(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 112100679 B
(45)授权公告日 2021.03.02

(21)申请号 202011276388.X

(22)申请日 2020.11.16

(65)同一申请的已公布的文献号
申请公布号 CN 112100679 A

(43)申请公布日 2020.12.18

(73)专利权人 支付宝(杭州)信息技术有限公司
地址 310000 浙江省杭州市西湖区西溪路556号8层B段801-11

(72)发明人 周亚顺 李漓春 应鹏飞

(74)专利代理机构 北京三友知识产权代理有限公司 11127
代理人 阚传猛 周达

(51)Int.Cl.
G06F 21/62(2013.01)
G06F 21/60(2013.01)
G06F 16/27(2019.01)
G06F 16/2457(2019.01)

审查员 朱江岩

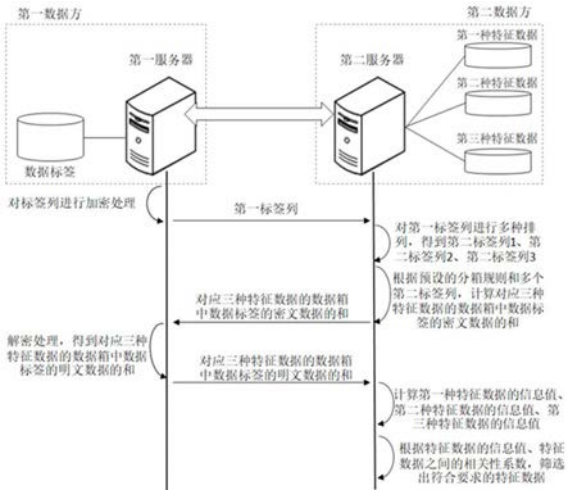
权利要求书5页 说明书18页 附图8页

(54)发明名称

基于隐私保护的数据处理方法、装置和服务器

(57)摘要

本说明书提供了基于隐私保护的数据处理方法、装置和服务器。基于该方法，持有多种特征数据的第二服务器在接收到包含有按照标识信息排列的数据标签的密文数据的第一标签列后，可以根据多种特征数据的数据值的排列顺序对第一标签列进行多种排列，得到对应多种特征数据的多个第二标签列；再基于上述多个第二标签列和预设的分箱规则，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和，并反馈给第二服务器：第二服务器通过解密，得到并向第一服务器反馈相应的明文数据的和。进而第一服务器可以根据上述明文数据的和，计算出各种特征数据的信息值；并根据特征数据的信息值、特征数据之间的相关性系数，筛选出符合要求的特征数据。



CN 112100679 B

1.一种基于隐私保护的数据处理方法,包括:

接收第一服务器发送的第一标签列;其中,所述第一标签列包含有与标识信息对应的数据标签的密文数据,所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列;

根据多个特征列,对所述第一标签列中的数据标签的密文数据进行多种排列,得到对应的多个第二标签列;其中,所述第二标签列与一种特征数据对应;所述特征列分别包含有一种特征数据,且所包含的特征数据根据特征数据的数据值排列;

根据预设的分箱规则和所述多个第二标签列,确定出对应多种特征数据的数据箱中的数据标签的密文数据的和;并将所述对应多种特征数据的数据箱中的数据标签的密文数据的和发送至第一服务器;其中,所述第一服务器对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密,得到对应多种特征数据的数据箱中的数据标签的明文数据的和;

根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和,确定出特征数据的信息值;

根据特征数据的信息值,以及特征数据之间的相关性系数,从多种特征数据中筛选出符合要求的特征数据;

其中,在得到对应的多个第二标签列之后,所述方法还包括:根据预设的分箱规则,分别在所述多个第二标签列中的各个第二标签列上确定并标记出数据箱的分位点,得到多个标记后的第二标签列;对所述多个标记后的第二标签列进行随机化操作,得到随机化操作后的第二标签列;将所述随机化操作后的第二标签列发送至第一服务器;其中,所述第一服务器用于根据所述随机化操作后的第二标签列,计算出对应多种特征数据的数据箱中的数据标签的明文数据的和。

2.根据权利要求1所述的方法,所述预设的分箱规则包括以下至少之一:等频分箱规则、卡方分箱规则、等宽分箱规则。

3.根据权利要求2所述的方法,根据预设的分箱规则和所述多个第二标签列,确定出对应多种特征数据的数据箱中的数据标签的密文数据的和,包括:

根据预设的分箱规则,分别在所述多个第二标签列中的各个第二标签列上确定出数据箱的分位点;

分别统计所述多个第二标签列中的各个第二标签列上的相邻的分位点之间的数据标签的密文数据的和,得到所述对应多种特征数据的数据箱中的数据标签的密文数据的和。

4.根据权利要求3所述的方法,在所述预设的分箱规则包括等频分箱规则的情况下,根据预设的分箱规则,分别在所述多个第二标签列中的各个第二标签列上确定出数据箱的分位点,包括:

按照以下方式在所述多个第二标签列中的当前第二标签列上确定出数据箱的分位点:

根据预设的分箱规则,确定出单个数据箱所包含的数据的目标数量;

从所述当前第二标签列的起始位置出发,沿所述当前第二标签列每间隔目标数量个的数据标签的密文数据确定一个数据箱的分位点。

5.根据权利要求4所述的方法,分别统计所述多个第二标签列中的各个第二标签列上的相邻的分位点之间的数据标签的密文数据的和,得到所述对应多种特征数据的数据箱中

的数据标签的密文数据的和,包括:

按照以下方式确定出对应多种特征数据中的当前特征数据的数据箱中的数据标签的密文数据的和:

从多个第二标签列中确定出与当前特征数据对应的当前第二标签列;

从所述当前第二标签列的起始位置出发,检测出当前第二标签列上的分位点;并将当前第二标签列上相邻的两个分位点之间的数据标签的密文数据划分进一个数据箱,得到对应当前特征数据的多个数据箱;

统计所述对应当前特征数据的多个数据箱中的各个数据箱的数据标签的密文数据的和,作为对应当前特征数据的数据箱中的数据标签的密文数据的和。

6.根据权利要求1所述的方法,根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和,确定出特征数据的信息值,包括:

按照以下方式确定出多种特征数据中的当前特征数据的信息值:

根据与当前特征数据对应的特征列,确定出对应当前特征数据的数据箱中的特征数据;

根据对应当前特征数据的数据箱中的数据标签的明文数据的和、所述对应当前特征数据的数据箱中的特征数据,计算对应当前特征数据的各个数据箱的权重证明;

根据所述对应当前特征数据的各个数据箱的权重证明,计算对应当前特征数据的数据箱的信息值;

根据所述对应当前特征数据的数据箱的信息值,计算当前特征数据的信息值。

7.根据权利要求1所述的方法,根据特征数据的信息值,以及特征数据之间的相关性系数,从多种特征数据中筛选出符合要求的特征数据,包括:

从多种特征数据中,筛选出特征数据之间的相关性系数大于预设的相关性系数阈值的特征数据组合,作为筛选组;

比较所述筛选组中的特征数据的信息值,保留所述筛选组中的信息值最大的特征数据作为所述符合要求的特征数据。

8.一种基于隐私保护的数据处理方法,包括:

对标签列进行加密处理,得到第一标签列,并将所述第一标签列发送至第二服务器;其中,所述第一标签列包含有与标识信息对应的数据标签的密文数据,所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列;

接收第二服务器发送的对应多种特征数据的数据箱中的数据标签的密文数据的和;其中,所述第二服务器持有与标识信息对应的多种特征数据;所述第二服务器根据所述多种特征数据对所述第一标签列中的数据标签的密文数据进行多种排列,得到对应的多个第二标签列;所述第二服务器根据预设的分箱规则和所述多个第二标签列,确定出对应多种特征数据的数据箱中的数据标签的密文数据的和;

对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密,得到对应多种特征数据的数据箱中的数据标签的明文数据的和;

将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器;其中,所述第二服务器根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和,确定出特征数据的信息值;所述第二服务器根据特征数据的信息值,以及特征数据之间

的相关性系数，从多种特征数据中筛选出符合要求的特征数据；

其中，在将所述第一标签列发送至第二服务器之后，所述方法还包括：接收第二服务器发送的随机化操作后的第二标签列；根据所述随机化操作后的第二标签列，计算出对应多种特征数据的数据箱中的数据标签的明文数据的和；其中，所述随机化操作后的第二标签列为第二服务器根据预设的分箱规则，在所述多个第二标签列中的各个第二标签列上确定并标记出分位点，得到多个标记后的第二标签列后，对所述多个标记后的第二标签列进行随机化操作得到的。

9.根据权利要求8所述的方法，对标签列进行加密处理，包括：利用同态加密算法对所述标签列中的数据标签进行加密处理。

10.根据权利要求8所述的方法，在将所述第一标签列发送至第二服务器之后，所述方法还包括：

接收第二服务器发送的多个标记后的第二标签列；

根据所述多个标记后的第二标签列，计算对应多种特征数据的数据箱中的数据标签的明文数据的和；

将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器。

11.根据权利要求10所述的方法，根据所述多个标记后的第二标签列，计算对应多种特征数据的数据箱中的数据标签的明文数据的和，包括：

按照以下方式计算出对应多种特征数据中的当前特征数据的数据标签的明文数据的和：

从多个第二标签列中确定出与当前特征数据对应的当前第二标签列；

对所述当前第二标签列中的数据标签的密文数据进行解密，得到解密后的当前第二标签列；其中，所述解密后的当前第二标签列包含有数据标签的明文数据；

从所述解密后的当前第二标签列的起始位置出发，检测出所述解密后的当前第二标签列上的分位点；并将相邻的两个分位点之间的数据标签的明文数据划分为一个数据箱，得到对应当前特征数据的多个数据箱；

统计所述对应当前特征数据的多个数据箱中的各个数据箱的数据标签的明文数据的和，作为对应当前特征数据的数据箱中的数据标签的明文数据的和。

12.根据权利要求8所述的方法，在将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器之后，所述方法还包括：

接收第二服务器发送的符合要求的特征数据；

利用所述符合要求的特征数据、所述数据标签，进行模型训练，以建立得到目标模型。

13.根据权利要求8所述的方法，在将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器之后，所述方法还包括：

以数据标签作为输出，与以所述符合要求的特征数据作为输出的第二服务器，通过多方安全计算，来建立目标模型。

14.根据权利要求13所述的方法，所述特征数据包括以下至少之一：用户的月收入数据、用户的月支出数据、用户的年龄数据。

15.根据权利要求14所述的方法，所述目标模型包括用户信用风险预测模型。

16.一种基于隐私保护的数据处理装置，包括：

接收模块，用于接收第一服务器发送的第一标签列；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列；

排列模块，用于根据多个特征列，对所述第一标签列中的数据标签的密文数据进行多种排列，得到对应的多个第二标签列；其中，所述第二标签列与一种特征数据对应；所述特征列分别包含有一种特征数据，且所包含的特征数据根据特征数据的数据值排列；

第一确定模块，用于根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和；并将所述对应多种特征数据的数据箱中的数据标签的密文数据的和发送至第一服务器；其中，所述第一服务器对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和；

第二确定模块，用于根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出特征数据的信息值；

筛选模块，用于根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据；

其中，所述装置在得到对应的多个第二标签列之后，还用于根据预设的分箱规则，分别在所述多个第二标签列中的各个第二标签列上确定并标记出数据箱的分位点，得到多个标记后的第二标签列；对所述多个标记后的第二标签列进行随机化操作，得到随机化操作后的第二标签列；将所述随机化操作后的第二标签列发送至第一服务器；其中，所述第一服务器用于根据所述随机化操作后的第二标签列，计算出对应多种特征数据的数据箱中的数据标签的明文数据的和。

17.一种基于隐私保护的数据处理装置，包括：

加密模块，用于对标签列进行加密处理，得到第一标签列，并将所述第一标签列发送至第二服务器；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列；

