



(12) 发明专利

(10) 授权公告号 CN 110036383 B

(45) 授权公告日 2023.04.04

(21) 申请号 201780075220.2

谢国彤 俞益琴

(22) 申请日 2017.11.03

(74) 专利代理机构 中国贸促会专利商标事务所
有限公司 11038

(65) 同一申请的已公布的文献号
申请公布号 CN 110036383 A

专利代理师 郑宗玉

(43) 申请公布日 2019.07.19

(51) Int.Cl.

(30) 优先权数据

G06F 16/901 (2019.01)

15/373,551 2016.12.09 US

H04L 67/12 (2022.01)

(85) PCT国际申请进入国家阶段日
2019.06.05

(56) 对比文件

JP 2015197761 A, 2015.11.09

(86) PCT国际申请的申请数据

PCT/CN2017/109286 2017.11.03

US 2014358585 A1, 2014.12.04

(87) PCT国际申请的公布数据

W02018/103479 EN 2018.06.14

US 2016042124 A1, 2016.02.11

CN 104699939 A, 2015.06.10

(73) 专利权人 国际商业机器公司
地址 美国纽约

张丽丽.对电子医疗记录有效的连接关键词
的搜索.Journal of software.2016,第2016,27
(6)卷全文.

审查员 宋若昀

(72) 发明人 郝碧波 胡岗 李静 孙文

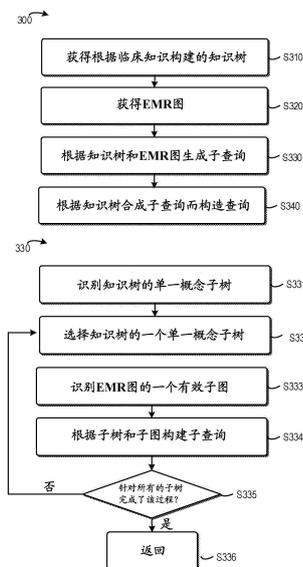
权利要求书4页 说明书13页 附图7页

(54) 发明名称

基于知识的电子病历特征自动提取

(57) 摘要

一种计算机实现的方法、装置和计算机程序产品,用于根据临床知识生成用于从电子病历(EMR)表集合提取临床特征的查询。按照一个临床知识数据集合构建知识树。获得对应于EMR表集合的EMR图。EMR图包括一个表节点集合和一个属性节点集合。表节点集合和属性节点集合表示EMR表集合中每个EMR表的结构和EMR表集合的属性之间的引用关系。基于知识树和EMR图生成多个子查询。通过按照知识树组合所述多个子查询来构造至少一个查询。



1. 一种计算机实现的方法,用于根据临床知识生成用于从电子病历 (EMR) 表集合提取临床特征的查询,所述计算机实现的方法包括:

确定要从EMR表集合中提取的至少一个临床特征;

获得至少一个临床特征的知识树,所述知识树是按照与至少一个临床特征相关联的一个临床知识数据集合进行构建的,其中知识树表示定义至少一个临床特征的一个或多个规则;

获得对应于EMR表集合的EMR图,其中EMR图包括一个表节点集合和一个属性节点集合,表节点集合和属性节点集合表示EMR表集合中每个EMR表的结构和EMR表集合的属性之间的引用关系;

基于知识树和EMR图生成多个子查询;和

通过按照知识树组合所述多个子查询来构造至少一个查询,用于从EMR表集合中提取至少一个临床特征。

2. 根据权利要求1所述的方法,其中,所述基于知识树和EMR图生成多个子查询包括:

识别知识树的多个单概念子树;和

对于所述多个单概念子树中的每个单概念子树:

选择知识树的一个单概念子树;

通过将该单概念子树的一个目标节点和一个概念节点分别映射到EMR图的一个表节点和至少一个属性节点来识别EMR图的一个有效子图;和

按照该单概念子树和该有效子图构建一个子查询。

3. 根据权利要求2所述的方法,其中知识树的单概念子树是包括从概念节点开始到目标节点的路径中的所有节点以及每个不包含概念节点的任何分支的子树。

4. 根据权利要求2所述的方法,其中,EMR图的有效子图是EMR图的这样一个子图,其中该单概念子树中有一个表节点与知识树的一个目标节点匹配,并且从该单概念子树中的每个属性节点到该表节点都有一个有向路径。

5. 根据权利要求2所述的方法,其中,所述将该单概念子树的一个目标节点和一个概念节点分别映射到EMR图的一个表节点和至少一个属性节点包括以下至少之一:

按属性映射;

按属性值映射;和

通过应用至少一个附加过滤器进行映射。

6. 根据权利要求1所述的方法,其中,所述获得对应于EMR表集合的EMR图包括将EMR表集合转换为等效图表示。

7. 根据权利要求1所述的方法,还包括:使用所述至少一个查询从EMR表集合中提取一个或多个临床特征。

8. 一种用于根据临床知识生成用于从电子病历 (EMR) 表集合提取临床特征的查询的装置,所述装置包括:

至少一个处理器;

存储器,可操作地耦合到处理器中的至少一个;和

存储在存储器中并由所述至少一个处理器以执行以下操作的一组计算机程序指令:

确定要从EMR表集合中提取的至少一个临床特征;

获得至少一个临床特征的知识树,所述知识树是按照与至少一个临床特征相关联的一个临床知识数据集进行构建的,其中知识树表示定义至少一个临床特征的一个或多个规则;

获得对应于EMR表集合的EMR图,其中EMR图包括一个表节点集合和一个属性节点集合,表节点集合和属性节点集合表示EMR表集合中每个EMR表的结构和EMR表集合的属性之间的引用关系;

基于知识树和EMR图生成多个子查询;和

通过按照知识树组合所述多个子查询来构造至少一个查询,用于从EMR表集合中提取至少一个临床特征。

9. 根据权利要求8所述的装置,其中,所述基于知识树和EMR图生成多个子查询包括:

识别知识树的多个单概念子树;和

对于所述多个单概念子树中的每个单概念子树:

选择知识树的一个单概念子树;

通过将该单概念子树的一个目标节点和一个概念节点分别映射到EMR图的一个表节点和至少一个属性节点来识别EMR图的一个有效子图;和

按照该单概念子树和该有效子图构建一个子查询。

10. 根据权利要求9所述的装置,其中知识树的单概念子树是包括从概念节点开始到目标节点的路径中的所有节点以及每个不包含概念节点的任何分支的子树。

11. 根据权利要求9所述的装置,其中,EMR图的有效子图是EMR图的这样一个子图,其中该单概念子树中有一个表节点与知识树的一个目标节点匹配,并且从该单概念子树中的每个属性节点到该表节点都有一个有向路径。

12. 根据权利要求9所述的装置,其中,所述将该单概念子树的一个目标节点和一个概念节点分别映射到EMR图的一个表节点和至少一个属性节点包括以下至少之一:

按属性映射;

按属性值映射;和

通过应用至少一个附加过滤器进行映射。

13. 根据权利要求8所述的装置,其中,所述获得对应于EMR表集合的EMR图包括将EMR表集合转换为等效图表示。

14. 根据权利要求8所述的装置,所述操作还包括:使用所述至少一个查询从EMR表集合中提取一个或多个临床特征。

15. 一种用于根据临床知识生成用于从电子病历(EMR)表集合提取临床特征的查询的计算机可读存储介质,所述计算机可读存储介质具有体现在其中的程序指令,所述程序指令可由至少一个处理器执行以使所述至少一个处理器执行以下方法:

确定要从EMR表集合中提取的至少一个临床特征;

获得至少一个临床特征的知识树,所述知识树是按照与至少一个临床特征相关联的一个临床知识数据集进行构建的,其中知识树表示定义至少一个临床特征的一个或多个规则;

获得对应于EMR表集合的EMR图,其中EMR图包括一个表节点集合和一个属性节点集合,表节点集合和属性节点集合表示EMR表集合中每个EMR表的结构和EMR表集合的属性之间的

引用关系；

基于知识树和EMR图生成多个子查询；和

通过按照知识树组合所述多个子查询来构造至少一个查询，用于从EMR表集合中提取至少一个临床特征。

16. 根据权利要求15所述的计算机可读存储介质，其中，所述基于知识树和EMR图生成多个子查询包括：

识别知识树的多个单概念子树；和

对于所述多个单概念子树中的每个单概念子树：

选择知识树的一个单概念子树；

通过将该单概念子树的一个目标节点和一个概念节点分别映射到EMR图的一个表节点和至少一个属性节点来识别EMR图的一个有效子图；和

按照该单概念子树和该有效子图构建一个子查询。

17. 根据权利要求16所述的计算机可读存储介质，其中知识树的单概念子树是包括从概念节点开始到目标节点的路径中的所有节点以及每个不包含概念节点的任何分支的子树。

18. 根据权利要求16所述的计算机可读存储介质，其中，EMR图的有效子图是EMR图的这样一个子图，其中该单概念子树中有一个表节点与知识树的一个目标节点匹配，并且从该单概念子树中的每个属性节点到该表节点都有一个有向路径。

19. 根据权利要求16所述的计算机可读存储介质，其中，所述将该单概念子树的一个目标节点和一个概念节点分别映射到EMR图的一个表节点和至少一个属性节点包括以下至少之一：

