(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 116521952 B
(45)授权公告日 2023.09.05

(21)申请号 202310812592.6

(22)申请日 2023.07.04

(65)同一申请的已公布的文献号
申请公布号 CN 116521952 A

(43)申请公布日 2023.08.01

(73)专利权人 北京富算科技有限公司
地址 102699 北京市大兴区黄村东大街38号院3号楼5层505

(72)发明人 尤志强 卞阳 王兆凯 张伟奇

(74)专利代理机构 北京慧加伦知识产权代理有限公司 16035
专利代理师 李永敏

(51)Int.Cl.
G06F 16/901(2019.01)
G06F 18/2415(2023.01)
G06F 18/243(2023.01)
G06F 17/16(2006.01)
G06N 20/20(2019.01)

(56)对比文件
CN 115630711 A,2023.01.20
CN 115438370 A,2022.12.06
CN 114819057 A,2022.07.29
CN 114372516 A,2022.04.19
WO 2021179720 A1,2021.09.16

审查员 李萌

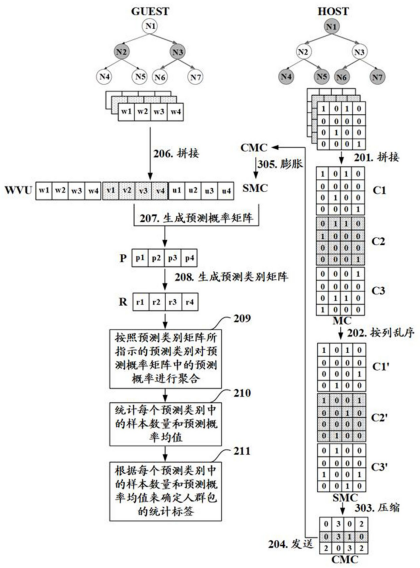
权利要求书3页 说明书13页 附图8页

(54)发明名称

使用联邦学习模型进行人群包统计的方法及装置

(57)摘要

本公开的实施例提供一种使用联邦学习模型进行人群包统计的方法及装置。联邦学习模型包括多棵树。参与联邦学习的第一参与方拥有多棵树中的每棵树的叶子节点的权重矩阵。参与联邦学习的第二参与方拥有多棵树中的每棵树针对人群包生成的预测结果矩阵。该方法由第一参与方执行。该方法包括:将多棵树的权重矩阵拼接成第一拼接矩阵;获得由第二参与方生成的第二拼接矩阵,第二拼接矩阵通过将多棵树的预测结果矩阵进行按列拼接并执行按列乱序操作来生成;将第一拼接矩阵与第二拼接矩阵进行矩阵相乘以获得预测概率矩阵;以及根据预测概率矩阵来确定人群包的统计信息。



CN 116521952 B

1.一种使用联邦学习模型进行人群包统计的方法,其特征在于,所述联邦学习模型包括多棵树,参与联邦学习的第一参与方拥有所述多棵树中的每棵树的叶子节点的权重矩阵,参与所述联邦学习的第二参与方拥有所述多棵树中的每棵树针对所述人群包生成的预测结果矩阵,所述方法由所述第一参与方执行,所述方法包括:

将所述多棵树的所述权重矩阵拼接成第一拼接矩阵;

获得由所述第二参与方生成的第二拼接矩阵,所述第二拼接矩阵通过将所述多棵树的所述预测结果矩阵进行按列拼接并执行按列乱序操作来生成;

将所述第一拼接矩阵与所述第二拼接矩阵进行矩阵相乘以获得预测概率矩阵;以及

根据所述预测概率矩阵来确定所述人群包的统计信息;

其中,根据所述预测概率矩阵来确定所述人群包的统计信息包括:

根据所述预测概率矩阵来生成预测类别矩阵,所述预测类别矩阵指示所述人群包中的每个样本的预测类别;

按照所述预测类别矩阵所指示的预测类别对所述预测概率矩阵中的预测概率进行聚合;

统计每个预测类别中的样本数量和预测概率均值;以及

根据每个预测类别中的样本数量和预测概率均值来确定所述人群包的统计标签。

2.根据权利要求1所述的方法,其特征在于,根据所述预测概率矩阵来生成预测类别矩阵包括:

在二分类场景下,确定所述预测概率矩阵中针对每个样本的预测概率是否超过预设的概率阈值;

响应于任一样本的预测概率高于所述概率阈值,确定该样本的预测类别为第一类别;

响应于任一样本的预测概率低于或者等于所述概率阈值,确定该样本的预测类别为第二类别;

在多分类场景下,确定所述预测概率矩阵中针对每个样本的多个预测概率中的最大预测概率,其中,所述多个预测概率中的每个预测概率对应一个类别;以及

针对每个样本,确定该样本的预测类别为针对该样本的最大预测概率所对应的类别。

3.根据权利要求1至2中任一项所述的方法,其特征在于,获得由所述第二参与方生成的第二拼接矩阵包括:

接收由所述第二参与方根据所述第二拼接矩阵生成的压缩矩阵;以及

根据所述压缩矩阵来生成所述第二拼接矩阵;

其中,所述多棵树中的每棵树所生成的预测结果矩阵对应所述压缩矩阵的一行,所述压缩矩阵的同一行中的每一列记录与该列相对应的样本在与该行相对应的预测结果矩阵中的预测结果。

4.一种使用联邦学习模型进行人群包统计的装置,其特征在于,所述联邦学习模型包括多棵树,参与联邦学习的第一参与方拥有所述多棵树中的每棵树的叶子节点的权重矩阵,参与所述联邦学习的第二参与方拥有所述多棵树中的每棵树针对所述人群包生成的预测结果矩阵,所述装置作为所述第一参与方,所述装置包括:

至少一个处理器;以及

存储有计算机程序的至少一个存储器;

其中，当所述计算机程序由所述至少一个处理器执行时，使得所述装置执行根据权利要求1至3中任一项所述的方法的步骤。

5.一种使用联邦学习模型进行人群包统计的方法，其特征在于，所述联邦学习模型包括多棵树，参与联邦学习的第一参与方拥有所述多棵树中的每棵树的叶子节点的权重矩阵，参与所述联邦学习的第二参与方拥有所述多棵树中的每棵树针对所述人群包生成的预测结果矩阵，所述方法由所述第二参与方执行，所述方法包括：

将所述多棵树的预测结果矩阵进行按列拼接以生成第三拼接矩阵；

对所述第三拼接矩阵执行按列乱序操作以生成第二拼接矩阵；以及

向所述第一参与方提供所述第二拼接矩阵的相关信息，以便所述第一参与方根据第一拼接矩阵和所述第二拼接矩阵来确定所述人群包的统计信息；

其中，所述第一拼接矩阵由所述第一参与方通过将所述多棵树的所述权重矩阵拼接来生成。

6.根据权利要求5所述的方法，其特征在于，向所述第一参与方提供所述第二拼接矩阵的相关信息包括：

根据所述第二拼接矩阵生成压缩矩阵；以及

向所述第一参与方发送所述压缩矩阵；

其中，所述多棵树中的每棵树所生成的预测结果矩阵对应所述压缩矩阵的一行，所述压缩矩阵的同一行中的每一列记录与该列相对应的样本在与该行相对应的预测结果矩阵中的预测结果。

7.根据权利要求5或6所述的方法，其特征在于，所述第二参与方通过以下操作来拥有每棵树的所述预测结果矩阵：

从所述第一参与方接收针对该棵树的第一样本索引，所述第一样本索引由所述第一参与方根据该棵树的第一节点分裂条件推理获得，所述第一样本索引指示所述人群包中的样本与该棵树的所述叶子节点的第一预测关系；

根据该棵树的第二节点分裂条件推理获得第二样本索引，所述第二样本索引指示所述人群包中的样本与该棵树的所述叶子节点的第二预测关系；

对所述第一样本索引和所述第二样本索引求交集以获得预测样本索引；以及

将所述预测样本索引转换成矩阵形式以获得该棵树的所述预测结果矩阵。

8.根据权利要求5或6所述的方法，其特征在于，所述第二参与方通过以下操作来拥有每棵树的所述预测结果矩阵：

获得所述第一参与方生成的第一样本索引的第一碎片矩阵，所述第一样本索引由所述第一参与方根据该棵树的第一节点分裂条件推理获得，所述第一样本索引指示所述人群包中的样本与该棵树的所述叶子节点的第一预测关系，所述第一样本索引被转换成第一样本索引矩阵，所述第一样本索引矩阵被碎片化成所述第一碎片矩阵和第二碎片矩阵；

根据该棵树的第二节点分裂条件推理获得第二样本索引，所述第二样本索引指示所述人群包中的样本与该棵树的所述叶子节点的第二预测关系；

将所述第二样本索引转换成矩阵形式以获得第二样本索引矩阵；

将所述第二样本索引矩阵碎片化成第三碎片矩阵和第四碎片矩阵；

获得所述第一参与方根据所述第二碎片矩阵和所述第三碎片矩阵生成的第一中间碎

片矩阵和第二中间碎片矩阵，其中，所述第三碎片矩阵由所述第二参与方发送给所述第一参与方或者由所述第一参与方生成；

根据所述第一碎片矩阵和所述第四碎片矩阵生成第三中间碎片矩阵和第四中间碎片矩阵；

向所述第一参与方发送所述第三中间碎片矩阵和所述第四中间碎片矩阵；

获得所述第一参与方根据所述第一中间碎片矩阵、所述第二中间碎片矩阵、所述第三中间碎片矩阵和所述第四中间碎片矩阵生成的第一交集碎片矩阵；

根据所述第一中间碎片矩阵、所述第二中间碎片矩阵、所述第三中间碎片矩阵和所述第四中间碎片矩阵生成第二交集碎片矩阵；以及

将所述第一交集碎片矩阵与所述第二交集碎片矩阵相加以获得该棵树的所述预测结果矩阵。

9.一种使用联邦学习模型进行人群包统计的装置，其特征在于，所述联邦学习模型包括多棵树，参与联邦学习的第一参与方拥有所述多棵树中的每棵树的叶子节点的权重矩阵，参与所述联邦学习的第二参与方拥有所述多棵树中的每棵树针对所述人群包生成的预测结果矩阵，所述装置作为所述第二参与方，所述装置包括：

至少一个处理器；以及

存储有计算机程序的至少一个存储器；

其中，当所述计算机程序由所述至少一个处理器执行时，使得所述装置执行根据权利要求5至8中任一项所述的方法的步骤。

使用联邦学习模型进行人群包统计的方法及装置

技术领域

[0001] 本公开的实施例涉及数据处理技术领域，具体地，涉及使用联邦学习模型进行人群包统计的方法及装置。

背景技术

[0002] 基于XGBoost的联邦学习模型(也可称为XGBoost模型)是常用隐私计算模型之一。如今在许多应用场景中XGBoost模型已被广泛使用，例如金融风控，广告营销，疾病预测等。在银行、电商等公司的应用场景中，往往会采用XGBoost模型来作为主要的机器学习模型。在实际应用中，有时需要对人群包做统计，例如预测人群包的兴趣偏好或者所属类别，从而为下游任务提供有意义的参考依据。在使用XGBoost模型进行人群包统计的过程中，如果人群包中的个体的模型预测值等信息被定位，则不能很好保护个体信息，难以满足合规需求。因此期望能够在不暴露个体信息的情况下进行人群包统计。

