

# (12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织  
国际局

(43) 国际公布日  
2019年1月3日 (03.01.2019)



(10) 国际公布号  
**WO 2019/000920 A1**

- (51) 国际专利分类号:  
**G06F 19/00** (2018.01)
- (21) 国际申请号: PCT/CN2018/073004
- (22) 国际申请日: 2018年1月17日 (17.01.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
201710507636.9 2017年6月28日 (28.06.2017) CN
- (71) 申请人: 京东方科技集团股份有限公司  
**(BOE TECHNOLOGY GROUP CO., LTD.)** [CN/CN];  
中国北京市朝阳区酒仙桥路10号,  
Beijing 100015 (CN)。
- (72) 发明人: 张振中 (**ZHANG, Zhenzhong**); 中国北京市经济技术开发区地泽路9号,  
Beijing 100176 (CN)。
- (74) 代理人: 北京银龙知识产权代理有限公司(**DRAGON INTELLECTUAL PROPERTY LAW FIRM**); 中国北京市海淀区西直门北大街32号院枫蓝国际中心2号楼10层, Beijing 100082 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** AUTOMATIC INFERENCE RULE DISCOVERY METHOD AND SYSTEM, DATABASE AND SEARCH METHOD

(54) 发明名称: 一种推理规则自动发现方法及系统、数据库及检索方法

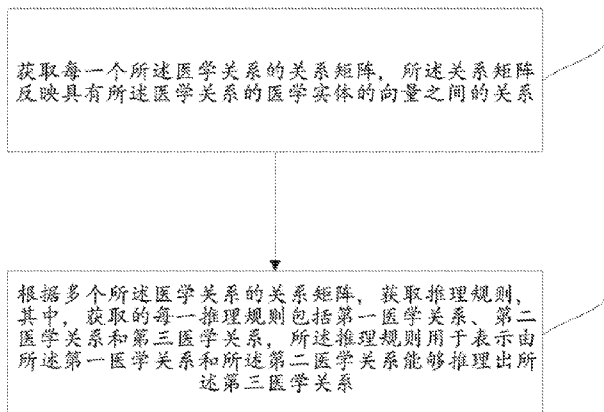


图 1

- 11 Acquire a relationship matrix of each medical relationship, the relationship matrix reflecting a relationship between vectors of medical entities having the medical relationships
- 12 According to the relationship matrices of the plurality of medical relationships, acquire inference rules, each of the acquired inference rules comprising a first medical relationship, a second medical relationship, and a third medical relationship, and the inference rule indicating that the third medical relationship can be inferred from the first medical relationship and the second medical relationship

(57) **Abstract:** An automatic inference rule discovery method and system, a medical knowledge database, and a search method. The automatic inference rule discovery method comprises: acquiring a relationship matrix of each medical relationship, the relationship matrix reflecting a relationship between vectors of medical entities having the medical relationships (11); according to the relationship matrices of the plurality of medical relationships, acquiring inference rules, each of the acquired inference rules comprising a first medical relationship, a second medical relationship, and a third medical relationship, and the inference rule indicating that the third medical relationship can be inferred from the first medical relationship and the second medical relationship (12). The method and system automatically construct inference rules in a medical knowledge database, without the need for manual definition, conserving manpower and material resources, reducing time costs, and facilitating search processes.

WO 2019/000920 A1

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,  
US, UZ, VC, VN, ZA, ZM, ZW。

- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

---

(57) 摘要: 一种推理规则自动发现方法及系统、医学知识数据库和检索方法。该推理规则自动发现方法包括: 获取每一个所述医学关系的关系矩阵, 所述关系矩阵反映具有所述医学关系的医学实体的向量之间的关系 (11); 根据多个所述医学关系的关系矩阵, 获取推理规则, 获取的每一个推理规则包括第一医学关系、第二医学关系和第三医学关系, 所述推理规则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系 (12)。该方法和系统能够自动构件医学知识库中的推理规则, 不需要人工定义, 节省了人力物力, 降低了时间成本; 同时方便了检索。

## 一种推理规则自动发现方法及系统、数据库及检索方法

### 相关申请的交叉参考

本申请主张在 2017 年 6 月 28 日在中国提交的中国专利申请号 No. 201710507636.9 的优先权，其全部内容通过引用包含于此。

### 技术领域

本公开涉及医学知识库数据挖掘技术领域，尤其涉及一种推理规则自动发现方法及系统、医学知识数据库及检索方法。

### 背景技术

当今，生物医学文献的数量以指数级的速度增长。海量的数据给研究者们带来丰富信息的同时，也让研究者们难以发现不同文献之间的关联信息，从而失去发现一些重要知识的机会。例如，在生物医学领域，生物医学文献数据库 MEDLINE 中的论文已经超过二千万，同时每年也有几百万篇的论文加入到数据库中。阅读如此海量的文献对于医学研究者来说是相当困难的。因此，从医学文献中自动发现知识的方法受到广泛关注。

### 发明内容

本公开提供一种推理规则自动发现方法及系统。还提供一种基于上述推理规则建立的医学知识数据库以及检索方法，使得检索人能够检索出更加完整的医学知识数据。

根据本公开的一个方面，本公开提供一种推理规则自动发现方法，其应用于医学知识库，所述医学知识库中包括多个医学知识，每一个所述医学知识为两个医学实体和一个医学关系的组合，所述方法包括：