按属性映射；

按属性值映射；和

通过应用至少一个附加过滤器进行映射。

20. 根据权利要求15所述的计算机可读存储介质，其中，所述获得对应于EMR表集合的EMR图包括将EMR表集合转换为等效图表示。

21. 根据权利要求15所述的计算机可读存储介质，所述方法还包括：使用所述至少一个查询从EMR表集合中提取一个或多个临床特征。

22. 一种计算机实现的方法，用于根据临床知识生成用于从电子病历 (EMR) 表集合提取临床特征的查询，所述计算机实现的方法包括：

确定要从EMR表集合中提取的至少一个临床特征；

获得至少一个临床特征的知识树，所述知识树是按照与至少一个临床特征相关联的一个临床知识数据集合进行构建的，其中知识树表示定义至少一个临床特征的一个或多个规则；

获得对应于EMR表集合、将EMR表集合转换为等效的图表示的EMR图，其中EMR图包括一个表节点集合和一个属性节点集合，表节点集合和属性节点集合表示EMR表集合中每个EMR表的结构和EMR表集合的属性之间的引用关系；

基于识别知识树的多个单概念子树而基于知识树和EMR图生成多个子查询；

通过按照知识树组合所述多个子查询来构造至少一个查询，用于从EMR表集合中提取

至少一个临床特征;和

利用至少一个查询在EMR表集合中检索至少一个临床特征。

23. 根据权利要求22所述的方法,其中,所述基于知识树和EMR图生成多个子查询包括:

对于所述多个单概念子树中的每个单概念子树:

选择知识树的一个单概念子树;

通过将该单概念子树的一个目标节点和一个概念节点分别映射到EMR图的一个表节点和至少一个属性节点来识别EMR图的一个有效子图;和

按照该单概念子树和该有效子图构建一个子查询。

24. 一种用于根据临床知识生成用于从电子病历 (EMR) 表集合提取临床特征的查询的计算机可读存储介质,所述计算机可读存储介质具有体现在其中的程序指令,所述程序指令可由至少一个处理器执行以使所述至少一个处理器执行以下方法:

确定要从EMR表集合中提取的至少一个临床特征;

获得至少一个临床特征的知识树,所述知识树是按照与至少一个临床特征相关联的一个临床知识数据集合进行构建的,其中知识树表示定义至少一个临床特征的一个或多个规则;

获得对应于EMR表集合、将EMR表集合转换为等效的图表示的EMR图,其中EMR图包括一个表节点集合和一个属性节点集合,表节点集合和属性节点集合表示EMR表集合中每个EMR表的结构和EMR表集合的属性之间的引用关系;

基于识别知识树的多个单概念子树而基于知识树和EMR图生成多个子查询;和

通过按照知识树组合所述多个子查询来构造至少一个查询,用于从EMR表集合中提取至少一个临床特征;和

利用至少一个查询在EMR表集合中检索至少一个临床特征。

25. 根据权利要求24所述的计算机可读存储介质,其中,所述基于知识树和EMR图生成多个子查询包括:

对于所述多个单概念子树中的每个单概念子树:

选择知识树的一个单概念子树;

通过将该单概念子树的一个目标节点和一个概念节点分别映射到EMR图的一个表节点和至少一个属性节点来识别EMR图的一个有效子图;和

按照该单概念子树和该有效子图构建一个子查询。

基于知识的电子病历特征自动提取

技术领域

[0001] 本发明涉及数据库领域,更具体地说,涉及一种用于生成查询消息以查询电子病历以提取临床特征的方法、系统和计算机程序产品。

背景技术

[0002] 目前,电子病历(EMR)系统中的临床信息量在迅速增加。临床信息通常以关系数据集的形式存储。不同的医院或医疗机构可能把电子病历数据存储到采用不同模式的EMR数据集或EMR表。对电子病历数据的特征提取/导出有通用要求。临床数据集的描述性分析和预测分析、特征提取或特征构造,通常是劳动密集型的。因此,健康专家在利用电子病历数据提取或评价临床知识方面受到困扰。不同来源的电子病历数据的模式不同,要把它们改变成统一的模式,以便生成查询消息,用于在检索系统中查询EMR数据集的信息,这样做的劳动强度大,又不可靠。需要整合这些不同模式的EMR数据集,并使用它们来检索临床特征。

发明内容

[0003] 针对EMR数据集模式差异的问题,一个或多个实施例提出了一种从任意关系模式的EMR数据集中自动提取临床知识和/或评估临床规则的技术。一个或多个实施例提出了一种生成用于从EMR数据集中提取临床特征的查询的方法、系统和计算机程序产品。

[0004] 在一个实施例中,提供一种用于基于临床知识生成用于从一个电子病历(EMR)表集合提取临床特征的查询的计算机实现的方法。该方法包括获取按照一个临床知识数据集构建的知识树。获得对应于EMR表集合的EMR图。EMR图包括一个表节点集合和一个属性节点集合。表节点集合和属性节点集合表示EMR表集合中每个EMR表的结构和EMR表的属性之间的引用关系。根据知识树和EMR图生成多个子查询。通过根据知识树合成多个子查询来生成至少一个查询。

[0005] 在另一实施例中,提供一种用于基于临床知识生成用于从一个电子病历(EMR)表集合提取临床特征的查询的计算机实现的装置。所述装置包括至少一个处理器和可操作地耦合到所述处理器中的至少一个的存储器。一组计算机程序指令存储在存储器中并由至少一个处理器执行以执行多个操作。该多个操作包括按照一个临床知识数据集构建的知识树。获得对应于EMR表集合的EMR图。EMR图包括一个表节点集合和一个属性节点集合。表节点集合和属性节点集合表示EMR表集合中每个EMR表的结构和EMR表的属性之间的引用关系。根据知识树和EMR图生成多个子查询。通过根据知识树合成多个子查询来生成至少一个查询。

[0006] 在又一个实施例中,提供一种用于基于临床知识生成用于从一个电子病历(EMR)表集合提取临床特征的查询的计算机程序产品。该计算机程序产品包含存储有程序指令的计算机存储介质,程序指令能被至少一个处理器执行以使该至少一个处理器执行一个方法。该方法按照一个临床知识数据集构建的知识树。获得对应于EMR表集合的EMR图。EMR图包括一个表节点集合和一个属性节点集合。表节点集合和属性节点集合表示EMR表集合

中每个EMR表的结构和EMR表的属性之间的引用关系。根据知识树和EMR图生成多个子查询。通过根据知识树合成多个子查询来生成至少一个查询。

[0007] 一个或多个实施例提供用于基于知识自动从任意关系模式的EMR数据集提取特征的新的解决方案。与临床数据分析领域的现有方法相比,实施例可以显著减少基于知识的特征工程所需的人力,因此非常适用于文献和异构EMR数据集中临床知识的数据挖掘。

附图说明

[0008] 图1表示按照本发明一个实施例的示例性计算机系统;

[0009] 图2示意性表示按照本发明一个实施例的表形式的EMR数据集的例子;

[0010] 图3是例示按照本发明一个实施例的生成对EMR数据集的查询的计算机实现的方法的流程图;

[0011] 图4A示意性表示按照本发明一个实施例的知识树;

[0012] 图4B表示按照本发明一个实施例的知识树的三个单一概念子树;

[0013] 图5示意性表示按照本发明一个实施例、从图2的EMR数据集转换的EMR图;

[0014] 图6A-6C示意性表示按照本发明一个实施例、通过映射知识树的元素到EMR图的元素而生成子查询和构造最终查询的过程。

具体实施方式

[0015] 本文按要求公开了详细的实施例;但是,应当理解,所公开的实施例只是示例,下文所述的系统和方法可以各种形式体现。因此,本文所公开的具体结构和功能细节不应被解释为限制性的,而仅仅是作为权利要求的基础,并且是教导本领域技术人员在几乎任何适当详细的结构和功能中以不同方式运用本发明主题的代表性基础。此外,本文中使用的术语和短语并非旨在限制,而是提供对概念的可理解的描述。

[0016] 现在参考图1,示出了适用于一个或多个实施例的计算机系统/服务器12的一个示例。计算机系统/服务器12只是说明性的,并非是要对本文讨论的各种实施例的使用范围或功能提出任何限制。至少在某些情况下,一个或多个实施例在计算机系统/服务器12内实现。图1显示了通用计算设备形式的计算机系统/服务器12。计算机系统/服务器12的组件可包括但不限于一个或多个处理器或处理单元16、系统存储器28和总线18,总线18将包括系统存储器28在内的各种系统组件与处理器16耦合。总线18表示几种总线结构中的一种或多种,包括存储器总线或存储器控制器、外围总线、加速图形端口以及使用各种总线结构的处理器或本地总线。举例来说,这些体系结构包括工业标准体系结构 (ISA) 总线、微通道体系结构 (MCA) 总线、增强型ISA (EISA) 总线、视频电子标准协会 (VESA) 本地总线和外围组件互连 (PCI) 总线。

[0017] 计算机系统/服务器12通常包括各种计算机系统可读介质。此类介质可以是计算机系统/服务器12可访问的任何可用介质,包括易失性介质和非易失性介质、可移动介质和不可移动介质。系统存储器28可以包括易失性存储器形式的计算机系统可读介质,例如随机存取存储器 (RAM) 30和/或高速缓冲存储器32。计算机系统/服务器12还可包括其他可移动/不可移动、易失性/非易失性计算机系统存储介质。仅举例来说,存储系统34可用于读取和写入不可移动、非易失性磁介质(未显示,通常称为“硬盘”)。尽管未显示,但可提供用于

从可移动非易失性磁盘(例如“软盘”)读取和写入的磁盘驱动器,以及用于从可移动非易失性磁盘(例如CD-ROM、DVD-ROM或其他光学介质)读取或写入的光盘驱动器。在这种情况下,每一个都可以通过一个或多个数据媒体接口连接到总线18。如下文将进一步描述和讨论的,存储器28可以包括至少一个程序产品,该程序产品具有一组(例如,至少一个)被配置为执行一个或多个实施例的功能的程序模块。