发明内容

[0003] 本文中描述的实施例提供了一种使用联邦学习模型进行人群包统计的方法、装置以及存储有计算机程序的计算机可读存储介质。

[0004] 根据本公开的第一方面，提供了一种使用联邦学习模型进行人群包统计的方法。联邦学习模型包括多棵树。参与联邦学习的第一参与方拥有多棵树中的每棵树的叶子节点的权重矩阵。参与联邦学习的第二参与方拥有多棵树中的每棵树针对人群包生成的预测结果矩阵。该方法由第一参与方执行。该方法包括：将多棵树的权重矩阵拼接成第一拼接矩阵；获得由第二参与方生成的第二拼接矩阵，第二拼接矩阵通过将多棵树的预测结果矩阵进行按列拼接并执行按列乱序操作来生成；将第一拼接矩阵与第二拼接矩阵进行矩阵相乘以获得预测概率矩阵；以及根据预测概率矩阵来确定人群包的统计信息。

[0005] 在本公开的一些实施例中，根据预测概率矩阵来确定人群包的统计信息包括：根据预测概率矩阵来生成预测类别矩阵，预测类别矩阵指示人群包中的每个样本的预测类别；按照预测类别矩阵所指示的预测类别对预测概率矩阵中的预测概率进行聚合；统计每个预测类别中的样本数量和预测概率均值；以及根据每个预测类别中的样本数量和预测概率均值来确定人群包的统计标签。

[0006] 在本公开的一些实施例中，根据预测概率矩阵来生成预测类别矩阵包括：在二分类场景下，确定预测概率矩阵中针对每个样本的预测概率是否超过预设的概率阈值；响应于任一样本的预测概率高于概率阈值，确定该样本的预测类别为第一类别；响应于任一样本的预测概率低于或者等于概率阈值，确定该样本的预测类别为第二类别。

[0007] 在本公开的一些实施例中，根据预测概率矩阵来生成预测类别矩阵包括：在多分类场景下，确定预测概率矩阵中针对每个样本的多个预测概率中的最大预测概率，其中，多个预测概率中的每个预测概率对应一个类别；以及针对每个样本，确定该样本的预测类别为针对该样本的最大预测概率所对应的类别。

[0008] 在本公开的一些实施例中，获得由第二参与方生成的第二拼接矩阵包括：从第二参与方直接接收第二拼接矩阵。

[0009] 在本公开的一些实施例中，获得由第二参与方生成的第二拼接矩阵包括：接收由第二参与方根据第二拼接矩阵生成的压缩矩阵；以及根据压缩矩阵来生成第二拼接矩阵；其中，多棵树中的每棵树所生成的预测结果矩阵对应压缩矩阵的一行，压缩矩阵的同一行中的每一列记录与该列相对应的样本在与该行相对应的预测结果矩阵中的预测结果。

[0010] 根据本公开的第二方面，提供了一种使用联邦学习模型进行人群包统计的装置。联邦学习模型包括多棵树。参与联邦学习的第一参与方拥有多棵树中的每棵树的叶子节点的权重矩阵。参与联邦学习的第二参与方拥有多棵树中的每棵树针对人群包生成的预测结果矩阵。该装置作为第一参与方。该装置包括至少一个处理器；以及存储有计算机程序的至少一个存储器。当计算机程序由至少一个处理器执行时，使得装置：将多棵树的权重矩阵拼接成第一拼接矩阵；获得由第二参与方生成的第二拼接矩阵，第二拼接矩阵通过将多棵树的预测结果矩阵进行按列拼接并执行按列乱序操作来生成；将第一拼接矩阵与第二拼接矩阵进行矩阵相乘以获得预测概率矩阵；以及根据预测概率矩阵来确定人群包的统计信息。

[0011] 在本公开的一些实施例中，计算机程序在由至少一个处理器执行时使得装置通过以下操作来根据预测概率矩阵来确定人群包的统计信息：根据预测概率矩阵来生成预测类别矩阵，预测类别矩阵指示人群包中的每个样本的预测类别；按照预测类别矩阵所指示的预测类别对预测概率矩阵中的预测概率进行聚合；统计每个预测类别中的样本数量和预测概率均值；以及根据每个预测类别中的样本数量和预测概率均值来确定人群包的统计标签。

[0012] 在本公开的一些实施例中，计算机程序在由至少一个处理器执行时使得装置通过以下操作来根据预测概率矩阵来生成预测类别矩阵：在二分类场景下，确定预测概率矩阵中针对每个样本的预测概率是否超过预设的概率阈值；响应于任一样本的预测概率高于概率阈值，确定该样本的预测类别为第一类别；响应于任一样本的预测概率低于或者等于概率阈值，确定该样本的预测类别为第二类别。

[0013] 在本公开的一些实施例中，计算机程序在由至少一个处理器执行时使得装置通过以下操作来根据预测概率矩阵来生成预测类别矩阵：在多分类场景下，确定预测概率矩阵中针对每个样本的多个预测概率中的最大预测概率，其中，多个预测概率中的每个预测概率对应一个类别；以及针对每个样本，确定该样本的预测类别为针对该样本的最大预测概率所对应的类别。

[0014] 在本公开的一些实施例中，计算机程序在由至少一个处理器执行时使得装置通过以下操作来获得由第二参与方生成的第二拼接矩阵：从第二参与方直接接收第二拼接矩阵。

[0015] 在本公开的一些实施例中，计算机程序在由至少一个处理器执行时使得装置通过以下操作来获得由第二参与方生成的第二拼接矩阵：接收由第二参与方根据第二拼接矩阵生成的压缩矩阵；以及根据压缩矩阵来生成第二拼接矩阵；其中，多棵树中的每棵树所生成的预测结果矩阵对应压缩矩阵的一行，压缩矩阵的同一行中的每一列记录与该列相对应的样本在与该行相对应的预测结果矩阵中的预测结果。

[0016] 根据本公开的第三方面，提供了一种存储有计算机程序的计算机可读存储介质，

其中，计算机程序在由处理器执行时实现根据本公开的第一方面所述的方法的步骤。

[0017] 根据本公开的第四方面，提供了一种使用联邦学习模型进行人群包统计的方法。联邦学习模型包括多棵树。参与联邦学习的第一参与方拥有多棵树中的每棵树的叶子节点的权重矩阵。参与联邦学习的第二参与方拥有多棵树中的每棵树针对人群包生成的预测结果矩阵。该方法由第二参与方执行。该方法包括：将多棵树的预测结果矩阵进行按列拼接以生成第三拼接矩阵；对第三拼接矩阵执行按列乱序操作以生成第二拼接矩阵；以及向第一参与方提供第二拼接矩阵的相关信息，以便第一参与方根据第一拼接矩阵和第二拼接矩阵来确定人群包的统计信息；其中，第一拼接矩阵由第一参与方通过将多棵树的权重矩阵拼接来生成。

[0018] 在本公开的一些实施例中，向第一参与方提供第二拼接矩阵的相关信息包括：向第一参与方直接提供第二拼接矩阵。

[0019] 在本公开的一些实施例中，向第一参与方提供第二拼接矩阵的相关信息包括：根据第二拼接矩阵生成压缩矩阵；以及向第一参与方发送压缩矩阵；其中，多棵树中的每棵树所生成的预测结果矩阵对应压缩矩阵的一行，压缩矩阵的同一行中的每一列记录与该列相对应的样本在与该行相对应的预测结果矩阵中的预测结果。

[0020] 在本公开的一些实施例中，第二参与方通过以下操作来拥有每棵树的预测结果矩阵：从第一参与方接收针对该棵树的第一样本索引，第一样本索引由第一参与方根据该棵树的第一节点分裂条件推理获得，第一样本索引指示人群包中的样本与该棵树的叶子节点的第一预测关系；根据该棵树的第二节点分裂条件推理获得第二样本索引，第二样本索引指示人群包中的样本与该棵树的叶子节点的第二预测关系；对第一样本索引和第二样本索引求交集以获得预测样本索引；以及将预测样本索引转换成矩阵形式以获得该棵树的预测结果矩阵。

[0021] 在本公开的一些实施例中，第二参与方通过以下操作来拥有每棵树的预测结果矩阵：获得第一参与方生成的第一样本索引的第一碎片矩阵，第一样本索引由第一参与方根据该棵树的第一节点分裂条件推理获得，第一样本索引指示人群包中的样本与该棵树的叶子节点的第一预测关系，第一样本索引被转换成第一样本索引矩阵，第一样本索引矩阵被碎片化成第一碎片矩阵和第二碎片矩阵；根据该棵树的第二节点分裂条件推理获得第二样本索引，第二样本索引指示人群包中的样本与该棵树的叶子节点的第二预测关系；将第二样本索引转换成矩阵形式以获得第二样本索引矩阵；将第二样本索引矩阵碎片化成第三碎片矩阵和第四碎片矩阵；获得第一参与方根据第二碎片矩阵和第三碎片矩阵生成的第一中间碎片矩阵和第二中间碎片矩阵，其中，第三碎片矩阵由第二参与方发送给第一参与方；根据第一碎片矩阵和第四碎片矩阵生成第三中间碎片矩阵和第四中间碎片矩阵；向第一参与方发送第三中间碎片矩阵和第四中间碎片矩阵；获得第一参与方根据第一中间碎片矩阵、第二中间碎片矩阵、第三中间碎片矩阵和第四中间碎片矩阵生成的第一交集碎片矩阵；根据第一中间碎片矩阵、第二中间碎片矩阵、第三中间碎片矩阵和第四中间碎片矩阵生成第二交集碎片矩阵；以及将第一交集碎片矩阵与第二交集碎片矩阵相加以获得该棵树的预测结果矩阵。

[0022] 在本公开的一些实施例中，第二参与方通过以下操作来拥有每棵树的预测结果矩阵：获得第一参与方生成的第一样本索引的第一碎片矩阵，第一样本索引由第一参与方根

据该棵树的第一节点分裂条件推理获得，第一样本索引指示人群包中的样本与该棵树的叶子节点的第一预测关系，第一样本索引被转换成第一样本索引矩阵，第一样本索引矩阵被碎片化成第一碎片矩阵和第二碎片矩阵；根据该棵树的第二节点分裂条件推理获得第二样本索引，第二样本索引指示人群包中的样本与该棵树的叶子节点的第二预测关系；将第二样本索引转换成矩阵形式以获得第二样本索引矩阵；将第二样本索引矩阵碎片化成第三碎片矩阵和第四碎片矩阵；获得第一参与方根据第二碎片矩阵和第三碎片矩阵生成的第一中间碎片矩阵和第二中间碎片矩阵，其中，第三碎片矩阵由第一参与方生成；根据第一碎片矩阵和第四碎片矩阵生成第三中间碎片矩阵和第四中间碎片矩阵；向第一参与方发送第三中间碎片矩阵和第四中间碎片矩阵；获得第一参与方根据第一中间碎片矩阵、第二中间碎片矩阵、第三中间碎片矩阵和第四中间碎片矩阵生成的第一交集碎片矩阵；根据第一中间碎片矩阵、第二中间碎片矩阵、第三中间碎片矩阵和第四中间碎片矩阵生成第二交集碎片矩阵；以及将第一交集碎片矩阵与第二交集碎片矩阵相加以获得该棵树的预测结果矩阵。

