



(12) 发明专利申请

(10) 申请公布号 CN 103136417 A

(43) 申请公布日 2013.06.05

(21) 申请号 201210509942.3

(22) 申请日 2012.12.03

(30) 优先权数据

13/308,745 2011.12.01 US

(71) 申请人 提升科技有限公司

地址 美国德克萨斯州普莱诺市水仙花路
4533 号

(72) 发明人 孟召兵 金祖胜 陈友明 宁平

(74) 专利代理机构 南京经纬专利商标代理有限
公司 32200

代理人 楼高潮

(51) Int. Cl.

G06F 17/50 (2006.01)

权利要求书4页 说明书7页 附图14页

(54) 发明名称

用于建模分析的基于互联网的托管系统和计算机可读介质

(57) 摘要

一个基于互联网的系统 and 包括代码的计算机可读的介质,用于确定与接收到的数据集相关的多个建模分析,其中至少有一个接收到的数据集是通过互联网接收的,接收多个建模分析任务的说明,通过互联网收集建模分析任务的说明,基于确定的多个建模分析任务和接收到的建模分析任务的说明在接收到的数据集上执行建模分析并通过互联网递送建模分析结果。

1. 一个储存在计算机可读介质上并用处理器来执行的基于互联网的计算机程序,此计算机程序包括计算机可读的程序代码,所述代码用于:

确定与至少一个接收到的数据集相关的多个建模分析任务,其中所述至少一个接收到的数据集通过互联网接收;

接收所述多个建模分析任务的说明;

通过互联网来收集所述建模分析任务的说明;

部分根据所述确定的多个建模分析任务以及所述建模分析任务的所述接收到的说明在所述的至少一个接收的数据集上执行建模分析;并

通过互联网递送所述建模分析的结果。

2. 根据权利要求 1 所述的计算机程序,其特征在于,其进一步包括:

收集所述至少一个接收到的数据集的目标变量名称;

收集所述建模分析任务的预测建模类型名称;

收集所述至少一个接收到的数据集的数据集格式;

在所述的至少一个接收到的数据集上进行预测建模,其部分基于所述多个确定的建模分析任务、所述数据集格式、所述收集的目标变量名称、所述预测建模类型名称、导致一个预测模型的预测建模;

通过互联网递送所述预测模型;

根据所述至少一个接收到的数据集评价预测模型;并

通过互联网递送所述预测模型评价结果。

3. 根据权利要求 2 所述的计算机程序,其特征在于,其进一步包括存储应用于多个接收到的数据集的所述预测模型。

4. 根据权利要求 2 所述的计算机程序,其特征在于,其进一步包括存储所述至少一个收到的数据集,以应用多个预测模型至所述接收到的数据集。

5. 根据权利要求 2 所述的计算机程序,其特征在于,所述预测建模包括回归建模分析。

6. 根据权利要求 2 所述的计算机程序,其特征在于,所述预测建模包括多个预测模型,其优化所述预测模型在所述至少一个接收到的数据集上的多个匹配。

7. 根据权利要求 2 所述的计算机程序,其特征在于,其进一步包含收集以下中的至少一个:

一组应用于所述预测模型的建模预设置;

用于所述预测建模递送的递送方式预设置;

所述至少一个接收到的数据集的数据共享预设置;以及

所述预测模型的结果共享预设置。

8. 根据权利要求 2 所述的计算机程序,其特征在于,其进一步包含:

将所述至少一个接收到的数据集分成训练数据集和验证数据集;

训练所述预测模型以优化所述预测模型在所述训练数据集上的匹配;以及

利用所述验证数据集验证所述预测模型。

9. 根据权利要求 2 所述的计算机程序,其特征在于,其进一步包括:

存储所述至少一个接收到的数据集的所述预测模型的结果记录;以及

存储所述至少一个接收到的数据集的预测模型的结果文件。

10. 根据权利要求 2 所述的计算机程序,其特征在于,其进一步包括:
确认所述建模的完成;
读取所述至少一个接收到的数据集的所述预测模型评价结果的所述结果记录;以及
读取所述至少一个接收到的数据集的所述预测模型评价结果的结果文件。
11. 根据权利要求 1 所述的计算机程序,其特征在于,其进一步包括:
确定所述至少一个接收到的数据集的分割类型,其中所述分割类型是从监督分割和非监督分割组成的组中选出的;
收集基于所述监督分割的所述至少一个接收到的数据集的目标变量名称;
确定多个分割任务;
执行以下步骤中的至少一个:
部分基于所述多个分割任务、所述目标变量名称和所述分割类型的所述至少一个接收到的数据集的监督分割分析;以及
部分基于所述确定的多个分割任务和所述分割类型的所述至少一个接收到的数据集的非监督分割分析;以及
通过互联网递送至少一个所述监督分割分析和所述非监督分割分析的结果。
12. 根据权利要求 1 所述的计算机程序,其特征在于,其进一步包括:
收集所述至少一个接收到数据集的数据集格式,其中所述数据集格式是统计系统的文件格式。
13. 根据权利要求 1 所述的计算机程序,其特征在于,其进一步包括:
收集所述至少一个接收到的数据集的数据集格式;
收集所述至少一个接收到的数据集的目标变量名称;
确认所述接收到的说明、所述数据集格式和所述目标变量名称;
发出所述建模分析任务的状态通知;
存储所述至少一个接收到的数据集的接收到的数据集记录;
存储所述至少一个接收到的数据集的所述目标变量名称;
存储所述至少一个接收到的数据集的所述数据集格式;以及
从所述至少一个接收到的数据集提取头信息。
14. 根据权利要求 1 所述的计算机程序,其特征在于,其进一步包括:
收集所述至少一个接收到的数据集的至少一个概要总体类别变量名称;
确定多个概要任务;
部分基于所述确定的多个概要任务和所述至少一个概要总体类别变量名称进行所述至少一个接收到的数据集的概要分析;以及
通过互联网递送所述概要分析的结果。
15. 根据权利要求 1 所述的计算机程序,其特征在于,其进一步包括:
收集所述至少一个接收到的数据集的至少一个方差分析变量名称;
确定多个方差分析任务;
部分基于所述至少一个方差分析变量名称和所述多个方差分析任务进行至少一个接收到的数据集的方差分析测试;以及
通过互联网递送所述方差分析测试的结果。

16. 根据权利要求 1 所述的计算机程序,其特征在于,其进一步包括收集以下中的至少一个:

所述至少一个接收到的数据集的加权变量名称;
所述至少一个接收到的数据集的包括变量名称;以及
所述至少一个接收到的数据集的排除变量名称。

17. 一个系统,包括:

一个内存,该内存用于储存:接收到的数据集,其中至少一个接收到的数据集是通过互联网接收的;应用于所述至少一个接收到的数据集的模型说明;

一个或多个处理器,该处理器共同地用于:确定与所述至少一个接收到的数据集相关的多个统计建模任务;部分基于所述确定的多个统计建模任务和所述建模说明对所述至少一个接收到的数据集建模;用所述模型评价所述至少一个接收到的数据集,产生一个评价数据集;以及通过互联网递送所述评价数据集。

18. 根据权利要求 17 所述的系统,其特征在于,所述内存用于进一步储存以下中的至少一个:

所述应用于多个接收到的数据集的模型;以及
所述至少一个接收到的数据集,用于将多个模型应用至所述接收到的数据集。

19. 根据权利要求 17 所述的系统,其特征在于,所述模型包括优化多个模型与所述至少一个接收到的数据集的匹配的多个模型。

20. 根据权利要求 17 所述的系统,其特征在于,所述内存进一步用于储存以下中的至少一个:

应用于所述模型的一组建模预设置;
用于所述建模和评价结果的递送的递送方法预设置;
所述至少一个接收到的数据集的数据共享预设置;以及
所述建模和评价结果的结果共享预设置。

21. 根据权利要求 17 所述的系统,其特征在于,所述模型说明进一步包括:

所述至少一个接收到的数据集的接收到的数据集格式;
所述至少一个接收到的数据集的目标变量名称;和至少一个:
所述至少一个接收到的数据集的加权变量名称;
所述至少一个接收到的数据集的包括变量名称;以及
所述至少一个接收到的数据集的排除变量名称。

22. 根据权利要求 17 所述的系统,其特征在于:

所述内存进一步用于储存所述至少一个接收到的数据集的接收到的数据集记录;并且
所述一个或多个处理器进一步用于从所述至少一个接收到的数据集中提取头信息。

23. 根据权利要求 17 所述的系统,其特征在于,所述一个或多个处理器可进一步用于:
确认所述建模的完成;

接收所述建模的结果请求;
递送所述建模的结果记录;以及
递送所述建模的结果文件。

24. 一个系统,包括:

- 一个内存,该内存用于储存 ;
接收到的数据集,其中至少一个接收到的数据集是通过互联网接收的 ;
应用至所述至少一个接收到的数据集的模型说明,其中所述模型说明至少包括 :
所述至少一个接收到的数据集的接收到的数据集格式 ;
所述至少一个接收到的数据集的目标变量名称 ;和至少一个 :
所述至少一个接收到的数据集的加权变量名称 ;
所述至少一个接收到的数据集的包括变量名称 ;以及
所述至少一个接收到的数据集的排除变量名称 ;
一个或多个处理器,其共同地用于 :
将所述至少一个接收到的数据集分成训练数据集和验证数据集 ;
使用多个统计建模任务训练多个模型以优化多个模型与所述训练数据集的匹配,其中
所述训练部分基于所述模型说明 ;
在所述验证数据集上验证多个所述训练模型的匹配 ;
评价所述多个验证模型 ;以及
通过互联网递送所述多个验证模型评价结果。
25. 根据权利要求 24 所述的系统,其特征在于,互联网通信接口包含互联网浏览器、
移动设备、数据集成服务和文件传输服务的至少一个。
26. 根据权利要求 24 所述的系统,其特征在于,互联网通信网络包含有线、无线和光通信的至少一个。

用于建模分析的基于互联网的托管系统和计算机可读介质

技术领域

[0001] 本方法和系统大体上涉及统计建模,更具体地涉及用于建模分析的基于互联网的托管系统和计算机可读介质。

背景技术

[0002] 目前,建模分析需要用户已经购买的建模分析软件。此外,目前的建模分析软件需要专家级的知识,才能够从分析中提取有用的信息。这两个条件,专家级的互动以及资本密集型的软件,成了公司在建模分析方面的重大障碍。

[0003] 因此,需要一种用于建模分析的基于互联网的托管系统和计算机可读介质。更具体地,需要一个基于互联网递送的、提供简化的用户交互的、根据请求进行回归建模的预测建模服务。该服务可递送到互联网浏览器、移动设备、数据集成服务(例如消息代理)或文件传输服务。

[0004] 此外,需要一个允许从可用互联网的任何地方远程访问预测建模的系统 and 计算机可读介质。该系统允许通过互联网上的订阅服务提供的建模服务的递送,并且允许可以通过每一个模型需求定价提供的建模服务的递送。系统和计算机可读介质可用一个数据集根据来自互联网的请求来建立多个模型,并且可用一个模型根据来自互联网的请求评价多个数据集。

[0005] 设想的是,本系统通过由用户指定的授权访问可以根据需求共享模型数据和结果。该系统允许在用户配置中定制建模预设置、建模递送方法、建模数据和结果的共享。这个基于互联网递送的用于建模的预测建模服务为多个客户保持适当的访问控制,其中一个客户不能查看另一个客户的数据或结果。

发明内容

[0006] 将描述一个基于互联网的计算机程序的例子,其存储于计算机可读介质并由处理器执行操作。所述计算机程序包含计算机可读的程序代码,此程序代码确定与接收到的数据集有关的多个建模分析任务,其中接收到的数据集是通过互联网接收到的。此程序代码包含下列指令:用于接收多个建模分析任务说明,通过互联网收集这些建模分析任务的说明,根据之前确定的多个建模分析任务和接收到的说明对接收到的数据集执行建模分析,以及通过互联网递送建模分析的结果。

[0007] 该程序可进一步具有用于收集接收到的数据集的目标变量名称(designation)、收集建模分析任务的预测模型类型名称、以及收集接收到的数据集的数据集格式的代码。对接收到的数据集的预测建模可基于已经确定的多种建模分析任务、数据集的格式、收集到的目标变量名称、以及预测建模类型名称。该代码还可包括用于通过互联网递送预测模型、评价预测模型,并通过互联网递送评价结果的指令。该收集到的项目可以通过互联网收集。

[0008] 此外,该程序可含有存储接收到的数据集的代码,其用于将多种预测模型应用至

接收到的数据集。该代码可含有用于通过互联网收集应用于预测模型的一组建模预设置(例如递送方法预设置、数据共享预设置和结果共享预设置)中的至少一个的指令。

[0009] 此外,可以想到的是,该程序可含有用于把收到的数据集分为一个训练数据集和一个验证数据集、训练预测模型以优化预测模型同训练数据集的匹配、并且利用验证数据集来验证该预测模型的代码。

[0010] 此外,该代码可含有用于存储预测模型的结果记录和结果文件、收到的数据集的记录、目标变量名称以及数据集的格式的指令。

[0011] 该代码还可含有用于读取如下内容的指令,所述内容包括预测模型的结果记录和预测模型的结果文件。当程序已经完成建模分析后,所述代码可具有用于确认建模完成的指令。该程序还可含有用于从接收到的数据集提取头信息的代码,其中数据集格式包含统计系统文件格式。通用的统计系统的文件格式可包括 CSV 文件格式、SAS 文件格式、SPSS 文件格式、S-Plus 文件格式、Stata (DTA) 文件格式、Systat (SYS) 文件格式、EpiInfo (REC) 文件格式、Minitab (Minitab 便携工作表) 文件格式以及 XML 格式等。

[0012] 变量的加权允许一些变量对结果有更大的影响。该代码可含有用于通过互联网收集接收到的数据集的至少一个加权变量的名称、接收到的数据集的包括变量的名称、和接收到的数据集的排除变量的名称的指令。

[0013] 该代码可含有用于分割分析的指令,例如确定接收到的数据集的分割类型,其中分割类型选自自由监督分割和非监督分割组成的组中。此外,该代码可含有基于监督分割收集接收到的数据集的目标变量名称(通过互联网收集)、确定多个分割任务、对收到的数据集基于多种分割任务、目标变量名称以及分割类型来执行监督分割分析的指令。对接收到的数据集的非监督分割分析可以基于确定的多个分割任务和分割类型。在这两种情况下,该代码含有用于通过互联网递送监督分割分析和非监督分割分析的结果的指令。

[0014] 此外,该代码还可含有用于概要(profiling)分析的指令,例如通过互联网收集接收到的数据集的概要总体类别变量名称,确定多个概要任务,基于已经确定的多个概要任务和概要总体类别变量名称执行接收到的数据集的概要分析,并通过互联网递送概要分析的结果。

[0015] 此外,该代码还可含有用于方差分析的指令,例如通过互联网收集接收到的数据集的方差分析变量名称,确定多个方差分析任务,基于方差分析变量名称和多个方差分析任务对接收到的数据集执行方差分析测试,并通过互联网递送方差分析测试的结果。

[0016] 在一个实施例中,系统包括一个内存,该内存用来存储通过互联网接收到的数据集和应用到接收到的数据集的模型说明。该系统具有一个或多个处理器,其共同地用来确定与接收到的数据集相关的多个统计建模任务,根据已经确定的多个统计建模任务和模型说明对该接收到的数据集进行建模,以所述模型对附加的接收到的数据集进行评价,并通过互联网递送建模和评价结果。

[0017] 在示例系统中,内存也可存储应用于多个接收到的数据集的至少一个模型和接收到的数据集,以将多个模型应用至接收到的数据集。该内存也可存储一组应用于模型的建模预设置的至少一个、用于建模和评价结果的递送的递送方法预设置、接收到的数据集的数据共享预设置、以及建模和评价结果的结果共享预设置。

[0018] 在示例系统中,模型可包括最佳化接收到的数据集的匹配的多个模型。该模型说

明可进一步包括接收到的数据集的接收到的数据集格式、接收到的数据集的目标变量名称、接收到的数据集的至少一个加权变量名称、接收到的数据集的包括变量名称和接收到的数据集的排除变量名称。内存可用于存储接收到的数据集的数据集记录,并且一个或多个处理器可用于从接收到的数据集提取头信息。处理器可用于确认建模的完成,接受建模结果的请求,递送建模的结果记录并递送建模的结果文件。

[0019] 在另一个实施例中,系统可包括一个内存,该内存用来存储通过互联网接收到的数据集和应用到接收到的数据集的模型说明。该模型说明至少包括接收到的数据集的接收到的数据集格式、接收到的数据集的目标变量名称和至少一个接收到的数据集的加权变量名称。该系统还可包含接收到的数据集的包括变量名称和接收到的数据集的排除变量名称。

[0020] 该系统可进一步包含一个或多个处理器,其共同用于把收到的数据集分成一个训练数据集和一个验证数据集,并用多个统计建模任务训练多个模型以优化训练数据集的匹配。该训练基于模型说明。该系统验证多个训练过的模型同验证数据集的匹配,评价多个验证过的模型,并通过互联网递送多个验证过的模型的评价结果。

[0021] 互联网通信接口可包括互联网浏览器、移动设备、数据集成服务和文件传输服务中的至少一个。互联网通信网络可包括有线、无线和光通信的至少一个。

附图说明

[0022] 图 1 描绘了第一个用于建模分析的基于互联网的托管系统;

[0023] 图 2 是第一个用于建模分析系统的基于互联网的托管系统的流程图;

[0024] 图 3 是软件流程框图;

[0025] 图 4 是软件流程框图的预测建模部分;

[0026] 图 5 是软件流程框图的用户预设置部分;

[0027] 图 6 是软件流程框图的用于未来应用的存储部分;

[0028] 图 7 是软件流程框图的训练和验证部分;

[0029] 图 8 是软件流程框图的预测模型存储部分;

[0030] 图 9 是软件流程框图的建模完成和数据读取部分;