获取每一个所述医学关系的关系矩阵，所述关系矩阵反映具有所述医学关系的医学实体的向量之间的关系；

根据多个所述医学关系的关系矩阵，获取所述推理规则，获取的每一个推理规则中包括第一医学关系、第二医学关系和第三医学关系，所述推理规

则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系。

可选地，所述获取每一个所述医学关系的关系矩阵的步骤包括：

针对每一个医学关系，构建一个初始的关系矩阵；

获取所述医学知识库中的所述医学关系对应的医学知识作为正确医学知识；

更换所述正确医学知识中的医学实体以得到新的医学知识作为错误医学知识；

采用打分函数为所述正确医学知识和错误医学知识分别打分；

采用预设目标函数，对所述初始的关系矩阵进行学习，使得所述正确医学知识的打分高于所述错误医学知识的打分，得到所述医学关系的关系矩阵。

可选地，对所述初始的关系矩阵进行学习时，采用的所述正确医学知识和错误医学知识的个数差小于预设阈值。

可选地，所述更换所述正确医学知识中的医学实体，得到新的医学知识作为错误医学知识的步骤包括：

采用除了包含在与所述医学关系对应的所有正确医学知识的医学实体之外的医学实体，更换所述正确医学知识中的医学实体，得到新的医学知识作为错误医学知识。

可选地，所述获取每一个所述医学关系的关系矩阵的步骤之前，还包括：

获取每一个所述医学实体的独热码向量：

$$\mathbf{x}_e = (a_1, \dots, a_i, \dots, a_N), \quad a_j = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases}$$

N 为所述医学知识库中的所有医学实体的个数， $\mathbf{x}_e$  为  $N \times 1$  维的向量；

根据所述每一个所述医学实体的独热码向量，获取每一个所述医学实体的向量：

$$\mathbf{y}_e = \sigma(W\mathbf{x}_e), \quad \sigma(x) = 1/(1 + e^{-x})$$

$\mathbf{y}_e$  为  $K \times 1$  维的向量，W 为映射矩阵，W 为  $K \times N$  维的向量，K 为预定的数值，K 小于 N。

可选地，所述打分函数为：

$$\text{Score}(r(e_1, e_2)) = y_{e_1}^T M_r y_{e_2}$$

其中， $r$  为所述医学关系， $e_1$ ， $e_2$  为所述医学实体， $r(e_1, e_2)$  为所述医学知识， $\text{Score}(\cdot)$  为打分函数， $M_r$  为所述医学关系  $r$  的关系矩阵。

可选地，所述预设目标函数为：

$$L = \sum_{r(e_1, e_2) \in T} \sum_{r(e_1', e_2') \in T'} \max\{\text{Score}(r(e_1', e_2')) - \text{Score}(r(e_1, e_2)) + 1, 0\}$$

其中， $L$  为目标函数， $T$  为所述正确医学知识集合， $T'$  为所述错误医学知识集合， $M_r$  为所述医学关系  $r$  的关系矩阵。

可选地，所述根据所述医学关系的关系矩阵，获取推理规则的步骤包括：从所有医学关系中，选择多个医学关系组，构建多个待验证推理规则，其中，每一个待验证推理规则对应一个医学关系组，每一个医学关系组包括三个医学关系；

计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵的相似度；

根据所述相似度，选择部分待验证推理规则作为最终得到的推理规则。

可选地，所述根据所述相似度，选择部分待验证推理规则作为最终得到的推理规则的步骤包括：

将所有所述医学关系组对应的相似度进行排序，得到相似度最大的预设数目个待验证推理规则；以及

将所述预设数目个待验证推理规则作为最终得到的推理规则。

可选地，每一个医学关系组中的三个医学关系满足如下条件：

- (1)  $L_p \cap H_q \neq \emptyset$ ；
- (2)  $H_p \cap H_r \neq \emptyset$ ；
- (3)  $L_q \cap L_r \neq \emptyset$ ；

其中， $H_p$ 、 $H_q$  和  $H_r$  分别是医学关系组中的三个医学关系  $p$ 、 $q$  和  $r$  对应的三元组中的所有第一个医学实体的集合， $L_p$ 、 $L_q$  和  $L_r$  分别是医学关系组中的三

个医学关系  $p$ 、 $q$  和  $r$  对应的三元组中的所有第二个医学实体的集合。

可选地，所述计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵的相似度的步骤包括：

计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵 L2 范数，将所述 L2 范数作为相似度。

根据本公开的另一个方面，本公开还提供一种推理规则自动发现系统，其应用于医学知识库，所述医学知识库中包括多个医学知识，每一个所述医学知识为两个医学实体和一个医学关系的组合，所述系统包括：

关系矩阵学习器，用于获取每一个所述医学关系的关系矩阵，所述关系矩阵反映具有所述医学关系的医学实体的向量之间的关系；

推理规则发现器，用于根据多个所述医学关系的关系矩阵，获取推理规则，其中，获取的每一个推理规则中包括第一医学关系、第二医学关系和第三医学关系，所述推理规则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系。

可选地，所述关系矩阵学习器，用于执行如下操作：针对每一医学关系，构建一个初始的关系矩阵；获取所述医学知识库中的所述医学关系对应的医学知识作为正确医学知识；更换所述正确医学知识中的医学实体，得到新的医学知识作为错误医学知识；采用打分函数为所述正确医学知识和错误医学知识分别打分；以及采用预设目标函数，对所述初始的关系矩阵进行学习，使得所述正确医学知识的打分高于所述错误医学知识的打分，得到所述医学关系的关系矩阵。

可选地，对所述初始的关系矩阵进行学习时，采用的所述正确医学知识和错误医学知识的个数差小于预设阈值。

可选地，所述关系矩阵学习器，用于采用除了包含在与所述医学关系对应的所有正确医学知识的医学实体之外的医学实体，更换所述正确医学知识中的医学实体，得到新的医学知识作为错误医学知识。

可选地，所述系统还包括：

实体向量学习器，用于执行如下操作：

获取每一所述医学实体的独热码向量：

$$\mathbf{x}_e = (a_1, \dots, a_i, \dots, a_N), \quad a_j = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases}$$

N 为所述医学知识库中的所有医学实体的个数， $\mathbf{x}_e$  为  $N \times 1$  维的向量；  
根据所述每一所述医学实体的独热码向量，获取每一所述医学实体的向量：

$$\mathbf{y}_e = \sigma(W\mathbf{x}_e), \quad \sigma(x) = 1/(1 + e^{-x})$$

$\mathbf{y}_e$  为  $K \times 1$  维的向量，W 为映射矩阵，W 为  $K \times N$  维的向量，K 为预定的数值，K 小于 N。

