

(12) 发明专利申请

(10) 申请公布号 CN 103365829 A

(43) 申请公布日 2013. 10. 23

(21) 申请号 201310092502. 7

(22) 申请日 2013. 03. 21

(30) 优先权数据

2012-072789 2012. 03. 28 JP

(71) 申请人 索尼公司

地址 日本东京都

(72) 发明人 矶崎隆司

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 朱胜 李春晖

(51) Int. Cl.

G06F 17/18 (2006. 01)

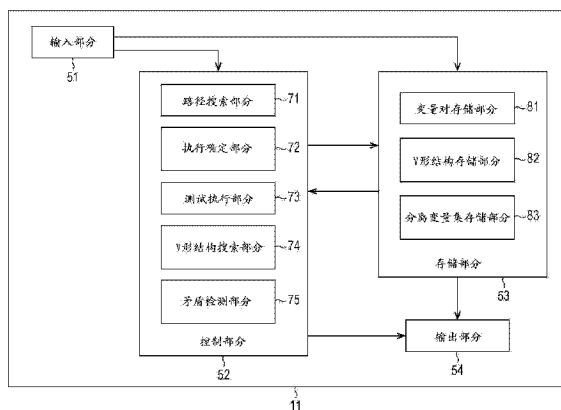
权利要求书2页 说明书14页 附图11页

(54) 发明名称

信息处理装置、信息处理方法和程序

(57) 摘要

本发明公开了一种信息处理装置、信息处理方法和程序。测试多个变量之间的独立性的信息处理装置包括执行部分和确定部分。执行部分在用作图形模型中至少不独立的两个变量之间的独立性的条件的条件变量设置在两个变量之间的路径上的情况下,执行两个变量之间的条件独立性的测试。确定部分确定V形结构是否存在于两个变量之间的路径上,V形结构是独立的第一变量和第二变量各自不独立于第三变量的图结构。在条件变量仅设置在被确定为具有V形结构的路径上的情况下,执行部分不执行两个变量之间的条件独立性的测试。



1. 一种测试多个变量之间的独立性的信息处理装置,包括:

执行部分,在用作图形模型中至少不独立的两个变量之间的独立性的条件的条件变量设置在所述两个变量之间的路径上的情况下,执行所述两个变量之间的条件独立性的测试;以及

确定部分,确定V形结构是否存在于所述两个变量之间的路径上,所述V形结构是独立的第一变量和第二变量各自不独立于第三变量的图结构,

其中,在所述条件变量仅设置在被确定为具有所述V形结构的路径上的情况下,所述执行部分不执行所述两个变量之间的条件独立性的测试。

2. 根据权利要求1所述的信息处理装置,

其中,在所述条件变量是所述V形结构中的所述第三变量的情况下,所述执行部分不执行所述两个变量之间的条件独立性的测试。

3. 根据权利要求1所述的信息处理装置,

其中,所述执行部分在执行条件独立性的测试之前执行所述第一变量和所述第二变量之间的无条件独立性的测试,

所述信息处理装置还包括:V形结构搜索部分,搜索所述第一变量和所述第二变量无条件独立并且所述第一变量和所述第二变量各自不独立于所述第三变量的图结构,作为所述V形结构。

4. 根据权利要求3所述的信息处理装置,

其中,所述V形结构搜索部分搜索所述第一变量和所述第二变量在除所述第三变量之外的变量的条件下条件独立并且所述第一变量和所述第二变量各自不独立于所述第三变量的图结构,作为所述V形结构。

5. 根据权利要求4所述的信息处理装置,

其中,在条件变量集的数量递增到超过预定数量的情况下,所述V形结构搜索部分不搜索所述V形结构。

6. 根据权利要求1所述的信息处理装置,

其中,所述确定部分针对所述两个变量之间的、数量与递增的条件变量集的数量相对应的路径,确定是否存在V形结构,以及

在所述条件变量中的任一个设置在被确定为具有所述V形结构的路径中的任一个上的情况下,所述执行部分不执行所述两个变量之间的条件独立性的测试。

7. 根据权利要求1所述的信息处理装置,还包括:

矛盾检测部分,检测多个所述V形结构中的有向边之间的矛盾,以将一些或全部所述有向边改变为无向边来解决所述矛盾。

8. 根据权利要求7所述的信息处理装置,

其中,所述矛盾检测部分使用在多个所述V形结构中的所述第一变量和所述第二变量之间的独立性的测试中使用的索引,将具有独立性较弱的所述第一变量和所述第二变量的所述V形结构的有向边改变为无向边。

9. 根据权利要求7所述的信息处理装置,

其中,在变量X和变量Y各自不独立于变量Z的第一V形结构、以及变量W和变量Y各自不独立于变量Z的第二V形结构中所述变量X和所述变量W不独立的情况下,所述矛盾

检测部分将所述第一 V 形结构中从所述变量 X 到所述变量 Z 的有向边和所述第二 V 形结构中从所述变量 W 到所述变量 Z 的有向边改变为无向边。

10. 根据权利要求 7 所述的信息处理装置，

其中，在变量 X 和变量 Y 各自不独立于变量 Z 的第一 V 形结构、以及变量 W 和变量 Y 各自不独立于变量 Z 的第二 V 形结构中所述变量 X 和所述变量 W 不独立的情况下，所述矛盾检测部分将所述第一 V 形结构和所述第二 V 形结构中的全部所述有向边改变为无向边。

11. 根据权利要求 7 所述的信息处理装置，

其中，在变量 X 和变量 Y 各自不独立于变量 Z 的第一 V 形结构、以及变量 W 和变量 Y 各自不独立于变量 Z 的第二 V 形结构中所述变量 X 和所述变量 W 不独立的情况下，所述矛盾检测部分使用两个变量之间的独立性的测试中使用的索引，比较所述变量 X 和所述变量 Y 之间的独立性以及所述变量 W 和所述变量 Y 之间的独立性，以将具有独立性较弱的两个变量的所述 V 形结构的有向边改变为无向边。

12. 一种用于测试多个变量之间的独立性的信息处理装置的信息处理方法，包括：

在用作图形模型中至少不独立的两个变量之间的独立性的条件的条件变量设置在所述两个变量之间的路径上的情况下，执行所述两个变量之间的条件独立性的测试；以及

确定 V 形结构是否存在于所述两个变量之间的路径上，所述 V 形结构是独立的第一变量和第二变量各自不独立于第三变量的图结构，

其中，在所述条件变量仅设置在被确定为具有所述 V 形结构的路径上的情况下，不执行所述两个变量之间的条件独立性的测试。

13. 一种使计算机执行用于测试多个变量之间的独立性的处理的程序，所述处理包括：

在用作图形模型中至少不独立的两个变量之间的独立性的条件的条件变量设置在所述两个变量之间的路径上的情况下，执行所述两个变量之间的条件独立性的测试；以及

确定 V 形结构是否存在于所述两个变量之间的路径上，所述 V 形结构是独立的第一变量和第二变量各自不独立于第三变量的图结构，

其中，在所述条件变量仅设置在被确定为具有所述 V 形结构的路径上的情况下，不执行所述两个变量之间的条件独立性的测试。

信息处理装置、信息处理方法和程序

技术领域

[0001] 本技术涉及信息处理装置、信息处理方法和程序,尤其涉及一种能够增强多个变量之间的估计因果关系的可靠性的信息处理装置、信息处理方法和程序。

背景技术

[0002] 根据现有技术的根据关于多元随机变量的观察数据来估计统计因果关系大致分为如下方法:通过信息标准、具有惩罚的最大似然法、或贝叶斯法将估计结果获得为分数并且使分数最大化的方法,以及通过变量之间的条件独立性的统计测试来估计变量之间的因果关系的方法。为了良好的结果可读性,作为结果的变量之间的因果关系常被表示为图形模型(非循环模型)。

[0003] 图 1 示出表示变量 X 和变量 Y 之间的因果关系的图形模型的两个示例。

[0004] 在图 1 的上部示出的图形模型中,变量 X 和变量 Y 之间的因果关系未经确认,并且变量 X 和变量 Y 用作由非定向边(无向边)链接的顶点。在图 1 的中部示出的图形模型中,变量 X 和变量 Y 之间的因果关系是变量 X 对应于因并且变量 Y 对应于果,以及变量 X 和变量 Y 用作由指示从因到果的方向的定向边(有向边)链接的顶点。在图 1 的下部示出的图形模型中,三个变量和用于链接变量的边链接的顶点。在图 1 的下部示出的图形模型中,三个变量和用于链接变量的边形成变量 X 和变量 Y 之间的路径,并且该路径可部分地包括有向边。

[0005] 通过变量之间的条件独立性的统计测试来估计变量之间的因果关系的方法已变得重要,因为该方法可能会估计潜在公共因变量的存在并且有向边的方向的理由基于物理背景。然而,另一方面,由于统计测试的检测能力不足,所以估计结果的可靠性低。

[0006] 在该方法的初始尝试中,为了执行 n 元随机变量中的两个变量之间的条件独立性的测试,考虑从 $(n-2)$ 个变量中提取所有变量组合作为用作条件独立性的条件的条件变量集,该变量组合的数量最小为 0 并且最大为 $(n-2)$,并且以循环(round-robin)的方式执行测试。然而,在这种情况下,变量组合的数量成指数增加,并且使用计算器执行计算是不实际的。

[0007] 稍后公开了一种显著降低测试所需计算量的算法(参见 P. Spirtes, C. Glymour, R. Scheines, "Causation, Prediction, and Search", MIT Press, 第二版, 2000 年)。如果假设变量之间的条件独立性由有向非循环图唯一地表示,则唯一地确定使两个特定变量条件地独立的条件变量集,并且条件变量不与关注的变量条件地独立。在这样的条件下,在以升序增加条件变量集的数量的同时执行测试,并且在独立性未被拒绝的情况下,立即移除两个变量之间的边。然而,这种算法仍涉及大量测试尝试,这可能导致频繁的测试错误。

[0008] 为了解决这种问题,P. Spirtes 等也公开了对以上讨论的算法的改进,其进一步降低了在条件变量设置在关注的两个变量之间的路径上的条件下的计算量。然而,在算法的执行期间,保留稍后被确定为独立的边,因此允许遵循长路径。因此,可能无法有效地显著降低测试中的尝试次数。

[0009] 提出了一种通过递归地将整个图划分为小的子图来降低具有大量条件变量集的独立性测试的尝试次数的算法(参见 R. Yehezkel, B. Lerner, "Bayesian Network Structure Learning by Recursive Autonomy Identification", Journal of Machine Learning Research, Vol. 10, pp. 1527-1570, 2009 年)。然而,该算法可能无法抑制如下测试错误的出现:在具有少量条件变量集的独立性测试中,内在地不独立的变量被确定为独立的。