接收模块，用于接收第二服务器发送的对应多种特征数据的数据箱中的数据标签的密文数据的和；其中，所述第二服务器持有与标识信息对应的多种特征数据；所述第二服务器根据所述多种特征数据对所述第一标签列中的数据标签的密文数据进行多种排列，得到对应的多个第二标签列；所述第二服务器根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和；

解密模块，用于对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和；

发送模块，用于将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器；其中，所述第二服务器根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出特征数据的信息值；所述第二服务器根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据；

其中，所述装置还用于接收第二服务器发送的随机化操作后的第二标签列；根据所述随机化操作后的第二标签列，计算出对应多种特征数据的数据箱中的数据标签的明文数据的和；其中，所述随机化操作后的第二标签列为第二服务器根据预设的分箱规则，在所述多

个第二标签列中的各个第二标签列上确定并标记出分位点,得到多个标记后的第二标签列后,对所述多个标记后的第二标签列进行随机化操作得到的。

18.一种服务器,包括处理器以及用于存储处理器可执行指令的存储器,所述处理器执行所述指令时实现权利要求1至7中任一项所述方法的步骤。

19.一种服务器,包括处理器以及用于存储处理器可执行指令的存储器,所述处理器执行所述指令时实现权利要求8至15中任一项所述方法的步骤。

基于隐私保护的数据处理方法、装置和服务器

技术领域

[0001] 本说明书属于互联网技术领域，尤其涉及基于隐私保护的数据处理方法、装置和服务器。

背景技术

[0002] 在许多数据处理场景中，不同的数据方可能分别拥有同一组数据对象的不同的特征数据。

[0003] 例如，数据方A拥有数据对象的数据标签，数据方B拥有同一组数据对象的多种不同的特征数据。当前，数据方A希望先从数据方B所持有多种不同的特征数据中筛选出效果相对较好的特征数据，再利用该特征数据进行例如联合统计等相关的数据处理。

[0004] 因此，亟需一种能够在不泄露双方所各自持有的数据信息、保护双方的数据隐私的前提下，安全、高效地从多种特征数据中筛选出符合要求的特征数据的方法。

发明内容

[0005] 本说明书提供了一种基于隐私保护的数据处理方法、装置和服务器，以在不泄露双方所各自持有的数据信息、保护双方的数据隐私的前提下，安全、高效地从多种特征数据中筛选出符合要求的特征数据。

[0006] 本说明书提供的一种基于隐私保护的数据处理方法、装置和服务器是这样实现的：

[0007] 一种基于隐私保护的数据处理方法，包括：接收第一服务器发送的第一标签列；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列；根据多个特征列，对所述第一标签列中的数据标签的密文数据进行多种排列，得到对应的多个第二标签列；其中，所述第二标签列与一种特征数据对应；所述特征列分别包含有一种特征数据，且所包含的特征数据根据特征数据的数据值排列；根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和；并将所述对应多种特征数据的数据箱中的数据标签的密文数据的和发送至第一服务器；其中，所述第一服务器对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和；根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出特征数据的信息值；根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据。

[0008] 一种基于隐私保护的数据处理方法，包括：对标签列进行加密处理，得到第一标签列，并将所述第一标签列发送至第二服务器；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列；接收第二服务器发送的对应多种特征数据的数据箱中的数据标签的密文数据的和；其中，所述第二服务器持有与标识信息对应的多种特征数据；所述第二服务器根据所述

多种特征数据对所述第一标签列中的数据标签的密文数据进行多种排列，得到对应的多个第二标签列；所述第二服务器根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和；对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和；将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器；其中，所述第二服务器根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出特征数据的信息值；所述第二服务器根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据。

[0009] 一种基于隐私保护的数据处理装置，包括：接收模块，用于接收第一服务器发送的第一标签列；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列；排列模块，用于根据多个特征列，对所述第一标签列中的数据标签的密文数据进行多种排列，得到对应的多个第二标签列；其中，所述第二标签列与一种特征数据对应；所述特征列分别包含有一种特征数据，且所包含的特征数据根据特征数据的数据值排列；第一确定模块，用于根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和；并将所述对应多种特征数据的数据箱中的数据标签的密文数据的和发送至第一服务器；其中，所述第一服务器对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和；第二确定模块，用于根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出特征数据的信息值；筛选模块，用于根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据。

[0010] 一种基于隐私保护的数据处理装置，包括：加密模块，用于对标签列进行加密处理，得到第一标签列，并将所述第一标签列发送至第二服务器；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列；接收模块，用于接收第二服务器发送的对应多种特征数据的数据箱中的数据标签的密文数据的和；其中，所述第二服务器持有与标识信息对应的多种特征数据；所述第二服务器根据所述多种特征数据对所述第一标签列中的数据标签的密文数据进行多种排列，得到对应的多个第二标签列；所述第二服务器根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和；解密模块，用于对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和；发送模块，用于将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器；其中，所述第二服务器根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出特征数据的信息值；所述第二服务器根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据。

[0011] 一种服务器，包括处理器以及用于存储处理器可执行指令的存储器，所述处理器执行所述指令时实现上述基于隐私保护的数据处理方法的相关步骤。

[0012] 本说明书提供的一种基于隐私保护的数据处理方法、装置和服务器，基于该方法，持有多种特征数据的第二服务器在接收到包含有按照标识信息排列的数据标签的密文数

据的第一标签列后，可以根据多种特征数据的数据值的排列顺序对第一标签列进行多种排列，得到对应多种特征数据的多个第二标签列；再基于上述多个第二标签列和预设的分箱规则，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和，并反馈给第二服务器：第二服务器通过解密，得到并向第一服务器反馈相应的明文数据的和。进而第一服务器可以根据上述明文数据的和，计算出各种特征数据的信息值；并根据特征数据的信息值、特征数据之间的相关性系数，筛选出符合要求的特征数据。从而可以在不泄露双方所各自持有的数据信息、保护双方的数据隐私的前提下，安全、高效地从多种特征数据中筛选出符合要求的特征数据。

附图说明

[0013] 为了更清楚地说明本说明书实施例，下面将对实施例中所需要使用的附图作简单地介绍，下面描述中的附图仅仅是本说明书中记载的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[0014] 图1是应用本说明书实施例提供的基于隐私保护的数据处理方法的系统的结构组成的一个实施例的示意图；

[0015] 图2是本说明书的一个实施例提供的基于隐私保护的数据处理方法的流程示意图；

[0016] 图3是在一个场景示例中，应用本说明书实施例提供的基于隐私保护的数据处理方法的一种实施例的示意图；

[0017] 图4是在一个场景示例中，应用本说明书实施例提供的基于隐私保护的数据处理方法的一种实施例的示意图；

[0018] 图5是本说明书的一个实施例提供的基于隐私保护的数据处理方法的流程示意图；

[0019] 图6是本说明书的一个实施例提供的服务器的结构组成示意图；

[0020] 图7是本说明书的一个实施例提供的基于隐私保护的数据处理装置的结构组成示意图；

[0021] 图8是本说明书的一个实施例提供的基于隐私保护的数据处理装置的结构组成示意图。

具体实施方式

[0022] 为了使本技术领域的人员更好地理解本说明书中的技术方案，下面将结合本说明书实施例中的附图，对本说明书实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本说明书一部分实施例，而不是全部的实施例。基于本说明书中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都应当属于本说明书保护的范围。

[0023] 本说明书实施例提供一种基于隐私保护的数据处理方法，该方法具体可以应用于包含有第一服务器和第二服务器的系统中。具体可以参阅图1所示。系统中的第一服务器、第二服务器可以通过有线或无线的方式相连，以进行相应的数据交互。

[0024] 其中，上述第一服务器具体可以理解为部署于第一数据方一侧的服务器，至少持

有第一数据方所拥有的与标识信息对应的数据标签。上述第二服务器具体可以理解为部署于第二数据方一侧的服务器，至少持有第二数据方所拥有的与同一组标识信息对应的多种不同的特征数据。例如，上述第二服务器持有第一种特征数据、第二种特征数据和第三种特征数据三种不同的特征数据。

[0025] 在本实施例中，上述第一服务器、第二服务器具体可以包括一种能够实现数据传输、数据处理等功能的后台服务器。具体的，上述第一服务器、第二服务器例如可以为一个具有数据运算、存储功能以及网络交互功能的电子设备。或者，上述第一服务器、第二服务器也可以为运行于该电子设备中，为数据处理、存储和网络交互提供支持的软件程序。在本实施例中，并不具体限定上述第一服务器、第二服务器所包含的服务器数量。上述第一服务器、第二服务器具体可以为一个服务器，也可以为几个服务器，或者，由若干服务器形成的服务器集群。

[0026] 当前要求在不泄露双方所各自持有的数据信息、保护双方的数据隐私的前提下，从第二服务器所持有的多种特征数据中筛选出效果较好的、符合要求的特征数据。

[0027] 具体实施时，第一服务器可以先对己方所持有的标签列进行加密处理，得到第一标签列。其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列。第一服务器将上述第一标签列发送至第二服务器。

[0028] 第二服务器接收第一标签列。第二服务器可以根据所持有的多个特征列，对所述第一标签列中的数据标签的密文数据进行多种排列，得到与所述多个特征列分别对应的多个第二标签列。其中，所述多个第二标签列中的每一个第二标签列与一种特征数据对应，所述第二标签列中所包含的数据标签的密文数据按照所对应的特征数据的数据值的排列顺序排列。所述多个特征列中的每一个特征列分别包含有一种特征数据，且每一个特征列所包含的特征数据根据特征数据的数据值的排列顺序排列。例如，第二服务器可以持有三个不同的特征列，即：对应第一种特征数据的特征列1、对应第二种特征数据的特征列2、对应第三种特征数据的特征列3。相应的，可以根据特征1对第一标签列中的数据标签的密文数据进行重新排列，得到对应第一种特征数据的第二标签列1；根据特征2对第一标签列中的数据标签的密文数据进行重新排列，得到对应第一种特征数据的第二标签列2；根据特征3对第一标签列中的数据标签的密文数据进行重新排列，得到对应第一种特征数据的第二标签列3。

[0029] 第二服务器可以根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和。具体的，第二服务器可以先根据预设的分箱规则，在对应多种特征数据的多个第二标签列上分别确定出各个数据箱的分位点，进行数据分箱。再针对多种特征数据中的每一种特征数据，分别计算各个数据箱中的数据标签的密文数据的和，得到对应多种特征数据的数据箱中的数据标签的密文数据的和。例如，第二服务器可以分别计算出对应第一种特征数据的数据箱中的数据标签的密文数据的和、对应第二种特征数据的数据箱中的数据标签的密文数据的和、对应第三种特征数据的数据箱中的数据标签的密文数据的和，即得到对应三种特征数据的数据箱中数据标签的密文数据的和。

[0030] 第二服务器可以将所述对应多种特征数据的数据箱中的数据标签的密文数据的

和发送至第一服务器。

[0031] 第一服务器对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密处理，得到对应多种特征数据的数据箱中的数据标签的明文数据的和；再将对应多种特征数据的数据箱中的数据标签的明文数据的和反馈给第二服务器。例如，第一服务器可以分别对上述对应第一种特征数据的数据箱中的数据标签的密文数据的和、对应第二种特征数据的数据箱中的数据标签的密文数据的和、对应第三种特征数据的数据箱中的数据标签的密文数据的和进行解密处理，得到对应第一种特征数据的数据箱中的数据标签的明文数据的和、对应第二种特征数据的数据箱中的数据标签的明文数据的和、对应第三种特征数据的数据箱中的数据标签的明文数据的和，即得到对应三种特征数据的数据箱中数据标签的明文数据的和。

[0032] 第二服务器根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出多种特征数据中的各种特征数据的信息值。例如，第二服务器可以根据上述对应第一种特征数据的数据箱中的数据标签的明文数据的和、对应第二种特征数据的数据箱中的数据标签的明文数据的和、对应第三种特征数据的数据箱中的数据标签的明文数据的和分别计算第一种特征数据的信息值、第二种特征数据的信息值、第三种特征数据的信息值。