可选地，所述打分函数为：

$$\text{Score}(r(e_1, e_2)) = \mathbf{y}_{e_1}^T \mathbf{M}_r \mathbf{y}_{e_2}$$

其中，r 为所述医学关系， $e_1, e_2$  为所述医学实体， $r(e_1, e_2)$  为所述医学知识，Score( $\cdot$ ) 为打分函数， $\mathbf{M}_r$  为所述医学关系 r 的关系矩阵。

可选地，所述预设目标函数为：

$$L = \sum_{r(e_1, e_2) \in T} \sum_{r(e'_1, e'_2) \in T'} \max\{\text{Score}(r(e'_1, e'_2)) - \text{Score}(r(e_1, e_2)) + 1, 0\}$$

其中，L 为目标函数，T 为所述正确医学知识集合， $T'$  为所述错误医学知识集合， $\mathbf{M}_r$  为所述医学关系 r 的关系矩阵。

可选地，所述推理规则发现器，用于执行如下操作：从所有医学关系中，选择多个医学关系组，构建多个待验证推理规则，其中，每一个待验证推理规则对应一个医学关系组，每一个医学关系组包括三个医学关系；计算所述医学关系组中的其中两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵的相似度；以及根据所述相似度，选择部分待验证推理规则作为最终得到的推理规则。

可选地，所述推理规则发现器，用于将所有所述医学关系组对应的相似度进行排序，得到相似度最大的预设数目个待验证推理规则；将所述预设数目个待验证推理规则作为最终得到的推理规则。

可选地，每一个医学关系组中的三个医学关系满足如下条件：

$$(1) \quad L_p \cap H_q \neq \emptyset;$$

$$(2) H_p \cap H_r \neq \emptyset;$$

$$(3) L_q \cap L_r \neq \emptyset;$$

其中， $H_p$ 、 $H_q$ 和 $H_r$ 分别是医学关系组中的三个医学关系 p、q 和 r 对应的三元组中的所有第一个医学实体的集合， $L_p$ 、 $L_q$ 和 $L_r$ 分别是医学关系组中的三个医学关系 p、q 和 r 对应的三元组中的所有第二个医学实体的集合。

可选地，所述推理规则发现器，用于计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵 L2 范数，将所述 L2 范数作为相似度。

根据本公开的再一个方面，提供一种医学知识库的数据检索方法，包括：获取每一个医学关系的关系矩阵，其中，所述关系矩阵反映具有所述医学关系的两个医学实体的向量之间的关系，所述医学知识库包括多个医学知识，每一个所述医学知识为两个医学实体以及所述两个医学实体的医学关系的组合；

根据多个所述医学关系的关系矩阵，获取推理规则，其中，获取的每一个推理规则包括第一医学关系、第二医学关系和第三医学关系，所述推理规则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系；

输入检索词或检索式进行检索；

根据所述检索词或检索式基于所述推理规则进行检索获得与该检索词或检索式相关的医学知识；以及

输出与该检索词或检索式相关的医学知识。

可选地，所述输出与该检索词或检索式相关的医学知识包括显示与该检索词或检索式相关的医学知识。

根据本公开的另一个方面，提供一种医学知识数据库，包括：

数据输入器，用于输入医学知识数据；

关系矩阵学习器，用于获取每一个所述医学关系的关系矩阵，其中，所述关系矩阵反映具有所述医学关系的医学实体的向量之间的关系，所述医学

知识库包括多个医学知识，每一个所述医学知识为两个医学实体以及所述两个医学实体的医学关系的组合；

推理规则发现器，用于根据多个所述医学关系的关系矩阵，获取推理规则，其中，获取的每一个推理规则中包括第一医学关系、第二医学关系和第三医学关系，所述推理规则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系；

输入器，用于输入检索词或检索式；

检索器，用于根据所述检索词或检索式基于所述推理规则进行检索以获得与该检索词或检索式相关的医学知识；以及

输出器，用于输出与该检索词或检索式相关的医学知识。

可选地，所述输出器用于显示与该检索词或检索式相关的医学知识。

## 附图说明

为了更清楚地说明本公开实施例的技术方案，下面将对本公开实施例的描述中所需要使用的附图作简单地介绍。显而易见地，下面描述中的附图仅仅是本公开的一些实施例，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。

图 1 为本公开一实施例的推理规则自动发现方法的流程示意图；

图 2 为本公开实施例的医学关系的关系矩阵的获取方法的流程示意图；

图 3 为本公开另一实施例的推理规则自动发现方法的流程示意图；

图 4 为本公开实施例的推理规则自动发现系统的结构框图；

图 5 为本公开实施例的数据库的结构框图；

图 6 为本公开实施例的检索方法的流程示意图。

## 具体实施方式

为使本公开实施例的目的、技术方案和优点更加清楚，下面将结合本公开实施例的附图，对本公开实施例的技术方案进行清楚、完整地描述。显然，所描述的实施例是本公开的一部分实施例，而不是全部的实施例。基于所描述的本公开的实施例，本领域普通技术人员所获得的所有其他实施例，都属

于本公开保护的范围内。

传统的医学知识自动发现方法需要人工定义推理规则。例如，人工定义推理规则“包含(A,B)<sup>^</sup>治疗(B,C) $\Rightarrow$ 预防(A,C)”。基于事实“金针菇含有朴菇素”和“朴菇素可以治疗癌症”，依据上面定义的推理规则，可以得出“金针菇可以预防癌症”。但是，目前医学领域包含的信息巨大，人工定义推理规则需要大量的时间和人力财力，也不利于检索出完整的医学知识数据。而在本公开的实施例中，通过将医学关系表示成矩阵，然后通过矩阵之间的运算，从海量的医学知识中自动发现推理规则，不需要人工定义推理规则，节省了大量的人力物力，降低了时间成本；基于上述推理规则建立的医学知识数据库以及检索方法，使得检索人能够检索出更加完整的医学知识数据。