[0010] 另外, X. Xie, Z. Geng, "A Recursive Method for Structural Learning of Directed Acyclic Graphs", Journal of Machine Learning Research, Vol. 9, pp. 459-483, 2008 年公开了以与根据 R. Yehezkel 等的方法不同的方法来执行递归处理。然而,在根据 X. Xie 等的方法中,独立性测试需要大量条件变量,这可能导致缺少计算稳定性。

发明内容

[0011] 因此,尚不存在通过条件独立性测试来充分地提高多个变量之间的因果关系的估计的低可靠性的方法。

[0012] 所以,期望增强多个变量之间的估计因果关系的可靠性。

[0013] 根据本技术的实施例,提供了一种测试多个变量之间的独立性的信息处理装置,包括:执行部分,在用作图形模型中至少不独立的两个变量之间的独立性的条件的条件变量设置在所述两个变量之间的路径上的情况下,执行所述两个变量之间的条件独立性的测试;以及确定部分,确定 V 形结构是否存在于所述两个变量之间的路径上, V 形结构是独立的第一变量和第二变量各自不独立于第三变量的图结构,其中,在条件变量仅设置在被确定为具有 V 形结构的路径上的情况下,所述执行部分不执行所述两个变量之间的条件独立性的测试。

[0014] 在条件变量是 V 形结构中的第三变量的情况下,所述执行部分可不执行所述两个变量之间的条件独立性的测试。

[0015] 所述执行部分可在执行条件独立性的测试之前执行第一变量和第二变量之间的无条件独立性的测试,以及所述信息处理装置还可包括:V 形结构搜索部分,搜索第一变量和第二变量无条件独立并且第一变量和第二变量各自不独立于第三变量的图结构,作为 V 形结构。

[0016] 所述 V 形结构搜索部分可搜索第一变量和第二变量在除第三变量之外的变量的条件下条件独立并且第一变量和第二变量各自不独立于第三变量的图结构,作为 V 形结构。

[0017] 在条件变量集的数量递增到超过预定数量的情况下,所述 V 形结构搜索部分可不搜索 V 形结构。

[0018] 所述确定部分可针对所述两个变量之间的、数量与递增的条件变量集的数量相对应的路径,确定是否存在 V 形结构,以及在条件变量中的任一个设置在被确定为具有 V 形结构的路径中的任一个上的情况下,所述执行部分可不执行所述两个变量之间的条件独立性的测试。

[0019] 所述信息处理装置还可包括:矛盾检测部分,检测多个 V 形结构中的有向边之间

的矛盾,以将一些或全部有向边改变为无向边来解决矛盾。

[0020] 所述矛盾检测部分可使用在所述多个 V 形结构中的第一变量和第二变量之间的独立性的测试中所使用的索引,将具有独立性较弱的第一变量和第二变量的 V 形结构的有向边改变为无向边。

[0021] 在变量 X 和变量 Y 各自不独立于变量 Z 的第一 V 形结构、以及变量 W 和变量 Y 各自不独立于变量 Z 的第二 V 形结构中变量 X 和变量 W 不独立的情况下,所述矛盾检测部分可将第一 V 形结构中从变量 X 到变量 Z 的有向边和第二 V 形结构中从变量 W 到变量 Z 的有向边改变为无向边。

[0022] 在变量 X 和变量 Y 各自不独立于变量 Z 的第一 V 形结构、以及变量 W 和变量 Y 各自不独立于变量 Z 的第二 V 形结构中变量 X 和变量 W 不独立的情况下,所述矛盾检测部分可将第一 V 形结构和第二 V 形结构中的全部有向边改变为无向边。

[0023] 在变量 X 和变量 Y 各自不独立于变量 Z 的第一 V 形结构、以及变量 W 和变量 Y 各自不独立于变量 Z 的第二 V 形结构中变量 X 和变量 W 不独立的情况下,所述矛盾检测部分可使用两个变量之间的独立性的测试中所使用的索引来比较变量 X 和变量 Y 之间的独立性、以及变量 W 和变量 Y 之间的独立性,以将具有独立性较弱的两个变量的 V 形结构的有向边改变为无向边。

[0024] 根据本技术的实施例,提供了一种用于测试多个变量之间的独立性的信息处理装置的信息处理方法,包括:在用作图形模型中至少不独立的两个变量之间的独立性的条件的条件变量设置在所述两个变量之间的路径上的情况下,执行所述两个变量之间的条件独立性的测试;以及确定 V 形结构是否存在于所述两个变量之间的路径上,V 形结构是独立的第一变量和第二变量各自不独立于第三变量的图结构,其中,在条件变量仅设置在被确定为具有 V 形结构的路径上的情况下,不执行所述两个变量之间的条件独立性的测试。

[0025] 根据本技术的实施例,提供了一种使计算机执行用于测试多个变量之间的独立性的处理的程序,所述处理包括:在用作图形模型中至少不独立的两个变量之间的独立性的条件的条件变量设置在所述两个变量之间的路径上的情况下,执行所述两个变量之间的条件独立性的测试;以及确定 V 形结构是否存在于所述两个变量之间的路径上,V 形结构是独立的第一变量和第二变量各自不独立于第三变量的图结构,其中,在条件变量仅设置在被确定为具有 V 形结构的路径上的情况下,不执行所述两个变量之间的条件独立性的测试。

[0026] 在本技术的实施例中,在用作两个变量之间的独立性的条件的条件变量设置在所述两个变量之间的路径上的情况下,执行所述两个变量之间的条件独立性的测试。确定 V 形结构是否存在于所述两个变量之间的路径上,V 形结构是独立的第一变量和第二变量各自不独立于第三变量的图结构。在条件变量仅设置在被确定为具有 V 形结构的路径上的情况下,不执行所述两个变量之间的条件独立性的测试。

[0027] 根据本技术的实施例,能够增强多个变量之间的估计因果关系的可靠性。

附图说明

[0028] 图 1 示出图形模型的示例;

[0029] 图 2 是示出根据本技术的实施例的信息处理装置的硬件配置示例的框图;

[0030] 图 3 是示出根据本技术的实施例的信息处理装置的功能配置示例的框图;

- [0031] 图 4 图示 V 形结构；
- [0032] 图 5 是图示独立性测试处理的流程图；
- [0033] 图 6 是图示独立性测试处理的流程图；
- [0034] 图 7 图示对 V 形结构的搜索；
- [0035] 图 8 图示两个变量之间的路径；
- [0036] 图 9 图示两个变量之间的路径；
- [0037] 图 10 是图示矛盾解决处理的流程图；
- [0038] 图 11 图示矛盾解决处理的具体示例；
- [0039] 图 12 图示矛盾解决处理的具体示例；
- [0040] 图 13 图示矛盾解决处理的具体示例；以及
- [0041] 图 14 图示矛盾解决处理的具体示例。

具体实施方式

- [0042] 以下将参照附图描述本技术的实施例。
- [0043] [信息处理装置的硬件配置示例]
- [0044] 图 2 示出根据本技术的实施例的信息处理装置 11 的硬件配置示例。
- [0045] 信息处理装置 11 测试离散多元随机变量之间的独立性或条件独立性，并将测试结果输出为贝叶斯网络，贝叶斯网络是一种图形模型。
- [0046] 信息处理装置 11 例如可由个人计算机组成，并可具有与个人计算机的配置类似的配置。
- [0047] 信息处理装置 11 包括 CPU（中央处理单元）21、ROM（只读存储器）22、RAM（随机存取存储器）23、总线 24、输入 / 输出接口 25、输入部分 26、输出部分 27、存储部分 28、通信部分 29 和驱动器 30。
- [0048] 在信息处理装置 11 中，CPU21、ROM22 和 RAM23 通过总线 24 彼此连接。输入 / 输出接口 25 还连接到总线 24。诸如键盘、鼠标和触摸板的输入部分 26、诸如显示器和扬声器的输出部分 27、诸如硬盘驱动器和非易失性存储器的存储部分 28、以及诸如网络接口的通信部分 29 连接到输入 / 输出接口 25。
- [0049] 此外，驱动器 30 根据需要连接到输入 / 输出接口 25。诸如磁盘、光盘、磁光盘和半导体存储器的可移除介质 31 视情况安装到驱动器 30。从可移除介质 31 读取的程序根据需要安装在存储部分 28 中。
- [0050] 替代地，可由通信部分 29 经由有线或无线传输介质接收程序以安装在存储部分 28 中。另外替代地，程序可预先安装在 ROM22 或存储部分 28 中。
- [0051] 由信息处理装置 11 执行的程序可根据本文中描述的顺序按时间先后顺序进行处理，或者可并行地或在合适的定时（例如，当进行调用时）进行处理。
- [0052] [信息处理装置的功能配置示例]
- [0053] 图 3 示出与本技术的实施例有关的信息处理装置 11 的一部分的功能配置示例。
- [0054] 图 3 的信息处理装置 11 包括输入部分 51、控制部分 52 以及存储部分 53 和输出部分 54。
- [0055] 输入部分 51 对应于图 2 的输入部分 26。输入部分 51 接收如下输入：用于在 N 个

离散随机变量之中指定要进行独立性测试的两个变量的自变量、用作条件独立性的条件的条件变量的变量集等,并将对应于该输入的内容的信息提供到控制部分 52。

[0056] 控制部分 52 对应于图 2 的 CPU21。控制部分 52 根据存储在对应于图 2 的存储部分 28 的存储部分 53 中的程序进行操作,并使用存储在存储部分 53 中的各种信息执行所关注的两个变量之间的独立性的测试。

[0057] 输出部分 54 对应于图 2 的输出部分 27。输出部分 54 在控制部分 52 的控制下将独立性测试的结果输出为图形模型。

[0058] 如图 3 中所示,控制部分 52 包括路径搜索部分 71、执行确定部分 72、测试执行部分 73、V 形结构搜索部分 74 和矛盾检测部分 75,以及存储部分 53 包括变量对存储部分 81、V 形结构存储部分 82 和分离变量集存储部分 83。

[0059] 例如,路径搜索部分 71 使用诸如深度优先搜索的通用算法来搜索所关注的两个变量之间的路径。

[0060] 执行确定部分 72 确定是否需要针对所关注的两个变量执行独立性测试。

[0061] 测试执行部分 73 针对所关注的两个变量执行独立性测试。然而,应该注意,在执行确定部分 72 确定不需要执行独立性测试的情况下,测试执行部分 73 不执行独立性测试。

[0062] V 形结构搜索部分 74 针对被确定为独立的两个变量搜索 V 形结构。例如,如图 4 所示,V 形结构是这样的图结构,其中变量 X 和变量 Y 独立,变量 X 和变量 Z 不独立,以及变量 Y 和变量 Z 不独立。