[0033] 进一步，第二服务器可以根据上述不同特征数据的信息值，以及特征数据之间的相关性系数，从所持有的多种特征数据中筛选出符合要求的特征数据。

[0034] 通过上述系统，可以安全、高效地完成相应的数据分箱，并计算出各种特征数据的信息值；进而可以根据各种特征数据的信息值、以及特征数据之间的相关性系数，准确地从多种特征数据中筛选出符合要求的特征数据。避免了在筛选过程中双方的数据信息遭到泄露，保护了双方的数据隐私的前提下，安全、高效地实现特征数据的筛选。

[0035] 参阅图2所示，本说明书实施例提供了一种基于隐私保护的数据处理方法。其中，该方法具体可以应用于第二服务器一侧。具体实施时，该方法可以包括以下内容。

[0036] S201：接收第一服务器发送的第一标签列；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列。

[0037] 在一个实施例中，上述第二服务器具体可以理解为部署于第二数据方一侧的服务器，其中，第二服务器可以持有并使用第二数据方所拥有的与标识信息对应的多种不同种类的特征数据。上述第一服务器具体可以理解为部署于第一数据方一侧的服务器，其中，第一服务器至少持有并使用第一数据方所拥有的与相同的标识信息对应的数据标签。可以将对应同一标识信息的特征数据和数据标签称为相互对应。

[0038] 需要补充的是，除了持有数据标签外，上述第一服务器也可以持有对应相同标识信息的特征数据。

[0039] 在一个实施例中，上述标识信息具体可以包括一种与数据对象一一对应的标识信息。具体的，上述标识信息可以是用户(对象)的身份ID、用户的账户名，或者用户的注册手机号码等。上述标识信息也可以是设备(对象)的设备编号、设备的物理地址，或者设备的IP地址等。上述标识信息还可以是通过哈希计算所得到的哈希值中的指定字段等。当然，上述所列举的标识信息只是一种示意性说明。具体实施时，根据具体的应用场景和处理需要上述标识信息还可以包括其他类型的标识信息。对此，本说明书不作限定。

[0040] 上述特征数据具体可以包括一种以数值的形式表征出所对应的标识信息所指示的数据对象的属性特征的数据。例如，上述特征数据具体可以是用户的年龄、用户的月收入、用户的违约次数等数据。在实施例中，上述第一服务器可以持有多种不同的特征数据。

[0041] 上述数据标签具体可以包括一种用于指示数据对象所属类型的标签。具体的，上述数据标签可以包括正标签和负标签。

[0042] 在不同的应用场景中，上述正标签和负标签具体又可以用于指示数据对象所属的不同类型。例如，在用户信用风险检测场景中，上述正标签可以用于指示不存在信用风险的用户，上述负标签可以用于指示存在信用风险的用户。又例如，在用户的购买意愿预测场景中，上述正标签可以指示具有购买意愿的用户，上述负标签可以用于指示不具有购买意愿的用户等等。

[0043] 当然，需要说明的是上述所列举的数据标签只是一种示意性说明。具体实施时，根据具体的应用场景，上述数据标签还可以包括除正标签、负标签以外其他类型的数据标签。对此，本说明书不作限定。

[0044] 在一个实施例中，具体的，例如，在用户的信用风险检测场景中，第一服务器可以持有与用户A的身份ID对应的用于指示用户A是否存在信用风险的数据标签。第二服务器可以同时持有与用户A的身份ID对应的，用户A的多种不同的特征数据。例如，第二服务器可以同时持有用户A的年龄数据、用户A的月收入数据、用户A的违约次数数据这三种特征数据。

[0045] 在一个实施例中，具体实施前，第二服务器可以根据特征数据所对应的标识信息的排列顺序排列所拥有的属于同一种类的特征数据，得到多个初始的特征数据列。其中，每一个初始的特征数据列对应一种特征数据，且每一个初始的特征数据列中所包含的属于同一种类的特征数据按照所对应的标识信息的排列顺序排列。

[0046] 具体的，例如，基于标识信息的排列顺序，用户A的身份ID排在第一位。相应的，在对应月收入数据的第一个初始的特征数据列中，用户A的月收入数据也排在第一位；在对应年龄数据的第二个初始的特征数据列中，用户A的年龄数据也排在第一位。

[0047] 在一个实施例中，具体实施前，第一服务器可以根据相同的标识信息的排列顺序排列所拥有的数据标签，得到初始的数据标签列，可以简记为标签列。

[0048] 具体的，例如，基于标识信息的排列顺序，用户A的身份ID排在第一位；相应的，在所述初始的数据标签列中，用户A的数据标签也排在第一位。

[0049] 在一个实施例中，具体实施时，第一服务器可以响应相应的数据处理请求，先对所拥有的标签列中的各个数据标签进行加密处理，得到对应的第一标签列(可以记为label_cipher)。其中，上述第一标签列包含有与标识信息对应的数据标签的密文数据，并且第一标签列中的数据标签的密文数据可以是根据标识信息的排列顺序排列的。再将上述第一标签列发送给第二服务器。这样第二服务器无法根据上述第一标签列知晓各个数据对象的数据标签的具体内容，从而可以避免向第二服务器泄露数据标签的数据信息，保护第一服务器一侧的数据隐私。

[0050] 在一个实施例中，第一服务器具体可以通过同态加密算法对上述标签列中的各个数据标签进行加密处理，得到各个数据标签的密文数据，得到对应的第一标签列。

[0051] 其中，所使用的同态加密算法具体可以包括elgamal同态加密算法。通过上述elgamal同态加密算法加密数据标签得到的密文数据的大小是固定的，且上述数据标签的

密文数据在第一标签列中还是按照之前的标识信息的排列顺序排列的。当然，上述所列举的同态加密算法只是一种示意性说明。具体实施时，根据具体的应用场景和处理需求，第一服务器还可以采用其他合适的同态加密算法对标签列进行加密处理。

[0052] 在一个实施例中，上述数据处理请求具体可以是一种请求对第二服务器所持有多种的特征数据进行筛选，以从多种特征数据中筛选出使用效果较好、质量较高的特征数据作为符合要求的特征数据的请求数据，当然，上述所列举的数据处理请求只是一种示意性说明。具体实施时，根据具体的应用场景和处理需求，上述数据处理请求还可以包括其他类型的数据处理请求。对此，本说明书不作限定。

[0053] 其中，上述数据处理请求具体可以是第一服务器发起的，也可以是第二服务器发起的，还可以是需要使用特征数据的第三方发起的。

[0054] 在一个实施例中，第二服务器可以接收第一服务器发送的上述第一标签列。

[0055] S202：根据多个特征列，对所述第一标签列中的数据标签的密文数据进行多种排列，得到对应的多个第二标签列；其中，所述第二标签列与一种特征数据对应；所述特征列分别包含有一种特征数据，且所包含的特征数据根据特征数据的数据值排列。

[0056] 在一个实施例中，第二服务器可以响应数据处理请求，根据特征数据的数据值重新排列多个初始的特征数据列中的特征数据，得到对应的多个特征列。其中，上述多个特征列中的各个特征列分别对应一种特征数据。且多个特征列中的各个特征列所包含的属于同一种类的特征数据是按照特征数据的数据值的排列顺序(例如数据值从小到大的排列顺序，或者数据值从大到小的排列顺序)排列。

[0057] 在一个实施例中，第二服务器在具体实施之前，也可以不需要构建多个初始的特征数据列，而是直接根据特征数据的数据值分别排列所拥有的属于不同种类的特征数据，得到对应的多个特征列。

[0058] 在一个实施例中，根据多个特征列，对所述第一标签列中的数据标签的密文数据进行多种排列，得到对应的多个第二标签列。其中，所述多个第二标签列中的每一个第二标签列对应于一种特征数据，该第二标签列中的数据标签的密文数据是根据所对应的特征数据的数据值的排列顺序排列的。

[0059] 具体的实施时，以得到对应当前特征数据的当前第二标签列为例，可以根据与当前特征数据对应的当前特征列，对第一标签列上的数据标签的密文数据进行重新排列。

[0060] 具体的，由于在第一标签列中的数据标签的密文数据是按照标识信息排列的。因此，第二服务器可以先根据各个数据标签的密文数据在第一标签列中的排列位置，确定出该数据标签的密文数据所对应的标识信息，即确定出该数据标签的密文数据所对应的数据对象。但是，由于第二服务器无法解密数据标签的密文数据，因此，第二服务器无法知晓各个与各个标识信息对应的数据标签的具体内容。

[0061] 进一步，第二服务器可以根据对应同一个标识信息的当前特征数据在当前特征列中的排列位置，对应调整数据标签的密文数据在第一标签列中的排列位置，从而实现对第一标签列中的数据标签的密文数据的重新排列，得到与当前特征数据对应的当前第二标签列。

[0062] 例如，参阅图3所示，在当前特征列中，对应于用户A的身份ID的当前特征数据排列位置为第四位，这时第二服务器可以将第一标签列中对应用户A的身份ID的数据标签的密

文数据从原来的第一位，调整到对应的第四位。按照类似的方式，对第一标签列上其他的数据标签的密文数据进行重新排列，得到对应当前特征数据的当前第二标签列。

[0063] 在一个实施例中，第二服务器可以按照上述方式分别根据对应不同特征数据的多个特征列，对第一标签列进行重新排列，得到分别对应多种不同特征数据的多个第二标签列。

[0064] S203：根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和；并将所述对应多种特征数据的数据箱中的数据标签的密文数据的和发送至第一服务器；其中，所述第一服务器对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和。

[0065] 在一个实施例中，第二服务器可以根据预设的分箱规则，基于所述多个第二标签列对所述多个第二标签列上的数据标签的密文数据分别进行数据分箱，得到对应多种特征数据的数据箱；进而可以统计对应多种特征数据的数据箱中的数据标签的密文数据的和。

[0066] 在一个实施例中，上述预设的分箱规则具体可以包含有在对第二标签列上的数据标签的密文数据进行数据分箱时所依据的相关参数。

[0067] 在一个实施例中，所述预设的分箱规则具体可以包括以下至少之一：等频分箱规则、卡方分箱规则、等宽分箱规则等。当然，上述所列举的分箱规则只是一种示意性说明。具体实施时，根据具体情况，上述预设的分箱规则还可以包括其他类型的分箱规则。例如，等正例分箱规则、等负例分箱规则等。

[0068] 在一个实施例中，上述根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和，具体实施时，可以包括：根据预设的分箱规则，分别在所述多个第二标签列中的各个第二标签列上确定出数据箱的分位点；分别统计所述多个第二标签列中的各个第二标签列上的相邻的分位点之间的数据标签的密文数据的和，得到所述对应多种特征数据的数据箱中的数据标签的密文数据的和。

[0069] 在一个实施例中，以等频分箱规则作为预设的分箱规则为例，预设的分箱规则中可以包含有单个数据箱所包含的数据的数量，记为目标数量。

[0070] 在一个实施例中，在所述预设的分箱规则包括等频分箱规则的情况下，根据预设的分箱规则，分别在所述多个第二标签列中的各个第二标签列上确定出数据箱的分位点，具体实施时，可以包括以下内容：

[0071] 按照以下方式在所述多个第二标签列中的当前第二标签列上确定出数据箱的分位点：根据预设的分箱规则，确定出单个数据箱所包含的数据的目标数量；从所述当前第二标签列的起始位置出发，沿所述当前第二标签列每间隔目标数量个的数据标签的密文数据确定一个数据箱的分位点。其中，当前第二标签列对应多种特征数据中的当前特征数据。

[0072] 在一个实施例中，上述分别统计所述多个第二标签列中的各个第二标签列上的相邻的分位点之间的数据标签的密文数据的和，得到所述对应多种特征数据的数据箱中的数据标签的密文数据的和，具体实施时，可以包括：按照以下方式确定出对应多种特征数据中的当前特征数据的数据箱中的数据标签的密文数据的和：