根据本公开实施例的推理规则自动发现方法及系统应用于医学知识库，可以从医学知识库中自动发现推理规则。

本公开实施例中的医学知识库包括多个医学知识，每一个所述医学知识为两个医学实体和一个医学关系的组合。可见，一个医学知识包括三个元素，本公开实施例中，可以采用三元组的方式表示医学知识，例如三元组  $r(e_1, e_2)$ ，其中， $r$ 表示医学关系， $e_1$ 和 $e_2$ 表示医学实体。举例来说，治疗(朴菇素，癌症)，医学关系是治疗，朴菇素和癌症是医学实体，即，朴菇素可以用于治疗癌症。

本公开实施例中，可以将医学知识库中的医学实体表示成向量。

例如，可以将医学实体表示成“one-hot（独热码）”向量。所谓“one-hot”向量是指只有相应的维度值为1，其余都为0。假设医学知识库中一共有N个医学实体，可以按照任意顺序排列这N个医学实体，假设医学实体“朴菇素”的位置为*i*，则“朴菇素”的“one-hot”向量  $x_e$  为  $(a_1, \dots, a_i, \dots, a_N)$ ，其中

$a_j = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases}$ ，即只有第*i*个元素为1，其余N-1个为0。可以看出，“one-hot”

向量的维度和医学实体的总数相等，每一维度表示一个医学实体，每个医学实体的“one-hot”向量只有自己对应的维度为1，其余维度为0。举例来说，假设一共有三个医学实体 {钙，朴菇素，锌}，对应的排列顺序为1，2，3，则

每个医学实体的“one-hot”向量都是3维，只有自己对应的维度为1，其余为0，如钙的“one-hot”向量为(1,0,0)。

从上面的描述可以看出，“one-hot”向量是 $N \times 1$ 维的向量，医学知识库中的医学实体的数量非常庞大，即 $N$ 的数值非常大，如果医学实体采用“one-hot”向量表示，则数量非常之大，难以处理。另外，由于医学知识库中可能存在不同名称的医学实体实际上是同一实体，或者，相似的医学实体的情况，例如医学知识库中包括医学实体“乙酰氨基酚”和“扑热息痛”，而两者实际上是同一实体，有必要将两者采用相同或相似的向量表示。

本公开实施例中，可以对高维空间的“one-hot”向量进行映射，将其映射为低维空间的向量。低维空间的向量可以有效降低向量的维度，且还可以将相同或相似的医学实体表示成相同或相似的向量。

例如，假设映射矩阵 $W$ 为 $K \times N$ 维的向量， $K$ 为小于 $N$ 的正整数，则医学实体 $e$ 的向量 $y_e$ 可以表示为： $y_e = \sigma(Wx_e)$ ，其中， $\sigma$ 为sigmoid函数， $x_e$ 为医学实体 $e$ 的“one-hot”向量， $\sigma(x) = 1 / (1 + e^{-x})$ ， $\sigma(x)$ 的公式中 $e$ 为数学常数。从上述 $y_e$ 的计算公式可以看出，由于 $W$ 为 $K \times N$ 维的向量， $x_e$ 为 $N \times 1$ 的向量，因而， $W * x_e$ 为 $K \times 1$ 维的向量， $y_e$ 为 $K \times 1$ 维的向量。由于 $K$ 小于 $N$ ，因而，映射后的向量 $y_e$ 的维数小于“one-hot”向量的维数。 $K$ 的数值可以根据需要设置，通常远小于 $N$ 。对于映射矩阵 $W$ ，可以通过预设目标函数学习其中每个元素的值，后续将详细介绍。

医学知识库中包括多个医学关系，例如，“治疗”，“预防”，“包含”等。本公开实施例中，每一个医学关系可以表示成一个关系矩阵 $M_r$ ，所述关系矩阵 $M_r$ 反映具有所述医学关系的医学实体的向量之间的关系。因为将医学实体从 $N$ 维的“one-hot”向量降维到 $K$ 维的向量，即所有医学实体都被映射到一个 $K$ 维空间（该 $K$ 维空间简记为 $S$ ），而关系矩阵用来反映医学实体之间的关系，所以关系矩阵的大小为 $K \times K$ ，即每个关系矩阵中有 $K^2$ 个元素。第 $i$ 行第 $j$ 列的元素 $M_r(i,j)$ （其中 $1 \leq i, j \leq K$ ）反映了空间 $S$ 的第 $i$ 维度和第 $j$ 维度的相互关系。本公开实施例中，关系矩阵 $M_r$ 可以通过预设目标函数学习其中每个元素的值，后续将详细介绍。

本公开实施例中，通过医学关系的关系矩阵之间的运算，来发现推理规

则。具体的，可以通过计算任意两个医学关系的关系矩阵的乘积，并计算得到的乘积与另一医学关系的关系矩阵的相似度。如果得到的乘积与另一医学关系的关系矩阵相似，则可以得到三个医学关系之间的推理规则。

请参考图1，图1为本公开一实施例的推理规则自动发现方法的流程示意图，该推理规则自动发现方法应用于医学知识库，所述医学知识库中包括多个医学知识，每一个所述医学知识为两个医学实体和一个医学关系的组合，所述方法包括：

步骤 11：获取每一个所述医学关系的关系矩阵，所述关系矩阵反映具有所述医学关系的医学实体的向量之间的关系；

步骤 12：根据多个所述医学关系的关系矩阵，获取推理规则，其中，获取的每一推理规则中包括第一医学关系、第二医学关系和第三医学关系，所述推理规则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系。

得到的推理规则为：第一医学关系 $\wedge$ 第二医学关系 $\Rightarrow$ 第三医学关系， $\wedge$ 为逻辑与符号。

本公开实施例中，通过将医学关系表示成矩阵，然后通过矩阵之间的运算，从海量的医学知识中自动发现推理规则，不需要人工定义推理规则，节省了大量的人力物力，降低了时间成本。

可选地，第一医学关系 $r_1$ 、第二医学关系 $r_2$ 和第三医学关系 $r_3$ 满足如下条件：

- (1)  $L_{r_1} \cap H_{r_2} \neq \emptyset$ ;
- (2)  $H_{r_1} \cap H_{r_2} \neq \emptyset$ ;
- (3)  $L_{r_2} \cap L_{r_3} \neq \emptyset$ ;

