



(12) 发明专利

(10) 授权公告号 CN 111985677 B

(45) 授权公告日 2024. 06. 21

(21) 申请号 202010618648.0

G06N 3/0464 (2023.01)

(22) 申请日 2020.06.30

G06N 3/049 (2023.01)

G06N 3/0895 (2023.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 111985677 A

(56) 对比文件

CN 108983710 A, 2018.12.11

JP 2008084039 A, 2008.04.10

(43) 申请公布日 2020.11.24

(73) 专利权人 联想(北京)有限公司

审查员 钟宇静

地址 100085 北京市海淀区上地西路6号2

幢2层201-H2-6

(72) 发明人 唐建权 杨帆 金继民 张成松

(74) 专利代理机构 北京乐知新创知识产权代理

事务所(普通合伙) 11734

专利代理师 周伟

(51) Int. Cl.

G06Q 10/0637 (2023.01)

G06Q 50/04 (2012.01)

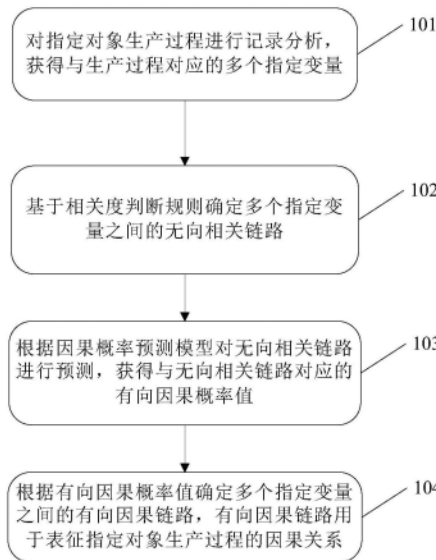
权利要求书2页 说明书9页 附图6页

(54) 发明名称

一种因果链路分析方法、设备及计算机可读存储介质

(57) 摘要

本发明公开了一种因果链路分析方法、设备及计算机可读存储介质,所述方法包括:对指定对象生产过程进行记录分析,获得与所述生产过程对应的多个指定变量;基于相关度判断规则确定所述多个指定变量之间的无向相关链路;根据因果概率预测模型对所述无向相关链路进行预测,获得与所述无向相关链路对应的有向因果概率值;根据所述有向因果概率值确定所述多个指定变量之间的有向因果链路,所述有向因果链路用于表征所述指定对象生产过程的因果关系;能知道在生产过程中各指定变量间互相影响的先后顺序。



1. 一种因果链路分析方法,其特征在于,所述方法包括:

对指定对象生产过程进行记录分析,获得与所述生产过程对应的多个指定变量;

基于相关度判断规则确定所述多个指定变量之间的无向相关链路;

根据因果概率预测模型对所述无向相关链路进行预测,获得与所述无向相关链路对应的有向因果概率值;

根据所述有向因果概率值确定所述多个指定变量之间的有向因果链路,所述有向因果链路用于表征所述指定对象生产过程的因果关系;

所述基于相关度判断规则确定所述多个指定变量之间的无向相关链路,包括:

建立与所述多个指定变量对应的有向完全图,所述有向完全图包含用于连接所述多个指定变量的多个双向链路,任一所述双向链路均包含两指向相反的有向链路;

根据所述多个指定变量确定与每一个有向链路对应的相关值;

筛选满足第一阈值的相关值,将与满足所述第一阈值的相关值对应的有向链路确定为第一有向链路;

根据所述多个指定变量和第一有向链路确定有向相关图,所述有向相关图用于表征所述指定变量之间的无向相关链路。

2. 根据权利要求1所述的方法,其特征在于,所述对指定对象生产过程进行记录分析,获得与所述生产过程对应的多个指定变量,包括:

通过分布式控制系统对指定对象运行过程进行记录,获得运行记录;

对所述运行记录进行筛选补充,获得与所述生产过程对应的关联信息;其中,所述生产过程包含在所述运行过程之中;

对所述关联信息进行标准化处理,获得指定变量。

3. 根据权利要求1所述的方法,其特征在于,所述因果概率预测模型为图形神经网络模型;

相应的,所述根据因果概率预测模型对所述无向相关链路进行预测,获得与所述无向相关链路对应的有向因果概率值,包括:

根据所述有向相关图确定变量集合与链路集合;其中,所述变量集合包含多个指定变量,所述链路集合包含第一有向链路;

通过图形神经网络模型对所述变量集合与链路集合进行预测,获得与所述链路集合对应的因果概率集,所述因果概率集包含与所述第一有向链路对应的有向因果概率值。

4. 根据权利要求1或3所述的方法,其特征在于,根据所述有向因果概率值确定所述多个指定变量之间的有向因果链路,包括:

筛选满足第二阈值的有向因果概率值,将与满足所述第二阈值的有向因果概率值对应的有向链路确定为第二有向链路;

根据所述多个指定变量和所述第二有向链路确定有向因果链路图,所述有向因果链路图用于表征所述指定变量之间的有向因果链路。

5. 根据权利要求1所述的方法,其特征在于,在所述基于相关度判断规则确定多个指定变量之间的无向相关链路之前,所述方法还包括:

对至少两个指定变量之间进行因果链路标注,确定已知有向因果链路;

基于所述已知有向因果链路对所述因果概率预测模型进行半监督学习,以获得因果概

率预测模型。

6.一种因果链路分析设备,其特征在于,所述设备包括:

分析模块,用于对指定对象的生产过程进行记录分析,获得多个指定变量;

确定模块,用于基于相关度判断规则确定所述多个指定变量之间的无向相关链路;

预测模块,用于根据因果概率预测模型对所述无向相关链路进行预测,获得与所述无向相关链路对应的有向因果概率值;

所述确定模块,还用于根据所述有向因果概率值确定所述多个指定变量之间的有向因果链路;

所述确定模块,包括:

建立子模块,用于建立与所述多个指定变量对应的有向完全图,所述有向完全图包含用于连接所述多个指定变量的多个双向链路,任一所述双向链路均包含两指向相反的有向链路;

确定子模块,用于根据所述多个指定变量确定与每一个有向链路对应的相关值;

筛选子模块,用于筛选满足第一阈值的相关值,将与满足所述第一阈值的相关值对应的有向链路确定为第一有向链路;