[0063] 矛盾检测部分 75 从 V 形结构搜索部分 74 找到的 V 形结构中检测彼此矛盾的 V 形结构,并解决 V 形结构之间的矛盾。

[0064] 变量对存储部分 81 存储这样的变量对,其中不独立或未进行独立性测试的两个变量通过边而链接。

[0065] V 形结构存储部分 82 存储 V 形结构搜索部分 74 找到的 V 形结构。执行确定部分 72 根据与存储在 V 形结构存储部分 82 中的 V 形结构相匹配的任意图结构是否设置在所关注的两个变量之间的路径上,确定是否需要针对两个变量执行独立性测试。

[0066] 分离变量集存储部分 83 存储设置在由测试执行部分 73 执行的独立性测试中被确定为条件地独立的两个变量之间的路径上的条件变量,作为用于分离两个变量的分离变量集。

[0067] [由信息处理装置执行的独立性测试处理]

[0068] 接下来,将参照图 5 和图 6 的流程图来描述由信息处理装置 11 执行的独立性测试处理。存储部分 53 预先存储随机变量(在下文中可被简称为“变量”)的数量 N、以及各变量可取的内部状态的数量(2 或更多)。当通过输入部分 51 输入用于描述所有变量的状态的 M 个数据时,独立性测试处理开始。

[0069] 在初始状态中,变量对存储部分 81 存储这样的变量对,其中 N 个变量与不同的 (N-1) 个变量成对(通过边而链接);以及 V 形结构存储部分 82 和分离变量集存储部分 83 不存储数据。

[0070] 在步骤 S11 中,控制部分 52 将条件变量集(在下文中可被称为“条件变量”)的数量 i 设置为 0,即使条件变量集为空集。

[0071] 在步骤 S12 中,控制部分 52 从存储在变量对存储部分 81 中的变量对中选择一个

变量对。

[0072] 在步骤 S13 中,测试执行部分 73 计算 G^2 统计量, G^2 统计量是用来执行独立性测试的值。 G^2 统计量由下式(1)给出。

$$[0073] \quad G^2 = 2M \sum_{x, y, z} \hat{P}(x, y, z) \log \frac{\hat{P}(x|y, z)}{\hat{P}(x|z)} \quad \dots (1)$$

[0074] 在式(1)中,M 指示数据的数量。附加有符号“^” (帽子(hat))的 $P(x, y, z)$ (例如,在下文中被称为“戴帽子的 $P(x, y, z)$ ”)、戴帽子的 $P(x|y, z)$ 和戴帽子的 $P(x|z)$ 分别表示变量 X 和 Y 以及变量 Z 的集合的特定状态 x、y、z 的估计联合概率、状态 x 在状态 y 和 z 的条件下的估计条件概率、状态 x 在状态 z 的条件下的估计条件概率。使用 M 个数据来估计这些概率。在步骤 S13 中,变量 Z 的集合为空集,因此不考虑状态 z。

[0075] 这里,在 p 值和预定的阈值 Th (例如,5% (0.05))之间进行比较,以测试所选择的变量对的两个变量之间的独立性。使用以上讨论的 G^2 统计量和 χ^2 分布来推导 p 值,以用作独立性测试的索引。

[0076] 即,在步骤 S14 中,测试执行部分 73 确定 p 值是否大于阈值 Th。

[0077] 如果在步骤 S14 中确定 p 值大于阈值 Th,则处理前进至步骤 S15,在步骤 S15 中测试执行部分 73 接受所选择的变量对的两个变量之间的独立性。

[0078] 然后,在步骤 S16 中,测试执行部分 73 从变量对存储部分 81 中删除所选择的变量对。

[0079] 如果在步骤 S14 中确定 p 值不大于阈值 Th,则处理前进至步骤 S17,在步骤 S17 中测试执行部分 73 拒绝所选择的变量对的两个变量之间的独立性。

[0080] 在以上提供的描述中,p 值用作独立性测试的索引。然而,由下式(2)指示的条件互信息量 MI 可用作独立性测试的索引。

$$[0081] \quad MI = \frac{G^2}{2M} \quad \dots (2)$$

[0082] 在这种情况下,在条件互信息量 MI 和预定的阈值(例如,0.05)之间进行比较。如果确定条件互信息量 MI 小于阈值,则接受两个变量之间的独立性。

[0083] 在步骤 S16 或步骤 S17 之后的步骤 S18 中,控制部分 52 确定是否已针对存储在变量对存储部分 81 中的所有变量执行了步骤 S12 至步骤 S17 中的处理,即,是否已针对存储在变量对存储部分 81 中的所有变量对执行了无条件独立性测试。

[0084] 如果在步骤 S18 中确定没有处理所有变量对,则处理返回到步骤 S12,以针对新选择的变量对重复步骤 S12 至步骤 S17 中的处理。

[0085] 另一方面,如果在步骤 S18 中确定已处理了所有变量对,则处理前进至步骤 S19。此时,变量对存储部分 81 仅存储具有这样的两个变量的变量对,这两个变量之间的独立性已被拒绝(不独立的两个变量)。

[0086] 在步骤 S19 中,控制部分 52 确定是否存在以上讨论的处理中找到的两个独立变量。

[0087] 如果在步骤 S19 中确定存在两个独立变量的至少一个集合,则处理前进至步骤 S20,在步骤 S20 中,V 形结构搜索部分 74 针对两个独立变量搜索 V 形结构。具体地,在如

图 4 所示的变量 X 和 Y 的特定集合独立并且各自不独立于变量 Z 的情况下,即,在变量 X 和 Y 具有路径(X-Z-Y)(其包括两条边以及在变量 X 和 Y 之间的变量 Z)、并且变量 Z 不作为条件变量有助于变量 X 和 Y 之间的独立性的情况下,V 形结构搜索部分 74 使 V 形结构存储部分 82 将这样的图结构存储为 V 形结构(例如,在下文中被表示为“ $X \rightarrow Z \leftarrow Y$ ”)。

[0088] 另一方面,如果在步骤 S19 中确定不存在两个独立变量的集合,则跳过步骤 S20。

[0089] 此后,处理前进至图 6 的步骤 S21,在步骤 S21 中,控制部分 52 将条件变量集(条件变量)的数量 i 递增 1。即,将条件变量集的数量设置为 1。

[0090] 在步骤 S22 中,控制部分 52 从存储在变量对存储部分 81 中的变量对中选择一个变量对,即,从具有不独立的两个变量的变量对中选择一个变量对。这里,假设已选择了具有变量 S 和变量 T 的变量对。

[0091] 在步骤 S23 中,控制部分 52 选择可用作所选择的变量对的两个变量之间的独立性的条件的条件变量。具体地,控制部分 52 参考存储在变量对存储部分 81 中的变量对,以将与所选择的变量对的两个变量中的任意变量(例如,变量 S)成对的变量之一选择为条件变量。这里,假设已选择了变量 W。

[0092] 在步骤 S24 中,路径搜索部分 71 搜索所选择的变量对的两个变量之间的路径,并确定条件变量是否设置在该路径上。

[0093] 例如,如果在步骤 S24 中确定变量 W 设置在变量 S 和变量 T 之间的路径上,则处理前进至步骤 S25,在步骤 S25 中,执行确定部分 72 确定与存储在 V 形结构存储部分 82 中的 V 形结构匹配的任意图结构是否设置在变量 S 和变量 T 之间的路径上。

[0094] 如果在步骤 S25 中确定存在任何匹配的 V 形结构,则变量 W 不能用作路径上变量 S 和 T 之间的独立性的条件变量。因此,在步骤 S26 中,路径搜索部分 71 搜索另一路径。

[0095] 在步骤 S27 中,确定是否存在任何其他路径。如果找到另一路径,则处理返回到步骤 S24,并且再次执行步骤 S24 和 S25 中的处理。如果在步骤 S24 中确定变量 W 没有设置在变量 S 和变量 T 之间的路径上,则处理返回到步骤 S23,在步骤 S23 中,选择另一条件变量。

[0096] 另一方面,如果在步骤 S25 中确定不存在匹配的 V 形结构,则变量 W 可能是路径上变量 S 和 T 之间的独立性的条件变量,并且处理前进至步骤 S28。

[0097] 步骤 S28 至步骤 S31、和步骤 S33 中的处理分别与以上讨论的步骤 S13 至步骤 S17 中的处理相同,因此将不进行描述。这里,例如,执行在变量 W 的条件下的变量 S 和 T 之间的条件独立性的测试。如果接受在变量 W 的条件下的变量 S 和 T 之间的条件独立性,则从变量对存储部分 81 中删除具有变量 S 和 T 的变量对。

[0098] 然后,在步骤 S32 中,测试执行部分 73 使得分离变量集存储部分 83 将条件变量(变量 W)存储为分离变量集。

[0099] 如果在步骤 S27 中确定是否存在任何其他路径并且没有找到其他路径,即,如果条件变量仅设置在被确定为具有 V 形结构的路径上,则跳过步骤 S28 至步骤 S33 中的处理。即,在这种情况下,不执行在变量 W 的条件下的变量 S 和 T 之间的条件独立性的测试,并且变量 S 和 T 被保留为变量对(通过边而链接)。

[0100] 这里,如果变量 S 和 T 条件地独立,则变量 S 和 T 将已在无条件独立性测试(步骤 S13 至步骤 S17 中的处理)中被确定为独立。这基于 P. Spirtes 等中公开的被称为“d-分离(d-separation)”和“忠实性(faithfulness)”的理论。此外,在这种情况下,变量 W 不能

是变量 S 和 T 之间的独立性的条件变量。因此,清楚的是,变量 S 和 T 不是条件地独立的,因此不执行变量 S 和 T 之间的条件独立性的测试。

[0101] 在步骤 S32、步骤 S33 或步骤 S27 之后的步骤 S34 中,控制部分 52 确定是否已针对存储在变量对存储部分 81 中的所有变量对执行了步骤 S22 至步骤 S33 中的处理,即,是否已针对存储在变量对存储部分 81 中的所有变量对执行了条件独立性测试。

[0102] 如果在步骤 S34 中确定未处理所有变量对,则处理返回到步骤 S22,以针对新选择的变量对重复步骤 S22 至步骤 S33 中的处理。

[0103] 另一方面,如果在步骤 S34 中确定已处理了所有变量对,则处理前进至步骤 S35。此时,变量对存储部分 81 仅存储具有这样的两个变量的变量对,这两个变量之间的独立性已被拒绝(不独立的两个变量)、或这两个变量未进行条件独立性测试。

[0104] 在步骤 S35 中,控制部分 52 确定是否存在在以上讨论的处理中找到的两个独立(条件地独立)变量。

[0105] 如果在步骤 S35 中确定存在两个独立(条件地独立)变量的至少一个集合,则处理前进至步骤 S36,在步骤 S36 中,V 形结构搜索部分 74 针对两个独立变量搜索 V 形结构。具体地,例如,在如图 7 中所示的变量 X 和 Y 的特定集合在变量 W 的条件下条件地独立并且各自不独立于变量 Z、以及变量 Z 未作为变量 X 和 Y 的条件变量(分离变量集)存储在分离变量集存储部分 83 中的情况下,V 形结构搜索部分 74 使得 V 形结构存储部分 82 将这样的图结构存储为 V 形结构。

