在维持模型训练的有效性的同时缩短训练时间,从而提高模型训练的效率。

[0107] 进一步地,从当前训练轮次对应的训练后模型序列中,由前往后确定当前训练轮次对应的目标训练后模型包括:

[0108] 基于预设筛选条件,确定当前训练轮次对应的筛选数量,其中,第一个训练轮次对应的筛选数量高于其他训练轮次的筛选数量;

[0109] 根据当前训练轮次对应的筛选数量,从训练后模型序列中由前到后确定多个训练后模型。

[0110] 在一些实施例中,根据预设的筛选条件,在每一个训练轮次结束对训练后模型序列进行筛选时,可以确定当前训练轮次对应的筛选条件,根据当前训练轮次对应的筛选条件,对训练后模型序列进行筛选,得到当前训练轮次对应的多个训练后模型。预设的筛选条件为第一个训练轮次对应的筛选数量高于其他训练轮次的筛选数量。

[0111] 在一种实施方式中,预设的筛选条件可以为除第一个训练轮次以外其他训练轮次后筛选得到的多个训练后模型的数量相同,其他训练轮次后筛选得到的多个训练后模型均小于第一个训练轮次后筛选得到的多个训练后模型。

[0112] 在另一种实施方式中,预设的筛选条件也可以为每一个训练轮次后筛选得到的多个训练后模型的数量大于每一个训练轮次后筛选得到的多个训练后模型的数量,也就是多个训练后模型的数量随训练轮次的增加而减少。

[0113] 在一个具体的实施例中,客户端根据当前训练数据集m,分别对N个目标机器学习模型进行训练,可以得到N个训练后模型。客户端根据这N个训练后模型的损失数据的大小,从小到大对训练后模型进行排序,得到训练后模型序列。客户端可以根据下述公式(3)计算得到目标训练后模型的筛选区间z,该筛选区间z为筛选数量对应的区间,从训练后模型序

列中确定目标训练后模型进行下一次训练。该公式为：

[0114] $z = N \times y^{\frac{a}{x}}$(3)

[0115] 其中,a表示当前已经训练的训练轮次数。a的最大值为x-1,N表示当前进行训练的目标训练后模型。

[0116] 例如,用户选择了N个目标机器学习模型,需要进行3次模型训练,客户端在第一个训练轮次中,根据第一个训练轮次对应的当前采样数据集,对N个目标机器学习模型进行训练,得到N个训练后模型。利用公式计算第一个训练轮次后的目标训练后模型的筛选区间z可得:

[0117] $z = N \times y^{\frac{a}{x}} = 0.1^{\frac{1}{3}}N = 46.77\%N$(4)

[0118] 则根据损失数据的大小从小到大排列训练后模型,留下前46.77%N的训练后模型作为第一个训练轮次后的目标训练后模型。

[0119] 客户端在第二个训练轮次中,根据第二个训练轮次对应的当前采样数据集,对46.77%N个目标训练后模型进行训练,得到46.77%N个训练后模型。利用公式计算第二个训练轮次后的目标训练后模型的筛选区间z可得:

[0120] $z = 46.77\%N \times y^{\frac{a}{x}} = 0.1^{\frac{2}{3}}N \times 46.77\% = 10\%N$(5)

[0121] 则根据损失数据的大小从小到大排列训练后模型,留下前10%N的训练后模型作为第二个训练轮次后的目标训练后模型。

[0122] 客户端在第三个训练轮次中,根据第三个训练轮次对应的当前采样数据集,对10%N个目标机器学习模型进行训练,得到10%N个训练后模型,此时已经耗尽训练次数。利用公式计算第三个训练轮次后的目标训练后模型的筛选区间z可得:

[0123] $z = 10\%N \times y^{\frac{a}{x}} = 0.1^{\frac{3}{3}}N \times 10\% = 1\%N$(6)

[0124] 客户端可以从1%N个目标训练后模型中确定目标训练模型。

[0125] 在每一个训练轮次后时更新训练后模型序列的筛选条件,逐次减少目标训练后模型中的数据,可以缩短模型筛选时间,从而提高模型筛选的效率。

[0126] 在数据集中采样部分数据,对多个目标机器学习模型进行训练,从训练后模型中选择损失数据最小的模型作为目标训练模型,从而以较小的数据集选择出较优的模型,提高了模型选优的效率。

[0127] S440.通过运行模型测试实例测试目标训练模型,得到测试结果;

[0128] 进一步地,通过运行模型测试实例,可以对目标训练模型进行测试,将目标训练模型输出的测试数据与模型预期参数进行比较,得到测试结果。测试结果包括目标训练模型的测试数据与模型预期参数不匹配和目标训练模型的测试数据与模型预期参数相匹配两种情况。

[0129] 在一个具体的实施例中,在客户端上运行模型训练实例,或者将模型训练实例发送到资源分配平台,资源分配平台接收资源分配平台为模型训练实例分配的虚拟机资源,基于虚拟机资源,运行模型训练实例,对目标机器学习模型进行训练,得到目标训练模型。

[0130] 进一步地,模型测试实例包括测试数据路径信息,运行模型测试实例,以测试目标训练模型,得到测试结果包括:

[0131] 根据测试数据路径信息,获取测试数据集;

[0132] 根据测试数据集,对目标训练模型进行模型测试,得到目标训练模型的测试结果。

[0133] 在一个具体的实施例中,客户端根据测试数据路径信息,从云端中获取测试数据集。客户端根据测试数据集,对目标训练模型进行模型测试,得到目标训练模型的测试结果。测试结果可以为输入测试数据集时目标训练模型的测试精确率、测试召回率和测试综合函数F1。

[0134] 模型预期参数可以包括预期的精确率、预期的召回率和预期的综合函数F1,F1为精确率和召回率的调和均值。在测试精确率与预期的精确率匹配,测试召回率与预期的召回率匹配,且测试综合函数F1与预期的综合函数F1匹配的情况下,将目标训练模型作为业务模型,执行业务模型的上线步骤。在测试结果与模型预期参数不匹配的情况下,可以停止上线步骤,转由人工处理。客户端可以在接收到更新后的模型处理配置指令的情况下,生成更新后的模型处理配置指令对应的工作流运行实例,重新执行训练步骤。