[0031] 图 10 是软件流程框图的分割分析部分;

[0032] 图 11 是软件流程框图的格式选择部分;

[0033] 图 12 是软件流程框图的信息存储和头信息提取部分;

[0034] 图 13 是软件流程框图的概要分析部分;

[0035] 图 14 是软件流程框图的方差分析(ANOVA)部分;

[0036] 图 15 是软件流程框图的变量名称部分;

[0037] 图 16 描绘了用于建模分析的第二个基于互联网的托管系统;

[0038] 图 17 描绘了该系统的用于未来应用的存储部分;

[0039] 图 18 描绘了该系统的用户预设置部分;

[0040] 图 19 描绘了该系统的模型说明部分;

[0041] 图 20 描绘了该系统的信息存储和头信息提取部分;

[0042] 图 21 描绘了该系统的建模分析完成和数据接收部分;

[0043] 图 22 描绘了第三个用于建模分析的基于互联网的托管系统 ;以及

[0044] 图 23 描绘了该系统的互联网通信接口和连接。

具体实施方式

[0045] 图 1 显示一个示例系统图 100。用户通过计算机 110 的互联网浏览器访问该系统。该服务可以递送至互联网浏览器、移动设备、数据集成服务(例如消息中介)或文件传输服务等。计算机 110 通过互联网 120 连接到系统的防火墙 130。通信协议可以是传输控制协议(TCP)、互联网协议(IP)、全球移动通信系统(GSM)等等。防火墙 130 联网到网络服务器 140,其依次联网到数据库服务器 150 和文件系统 160。数据库服务器 150 和文件系统 160 联网到分析建模服务器 170。通信结合可以包括有线、无线、光以及它们的混合等等。

[0046] 图 2 显示该系统的一个示例操作 200。操作 200 显示用户 210、网络服务器 220、数据库服务器 230、文件系统 240 和分析建模服务器 250 的交互。网络服务器 220 在网络服务器 220 接收到来自用户 210 的一个数据集 252。数据集由用户编辑并提供,并可以作为例如电子表格或数据库文件等传送。网络服务器 220 也可接收来自用户的模型说明 262。

[0047] 模型说明包括目标变量、预测属性、建模模式等。该数据集有一个存储在数据库服务器 230 的相关记录 254。从网络服务器 220 收到的数据集 256 存储在文件系统 240 上。通过分析建模服务器 250 从数据集中提取头信息 258。

[0048] 用户 210 从网络服务器 220 收到一个属性 260 的显示。从网络服务器 220 收到模型说明并将其存储 264 在数据库服务器 230。当从网络服务器 220 收到模型说明时,其触发建模过程 266。

[0049] 通过分析建模服务器 250 从数据库服务器 230 读取模型说明 268。通过分析建模服务器 250 从文件系统 240 读取数据文件 270。建模过程被执行 272。从分析建模服务器 250 把结果记录存储 274 在数据库服务器 230。从分析建模服务器 250 把结果文件存储 276 在文件系统 240。

[0050] 建模过程后,用户 210 从分析建模服务器 250 收到确认 278。网络服务器 220 收到来自用户 210 的结果请求 280。网络服务器 220 从数据库服务器 230 读取结果记录 282,从文件系统 240 读取结果文件 284。通过网络服务器 220 将结果显示 286 给用户 210。

[0051] 图 3 中描述一个基于互联网的、存储于计算机可读介质和并由处理器执行的计算机程序的实施例。该程序包括确定 310 与通过互联网收到的数据集关联的多个建模分析任务,接收 320 多个建模分析任务的说明并通过互联网收集 330 一组建模预设的代码。此外,该程序包括依据确定的多个建模分析任务以及收到的建模分析任务的说明在接收到的数据集上进行 340 建模分析,并通过互联网递送 350 建模分析结果的代码。

[0052] 如图 4 所示,计算机可读介质还可以包括用于收集 410 接收到的数据集的目标变量名称的指令,收集 420 建模分析任务的预测建模类型名称的指令,和收集 430 接收到的数据集的数据集格式的指令。该程序还可包括用于依据确定的多个建模分析任务、数据集格式、收集到的目标变量名称和预测建模类型名称执行 440 接收到的数据集的预测建模并导致预测模型的指令。该计算机可读介质还可包括通过互联网递送 450 预测模型、评价 460 接收的数据集和通过互联网递送 470 模型评价结果的指令。收集到的项目可通过互联网收集。

[0053] 用户预设置也可以如图 5 的 500 所示的收集。用户预设置收集可包括通过互联网收集 510 可由用户设定的一组建模预设置 520、用于预测模型的递送的递送方法预设置 530、接收到的数据集的数据共享预设置 540 和预测模型的结果共享预设置 550 中的至少一个。

[0054] 参考图 6, 代码可包含存储 610 预测模型以便应用至多个接收到的数据集, 和存储 620 接收到的数据集以便将多个预测模型应用至接收到的数据集。

[0055] 训练集是接收数据的子集, 用于发现潜在预测关系。验证集也是接收数据的子集, 用于确定该预测模型的匹配度。图 7 描绘了软件流程框图的训练和验证部分, 包含指令 700, 用于将接收到的数据集分成 710 一个训练数据集和一个验证数据集, 训练 720 预测模型以优化预测模型对训练数据集匹配, 并利用验证数据集验证 730 预测模型。

[0056] 图 8 描绘了与预测模型的结果相关的代码的预测模型部分的存储 800。在本实施例中, 计算机可读介质可包含用于存储 810 接收到的数据集的预测模型的结果记录并存储 820 接收到的数据集的预测模型的结果文件的指令。

[0057] 在已经完成了接收到的数据集的预测建模后, 计算机可读介质可包含用于将建模完成的确认发送 910 给用户的指令, 读取 920 预测模型的结果记录的指令, 和如图 9 中所示读取 930 预测模型的结果文件。

[0058] 分割分析是对接收到的数据集的指定部分分组的任务, 使得在同一组的对象相对在其他组的那些彼此之间更加类似。接收到的数据集的分割也可以如图 10 中描述的来执行 1000。确定 1010 接收到的数据集的分割类型。分割类型可以是多个类型中的一个, 包括监督分割和非监督分割。对于监督分割, 可通过互联网 1020 收集目标变量名称。确定 1030 多个分割任务, 并执行 1040 监督或非监督分割分析。如果在接收到的数据集执行监督分割分析 1050, 它是依据多个分割任务、目标变量名称和分割类型。如果在接收到的数据集执行 1060 非监督分割分析, 它将依据多个确定的分割任务和分割类型。然后监督分割分析和/或非监督分割分析的结果通过互联网递送 1070。

[0059] 如图 11 中所示, 用户提供的数据集格式可包含 1100 统计系统文件格式, 例如 CSV 格式 1110、SAS 文件格式 1120、SPSS 文件格式 1130、S-Plus 文件格式 1140、Stata 文件格式 1150、Systat 文件格式 1160、EpiInfo 文件格式 1170、Minitab 文件格式 1180 和 XML 格式 1190 等等。

[0060] 如图 12 中所示, 各种类型的信息可以从接收到的数据集中加以收集和存储。通过互联网收集接收到的数据集的数据集格式 1206 和目标变量名称 1207。代码也可确认 1208 用户收到的任务指令, 例如说明、数据集格式和目标变量名称的确收。关于建模分析任务的状态通知也可以发送 1209 给用户。可以存储接收到的数据集的接收数据集记录 1210, 目标变量名称 1220 和数据集格式 1230, 并且可提取 1240 头信息。

[0061] 总体概要是对接收到的数据的分析, 使它的结构、内容和关系明晰。概要分析将允许用户深入了解多个可比较的总体领域。概要分析将通过在多个选定的尺寸的指定总体比较和对比属性。

[0062] 如图 13 中的 1300 所示, 代码可包含可以进行分析的接收到的数据集的概要数据。这种分析可以包括通过互联网收集 1310 接收到的数据集的概要总体类别变量名称, 确定 1320 多个概要任务, 根据所确定的多个概要任务和概要总体类别变量名称对接收到的数据

集执行 1330 概要分析,并通过互联网递送 1340 概要分析的结果。

[0063] 方差分析是一种同时在两个或更多的平均值之间进行比较的统计方法。这种统计方法产生的值可进行测试,以确定变量之间是否存在显著关系。如图 14 中一个实例中所示,方差分析(ANOVA)信息也可以从所接收到的数据集确定 1400。可通过互联网收集 1410 接收到的数据集的方差分析变量名称,确定 1420 与方差分析相关的多个任务,然后基于方差分析变量名称和多个方差分析任务执行方差分析测试的实际进行 1430。可通过互联网递送 1440 方差分析测试的结果。

[0064] 如图 15 中所示,也可通过互联网收集 1510 接收到的数据集的各种类型的变量名称。变量名称类型可包括加权变量 1520、包括变量 1530 和排除变量 1540。