[0073] 从多个第二标签列中确定出与当前特征数据对应的当前第二标签列；从所述当前第二标签列的起始位置出发，检测出当前第二标签列上的分位点；并将当前第二标签列上

相邻的两个分位点之间的数据标签的密文数据划分进一个数据箱，得到对应当前特征数据的多个数据箱；统计所述对应当前特征数据的多个数据箱中的各个数据箱的数据标签的密文数据的和，作为对应当前特征数据的数据箱中的数据标签的密文数据的和。

[0074] 这样可以先在上述当前第二标签列上确定出多个数据箱的分位点；再将当前第二标签列上的相邻的两个分位点之间的所包含的数据标签的密文数据划分进行一个数据箱。从而可以完成对应当前特征数据的数据分箱，得到对应当前特征数据的数据箱。按照类似的方式，可以根据预设的分箱规则，在其他第二标签列上分别确定出相应数据箱的分位点，完成对应多种特征数据的数据分箱，得到对应多种特征数据的数据箱。

[0075] 进一步，在针对多种特征数据中的各种特征数据的数据箱，分别计算对应各种特征数据的各个数据箱中数据标签的密文和，得到对应各种特征数据的数据箱的密文和，从而得到对应多种特征数据的数据箱中的数据标签密文数据的和。

[0076] 在一个实施例中，第二服务器可以将上述对应多种特征数据的数据箱中的数据标签的密文数据的和发送至第一服务器。

[0077] 在一个实施例中，第一服务器在接收到上述对应多种特征数据的数据箱中的数据标签的密文数据的和之后，可以对上述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密处理，得到对应多种特征数据的数据箱中的数据标签的明文数据的和。第一服务器可以将对应多种特征数据的数据箱中的数据标签的明文数据的和反馈给第二服务器。第二服务器可以获得对应多种特征数据的数据箱中的数据标签的明文数据的和。

[0078] 需要说明的是，由于数据标签的密文数据是通过采用同态加密算法对数据标签的明文数据加密得到的。因此，上述加密、解密的过程并不会影响最终得到的对应多种特征数据的数据箱中的数据标签的明文数据的和的具体数值。

[0079] 在一个实施例中，第二服务器还可以采用其他的方式来获取对应多种特征数据的数据箱中的数据标签的明文数据的和。

[0080] 在一个实施例中，在根据预设的分箱规则，分别在所述多个第二标签列中的各个第二标签列上确定出数据箱的分位点之后，所述方法具体实施时，还可以包括以下内容：在所述多个第二标签列中的各个第二标签列上标记出分位点，得到多个标记后的第二标签列；将所述多个标记后的第二标签列发送至第一服务器；其中，所述第一服务器用于根据所述多个标记后的第二标签列，计算出对应多种特征数据的数据箱中的数据标签的明文数据的和。

[0081] 在一个实施例中，第一服务器在接收到上述多个标记后的第二标签列后，可以在第一服务器一侧对多个标记后的第二标签列进行解密，得到多个解密后的第二标签列；其中，解密后的第二标签列包含有数据标签的明文数据。进一步，第一服务器可以根据多个标记后的第二标签列上标记出的数据箱的分位点，对各个标记后的第二标签列上的数据标签的明文数据分别进行数据分箱，得到对应多种特征数据的数据箱；其中，上述数据箱中包含的是数据标签的明文数据。进而第一服务器可以通过统计，确定出对应多种特征数据的数据箱中的数据标签的明文数据的和，再将对应多种特征数据的数据箱中的数据标签的明文数据的和发送给第二服务器。

[0082] 在一个实施例中，在得到多个标记后的第二标签列之后，所述方法具体实施时，还可以包括：对所述多个标记后的第二标签列进行随机化操作，得到随机化操作后的第二标

签列;相应的,将所述随机化操作后的第二标签列发送至第一服务器。

[0083] 在一个实施例中,上述随机化操作具体可以包括rerandom操作。具体实施时,上述对所述第二标签列中数据标签的密文数据进行随机化操作可以包括:在所述第二标签列中的数据标签的密文数据分别加上一个同态0的密文(例如,E(0),非确定性密文),得到随机化操作后的第二标签列,可以记为rerandom_label_cipher。

[0084] 通过上述随机化操作,可以使得随机化操作后的第二标签列中数据标签的密文数据在形式与第一标签列中的数据标签的密文数据存在差别,从而可以使得第一服务器无法根据第二标签列中数据标签的密文数据反推出各个数据标签的密文数据所对应的标识信息。能够有效地避免第一服务器根据第二标签列反推出对应不同标识信息的特征数据的数据值的排列顺序,从而可以避免向第一服务器泄露第二服务器所拥有的特征数据的相关信息,进一步更好地保护第二服务器一侧的数据隐私。此外,通过上述随机化操作,不影响第一服务器正常的解密处理。

[0085] S204:根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和,确定出特征数据的信息值。

[0086] 在一个实施例中,上述特征数据的信息值(Information Value,IV)具体可以理解为一种能够衡量特征数据整体的预测能力的参数值。通常如果特征数据的信息值越大,表明该组特征数据的预测能力越高,用于模型训练或者数据统计的效果越好。相反,如果特征数据的信息值越小,表明该组特征数据的预测能力越低,用于模型训练或者数据统计的效果越差。

[0087] 在一个实施例中,第二服务器在接收对应多种特征数据的数据箱中的数据标签的明文数据的和之后,可以结合对应多种特征数据的特征列,分别计算出多种特征数据中的各种特征数据的信息值。

[0088] 在一个实施例中,上述根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和,确定出特征数据的信息值,具体实施时,可以包括:按照以下方式确定出多种特征数据中的当前特征数据的信息值:

[0089] 根据与当前特征数据对应的特征列,确定出对应当前特征数据的数据箱中的特征数据;根据对应当前特征数据的数据箱中的数据标签的明文数据的和、所述对应当前特征数据的数据箱中的特征数据,计算对应当前特征数据的各个数据箱的权重证明;根据所述对应当前特征数据的各个数据箱的权重证明,计算对应当前特征数据的数据箱的信息值;根据所述对应当前特征数据的数据箱的信息值,计算当前特征数据的信息值。

[0090] 其中,上述数据箱的权重证明(Weight of Evidence,WOE)具体可以理解为对数据箱中所包含的特征数据的编码形式。

[0091] 在一个实施例中,具体的,针对当前特征数据,可以先根据数据箱中的数据标签的明文数据的和,计算出各个数据箱中所包含的目标标签的数量。其中,上述目标标签根据具体情况,可以是正标签,也可以是负标签。再根据各个数据箱中所包含的目标标签的数量、各个数据箱中所包含的特征数据的数量,按照以下算式计算出各个数据箱的权重证明:

[0092] $$WOE(i)=\ln(\frac{\#y_i/\#y_T}{\#n_i/\#n_T})=\ln(\frac{\#y_i/\#n_i}{\#y_T/\#n_T})$$

[0093] 其中，i为对应当前特征数据的多个数据箱中的数据箱的编号，$WOE(i)$具体可以表示为编号为i的数据箱的权重证明，$\#y_i$具体可以表示为编号为i的数据箱中对应目标标签的特征数据的数量(即目标标签的数量)，$\#y_T$具体可以表示为所有数据箱中对应目标标签的特征数据的数量，$\#n_i$具体可以表示为编号为i的数据箱中对应的数据标签不是目标标签的特征数据的数量，$\#n_T$具体可以表示为所有数据箱中对应的数据标签不是目标标签的特征数据的数量。

[0094] 进一步，可以按照以下算式，根据数据箱的权重证明，计算对应当前特征数据的多个数据箱中各个数据箱的信息值：$IV(i)=(\#y_i/\#y_T-\#n_i/\#n_T)\times WOE(i)$。其中，$IV(i)$具体可以表示为编号为i的数据箱的信息值。

[0095] 接着，可以按照以下算式，根据多个数据箱的信息值，计算出当前特征数据的信息值：$IV=\sum_i^N IV(i)$。其中，IV具体可以表示为特征数据的信息值，N具体可以表示为数据箱的个数。

[0096] 按照上述类型方式，第二服务器可以分别计算出所持有的多种特征数据中各种特征数据的信息值。

[0097] 通过上述方式，参与数据处理的第一服务器和第二服务器可以在不向对方泄露己方所持有的数据信息，或者只需要向对方泄露较少的数据信息的前提下，通过合作，安全地确定出第二服务器所持有的多种特征数据中的各种特征数据的信息值。

[0098] S205：根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据。

[0099] 在一个实施例中，上述相关性系数具体可以用于表征两种特征数据之间的线性相关程度。通常，如果两种特征数据之间的相关性系数的数值越大，说明两种特征数据之间的线性相关程度越高；相应的，两种特征数据的使用效果、作用的差异性越小，重叠率越高。相反，如果两种特征数据之间的相关性系数的数值越小，说明两种特征数据之间的线性相关程度越低，相应的，两种特征数据的使用效果、作用的差异性越大，重叠率越低。

[0100] 在一个实施例中，上述相关性系数具体可以按照以下方式获取：将所述多种特征数据进行两两组合，得到多个特征数据组合；其中，多个特征数据组合中的各个特征数据组合分别包含有两种不同的特征数据；根据特征数据组合中所包含的两种不同的特征数据，通过计算皮尔逊相关系数，确定出该组合中两种特征数据之间的相关性系数。

[0101] 当然，上述所列举的计算特征数据之间的相关性系数的方式只是一种示意性说明。具体实施时，根据具体情况和处理需求，还可以采用其他合适的方式来计算特征数据之间的相关性系数。对此，本说明书不作限定。

[0102] 在一个实施例中，上述根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据，具体实施时，可以包括以下内容：从多种特征数据中，筛选出特征数据之间的相关性系数大于预设的相关性系数阈值的特征数据组合，作为筛选组；比较所述筛选组中的特征数据的信息值，保留所述筛选组中的信息值最大的特征数据作为所述符合要求的特征数据。

[0103] 其中，上述预设的相关性系数阈值的具体数值可以根据第一服务器，或第三方对

特征数据的要求灵活设置。对此，本说明书不作限定。

[0104] 在本实施例中，通过上述方式可以根据相关性系数从多种特征数据中使用效果、作用的差异性较小的特征数据组合；再根据该特征数据组合中两种特征数据的信息值，只保留下信息值较大的特征数据，去除信息值较小的特征数据，从而可以过滤掉使用效果、作用相近，重叠率较高的特征数据，保留下使用效果较好的特征数据，筛选出符合要求的特征数据。

[0105] 这样后续第一服务器，或者第三方需要使用特征数据时，第二服务器可以将所持有的多种特征数据中的符合要求的特征数据提供给第一服务器，或者第三方使用，以使得第一服务器，或者第三方可以得到效果更好的特征数据进行诸如联合统计或者联合建模等数据处理。

[0106] 具体的，例如，参阅图4所示，第二服务器持有三种特征数据，分别为第一种特征数据、第二种特征数据和第三种特征数据。可以按照上述方式通过两两组合，得到共三个不同特征数据组合，分别记为为：组合1(包含有第一种特征数据和第二种特征数据)、组合2(包含有第一种特征数据和第三种特征数据)、组合3(包含有第三种特征数据和第二种特征数据)。

[0107] 进而可以分别基于上述三个特征数据组合计算得到特征数据之间的相关性系数，包括：第一种特征数据和第二种特征数据之间的相关性系数1、第一种特征数据和第三种特征数据之间的相关性系数2、第三种特征数据和第二种特征数据之间的相关性系数3。

[0108] 再将上述三个相关性系数分别与预设的相关性系数阈值进行比较，发现只有相关性系数1大于上述预设的相关性系数阈值。因此，可以将组合1确定为筛选组。

[0109] 进一步，可以比较上述筛选组中所包含的第一种特征数据的信息值(记为IV1)与第二种特征数据的信息值(记为IV2)，发现IV1大于IV2，相应的可以保留该筛选组中的第一种特征数据，去除第二种特征数据。