[0018] 具有一组(至少一个)程序模块42的程序/实用程序40,可以例如而不同于与操作系统、一个或多个应用程序、其他程序模块和程序数据一样存储在存储器28。操作系统、一个或多个应用程序、其他程序模块和程序数据或它们的某些组合中的每一个都可以包括网络环境的实现。程序模块42通常执行本文讨论的一个或多个实施例的功能和/或方法。

[0019] 计算机系统/服务器12还可以与诸如键盘、定点设备、显示器24等的一个或多个外部设备14通信;与使用户能够与计算机系统/服务器12交互的一个或多个设备通信;以及/或者与使计算机系统/服务器12能够与一个或多个多其他计算设备通信的任何设备(例如网卡、调制解调器等)通信。此类通信可通过输入/输出(I/O)接口22进行。此外,计算机系统/服务器12还可以通过网络适配器20与一个或多个网络—例如局域网(LAN)、通用广域网(WAN)和/或公用网络(例如Internet)—通信。如图所示,网络适配器20通过总线18与计算机系统/服务器12的其他组件通信。应理解,尽管未显示,其他硬件和/或软件组件可与计算机系统/服务器12一起使用。示例包括但不限于:微码、设备驱动程序、冗余处理单元、外部磁盘驱动器阵列、RAID系统、磁带驱动器和数据存档存储系统等。

[0020] 为了更好地理解本文讨论的各种实施例,最好介绍一些相关知识,提供本公开的上下文中电子病历和临床知识的介绍。在一个实施例中,基于临床知识自动构造对任意形式的电子病历的查询。然后,可将查询用于从电子病历中提取临床特征。

[0021] I. 电子病历

[0022] 电子病历(EMR)是患者和人口的医疗历史和/或健康信息的数字记录。在一个实施例中,电子病历(EMRs)存储在正式的EMR数据集中。EMR数据集的模式可能因不同的数据库系统和/或不同的医院/医疗机构而不同。图2显示一个EMR数据集的例子。该数据集由正式的表组成。为简单起见,只例示了四个表T1、T2、T3和T4。

[0023] 表T1叫作“患者表”,含有患者的基本数据。表T1的第一行含有属性(也称“属性名称”),第二行和其它行(未予显示)包含属性的值(也称“属性值”)。表T1中,属性“PID”指唯一地标识一个患者的患者标识符(ID)。“PID”附有符号“+”,表示“PID”作为表T1的主键。属性“gender”是指患者的性别,属性“age”指患者的年龄。第二行代表一个患者的数据记录。它包括分别对应三种属性“PID”、“gender”、“age”的属性值“P1”、“M”和“67”。值“P1”是一个患者的ID,“M”指该患者是男性,“67”指该患者的年龄为67岁。T1表可能包含不同患者的基本资料。图2只例示了两行三列,在实际实现中,列数和行数不限于此。

[0024] 表T2叫作“就诊表”,其存储患者在医院就诊的数据。表T2的第一行包含属性,其它行包含属性值。第一行中显示有四个属性。属性“EID”指唯一地标识患者的一次就诊的就诊标识符(ID)。“EID”附有符号“+”,表示“EID”作为表T2的主键。属性“diag”是指一个医院就诊的诊断结果。属性“dept”是医院的科室。“PID”与表T1中的意思相同,只是此处附有符号“*”,表示作为表T2的一个外键(foreign key)。用外键“PID”,可以引用表T1。第二行代表患者就诊的一个数据记录。第二行中的属性值“E1”、“DM”、“D1”和“P1”分别对应于四个属性”

EID”、“diag”、“dept”和“PID”。“E1”是就诊ID，“P1”是患者ID，“DM”是患者“P1”的诊断结果，这是“diabetes mellitus”（糖尿病）的缩写，“D1”是患者“P1”就诊所在的临床科室。表T2可包含不同患者的信息，并且可以包含特定患者的多个不同就诊记录行。图2仅例示了四列二行，在实际实现中，列数和行数不限于此。

[0025] 表T3叫作“化验表”，其存储患者一次或多次就诊中的化验数据。表T3的第一行包含属性，其它行包含属性值。第一行的四列中显示了四个属性。属性“LID”是唯一地标识化验的化验标识符（ID）。“LID”附有符号“+”，表示“LID”作为T3的主键。“EID”的意思与表T2中的相同，如符号“*”所示，其可作为表T3的外键。属性“date”是化验的日期。属性“tech”指负责化验的化验员。第二行代表化验的一个数据记录，包括分别对应于四个属性“LID”、“EID”、“date”和“tech”的值“L1”、“E1”、“15-10-11”和“Tech1”。“L1”是化验ID，“E1”是就诊ID，“15-10-11”是化验日期，“Tech1”是负责化验“L1”的化验员的ID。表T3可以含有不同患者的信息，并可能包含一次就诊期间可能进行几项化验的特定患者的多行记录。同样，表T3中的行和列的数量没有限制。

[0026] 表T4称为“化验指标表”，其存储化验的化验指标的数据。表T4的第一行包含属性，其它行包含属性值。第一行的三列中显示了三个属性。属性“LName”是唯一地标识化验指标的化验指标名称。“LValue”是化验指标的值。“LID”与表T3中的意思相同，如所附的符号“*”所示，其可以作为表T4的外键。第二行代表一个化验指标的数据记录。值“HbA1C”（糖化血红蛋白）、“10.5”、“L1”表示在化验“L1”中进行了对名称为“HbA1C”的化验指标的化验，“HbA1C”的结果值是“10.5”。表T4可以含有不同化验和不同患者的信息，并且可以包含单一化验的多行记录。同样，表T4可容纳的行和列的数量没有限制。

[0027] II. 临床知识

[0028] 临床知识是指关于疾病、机理和发病机制、治疗和相互作用以及对化验的解释的信息，其广泛适用于有关患者和公共卫生政策的决策。临床知识可以用一组临床规则的形式来表达。例如，下列规则是一组临床规则。

[0029] 规则R1：血糖控制达标：80%的血糖值达标（空腹血糖 $<7.5\text{mmol/L}$ 或2h血糖 $<10\text{mmol/L}$ ）在一个月；

[0030] 规则R2：血糖持续高：80%的血糖值高（空腹血糖 $\geq 9\text{mmol/L}$ 或2h血糖 $\geq 13\text{mmol/L}$ ）三个月；

[0031] 规则R3：低血糖：最近的新血糖低（血糖 $<3.9\text{mmol/L}$ ）；

[0032] 规则R4：高危患者：年龄大于65，有糖尿病，且过去的三个月的HbA1c（糖化血红蛋白） >9 。

[0033] 在实施例中，临床知识将被理解为如以上所示的正式临床规则。存在将临床知识转化为正式临床规则的现有技术。为了避免模糊对各实施例的描述，此处省略了关于临床规则生成的细节。

[0034] 给定一个EMR数据集，最好能根据已知的临床知识从EMR数据集中提取临床特征。例如，最好能根据例如临床规则R4从图2所示的表中找出高危患者，规则R4可被解释为：

[0035] “高危患者是那些年龄超过65岁、患糖尿病、三个月内化验的“HbA1c”（糖化血红蛋白）化验指标大于9的患者。

[0036] 首先构造一个查询，用于自动提取临床特征。现在将参照图3，描述本发明的构建

这样的查询的实施例。图3是例示按照本发明实施例的生成对一个EMR表集合的查询的计算机实现的方法300的流程图。总的来说,方法300包括以下步骤:

[0037] 获得根据临床知识构建的知识树,如S310所示;

[0038] 获得对应于该EMR表集合的EMR图,其中EMR图包括表节点和属性节点,用以表示每个EMR表的结构和EMR表的属性之间的引用关系,如S320所示;

[0039] 根据知识树和EMR图生成子查询,如S330所示;

[0040] 根据知识树合成子查询而构造查询,如S340所示。

[0041] 现在将详细描述上述各步骤。图4A例示按照本发明实施例的一个知识树400。在实施例中,用知识树来表达树结构的知识和/或规则。假设方法300涉及的临床知识是临床规则R4。根据临床规则R4,如果某患者年龄超过65岁,患有糖尿病,并且三个月内化验的“HbA1c”(糖化血红蛋白)化验指标大于9,则该患者会被归为“高危患者”的范畴。因此,获得根据临床知识构建的知识树的步骤S310,将获得如图4A所示的、对应于规则R4的知识树400。

[0042] 如图4A中所示、知识树400由不同形状的结点组成。五边形节点代表目标节点,例如节点t-1。椭圆形节点代表算术/逻辑运算符,例如节点op2、op3、op4、op8和op11。矩形节点代表临床概念,例如概念节点c5、c9和c13。棱形节点表示聚合(aggregation)算子,例如聚合节点ag7和ag12。平行四边形节点表示概念过滤器(filter),如过滤器节点f14。最后,三角形节点代表常数值,如值节点v6、v10和v15。

[0043] 如图4A中所示,目标节点t-1是知识树400的根。在本发明的上下文中,目标节点代表临床特征提取的目标。在这里它代表“patient”(患者)。一般来说,目标节点可被视为规则或知识树的主题。例如,“patient”是规则“patients with condition of ABC are high-risk patients”(具备条件ABC的患者,是高危患者)的目标,而要构建的特征将是针对每个“patient”的布尔型(boolean-typed)的特征,诸如“whether the patient is high risk”(该患者是否高危患者)。“Encounter”是规则“Encounters with condition of XYZ are outpatient followup encounters”(具备条件XYZ的就诊,是门诊跟踪就诊)的目标,而要构建的特征将是针对每个“encounter”的布尔型特征,诸如“whether the encounter is an outpatient followup encounter”(该就诊是否门诊跟踪就诊)。