所述确定子模块,还用于根据所述多个指定变量和第一有向链路确定有向相关图,所述有向相关图用于表征所述指定变量之间的无向相关链路。

7.根据权利要求6所述的设备,其特征在于,所述分析模块,包括:

记录子模块,用于通过分布式控制系统对所述指定对象的运行过程进行记录,获得运行记录;

筛选补充子模块,用于对所述运行记录进行筛选补充,获得与所述生产过程对应的关联信息;其中,所述生产过程包含在所述运行过程之中;

处理子模块,用于对所述关联信息进行标准化处理,获得指定变量。

8.一种计算机可读存储介质,所述存储介质包括一组计算机可执行指令,当所述指令被执行时用于执行权利要求1-5任一项所述的因果链路分析方法。

一种因果链路分析方法、设备及计算机可读存储介质

技术领域

[0001] 本发明涉及数据处理技术领域,尤其涉及一种因果链路分析方法、设备及计算机可读存储介质。

背景技术

[0002] 工业领域在生产过程中均存在繁多的工艺参数,并且工艺参数互相强耦合,一个工艺参数的变化往往会引起其他一个或多个工艺参数的改变,从而影响最终生成的产物。当人们根据经验针对性地调整工艺参数时,可能会引起“连锁反应”,比如提高A处的温度,可能导致B处温度过高。目前在工艺参数建模分析过程中,通常通过分析各工艺参数之间的相关性来调整工艺参数,但是通过相关性分析只能确定工艺参数之间具有相关性,无法精确预测调整工艺参数后会导致什么结果,从而不能对工艺参数进行精确控制,在提升产物质量、装置智能预警等方面效果并不理想。

发明内容

[0003] 本发明实施例提供了一种因果链路分析方法、设备及计算机可读存储介质,能知道在生产过程中各指定变量间互相影响的先后顺序。

[0004] 本发明一方面提供一种因果链路分析方法,所述方法包括:对指定对象生产过程进行记录分析,获得与所述生产过程对应的多个指定变量;基于相关度判断规则确定所述多个指定变量之间的无向相关链路;根据因果概率预测模型对所述无向相关链路进行预测,获得与所述无向相关链路对应的有向因果概率值;根据所述有向因果概率值确定所述多个指定变量之间的有向因果链路,所述有向因果链路用于表征所述指定对象生产过程的因果关系。

[0005] 在一可实施方式中,所述对指定对象生产过程进行记录分析,获得与所述生产过程对应的多个指定变量,包括:通过分布式控制系统对指定对象运行过程进行记录,获得运行记录;对所述运行记录进行筛选补充,获得与所述生产过程对应的关联信息;其中,所述生产过程包含在所述运行过程之中;对所述关联信息进行标准化处理,获得指定变量。

[0006] 在一可实施方式中,所述基于相关度判断规则确定所述多个指定变量之间的无向相关链路,包括:建立与所述多个指定变量对应的有向完全图,所述有向完全图包含用于连接所述多个指定变量的多个双向链路,任一所述双向链路均包含两指向相反的有向链路;根据所述多个指定变量确定与每一个有向链路对应的相关值;筛选满足第一阈值的相关值,将与满足所述第一阈值的相关值对应的有向链路确定为第一有向链路;根据所述多个指定变量和第一有向链路确定有向相关图,所述有向相关图用于表征所述指定变量之间的无向相关链路。

[0007] 在一可实施方式中,所述因果概率预测模型为图形神经网络模型;相应的,所述根据因果概率预测模型对所述无向相关链路进行预测,获得与所述无向相关链路对应的有向因果概率值,包括:根据所述有向相关图确定变量集合与链路集合;其中,所述变量集

合包含多个指定变量,所述链路集合包含第一有向链路;通过图形神经网络模型对所述变量集合与链路集合进行预测,获得与所述链路集合对应的因果概率集,所述因果概率集包含与所述第一有向链路对应的有向因果概率值。

[0008] 在一可实施方式中,根据所述有向因果概率值确定所述多个指定变量之间的有向因果链路,包括:筛选满足第二阈值的有向因果概率值,将与满足所述第二阈值的有向因果概率值对应的有向链路确定为第二有向链路;根据所述多个指定变量和所述第二有向链路确定有向因果链路图,所述有向因果链路图用于表征所述指定变量之间的有向因果链路。

[0009] 在一可实施方式中,在所述基于相关度判断规则确定多个指定变量之间的无向相关链路之前,所述方法还包括:对至少两个指定变量之间进行因果链路标注,确定已知有向因果链路;基于所述已知有向因果链路对所述因果概率预测模型进行半监督学习,以获得因果概率预测模型。

[0010] 本发明另一方面提供一种因果链路分析设备,所述设备包括:分析模块,用于对指定对象的生产过程进行记录分析,获得多个指定变量;确定模块,用于基于相关度判断规则确定所述多个指定变量之间的无向相关链路;预测模块,用于根据因果概率预测模型对所述无向相关链路进行预测,获得与所述无向相关链路对应的有向因果概率值;所述确定模块,还用于根据所述有向因果概率值确定所述多个指定变量之间的有向因果链路。

[0011] 在一可实施方式中,所述分析模块,包括:记录子模块,用于通过分布式控制系统对所述指定对象的运行过程进行记录,获得运行记录;筛选补充子模块,用于对所述运行记录进行筛选补充,获得与所述生产过程对应的关联信息;其中,所述生产过程包含在所述运行过程之中;处理子模块,用于对所述关联信息进行标准化处理,获得指定变量。

[0012] 在一可实施方式中,所述确定模块,包括:建立子模块,用于建立与所述多个指定变量对应的有向完全图,所述有向完全图包含用于连接所述多个指定变量的多个双向链路,任一所述双向链路均包含两指向相反的有向链路;确定子模块,用于根据所述多个指定变量确定与每一个有向链路对应的相关值;筛选子模块,用于筛选满足第一阈值的相关值,将与满足所述第一阈值的相关值对应的有向链路确定为第一有向链路;所述确定子模块,还用于根据所述多个指定变量和第一有向链路确定有向相关图,所述有向相关图用于表征所述指定变量之间的无向相关链路。