[0110] 通过上述方式，最终保留下符合要求的特征数据可以包括：第一种特征数据和第三种特征数据这两种特征数据。

[0111] 由上可见，基于本说明书实施例提供的基于隐私保护的数据处理方法，持有多种特征数据的第二服务器在接收到包含有按照标识信息排列的数据标签的密文数据的第一标签列后，可以根据多种特征数据的数据值的排列顺序对第一标签列进行多种排列，得到对应多种特征数据的多个第二标签列；再基于上述多个第二标签列和预设的分箱规则，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和，并反馈给第二服务器：第二服务器通过解密，得到并向第一服务器反馈相应的明文数据的和。进而第一服务器可以根据上述明文数据的和，计算出各种特征数据的信息值；并根据特征数据的信息值、特征数据之间的相关性系数，筛选出符合要求的特征数据。从而可以在不泄露双方所各自持有的数据信息、保护双方的数据隐私的前提下，安全、高效地从多种特征数据中筛选出符合要求的特征数据。

[0112] 参阅图5所示，本说明书实施例还提供了一种基于隐私保护的数据处理方法。该方法具体应用于第一服务器一侧。该方法具体实施时，可以包括以下内容。

[0113] S501：对标签列进行加密处理，得到第一标签列，并将所述第一标签列发送至第二服务器；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标

签列中的数据标签的密文数据根据标识信息的排列顺序排列。

[0114] S502:接收第二服务器发送的对应多种特征数据的数据箱中的数据标签的密文数据的和;其中,所述第二服务器持有与标识信息对应的多种特征数据;所述第二服务器根据所述多种特征数据对所述第一标签列中的数据标签的密文数据进行多种排列,得到对应的多个第二标签列;所述第二服务器根据预设的分箱规则和所述多个第二标签列,确定出对应多种特征数据的数据箱中的数据标签的密文数据的和。

[0115] S503:对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密,得到对应多种特征数据的数据箱中的数据标签的明文数据的和。

[0116] S504:将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器;其中,所述第二服务器根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和,确定出特征数据的信息值;所述第二服务器根据特征数据的信息值,以及特征数据之间的相关性系数,从多种特征数据中筛选出符合要求的特征数据。

[0117] 在一个实施例中,上述对标签列进行加密处理,具体可以包括:利用同态加密算法对所述标签列中的数据标签进行加密处理。其中,所述同态加密算法具体可以包括elgamal同态加密算法等。

[0118] 在一个实施例中,在将所述第一标签列发送至第二服务器之后,所述方法具体实施时,还可以包括以下内容:接收第二服务器发送的多个标记后的第二标签列;根据所述多个标记后的第二标签列,计算对应多种特征数据的数据箱中的数据标签的明文数据的和;将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器。

[0119] 在一个实施例中,上述根据所述多个标记后的第二标签列,计算对应多种特征数据的数据箱中的数据标签的明文数据的和,具体实施时,可以包括以下内容:按照以下方式计算出对应多种特征数据中的当前特征数据的数据标签的明文数据的和:从多个第二标签列中确定出与当前特征数据对应的当前第二标签列;对所述当前第二标签列中的数据标签的密文数据进行解密,得到解密后的当前第二标签列;其中,所述解密后的当前第二标签列包含有数据标签的明文数据;从所述解密后的当前第二标签列的起始位置出发,检测出所述解密后的当前第二标签列上的分位点;并将相邻的两个分位点之间的数据标签的明文数据划分为一个数据箱,得到对应当前特征数据的多个数据箱;统计所述对应当前特征数据的多个数据箱中的各个数据箱的数据标签的明文数据的和,作为对应当前特征数据的数据箱中的数据标签的明文数据的和。

[0120] 在一个实施例中,在将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器之后,所述方法具体实施时,还可以包括:接收第二服务器发送的符合要求的特征数据;利用所述符合要求的特征数据、所述数据标签,进行模型训练,以建立得到目标模型。

[0121] 在一个实施例中,在将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器之后,所述方法具体实施时,还可以包括:以数据标签作为输出,与以所述符合要求的特征数据作为输出的第二服务器,通过多方安全计算,来建立目标模型。

[0122] 通过上述采用安全多方计算的方式来联合进行模型训练,第二服务器无法直接获取得到第一服务器持有的数据标签,同时第一服务器也无法直接获取得到第二服务器持有的特征数据,从而可以更加安全地利用双方所持有的数据来训练目标模型。

[0123] 在一个实施例中，在用户的信用风险检测场景中，上述特征数据具体可以包括以下至少之一：用户的月收入数据、用户的月支出数据、用户的年龄数据等。当然，上述所列举的特征数据只是一种示意性说明。具体实施时，根据具体的应用场景和处理需求，上述特征数据还可以包括例如用户的违约次数等其他数值型数据。对此，本说明书不作限定。

[0124] 在一个实施例中，上述目标模型具体可以包括用户信用风险预测模型。具体的，在用户的信用风险检测场景中，第一服务器或者其他第三方可以按照上述方式使用第二服务器所持有的符合要求的特征数据训练得到上述目标模型。进而可以利用上述目标模型根据用户的属性数据检测用户是否存在信用风险，并对存在信用风险的用户设置对应的风险标记。

[0125] 在一个实施例中，在根据所述多个标记后的第二标签列，计算对应多种特征数据的数据箱中的数据标签的明文数据的和之后，所述方法具体实施时，还可以包括以下内容：利用私钥数据对所述对应多种特征数据的数据箱中的数据标签的明文数据的和进行加密处理；将所述加密后的对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器；其中，所述第二服务器持有对应的公钥数据。

[0126] 在本实施例中，具体实施前，第一服务器可以预先生成相互对应的公钥数据和私钥数据，并将上述公钥数据发送至第二服务器。

[0127] 由上可见，基于本说明书实施例提供的基于隐私保护的数据处理方法，可以在不泄露双方所各自持有的数据信息、保护双方的数据隐私的前提下，安全、高效地从多种特征数据中筛选出符合要求的特征数据。

[0128] 本说明书实施例还提供一种服务器，包括处理器以及用于存储处理器可执行指令的存储器，所述处理器具体实施时可以根据指令执行以下步骤：接收第一服务器发送的第一标签列；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列；根据多个特征列，对所述第一标签列中的数据标签的密文数据进行多种排列，得到对应的多个第二标签列；其中，所述第二标签列与一种特征数据对应；所述特征列分别包含有一种特征数据，且所包含的特征数据根据特征数据的数据值排列；根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和；并将所述对应多种特征数据的数据箱中的数据标签的密文数据的和发送至第一服务器；其中，所述第一服务器对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和；根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出特征数据的信息值；根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据。

[0129] 为了能够更加准确地完成上述指令，参阅图6所示，本说明书实施例还提供了另一种具体的服务器，其中，所述服务器包括网络通信端口601、处理器602以及存储器603，上述结构通过内部线缆相连，以便各个结构可以进行具体的数据交互。

[0130] 其中，所述网络通信端口601，具体可以用于接收第一服务器发送的第一标签列；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列。

[0131] 所述处理器602，具体可以用于根据多个特征列，对所述第一标签列中的数据标签

的密文数据进行多种排列，得到对应的多个第二标签列；其中，所述第二标签列与一种特征数据对应；所述特征列分别包含有一种特征数据，且所包含的特征数据根据特征数据的数据值排列；根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和；并将所述对应多种特征数据的数据箱中的数据标签的密文数据的和发送至第一服务器；其中，所述第一服务器对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和；根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出特征数据的信息值；根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据。

[0132] 所述存储器603，具体可以用于存储相应的指令程序。

[0133] 在本实施例中，所述网络通信端口601可以是与不同的通信协议进行绑定，从而可以发送或接收不同数据的虚拟端口。例如，所述网络通信端口可以是负责进行web数据通信的端口，也可以是负责进行FTP数据通信的端口，还可以是负责进行邮件数据通信的端口。此外，所述网络通信端口还可以是实体的通信接口或者通信芯片。例如，其可以为无线移动网络通信芯片，如GSM、CDMA等；其还可以为Wifi芯片；其还可以为蓝牙芯片。

[0134] 在本实施例中，所述处理器602可以按任何适当的方式实现。例如，处理器可以采取例如微处理器或处理器以及存储可由该(微)处理器执行的计算机可读程序代码(例如软件或固件)的计算机可读介质、逻辑门、开关、专用集成电路(Application Specific Integrated Circuit，ASIC)、可编程逻辑控制器和嵌入微控制器的形式等等。本说明书并不作限定。

[0135] 在本实施例中，所述存储器603包括多个层次，在数字系统中，只要能保存二进制数据的都可以是存储器；在集成电路中，一个没有实物形式的具有存储功能的电路也叫存储器，如RAM、FIFO等；在系统中，具有实物形式的存储设备也叫存储器，如内存条、TF卡等。

[0136] 本说明书实施例还提供了一种基于上述基于隐私保护的数据处理方法的计算机存储介质，所述计算机存储介质存储有计算机程序指令，在所述计算机程序指令被执行时实现：接收第一服务器发送的第一标签列；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列；根据多个特征列，对所述第一标签列中的数据标签的密文数据进行多种排列，得到对应的多个第二标签列；其中，所述第二标签列与一种特征数据对应；所述特征列分别包含有一种特征数据，且所包含的特征数据根据特征数据的数据值排列；根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和；并将所述对应多种特征数据的数据箱中的数据标签的密文数据的和发送至第一服务器；其中，所述第一服务器对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和；根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出特征数据的信息值；根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据。

[0137] 在本实施例中，上述存储介质包括但不限于随机存取存储器(Random Access Memory, RAM)、只读存储器(Read-Only Memory, ROM)、缓存(Cache)、硬盘(Hard Disk

Drive, HDD)或者存储卡(Memory Card)。所述存储器可以用于存储计算机程序指令。网络通信单元可以是依照通信协议规定的标准设置的,用于进行网络连接通信的接口。

[0138] 在本实施例中,该计算机存储介质存储的程序指令具体实现的功能和效果,可以与其它实施方式对照解释,在此不再赘述。

[0139] 本说明书实施例还提供另一种服务器,包括处理器以及用于存储处理器可执行指令的存储器,所述处理器具体实施时可以根据指令执行以下步骤:对标签列进行加密处理,得到第一标签列,并将所述第一标签列发送至第二服务器;其中,所述第一标签列包含有与标识信息对应的数据标签的密文数据,所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列;接收第二服务器发送的对应多种特征数据的数据箱中的数据标签的密文数据的和;其中,所述第二服务器持有与标识信息对应的多种特征数据;所述第二服务器根据所述多种特征数据对所述第一标签列中的数据标签的密文数据进行多种排列,得到对应的多个第二标签列;所述第二服务器根据预设的分箱规则和所述多个第二标签列,确定出对应多种特征数据的数据箱中的数据标签的密文数据的和;对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密,得到对应多种特征数据的数据箱中的数据标签的明文数据的和;将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器;其中,所述第二服务器根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和,确定出特征数据的信息值;所述第二服务器根据特征数据的信息值,以及特征数据之间的相关性系数,从多种特征数据中筛选出符合要求的特征数据。

[0140] 参阅图7所示,在软件层面上,本说明书实施例还提供了一种基于隐私保护的数据处理装置,该装置具体可以包括以下的结构模块。

[0141] 接收模块701,具体可以用于接收第一服务器发送的第一标签列;其中,所述第一标签列包含有与标识信息对应的数据标签的密文数据,所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列;

[0142] 排列模块702,具体可以用于根据多个特征列,对所述第一标签列中的数据标签的密文数据进行多种排列,得到对应的多个第二标签列;其中,所述第二标签列与一种特征数据对应;所述特征列分别包含有一种特征数据,且所包含的特征数据根据特征数据的数据值排列;