其中， $H_{r_1}$ 、 $H_{r_2}$ 和 $H_{r_3}$ 分别是第一医学关系  $r_1$ 、第二医学关系  $r_2$ 和第三医学关系  $r_3$ 对应的三元组中的所有第一个医学实体的集合， $L_{r_1}$ 、 $L_{r_2}$ 和 $L_{r_3}$ 分别是第一医学关系  $r_1$ 、第二医学关系  $r_2$ 和第三医学关系  $r_3$ 对应的三元组中的所有第二个医学实体的集合。

下面对如何确定医学关系的关系矩阵进行详细说明。

上述实施例中提到，可以采用预设目标函数对医学关系的关系矩阵进行学习。

本公开实施例中，预设目标函数是基于医学关系对应的医学知识的打分对医学关系的关系矩阵进行优化，采用预设目标函数学习的目的是使得正确的医学知识的打分高于错误的医学知识的打分。可以理解的是，医学知识库中的医学知识是经过实验和人工检验的，因此，本公开实施例中，使用医学知识库中的医学知识作为正确的医学知识。而，随机构建的医学知识可以认为是错误的医学知识。举例来说，医学知识库中的医学知识“治疗(朴菇素，癌症)”作为正确的医学知识，随机构建的医学知识“治疗(维生素 A，癌症)”作为错误的医学知识。

请参考图 2，图 2 为本公开实施例的医学关系的关系矩阵的获取方法的流程示意图，该方法包括：

步骤 111：针对每一个医学关系，构建一个初始的关系矩阵；

初始的关系矩阵中，每一个元素可以是任意的数值。

步骤 112：获取所述医学知识库中的所述医学关系对应的医学知识作为正确医学知识；

例如，获取到的医学知识库中医学关系“治疗”对应的医学知识“治疗(朴菇素，癌症)”，“治疗(乙酰氨基酚，发烧)”，“治疗(炉甘石，止痒)”等。

步骤 113：更换所述正确医学知识中的医学实体，得到新的医学知识作为错误医学知识；

举例来说，可以采用“维生素 A”更换医学知识“治疗(朴菇素，癌症)”中的元素“朴菇素”，得到错误的医学知识“治疗(维生素 A，癌症)”。

步骤 114：采用打分函数为所述正确医学知识和错误医学知识分别打分；

步骤 115：采用预设目标函数，对所述初始的关系矩阵进行学习，使得所述正确医学知识的打分高于所述错误医学知识的打分，得到所述医学关系的关系矩阵。

本公开实施例中，基于正确的医学知识的打分要高于错误的医学知识的打分的原则，采用预设目标函数对医学关系的关系矩阵进行学习，学习方法

简单，有效。

可选地，对所述初始的关系矩阵进行学习时，采用的所述正确医学知识和错误医学知识的个数差小于预设阈值。进一步可选地，采用的所述正确医学知识和错误医学知识的个数相同。即，对所述初始的关系矩阵进行学习时，采用正确医学知识的个数和错误医学知识的个数相同或基本相同，从而保证正确的学习结果。

本公开实施例中，可以采用下述打分函数对医学知识进行打分：

$$\text{Score}(r(e_1, e_2)) = y_{e_1}^T M_r y_{e_2}$$

其中， $r$  为所述医学关系， $e_1$ ， $e_2$  为所述医学实体， $r(e_1, e_2)$  为所述医学知识， $\text{Score}(\cdot)$  为打分函数。

当然，在本公开的其他一些实施例中，也不排除采用其他类型的打分函数的可能。

本公开实施例中，可以采用如下预设目标函数对医学关系的关系矩阵进行学习：

$$L = \sum_{r(e_1, e_2) \in T} \sum_{r(e'_1, e'_2) \in T'} \max\{\text{Score}(r(e'_1, e'_2)) - \text{Score}(r(e_1, e_2)) + 1, 0\}$$

其中， $L$  为目标函数， $T$  为所述正确医学知识集合， $T'$  为所述错误医学知识集合， $M_r$  为所述医学关系  $r$  的关系矩阵。

本公开实施例中，是采用随机梯度下降方法最小化预设目标函数的方式，学习医学关系的关系矩阵。当然，在本公开的其他一些实施例中，也不排除采用其他优化算法。

本公开实施例中，在构建错误的医学知识时，例如在构建医学关系  $r$  的错误的医学知识时，可以从正确的医学知识集合  $T$  中选取一个医学知识  $r(e_1, e_2)$ ，固定  $r(e_1, e_2)$  中的一个医学实体，随机变换另一个医学实体，例如保持  $e_1$  不变，将  $e_2$  随机变换为  $e'$ ，来构建错误的医学知识集合  $T'$ 。

也就是说，本公开实施例中，在更换所述医学关系对应的正确医学知识中的医学实体时，可以将所述正确医学知识中的医学实体随机变换成另一个医学实体，得到错误的医学知识。

可以理解的是，采用随机替换的方法得到的医学知识，也不一定绝对是

错误的，有可能会得到一个正确的医学知识。例如将正确医学知识中“治疗（朴菇素，癌症）”中的朴菇素替换成阿糖胞苷，得到医学知识“治疗（阿糖胞苷，癌症）”，这有可能是正确的。然而，本公开实施例中所说的错误医学知识是指统计意义上的，即绝大部分随机得到的医学知识都是错误的。举例说明，假设医学知识库中存在 100000 个医学实体，其中用来治疗癌症的医学实体其实很少，例如由 1000 个医学实体，随机替换后只有  $1000/100000=1\%$  的概率是对的，99%的概率是错误的。如果我们替换很多次，例如 10000 次，依据大数定律则这 10000 次替换的结果中 99%的都是错误的，即使有很小一部分是对的也不会造成很大影响。

可选地，在更换与所述医学关系对应的正确医学知识中的医学实体时，采用除了包含在与所述医学关系对应的正确医学知识的医学实体之外的医学实体，更换所述正确医学知识中的医学实体，得到新的医学知识作为错误医学知识。举例来说，医学关系“治疗”所对应的所有医学知识中，不包括医学实体“金针菇”“生姜”等，则可以采用“金针菇”“生姜”等更换“治疗”所对应的所有医学知识中的医学实体，使得得到的错误的医学知识是正确医学知识的概率减少。