[0065] 图 16 示出了一个用于实施建模分析的优选系统 1600。该系统将包括一个内存 1610,该内存用来存储通过互联网收到的接收到的数据集 1620 以及用于接收到的数据集的模型说明 1630 和建模类型 1632。该系统将包括一个或多个处理器 1640,其共同用于确定与接收到的数据集关联的多个统计建模任务 1650,根据确定的多个统计建模任务、建模说明和模型预设置为接收到的数据集建模 1660,对附加的接收到的数据集进行评价 1670,并通过互联网递送 1680 建模和评价结果。

[0066] 关于图 17,系统还可包括一个内存,其构造成存储 1710 至少一个应用至多个接收到的数据集的模型 1720 以及将多个模型应用至接收到的数据集的接受的数据集 1730。该模型可包括多个模型以便优化多个模型对接收到的数据集 1740 的匹配。

[0067] 如图 18 中所示,该系统可进一步包括一个内存,其构造成存储 1810 应用至模型的一组模型预设置 1820 的至少一个,用于递送建模和评价结果的递送方法预设置 1830,接收到的数据集的数据分享预设置 1840 和建模和评分结果的结果分享预设置 1850 等等。

[0068] 图 19 示出了模型说明 191 可进一步包括接收到的数据集格式 1920,一个目标变量名称 1930 和至少一个 1940 加权变量名称 1950,包括变量名称 1960 以及排除变量名称 1970。

[0069] 如图 20 中所示,内存可用于进一步存储接收到的数据集记录 2010,并且一个或多个处理器可用于从接收到的数据集提取头信息 2020。

[0070] 该处理器可另外被用来 2110 确认建模的完成 2120。至于建模结果,该系统可用来接收对结果的请求 2130,递送结果记录 2140 并递送结果文件 2150。

[0071] 图 22 示出了另一个实施建模分析的系统实施例 2200。该系统包括两个主系统,内存 2210 和一个或多个处理器 2230。该系统包括内存,其用来存储 2210 通过互联网收到的数据集 2212 和被应用到接收到的数据集的模型说明 2214。模型说明包括至少以下与接收到的数据集有关的信息:接收到的数据集格式 2216,目标变量名称 2218,建模类型 2219,和加权变量名称 2222、包括变量名称 2224 以及排除变量名称 2226 中的至少一个。

[0072] 系统 2200 还包括一个或多个处理器,其共同地用于 2230 将接收到的数据集分割 2232 成为训练数据集和验证数据集,用多个统计建模任务训练多个模型 2234 以最佳化训练数据集的多个模型匹配,其中训练根据模型说明进行。此外,该系统还将用训练的多个模型与验证数据集的匹配验证多个训练模型 2236。该系统然后对多个验证过的模型评价 2238,并通过互联网递送 2240 多个验证过的模型的评价结果。

[0073] 图 23 示出了互联网通信接口和网络。互联网通信接口 2310 包括互联网浏览器

2320、移动设备 2330、数据集成服务 2340 和文件传输服务 2350 的至少一个。互联网通信网络 2360 包括有线 2370、无线 2380 和光通信 2390 的至少一个。建模过程 2392 完成的确认可包括电子邮件 2394 和 SMS 消息 2396 的至少一个。