[0023] 根据本公开的第五方面，提供了一种使用联邦学习模型进行人群包统计的装置。联邦学习模型包括多棵树。参与联邦学习的第一参与方拥有多棵树中的每棵树的叶子节点的权重矩阵。参与联邦学习的第二参与方拥有多棵树中的每棵树针对人群包生成的预测结果矩阵。该装置作为第二参与方。该装置包括至少一个处理器；以及存储有计算机程序的至少一个存储器。当计算机程序由至少一个处理器执行时，使得装置：将多棵树的预测结果矩阵进行按列拼接以生成第三拼接矩阵；对第三拼接矩阵执行按列乱序操作以生成第二拼接矩阵；以及向第一参与方提供第二拼接矩阵的相关信息，以便第一参与方根据第一拼接矩阵和第二拼接矩阵来确定人群包的统计信息；其中，第一拼接矩阵由第一参与方通过将多棵树的权重矩阵拼接来生成。

[0024] 在本公开的一些实施例中，计算机程序在由至少一个处理器执行时使得装置通过以下操作来向第一参与方提供第二拼接矩阵的相关信息：向第一参与方直接提供第二拼接矩阵。

[0025] 在本公开的一些实施例中，计算机程序在由至少一个处理器执行时使得装置通过以下操作来向第一参与方提供第二拼接矩阵的相关信息：根据第二拼接矩阵生成压缩矩阵；以及向第一参与方发送压缩矩阵；其中，多棵树中的每棵树所生成的预测结果矩阵对应压缩矩阵的一行，压缩矩阵的同一行中的每一列记录与该列相对应的样本在与该行相对应的预测结果矩阵中的预测结果。

[0026] 在本公开的一些实施例中，计算机程序在由至少一个处理器执行时使得装置通过以下操作来拥有每棵树的预测结果矩阵：从第一参与方接收针对该棵树的第一样本索引，第一样本索引由第一参与方根据该棵树的第一节点分裂条件推理获得，第一样本索引指示人群包中的样本与该棵树的叶子节点的第一预测关系；根据该棵树的第二节点分裂条件推理获得第二样本索引，第二样本索引指示人群包中的样本与该棵树的叶子节点的第二预测关系；对第一样本索引和第二样本索引求交集以获得预测样本索引；以及将预测样本索引转换成矩阵形式以获得该棵树的预测结果矩阵。

[0027] 在本公开的一些实施例中，计算机程序在由至少一个处理器执行时使得装置通过以下操作来拥有每棵树的预测结果矩阵：获得第一参与方生成的第一样本索引的第一碎片矩阵，第一样本索引由第一参与方根据该棵树的第一节点分裂条件推理获得，第一样本索

引指示人群包中的样本与该棵树的叶子节点的第一预测关系，第一样本索引被转换成第一样本索引矩阵，第一样本索引矩阵被碎片化成第一碎片矩阵和第二碎片矩阵；根据该棵树的第二节点分裂条件推理获得第二样本索引，第二样本索引指示人群包中的样本与该棵树的叶子节点的第二预测关系；将第二样本索引转换成矩阵形式以获得第二样本索引矩阵；将第二样本索引矩阵碎片化成第三碎片矩阵和第四碎片矩阵；获得第一参与方根据第二碎片矩阵和第三碎片矩阵生成的第一中间碎片矩阵和第二中间碎片矩阵，其中，第三碎片矩阵由第二参与方发送给第一参与方；根据第一碎片矩阵和第四碎片矩阵生成第三中间碎片矩阵和第四中间碎片矩阵；向第一参与方发送第三中间碎片矩阵和第四中间碎片矩阵；获得第一参与方根据第一中间碎片矩阵、第二中间碎片矩阵、第三中间碎片矩阵和第四中间碎片矩阵生成的第一交集碎片矩阵；根据第一中间碎片矩阵、第二中间碎片矩阵、第三中间碎片矩阵和第四中间碎片矩阵生成第二交集碎片矩阵；以及将第一交集碎片矩阵与第二交集碎片矩阵相加以获得该棵树的预测结果矩阵。

[0028] 在本公开的一些实施例中，计算机程序在由至少一个处理器执行时使得装置通过以下操作来拥有每棵树的预测结果矩阵：获得第一参与方生成的第一样本索引的第一碎片矩阵，第一样本索引由第一参与方根据该棵树的第一节点分裂条件推理获得，第一样本索引指示人群包中的样本与该棵树的叶子节点的第一预测关系，第一样本索引被转换成第一样本索引矩阵，第一样本索引矩阵被碎片化成第一碎片矩阵和第二碎片矩阵；根据该棵树的第二节点分裂条件推理获得第二样本索引，第二样本索引指示人群包中的样本与该棵树的叶子节点的第二预测关系；将第二样本索引转换成矩阵形式以获得第二样本索引矩阵；将第二样本索引矩阵碎片化成第三碎片矩阵和第四碎片矩阵；获得第一参与方根据第二碎片矩阵和第三碎片矩阵生成的第一中间碎片矩阵和第二中间碎片矩阵，其中，第三碎片矩阵由第一参与方生成；根据第一碎片矩阵和第四碎片矩阵生成第三中间碎片矩阵和第四中间碎片矩阵；向第一参与方发送第三中间碎片矩阵和第四中间碎片矩阵；获得第一参与方根据第一中间碎片矩阵、第二中间碎片矩阵、第三中间碎片矩阵和第四中间碎片矩阵生成的第一交集碎片矩阵；根据第一中间碎片矩阵、第二中间碎片矩阵、第三中间碎片矩阵和第四中间碎片矩阵生成第二交集碎片矩阵；以及将第一交集碎片矩阵与第二交集碎片矩阵相加以获得该棵树的预测结果矩阵。

[0029] 根据本公开的第六方面，提供了一种存储有计算机程序的计算机可读存储介质，其中，计算机程序在由处理器执行时实现根据本公开的第四方面所述的方法的步骤。

附图说明

[0030] 为了更清楚地说明本公开的实施例的技术方案，下面将对实施例的附图进行简要说明，应当知道，以下描述的附图仅仅涉及本公开的一些实施例，而非对本公开的限制，其中：

[0031] 图1是根据本公开的实施例的联邦学习模型在第一参与方与第二参与方处的示例性存储结构图；

[0032] 图2是根据本公开的实施例的使用联邦学习模型进行人群包统计的过程的示意性组合流程图和信令方案；

[0033] 图3是根据本公开的实施例的使用联邦学习模型进行人群包统计的过程的另一示

意性组合流程图和信令方案；

[0034] 图4是根据本公开的实施例的生成单棵树的预测结果矩阵的示意性组合流程图和信令方案；

[0035] 图5是根据本公开的实施例的生成单棵树的预测结果矩阵的另一示意性组合流程图和信令方案；

[0036] 图6是根据本公开的实施例的由第一参与方执行的使用联邦学习模型进行人群包统计的方法的示意性流程图；

[0037] 图7是根据本公开的实施例的由第二参与方执行的使用联邦学习模型进行人群包统计的方法的示意性流程图；

[0038] 图8是根据本公开的实施例的作为第一参与方的使用联邦学习模型进行人群包统计的装置的示意性框图；以及

[0039] 图9是根据本公开的实施例的作为第二参与方的使用联邦学习模型进行人群包统计的装置的示意性框图。

[0040] 需要注意的是，附图中的元素是示意性的，没有按比例绘制。

具体实施方式

[0041] 为了使本公开的实施例的目的、技术方案和优点更加清楚，下面将结合附图，对本公开的实施例的技术方案进行清楚、完整的描述。显然，所描述的实施例是本公开的一部分实施例，而不是全部的实施例。基于所描述的本公开的实施例，本领域技术人员在无需创造性劳动的前提下所获得的所有其它实施例，也都属于本公开保护的范围。

[0042] 除非另外定义，否则在此使用的所有术语(包括技术和科学术语)具有与本公开主题所属领域的技术人员所通常理解的相同含义。进一步将理解的是，诸如在通常使用的词典中定义的那些的术语应解释为具有与说明书上下文和相关技术中它们的含义一致的含义，并且将不以理想化或过于正式的形式来解释，除非在此另外明确定义。另外，诸如“第一”和“第二”的术语仅用于将一个部件(或部件的一部分)与另一个部件(或部件的另一部分)区分开。

[0043] 图1示出根据本公开的实施例的联邦学习模型在第一参与方与第二参与方处的示例性存储结构图。在图1的示例中，第一参与方GUEST为标签拥有方，第二参与方HOST为数据合作方。一般而言，联邦学习中数据合作方的数量可以有多个，但标签拥有方的数量为一个。在图1中以两个参与方为例来进行说明。图1中的第一参与方GUEST和第二参与方HOST拥有完整的模型节点关系结构。联邦学习模型可包括多棵树(即，多个树模型)。在图1中以一棵树为例来进行说明。第一参与方GUEST拥有该棵树的非叶子节点N1和所有叶子节点N4、N5、N6和N7的信息，不拥有该棵树的非叶子节点N2和N3的信息。第二参与方HOST拥有该棵树的非叶子节点N2和N3的信息，不拥有该棵树的非叶子节点N1和任何叶子节点N4、N5、N6和N7的信息。

[0044] 假设有人群包中有四个样本a、b、c和d。四个样本a、b、c和d被分别输入第一参与方GUEST的树模型和第二参与方HOST的树模型，并在每个树模型上都进行了路径推理。在图1的示例中，第一参与方GUEST的预测结果是：叶子节点N4有样本a和c，叶子节点N5有样本a和c，叶子节点N6有样本b和d，叶子节点N7有样本b和d。第二参与方HOST的预测结果是：叶子节

点N4有样本a、b、c和d，叶子节点N5没有样本，叶子节点N6有样本a和b，叶子节点N7有样本c和d。

[0045] 第一参与方GUEST拥有的叶子节点N4、N5、N6和N7的信息包括：叶子节点N4、N5、N6和N7的权重。叶子节点N4、N5、N6和N7的权重可按照叶子节点的编号顺序来组成权重矩阵。第二参与方HOST拥有联邦学习模型针对人群包中的样本生成的预测结果矩阵。预测结果矩阵根据第一参与方GUEST的预测结果与第二参与方HOST的预测结果的交集来生成。在本公开的实施例中，预测结果矩阵的每一行对应一个叶子节点，预测结果矩阵的每一列对应一个样本。

[0046] 在本文中以XGBoost模型为例来进行说明。本领域技术人员应理解图1中的存储结构只是示例性的，本公开的实施例不限制联邦学习模型在各参与方处的存储结构。

[0047] 图2示出根据本公开的实施例的使用联邦学习模型进行人群包统计的过程的示意性组合流程图和信令方案。为便于描述，下文以联邦学习模型包括3棵树为例来进行说明。本领域技术人员应理解，联邦学习模型中树的数量也可以为其他值。