[0106] 然而,应该注意,独立性测试中的错误可能导致这里获得的 V 形结构中的错误,并且如稍后所讨论的,条件变量集的数量增加可能使测试中的这样的错误更容易出现。因此,即使这里获得的 V 形结构设置在正在测试的两个变量之间的路径上,也可执行测试。

[0107] 另一方面,如果在步骤 S35 中确定不存在两个独立(条件地独立)变量的集合,则跳过步骤 S36。

[0108] 在步骤 S37 中,控制部分 52 确定条件变量集(条件变量)的数量 i 是否为最大数量(变量集中所包括的所有集的数量) $i_{\max}$ ($i=i_{\max}$)。

[0109] 如果在步骤 S37 中确定 $i=i_{\max}$ 不为真,则处理返回到步骤 S21,在步骤 S21 中,条件变量集(条件变量)的数量 i 递增 1。然后,执行后续处理。

[0110] 例如,在 $i=2$ 为真的情况下,对两个条件变量中的每个执行与 $i=1$ 为真的情况类似的处理,并且仅在两个条件变量均满足与具有一个条件变量的情况相同的条件的情况下,执行条件独立性测试。即,在条件变量中的任一个设置在具有 V 形结构的路径中的任一个上的情况下,不执行条件独立性测试。因此,可避免执行不需要的条件独立性测试,以降低错误出现的次数。

[0111] 然后,如果在步骤 S37 中确定 $i=i_{\max}$ 为真,则处理结束。当所有可执行的测试以这种方式结束时,具有这样的两个变量的变量对保留在变量对存储部分 81 中,这两个变量在测试中已被连续地确定为不独立,或这两个变量在不进行测试的情况下呈现为不独立。V 形结构存储部分 82 存储用于形成 V 形结构的变量的集合。分离变量集存储部分 83 存储两个独立变量的条件变量集。

[0112] 存储在存储部分 53 中的内容在控制部分 52 的控制下输出到输出部分 54。具体地,输出包括有向边和无向边的组合的部分有向非循环图(贝叶斯网络),以允许对变量之间的

依赖关系或因果关系进行估计。

[0113] 根据上述处理,在条件变量仅设置在两个变量之间的路径中的具有 V 形结构的路径上的情况下,不执行两个变量之间的条件独立性的测试。

[0114] 例如,在仅找到图 8 中所示的路径 1 作为变量 X 和变量 Y 之间的路径的情况下(路径 1 具有 V 形结构并且条件变量 Z 设置在路径 1 上),不执行变量 X 和变量 Y 之间的条件独立性的测试。然而,应该注意,在找到不具有 V 形结构但是其上设置有条件变量 Z 的路径 2 作为变量 X 和变量 Y 之间的路径的情况下,执行条件独立性测试。

[0115] 此外,在仅找到图 9 中所示的路径作为变量 X 和变量 Y 之间的路径的情况下(路径具有 V 形结构,并且条件变量 Z 是不与 V 形结构的两个独立变量中的每个相独立的变量),不执行变量 X 和变量 Y 之间的条件独立性的测试。

[0116] 因此,根据上述处理,可降低条件独立性测试的执行次数,这可降低测试错误的发生频率,因此增强变量之间的估计因果关系的可靠性。

[0117] 以上描述了用于执行具有离散值的变量之间的独立性测试的处理。然而,在变量具有连续值的情况下,可使用用作条件独立性索引的部分相关系数和 Fisher 的 Z 变换来执行独立性测试。替代地,可使用内核方法来执行独立性测试,内核方法是用于统计模式识别的技术。

[0118] 在以上讨论的独立性测试处理中,如果条件变量集的数量变得更大,则条件独立性测试的可靠性不可避免地降低。因此,V 形结构搜索的精度也可能劣化,这可能导致变量之间的估计因果关系的可靠性的降低。

[0119] 因此,在条件变量集(条件变量)的数量 i 递增到超过预定数量的情况下,即使在步骤 S35 中确定存在两个条件地独立的变量,也可跳过步骤 S36 中的处理(搜索 V 形结构)。

[0120] 因此,不执行由于条件独立性测试的可靠性降低而导致的精度差的 V 形结构搜索,从而避免变量之间的估计因果关系的可靠性的降低。在完成所有条件独立性测试之后,执行已跳过的 V 形结构搜索。

[0121] 在以上讨论的独立性测试处理中,V 形结构存储部分 82 存储所找到的 V 形结构。然而,因为独立地对每对两个变量执行独立性测试处理,所以在发生测试错误的情况下,可能存储具有相互矛盾关系的 V 形结构。

[0122] [矛盾解决处理]

[0123] 将参照图 10 的流程图来描述用于解决 V 形结构之间的矛盾的矛盾解决处理。

[0124] 在步骤 S51 中,矛盾检测部分 75 从存储在 V 形结构存储部分 82 中的 V 形结构中检测具有矛盾的 V 形结构。

[0125] 在步骤 S52 中,矛盾检测部分 75 根据预定标准来解决检测到的 V 形结构之间的矛盾。

[0126] [矛盾解决处理的第一具体示例]

[0127] 图 11 图示具有相互矛盾关系的 V 形结构的矛盾解决处理的具体示例。

[0128] 在图 11 的上部,示出了其中独立的变量 X 和变量 Y 各自不与变量 Z 独立的 V 形结构($X \rightarrow Z \leftarrow Y$)、和其中独立的变量 W 和变量 Z 各自不与变量 X 独立的 V 形结构($W \rightarrow X \leftarrow Z$)。

[0129] 在这种情况下,在关系 $X \rightarrow Z$ 和关系 $X \leftarrow Z$ 之间存在矛盾。

[0130] 为了解决这样的矛盾,矛盾检测部分 75 使用在每个 V 形结构的两个独立变量的独

立性测试中使用的条件独立性的索引(具体为 p 值或条件互信息量 MI),以将具有两个条件独立性较弱变量的 V 形结构的有向边改变为无向边。

[0131] 具体地,比较变量 X 和变量 Y 、以及变量 W 和变量 Z 的相应 p 值,以确定哪对两个变量条件独立性较弱。例如,在变量 X 和变量 Y 的 p 值是 0.50 且变量 W 和变量 Z 的 p 值是 0.20 的情况下,接受具有变量 X 和变量 Y 的 V 形结构 $X \rightarrow Z \leftarrow Y$,以及将表示变量 Z 对应于因并且变量 X 对应于果的因果关系的有向边(例如,在下文中被表示为“ $Z \rightarrow X$ 关系”)改变为如图 11 的下部所示的无向边。此外,从 V 形结构存储部分 82 中删除 V 形结构 $W \rightarrow X \leftarrow Z$ 。

[0132] [矛盾解决处理的第二具体示例]

[0133] 图 12 图示具有相互矛盾关系的 V 形结构的矛盾解决处理的另一具体示例。

[0134] 在图 12 的上部,示出了其中独立的变量 X 和变量 Y 各自不与变量 Z 独立的 V 形结构($X \rightarrow Z \leftarrow Y$)、和其中独立的变量 W 和变量 Y 各自不与变量 Z 独立的 V 形结构($W \rightarrow Z \leftarrow Y$)。在图 12 中,变量 X 和变量 W 不独立。

[0135] 在这种情况下,存在这样的矛盾:需要的是,变量 X 和变量 W 应该与除了用作条件变量的变量 Z 之外的变量内在地条件独立。

[0136] 为了解决这样的矛盾,矛盾检测部分 75 将表示 V 形结构 $X \rightarrow Z \leftarrow Y$ 中的 $X \rightarrow Z$ 关系和 V 形结构 $W \rightarrow Z \leftarrow Y$ 中的 $W \rightarrow Z$ 关系的有向边改变为如图 12 的下部所示的无向边,并从 V 形结构存储部分 82 中删除 V 形结构 $X \rightarrow Z \leftarrow Y$ 和 $W \rightarrow Z \leftarrow Y$ 。

[0137] [矛盾解决处理的第三具体示例]

[0138] 图 13 图示具有相互矛盾关系的 V 形结构的矛盾解决处理的又一具体示例。

[0139] 在图 13 的上部,如图 12 中一样,示出了其中独立的变量 X 和变量 Y 各自不与变量 Z 独立的 V 形结构($X \rightarrow Z \leftarrow Y$)、和其中独立的变量 W 和变量 Y 各自不与变量 Z 独立的 V 形结构($W \rightarrow Z \leftarrow Y$)。同样在图 13 中,变量 X 和变量 W 不独立。

[0140] 同样在这种情况下,存在这样的矛盾:需要的是,变量 X 和变量 W 应该与除了用作条件变量的变量 Z 之外的变量内在地条件独立。

[0141] 为了解决这样的矛盾,矛盾检测部分 75 将 V 形结构 $X \rightarrow Z \leftarrow Y$ 和 V 形结构 $W \rightarrow Z \leftarrow Y$ 中的全部有向边改变为如图 13 的下部所示的无向边,并从 V 形结构存储部分 82 中删除 V 形结构 $X \rightarrow Z \leftarrow Y$ 和 $W \rightarrow Z \leftarrow Y$ 。

[0142] [矛盾解决处理的第四具体示例]

[0143] 图 14 图示具有相互矛盾关系的 V 形结构的矛盾解决处理的又一具体示例。

[0144] 在图 14 的上部,如图 12 中一样,示出了其中独立的变量 X 和变量 Y 各自不与变量 Z 独立的 V 形结构($X \rightarrow Z \leftarrow Y$)、和其中独立的变量 W 和变量 Y 各自不与变量 Z 独立的 V 形结构($W \rightarrow Z \leftarrow Y$)。同样在图 14 中,变量 X 和变量 W 不独立。

[0145] 同样在这种情况下,存在这样的矛盾:需要的是,变量 X 和变量 W 应该与除了用作条件变量的变量 Z 之外的变量内在地条件独立。

[0146] 为了解决这样的矛盾,矛盾检测部分 75 使用在每个 V 形结构的两个独立变量的独立性测试中使用的条件独立性的索引(具体为 p 值或条件互信息量 MI),以将具有未被其他 V 形结构共享的两个条件独立性较弱变量的 V 形结构的有向边改变为无向边。

[0147] 具体地,比较变量 X 和变量 Y 、以及变量 W 和变量 Y 的相应 p 值,以确定哪对两个变量条件独立性较弱。例如,在变量 X 和变量 Y 的 p 值是 0.50 且变量 W 和变量 Y 的 p 值

是 0.30 的情况下,接受具有变量 X 和变量 Y 的 V 形结构 $X \rightarrow Z \leftarrow Y$, 以及将表示 V 形结构 $W \rightarrow Z \leftarrow Y$ 中未被 V 形结构 $X \rightarrow Z \leftarrow Y$ 共享的 $W \rightarrow Z$ 关系的有向边改变为如图 14 的下部所示的无向边。此外,从 V 形结构存储部分 82 中删除 V 形结构 $W \rightarrow Z \leftarrow Y$ 。