[0044] 操作符节点op2和op3表示逻辑运算符“&&”,即“AND”(逻辑与)。操作符节点op4和op11代表逻辑运算符“>”或“大于”。操作符节点op8代表逻辑算子“=”或“等于”。概念节点c5、c9和c13分别代表临床概念“age”、“DM”、“HbA1c”。聚合节点ag7和ag12分别代表聚合算子“any”(任何)和“max”(最大)。值节点的v10代表常数值“Y”,本例中其意思为“是”或逻辑“真”。最后,值节点v15代表常数值“9”,本例中其为化验指标“HbA1c”的测量值。

[0045] 所属技术领域的熟练人员应当明白,聚合算子被用来对相同性质的元素的集合执行一个操作。除了上述的“max”和“any”,聚合算子的另一个例子是“average”(平均),其通常被用来计算数字型元素的集合的平均值。总的来说,聚合节点表示在特征构建中如何处理“一对多”或“多对一”的关系。例如,当目标是“patient”时,我们想构建特征“whether the patient has ever lived in Beijing”(患者是否曾在北京生活)。每一次就诊,可能都留有患者的地址记录。因此,通过使用聚合算子“any”(任何)来处理这种一对多的关系。同样,如果特征变为“the patient's maximum body weight in the past”(患者过去的最

大体重),则需要将聚合算子“max”应用到多个重量测量记录。通常情况下,当“目标”不是关系表中的唯一键时,或具有“目标”属性的表将与具有“一对多”或“多对一”的关系的另一个表连接(join)时,要运用聚合算子。

[0046] 过滤器节点f14代表概念过滤器“last 3months”(最后3个月)。这里所用的过滤器节点表示其父节点的关联属性要满足的附加条件。本例中,过滤器节点f14表示“last 3months”,所指定的限制是,其父节点C13需要某种日期类型的相关属性,且属性值应在最后三个月内。

[0047] 熟悉知识表达树的技术人员明白,知识树400包含由两个节点(op2和op3)组合的三个条件:(由节点c5、op4和v6表示的)“age>65”、(由节点c9、ag7、op8和v10表示的)“has diabetes”和(由节点c13、op11、ag12、v15和f14表示的)“HbA1c>9.0in last three months”(过去三个月糖化血红蛋白>9)。要注意的是,使用不同的形状是为了直观地区分不同类型的节点。所属技术领域的技术人员明白,可以有其它类似的或等同的方式来表示不同类型的节点。

[0048] 现有技术中存在可以构建知识树来表示各个领域的领域知识的技术。所属技术领域的技术人员基于上述说明,很容易知道如何以各种方式根据临床知识来构建知识树。本发明的实施例可以利用这些技术来构建代表临床知识的知识树,从而实现步骤S310,即获得根据临床知识构建的知识树。

[0049] 返回图3。描述了步骤S310后,现在将描述方法300的步骤320。步骤S320包括获得对应于该EMR表集合的EMR图,其中EMR图包括表节点和属性节点,用以表示每个EMR表的结构和EMR表的属性之间的引用关系。按照本发明的一个实施例,获得获得对应于该EMR表集合的EMR图包括将该EMR表集合转换为其等效图表示。

[0050] 图5示意性地表示按照本发明实施例从图2的EMR表集合转换的EMR图500。EMR图500包括四个表节点t1、t2、t3和t4以及双向连接到相关联的表节点的属性节点,它代表图2中的四个EMR表T1、T2、T3和T4的每个的结构以及这些表的属性之间的引用关系。

[0051] 如EMR图500中所示,表节点t1代表表T1的主键(primary key),即患者ID属性“PID”。因此,表节点也称“键节点”。表节点t1通过两个平实边双向连接到三个属性节点“PID”、“Gender”、“Age”。这三个属性节点对应于表T1的三个属性“PID”、“gender”、“age”。同样,表节点t2代表表T2的主键,即“EID”。表节点t2通过两个(在图5中以双向实线表示的)平实边,双向连接到四个属性节点“EID”、“Diag”、“Dept”和“PID”。四个属性节点对应于表T2的四个属性“EID”、“diag”、“dept”和“PID”。表节点t3代表表T3的主键,即“LID”。表节点t3通过两个平实边,双向连接到四个属性节点“LID”、“EID”、“Date”和“Tech”。这四个属性节点对应于表T3的四个属性“LID”、“EID”、“date”和“tech”。表节点t4通过两个平实边双向连接到两个属性节点“LName”、“LValue”。这两个属性节点对应于表T4的两个属性“LName”、“LValue”。

[0052] 表节点t1、t2、t3和t4是按照以下原则连接的。如果有从某表TA到某表TB的外键引用,则有从表节点tB到表节点tA的平实边。在EMR图500中也有以虚线箭头线表示的、从表节点tA到表节点tB的聚合边(aggregation edge)。在实施例中,“聚合边”意味着输入节点和输出节点的表有一对多或多对多的关系,如果在特征构建时需要用输出节点的属性(与输入节点连接),则需要“聚合节点”。聚合边是可传递的(transitive),换言之,如果表节点tY

与表节点tX不直接连接,但是从表节点tX可以通过遍历若干个聚合边而到达表节点tY,那么就有一个从表节点tX到表节点tY的聚合边。例如,在图5中,表T2和T1之间有一个外键“PID”引用(reference),所以,除了从表节点t1到表节点t2的平实边,还有从表节点t2到表节点t1的聚合边。

[0053] 同样,表T3和T2之间有一个外键“EID”引用,所以,除了从表节点t2到表节点t3的平实边,还有从表节点t3到表节点t2的聚合边。表T4和T3之间有一个外键“LID”引用,所以,除了从表节点t3到表节点t4的平实边,还有从表节点t4到表节点t3的聚合边。此外,表节点t3虽然不直接连接到表节点t1,但可以通过遍历两个聚合边到达表节点t1,所以从表节点t3到t1有一个聚合边。出于同样的原因,从表节点t4到t2有一个聚合边。同理,从表节点t4到t1有一个聚合边。

[0054] 以上段落描述了图2中的EMR表T1至T4所代表的EMR数据集与图5中的EMR图500之间的关系。一般来说,EMR图表示一个相关EMR表集合的属性关系。EMR图由表节点和属性节点组成,表节点分别对应于EMR表,每个表节点双向连接到对应于该表节点的EMR表的所有属性节点,分别对应于相应的表节点的病历表属性。表节点之间根据相应的EMR表引用关系而由平实边和聚合边相互连接。

[0055] 所属技术领域的熟练人员明白,按照关系进行适当编程,可以将表t1至t4所代表的EMR数据集自动转换为EMR图500。进一步,可以将任意关系模式的EMR数据集自动转换与图5中所示的类似的EMR图。这样,在步骤S320中就可以获得与如图2中所示的EMR表集合对应的EMR图500。

[0056] 返回图3,在步骤S310和S320之后,下一个步骤S330是基于知识树和EMR图生成子查询。在该步骤中,将从EMR图500生成与知识树400的每个概念节点对应的子查询。按照一个实施例,基于知识树和EMR图生成子查询的步骤S330包含:

[0057] 识别知识树的单一概念子树,如S331所示;

[0058] 对于所识别的单一概念子树中的每个概念子树:

[0059] 选择知识树中的一个单一概念子树,如S332所示;

[0060] 通过将该单一概念子树的目标节点和概念节点分别映射到EMR图的表节点和至少一个属性节点而识别EMR图的一个有效子图,如S333所示;和

[0061] 根据单一概念子树和有效子图构建子查询,如S334所示。

[0062] 换言之,步骤S330可通过如图3中所示的、包括步骤S331到S336的过程来执行。首先,执行步骤S331,以识别知识树的单一概念子树。实施例中,单一概念子树是知识树的一个子树,它包括从一个概念节点到目标节点或根节点的路径中的所有节点;此外,它还包括该路径中每个节点的不含其它概念节点的任何分支。换言之,单一概念子树应该符合以下标准:(1)该单一概念子树只有一个概念节点;(2)该单一概念子树包括从该一个概念节点起到目标节点止的路径中的所有节点;和(3)该单一概念子树包含该路径中的节点的任何子节点——只要该子节点中不含概念节点。

[0063] 图4B显示了知识树400的三个单一概念子树401、402和403。如图4B中所示,单一概念子树401包含节点{t-1,op2,op3,op4,c5,v6}。单一概念子树402包含节点{t-1,op2,op3,ag7,op8,c9,v10}。单一概念子树403包含节点{t-1,op2,op11,ag12,c13,f14,v15}。在实施例中,识别知识树的单一概念子树的步骤S331,可以通过以下面的方式分解知识树400来执

行。对于每个概念节点c5,找出从概念节点c5开始到目标节点t-1的路径。这条路径形成一个分解成的连通部分(或子树)的骨干。针对这条路径中的每个节点,检查其所有的子节点。如果某子节点不含任何概念节点,则将该子节点到连通部分(或子树)。当检查了该路径的所有节点后,该子树就是整个知识树的分解部分的其中之一。这样,就可识别一个单一概念子树(以下也称“子树”)。要指出的是,对于具有k个概念节点的知识树来说,可以识别出k个子树。例如,对于知识树400来说,可以确定三个子树401、402和403。