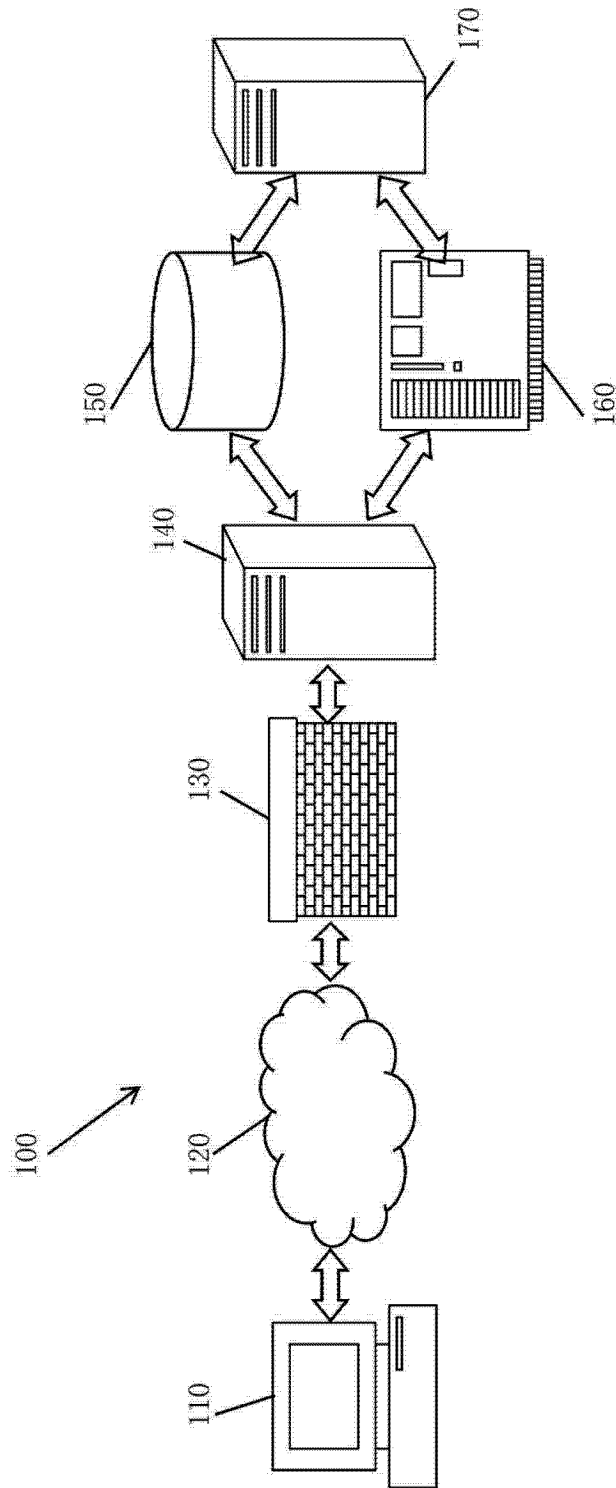


图 1

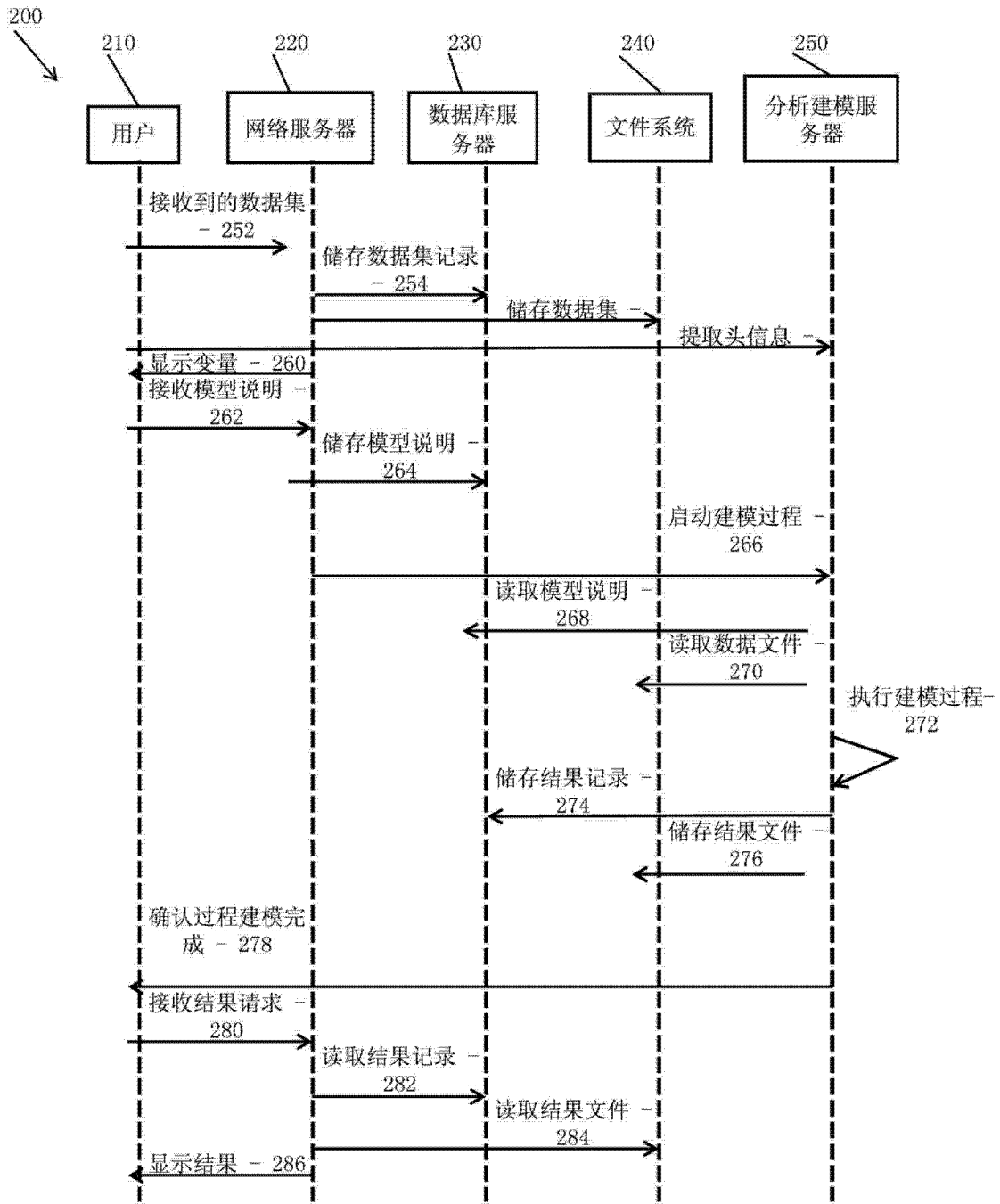


图 2

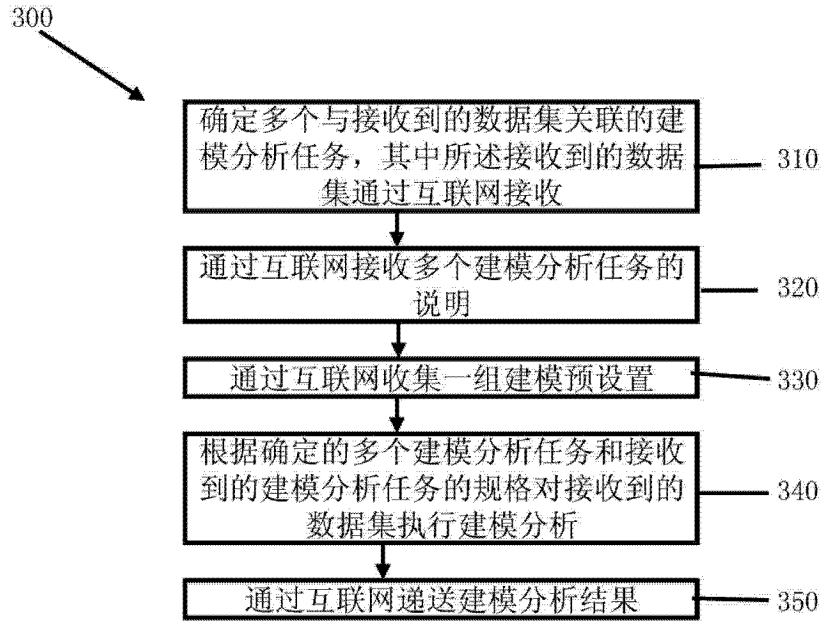


图 3

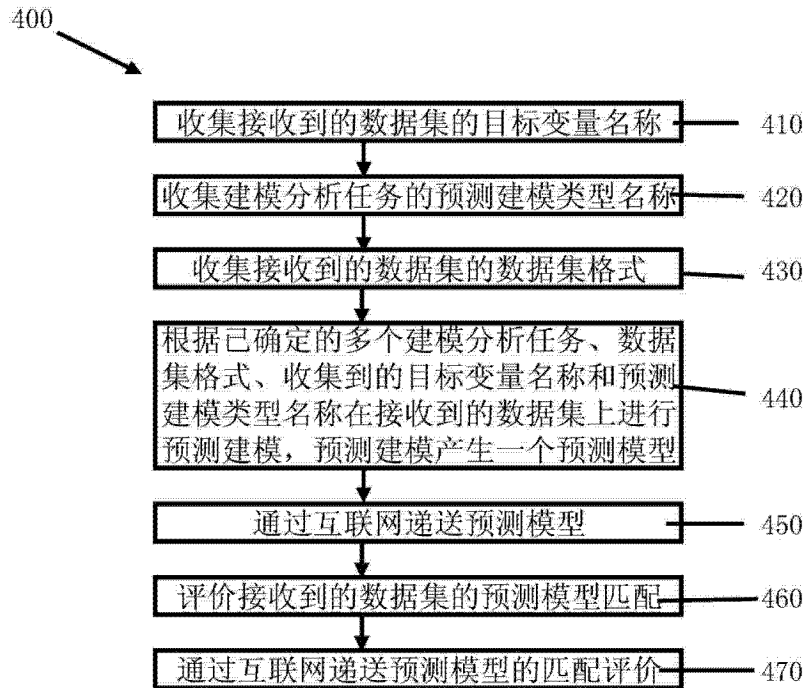


图 4

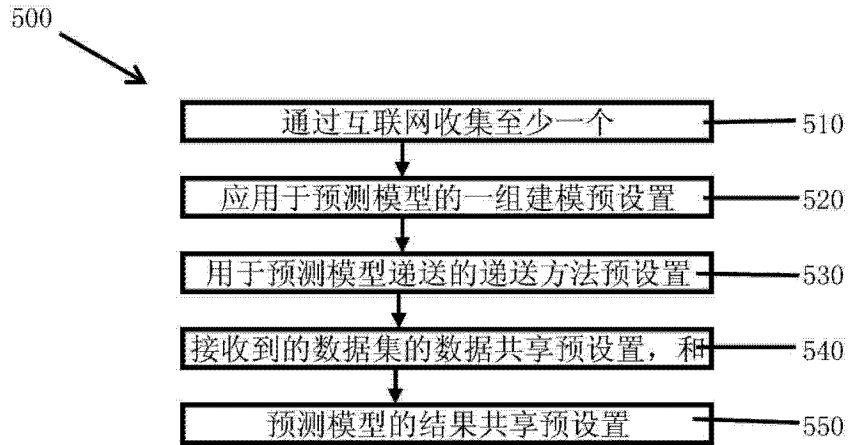


图 5

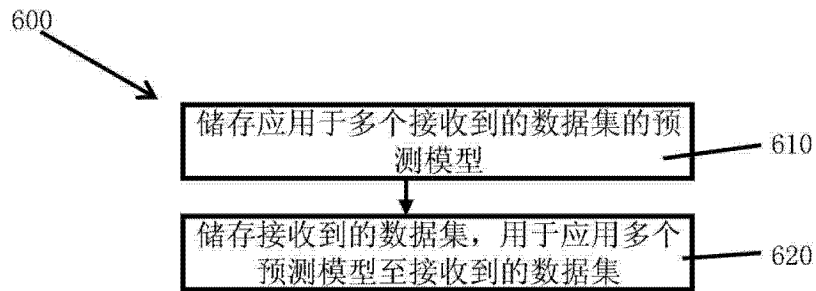