[0013] 在一可实施方式中,所述因果概率预测模型为图形神经网络模型;相应的,所述根据因果概率预测模型对所述无向相关链路进行预测,获得与所述无向相关链路对应的有向因果概率值,包括:根据所述有向相关图确定变量集合与链路集合;其中,所述变量集合包含多个指定变量,所述链路集合包含第一有向链路;通过图形神经网络模型对所述变量集合与链路集合进行预测,获得与所述链路集合对应的因果概率集,所述因果概率集包含与所述第一有向链路对应的有向因果概率值。

[0014] 在一可实施方式中,所述筛选子模块,还用于筛选满足第二阈值的有向因果概率值,将与满足所述第二阈值的有向因果概率值对应的有向链路确定为第二有向链路;所述确定子模块,还用于根据所述多个指定变量和所述第二有向链路确定有向因果链路图,所述有向因果链路图用于表征所述指定变量之间的有向因果链路。

[0015] 在一可实施方式中,所述设备还包括:标注模块,用于对至少两个指定变量之间

进行因果链路标注,确定已知有向因果链路;学习模块,用于基于所述已知有向因果链路对所述因果概率预测模型进行半监督学习,以获得因果概率预测模型。

[0016] 本发明另一方面提供一种计算机可读存储介质,所述存储介质包括一组计算机可执行指令,当所述指令被执行时用于执行上述任一项所述的因果链路分析方法。

[0017] 在本发明实施例中,本发明实施例提供的因果链路分析方法用于对生产过程中的多个指定变量进行分析,以确定多个变量之间的因果关系,所获得的因果关系有利于确定生产过程中任一指定变量对整个生产过程的影响,从而能够通过精确地控制各项指定变量,保证生产过程的稳定性,保证由生产过程生成得到的产物的质量,有效地提升产物的收率,还可以达到节能减排的目的。

附图说明

[0018] 通过参考附图阅读下文的详细描述,本发明示例性实施方式的上述以及其他目的、特征和优点将变得易于理解。在附图中,以示例性而非限制性的方式示出了本发明的若干实施方式,其中:

[0019] 在附图中,相同或对应的标号表示相同或对应的部分。

[0020] 图1为本发明实施例一种因果链路分析方法的实现流程示意图;

[0021] 图2为本发明实施例一种因果链路分析方法分析指定变量的实现流程示意图;

[0022] 图3为本发明实施例一种因果链路分析方法相关度判断的实现流程示意图;

[0023] 图4为本发明实施例一种因果链路分析方法预测有向因果概率值的实现流程示意图;

[0024] 图5为本发明实施例一种因果链路分析方法确定有向因果链路的实现流程示意图;

[0025] 图6为本发明实施例一种因果链路分析方法建立有向完全图的场景示意图;

[0026] 图7为本发明实施例一种因果链路分析方法模型更新的场景示意图;

[0027] 图8为本发明实施例一种因果链路分析方法获得因果关系链路图的场景示意图;

[0028] 图9为本发明实施例一种因果链路分析设备的实现流程示意图。

具体实施方式

[0029] 为使本发明的目的、特征、优点能够更加的明显和易懂,下面将结合本发明实施例中的附图,对本发明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发明一部分实施例,而非全部实施例。基于本发明中的实施例,本领域技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0030] 图1为本发明实施例一种因果链路分析方法的实现流程示意图。

[0031] 参见图1,本发明实施例一方面提供一种因果链路分析方法,方法包括:操作101,对指定对象生产过程进行记录分析,获得与生产过程对应的多个指定变量;操作102,基于相关度判断规则确定多个指定变量之间的无向相关链路;操作103,根据因果概率预测模型对无向相关链路进行预测,获得与无向相关链路对应的有向因果概率值;操作104,根据有向因果概率值确定多个指定变量之间的有向因果链路,有向因果链路用于表征指定对象生产过程的因果关系。

[0032] 本发明实施例提供的因果链路分析方法用于对生产过程中的多个指定变量进行分析,以确定多个变量之间的因果关系,所获得的因果关系有利于确定生产过程中任一指定变量对整个生产过程的影响,从而能够通过精确地控制各项指定变量,保证生产过程的稳定性,保证由生产过程生成得到的产物的质量,有效地提升产物的收率,还可以达到节能减排的目的。

[0033] 其中,本方法的生产过程可以是各类领域中的生产过程,如石油化工领域、新材料领域、生物工程领域等。尤其适用于生产周期性长,生产过程中指定变量多,且指定变量之间具有互相强耦合的生产过程,即一个指定变量的变化往往会引起多个指定变量的改变,从而影响最终产物。例如,当该方法应用在催化裂化反应中,研究多个指定变量,即多个工艺参数和产物收率之间的因果链路关系,能够精确控制工艺参数,有效地提升收率;同时,可以实现装置故障智能检测及预警,提高企业效益。在本方法中,指定变量包括以下至少之一:用于表征工艺参数的变量、用于表征生成产物的变量。其中,本方法所指代的工艺参数包括以下至少之一:包括用于表征原材料参数的变量、用于表征设备参数的变量、用于表征环境参数的变量。

[0034] 具体的,在本方法中,首先,对指定对象生产过程进行记录分析,获得与生产过程对应的多个指定变量。其中,指定对象用于指代某一特定范围内的生产过程。该特定范围可以基于生产产物进行划分,也可以基于生产周期进行划分,还可以基于特定装置进行划分。例如,在一种情况下,本方法的指定对象生产过程可以为将特定原料通过生产获得特定产物的生产过程,其中,特定原料和特定产物可根据实际情况进行确定。在另一种情况下,本方法的指定对象生产过程可以为在某一特定产物在其中一个时间阶段内的生产过程。在再一种情况下,本方法的指定对象生成过程还可以为应用某一特定设备和/或系统所对应的生产过程。本方法以将特定原料通过生产获得特定产物的生产过程进行说明。其中,可以理解的是,特定原料和特定产物所对应的参数均包含在多个指定变量中。

[0035] 在确定指定对象生产过程为将特定原料通过生产获得特定产物的生产过程后,指定变量用于表征在生产过程中自身可以发生变化的参数信息、也用于表征可以导致其他变量发生变化的参数信息。可以通过对该生产过程所对应的生产记录进行分析,以确定在生产过程中的指定变量。可以理解的是,部分与生产过程无关的工艺参数也可能发生改变,在该操作中,这些不影响生产过程的工艺参数不属于指定变量。例如与生产过程无关的维修参数。