[0064] 在步骤S331后,循环地为每个子树执行步骤S332、S333和S334,直到所有子树401、402和403都被处理。在步骤S332,选择子树401。然后在步骤S333,识别EMR图500的一个有效子图。这可以通过将子树401的目标节点和概念节点映射到EMR图500的一个表节点和至少一个属性节点(以下也称“概念映射”)而完成。

[0065] 在一个实施例中,EMR图的有效子图是EMR图的符合以下标准的子图:(1)子图中有一个表节点与目标节点匹配;和(2)子图中的每个属性节点可以直接地或通过一个有向路径间接地到达该表节点。在一个实施例中,概念映射包括以下的任何一种或任意组合:(1)按属性的映射;(2)按属性值的映射;和(3)应用附加滤波器的映射。对于第一种映射,概念节点将被映射到一个或多个与概念节点的概念匹配的(即相同或视为相同的)属性节点。对于第二种映射,概念节点将被映射到一个或多个属性的一些特定值,并且可能基于概念数据类型和被映射属性的数据类型而与额外的属性相关联。对于第三种映射,概念节点将如第一种或第二种映射那样被映射,但是还根据过滤器的定义而与一个或多个属性关联。

[0066] 参看图6A,在步骤S333,以如下方式进行针对子树401的概念映射。首先,将子树401的目标节点t-1映射到EMR图500的表节点“t1”,如箭头线601所示的那样。然后,因为表T1的属性“age”与子树401的概念节点c5的概念“age”相同,概念节点c5将被映射到EMR图500的属性节点“Age”,如箭头线602所示的那样。由于属性节点“Age”可以直接到达与目标节点t-1匹配的表节点“t1”,包含属性节点“Age”和表节点“t1”的子图被确定为有效子图(以下也称“子图”),该子图在图6A中被高亮显示,记为{t1, Age}。

[0067] 然后,在步骤S334,根据子树401和有效子图{t1, Age},构造一个子查询。本例中,可以根据子树401的结构来构建子查询。具体来说,子树401的结构代表一个条件“age>65”(年龄大于65),因此导出该子查询是如下的子查询(1):

[0068] “t1.age>65” (1)

[0069] 在一个实施例中,根据子树的结构以及概念节点与属性节点的映射导出子查询。首先,生成子查询的主题(subject)。如果映射的属性节点(例如,与概念节点c5匹配的属性节点“Age”)直接连接到目标节点(例如,映射到表节点t-1的目标节点的t-1),则对应于属性节点的属性即为主题。在上面的例子中,主题是“age”(年龄)。

[0070] 图6A中显示过程是第一种概念映射的一个例子。在这个例子中,概念节点c5被映射到一个属性节点。然而,一个概念节点可能被映射到一个以上的属性节点。如果可以映射到若干个属性节点,则根据匹配的得分枚举每个映射。例如,如果概念“age”能映射到属性“Age”以及“first pregnancy age”(首次怀孕的年龄),则先评估(得分更高的)到前一个属性的映射;如果需要更多的查询建议,再评估后一个。

[0071] 在进行概念映射时,如果某个属性节点是通过跨多个表节点的一个或多个平实边连接到目标节点的,则根据该一个或多个平实边对所涉及的表节点应用一个或多个“join”

(联接),联接的表中被映射的属性将是主题。如果连通子图中包含有聚合边,则在表节点联接中使用相应的聚合算子(aggregation operators)。

[0072] 在针对子树401完成了子查询的生成的过程后,判断是否针对所有的子树完成了该过程,如方框S335所示的那样。如果是,则S330的过程结束,控制将返回一如方框S336所示的那样—以便执行步骤S340。否则,过程前进到步骤S332,以选择另一个子树,本例中选择的是子树402。

[0073] 参看图6B。在步骤S333,子树402的目标节点t-1被映射到EMR图500的表节点“t1”,如箭头线611所示的那样。然后,因为对应于表T2的属性“diag”的属性值“DM”被视为与子树402的概念节点c9的概念“Diabetes”相同,且EMR图的属性节点“Diag”与属性“diag”对应,所以概念节点c9将被映射到属性节点“Diag”—如箭头线612所示的那样,该属性节点是与表节点t2连通的。此时,两个被映射的节点“Diag”和“t1”并不连通。然而,子树402有聚合节点ag7,将其应用于激活从表节点“t2”到“t1”的聚合边<t2,t1>—如箭头线613所示的那样,这意味着表节点“t2”现在可以到达表节点“t1”了。此时,由于属性节点“Diag”可以通过平实边<Diag,t2>和被激活的聚合边<t2,t1>间接地到达表节点“t1”,表节点“t1”和表节点“t2”以及属性节点“Diag”构成EMR图500的一个连通子图,因此可确定其是一个有效子图,该子图在图6B中被高亮显示,记为{t1,t2,Diag}。

[0074] 然后,在步骤S334,根据子树402和{t1,t2,Diag}构建一个子查询。具体来说,子树402的结构表示一个条件“any patient who has diabetes”(任何糖尿病患者),因此导出该子查询是如下的子查询(2):

[0075] “t1join t2on PID,any(t2.diag= ‘DM’)group by PID” (2)

[0076] 子查询(2)的含义可以解释如下。利用属性“PID”连接表T1和T2(因此每个具有唯一性ID的患者在新表中单独成行),生成是否一个患者的任何诊断等于“DM”(糖尿病)的新的(布尔型)属性。

[0077] 在上述概念映射过程中,聚合节点ag7被用来通过激活聚合边<t2,t1>而连接被映射的表节点“t2”和“t1”。例如,如果子树中有k个聚合节点,则可用k个聚合边来连接EMR图500中被映射的节点,以使它们连接。图6B所示的过程是第二种概念映射的例子。对于第二种概念映射,子树的概念节点将被映射到一个或多个属性的一些特定值,并且可能基于概念数据类型和被映射属性的数据类型而与额外的属性相关联。

[0078] 在图6B的例子中,子图{t1,t2,Diag}构成一个连通子图,并且其可被确定为是一个有效子图{t1,t2,Diag},因为其符合以下标准:(1)子图中有一个表节点与目标节点匹配;和(2)子图中的每个属性节点可以直接地或通过一个有向路径间接地到达该表节点。在实施例中,也用“连通子图”一词来指有效子图。一般来说,如果有一条从一个映射的节点到目标节点的有向路径,则称该映射的节点与目标节点是连接的,或者说该映射的节点可以到达目标节点。如果多个节点被映射(即一个键节点和一个或多个属性节点被映射),那么,如果(1)从键节点到目标节点有一个有向路径;(2)从每个属性节点到键节点都有一个有向路径,则由该多个映射的节点组成的子图是个连通子图。这里,将目标节点(例如t-1)与目标节点映射的节点(例如t1)等同地看待。

[0079] 参看图6C,生成了子查询(2)后,在步骤S332选择子树403。然后,在步骤S333,以如下方式进行针对子树403的概念映射。首先,将子树403的目标节点t-1映射到EMR图500的表

节点“t1”，如箭头线621所示的那样。然后，由于对应于表T4的属性“LName”的属性值“HbA1c”（糖化血红蛋白）与子树403的概念节点c13的概念“HbA1c”是相同的，将概念节点c13映射到EMR图500的节点“LName”和节点“LValue”，如分裂的箭线622所示的那样。此外，由于与概念节点c13相关联的过滤器节点f14定义一个时间相关的条件，概念节点c13被映射到日期类型的属性节点“Date”，如箭头线623所示的那样。然后，将子树403中的聚合节点ag12映射到从表节点“t4”到表节点“t1”的聚合边，如箭头线624所示的那样。于是，聚合边<t4,t1>被激活。此时，属性节点“Date”可以通过平实边<Date,t3>、平实边<t3,t4>和聚合边<t4,t1>到达表节点“t1”，“LName”可以通过路径“LName”->“t4”->“t1”到达表节点“t1”，“LValue”可以通过路径“LValue”->“t4”->“t1”到达表节点“t1”。所以，对子树403的概念映射的结果，是识别出如图6C中高亮显示的一个有效子图{t1,t3,t4,Date,LName,LValue}。

[0080] 然后，在步骤S334、根据子树403和子图{t1,t3,t4,Date,LName,LValue}构建子查询。具体来说，按照子树403的结构所表示的条件，导出该子查询是如下的子查询(3)：

[0081] “t3join t4on LID,t4.LName=‘HbA1c’ and t3.Date>‘2015-07-01’ as t5; t1join t5,max(t5.LValue)>9.0group by PID (3)

[0082] 其查询近期“HbA1C”（糖化血红蛋白）检测结果的最大值大于9的所有患者。

[0083] 图6C中所示的过程是第三种概念映射的一个例子。一般来说，对于第三种映射，概念节点将如第一种或第二种映射那样被映射，还根据过滤器的定义而与一个或多个属性关联。

[0084] 回到图3。步骤S334执行后，过程前进到步骤S340中。在步骤S340中，根据知识树400来合成子查询(1)、(2)和(3)而构建查询。从知识树400中的节点op2的操作符“&&”和节点op3的操作符“&&”可知，该查询是上述三个子查询—子查询(1)、(2)和(3)—的交集(intersection)，即“t1.age>65”&&“t1join t2on PID,any(t2.diag=‘DM’)group by PID”&&“t3join t4on LID,t4.LName=‘HbA1c’ and t3.Date>‘2015-07-01’ as t5;t1join t5,max(t5.LValue)>9.0group by PID”。