[0048] 第一参与方GUEST拥有每棵树的叶子节点的权重矩阵。其中，第一棵树的叶子节点的权重矩阵被表示为[w1 w2 w3 w4]。其中，w1表示第一棵树的叶子节点N4的权重。w2表示第一棵树的叶子节点N5的权重。w3表示第一棵树的叶子节点N6的权重。w4表示第一棵树的叶子节点N7的权重。类似地，第二棵树的叶子节点的权重矩阵被表示为[v1 v2 v3 v4]。其中，v1表示第二棵树的叶子节点N4的权重。v2表示第二棵树的叶子节点N5的权重。v3表示第二棵树的叶子节点N6的权重。v4表示第二棵树的叶子节点N7的权重。第三棵树的叶子节点的权重矩阵被表示为[u1 u2 u3 u4]。其中，u1表示第三棵树的叶子节点N4的权重。u2表示第三棵树的叶子节点N5的权重。u3表示第三棵树的叶子节点N6的权重。u4表示第三棵树的叶子节点N7的权重。在二分类场景下，每个叶子节点的权重为一个值。在多分类场景下，每个叶子节点的权重为K个值，其中K为类别数量。

[0049] 第二参与方HOST拥有每棵树的预测结果矩阵。其中，第一棵树的叶子节点的预测结果矩阵被表示为C1。第二棵树的叶子节点的预测结果矩阵被表示为C2。第三棵树的叶子节点的预测结果矩阵被表示为C3。其中，预测结果矩阵C1、C2和C3的第一行对应叶子节点N4。预测结果矩阵C1、C2和C3的第二行对应叶子节点N5。预测结果矩阵C1、C2和C3的第三行对应叶子节点N6。预测结果矩阵C1、C2和C3的第四行对应叶子节点N7。预测结果矩阵C1、C2和C3的第一列对应样本a。预测结果矩阵C1、C2和C3的第二列对应样本b。预测结果矩阵C1、C2和C3的第三列对应样本c。预测结果矩阵C1、C2和C3的第四列对应样本d。预测结果矩阵C1的第一行第一列的元素“1”表示叶子节点N4有样本a。预测结果矩阵C1的第一行第二列的元素“0”表示叶子节点N4没有样本b。预测结果矩阵C1的第二行第一列的元素“0”表示叶子节点N5没有样本a。以此类推。

[0050] 第二参与方HOST在动作201处将多棵树的预测结果矩阵进行按列拼接以生成第三拼接矩阵MC。在图2的示例中，第三拼接矩阵MC按照C1-C2-C3的顺序在列的方向上拼接而成。在图2的替代示例中，第三拼接矩阵MC也可以按照其他顺序在列的方向上拼接而成。为了使得个体预测结果不被泄露，第二参与方HOST在动作202处对第三拼接矩阵MC执行按列乱序操作以生成第二拼接矩阵SMC。在第二拼接矩阵SMC中，预测结果矩阵C1被按列乱序成预测结果矩阵C1’，预测结果矩阵C2被按列乱序成预测结果矩阵C2’，预测结果矩阵C3被按

列乱序成预测结果矩阵C3’。

[0051] 然后，第二参与方HOST在动作204处向第一参与方GUEST发送第二拼接矩阵SMC。第一参与方GUEST接收到的第二拼接矩阵SMC经过了按列乱序处理，因此，第一参与方GUEST不能够根据列号来确定该列对应哪个样本，所以无法定位个体预测结果。

[0052] 第一参与方GUEST在动作206处对多棵树的叶子节点的权重矩阵进行拼接，以获得第一拼接矩阵WVU=[w1 w2 w3 w4 v1 v2 v3 v4 u1 u2 u3 u4]。在这里，权重矩阵的拼接顺序应与预测结果矩阵在动作201处的拼接顺序相同。例如，都是按照第一棵树-第二棵树-第三棵树的顺序进行拼接。拼接顺序可以是默认值，也可以是第一参与方GUEST和第二参与方HOST预先商议好的。

[0053] 动作206可与动作201、动作202或动作204并行地执行，也可以先于动作201、动作202和动作204中的任一个执行。

[0054] 第一参与方GUEST在动作207处将第一拼接矩阵WVU与第二拼接矩阵SMC进行矩阵相乘以获得预测概率矩阵P=[p1 p2 p3 p4]。在二分类场景下，每个叶子节点的权重为一个值，因此预测概率矩阵P的每一列包括一个值。在多分类场景下，每个叶子节点的权重为K个值，因此预测概率矩阵P的每一列包括K个值。

[0055] 第一参与方GUEST在动作208处根据预测概率矩阵P来生成预测类别矩阵R=[r1 r2 r3 r4]。预测类别矩阵R指示人群包中的每个样本的预测类别。

[0056] 在二分类场景下，确定预测概率矩阵R中针对每个样本的预测概率是否超过预设的概率阈值。在图2的示例中，分别确定p1、p2、p3和p4是否超过预设的概率阈值。如果任一样本的预测概率高于概率阈值，确定该样本的预测类别为第一类别。如果任一样本的预测概率低于或者等于概率阈值，确定该样本的预测类别为第二类别。例如，如果p1高于概率阈值，则r1=1。如果p2低于或者等于概率阈值，则r2=0。其中，用1代表第一类别，用0代表第二类别。

[0057] 在多分类场景下，确定预测概率矩阵中针对每个样本的多个预测概率中的最大预测概率。其中，多个预测概率中的每个预测概率对应一个类别。针对每个样本，确定该样本的预测类别为针对该样本的最大预测概率所对应的类别。假设样本a属于第一分类的概率为0.3，样本a属于第二分类的概率为0.6，样本a属于第三分类的概率为0.1，则样本a的最大预测概率为0.6，因此样本a的预测类别为第二分类。r1=2。其中，用1代表第一类别，用2代表第二类别，用3代表第三类别。

[0058] 第一参与方GUEST在动作209处按照预测类别矩阵R所指示的预测类别对预测概率矩阵P中的预测概率进行聚合。例如，将对应第一类别的预测概率归入第一集合，将对应第二类别的预测概率归入第二集合，以此类推。

[0059] 第一参与方GUEST在动作210处统计每个预测类别中的样本数量和预测概率均值。每个预测类别中的样本数量等于该预测类别对应的集合中的预测概率的数量。该集合中的预测概率均值等于该集合中的所有预测概率之和除以该集合中的预测概率的数量。

[0060] 第一参与方GUEST在动作211处根据每个预测类别中的样本数量和预测概率均值来确定人群包的统计标签。例如，在二分类场景下，假设需要标记人群包是否会贷款逾期，则可根据预测类别为贷款逾期的集合中的样本数量和预测概率均值来指示人群包是否会贷款逾期。在多分类场景下，假设需要标记人群包针对啤酒、滑板和书籍的偏好，根据这三

个预测类别的集合中的样本数量和预测概率均值来指示人群包的偏好，从而例如进行定向营销。

[0061] 在图2的示例的替代实施例中，第二参与方HOST可向第一参与方GUEST发送第二拼接矩阵SMC的压缩矩阵，从而减小与第一参与方GUEST进行通信的数据量，提高通讯效率。图3示出这种情况下使用联邦学习模型进行人群包统计的过程的示意性组合流程图和信令方案。在图2的示例的基础上，第二参与方HOST在动作303处根据第二拼接矩阵SMC生成压缩矩阵CMC。其中，多棵树中的每棵树所生成的预测结果矩阵对应压缩矩阵CMC的一行(不同行)。压缩矩阵CMC的同一行中的每一列记录与该列相对应的样本在与该行相对应的预测结果矩阵中的预测结果。例如，压缩矩阵CMC的第一行对应预测结果矩阵C1'。压缩矩阵CMC的第一行第一列的元素“0”指示预测结果矩阵C1'的第一列中的“1”位于第一行(行号为“0”)。压缩矩阵CMC的第一行第二列的元素“3”指示预测结果矩阵C1'的第二列中的“1”位于第四行(行号为“3”)。压缩矩阵CMC的第一行第三列的元素“0”指示预测结果矩阵C1'的第三列中的“1”位于第一行(行号为“0”)。压缩矩阵CMC的第一行第四列的元素“2”指示预测结果矩阵C1'的第四列中的“1”位于第三行(行号为“2”)。类似地，压缩矩阵CMC的第二行对应预测结果矩阵C2'。压缩矩阵CMC的第三行对应预测结果矩阵C3'。

[0062] 第二参与方HOST在动作204处向第一参与方GUEST发送压缩矩阵CMC。

[0063] 第一参与方GUEST在动作305处对压缩矩阵CMC进行膨胀操作以生成第二拼接矩阵SMC。第一参与方GUEST预先知道第二参与方HOST的压缩规则，因此，可使用该压缩规则的逆规则来从压缩矩阵CMC复原第二拼接矩阵SMC。例如，压缩矩阵CMC的第一行第一列的元素“0”指示预测结果矩阵C1'的第一列中的“1”位于第一行，因此可将预测结果矩阵C1'的第一列恢复为$[1\ 0\ 0\ 0]^T$。以此类推，可完全恢复出第二拼接矩阵SMC。

[0064] 本公开的实施例能够分别在基于低带宽和基于MPC(Multi-Party Computation)高带宽的应用场景下使用联邦学习模型来进行联合预测。基于低带宽的预测方案有较高的计算性能，在半诚实场景下，可安全运行。基于MPC高带宽的预测方案则有更强的安全性保障。图4示出基于低带宽的预测方案。图5示出基于MPC高带宽的预测方案。为便于描述，图4和图5均以单棵树为例来进行说明。在下文中以“目标树”来指代该棵树。

[0065] 在图4的示例中，第一参与方GUEST在动作441处根据目标树的第一节点分裂条件推理生成第一样本索引[[a, c][a, c] [b, d][b, d]]。第一样本索引指示样本a、b、c和d与目标树的叶子节点N4、N5、N6和N7的第一预测关系。在图4的示例中，第一节点分裂条件由目标树的非叶子节点N1和所有叶子节点N4、N5、N6和N7的节点分裂条件来组成。第一预测关系指示：目标树的叶子节点N4有样本a和c，目标树的叶子节点N5有样本a和c，目标树的叶子节点N6有样本b和d，目标树的叶子节点N7有样本b和d。

[0066] 第一参与方GUEST在动作442处向第二参与方HOST发送第一样本索引[[a, c] [a, c][b, d][b, d]]。

[0067] 第二参与方HOST在动作443处根据目标树的第二节点分裂条件推理获得第二样本索引[[a, b, c, d] [ ][a, b] [c, d]]。第二样本索引指示样本a、b、c和d与目标树的叶子节点N4、N5、N6和N7的第二预测关系。在图4的示例中，第二节点分裂条件由目标树的非叶子节点N2和N3的节点分裂条件来组成。第二预测关系指示：目标树的叶子节点N4有样本a、b、c和d，目标树的叶子节点N5没有样本，目标树的叶子节点N6有样本a和b，目标树的叶子节

点N7有样本c和d。

[0068] 动作443可与动作441或动作442并行地执行，也可以在动作441或动作442之前执行。

[0069] 第二参与方HOST在动作444处对第一样本索引和第二样本索引求交集以获得预测样本索引[[a, c] [ ][b] [d]]。然后，第二参与方HOST在动作445处将预测样本索引[[a, c][ ] [b][d]]转换成矩阵形式以获得目标树的预测结果矩阵C。在本公开的实施例中，预测结果矩阵C的每一行对应一个叶子节点，预测结果矩阵的每一列对应一个样本标签。在图4的示例中，预测结果矩阵C表示：叶子节点N4有样本a和c(第一行对应叶子节点N4，第一行的第一列和第三列为1，其余列为0)，叶子节点N5没有样本(第二行对应叶子节点N5，第二行的每列都为0)，叶子节点N6有样本b(第三行对应叶子节点N6，第三行的第二列为1，其余列为0)，叶子节点N7有样本d(第四行对应叶子节点N7，第四行的第四列为1，其余列为0)。