[0036] 在确定多个指定变量后,本方法基于相关度判断规则确定多个指定变量之间的无向相关链路。其中,相关度判断规则用于确定指定变量之间是否相关。需要解释的是,指定变量之间的是否相关定义为在其中一变量发生改变的时候,另一变量也会发生改变,此时定义为指定变量之间相关。例如,当第二变量发生改变时,第一变量和第三变量也会发生改变,此时可以确定第二变量和第一变量、第二变量和第三变量之间具有相关性。根据相关性判断规则确定所有指定变量之间是否具有相关性。根据具有相关性的指定变量确定无向相关链路。其中,无向相关链路有多个,且每一个无向相关链路分别用于连接两个具有相关性的指定变量。即,无向相关链路用于表征位于链路两端的指定变量之间具有相关性。

[0037] 在确定无向相关链路之后,本方法还包括,根据因果概率预测模型对无向相关链

路进行预测,获得与无向相关链路对应的有向因果概率值。因果概率预测模型用于预测无向相关链路和对应在无向相关链路两端的指定变量作为输入,根据预测结果输出因果概率值,有向因果概率值用于预测位于无向相关链路两端的指定变量之间是否具有因果关系。根据上述方式,可以对每一条无向相关链路和对应的指定变量进行预测,以确定对应无向相关链路的有向因果概率值。

[0038] 根据有向因果概率值确定多个指定变量之间的有向因果链路,有向因果链路用于表征指定对象生产过程的因果关系。

[0039] 其中,有向因果概率值的方向可通过正负进行表示。通过有向因果概率值,可以确定对应在无向相关链路两端的指定变量之间的方向性,即是否具有因果关系。例如,定义为数值满足预设范围内的有向因果概率值可以确定对应在无向相关链路两端的指定变量之间具有因果关系,并定义方向用于表征两指定变量之间的因果关系方向,在满足预设范围的情况下,当有向因果概率值为正数时,即确定为由第一变量导致第二变量发生改变,当有向因果概率值为负数时,即确定为由第二变量导致第一变量发生改变。根据上述方式,对每一个有向因果概率值进行确定,从而获得多个指定变量之间的有向因果链路。根据该有向因果链路可以确定指定变量之间的因果关系,从而可以基于该有向因果链路精确地控制各项指定变量,保证生产过程的稳定性。

[0040] 图2为本发明实施例一种因果链路分析方法分析指定变量的实现流程示意图。

[0041] 参见图2,在一可实施方式中,操作101,对指定对象生产过程进行记录分析,获得与生产过程对应的多个指定变量,包括:操作1011,通过分布式控制系统对指定对象运行过程进行记录,获得运行记录;操作1012,对运行记录进行筛选补充,获得与生产过程对应的关联信息;其中,生产过程包含在运行过程之中;操作1013,对关联信息进行标准化处理,获得指定变量。

[0042] 本方法对指定对象生产过程的分析包括,在生产过程中设置分布式控制系统(DCS系统)以对生产过程中的所有生产信息进行采集记录,获得运行记录。可以理解的是,运行记录包含了与生产过程有关的记录和与生产过程无关的记录。

[0043] 因此需要对运行记录进行筛选补充,具体的,本方法包括多种筛选补充方法。可以理解的是,本方法应用了以下筛选补充方法的至少之一:

[0044] 在第一种筛选补充方法中,本方法根据生产过程运行日报删除与生产过程无关的记录。例如,根据装置运行日报删除装置损坏及装置维修期间的所有数据,获得与生产过程对应的关联信息。

[0045] 在第二种筛选补充方法中,本方法统计多个指定变量的缺失值数量,删除缺失值占比超过预设阈值的所有变量,例如,设置预设阈值为10%,删除超过10%的所有变量,使用向上填充的方法填充各变量中存在的缺失值,以获得与生产过程对应的关联信息。

[0046] 在第三种筛选补充方法中,本方法通过对多个指定变量进行离群点处理,如采用 3σ 方法处理离群点,以获得与生产过程对应的关联信息。

[0047] 可以理解的是,上述筛选补充方法可以根据实际需要进行结合使用。例如在一种实施场景下,本方法首先通过分布式控制系统,收集汇总DCS数据;然后根据装置运行日报删除装置损坏及装置维修期间的所有数据,获得筛选变量;之后,统计筛选变量中各变量的缺失值数量,删除缺失值占比超过10%的所有变量,使用向上填充的方法填充各变量中

存在的缺失值,获得补充变量;然后对补充变量,使用 3σ 方法处理离群点,将处理后的变量确定为与生产过程对应的关联信息。

[0048] 之后,对关联信息进行标准化处理,获得指定变量。可以理解的是,关联信息中包含生产过程中产物的收率。标准化处理的具体内容为,根据预设的换算单位,对每一个指定变量进行单位换算,以使所有指定变量采用同一换算单位进行表征,将采用同一换算单位进行表征的变量确定为指定变量。

[0049] 图3为本发明实施例一种因果链路分析方法相关度判断的实现流程示意图。

[0050] 参见图3,在一可实施方式中,操作102,基于相关度判断规则确定多个指定变量之间的无向相关链路,包括:操作1021,建立与多个指定变量对应的有向完全图,有向完全图包含用于连接多个指定变量的多个双向链路,任一双向链路均包含两指向相反的有向链路;操作1022,根据多个指定变量确定与每一个有向链路对应的相关值;操作1023,筛选满足第一阈值的相关值,将与满足第一阈值的相关值对应的有向链路确定为第一有向链路;操作1024,根据多个指定变量和第一有向链路确定有向相关图,有向相关图用于表征指定变量之间的无向相关链路。

[0051] 本方法在进行相关度判断时,采用图表表征指定变量之间的无向相关链路,具体的,本方法采用有向相关图进行表征。本方法将指定变量作为顶点,对每个顶点通过双向链路进行连接,形成有向完全图。在有向完全图上,通过皮尔逊相关系数、互信息等方法来计算每两个指定变量之间的相关度,获得相关值,相关值用于对应每两个指定变量之间的双向链路。当相关值不满足第一阈值时,筛除与不满足第一阈值阈值的双向链路,即当相关值满足第一阈值时,保留与该相关值对应的双向链路,将保留的双向链路和顶点确定为有向相关途径。即有向相关图包含与多个指定变量对应的顶点、与满足第一阈值的相关值对应的双向链路,且每一个双向链路用于连接其中两个具有相关性的指定变量。其中,第一阈值用于表征满足相关性关系的阈值,根据实际情况进行预设。有向相关图可以表征指定变量之间的无向相关链路。