图 6

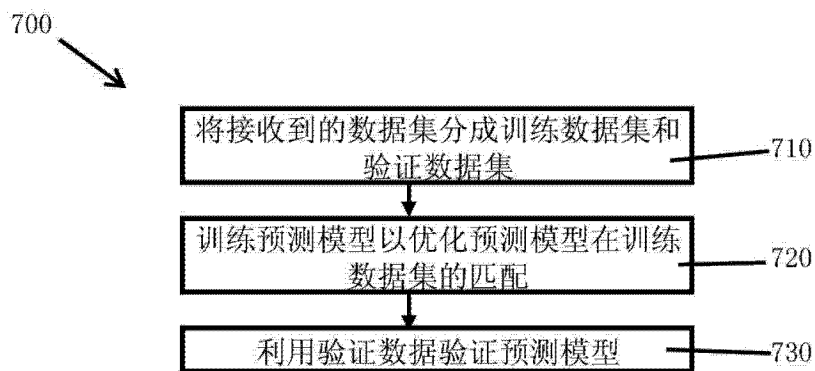


图 7

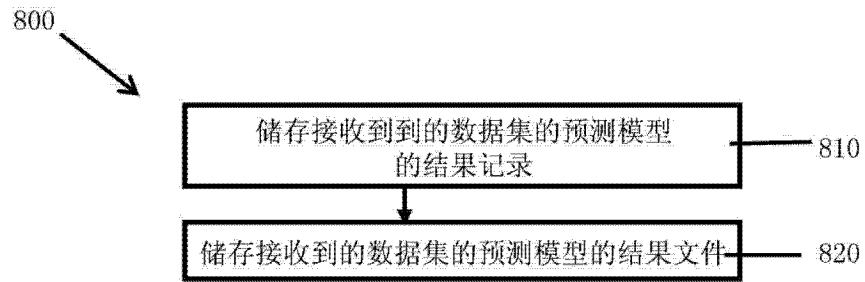


图 8

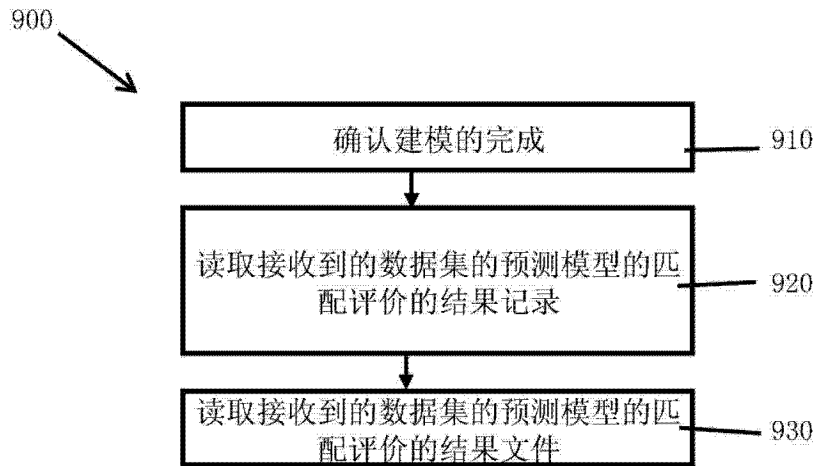


图 9

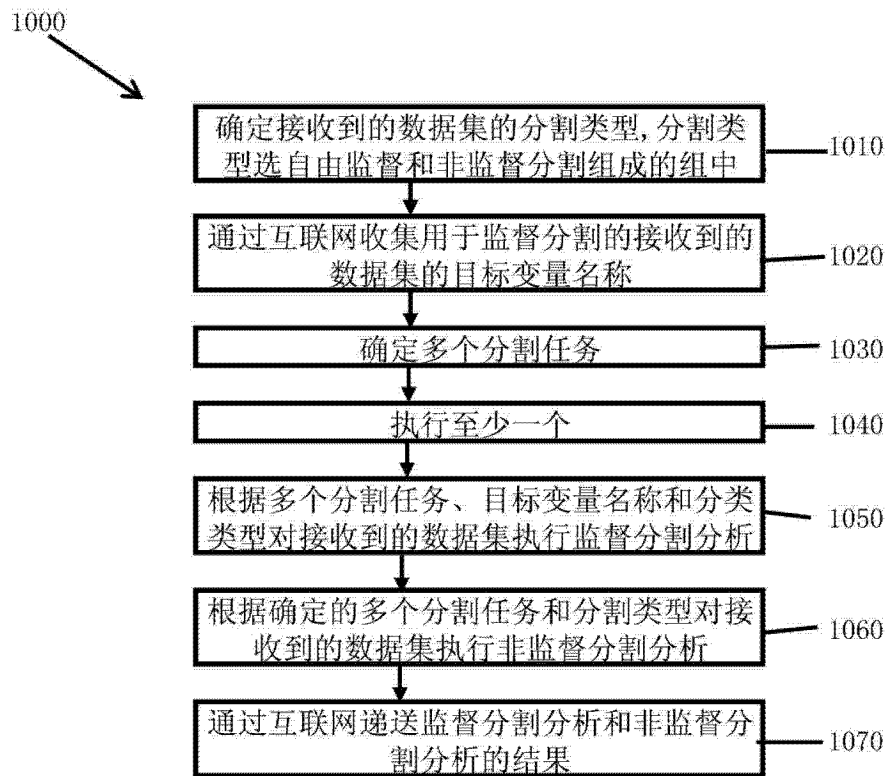


图 10

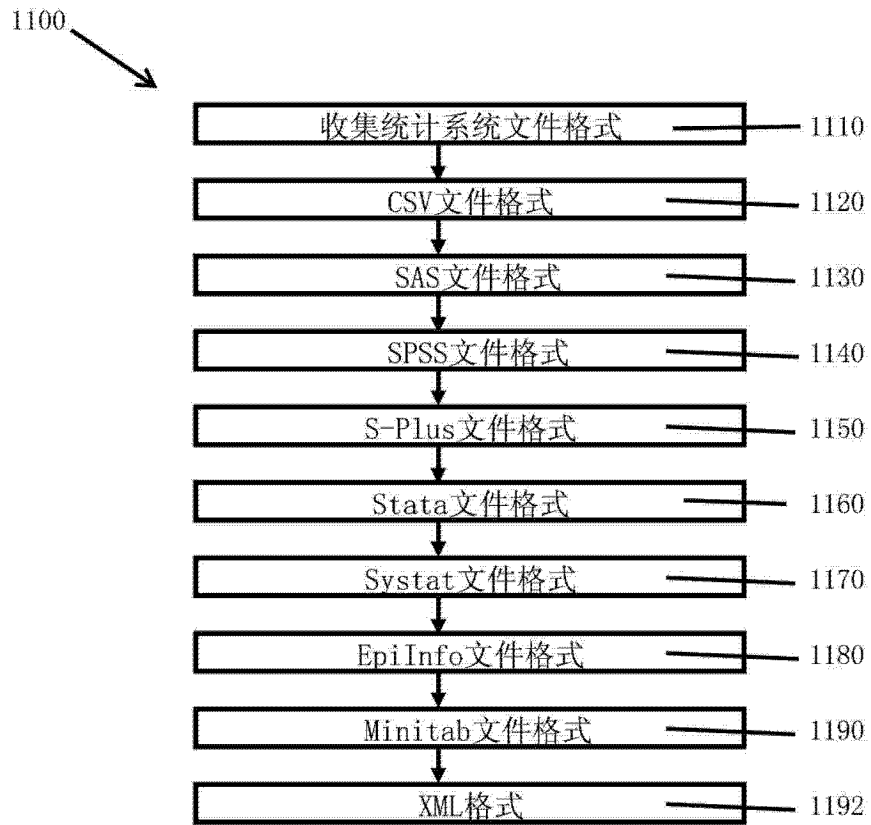


图 11

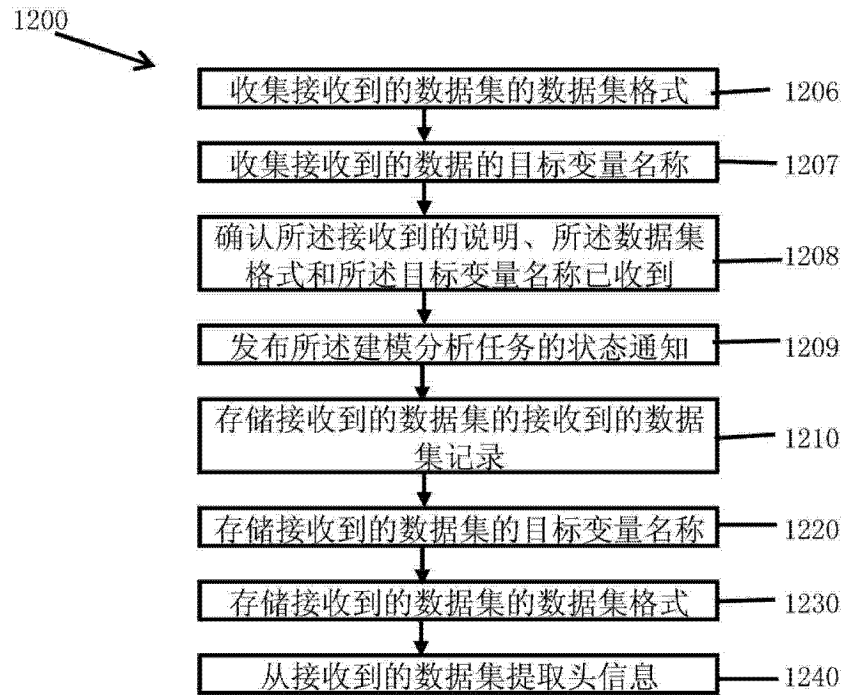