[0135] 在完成训练步骤后,客户端自动执行模型测试步骤,提高了模型处理的效率,且在测试步骤未通过时可以通过重新设置模型处理配置指令的方式,重新执行训练步骤,提高了模型重训练的效率,简化了模型重训练中用户的操作。

[0136] S450.在测试结果与模型预期参数相匹配的情况下,将目标训练模型作为业务模型,业务模型用于执行目标业务。进一步地,在得到业务模型后,可以对业务模型进行应用处理。业务模型的应用可以包括直接上线、推送给用户进行应用、推送到云端等。

[0137] 进一步地,模型处理配置指令包括服务器标识信息,对业务模型进行上线处理包括:

[0138] 将业务模型发送到服务器标识信息对应的上线服务器中,以使得上线服务器将业务模型配置到线上环境中。

[0139] 在一个具体的实施例中,在模型处理配置指令中包括服务器标识信息的情况下,在得到业务模型后,可以直接将业务模型发送到服务器标识信息对应的上线服务器中。上线服务器可以进行业务模型的上线处理,将业务模型配置到线上环境中,使得模型运行在对应的目标业务场景中,执行目标业务。

[0140] 客户端在目标训练模型通过测试后可以自动执行模型上线步骤,并简化了上线步骤,提高了模型处理的效率。

[0141] 进一步地,请参见图11,接收业务模型的模型处理配置指令之前,该方法还包括:

[0142] S1110.接收初始数据集;

[0143] S1120.将初始数据集配置到预设的数据解析流程信息中,生成初始数据集对应的数据集解析实例;

[0144] S1130.通过运行数据集解析实例解析初始数据集,得到数据解析结果;

[0145] S1140.将数据解析结果发送到云端进行存储。

[0146] 在一个具体的实施例中,请参见图12,如图12所示为初始数据集的上传界面。初始数据集的相关信息可以包括应用组、业务类型、数据来源、码率、数据集名称、数据集描述和维护人等信息。客户端接收初始数据集,客户端可以保存初始数据集的源信息,并将初始数据集传输到云端进行存储。客户端将初始数据集配置到预设的数据解析流程信息中,生成初始数据集对应的数据集解析实例。

[0147] 客户端运行数据集解析实例,对初始数据集进行数据解析。数据解析流程信息可以包括采样→采样数据预处理→数据同步→抽帧→采样数据处理。其中,执行数据同步步骤时可以将采样数据同步到文件系统中,例如Ceph文件系统。抽帧步骤抽取的信息可以根据初始数据集的类型进行选择。若输入的初始数据集为图像数据集,则在抽帧步骤时抽取图像数据,若输入的初始数据集为视频数据集,则在抽帧步骤时抽取视频数据,若输入的初始数据集为音频数据集,则在抽帧步骤时抽取音频数据。

[0148] 在一个具体的实施例中,如图13所示,图13为在资源管理平台对初始数据集进行数据解析的流程示意图。客户端可以将数据集解析实例发送到资源管理平台,资源管理平台从云端获取初始数据集,并运行数据集解析实例对初始数据集进行解析。客户端可以定时检查资源管理平台对初始数据集的解析进度,在资源管理平台解析完成时,资源管理平台将解析结果发送到云端,客户端从云端获取解析结果,对解析结果进行格式加工后,再次发送到云端进行存储。响应于用户输入的解析结果查看请求,客户端从云端中获取格式加工后的解析结果,并在前端界面上显示格式加工后的解析结果。

[0149] 在一个具体的实施例中,如图14所示,图14为展示数据集解析结果的页面示意图。客户端在检测到数据集解析结果查询指令的情况下,显示数据集解析结果展示页面,在数据集解析结果展示页面中展示每个数据集的数据集参数信息,数据集参数信息包括名称信息、业务类型信息、解析状态、更新时间和解析操作状态。解析状态信息可以包括解析成功和解析失败两种状态,解析操作状态可以显示为该数据集对应的数据集解析的工作流信息和数据集解析后得到的样本信息。

[0150] 在对模型处理前,客户端可以对初始数据集进行解析,从而提高模型处理时特征数据的有效性。进一步地,请参见图15,该方法还包括:

[0151] S1510.接收机器学习模型;

[0152] S1520.生成机器学习模型的初始应用状态信息,初始应用状态信息表征机器学习模型不可应用;

[0153] S1530.获取机器学习模型的训练数据路径信息;

[0154] S1540.将训练数据路径信息配置到预设的工作流实例中,生成模型验证实例;

[0155] S1550.通过运行模型验证实例,训练机器学习模型,得到机器学习模型对应的验证数据;

[0156] S1560.在验证数据满足预设应用条件时,将机器学习模型的初始应用状态信息更新为可应用。

[0157] 在一个具体的实施例中,请参见图16,如图16所示为机器学习模型的状态标注界面的示意图。客户端接收用户上传的机器学习模型。当一个机器学习模型为新上传的模型时,生成机器学习模型的初始应用状态信息,初始应用状态信息表征该机器学习模型不可应用。客户端获取该机器学习模型的训练数据路径信息,将训练数据路径信息配置到预设的工作流实例中,生成模型验证实例。模型验证实例中只包括模型训练步骤和模型评估步骤,不进行模型上线步骤。客户端运行模型验证实例,以对机器学习模型进行模型训练,得到机器学习模型的验证数据。在验证数据满足预设应用条件时,将机器学习模型的初始应用状态信息更新为可应用。

[0158] 验证数据可以为机器学习模型的损失数据,在验证数据为机器学习模型的损失数

据时,预设应用条件对应为预设的损失数据。将机器学习模型的损失数据与预设的损失数据进行比对。在损失数据小于或等于预设的损失数据的情况下,将机器学习模型的初始应用状态信息更新为可应用。在损失数据大于预设的损失数据的情况下,不更新机器学习模型的初始应用状态信息。

[0159] 验证数据还可以为机器学习模型的召回率和精确率,在验证数据为机器学习模型的召回率和精确率时,预设应用条件对应为预设的召回率和精确率。将机器学习模型的召回率和精确率与预设的召回率和精确率进行比对。在召回率和精确率大于或等于预设的召回率和精确率的情况下,将机器学习模型的初始应用状态信息更新为可应用。在召回率和精确率小于预设的召回率和精确率的情况下,不更新机器学习模型的初始应用状态信息。