[0148] 根据上述处理,检测具有矛盾的 V 形结构以解决 V 形结构之间的矛盾。因此,与独立性测试处理并发地执行矛盾解决处理可降低由于其中使用错误的 V 形结构的条件独立性测试而导致的错误的发生概率,因此增强变量之间的估计因果关系的可靠性。

[0149] [本技术的应用示例]

[0150] 本技术可应用于以下示例。

[0151] (1) 统计因果关系估计设备

[0152] 根据本技术的实施例的独立性测试处理可应用于用于根据关于多元随机变量的观察数据来估计统计因果关系的因果关系估计处理。由用户来定义可具有离散值或连续值的多元随机变量,并且已预先准备了数据集。因此,向用作输出部分 54 的监视器等输出包括有向边和无向边的组合的图形模型作为图像。在这种情况下,可根据用于考虑条件独立性来确定有向边的取向的取向规则,尽可能地用有向边替换图形模型中的无向边。替代地,可将图形模型作为表示等同关系的文本数据输出到监视器等。

[0153] (2) 医疗诊断辅助设备

[0154] 根据本技术的实施例的独立性测试处理可应用于用于估计疾病和症状之间的因果关系的因果关系估计处理。已将 { 肺结核 T, 肺癌 L, 支气管炎 B, 胸部 X 射线检查结果 X, 呼吸困难 D, 和吸烟状态 S } 定义为随机变量,并且已将每个随机变量的内部状态的数量准备为数据。因此,揭示具有 V 形结构和无向边的变量对,并且向用作输出部分 54 的监视器等输出包括有向边和无向边的组合的图形模型作为图像。此外,可根据用于考虑条件独立性来确定有向边的取向的取向规则,尽可能地用有向边替换图形模型中的无向边。在示例中,例如输出 {S-L}、{S-B}、{L → D}、{B → D}、{L → X}、{T → X} 等作为变量之间的因果关系。

[0155] (3) 产品管理设备

[0156] 根据本技术的实施例的独立性测试处理可应用于用于估计工厂中的各种测量项和产量之间的因果关系的因果关系估计处理。已将 { 是否为好产品 Y, 工厂温度 T, 工厂湿度 M, 测量项 1A, 测量项 2B, 和测量项 3C } 定义为随机变量。这里,变量 Y 具有用于指示产品是否好的两个离散值,并且所有其他变量具有连续值。因此,使用根据现有技术的使用具有离散值的变量 Y 作为参考变量来离散化所有其他变量(连续值)的技术来离散化所有变量。因此,向用作输出部分 54 的监视器等输出包括有向边和无向边的组合的图形模型作为图像。

[0157] (4) 文档分类设备

[0158] 根据本技术的实施例的独立性测试处理可应用于用于将输入文档分类为用户预先定义的类别的文档分类设备中执行的文档分类的学习处理。当输入预先分类的或附有分类标签的文档时,文档分类设备通过词法分析(morphological analysis)等将文档分解为词类(parts of speech)。通过以上处理,已准备了表示类别的变量 C 和表示词类出现频率的变量 S。这里,变量 C 具有离散值,而变量 S 具有连续值。因此,参考变量 C 将变量 S 离散化(归类)。然后,通过对具有离散值的变量 S 执行根据本技术的实施例的独立性测试处理

来获得包括有向边和无向边的组合的图形模型。在这种情况下,可使用信息标准用有向边替换有向边。另外,可通过用有向边来链接表示类别的变量 C 和所有变量 S,将条件概率分布的乘积表示为贝叶斯网络。因此,通过使用相同的数据估计条件概率分布来执行文档分类的学习处理。通过以这种方式估计(学习)词类之间的依赖关系,生成以高准确度对未知文档进行分类的分类器。

[0159] 当新的未分类文档输入到文档分类设备时,文档被分解为词类,并且关于词类出现频率的信息在分类器中被映射在变量 S 的频率信息类别中。因此,为输入文档提供有作为证据的用于表示每个词类出现概率的变量 S。然后,通过将变量 S 边缘化来计算未提供有证据的类别的概率值,并将输入文档分类为已计算最大概率值的类别。

[0160] (5) 图像识别设备

[0161] 根据本技术的实施例的独立性测试处理可应用于用于将输入图像分类为用户预先定义的类别的图像识别设备中执行的图像分类的学习处理。当输入预先分类的或附有分类标签的多个图像(用于学习的图像)时,图像识别设备对图像进行聚类。另外,图像识别设备使用每个聚类的图像的 RGB 值的均值,以无监督的方式进一步对图像进行聚类。通过以上处理,已准备了表示类别的变量 C 和表示 RGB 值的均值的变量 V。这里,变量 C 具有离散值,而变量 V 具有连续值。因此,参考变量 C 将变量 V 离散化(归类)。然后,通过对具有离散值的变量 V 执行根据本技术的实施例的独立性测试处理来获得包括有向边和无向边的组合的图形模型。在这种情况下,可使用信息标准用有向边替换无向边。另外,可通过用有向边来链接表示类别的变量 C 和所有变量 V,将条件概率分布的乘积表示为贝叶斯网络。因此,通过使用相同的数据估计条件概率分布来执行图像分类的学习处理。通过以这种方式估计(学习)RGB 值之间的依赖关系,生成以高准确度对未知图像进行分类的分类器。

[0162] 当新的未分类图像输入到图像识别设备时,计算 RGB 值的平均值,并且图像的 RGB 值的均值在分类器中被映射在变量 V 的 RGB 值类别中。因此,为输入的图像提供有作为证据的变量 V。然后,通过将变量 V 边缘化来计算未提供有证据的类别的概率值,并将输入图像分类为已计算最大概率值的类别。

[0163] 除了以上讨论的示例之外,本技术也可应用于用于辅助用户进行选择的决策支持设备、用于执行诸如遗传分析和心理学分析的社会科学分析的分析设备、用于分析从传感器输入的信息的分析设备、用于根据用户的品味等进行推荐的推荐设备、其他代理系统等。

[0164] 本技术不限于上述实施例,并且在不脱离本技术的范围和精神的情况下,可以以各种方式进行修改。

[0165] 例如,可通过云计算来实现本技术,在云计算中,经由网络由多个设备以分布式的方式联合地处理单个功能。

[0166] 此外,在以上讨论的流程图中描述的每个步骤除了由单个设备执行之外,也可以由多个设备以分布式的方式执行。

[0167] 另外,在单个步骤中包括多个处理的情况下,单个步骤中包括的多个处理除了由单个设备执行之外,也可以由多个设备以分布式的方式执行。

[0168] 本技术可如下配置。

[0169] (1) 一种测试多个变量之间的独立性的信息处理装置,包括:执行部分,在用作图形模型中至少不独立的两个变量之间的独立性的条件的条件变量设置在所述两个变量之

间的路径上的情况下,执行所述两个变量之间的条件独立性的测试;以及确定部分,确定 V 形结构是否存在于所述两个变量之间的路径上,所述 V 形结构是独立的第一变量和第二变量各自不独立于第三变量的图结构,其中,在所述条件变量仅设置在被确定为具有所述 V 形结构的路径上的情况下,所述执行部分不执行所述两个变量之间的条件独立性的测试。

[0170] (2) 根据(1)的信息处理装置,其中,在所述条件变量是所述 V 形结构中的所述第三变量的情况下,所述执行部分不执行所述两个变量之间的条件独立性的测试。

[0171] (3) 根据(1)或(2)的信息处理装置,其中,所述执行部分在执行条件独立性的测试之前执行所述第一变量和所述第二变量之间的无条件独立性的测试,以及所述信息处理装置还包括:V 形结构搜索部分,搜索所述第一变量和所述第二变量无条件独立并且所述第一变量和所述第二变量各自不独立于所述第三变量的图结构,作为所述 V 形结构。

[0172] (4) 根据(3)的信息处理装置,其中,所述 V 形结构搜索部分搜索所述第一变量和所述第二变量在除所述第三变量之外的变量的条件下条件独立并且所述第一变量和所述第二变量各自不独立于所述第三变量的图结构,作为所述 V 形结构。

[0173] (5) 根据(4)的信息处理装置,其中,在条件变量集的数量递增到超过预定数量的情况下,所述 V 形结构搜索部分不搜索所述 V 形结构。

[0174] (6) 根据(1)至(5)中的任一项的信息处理装置,其中,所述确定部分针对所述两个变量之间的、数量与递增的条件变量集的数量相对应的路径,确定是否存在所述 V 形结构,以及在所述条件变量中的任一个设置在被确定为具有所述 V 形结构的路径中的任一个上的情况下,所述执行部分不执行所述两个变量之间的条件独立性的测试。

[0175] (7) 根据(1)至(6)中的任一项的信息处理装置,还包括:矛盾检测部分,检测多个所述 V 形结构中的有向边之间的矛盾,以将一些或全部所述有向边改变为无向边来解决所述矛盾。

[0176] (8) 根据(7)的信息处理装置,其中,所述矛盾检测部分使用在多个所述 V 形结构中的所述第一变量和所述第二变量之间的独立性的测试中使用的索引,将具有独立性较弱的所述第一变量和所述第二变量的所述 V 形结构的有向边改变为无向边。

[0177] (9) 根据(7)的信息处理装置,其中,在变量 X 和变量 Y 各自不独立于变量 Z 的第一 V 形结构、以及变量 W 和变量 Y 各自不独立于变量 Z 的第二 V 形结构中所述变量 X 和所述变量 W 不独立的情况下,所述矛盾检测部分将所述第一 V 形结构中从所述变量 X 到所述变量 Z 的有向边和所述第二 V 形结构中从所述变量 W 到所述变量 Z 的有向边改变为无向边。

[0178] (10) 根据(7)的信息处理装置,其中,在变量 X 和变量 Y 各自不独立于变量 Z 的第一 V 形结构、以及变量 W 和变量 Y 各自不独立于变量 Z 的第二 V 形结构中所述变量 X 和所述变量 W 不独立的情况下,所述矛盾检测部分将所述第一 V 形结构和所述第二 V 形结构中的全部有向边改变为无向边。

[0179] (11) 根据(7)的信息处理装置,其中,在变量 X 和变量 Y 各自不独立于变量 Z 的第一 V 形结构、以及变量 W 和变量 Y 各自不独立于变量 Z 的第二 V 形结构中所述变量 X 和所述变量 W 不独立的情况下,所述矛盾检测部分使用两个变量之间的独立性的测试中使用的索引,比较所述变量 X 和所述变量 Y 之间的独立性以及所述变量 W 和所述变量 Y 之间的独立性,以将具有独立性较弱的两个变量的所述 V 形结构的有向边改变为无向边。