[0143] 第一确定模块703,具体可以用于根据预设的分箱规则和所述多个第二标签列,确定出对应多种特征数据的数据箱中的数据标签的密文数据的和;并将所述对应多种特征数据的数据箱中的数据标签的密文数据的和发送至第一服务器;其中,所述第一服务器对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密,得到对应多种特征数据的数据箱中的数据标签的明文数据的和;

[0144] 第二确定模块704,具体可以用于根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和,确定出特征数据的信息值;

[0145] 筛选模块705,具体可以用于根据特征数据的信息值,以及特征数据之间的相关性系数,从多种特征数据中筛选出符合要求的特征数据。

[0146] 需要说明的是,上述实施例阐明的单元、装置或模块等,具体可以由计算机芯片或实体实现,或者由具有某种功能的产品来实现。为了描述的方便,描述以上装置时以功能分为各种模块分别描述。当然,在实施本说明书时可以把各模块的功能在同一个或多个软件

和/或硬件中实现，也可以将实现同一功能的模块由多个子模块或子单元的组合实现等。以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，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。

[0147] 本说明书实施例还提供了另一种基于隐私保护的数据处理装置。参阅图8所示，该装置具体可以包括以下的结构模块。

[0148] 加密模块801，具体可以用于对标签列进行加密处理，得到第一标签列，并将所述第一标签列发送至第二服务器；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列；

[0149] 接收模块802，具体可以用于接收第二服务器发送的对应多种特征数据的数据箱中的数据标签的密文数据的和；其中，所述第二服务器持有与标识信息对应的多种特征数据；所述第二服务器根据所述多种特征数据对所述第一标签列中的数据标签的密文数据进行多种排列，得到对应的多个第二标签列；所述第二服务器根据预设的分箱规则和所述多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和；

[0150] 解密模块803，具体可以用于对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和；

[0151] 发送模块804，具体可以用于将所述对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器；其中，所述第二服务器根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出特征数据的信息值；所述第二服务器根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据。

[0152] 由上可见，本说明书实施例提供的基于隐私保护的数据处理装置，可以在不泄露双方所各自持有的数据信息、保护双方的数据隐私的前提下，安全、高效地从多种特征数据中筛选出符合要求的特征数据。

[0153] 虽然本说明书提供了如实施例或流程图所述的方法操作步骤，但基于常规或者无创造性的手段可以包括更多或者更少的操作步骤。实施例中列举的步骤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。在实际中的装置或客户端产品执行时，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，甚至为分布式数据处理环境）。术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、产品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、产品或者设备所固有的要素。在没有更多限制的情况下，并不排除在包括所述要素的过程、方法、产品或者设备中还存在另外的相同或等同要素。第一，第二等词语用来表示名称，而并不表示任何特定的顺序。

[0154] 本领域技术人员也知道，除了以纯计算机可读程序代码方式实现控制器以外，完全可以通过将方法步骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种

硬件部件，而对其内部包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至，可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又可以是硬件部件内的结构。

[0155] 本说明书可以在由计算机执行的计算机可执行指令的一般上下文中描述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构、类等等。也可以在分布式计算环境中实践本说明书，在这些分布式计算环境中，由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中，程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0156] 通过以上的实施例的描述可知，本领域的技术人员可以清楚地了解到本说明书可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解，本说明书的技术方案本质上可以以软件产品的形式体现出来，该计算机软件产品可以存储在存储介质中，如ROM/RAM、磁碟、光盘等，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，移动终端，服务器，或者网络设备等)执行本说明书各个实施例或者实施例的某些部分所述的方法。

[0157] 本说明书中的各个实施例采用递进的方式描述，各个实施例之间相同或相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。本说明书可用于众多通用或专用的计算机系统环境或配置中。例如：个人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处理器系统、基于微处理器的系统、置顶盒、可编程的电子设备、网络PC、小型计算机、大型计算机、包括以上任何系统或设备的分布式计算环境等等。

[0158] 虽然通过实施例描绘了本说明书，本领域普通技术人员知道，本说明书有许多变形和变化而不脱离本说明书的精神，希望所附的权利要求包括这些变形和变化而不脱离本说明书的精神。

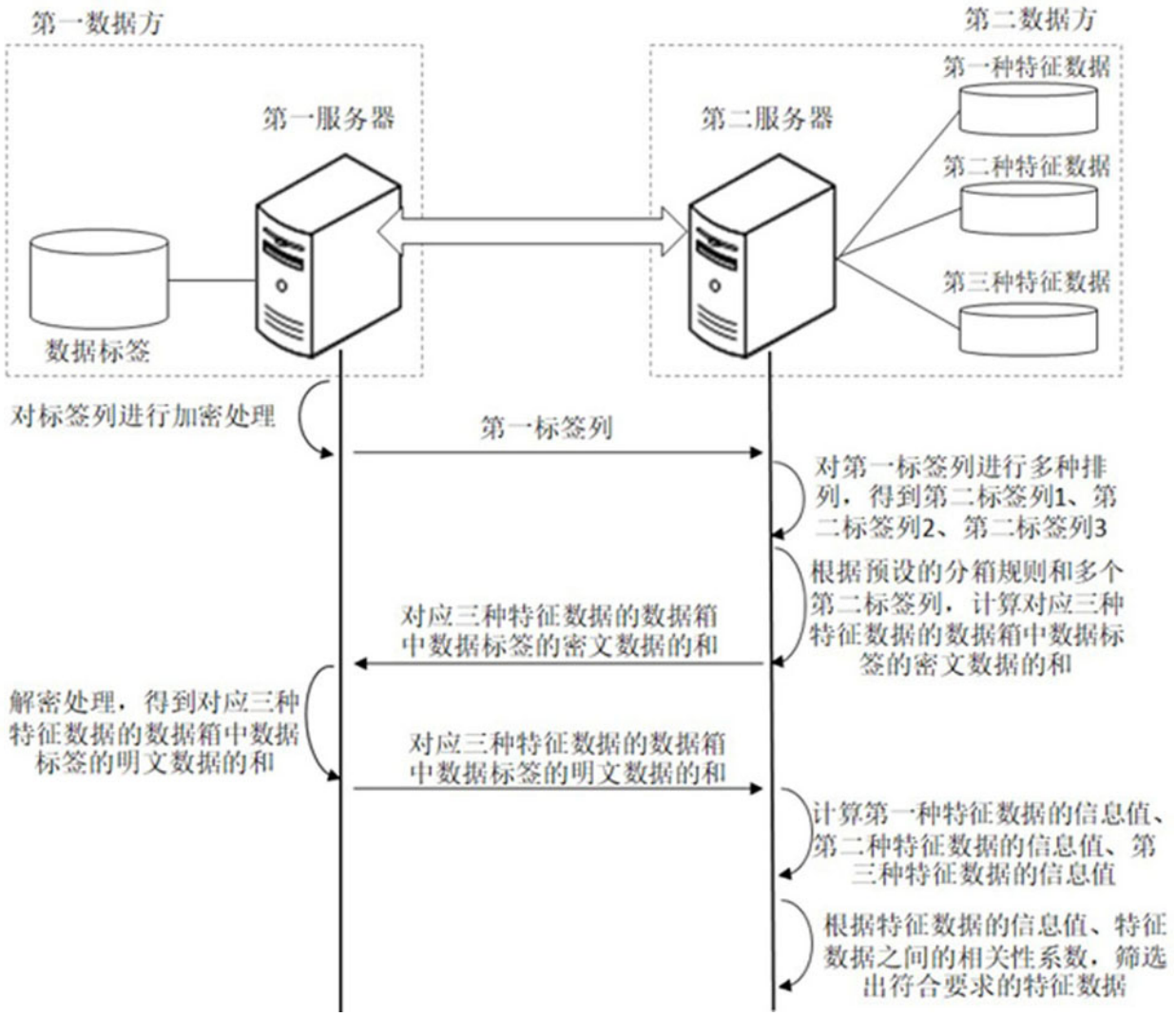

第一数据方
第二数据方
第一种特征数据
第一服务器
第二服务器
第二种特征数据
第三种特征数据
数据标签
对标签列进行加密处理
第一标签列
对第一标签列进行多种排列，得到第二标签列1、第二标签列2、第二标签列3
根据预设的分箱规则和多个第二标签列，计算对应三种特征数据的数据箱中数据标签的密文数据的和
对应三种特征数据的数据箱中数据标签的密文数据的和
解密处理，得到对应三种特征数据的数据箱中数据标签的明文数据的和
对应三种特征数据的数据箱中数据标签的明文数据的和
计算第一种特征数据的信息值、第二种特征数据的信息值、第三种特征数据的信息值
根据特征数据的信息值、特征数据之间的相关性系数，筛选出符合要求的特征数据

图1

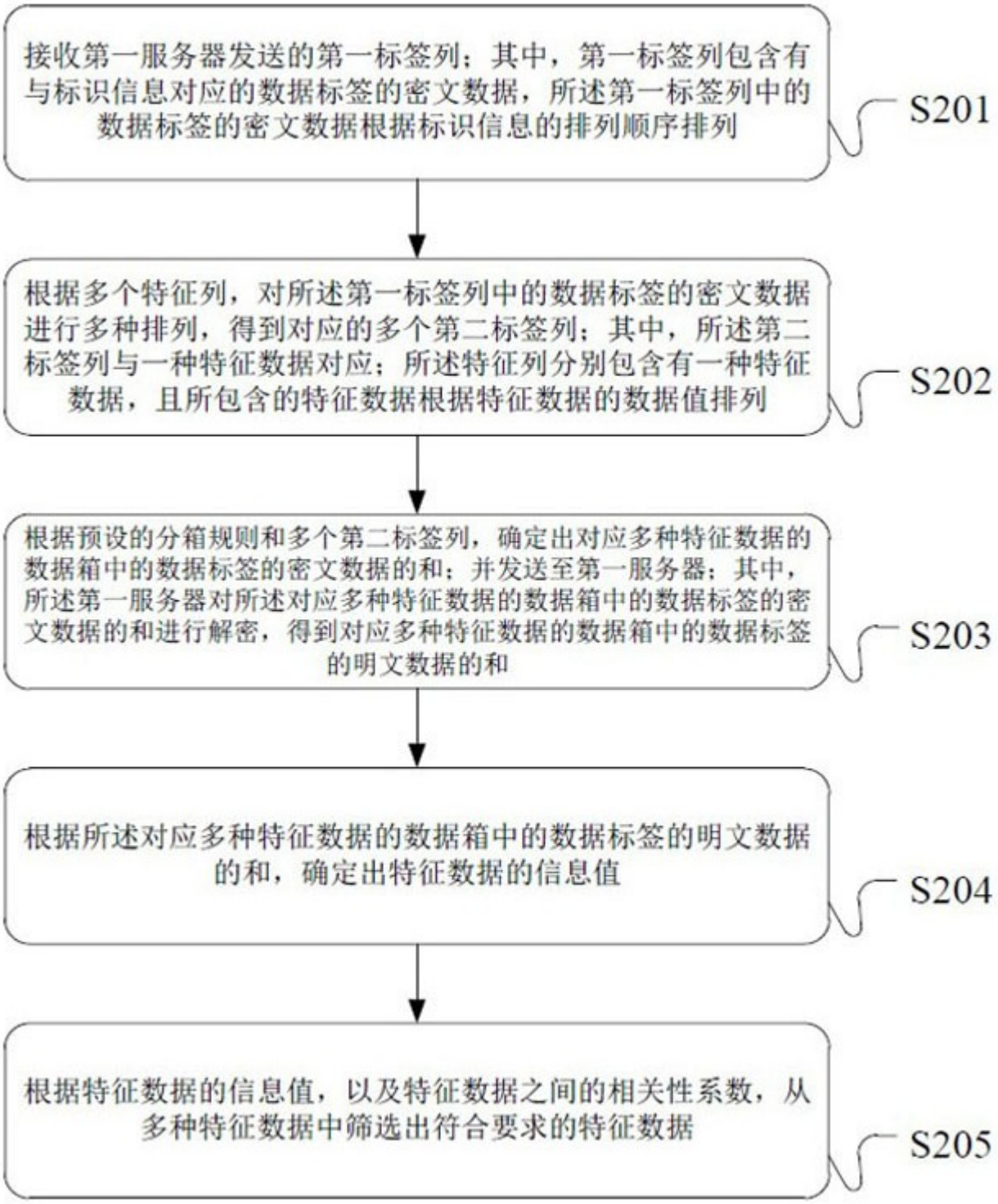

接收第一服务器发送的第一标签列；其中，第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列
S201
根据多个特征列，对所述第一标签列中的数据标签的密文数据进行多种排列，得到对应的多个第二标签列；其中，所述第二标签列与一种特征数据对应；所述特征列分别包含有一种特征数据，且所包含的特征数据根据特征数据的数据值排列
S202
根据预设的分箱规则和多个第二标签列，确定出对应多种特征数据的数据箱中的数据标签的密文数据的和；并发送至第一服务器；其中，所述第一服务器对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和
S203
根据所述对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出特征数据的信息值
S204
根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据
S205