[0052] 图4为本发明实施例一种因果链路分析方法预测有向因果概率值的实现流程示意图。

[0053] 参见图4,在一可实施方式中,因果概率预测模型为图形神经网络模型;相应的,操作103,根据因果概率预测模型对无向相关链路进行预测,获得与无向相关链路对应的有向因果概率值,包括:操作1031,根据有向相关图确定变量集合与链路集合;其中,变量集合包含多个指定变量,链路集合包含第一有向链路;操作1032,通过图形神经网络模型对变量集合与链路集合进行预测,获得与链路集合对应的因果概率集,因果概率集包含与第一有向链路对应的有向因果概率值。

[0054] 本方法的因果概率预测模型为图形神经网络模型(Graph Neural Network,GNN),当本方法涉及指定变量较多时,通过图形神经网络模型对生产过程进行预测,能够更为直观、清楚地体现框架结果和预测结果。其中,本方法在利用因果概率预测模型进行预测时,以变量集合与链路集合作为输入,与第一有向链路对应的有向因果概率值作为输出。

[0055] 具体的,本方法建立与指定变量集合和无向相关链路集合对应的图模型 $G=(V,E)$,其中, V 为在有向相关图中,与指定变量对应的所有顶点的集合,即变量集合; E 为在有向相关图中,用于连接所有顶点的的所有第一有向链路的集合,即链路集合。需要说明的是,

第一有向链路包含第一正向链路和第一负向链路。图形神经网络模型对变量集合与链路集合进行预测,具体的,以获得有向因果概率值。进一步的,每一个有向因果概率值与有向链路对应。

[0056] 图5为本发明实施例一种因果链路分析方法确定有向因果链路的实现流程示意图。

[0057] 参见图5,在一可实施方式中,操作104,根据有向因果概率值确定多个指定变量之间的有向因果链路,包括:操作1041,筛选满足第二阈值的有向因果概率值,将与满足第二阈值的有向因果概率值对应的有向链路确定为第二有向链路;操作1042,根据多个指定变量和第二有向链路确定有向因果链路图,有向因果链路图用于表征指定变量之间的有向因果链路。

[0058] 第二阈值用于表征满足因果关系的阈值条件。即在有向因果概率值满足第二阈值的情况下,可以判断为与该有向因果概率值对应的有向链路用于表征指定变量之间具有因果关系,如此,将与满足第二阈值的有向因果概率值对应的有向链路确定为第二有向链路。其中,第二有向链路用于表征指定变量之间具有的因果关系及因果关系的方向。将多个指定变量确定为顶点,第二有向链路确定用于连接顶点,形成有向因果链路图,通过有向因果链路图,可以确定多个指定变量之间的因果关系。从而确定在指定对象的生产过程中,有向生产过程的所有指定变量之间的有向因果链路。

[0059] 在一可实施方式中,在操作102,基于相关度判断规则确定多个指定变量之间的无向相关链路之前,方法还包括:首先,对至少两个指定变量之间进行因果链路标注,确定已知有向因果链路;然后,基于已知有向因果链路对因果概率预测模型进行半监督学习,以获得因果概率预测模型。

[0060] 具体的,本方法设计的图形神经网络模型更新的示意图如图7所示:其中,通过设置 ϕ^v 进行卷积、池化操作,通过设置 $p^{v \rightarrow e}$ 为一条有向链路两端顶点数据进行堆叠,通过设置 ϕ^e 为单层LSTM网络,输出为因果关系概率值。进一步的,该模型还设置损失函数为交叉熵损失,设置迭代次数,进行半监督学习。其中,半监督学习的训练样本可以邀请专家利用对应指定对象生产过程的领域知识标注少量的直接因果关系,以及少量的无关关系。并对少量的直接因果关系,以及少量的无关关系进行标准化确定。对图形神经网络模型更新的示意图进行设计,包括但不限于,设置顶点更新、边更新的函数或者网络结构。

[0061] 为方便上述实施方式的理解,以下提供一种具体实施场景进行说明。在该场景中,本发明实施例提供的因果链路分析方法应用于因果链路分析设备,该设备用于对石油化工领域中生产过程复杂,工艺参数繁多,且工艺参数就有互相强耦合的生产过程进行分析,例如,石油炼化工艺。

[0062] 在分析过程中,设备首先通过分布式控制系统收集汇总分布式控制系统数据(DCS数据)。然后根据生产过程中所使用装置的运行日报删除装置损坏及装置维修期间的的数据,获得生产数据,对生产数据进行分析确定生产变量。其中,运行日报包含在分布式控制系统数据中。

[0063] 之后,统计生产变量中的缺失值数量,删除缺失值占比超过10%的变量,使用向上填充的方法填充删除后生产变量中存在的缺失值,获得填充变量。对填充变量使用 3σ 方法处理离群点,获得处理变量。对处理变量进行标准化处理,确定与生产过程对应的指定

变量,其中,指定变量为多个。

[0064] 再后,设备将指定变量作为顶点,通过双向链路连接指定变量,建立如图6所示的有向完全图,其中,图6中的圆点表征顶点,连接两圆点之间的两条箭头方向相反的线段表示双向链路。采用皮尔逊相关系数、互信息等方法计算与双向链路对应的相关度,以表征指定变量之间的相关度。相关度通过阈值th1进行评价,当两顶点之间的相关度不满足预设公式时,删除两顶点之间的双向链路,公式为: $\text{abs}(\text{corr}) \geq \text{th1}$,公式中,corr表示两顶点间的相关度,abs代表取绝对值,th1表示设定的阈值,将删除完与不满足预设公式相关度对应的双向链路所获得的图确定为有向相关图。