[0180] (12) 一种用于测试多个变量之间的独立性的信息处理装置的信息处理方法,包

括：在用作图形模型中至少不独立的两个变量之间的独立性的条件的条件变量设置在所述两个变量之间的路径上的情况下，执行所述两个变量之间的条件独立性的测试；以及确定 V 形结构是否存在于所述两个变量之间的路径上，所述 V 形结构是独立的第一变量和第二变量各自不独立于第三变量的图结构，其中，在所述条件变量仅设置在被确定为具有所述 V 形结构的路径上的情况下，不执行所述两个变量之间的条件独立性的测试。

[0181] (13) 一种使计算机执行用于测试多个变量之间的独立性的处理的程序，所述处理包括：在用作图形模型中至少不独立的两个变量之间的独立性的条件的条件变量设置在所述两个变量之间的路径上的情况下，执行所述两个变量之间的条件独立性的测试；以及确定 V 形结构是否存在于所述两个变量之间的路径上，所述 V 形结构是独立的第一变量和第二变量各自不独立于第三变量的图结构，其中，在所述条件变量仅设置在被确定为具有所述 V 形结构的路径上的情况下，不执行所述两个变量之间的条件独立性的测试。

[0182] 本公开包含与 2012 年 3 月 28 日在日本专利局提交的日本优先权专利申请 JP2012-072789 中公开的主题有关的主题，其全部内容通过引用合并于此。