[0085] 一个或多个实施例也提供对应的装置。在一个实施例中，提供一种用于根据临床知识生成用于从一个EMR表集合提取临床特征的查询的计算机实现的装置。该装置包含一个或多个处理器、与至少一个处理器耦合的存储器、存储在存储器中的计算机程序指令，其在被该至少一个处理器执行时执行以下操作：获取根据临床知识构建的知识树；获得对应于该EMR表集合的EMR图，其中EMR图包括表节点和属性节点，用以表示每个EMR表的结构和EMR表的属性之间的引用关系；根据知识树和EMR图生成子查询；根据知识树合成子查询而构造查询。

[0086] 根据一个实施例，基于知识树和EMR图的子查询的生成包括：

[0087] 识别知识树的单概念子树；

[0088] 对于单概念子树的每个单概念子树：

[0089] 选择知识树的单概念子树；

[0090] 通过将目标节点和单概念子树的概念节点分别映射到EMR图的表节点和至少一个属性节点来识别EMR图的有效子图；和

[0091] 根据单概念子树和有效子图组成子查询。

[0092] 根据一个实施例，所述获得对应于EMR表集合的EMR图包括将EMR表集合转换为其

等效图表示。在一个实施例中，知识树的单概念子树是包括从概念节点开始到目标节点的路径中的所有节点以及每个不包含概念节点的节点的任何分支的子树。

[0093] 根据一个实施例，EMR图的有效子图是EMR图的满足以下标准的子图，(1) 该子图中有与目标节点匹配的表节点；(2) 该子图中的每个属性节点可以直接或通过有向路径间接到达表节点。

[0094] 根据一个实施例，将单概念子树的一个目标节点和一个概念节点分别映射到EMR图的表节点和至少一个属性节点包括以下中的任何一个或任何组合：按属性映射；按属性值映射；已经通过应用附加过滤器进行映射。

[0095] 按照一个实施例，提供一种用于根据临床知识生成用于从EMR表集合提取临床特征的查询的计算机程序产品。该计算机程序产品包含存储有程序指令的计算机存储介质，程序指令能被至少一个处理器执行以使该至少一个处理器执行包含以下步骤的方法：

[0096] 获取根据临床知识构建的知识树；

[0097] 获得对应于该EMR表集合的EMR图，其中该EMR图包括表节点和属性节点，用以表示每个EMR表的结构和EMR表的属性之间的引用关系；

[0098] 根据知识树和EMR图生成子查询；和

[0099] 根据知识树合成子查询而构造查询。

[0100] 根据一个实施例，基于知识树和EMR图生成多个子查询包括：

[0101] 识别知识树的多个单概念子树；和

[0102] 对于所述多个单概念子树中的每个单概念子树：

[0103] 选择知识树的一个单概念子树；

[0104] 通过将该单概念子树的一个目标节点和一个概念节点分别映射到EMR图的一个表节点和至少一个属性节点来识别EMR图的一个有效子图；和

[0105] 按照该单概念子树和该有效子图构建一个子查询

[0106] 根据一个实施例，获得对应于EMR表集合的EMR图包括将EMR表集合转换为其等效图表示。

[0107] 根据一个实施例，知识树的单概念子树是包括从概念节点开始到目标节点的路径中的所有节点以及每个不包含概念节点的节点的任何分支的子树。

[0108] 根据一个实施例，EMR图的有效子图是EMR图的满足以下标准的子图，(1) 该子图中有与目标节点匹配的表节点；(2) 该子图中的每个属性节点可以直接或通过有向路径间接到达表节点。

[0109] 根据一个实施例，将单概念子树的一个目标节点和一个概念节点分别映射到EMR图的表节点和至少一个属性节点包括以下中的任何一个或任何组合：按属性映射；按属性值映射；已经通过应用附加过滤器进行映射。

[0110] 在任何可能的技术细节集成层面，本发明可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括计算机可读存储介质，其上载有用于使处理器实现本发明的各个方面的计算机可读程序指令。

[0111] 计算机可读存储介质是可以保持和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质例如可以是一一但不限于一一电存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或者上述的任意合适的组合。计算机可读存储介质的

更具体的例子(非穷举的列表)包括:便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、静态随机存取存储器(SRAM)、便携式压缩盘只读存储器(CD-ROM)、数字多功能盘(DVD)、记忆棒、软盘、机械编码设备、例如其上存储有指令的打孔卡或凹槽内凸起结构、以及上述的任意合适的组合。这里所使用的计算机可读存储介质不被解释为瞬时信号本身,诸如无线电波或者其他自由传播的电磁波、通过波导或其他传输媒介传播的电磁波(例如,通过光纤电缆的光脉冲)、或者通过电线传输的电信号。

[0112] 这里所描述的计算机可读程序指令可以从计算机可读存储介质下载到各个计算/处理设备,或者通过网络、例如因特网、局域网、广域网和/或无线网下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光纤传输、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或者网络接口从网络接收计算机可读程序指令,并转发该计算机可读程序指令,以供存储在各个计算/处理设备中的计算机可读存储介质中。

[0113] 用于执行本发明操作的计算机程序指令可以是汇编指令、指令集架构(ISA)指令、机器指令、机器相关指令、微代码、固件指令、状态设置数据、集成电路配置数据或者以一种或多种编程语言的任意组合编写的源代码或目标代码,所述编程语言包括面向对象的编程语言—诸如Smalltalk、C++等,以及过程式编程语言—诸如“C”语言或类似的编程语言。计算机可读程序指令可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络—包括局域网(LAN)或广域网(WAN)—连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。在一些实施例中,通过利用计算机可读程序指令的状态信息来个性化定制电子电路,例如可编程逻辑电路、现场可编程门阵列(FPGA)或可编程逻辑阵列(PLA),该电子电路可以执行计算机可读程序指令,从而实现本发明的各个方面。

[0114] 这里参照根据本发明实施例的方法、装置(系统)和计算机程序产品的流程图和/或框图描述了本发明的各个方面。应当理解,流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合,都可以由计算机可读程序指令实现。

[0115] 这些计算机可读程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理器,从而生产出一种机器,使得这些指令在通过计算机或其它可编程数据处理装置的处理器执行时,产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中,这些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作,从而,存储有指令的计算机可读介质则包括一个制品,其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

[0116] 也可以把计算机可读程序指令加载到计算机、其它可编程数据处理装置、或其它设备上,使得在计算机、其它可编程数据处理装置或其它设备上执行一系列操作步骤,以产生计算机实现的过程,从而使得在计算机、其它可编程数据处理装置、或其它设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

[0117] 附图中的流程图和框图显示了根据本发明的多个实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分,所述模块、程序段或指令的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0118] 为了示例的目的,描述了本发明的各种实施例,对各种实施例的描述并非穷尽性的,也并非是要把本发明仅限于所公开的实施例。对于本领域的普通技术人员而言,许多修改和变化将是显而易见的,而不背离所述实施例的范围和精神。选择本文所用术语是为了最好地解释实施例的原理、在市场上发现的技术的实际应用或技术改进,或使本领域的其他普通技术人员能够理解本文所公开的实施例。

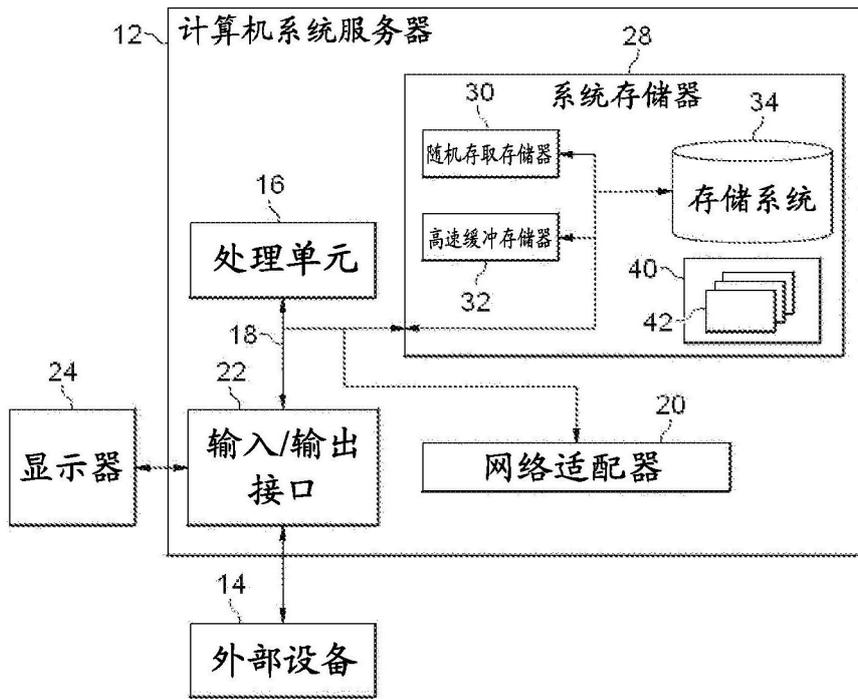


图1

T1: 患者表

PID+	性别 (gender)	年龄 (age)	...
P1	M	67	...
...

T2: 就诊表

EID+	diag	dept	PID*	...
E1	DM	D1	P1	...
...

T3: 化验表

LID+	EID*	日期 (date)	tech	...
L1	E1	15-10-11	Tech1	...
...

T4: 化验指标表

LName	LValue	LID*	...
HbA1C	10.5	L1	...
...