[0065] 然后,如图7所示,建立图模型 $G=(V,E)$,V为有向相关图中所有顶点的集合,每个顶点的初始值为顶点所对应变量的所有时间序列值;E为有向相关图中所有边的集合,设置初始值为全1矩阵。设计GNN模型的更新示意图,设置顶点更新、边更新的函数或者网络结构,可以设置 Φ^v 对顶点集合V和边集合E进行卷积、池化操作。设置 $\rho^{v,e}$ 为一条边的两端顶点数据进行堆叠,即对进行卷积、池化操作后的边集合中一条边的两端顶点数据进行堆叠,获得堆叠数据。可以设置 Φ^e 为单层LSTM网络,即用于根据输入的堆叠数据和边集合,输出对应的因果关系概率值 E' 。其中,可以邀请专家利用领域知识标注少量的直接因果关系,以及少量的无关关系,将专家标注的因果关系和无关关系标准化后作为训练样本,设置损失函数为交叉熵损失,设置迭代次数,对GNN模型进行半监督学习。通过GNN模型对与有向相关图对应的顶点集合和边集合进行预测,获得与每条边对应的有向因果概率值。

[0066] 最后,设置阈值th2,如图8所示,保留有向相关图中有向因果概率值大于阈值的边th2,从而得到因果关系链路图。其中,图8中的圆点与图6中的圆点指代相同的指定变量,图8中位于两圆点之间的有向线段用于表征两指定变量之间的因果关系,具体的,背向线段箭头一端的指定变量改变时,会导致朝向线段箭头的指定变量发生改变。

[0067] 图9为本发明实施例一种因果链路分析设备的实现流程示意图。

[0068] 参见图9,本发明另一方面提供一种因果链路分析设备,设备包括:分析模块601,用于对指定对象的生产过程进行记录分析,获得多个指定变量;确定模块602,用于基于相关度判断规则确定多个指定变量之间的无向相关链路;预测模块603,用于根据因果概率预测模型对无向相关链路进行预测,获得与无向相关链路对应的有向因果概率值;确定模块602,还用于根据有向因果概率值确定多个指定变量之间的有向因果链路。

[0069] 在一可实施方式中,分析模块601,包括:记录子模块6011,用于通过分布式控制系统对指定对象的运行过程进行记录,获得运行记录;筛选补充子模块6012,用于对运行记录进行筛选补充,获得与生产过程对应的关联信息;其中,生产过程包含在运行过程之中;处理子模块6013,用于对关联信息进行标准化处理,获得指定变量。

[0070] 在一可实施方式中,确定模块602,包括:建立子模块6021,用于建立与多个指定变量对应的有向完全图,有向完全图包含用于连接多个指定变量的多个双向链路,任一双向链路均包含两指向相反的有向链路;确定子模块6021,用于根据多个指定变量确定与每一个有向链路对应的相关值;筛选子模块6022,用于筛选满足第一阈值的相关值,将与满足第一阈值的相关值对应的有向链路确定为第一有向链路;确定子模块6021,还用于根据多个指定变量和第一有向链路确定有向相关图,有向相关图用于表征指定变量之间的无向相关链路。

[0071] 在一可实施方式中,因果概率预测模型为图形神经网络模型;相应的,预测模块603,包括:根据有向相关图确定变量集合与链路集合;其中,变量集合包含多个指定变量,链路集合包含第一有向链路;通过图形神经网络模型对变量集合与链路集合进行预测,获得与链路集合对应的因果概率集,因果概率集包含与第一有向链路对应的有向因果概率值。

[0072] 在一可实施方式中,筛选子模块6022,还用于筛选满足第二阈值的有向因果概率值,将与满足第二阈值的有向因果概率值对应的有向链路确定为第二有向链路;确定子模块6021,还用于根据多个指定变量和第二有向链路确定有向因果链路图,有向因果链路图用于表征指定变量之间的有向因果链路。

[0073] 在一可实施方式中,设备还包括:标注模块604,用于对至少两个指定变量之间进行因果链路标注,确定已知有向因果链路;学习模块605,用于基于已知有向因果链路对因果概率预测模型进行半监督学习,以获得因果概率预测模型。

[0074] 本发明另一方面提供一种计算机可读存储介质,存储介质包括一组计算机可执行指令,当指令被执行时用于执行上述任一项的因果链路分析方法。

[0075] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。而且,描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外,在不相互矛盾的情况下,本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0076] 此外,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”的特征可以明示或隐含地包括至少一个该特征。在本发明的描述中,“多个”的含义是两个或两个以上,除非另有明确具体的限定。