[0070] 在图5的示例中，第一参与方GUEST根据目标树的第一节点分裂条件推理生成第一样本索引[[a, c] [a, c][b, d] [b, d]]并在动作551处将第一样本索引[[a, c][a, c][b, d] [b, d]]转换成矩阵形式，以获得第一样本索引矩阵P。

[0071] 第二参与方HOST根据目标树的第二节点分裂条件推理生成第二样本索引[[a, b, c, d][ ] [a, b][c, d]]并在动作552处将第二样本索引[[a, b, c, d] [ ][a, b] [c, d]]转换成矩阵形式，以获得第二样本索引矩阵Q。

[0072] 第一参与方GUEST在动作553处将第一样本索引矩阵P碎片化成第一碎片矩阵p2和第二碎片矩阵p1。例如，可随机生成第一碎片矩阵p2，然后根据p1=P-p2来计算p1。

[0073] 第二参与方HOST在动作554处将第二样本索引矩阵Q碎片化成第三碎片矩阵q1和第四碎片矩阵q2。例如，可随机生成第三碎片矩阵q1，然后根据q2=Q-q1来计算q2。

[0074] 动作553可与动作552或动作554并行地执行，也可以在动作552或动作554之前执行。动作554可与动作551或动作553并行地执行，也可以在动作551或动作553之前执行。

[0075] 在动作555处，第一参与方GUEST与第二参与方HOST共享第一碎片矩阵p2，第二参与方HOST与第一参与方GUEST共享第三碎片矩阵q1。

[0076] 第一参与方GUEST在动作556处根据第二碎片矩阵p1和第三碎片矩阵q1生成第一中间碎片矩阵f1和第二中间碎片矩阵e1。在本公开的一些实施例中，第一参与方GUEST可预先生成三元组碎片矩阵<a1, b1, c1>。第一参与方GUEST可根据第二碎片矩阵p1、第三碎片矩阵q1和三元组碎片矩阵<a1, b1, c1>来生成第一中间碎片矩阵f1和第二中间碎片矩阵e1。其中，f1= p1-a1，e1=q1-b1。

[0077] 第二参与方HOST在动作557处根据第一碎片矩阵p2和第四碎片矩阵q2生成第三中间碎片矩阵f2和第四中间碎片矩阵e2。在本公开的一些实施例中，第二参与方HOST可预先生成三元组碎片矩阵<a2, b2, c2>。第二参与方HOST可根据第一碎片矩阵p2、第四碎片矩阵q2和三元组碎片矩阵<a2, b2, c2>来生成第三中间碎片矩阵f2和第四中间碎片矩阵e2。其中，f2= p2-a2，e2=q2-b2。(a1 + a2) × (b1 + b2) = (c1 + c2)。

[0078] 在动作558处，第一参与方GUEST与第二参与方HOST共享(向第二参与方HOST发送)第一中间碎片矩阵f1和第二中间碎片矩阵e1，第二参与方HOST与第一参与方GUEST(向第一参与方GUEST发送)共享第三中间碎片矩阵f2和第四中间碎片矩阵e2。

[0079] 第一参与方GUEST在动作559处根据第一中间碎片矩阵f1、第二中间碎片矩阵e1、

第三中间碎片矩阵f2和第四中间碎片矩阵e2生成第一交集碎片矩阵z1。在一个示例中，z1=e×f+a1×f+b1×e+c1，其中，f=f1+f2，e=e1+e2。

[0080] 第二参与方HOST在动作560处根据第一中间碎片矩阵f1、第二中间碎片矩阵e1、第三中间碎片矩阵f2和第四中间碎片矩阵e2生成第二交集碎片矩阵z2。在一个示例中，z2=a2×f+b2×e+c2，其中，f=f1+f2，e=e1+e2。

[0081] 第一参与方GUEST在动作561处向第二参与方HOST发送第一交集碎片矩阵z1。第二参与方HOST在动作562处将第一交集碎片矩阵z1与第二交集碎片矩阵z2相加以获得目标树的预测结果矩阵C。

[0082] 通过对第一参与方GUEST和第二参与方HOST的预测结果执行碎片化操作，并只共享预测结果的一部分(碎片)，第一参与方GUEST和第二参与方HOST都不知道对方的预测结果，因此有更强的安全性保障。

[0083] 在图5的示例的替代实施例中，在动作555处，第一参与方GUEST与第二参与方HOST不共享第一碎片矩阵p2和第三碎片矩阵q1。第一参与方GUEST与第二参与方HOST可先执行DH密钥交换，然后共享随机种子。接着，第一参与方GUEST和第二参与方HOST根据共享的随机种子分别生成第三碎片矩阵q1和第一碎片矩阵p2。这样可以减少第一参与方GUEST与第二参与方HOST的数据交换量，从而节约网络资源。

[0084] 图6示出根据本公开的实施例的由第一参与方执行的使用联邦学习模型进行人群包统计的方法600的示意性流程图。联邦学习模型包括多棵树。参与联邦学习的第一参与方拥有多棵树中的每棵树的叶子节点的权重矩阵。参与联邦学习的第二参与方拥有多棵树中的每棵树针对人群包生成的预测结果矩阵。

[0085] 在框S602处，第一参与方将多棵树的权重矩阵拼接成第一拼接矩阵。

[0086] 在框S604处，第一参与方获得由第二参与方生成的第二拼接矩阵。第二拼接矩阵通过将多棵树的预测结果矩阵进行按列拼接并执行按列乱序操作来生成。在本公开的一些实施例中，第一参与方可从第二参与方直接接收第二拼接矩阵。在本公开的另一些实施例中，第一参与方接收由第二参与方根据第二拼接矩阵生成的压缩矩阵。然后，第一参与方根据压缩矩阵来生成第二拼接矩阵。其中，多棵树中的每棵树所生成的预测结果矩阵对应压缩矩阵的一行。压缩矩阵的同一行中的每一列记录与该列相对应的样本在与该行相对应的预测结果矩阵中的预测结果。

[0087] 在框S606处，第一参与方将第一拼接矩阵与第二拼接矩阵进行矩阵相乘以获得预测概率矩阵。

[0088] 在框S608处，第一参与方根据预测概率矩阵来确定人群包的统计信息。在本公开的一些实施例中，第一参与方根据预测概率矩阵来生成预测类别矩阵。预测类别矩阵指示人群包中的每个样本的预测类别。在二分类场景下，第一参与方确定预测概率矩阵中针对每个样本的预测概率是否超过预设的概率阈值。如果任一样本的预测概率高于概率阈值，第一参与方确定该样本的预测类别为第一类别。如果任一样本的预测概率低于或者等于概率阈值，第一参与方确定该样本的预测类别为第二类别。在多分类场景下，第一参与方确定预测概率矩阵中针对每个样本的多个预测概率中的最大预测概率。其中，多个预测概率中的每个预测概率对应一个类别。针对每个样本，第一参与方确定该样本的预测类别为针对该样本的最大预测概率所对应的类别。

[0089] 然后，第一参与方按照预测类别矩阵所指示的预测类别对预测概率矩阵中的预测概率进行聚合。第一参与方统计每个预测类别中的样本数量和预测概率均值。之后，第一参与方根据每个预测类别中的样本数量和预测概率均值来确定人群包的统计标签。

[0090] 图7示出根据本公开的实施例的由第二参与方执行的使用联邦学习模型进行人群包统计的方法700的示意性流程图。联邦学习模型包括多棵树。参与联邦学习的第一参与方拥有多棵树中的每棵树的叶子节点的权重矩阵。参与联邦学习的第二参与方拥有多棵树中的每棵树针对人群包生成的预测结果矩阵。

[0091] 在框S702处，第二参与方将多棵树的预测结果矩阵进行按列拼接以生成第三拼接矩阵。

[0092] 在框S704处，第二参与方对第三拼接矩阵执行按列乱序操作以生成第二拼接矩阵。

[0093] 在框S706处，第二参与方向第一参与方提供第二拼接矩阵的相关信息，以便第一参与方根据第一拼接矩阵和第二拼接矩阵来确定人群包的统计信息。其中，第一拼接矩阵由第一参与方通过将多棵树的权重矩阵拼接来生成。

[0094] 在本公开的一些实施例中，第二参与方向第一参与方直接提供第二拼接矩阵。第二拼接矩阵的相关信息指的是第二拼接矩阵本身。在本公开的另一些实施例中，第二参与方根据第二拼接矩阵生成压缩矩阵。然后第二参与方向第一参与方发送压缩矩阵。其中，多棵树中的每棵树所生成的预测结果矩阵对应压缩矩阵的一行。压缩矩阵的同一行中的每一列记录与该列相对应的样本在与该行相对应的预测结果矩阵中的预测结果。第二拼接矩阵的相关信息指的是第二拼接矩阵的压缩矩阵。

[0095] 图8示出根据本公开的实施例的作为第一参与方的使用联邦学习模型进行人群包统计的装置800的示意性框图。如图8所示，该装置800可包括处理器810和存储有计算机程序的存储器820。当计算机程序由处理器810执行时，使得装置800可执行如图6所示的方法600的步骤。在一个示例中，装置800可以是计算机设备或云计算节点等。装置800可将多棵树的权重矩阵拼接成第一拼接矩阵。装置800可获得由第二参与方生成的第二拼接矩阵。第二拼接矩阵通过将多棵树的预测结果矩阵进行按列拼接并执行按列乱序操作来生成。装置800可将第一拼接矩阵与第二拼接矩阵进行矩阵相乘以获得预测概率矩阵。装置800可根据预测概率矩阵来确定人群包的统计信息。

[0096] 在本公开的实施例中，处理器810可以是例如中央处理单元(CPU)、微处理器、数字信号处理器(DSP)、基于多核的处理器架构的处理器等。存储器820可以是使用数据存储技术实现的任何类型的存储器，包括但不限于随机存取存储器、只读存储器、基于半导体的存储器、闪存、磁盘存储器等。

[0097] 此外，在本公开的实施例中，装置800也可包括输入设备830，例如键盘、鼠标等，用于输入人群包。另外，装置800还可包括输出设备840，例如显示器等，用于输出人群包的统计信息。

[0098] 图9示出根据本公开的实施例的作为第二参与方的使用联邦学习模型进行人群包统计的装置900的示意性框图。如图9所示，该装置900可包括处理器910和存储有计算机程序的存储器920。当计算机程序由处理器910执行时，使得装置900可执行如图7所示的方法700的步骤。在一个示例中，装置900可以是计算机设备或云计算节点等。装置900可将多棵

树的预测结果矩阵进行按列拼接以生成第三拼接矩阵。装置900可对第三拼接矩阵执行按列乱序操作以生成第二拼接矩阵。装置900可向第一参与方提供第二拼接矩阵的相关信息，以便第一参与方根据第一拼接矩阵和第二拼接矩阵来确定人群包的统计信息。其中，第一拼接矩阵由第一参与方通过将多棵树的权重矩阵拼接来生成。

[0099] 在本公开的实施例中，处理器910可以是例如中央处理单元(CPU)、微处理器、数字信号处理器(DSP)、基于多核的处理器架构的处理器等。存储器920可以是使用数据存储技术实现的任何类型的存储器，包括但不限于随机存取存储器、只读存储器、基于半导体的存储器、闪存、磁盘存储器等。