图2

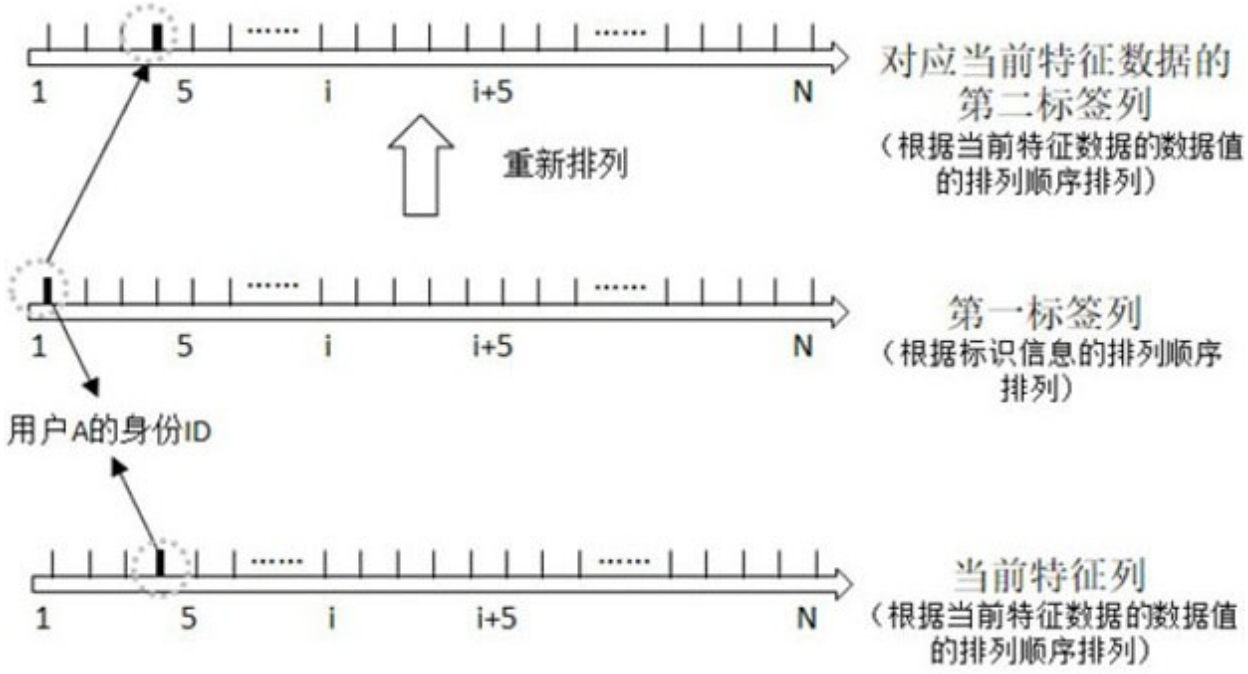

1
5
i
i+5
N
重新排列
对应当前特征数据的第二标签列
（根据当前特征数据的数据值的排列顺序排列）
第一标签列
（根据标识信息的排列顺序排列）
用户A的身份ID
当前特征列
（根据当前特征数据的数据值的排列顺序排列）

图3

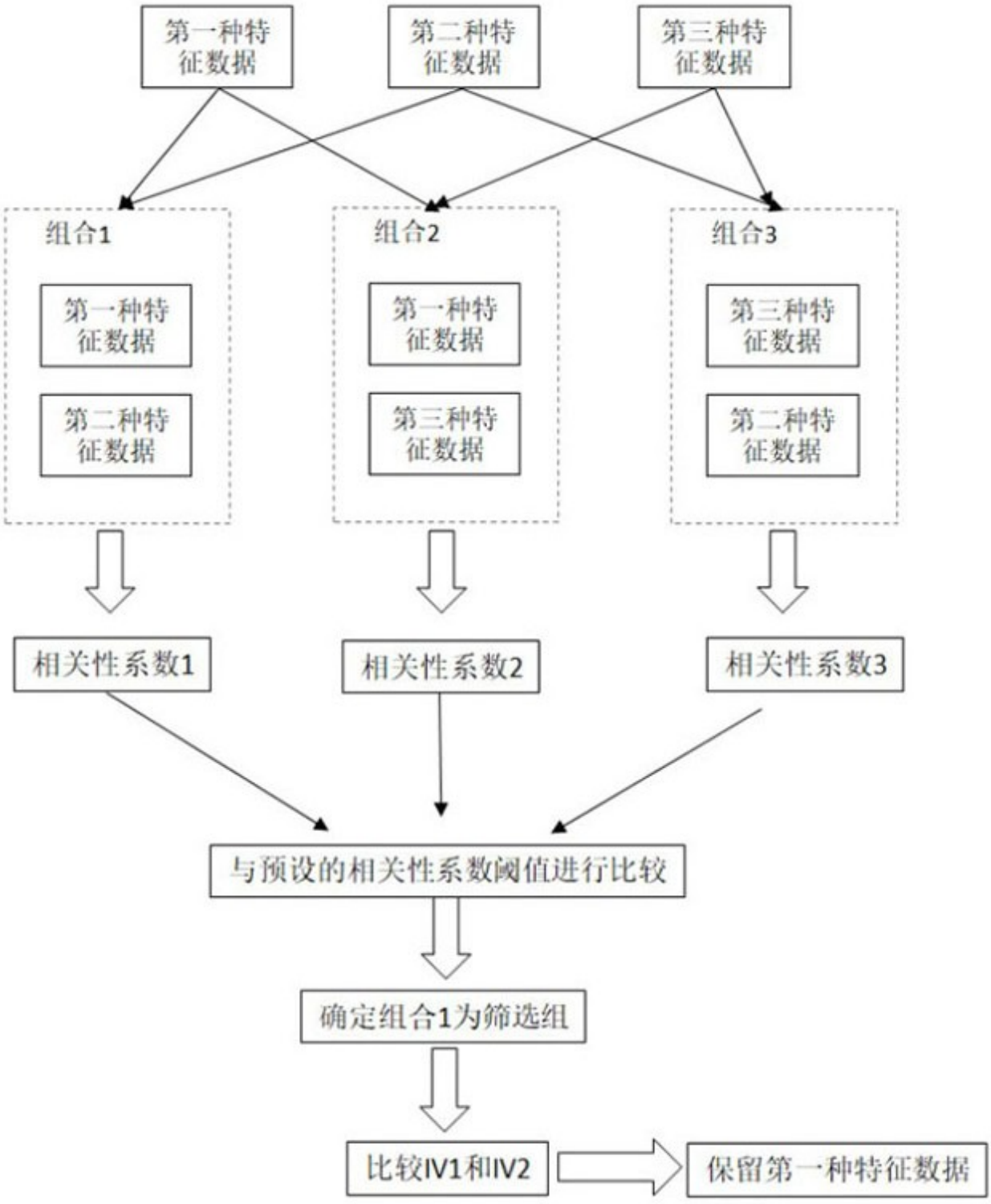

第一种特征数据
第二种特征数据
第三种特征数据
组合1
第一种特征数据
第二种特征数据
组合2
第一种特征数据
第三种特征数据
组合3
第三种特征数据
第二种特征数据
相关性系数1
相关性系数2
相关性系数3
与预设的相关性系数阈值进行比较
确定组合1为筛选组
比较IV1和IV2
保留第一种特征数据

图4

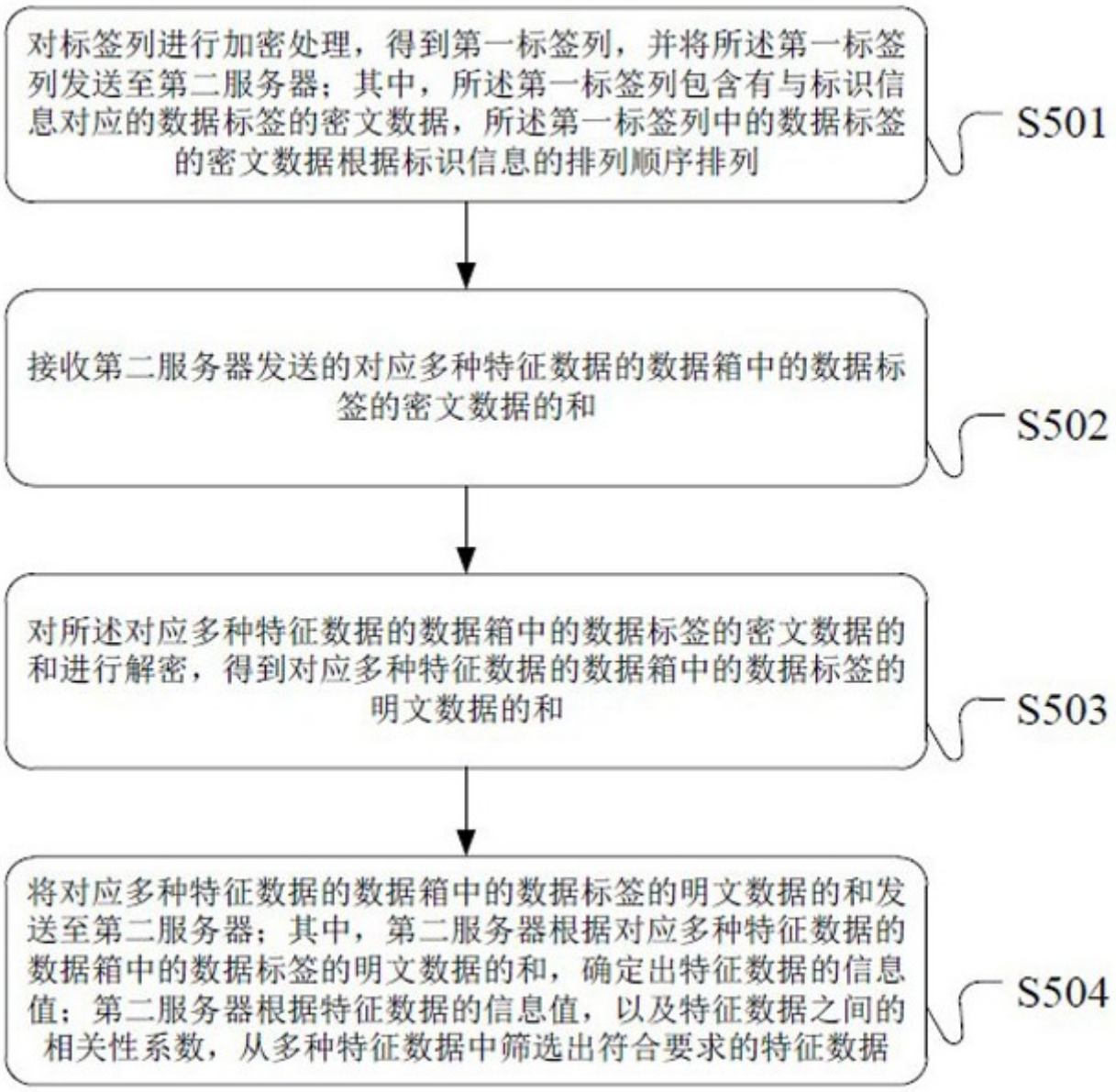

对标签列进行加密处理，得到第一标签列，并将所述第一标签列发送至第二服务器；其中，所述第一标签列包含有与标识信息对应的数据标签的密文数据，所述第一标签列中的数据标签的密文数据根据标识信息的排列顺序排列
S501
接收第二服务器发送的对应多种特征数据的数据箱中的数据标签的密文数据的和
S502
对所述对应多种特征数据的数据箱中的数据标签的密文数据的和进行解密，得到对应多种特征数据的数据箱中的数据标签的明文数据的和
S503
将对应多种特征数据的数据箱中的数据标签的明文数据的和发送至第二服务器；其中，第二服务器根据对应多种特征数据的数据箱中的数据标签的明文数据的和，确定出特征数据的信息值；第二服务器根据特征数据的信息值，以及特征数据之间的相关性系数，从多种特征数据中筛选出符合要求的特征数据
S504