[0077] 以上,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以权利要求的保护范围为准。

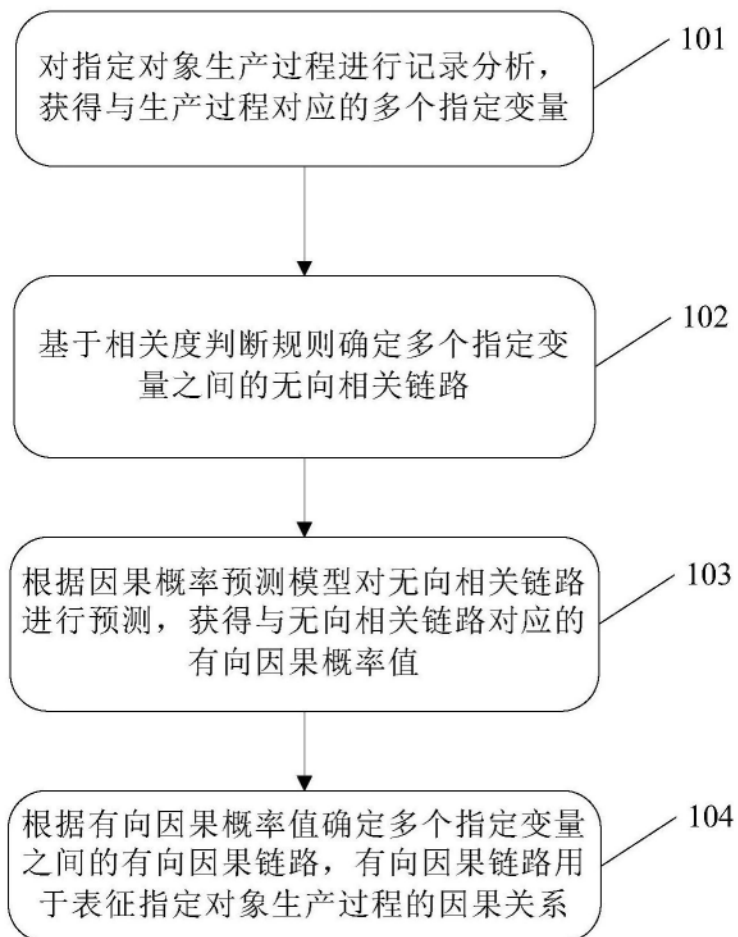


图1

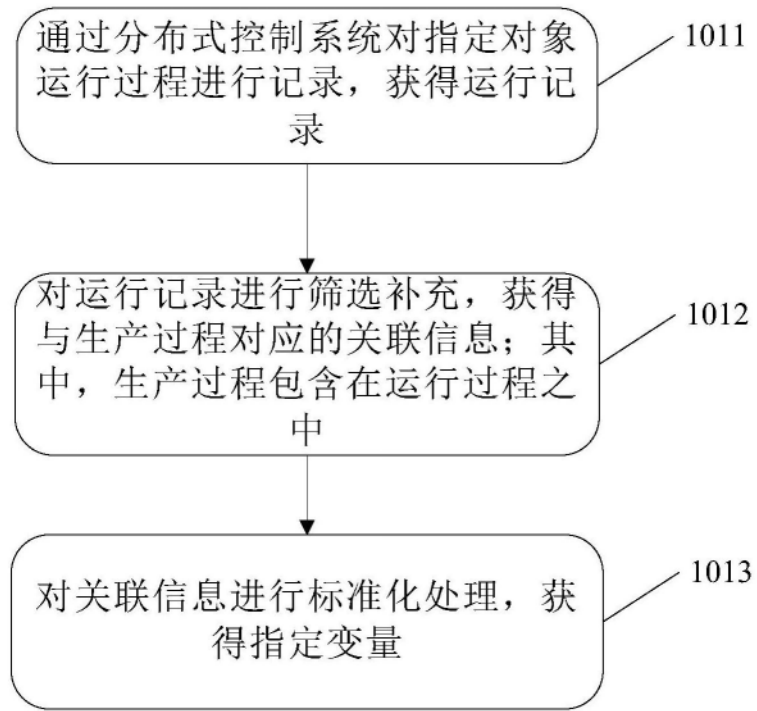


图2

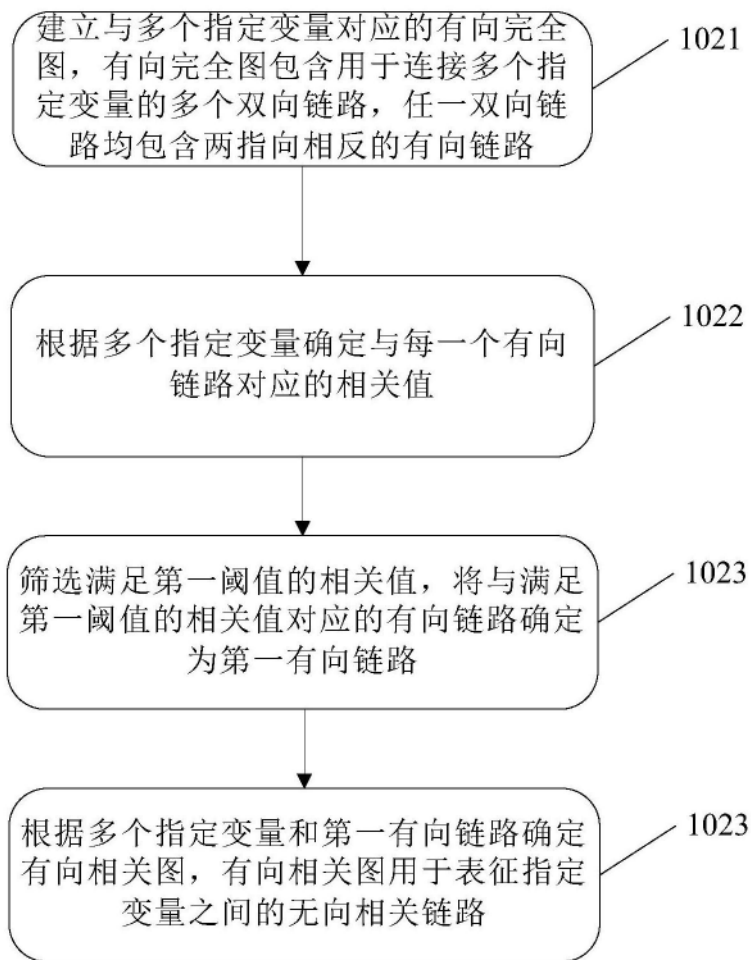


图3

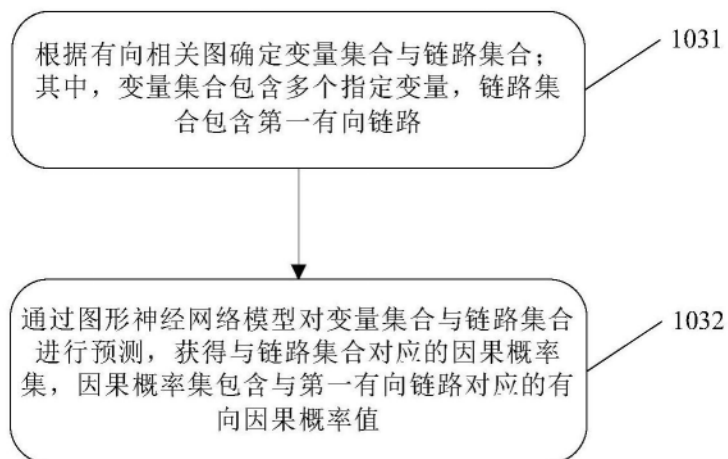