[0100] 此外，在本公开的实施例中，装置900也可包括输入设备930，例如键盘、鼠标等，用于输入人群包。另外，装置900还可包括输出设备940，例如显示器等，用于输出第二拼接矩阵或压缩矩阵。

[0101] 在本公开的其它实施例中，还提供了一种存储有计算机程序的计算机可读存储介质，其中，计算机程序在由处理器执行时能够实现如图6至图7所示的方法的步骤。

[0102] 综上所述，根据本公开的实施例的使用联邦学习模型进行人群包统计的方法及装置能够在对人群包进行统计的时候避免个体预测结果泄露，满足合规需求。根据本公开的实施例的使用联邦学习模型进行人群包统计的方法及装置能够适用于不同带宽的应用场景。

[0103] 附图中的流程图和框图显示了根据本公开的多个实施例的装置和方法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、程序段或指令的一部分，所述模块、程序段或指令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。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，两个连续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。也要注意的是，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动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。

[0104] 除非上下文中另外明确地指出，否则在本文和所附权利要求中所使用的词语的单数形式包括复数，反之亦然。因而，当提及单数时，通常包括相应术语的复数。相似地，措辞“包含”和“包括”将解释为包含在内而不是独占性地。同样地，术语“包括”和“或”应当解释为包括在内的，除非本文中明确禁止这样的解释。在本文中使用术语“示例”之处，特别是当其位于一组术语之后时，所述“示例”仅仅是示例性的和阐述性的，且不应当被认为是独占性的或广泛性的。

[0105] 适应性的进一步的方面和范围从本文中提供的描述变得明显。应当理解，本申请的各个方面可以单独或者与一个或多个其它方面组合实施。还应当理解，本文中的描述和特定实施例旨在仅说明的目的并不旨在限制本申请的范围。

[0106] 以上对本公开的若干实施例进行了详细描述，但显然，本领域技术人员可以在不脱离本公开的精神和范围的情况下对本公开的实施例进行各种修改和变型。本公开的保护范围由所附的权利要求限定。

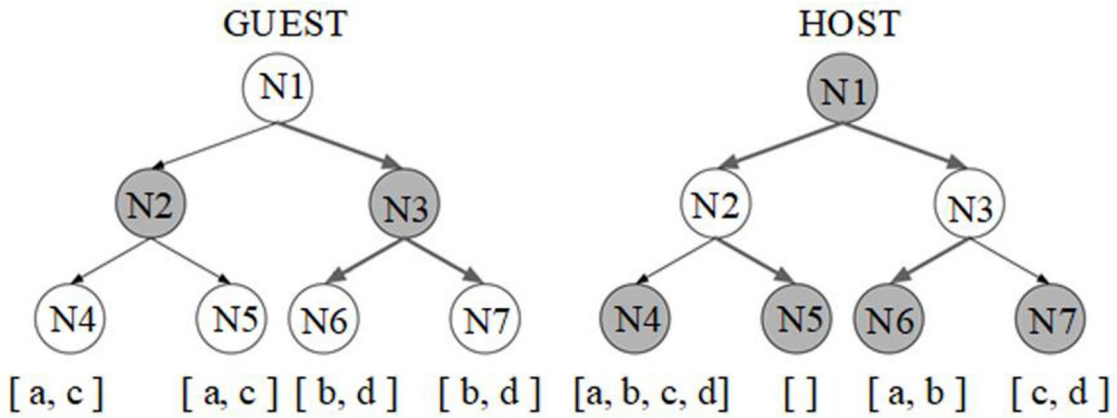

GUEST
N1
N2
N3
N4
N5
N6
N7
[a, c]
[a, c]
[b, d]
[b, d]
HOST
N1
N2
N3
N4
N5
N6
N7
[a, b, c, d]
[ ]
[a, b]
[c, d]

图 1

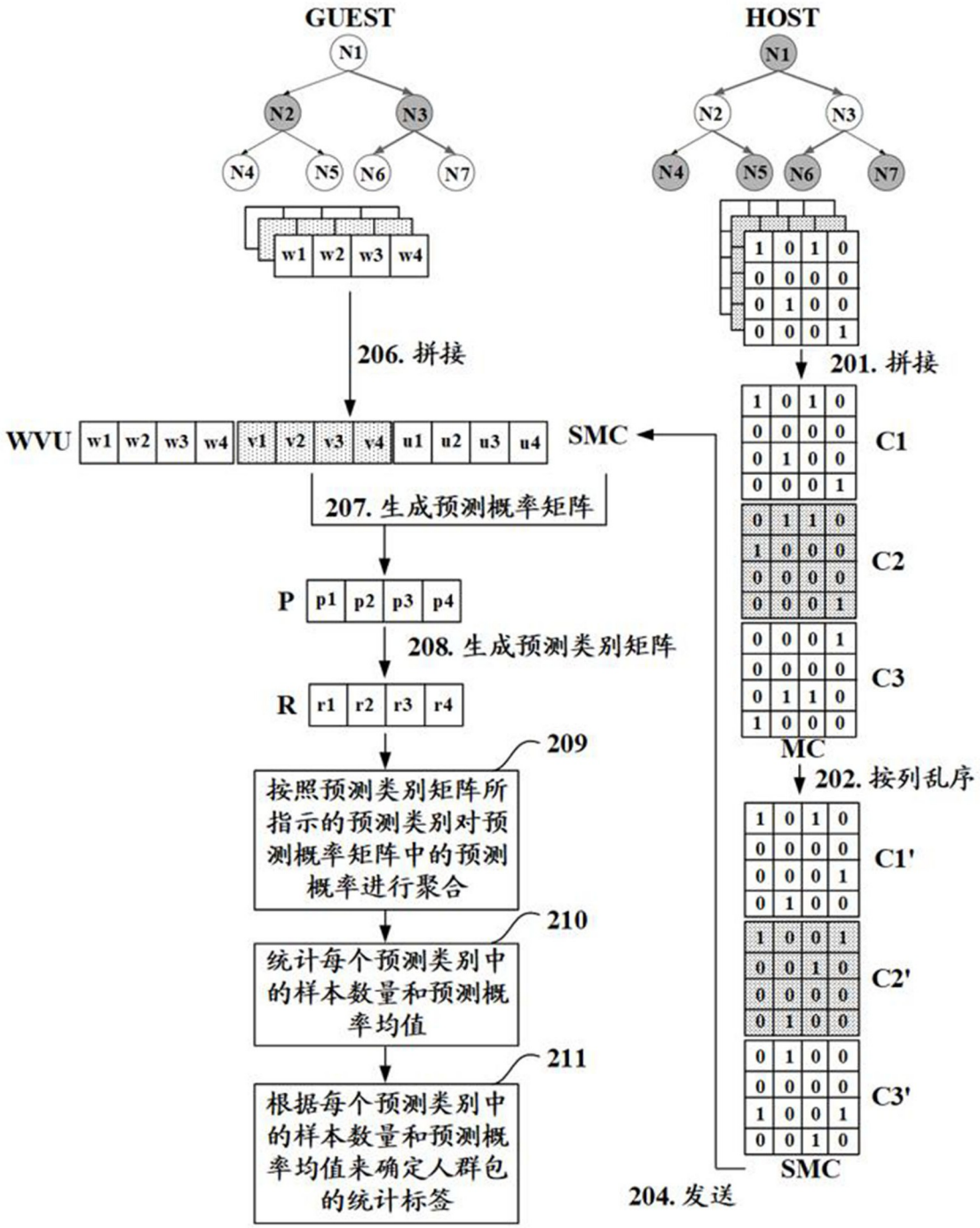

GUEST
HOST
N1
N2
N3
N4
N5
N6
N7
w1 w2 w3 w4
206. 拼接
WVU
w1 w2 w3 w4 v1 v2 v3 v4 u1 u2 u3 u4
SMC
207. 生成预测概率矩阵
P
p1 p2 p3 p4
208. 生成预测类别矩阵
R
r1 r2 r3 r4
209
按照预测类别矩阵所指示的预测类别对预测概率矩阵中的预测概率进行聚合
210
统计每个预测类别中的样本数量和预测概率均值
211
根据每个预测类别中的样本数量和预测概率均值来确定人群包的统计标签
201. 拼接
C1
C2
C3
MC
202. 按列乱序
C1'
C2'
C3'
SMC
204. 发送

图 2

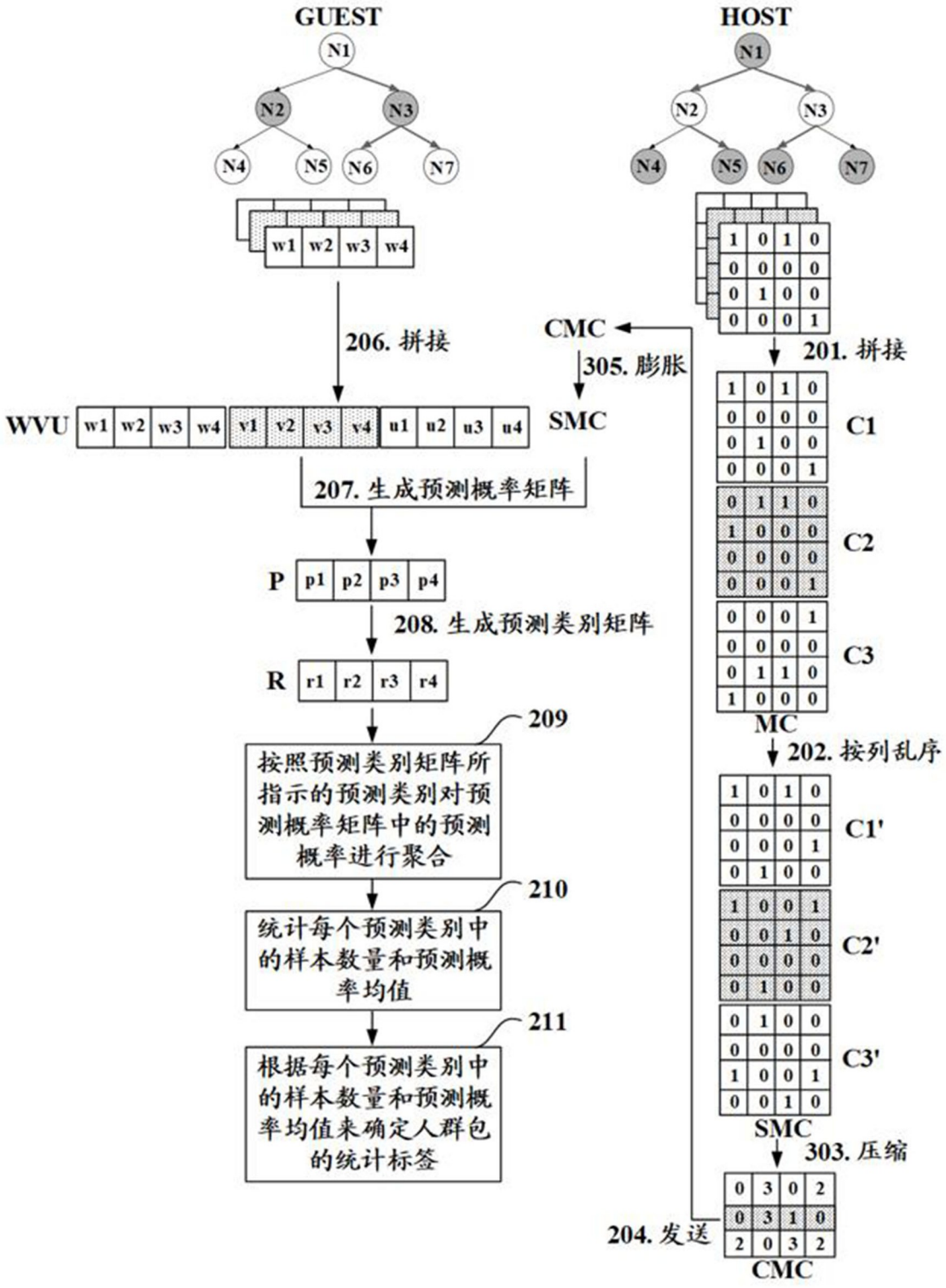

GUEST
N1
N2
N3
N4
N5
N6
N7
w1 w2 w3 w4
206. 拼接
WVU
w1 w2 w3 w4 v1 v2 v3 v4 u1 u2 u3 u4
207. 生成预测概率矩阵
P
p1 p2 p3 p4
208. 生成预测类别矩阵
R
r1 r2 r3 r4
209
按照预测类别矩阵所指示的预测类别对预测概率矩阵中的预测概率进行聚合
210
统计每个预测类别中的样本数量和预测概率均值
211
根据每个预测类别中的样本数量和预测概率均值来确定人群包的统计标签
HOST
CMC
305. 膨胀
SMC
204. 发送
201. 拼接
C1
C2
C3
MC
202. 按列乱序
C1'
C2'
C3'
SMC
303. 压缩
CMC

