

权 利 要 求 书

1、一种用于处理模型参数的方法，包括：

5 获取待分片的模型参数集合，其中，所述模型参数集合包括特征表示对应的多维数组；

获取用于存储所述待分片的模型参数集合的存储系统的属性信息，其中，所述用于存储所述待分片的模型参数集合的存储系统有别于所述待分片的模型参数集合对应的模型运行时所在的系统；

10 根据所述属性信息，将所述待分片的模型参数集合存入所述存储系统。

2、根据权利要求 1 所述的方法，其中，所述获取待分片的模型参数集合，包括：

获取所述待分片的模型参数集合对应的模型文件；

15 对所述模型文件进行解析，生成对应的解析图，其中，所述解析图中包括变量和对变量的操作；

根据预设的字段，从所述解析图中提取变量作为目标变量；

根据所述目标变量对应的特征表示，获取所述待分片的模型参数集合。

20

3、根据权利要求 2 所述的方法，其中，所述根据所述目标变量对应的特征表示，获取所述待分片的模型参数集合，包括：

获取所述根据所述目标变量对应的特征表示对应的多维数组的数据量；

25 响应于确定获取的数据量大于预设数据量阈值，获取所述根据所述目标变量对应的特征表示对应的多维数组作为所述待分片的模型参数集合。

4、根据权利要求 1 所述的方法，其中，所述方法还包括：

30 接收所述待分片的模型参数集合对应的特征表示查询操作；

根据所述特征表示查询操作，在所述存储系统中查询与所述特征

表示查询操作匹配的结果；

发送所述匹配的结果。

5 5、根据权利要求 1-4 之一所述的方法，其中，所述属性信息包括
存储系统的存储空间分布；以及

所述根据所述属性信息，将所述待分片的模型参数集合存入所述
存储系统，包括：

根据所述属性信息，生成模型参数分配表，其中，所述模型参数
分配表用于表征所述待分片的模型参数集合与所述存储系统的存储位
10 置之间的对应关系；

根据所述模型参数分配表，将所述待分片模型参数集合存入所述
存储系统对应的存储位置，以及生成查询接口。

15 6、根据权利要求 1-3 之一所述的方法，其中，所述属性信息包括
键值存储系统的标识；以及

所述根据所述属性信息，将所述待分片的模型参数集合存入所述
存储系统，包括：

获取前次更新的模型参数集合作为参考模型参数集合；

确定所述待分片的模型参数集合与所述参考模型参数集合之间的
20 差异，生成更新模型参数集合；

根据所述属性信息，将所述更新模型参数集合存入所述存储系统。

7、根据权利要求 6 所述的方法，其中，所述方法还包括：

接收所述待分片的模型参数集合对应的特征表示查询操作；

25 根据所述特征表示查询操作，生成相应数量的远程调用请求；

向所述键值存储系统发送所述远程调用请求；

响应于接收到所述键值存储系统发送的包括与所述特征表示查询
操作匹配的结果的响应信息，发送所述匹配的结果。

30 8、一种用于处理模型参数的装置，包括：

第一获取单元，被配置成获取待分片的模型参数集合，其中，所述模型参数集合包括特征表示对应的多维数组；

第二获取单元，被配置成获取用于存储所述待分片的模型参数集合的存储系统的属性信息，其中，所述用于存储所述待分片的模型参数集合的存储系统有别于所述待分片的模型参数集合对应的模型运行时所在的系统；

存储单元，被配置成根据所述属性信息，将所述待分片的模型参数集合存入所述存储系统。

10 9、一种服务器，包括：

一个或多个处理器；

存储装置，其上存储有一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求 1-7 中任一所述的方法。

15

10、一种计算机可读介质，其上存储有计算机程序，其中，该程序被处理器执行时实现如权利要求 1-7 中任一所述的方法。

20 11、一种计算机程序产品，包括计算机程序，所述计算机程序在被处理器执行时实现如权利要求 1-7 中任一所述的方法。

12、一种计算机程序，所述计算机程序在被处理器执行时实现如权利要求 1-7 中任一所述的方法。