[0160] 机器学习模型为不可应用的状态时,在用户进行模型处理配置时,不能选择不可应用状态的机器学习模型。

[0161] 在一个具体的实施例中,客户端可以设置模型市场的功能。在模型市场中,用户可以进行机器学习模型的上传和目标机器学习模型对应的模型业务的选择。

[0162] 在进行机器学习模型的上传时,客户端检测到用户输入的模型上传指令,接收用户上传的机器学习模型,并对应该机器学习模型,增加机器学习模型的状态信息,机器学习模型的状态信息可以包括上传状态信息、应用状态信息和测试报告信息,上传状态信息可以表征机器学习模型的上传状态,包括上传中状态和已上传状态,应用状态信息可以表征机器学习模型的可应用状态,初始应用状态信息为不可应用,在机器学习模型通过模型验证后,可以修改应用状态信息为可应用。测试报告信息在机器学习模型通过模型验证后生成,可以表征机器学习在模型验证中的表现情况。

[0163] 在选择目标机器学习模型对应的模型业务时,客户端检测到用户输入的模型业务查询指令时,跳转到模型市场界面的查询结果列表。在模型市场界面的查询结果列表中,可以展示目标机器学习模型对应的业务类型。响应于用户输入的模型选择指令,可以从查询结果列表中获取对应的目标机器学习模型。

[0164] 客户端可以获得用户上传的机器学习模型,从而丰富了机器学习模型的多样性。

[0165] 在一个具体的实施例中,请参见图17,如图17所示为模型管理的产品结构,包括网页服务模块、定时调度模块和数据访问模块这些通用模块,以及 workflow 配置并执行的业务模块。网页服务模块,可以响应前端请求。定时调度模块,可以完成预设的周期性的任务。数据访问模块,可以对接各种外部数据源,例如云端,在从云端获取训练数据集和测试数据集时,可以通过数据访问模块访问云端,获取对应的数据集。业务模块包括数据集模块、模型模块、训练模块、测试模块、workflow 配置模块和 workflow 实例生成模块等,可以执行模型处理配置指令,对模型处理的工作流进行配置,生成对应的工作流实例,对目标机器学习模型进行模型处理。

[0166] 本申请实施例提供的一种模型处理方法可以应用于图像分类、视频分类、物体检测等多个业务方向。在一个具体的实施例中,若用户需要图像分类模型,则用户在客户端的模型处理配置界面上选择训练数据集,选择任务类型为图像分类,语料特征为图像特征,选择基础模型,设置训练轮次,选择评估数据集,设置服务器标识信息和模型预期参数。根据用户输入的信息,客户端生成模型处理配置指令。客户端将模型处理配置指令中的内容配置到 workflow 流程信息中,生成 workflow 运行实例。客户端运行 workflow 运行实例,自动执行模型

训练,得到初始图像分类模型,客户端运行 workflow 运行实例,对初始图像分类模型进行模型测试,并在模型测试的结果满足模型预期参数时,将初始图像分类模型作为图像分类模型。客户端执行上线步骤,将图像分类模型发送到服务器标识信息对应的上线服务器中。上线服务器自动对图像分类模型进行上线。

[0167] 本申请实施例提供了一种模型处理方法,该方法包括:接收业务模型的模型处理配置指令,将模型处理配置指令中的内容配置到预设的工作流流程信息中,生成业务模型对应的目标 workflow 运行实例,目标 workflow 运行实例包括模型训练实例和模型测试实例。运行模型训练实例,以训练目标机器学习模型,得到目标训练模型。运行模型测试实例,以测试目标训练模型,得到测试结果。在测试结果与模型预期参数相匹配的情况下,将目标训练模型作为业务模型。该方法可以自动完成模型训练、测试等流程,提高了模型处理的效率,简化了模型处理的操作,从而节省了人工成本和运维成本。

[0168] 本申请实施例还提供了一种模型处理装置,请参见图18,该装置包括:配置信息获取模块1810、workflow 运行实例生成模块1820、模型训练模块1830、模型测试模块1840和业务模型确定模块1850;

[0169] 配置信息获取模块1810用于接收目标业务的模型处理配置指令,模型处理配置指令包括目标业务对应的目标机器学习模型的模型关联特征信息,以及目标机器学习模型在目标业务下的数据集路径信息、训练轮次和模型预期参数;

[0170] workflow 运行实例生成模块1820用于将模型关联特征信息、数据集路径信息和训练轮次配置到预设的工作流流程信息中,生成目标业务对应的目标 workflow 运行实例,目标 workflow 运行实例包括模型训练实例和模型测试实例;

[0171] 模型训练模块1830用于通过运行模型训练实例训练目标机器学习模型,得到目标训练模型;

[0172] 模型测试模块1840用于通过运行模型测试实例测试目标训练模型,得到测试结果;

[0173] 业务模型确定模块1850用于在测试结果与模型预期参数相匹配的情况下,将目标训练模型作为业务模型,业务模型用于执行目标业务。

[0174] 进一步地,数据集路径信息包括训练数据路径信息,目标机器学习模型的数目有多个,模型训练模块1830包括:

[0175] 训练数据获取单元,用于根据训练数据路径信息,获取训练数据集;

[0176] 当前训练集获取单元,用于从训练数据集中获取当前训练轮次对应的训练数据集;

[0177] 模型训练单元,用于基于当前训练轮次对应的训练数据集,分别对上一训练轮次对应的目标训练后模型进行训练,得到当前训练轮次对应的训练后模型;其中,在当前训练轮次为第一训练轮次时,上一训练轮次对应的目标训练后模型为多个目标机器学习模型;

[0178] 模型排序单元,用于对当前训练轮次对应的训练后模型,按照每个训练后模型对应的训练效果数据由大至小依次排序,得到当前训练轮次对应的训练后模型序列,训练效果数据表征训练后模型与目标业务的匹配程度;

[0179] 模型筛选单元,用于从当前训练轮次对应的训练后模型序列中,由前往后确定当前训练轮次对应的目标训练后模型;

[0180] 目标训练模型确定单元,用于若当前训练轮次对应的目标训练后模型不满足预设训练条件,则返回执行从训练数据集中获取当前训练轮次对应的训练数据集及后续步骤,直至当前训练轮次对应的目标训练后模型满足预设训练条件时,从当前训练轮次对应的目标训练后模型中确定目标训练模型。

[0181] 进一步地,当前训练集获取单元包括:

[0182] 采样数量确定单元,用于基于预设采样条件,确定当前训练轮次对应的采样数量,其中,第一个训练轮次对应的采样数量低于其他训练轮次的采样数量;

[0183] 训练数据集采样单元,用于基于当前训练轮次对应的采样数量,从训练数据集中获取采样数量个训练数据,得到当前训练轮次对应的当前训练数据集。。

[0184] 进一步地,模型筛选单元包括:

[0185] 筛选数量确定单元,用于基于预设筛选条件,确定当前训练轮次对应的筛选数量,其中,第一个训练轮次对应的筛选数量高于其他训练轮次的筛选数量;

[0186] 训练后模型筛选单元,用于根据所述当前训练轮次对应的筛选数量,从所述训练后模型序列中由前到后确定多个训练后模型。

[0187] 进一步地,该装置还包括:

[0188] 初始数据集接收模块,用于接收初始数据集;

[0189] 数据集解析实例生成模块,用于将初始数据集配置到预设的数据解析流程信息中,生成初始数据集对应的数据集解析实例;

[0190] 数据解析模块,用于通过运行数据集解析实例解析初始数据集,得到数据解析结果;

[0191] 解析结果存储模块,用于将数据解析结果发送到云端进行存储。

[0192] 进一步地,该装置还包括:

[0193] 机器学习模型接收模块,用于接收机器学习模型;

[0194] 应用状态生成模块,用于生成机器学习模型的初始应用状态信息,初始应用状态信息表征机器学习模型不可应用;

[0195] 验证数据获取模块,用于获取机器学习模型的训练数据路径信息;

[0196] 模型验证实例生成模块,用于将训练数据路径信息配置到预设的工作流实例中,生成模型验证实例;

[0197] 模型验证模块,用于通过运行模型验证实例训练机器学习模型,得到机器学习模型对应的验证数据;

[0198] 应用状态更新模块,用于在验证数据满足预设应用条件时,将机器学习模型的初始应用状态信息更新为可应用。

[0199] 进一步地,配置信息获取模块1810包括:

[0200] 界面显示单元,用于显示模型处理配置界面;

[0201] 模型配置确认单元,用于响应于模型处理配置界面中的模型配置确认指令,从初始应用状态信息更新为可应用的机器学习模型中获取目标业务对应的目标机器学习模型的模型关联特征信息,并获取目标机器学习模型在目标业务下的数据集路径信息、训练轮次和模型预期参数;

[0202] 配置指令生成单元,用于基于模型关联特征信息、数据集路径信息、训练轮次和模

型预期参数,生成模型处理配置指令。

[0203] 上述实施例中提供的装置可执行本申请任意实施例所提供方法,具备执行该方法相应的功能模块和有益效果。未在上述实施例中详尽描述的技术细节,可参见本申请任意实施例所提供的一种模型处理方法。

[0204] 本实施例还提供了一种计算机可读存储介质,存储介质中存储有计算机可执行指令,计算机可执行指令由处理器加载并执行本实施例上述的一种模型处理方法。

[0205] 本实施例还提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述模型处理的各种可选实现方式中提供的方法。

[0206] 本实施例还提供了一种电子设备,该电子设备包括处理器和存储器,其中,存储器存储有计算机程序,计算机程序适于由处理器加载并执行本实施例上述的一种模型处理方法。

[0207] 设备可以为计算机终端、移动终端或服务器,设备还可以参与构成本申请实施例所提供的装置或系统。如图19所示,计算机终端19可以包括一个或多个(图中采用1902a、1902b,……,1902n来示出)处理器1902(处理器1902可以包括但不限于微处理器MCU或可编程逻辑器件FPGA等的处理装置)、用于存储数据的存储器1904、以及用于通信功能的传输装置1906。除此以外,还可以包括:显示器、输入/输出接口(I/O接口)、网络接口、电源和/或相机。本领域普通技术人员可以理解,图19所示的结构仅为示意,其并不对上述电子装置的结构造成限定。例如,计算机终端19还可包括比图19中所示更多或者更少的组件,或者具有与图19所示不同的配置。

[0208] 应当注意到的是上述一个或多个处理器1902和/或其他数据处理电路在本文中通常可以被称为“数据处理电路”。该数据处理电路可以全部或部分的体现为软件、硬件、固件或其他任意组合。此外,数据处理电路可为单个独立的处理模块,或全部或部分的结合到计算机终端19中的其他元件中的任意一个内。

[0209] 存储器1904可用于存储应用软件的程序以及模块,如本申请实施例中的方法对应的程序指令/数据存储装置,处理器1902通过运行存储在存储器1904内的软件程序以及模块,从而执行各种功能应用以及数据处理,即实现上述的一种基于自注意力网络的时序行为捕捉框生成方法。存储器1904可包括高速随机存储器,还可包括非易失性存储器,如一个或者多个磁性存储装置、闪存、或者其他非易失性固态存储器。在一些实例中,存储器1904可进一步包括相对于处理器1902远程设置的存储器,这些远程存储器可以通过网络连接至计算机终端19。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0210] 传输装置1906用于经由一个网络接收或者发送数据。上述的网络具体实例可包括计算机终端19的通信供应商提供的无线网络。在一个实例中,传输装置1906包括一个网络适配器(Network Interface Controller, NIC),其可通过基站与其他网络设备相连从而可与互联网进行通讯。在一个实例中,传输装置1906可以为射频(Radio Frequency, RF)模块,其用于通过无线方式与互联网进行通讯。

[0211] 显示器可以例如触摸屏式的液晶显示器(LCD),该液晶显示器可使得用户能够与

计算机终端19的用户界面进行交互。

[0212] 本说明书提供了如实施例或流程图的方法操作步骤,但基于常规或者无创造性的劳动可以包括更多或者更少的操作步骤。实施例中列举的步骤和顺序仅仅为众多步骤执行顺序中的一种方式,不代表唯一的执行顺序。在实际中的系统或中断产品执行时,可以按照实施例或者附图所示的方法顺序执行或者并行执行(例如并行处理器或者多线程处理的环境)。

[0213] 本实施例中所示出的结构,仅仅是与本申请方案相关的部分结构,并不构成对本申请方案所应用于其上的设备的限定,具体的设备可以包括比示出的更多或更少的部件,或者组合某些部件,或者具有不同的部件的布置。应当理解到,本实施例中所揭露的方法、装置等,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,模块的划分仅仅为一种逻辑功能的划分,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,装置或单元模块的间接耦合或通信连接。