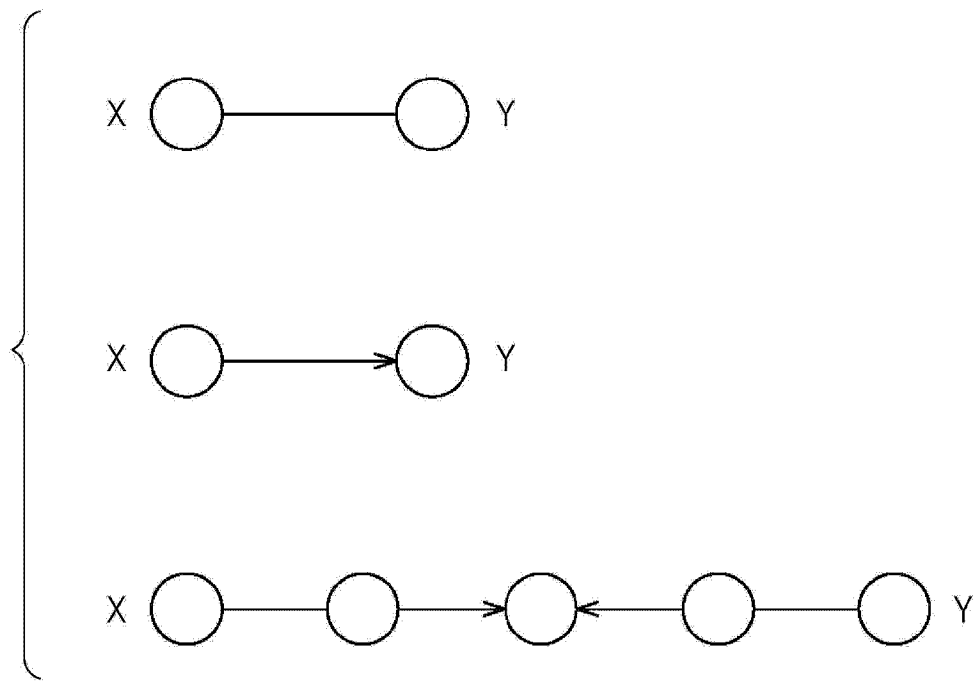


图 1

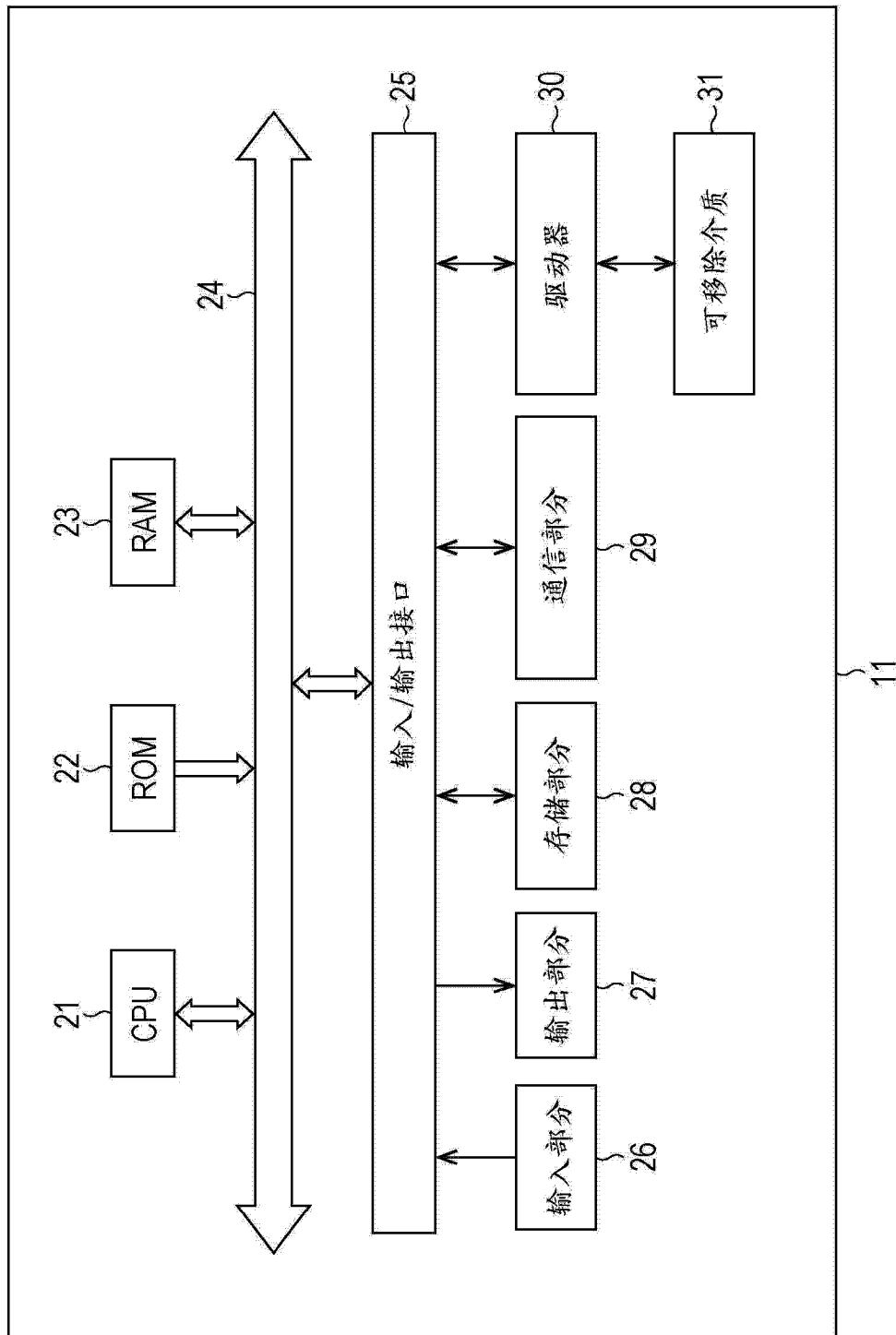


图 2

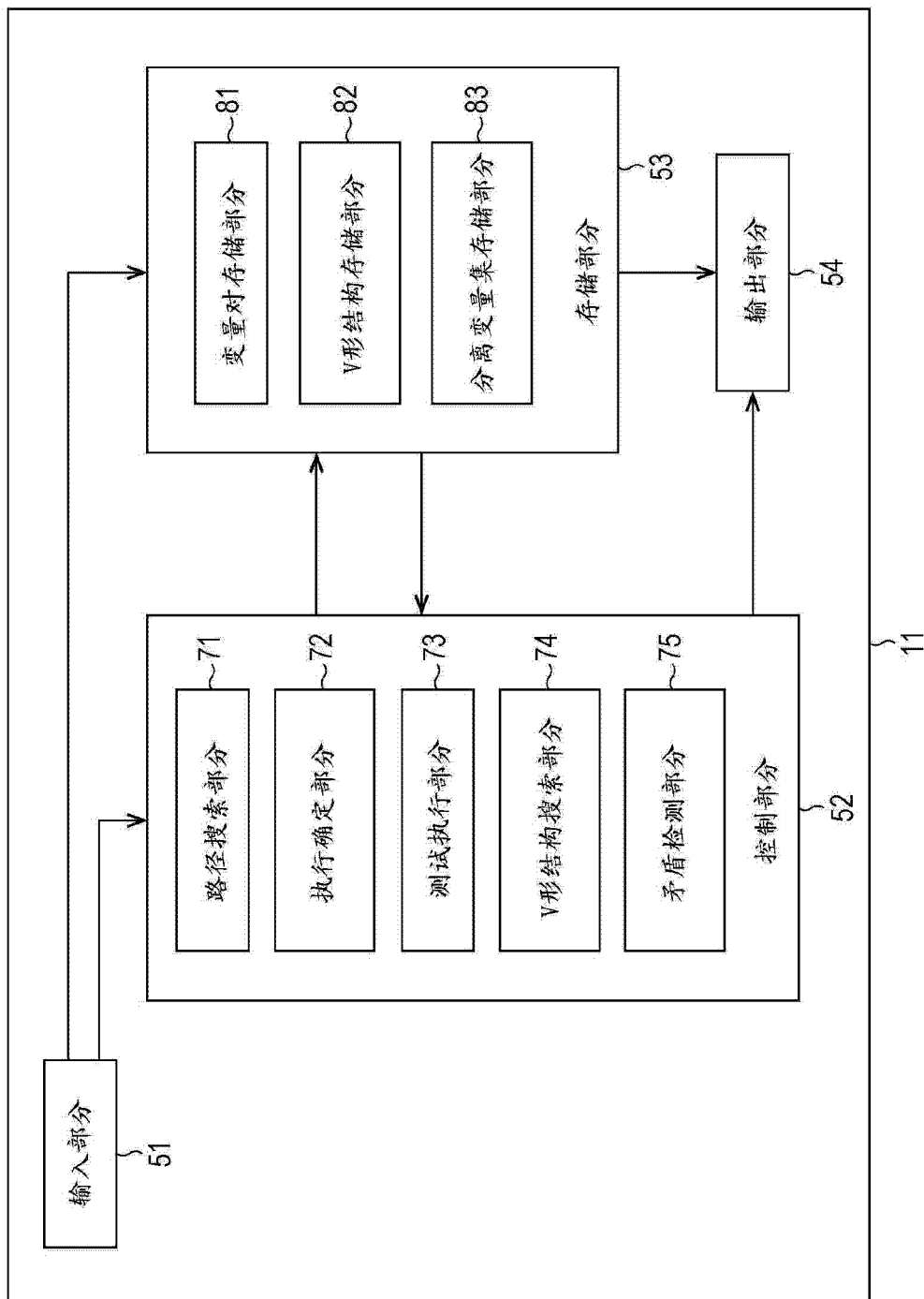


图 3

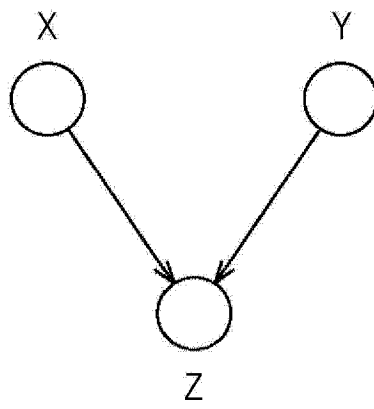


图 4

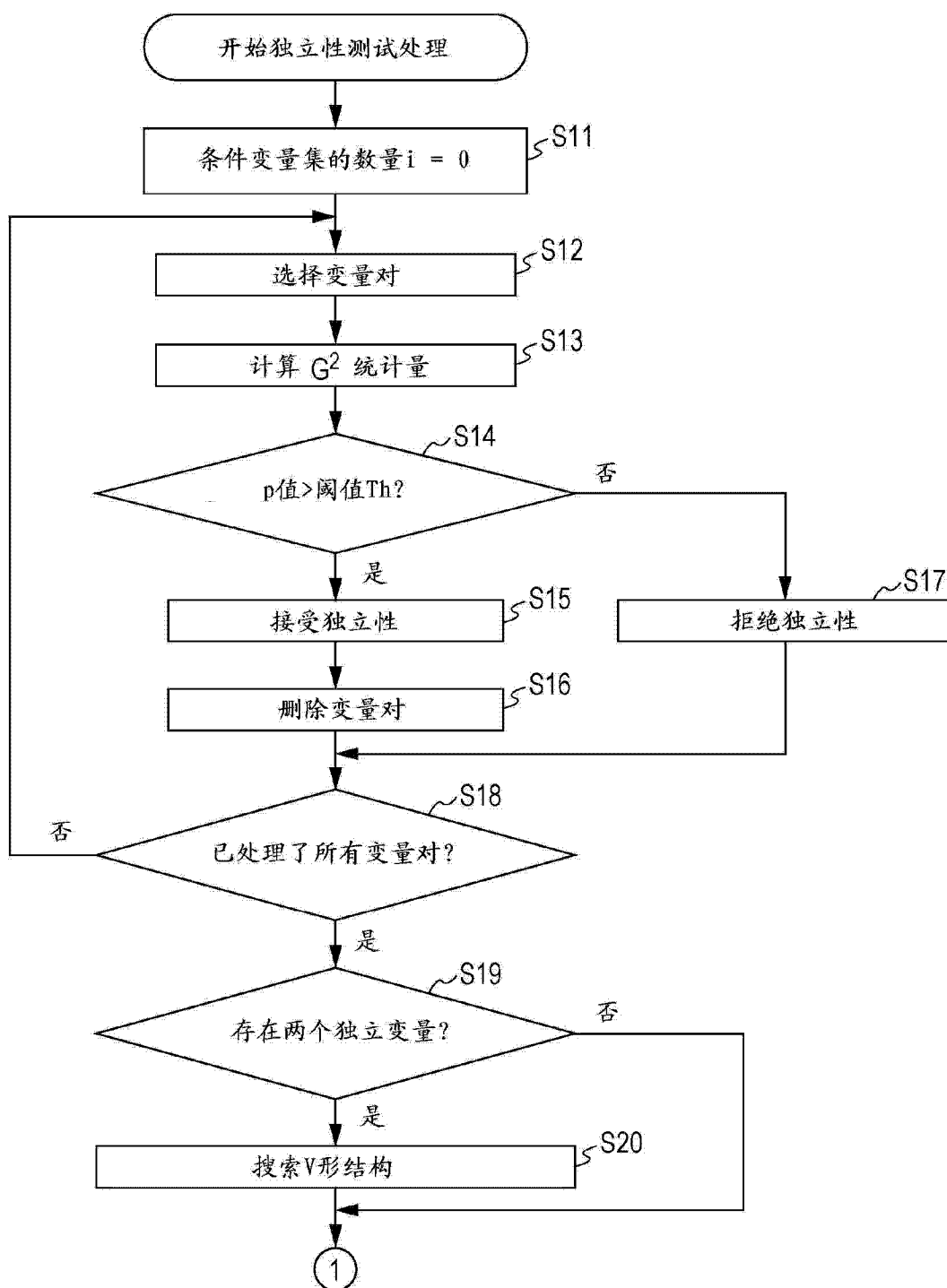


图 5

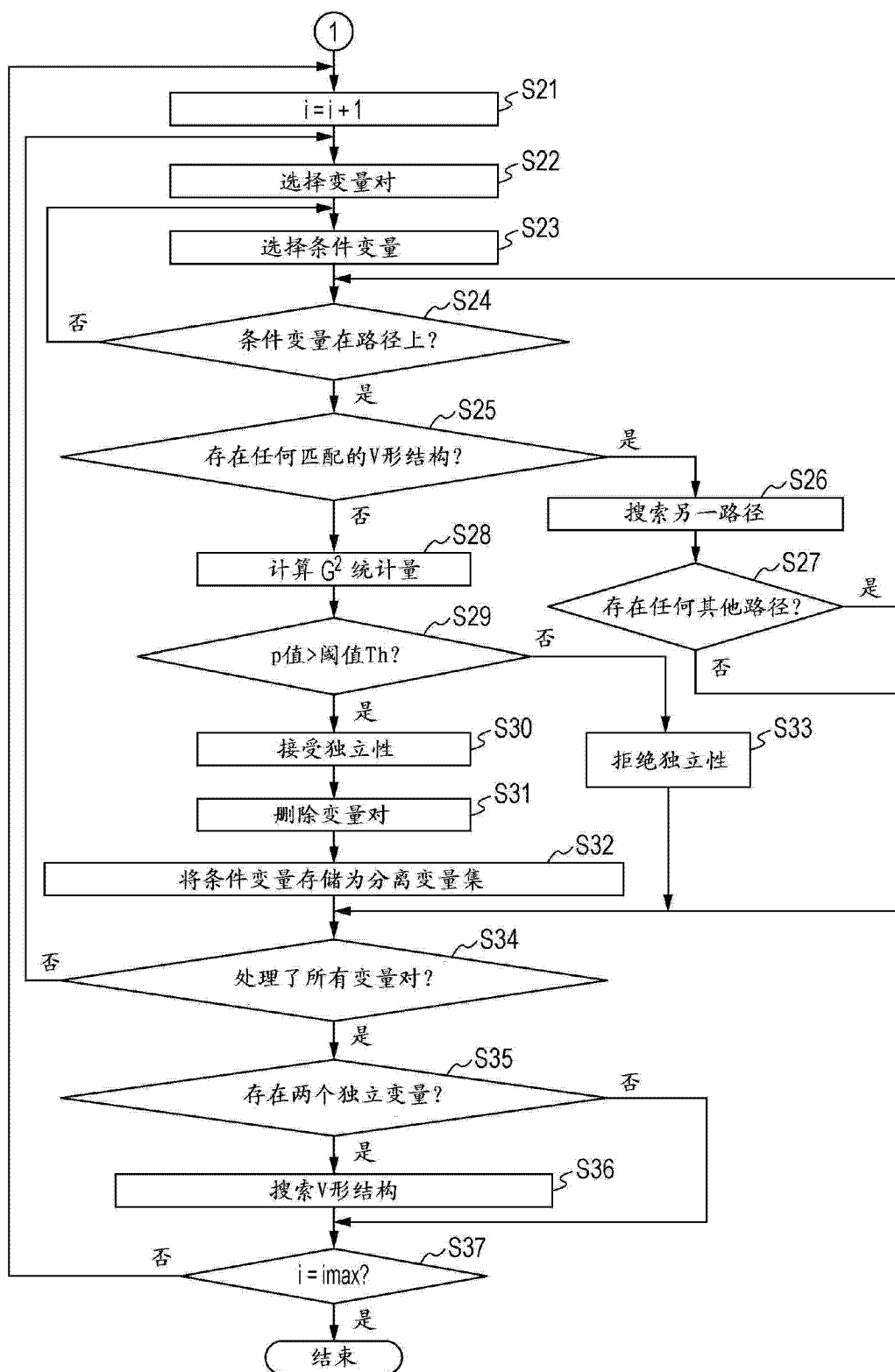


图 6

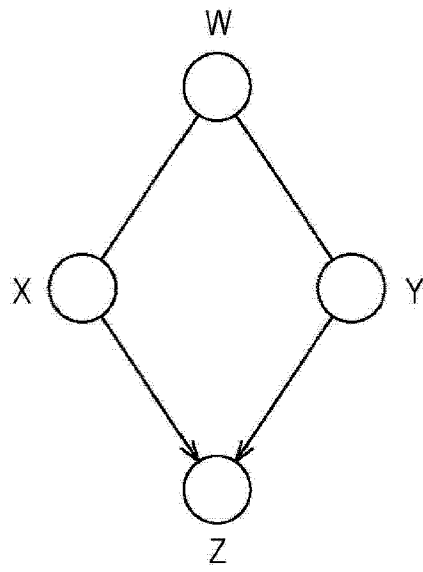


图 7