图2

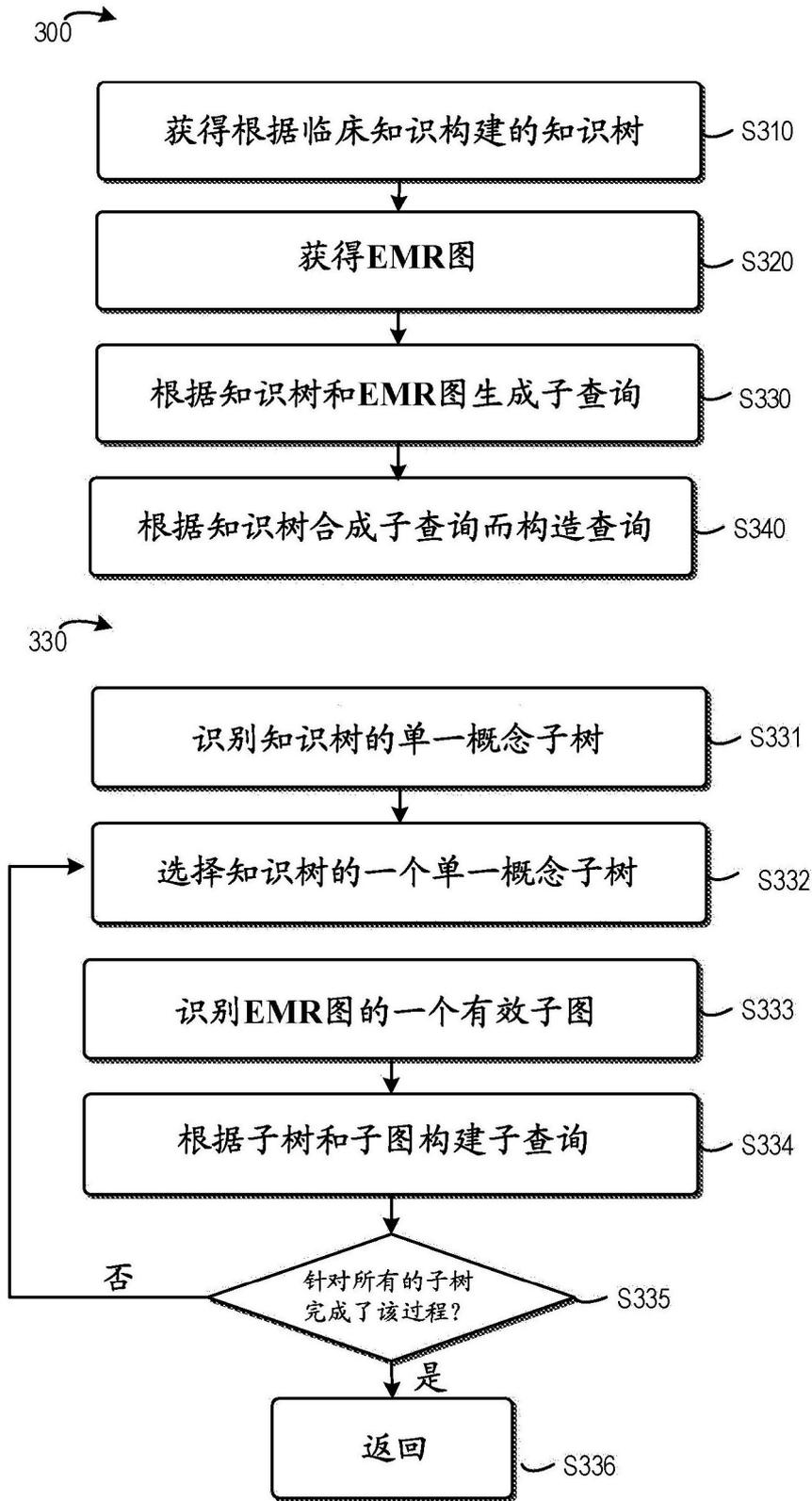


图3

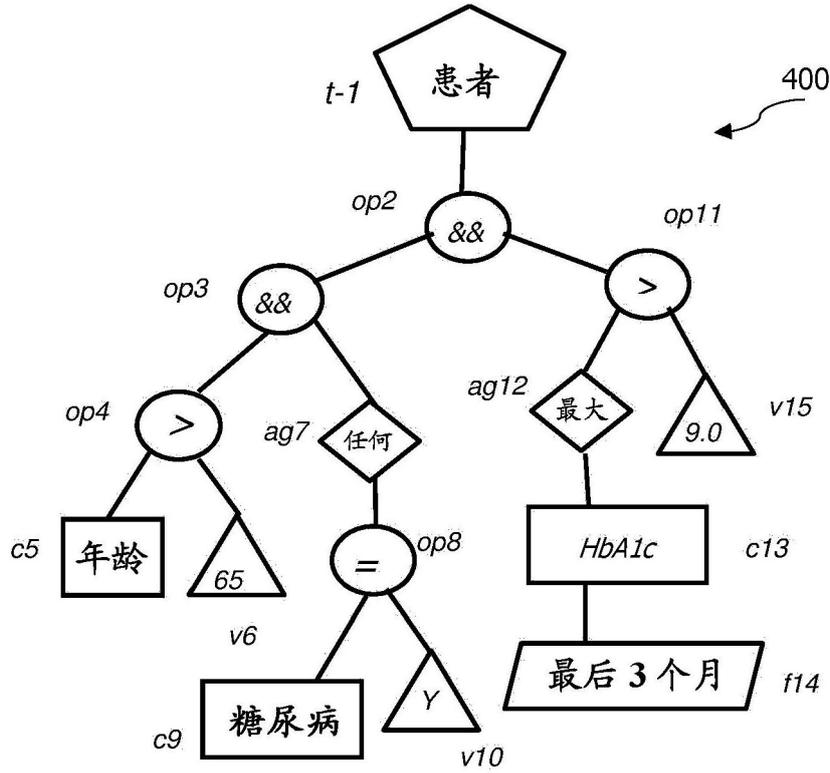


图4A

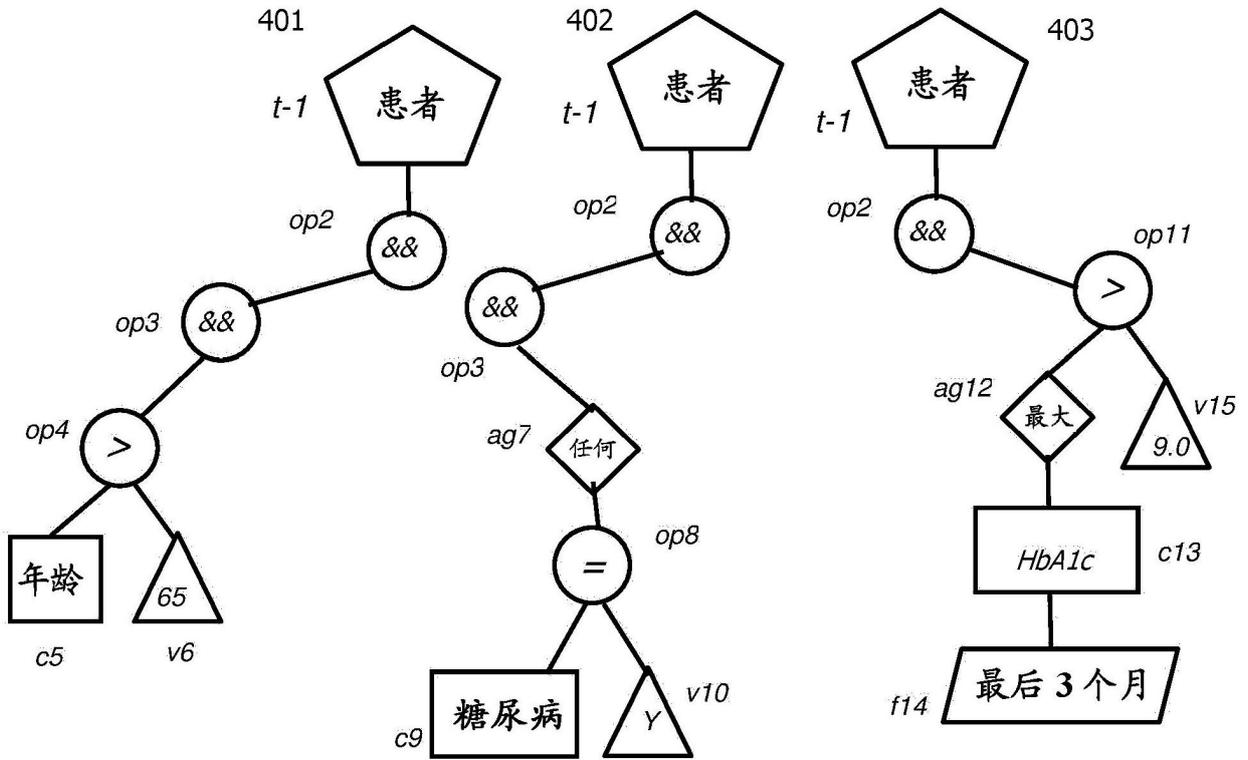


图4B