图 12

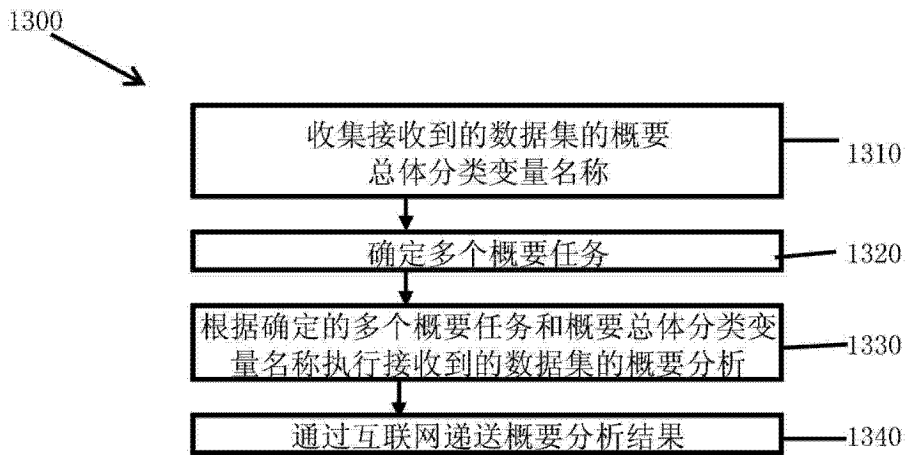


图 13

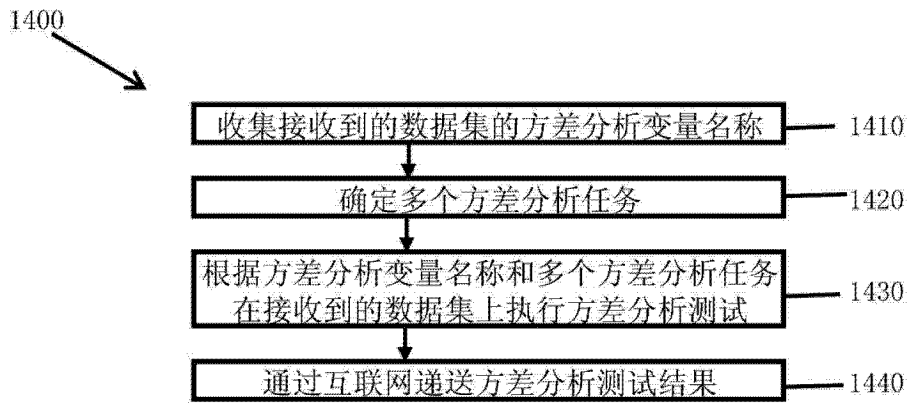


图 14

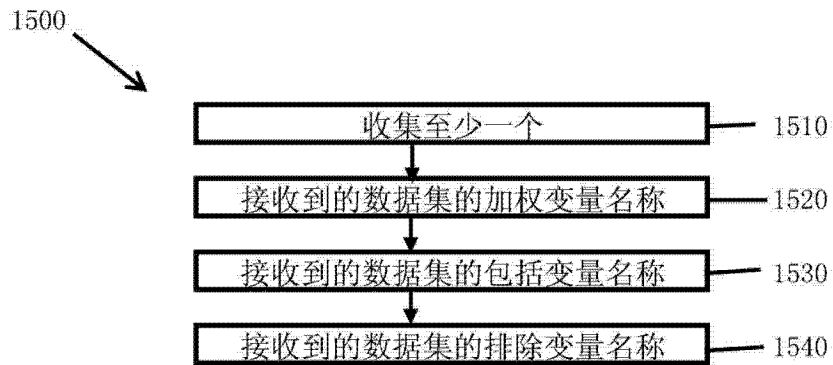


图 15

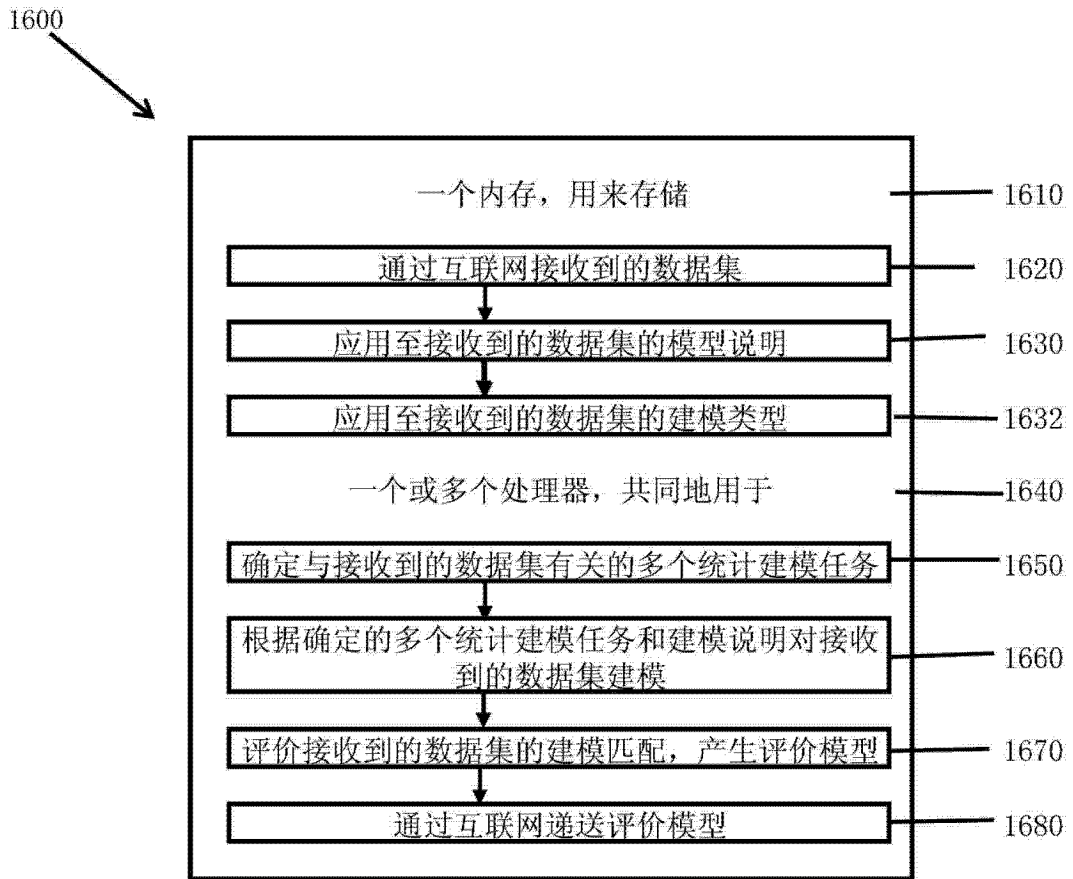


图 16

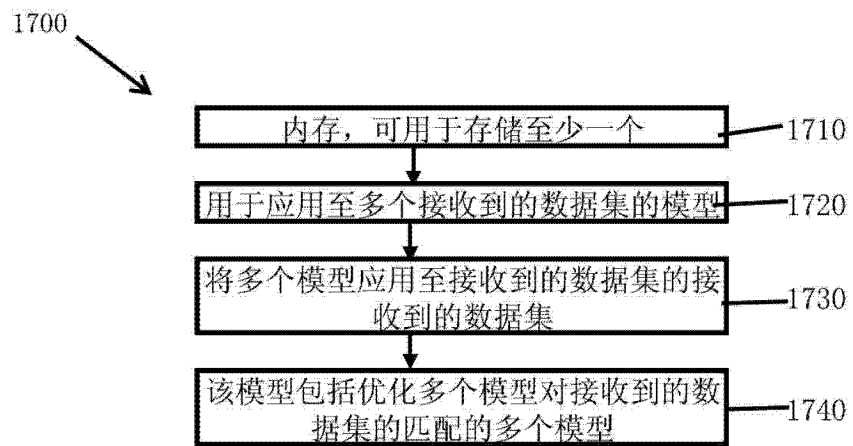


图 17