图 3

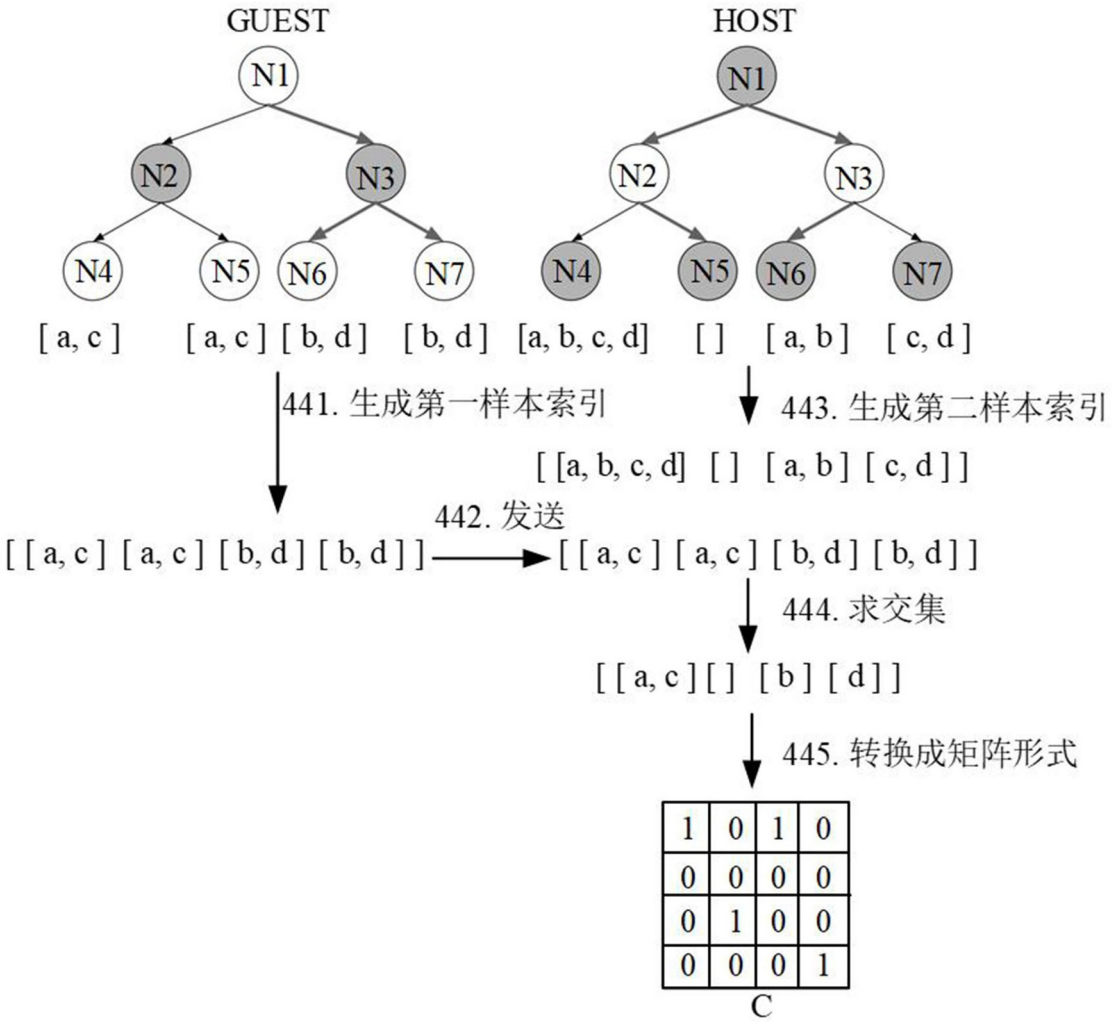

GUEST
HOST
N1
N2
N3
N4
N5
N6
N7
[a, c]
[a, c] [b, d]
[b, d]
[a, b, c, d]
[ ]
[a, b]
[c, d]
441. 生成第一样本索引
443. 生成第二样本索引
[[a, b, c, d] [ ] [a, b] [c, d]]
442. 发送
[[a, c] [a, c] [b, d] [b, d]]
[[a, c] [a, c] [b, d] [b, d]]
444. 求交集
[[a, c] [ ] [b] [d]]
445. 转换成矩阵形式
1 0 1 0
0 0 0 0
0 1 0 0
0 0 0 1
C

图 4

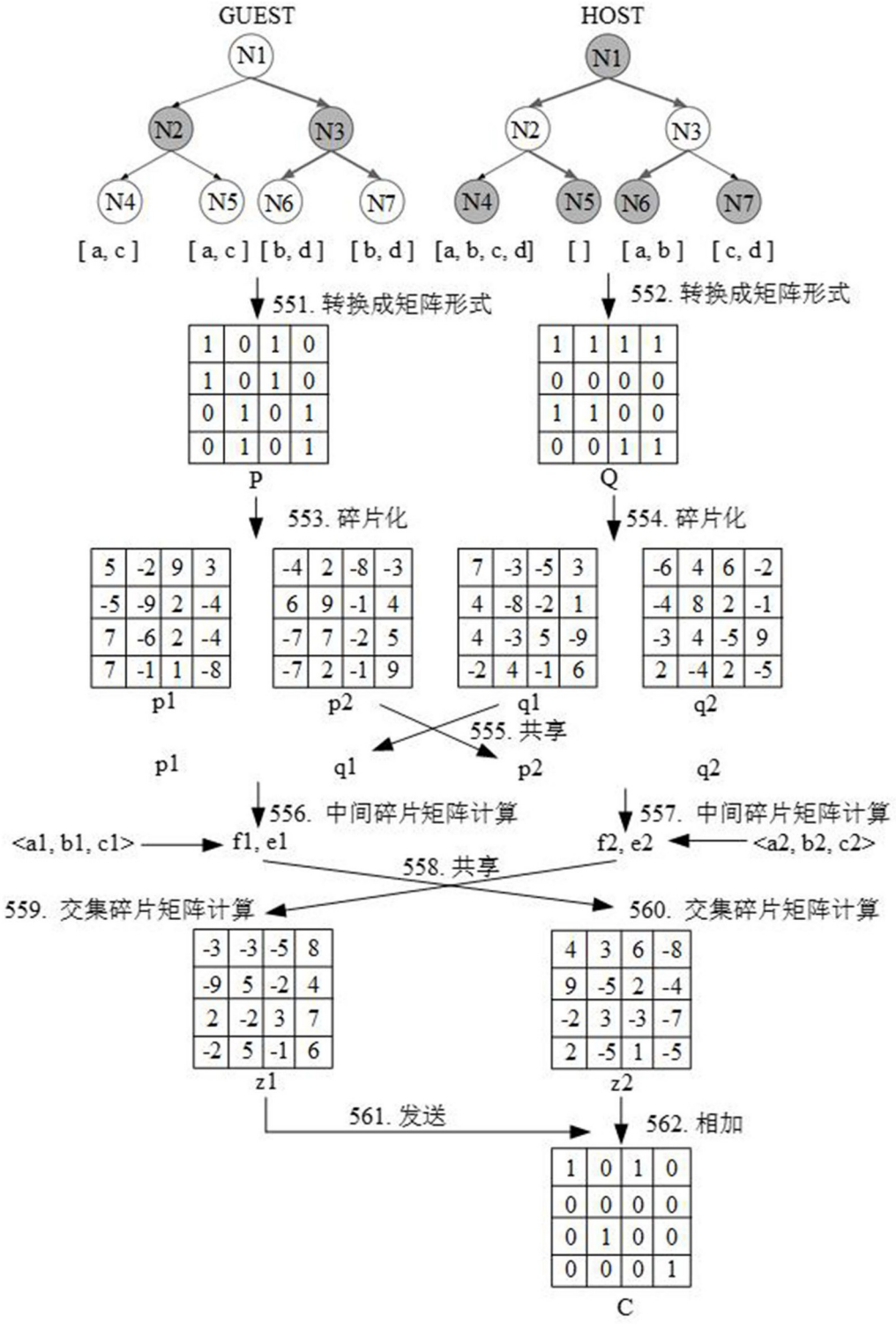

GUEST
HOST
N1
N2
N3
N4
N5
N6
N7
[a, c]
[a, c]
[b, d]
[b, d]
[a, b, c, d]
[ ]
[a, b]
[c, d]
551. 转换成矩阵形式
552. 转换成矩阵形式
P
Q
553. 碎片化
554. 碎片化
p1
p2
q1
q2
555. 共享
556. 中间碎片矩阵计算
557. 中间碎片矩阵计算
<a1, b1, c1>
f1, e1
f2, e2
<a2, b2, c2>
558. 共享
559. 交集碎片矩阵计算
560. 交集碎片矩阵计算
z1
z2
561. 发送
562. 相加
C