[0214] 基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本申请各个实施例方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0215] 本领域技术人员还可以进一步意识到,结合本说明书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,但这种实现不应认为超出本申请的范围。

[0216] 以上,以上实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围。

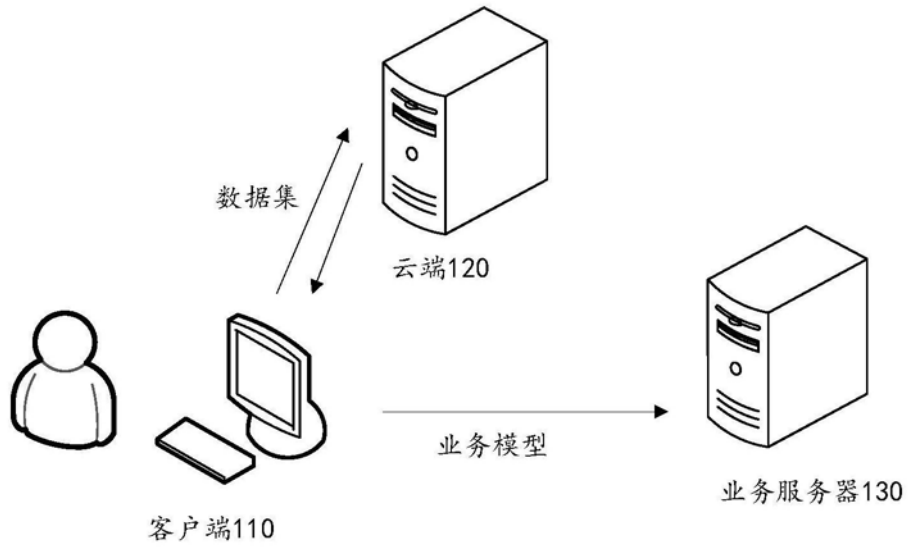