上述实施例中提到，在进行医学关系的关系矩阵学习之前，还需要获取医学实体的向量，本公开实施例中采用以下方式得到医学实体的向量：

获取每一个所述医学实体的独热码向量：

$$\mathbf{x}_e = (a_1, \dots, a_i, \dots, a_N), \quad a_j = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases}$$

$N$  为所述医学知识库中的所有医学实体的个数， $\mathbf{x}_e$  为  $N \times 1$  维的向量；

根据所述每一个所述医学实体的独热码向量，获取每一个所述医学实体的向量：

$$\mathbf{y}_e = \sigma(W\mathbf{x}_e), \quad \sigma(\mathbf{x}) = 1/(1 + e^{-\mathbf{x}})$$

$\mathbf{y}_e$  为  $K \times 1$  维的向量， $W$  为映射矩阵， $W$  为  $K \times N$  维的向量， $K$  为预定的数值， $K$  小于  $N$ 。

映射矩阵  $W$  的初始矩阵中的元素可以是随机的数值。从上述预设目标函数的公式可以看出，在学习得到关系矩阵，也可以同时学习到最终的映射矩

阵 W。

下面对本公开实施例中自动发现推理规则的原理进行说明。

可以将医学实体的向量归一化成单位向量（即模为 1 的向量），即对任一个医学实体  $e$ ，都有  $\|e\|=1$ （ $\|\cdot\|$  表示长度，也即欧式距离）。同时，将向量  $y_{e_1}^T M_{r_1}$  也归一化为单位向量。

因为将向量  $y_{e_1}^T M_{r_1}$  和实体向量  $y_{e_2}$  归一化为单位向量后， $y_{e_1}^T M_{r_1} y_{e_2}$  表示的就是向量  $y_{e_1}^T M_{r_1}$  和  $y_{e_2}$  之间夹角的  $\cos$  值。按照前面的预设目标函数，正确的医学知识的打分至少要比错误的医学知识的打分要大 1，此时  $y_{e_1}^T M_{r_1} y_{e_2}$  的值越大越好，但  $\cos$  的最大值为 1，所以优化预设目标函数的一个结果就是会使正确的医学知识的打分趋近 1，即  $y_{e_1}^T M_{r_1} y_{e_2} \approx 1$ 。因此，对于正确的医学知识  $r_1(e_1, e_2)$ ，则有  $y_{e_1}^T M_{r_1} y_{e_2} \approx 1$ 。

因为  $y_{e_2}$  归一化为单位向量，所以  $y_{e_1}^T y_{e_2} \approx 1$ ，又因为  $y_{e_1}^T M_{r_1} y_{e_2} \approx 1$ ，并且  $y_{e_1}^T M_{r_1}$  也是归一化后的单位向量，所以得到  $y_{e_1}^T M_{r_1} \approx y_{e_2}^T$ 。直观上就是任何一个非 0 单位向量只有与自身相同的单位向量的夹角余弦值为 1。

同理，对于正确的医学知识  $r_2(e_2, e_3)$ ，可以得到  $y_{e_2}^T M_{r_2} \approx y_{e_3}^T$ 。

从而得到， $y_{e_1}^T (M_{r_1} M_{r_2}) \approx y_{e_3}^T$ 。

因为  $r_3(e_1, e_3)$  成立，所以  $y_{e_1}^T M_{r_1} \approx y_{e_3}^T$ ，所以有  $M_{r_1} M_{r_2} \approx M_{r_3}$ 。

也就是说，对于医学关系  $r_1$ 、 $r_2$  和  $r_3$ ，如果  $M_{r_1} M_{r_2} \approx M_{r_3}$ ，则可以得到推理规则  $r_1 \wedge r_2 \Rightarrow r_3$ 。

同样的，对于医学关系  $r_4$ 、 $r_5$  和  $r_6$ ，如果  $M_{r_4} M_{r_5} \approx M_{r_6}$ ，则可以得到推理规则  $r_4 \wedge r_5 \Rightarrow r_6$ 。

基于上述分析，本公开实施例可以通过矩阵之间的乘法来发现推理规则。

本方实施例中，如果计算得到多个推理规则，也可以对多个推理规则对应的相似度进行排序，选取相似度最大的预设数目个推理规则，作为最终的

推理规则。

可选地，所述根据所述医学关系的关系矩阵，获取推理规则的步骤包括：从所有医学关系中，选择多个医学关系组，构建多个待验证推理规则，其中，每一个待验证推理规则对应一个医学关系组，每一个医学关系组包括三个医学关系；

计算所述医学关系组中的其中两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵的相似度；

根据所述相似度，选择部分待验证推理规则作为最终得到的推理规则。

请参考图3，图3为本公开另一实施例的推理规则自动发现方法的流程示意图，该推理规则自动发现方法应用于医学知识库，所述医学知识库中包括多个医学知识，每一个所述医学知识为两个医学实体和一个医学关系的组合，所述方法包括：

步骤 31：获取每一个所述医学关系的关系矩阵，所述关系矩阵反映具有所述医学关系的医学实体的向量之间的关系；

步骤 32：从所有医学关系中，选择多个医学关系组，构建多个待验证推理规则，其中，每一待验证推理规则对应一个医学关系组，每一个医学关系组包括三个医学关系；

步骤 33：计算所述医学关系组中的其中两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵的相似度；

步骤 34：将所有所述医学关系组对应的相似度进行排序，得到相似度最大的预设数目个待验证推理规则；

步骤 35：将所述预设数目个待验证推理规则作为最终得到的推理规则。

本公开实施例中，对所有推理规则对应的相似度进行排序，得到相似度最大的几个推理规则作为最终得到的推理规则，使得得到的推理规则更准确。

除了根据相似度进行排序之外，也可以采用其他方法从多个待验证推理规则中选择最终的推理规则。例如，选取相似度大于预设阈值的待验证推理规则作为最终的推理规则。

本公开实施例中，可以从所有医学关系中，任意选择三个医学关系构建一待验证推理规则。

优选地，每一医学关系组中的三个医学关系满足如下条件：

- (1)  $L_p \cap H_q \neq \emptyset$ ;
- (2)  $H_p \cap H_r \neq \emptyset$ ;
- (3)  $L_q \cap L_r \neq \emptyset$ ;