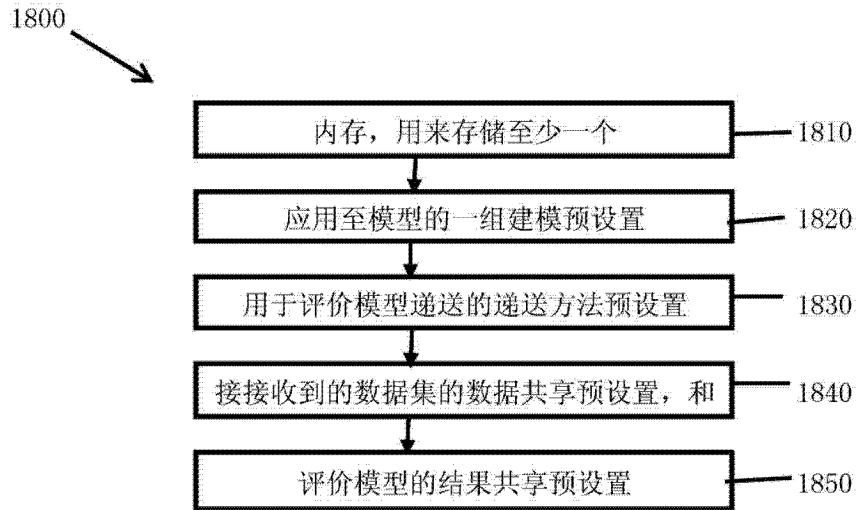


图 18

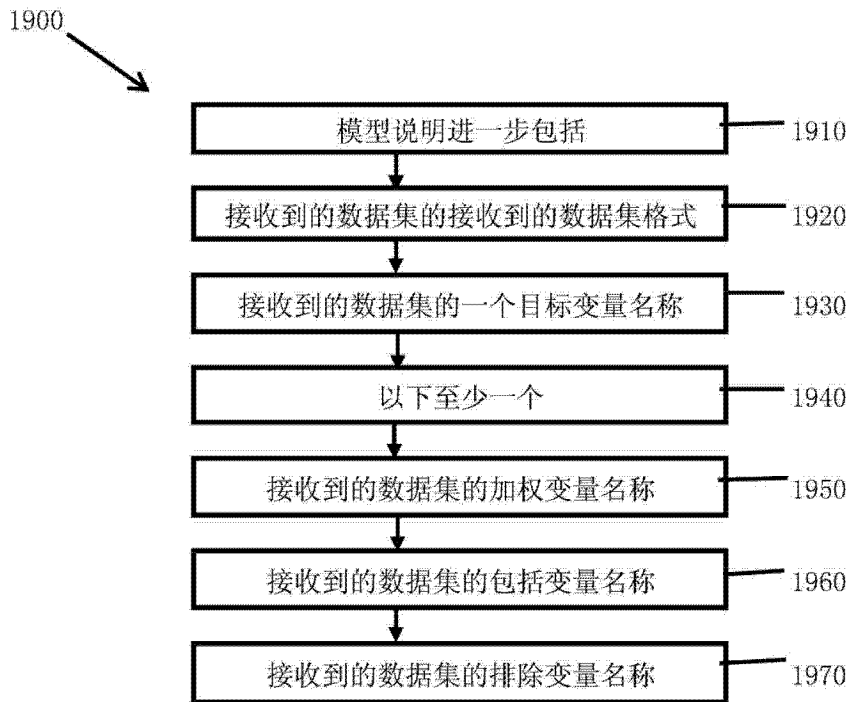


图 19

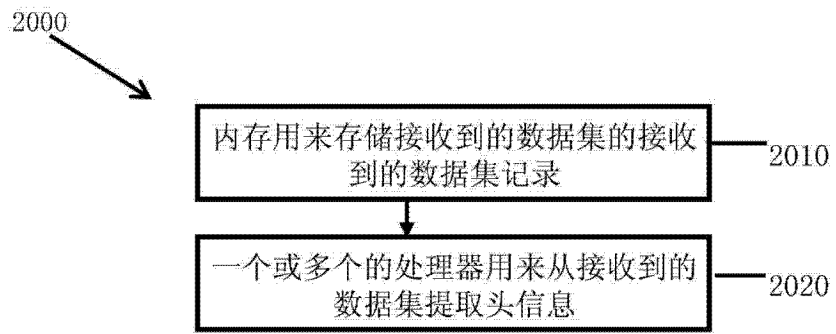


图 20

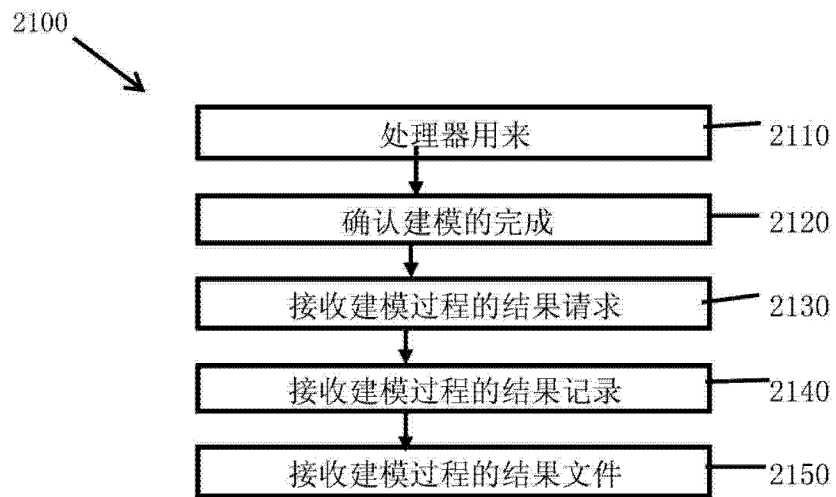


图 21

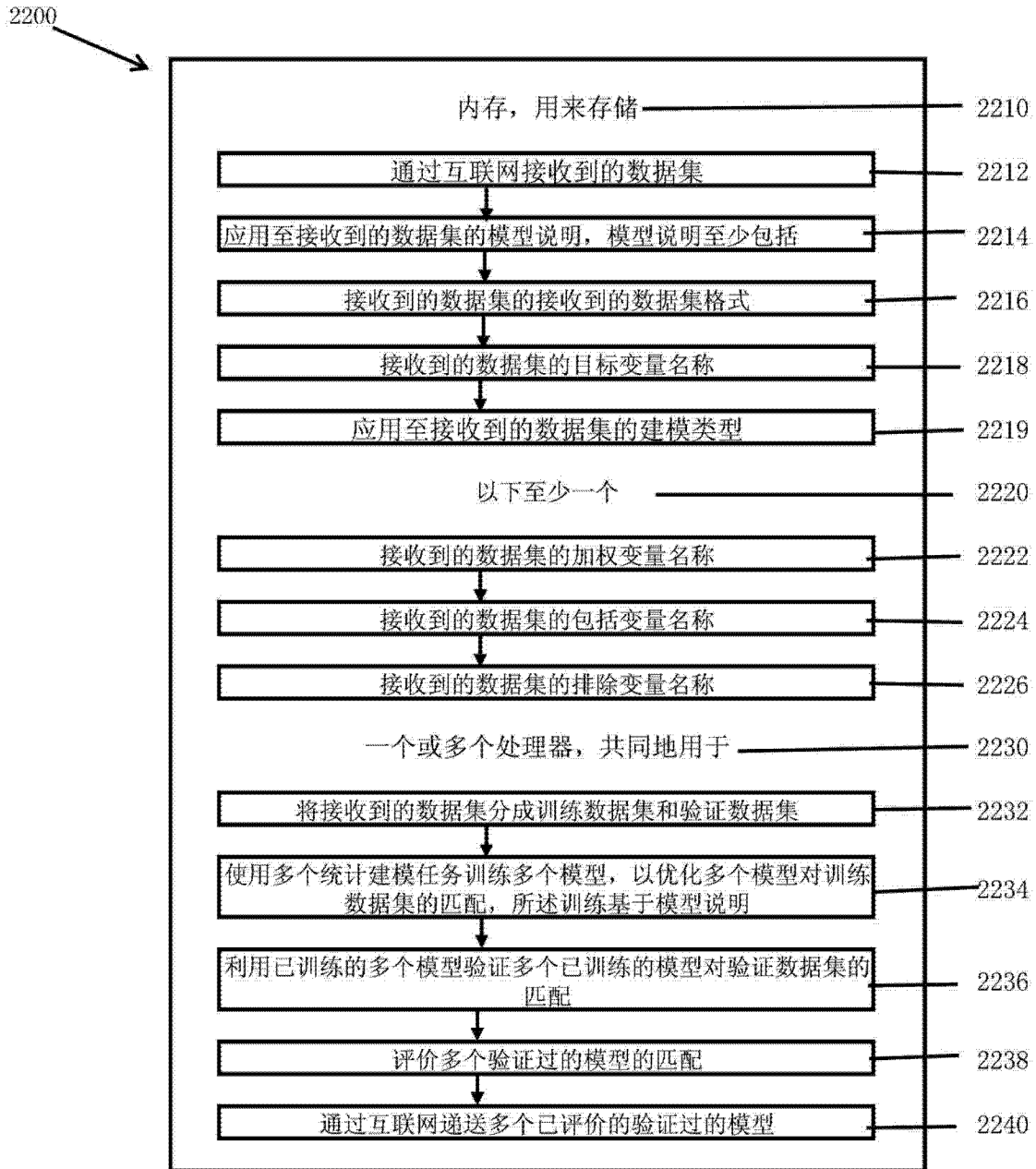


图 22

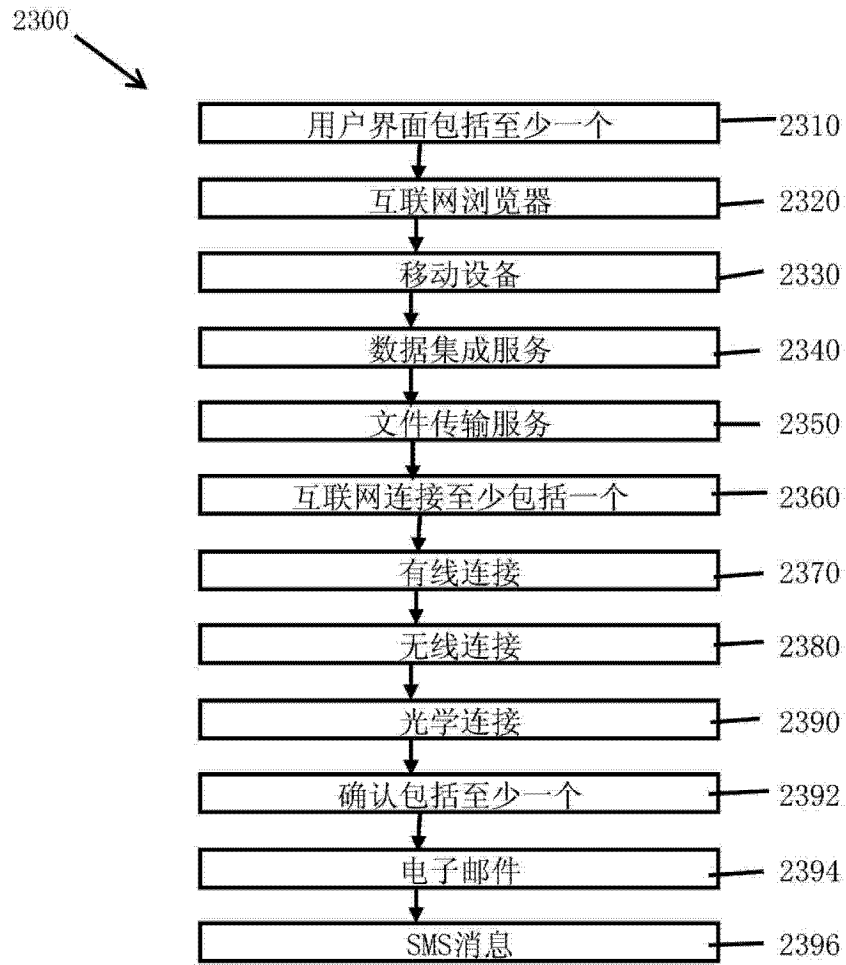


图 23