图5

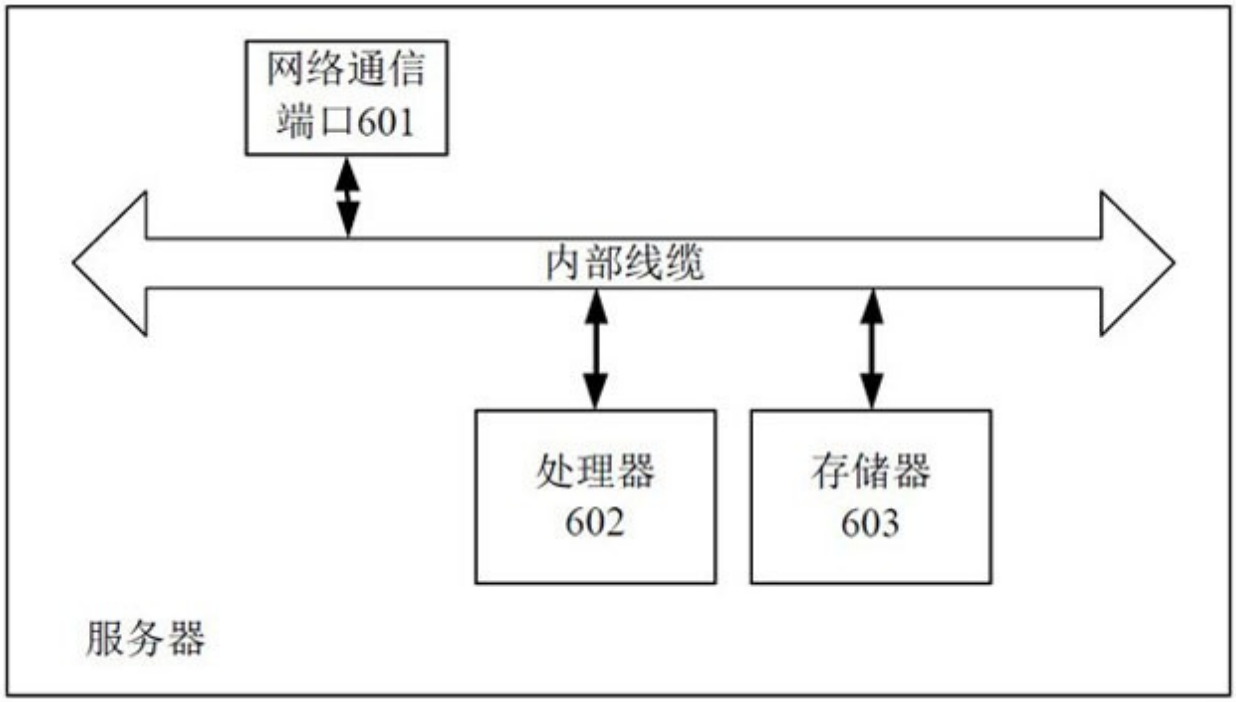


图6

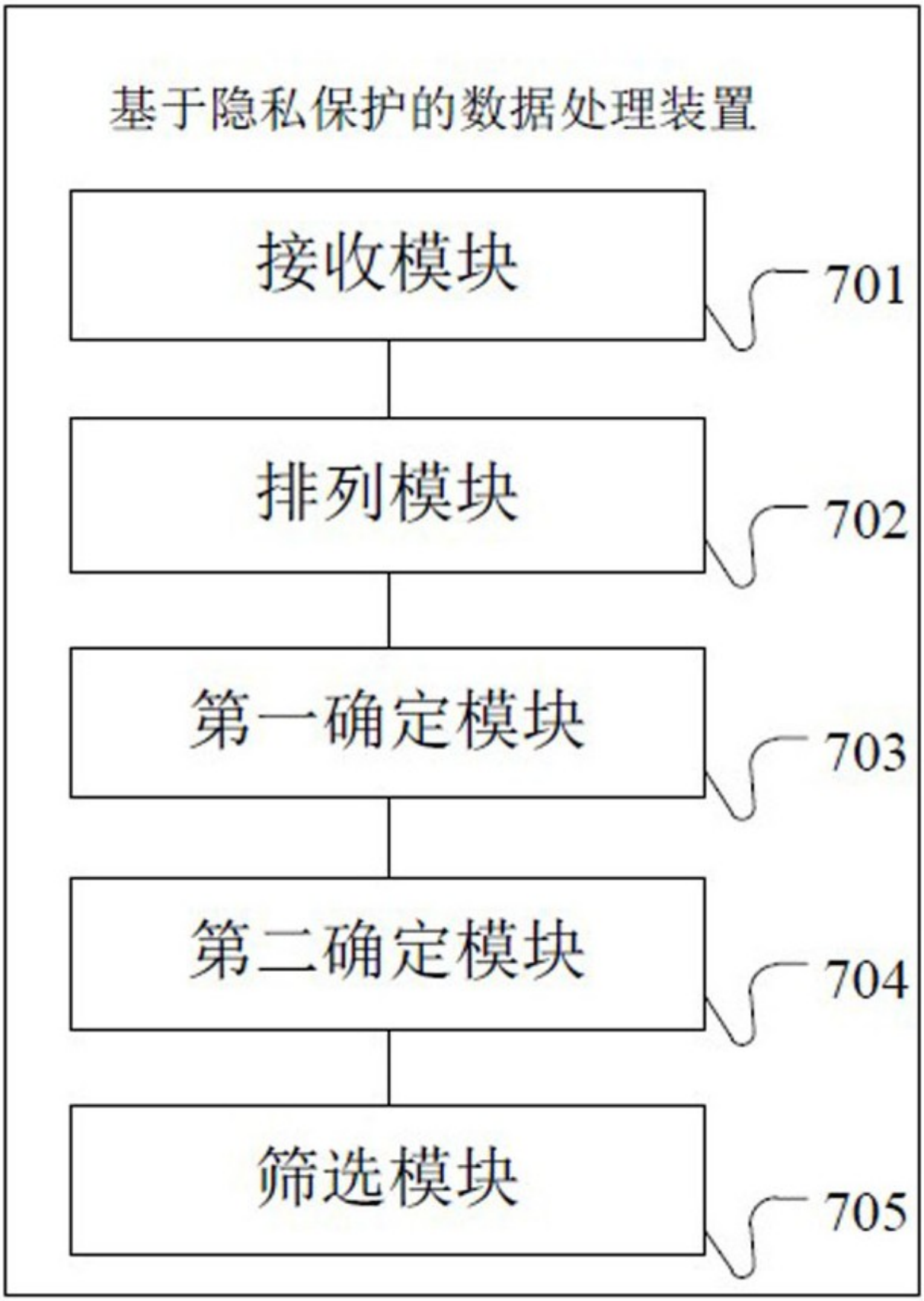

基于隐私保护的数据处理装置
接收模块
701
排列模块
702
第一确定模块
703
第二确定模块
704
筛选模块
705

图7

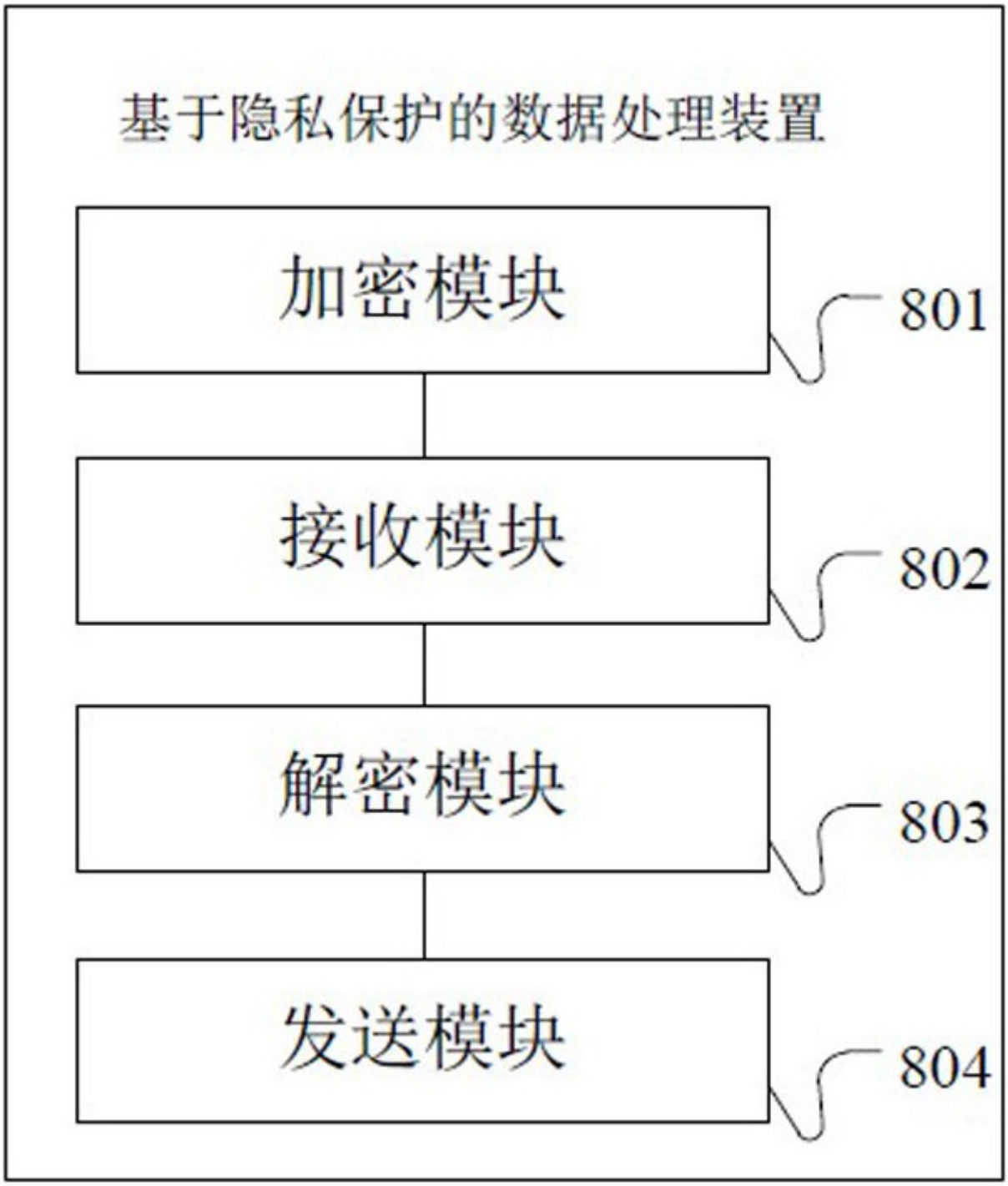


图8