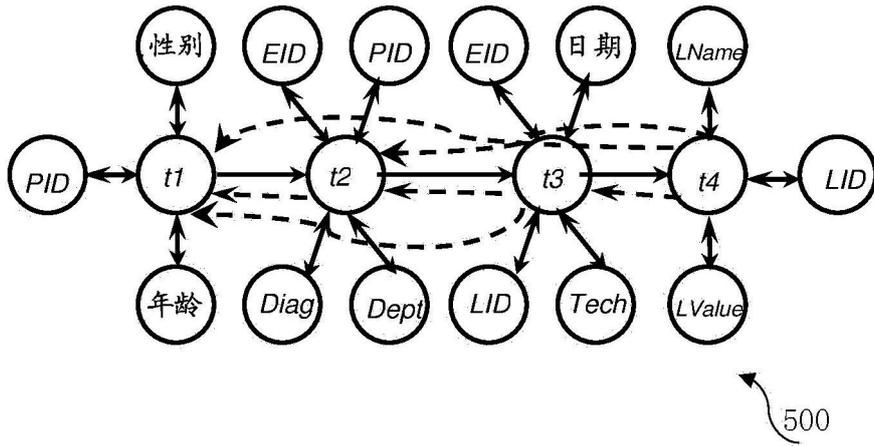


图5

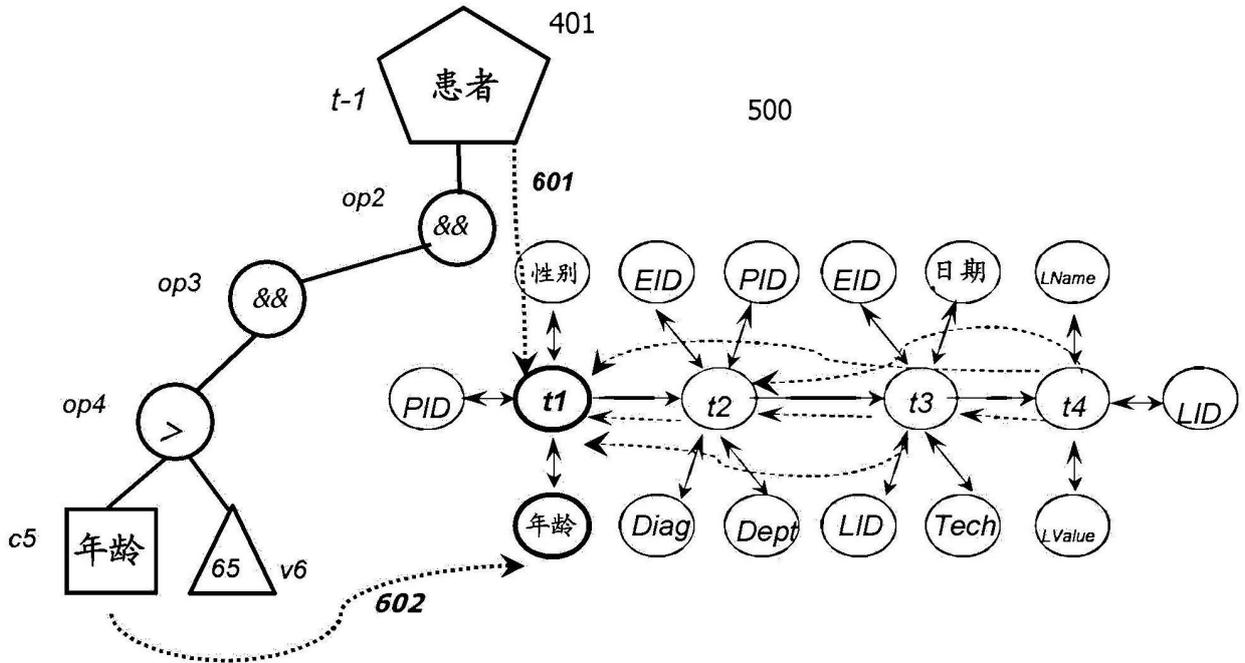


图6A

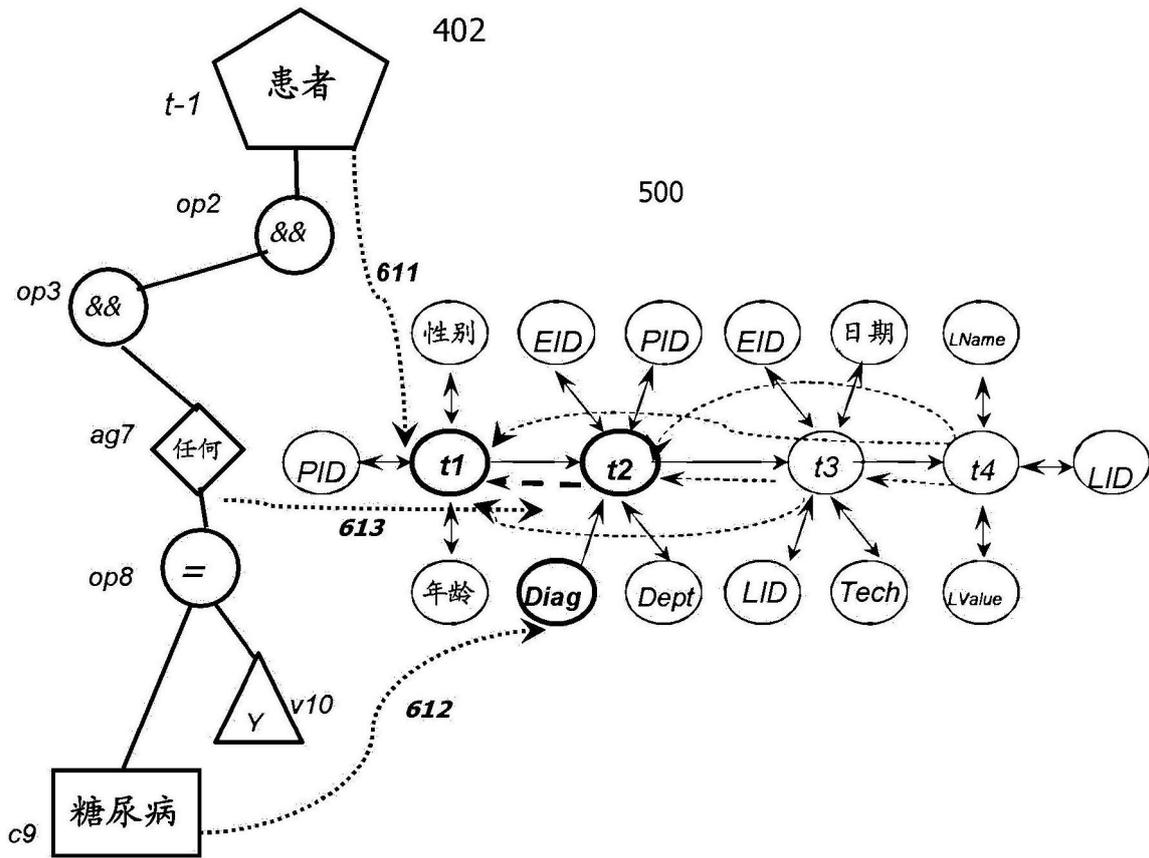


图6B

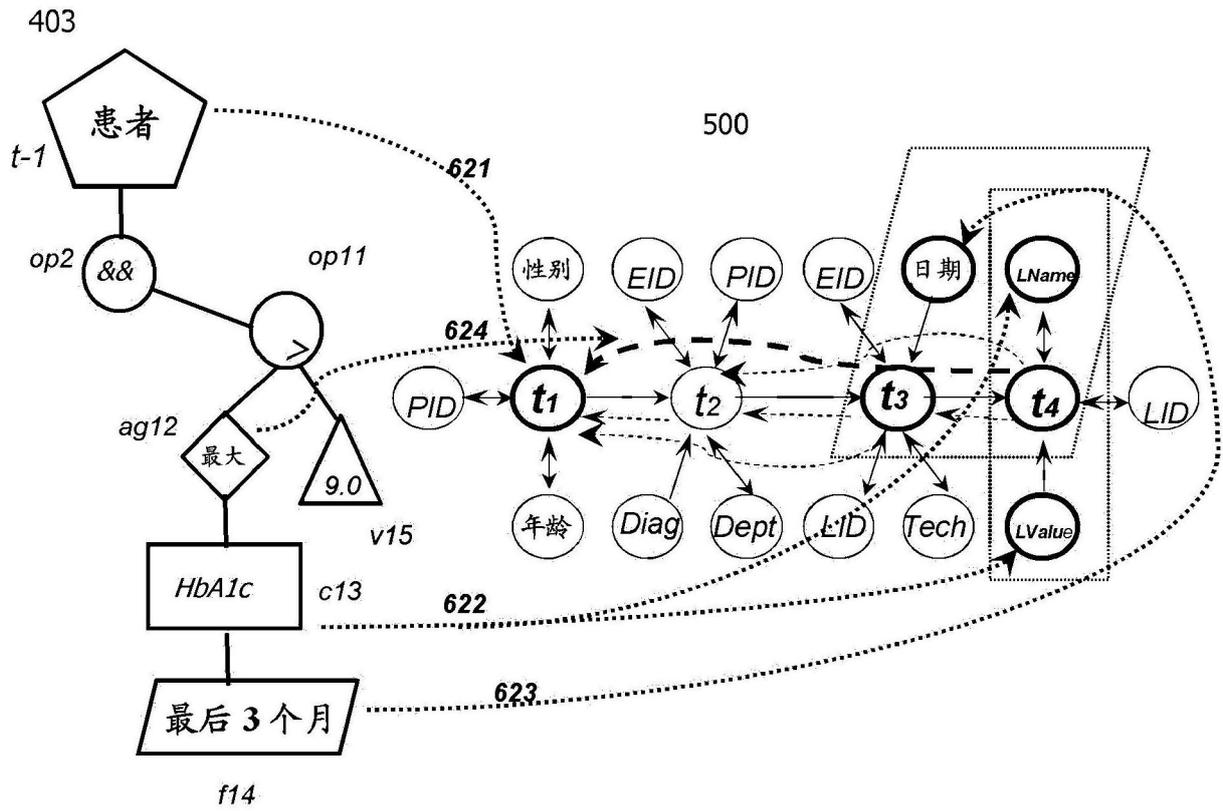


图6C