其中， $H_p$ 、 $H_q$ 和 $H_r$ 分别是医学关系组中的三个医学关系 p、q 和 r 对应的三元组中的所有第一个医学实体的集合， $L_p$ 、 $L_q$ 和 $L_r$ 分别是医学关系组中的三个医学关系 p、q 和 r 对应的三元组中的所有第二个医学实体的集合。

可选地，所述计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵的相似度的步骤包括：

计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵 L2 范数，将所述 L2 范数作为相似度。

具体的计算方法可以如下：

对于给定医学关系 r 的三元组 $r(e_1, e_2)$ ，设医学关系 r 的三元组中所有医学实体 $e_1$ 组成的集合为 $H_r$ ，所有医学实体 $e_2$ 组成的集合为 $L_r$ 。对于关系 r，本公开实施例中，从满足如下条件的医学关系对 (p, q) 中找出可能的推理规则  $p \wedge q \Rightarrow r$ ；

- (1)  $L_p \cap H_q \neq \emptyset$ ;
- (2)  $H_p \cap H_r \neq \emptyset$ ;
- (3)  $L_q \cap L_r \neq \emptyset$ ;

对于满足上述条件的所有医学关系对 (p, q)，本公开实施例通过如下计算公式计算 L2 范数，并按照从小到大的顺序排列计算出的 L2 范数对应的医学关系对： $\|M_p M_q - M_r\|_2$ ，其中  $\|\cdot\|_2$  是矩阵的 L2 范数。

选取前预设数目个 L2 范数最大的医学关系对 (p, q)，得出推理规则  $p \wedge q \Rightarrow r$ 。

基于同一发明构思，请参考图 4，本公开实施例还提供一种推理规则自动发现系统，其应用于医学知识库，所述医学知识库中包括多个医学知识，

每一个所述医学知识为两个医学实体和一个医学关系的组合，所述系统包括：

关系矩阵学习器，用于获取每一个所述医学关系的关系矩阵，所述关系矩阵反映具有所述医学关系的医学实体的向量之间的关系；

推理规则发现器，用于根据多个所述医学关系的关系矩阵，获取推理规则，其中，获取的每一推理规则中包括第一医学关系、第二医学关系和第三医学关系，所述推理规则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系。

本公开实施例中，通过将医学关系表示成矩阵，然后通过矩阵之间的运算，从海量的医学知识中自动发现推理规则，不需要人工定义推理规则，节省了大量的人力物力，降低了时间成本。

可选地，所述关系矩阵学习器，用于执行如下操作：针对每一医学关系，构建一个初始的关系矩阵；获取所述医学知识库中的所述医学关系对应的医学知识作为正确医学知识；更换所述正确医学知识中的医学实体，得到新的医学知识作为错误医学知识；采用打分函数为所述正确医学知识和错误医学知识分别打分；以及采用预设目标函数，对所述初始的关系矩阵进行学习，使得所述正确医学知识的打分高于所述错误医学知识的打分，得到所述医学关系的关系矩阵。

可选地，对所述初始的关系矩阵进行学习时，采用的所述正确医学知识和错误医学知识的个数差小于预设阈值。

可选地，所述关系矩阵学习器，用于采用除了包含在与所述医学关系对应的所有正确医学知识的医学实体之外的医学实体，更换所述正确医学知识中的医学实体，得到新的医学知识作为错误医学知识。

可选地，所述推理规则自动发现系统还包括：

实体向量学习器，用于执行如下操作：

获取每一个所述医学实体的独热码向量：

$$\mathbf{x}_e = (a_1, \dots, a_i, \dots, a_N), \quad a_j = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases}$$

N 为所述医学知识库中的所有医学实体的个数， $\mathbf{x}_e$  为  $N \times 1$  维的向量；

根据所述每一所述医学实体的独热码向量，获取每一所述医学实体的向

量:

$$y_e = \sigma(Wx_e), \sigma(x) = 1/(1 + e^x)$$

$y_e$ 为  $K \times 1$  维的向量,  $W$  为映射矩阵,  $x_e$  为  $K \times N$  维的向量,  $K$  为预定的数值,  $K$  小于  $N$ 。

以上实施例可以由硬件、软件或者硬件和软件的结合实现。例如, 本申请实施例中描述的各种方法、步骤和功能器(或功能单元)可以由处理器实现(处理器是指广义上的处理器, 包括 CPU, 处理单元, ASIC, 逻辑单元或者可编程逻辑阵列, 等等)。本申请实施例中描述的过程、方法和功能模块可以由一个单独的处理器实现也可以分别由多个处理器实现。本申请实施例或权利要求中所述的处理器应当理解为一个或者多个处理器。此外, 以上描述的实施例可以以软件产品的方式实现。该计算机软件产品存储在非易失性存储介质中并包括一系列指令用于使得计算机设备(例如个人计算机, 服务器或者网络设备如路由器、交换机、接入点、等等)来执行本申请实施例中所述的方法。在本公开的实施例中, 所述关系矩阵学习器、推理规则发现器和实体向量学习器例如可以由上述的处理器(包括 CPU、存储器和总线等)来实现。本申请中使用的计算机可读指令由多个处理器存储在可读存储介质中, 例如硬盘、CD-ROM、DVD、光盘、软盘、磁带、RAM、ROM 或其它合适的存储设备。或者, 至少部分计算机可读指令可以由具体硬件替换, 例如, 定制集成线路、门阵列、FPGA、PLD 和具体功能的计算机等等。

可选地, 所述打分函数为:

$$\text{Score}(r(e_1, e_2)) = y_{e_1}^T M_r y_{e_2}$$

其中,  $r$  为所述医学关系,  $e_1, e_2$  为所述医学实体,  $r(e_1, e_2)$  为所述医学知识,  $\text{Score}(\cdot)$  为打分函数。