图 5

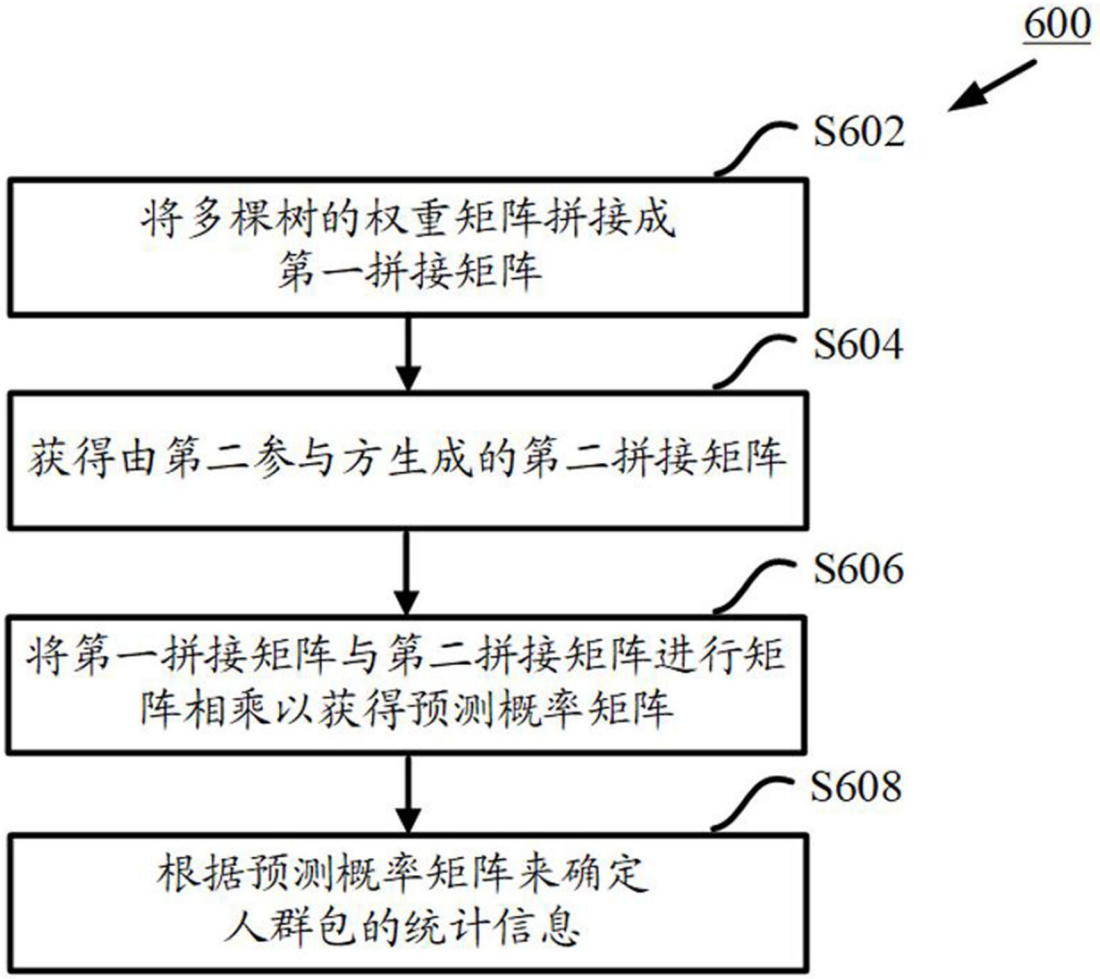

600
S602
将多棵树的权重矩阵拼接成第一拼接矩阵
S604
获得由第二参与方生成的第二拼接矩阵
S606
将第一拼接矩阵与第二拼接矩阵进行矩阵相乘以获得预测概率矩阵
S608
根据预测概率矩阵来确定人群包的统计信息

图 6

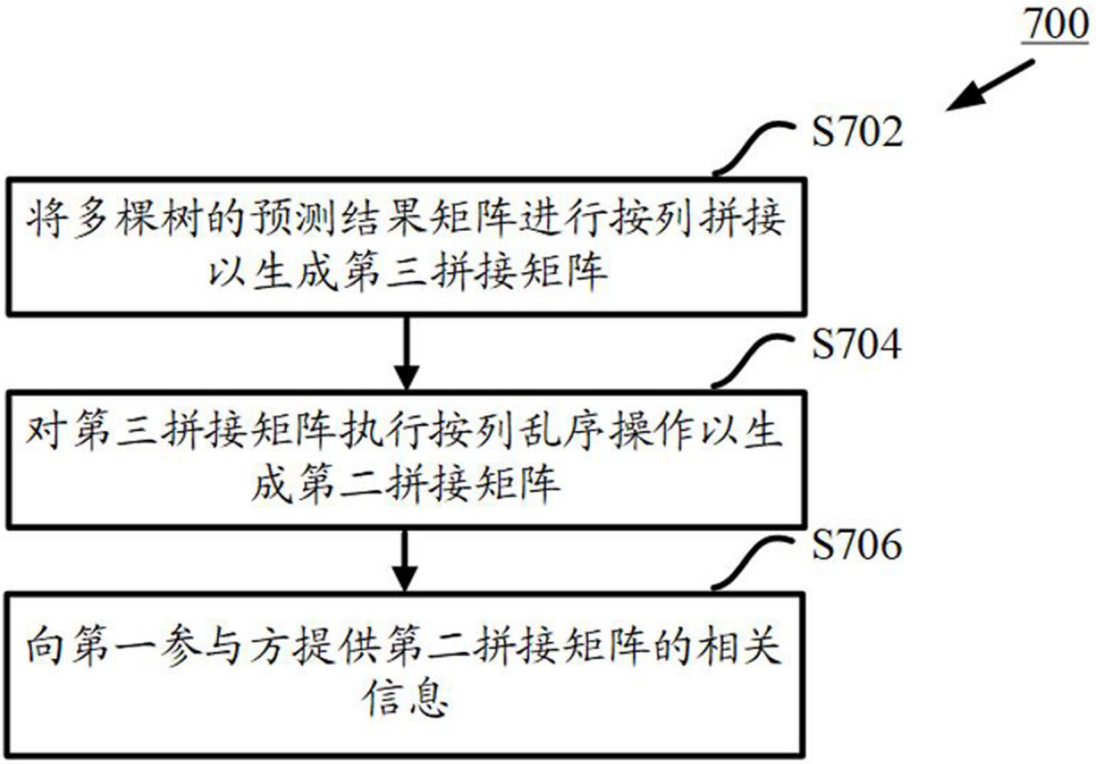

700
S702
将多棵树的预测结果矩阵进行按列拼接以生成第三拼接矩阵
S704
对第三拼接矩阵执行按列乱序操作以生成第二拼接矩阵
S706
向第一参与方提供第二拼接矩阵的相关信息

图 7

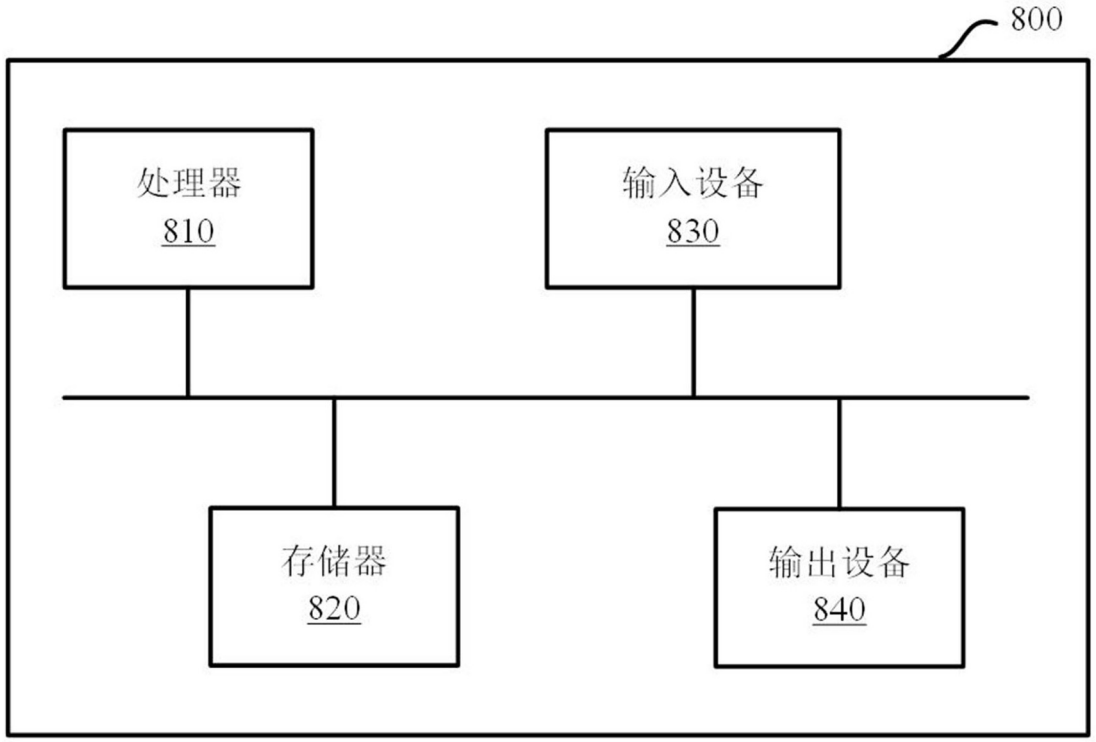

800
处理器
810
输入设备
830
存储器
820
输出设备
840

图 8

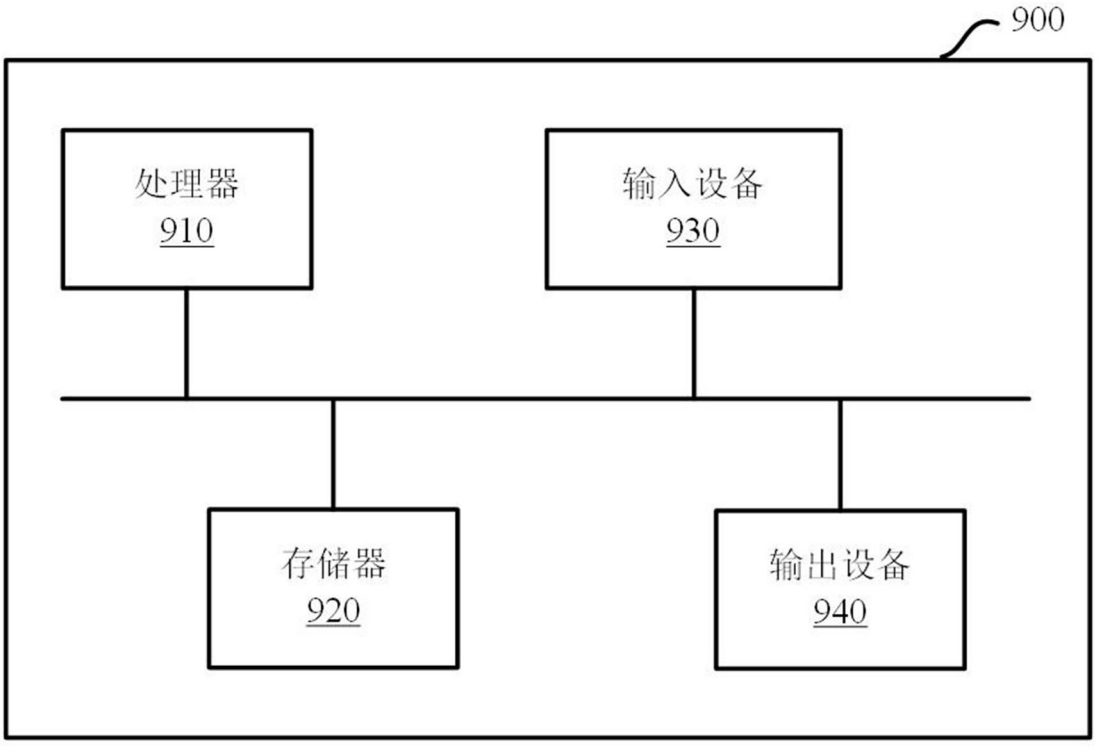


图 9