图1

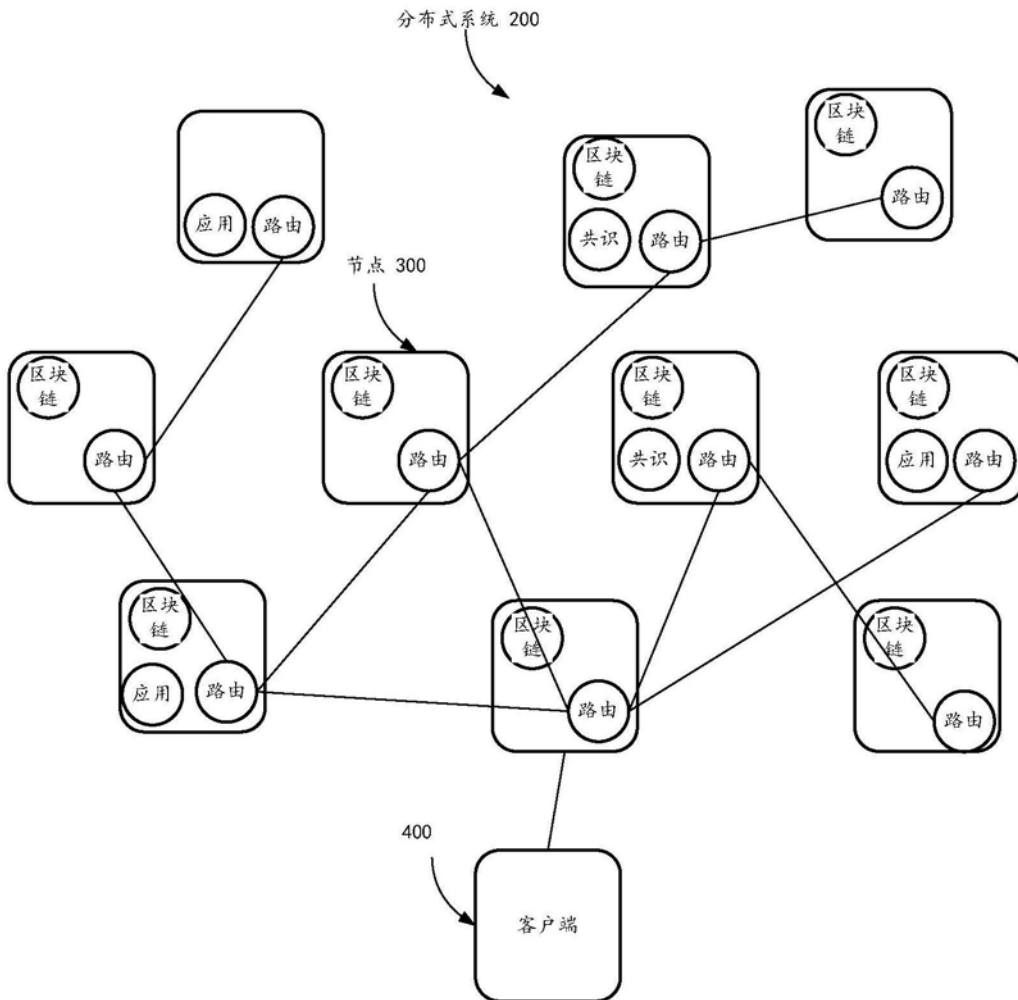


图2

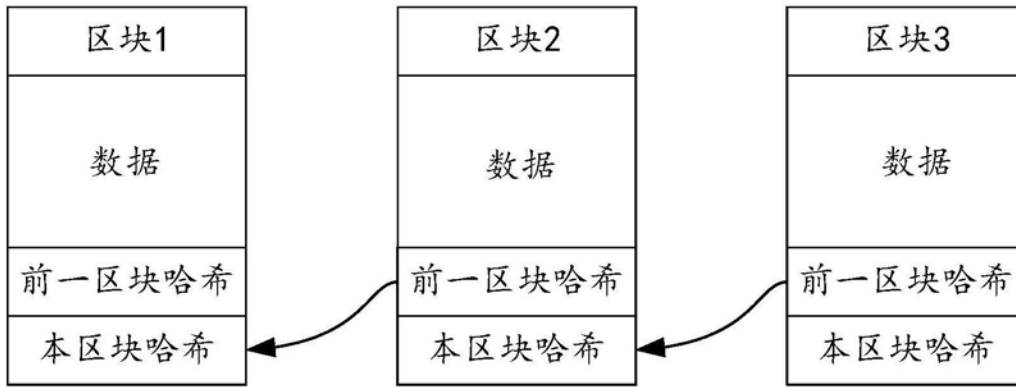


图3

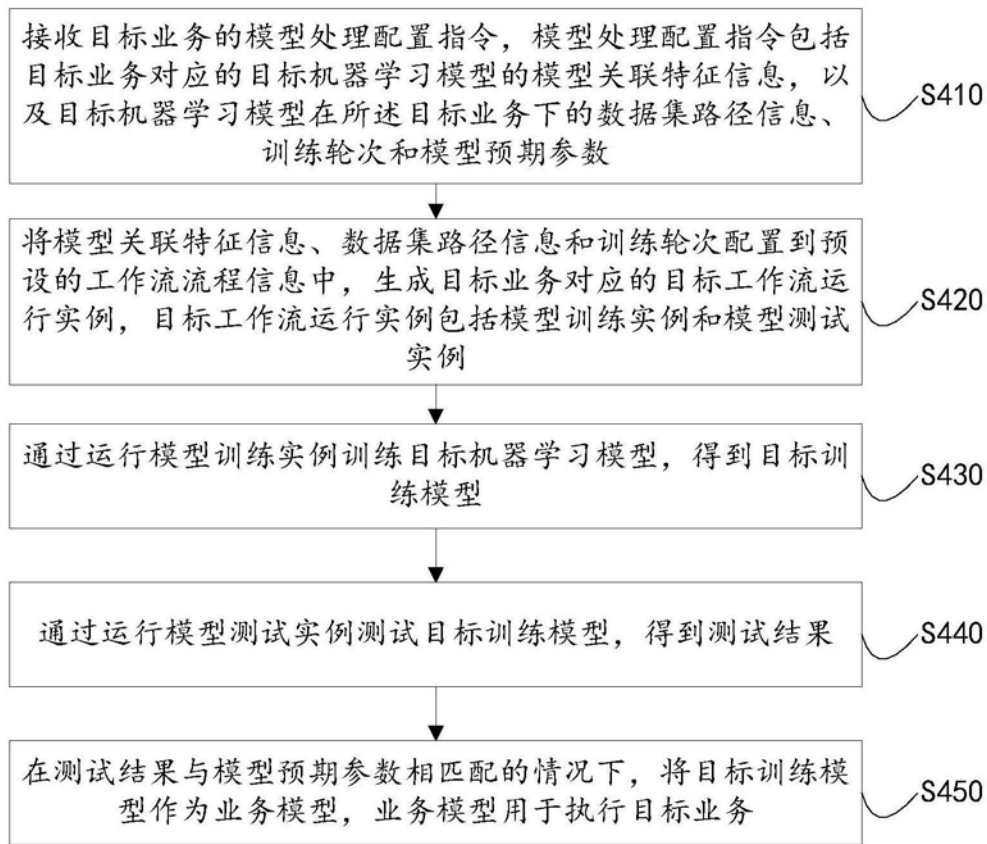


图4

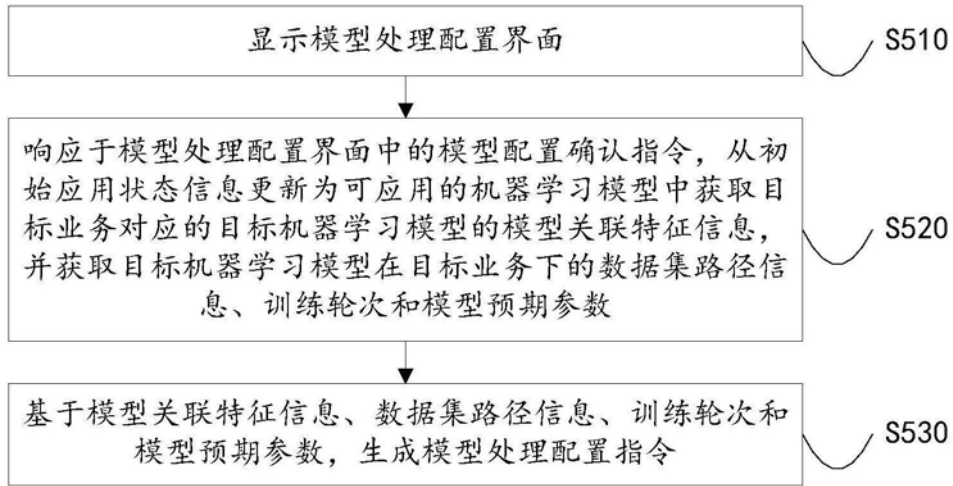


图5



图6

模型列表 / 训练配置

训练配置 评估训练时长

* 训练数据集:

* 任务类型:

* 选料特征:

* 基础模型:

高级配置: 高级模式

模型评估 训练后自动评估

* 评估数据集:

模型上线 评估达标后自动上线

* 服务名称:

* 预期 精确率: %

* 预期 召回率: %

* 预期 F1: %

图7

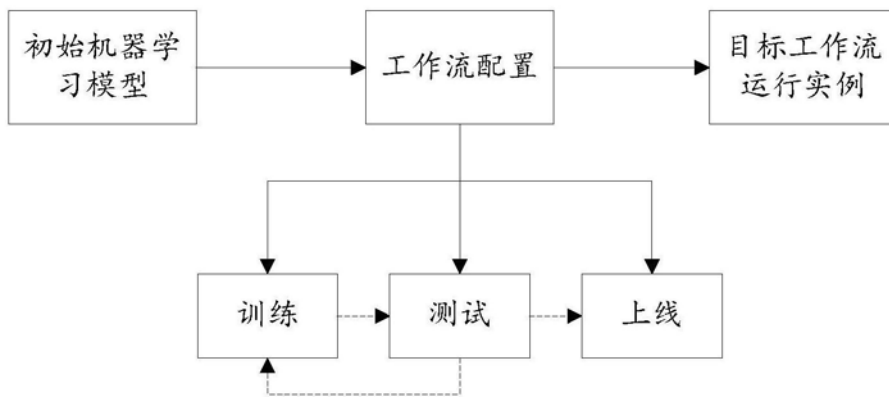


图8



图9

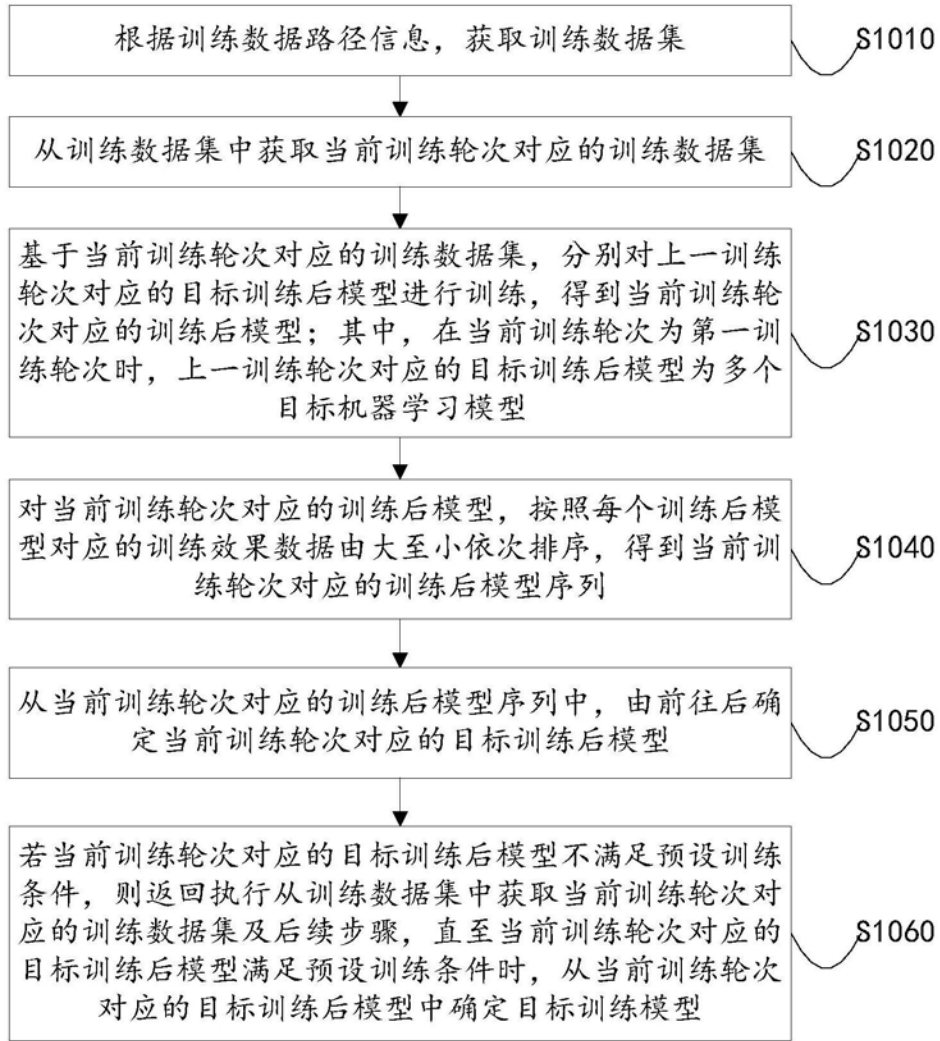


图10

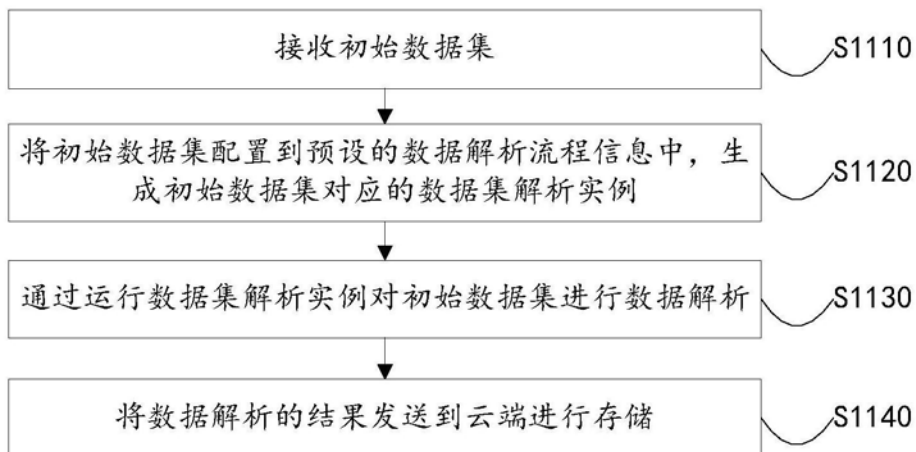


图11

数据上传 ×

* 应用组:

* 业务类型:

* 数据来源:

* 视频来源:

* 码率Id:

* 数据集名称:
数据集名称只能由小写英文、下划线组成

* 文件选择:

数据集描述:

维护人:

图12

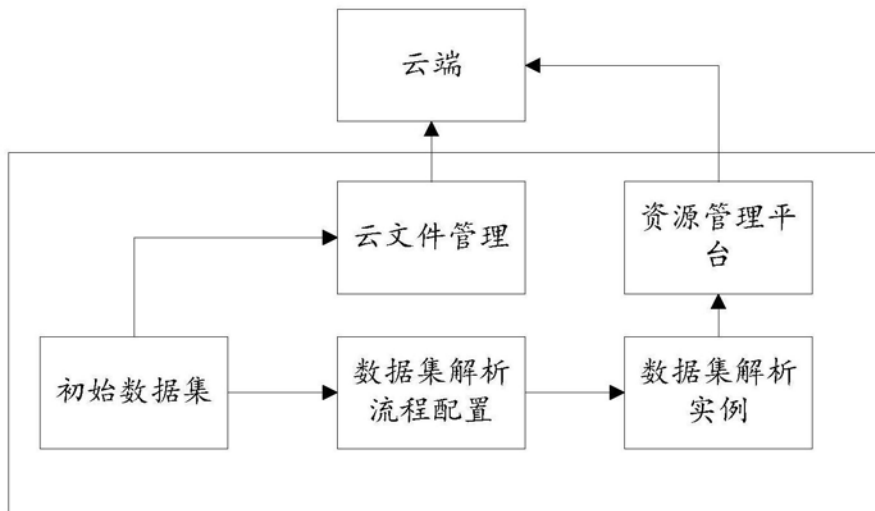


图13

数据源名称	业务类型	描述	状态	维护人	更新时间	操作
ferkel_10_old_predict_1210	video		• 解析成功	saqyan	2020-12-10 22:31:31	工作流 查看样本
ferkel_train_10_1210	video		• 解析成功	saqyan	2020-12-10 22:22:30	工作流 查看样本
ferkel_train_8_1210	video		• 解析成功	saqyan	2020-12-10 22:18:31	工作流 查看样本
ferkel_8_old_predict_1210	video		• 解析成功	saqyan	2020-12-10 21:41:32	工作流 查看样本
quality_train_all_1204_3	video		• 解析成功	saqyan	2020-12-08 08:19:11	工作流 查看样本
qqal_predict_1207	video		• 解析成功	saqyan	2020-12-07 19:22:10	工作流 查看样本
qqal_train_1207_1	video		• 解析成功	saqyan	2020-12-07 19:12:10	工作流 查看样本
qqal_train_1207	video		• 解析失败	saqyan	2020-12-07 18:44:11	工作流
quality_train_junheng_1204	video		• 解析成功	saqyan	2020-12-06 22:09:10	工作流 查看样本
quality_train_suji_1204_1	video		• 解析成功	saqyan	2020-12-05 13:23:10	工作流 查看样本
baoshe_test_data_5w	video	测试	• 解析成功	baoshechen	2020-12-04 16:35:11	编辑 工作流 查看样本
quality_train_all_1204_1	video		• 解析失败	saqyan	2020-12-04 15:13:10	工作流
quality_train_all_1204	video		• 解析成功	saqyan	2020-12-04 12:58:42	工作流 查看样本
quality_junheng_train_1204	video		• 解析失败	saqyan	2020-12-04 11:05:42	工作流
whiskoning_train_1203	video	dfs	• 解析成功	saqyan	2020-12-03 21:43:42	工作流 查看样本
quality_train_junheng_1203	video		• 解析成功	saqyan	2020-12-03 21:29:42	工作流 查看样本

图14



图15

基础模型名: 负责人:

查询结果

基础模型名	预训练模型名	路径	负责人	上传时间	上传状态	模型状态	操作
A	test1	xxxx/xxxx/xxxx			已上传	待测试	发起测试
B	test2	xxxx/xxxx/xxxx			上传中	-	-
C	test3	xxxx/xxxx/xxxx			已上传	可用	测试报告
D	test2	xxxx/xxxx/xxxx			上传中	-	-
E	test3	xxxx/xxxx/xxxx			已上传	待测试	发起测试

图16

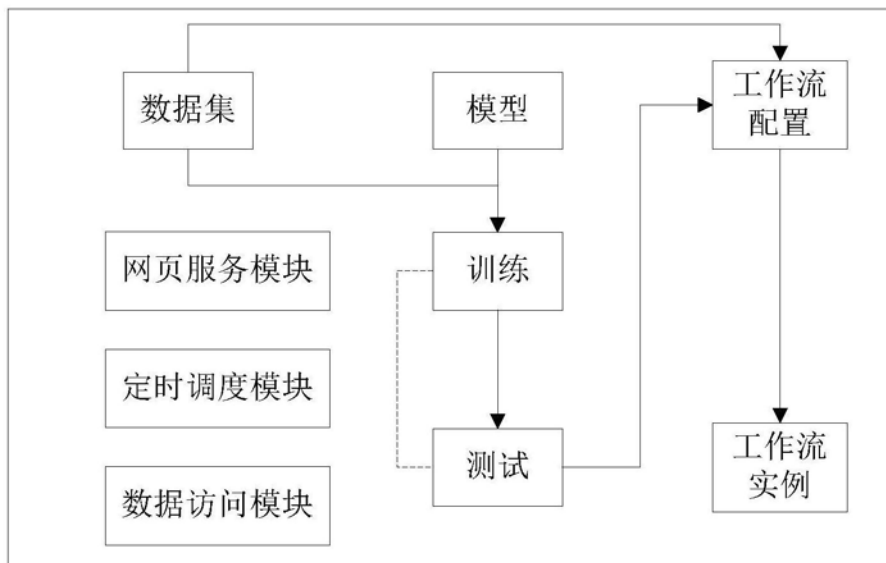


图17



图18

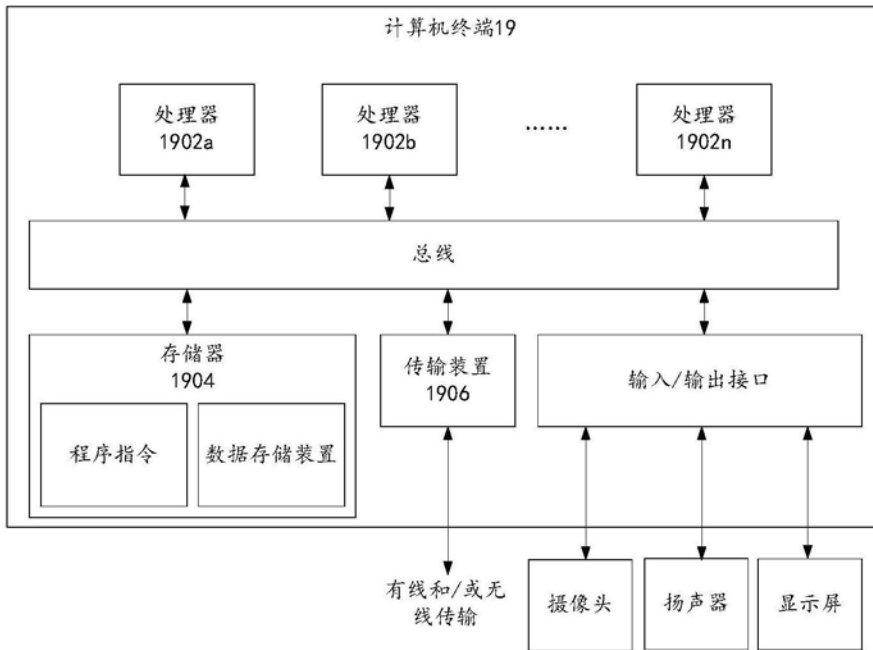


图19