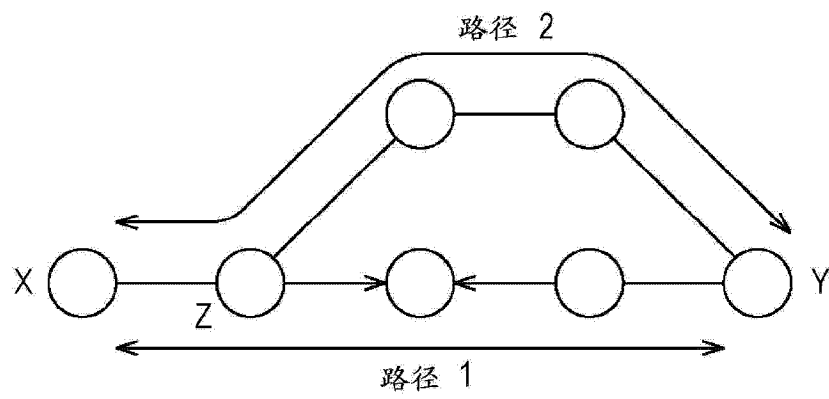


图 8

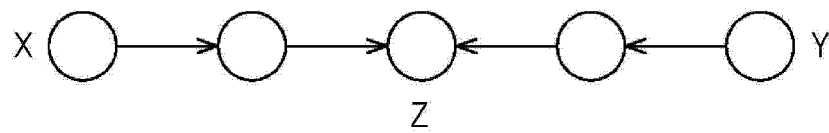


图 9

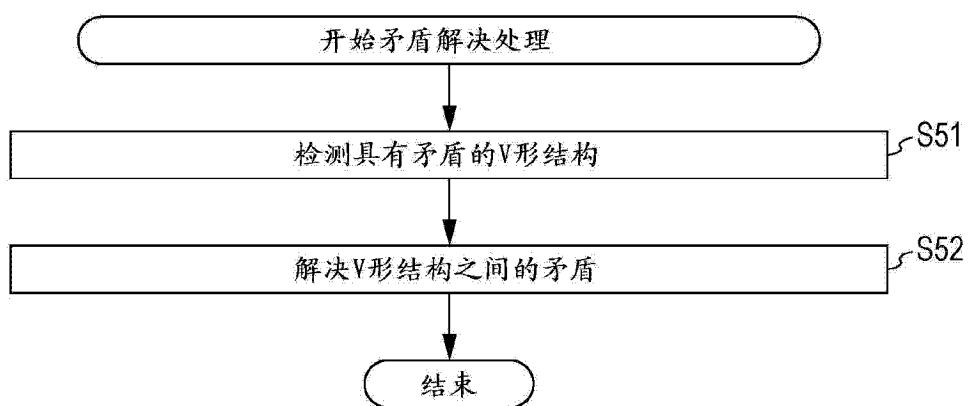


图 10

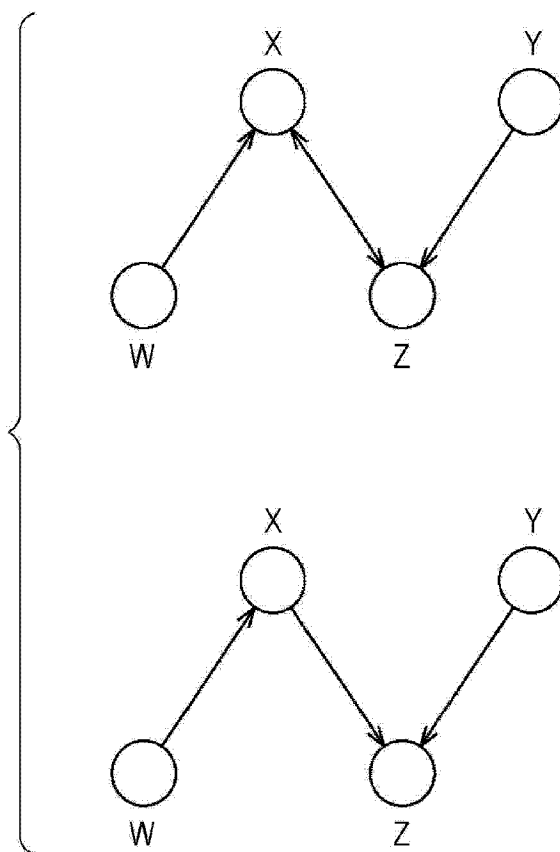


图 11

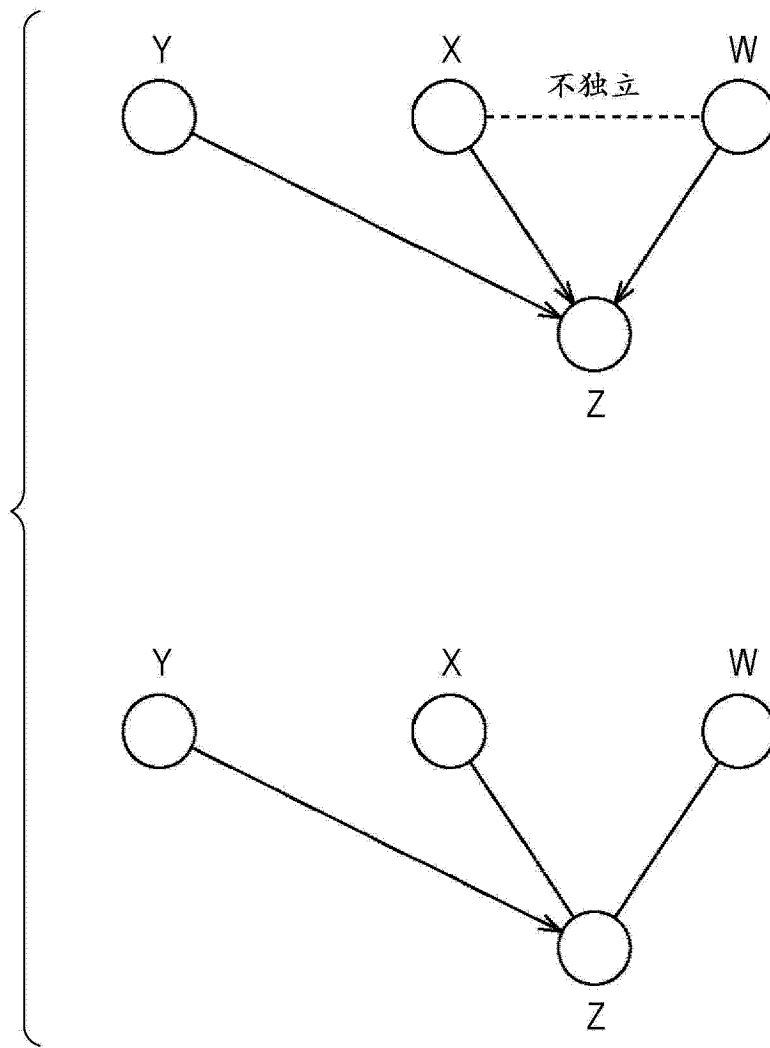


图 12

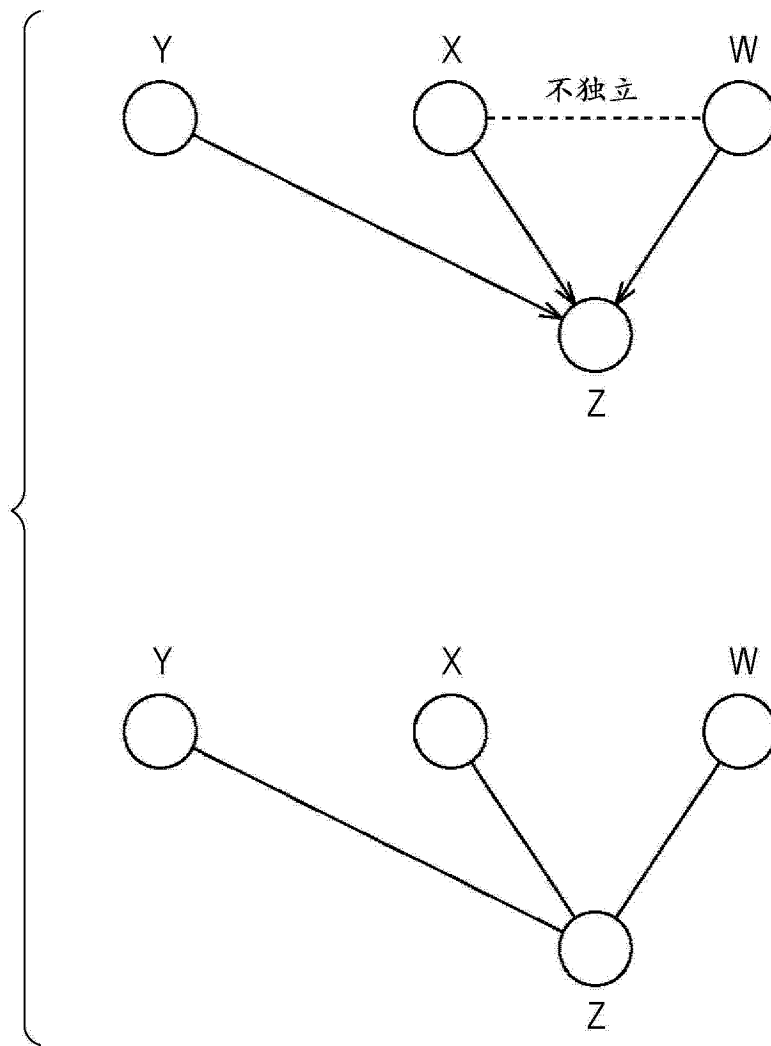


图 13

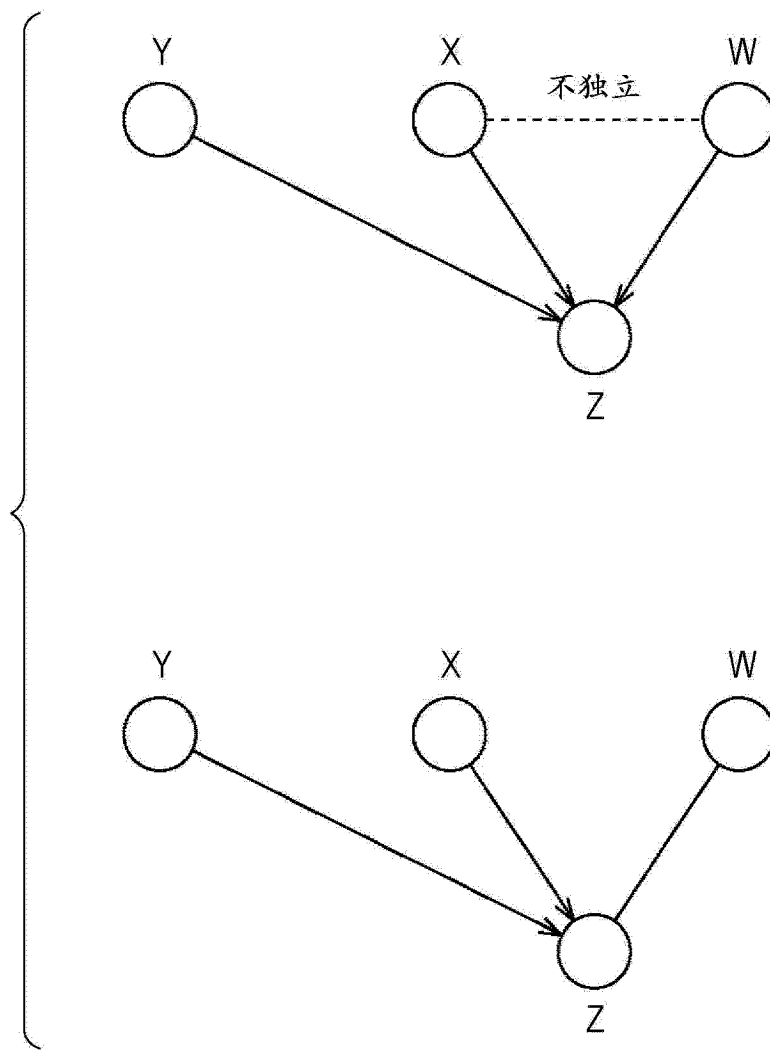


图 14