图4

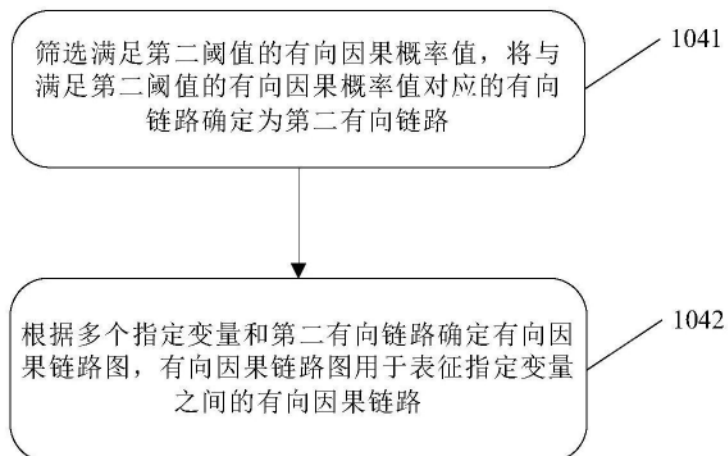


图5

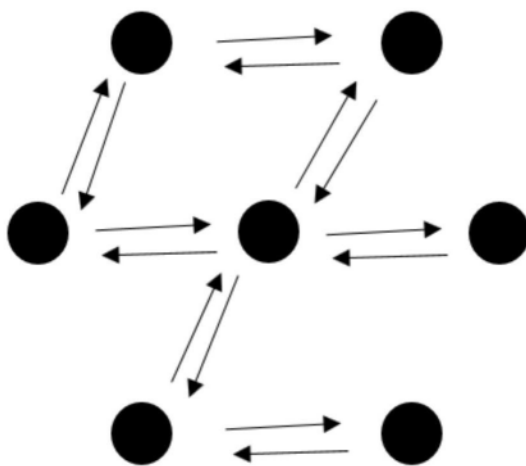


图6

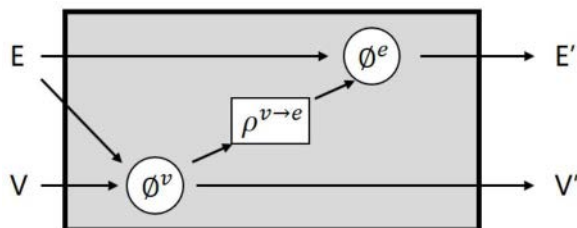


图7

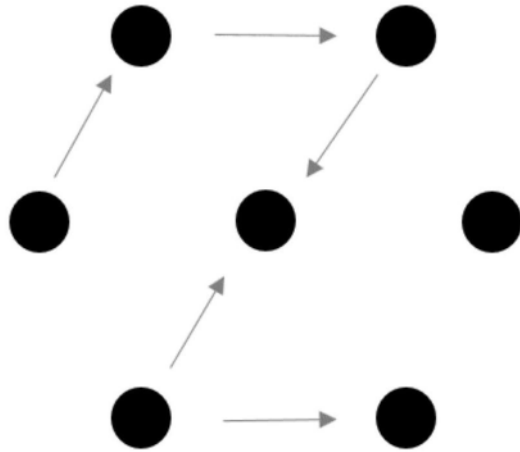


图8

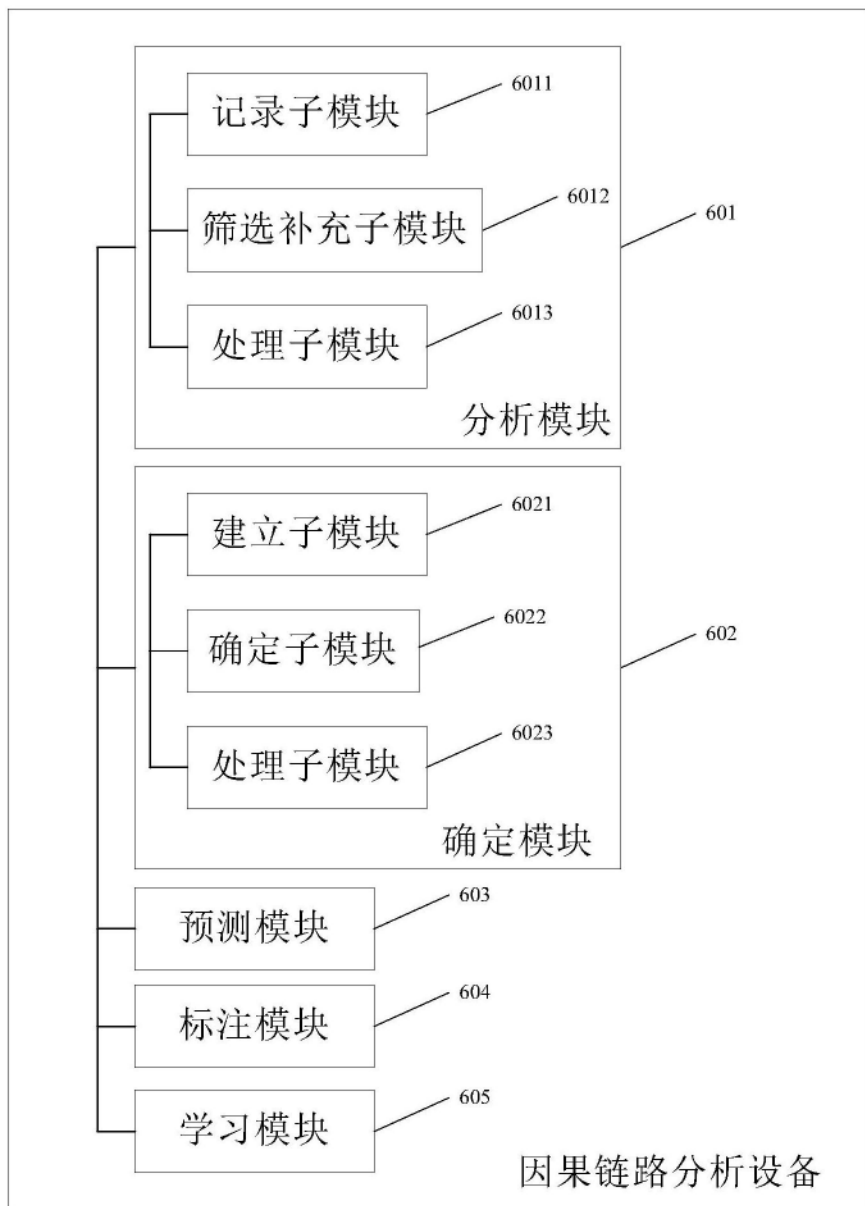


图9