可选地, 所述预设目标函数为:

$$L = \sum_{r(e_1, e_2) \in T} \sum_{r(e'_1, e'_2) \in T'} \max\{\text{Score}(r(e'_1, e'_2)) - \text{Score}(r(e_1, e_2)) + 1, 0\}$$

其中,  $L$  为目标函数,  $T$  为所述正确医学知识集合,  $T'$  为所述错误医学知识集合,  $M_r$  为所述医学关系  $r$  的关系矩阵。

可选地，所述推理规则发现器，用于执行如下操作：从所有医学关系中，选择多个医学关系组，构建多个待验证推理规则，其中，每一个待验证推理规则对应一个医学关系组，每一个医学关系组包括三个医学关系；计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵的相似度；根据所述相似度，选择部分待验证推理规则作为最终得到的推理规则。

可选地，所述推理规则发现器，用于执行如下操作：将所有所述医学关系组对应的相似度进行排序，得到相似度最大的预设数目个待验证推理规则；将所述预设数目个待验证推理规则作为最终得到的推理规则。

可选地，每一个医学关系组中的三个医学关系满足如下条件：

- (1)  $L_p \cap H_q \neq \emptyset$ ；
- (2)  $H_p \cap H_r \neq \emptyset$ ；
- (3)  $L_q \cap L_r \neq \emptyset$ ；

其中， $H_p$ 、 $H_q$ 和 $H_r$ 分别是医学关系组中的三个医学关系 p、q 和 r 对应的三元组中的所有第一个医学实体的集合， $L_p$ 、 $L_q$ 和 $L_r$ 分别是医学关系组中的三个医学关系 p、q 和 r 对应的三元组中的所有第二个医学实体的集合。

可选地，所述推理规则发现器，用于计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵 L2 范数，将所述 L2 范数作为相似度。

本公开的实施例还提供一种医学知识库的数据检索方法，包括：获取每一个医学关系的关系矩阵，其中，所述关系矩阵反映具有所述医学关系的两个医学实体的向量之间的关系，所述医学知识库包括多个医学知识，每一个所述医学知识为两个医学实体以及所述两个医学实体的医学关系的组合；根据多个所述医学关系的关系矩阵，获取推理规则，其中，获取的每一个推理规则包括第一医学关系、第二医学关系和第三医学关系，所述推理规则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系；输入检索词或检索式进行检索；根据所述检索词或检索式基于所述推理

规则进行检索获得与该检索词或检索式相关的医学知识；以及输出与该检索词或检索式相关的医学知识。该方法的具体流程示意图如图6所示。

本公开的实施例还提供一种医学知识数据库，包括：数据输入器，用于输入医学知识数据；关系矩阵学习器，用于获取每一个所述医学关系的关系矩阵，其中，所述关系矩阵反映具有所述医学关系的医学实体的向量之间的关系，所述医学知识库包括多个医学知识，每一个所述医学知识为两个医学实体以及所述两个医学实体的医学关系的组合；推理规则发现器，用于根据多个所述医学关系的关系矩阵，获取推理规则，其中，获取的每一个推理规则中包括第一医学关系、第二医学关系和第三医学关系，所述推理规则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系；输入器，用于输入检索词或检索式；检索器，用于根据所述检索词或检索式基于所述推理规则进行检索以获得与该检索词或检索式相关的医学知识；以及输出器，用于输出与该检索词或检索式相关的医学知识。该医学知识数据库的结构框图如图5所示。

在本公开的实施例中，所述数据输入器可以为例如网络输入、USB 存储设备、光盘或其他存储设备；所述输入器可以为例如键盘、鼠标、摄像头、扫描仪、光笔、语音输入装置、手写输入板和触摸屏等；所述检索器可以为例如医学技术领域常见的检索器；所述输出器可以为例如显示器、打印机、绘图仪、影像输出系统、语音输出系统、磁记录设备等。

本公开实施例中，通过将医学关系表示成矩阵，然后通过矩阵之间的运算，从海量的医学知识中自动发现推理规则，不需要人工定义推理规则，节省了大量的人力物力，降低了时间成本；基于上述推理规则建立的医学知识数据库以及检索方法，使得检索人能够检索出更加完整的医学知识数据。

以上所述是本公开的可选实施方式。应当指出，对于本技术领域的普通技术人员来说，在不脱离本公开所述原理的前提下，还可以作出若干改进和润饰，这些改进和润饰也应视为本公开的保护范围。

## 权利要求书

1. 一种推理规则自动发现方法，其应用于医学知识库，所述医学知识库包括多个医学知识，每一个所述医学知识为两个医学实体以及所述两个医学实体的医学关系的组合，其特征在于，所述方法包括：

获取每一个所述医学关系的关系矩阵，所述关系矩阵反映具有所述医学关系的两个医学实体的向量之间的关系；

根据多个所述医学关系的关系矩阵，获取推理规则，其中，获取的每一个推理规则包括第一医学关系、第二医学关系和第三医学关系，所述推理规则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系。

2. 根据权利要求1所述的方法，其特征在于，所述获取每一个所述医学关系的关系矩阵的步骤包括：

针对每一个医学关系，构建一个初始的关系矩阵；

获取与所述医学知识库中的所述医学关系对应的医学知识作为正确医学知识；

更换所述正确医学知识中的医学实体以得到新的医学知识作为错误医学知识；

采用打分函数为所述正确医学知识和错误医学知识分别打分；

采用预设目标函数，对所述初始的关系矩阵进行学习，使得所述正确医学知识的打分高于所述错误医学知识的打分，得到所述医学关系的关系矩阵。

3. 根据权利要求2所述的方法，其特征在于，对所述初始的关系矩阵进行学习时，采用的所述正确医学知识和错误医学知识的个数差小于预设阈值。

4. 根据权利要求2所述的方法，其特征在于，所述更换所述正确医学知识中的医学实体以得到新的医学知识作为错误医学知识的步骤包括：

采用除了包含在与所述医学关系对应的所有正确医学知识中的医学实体之外的医学实体来更换所述正确医学知识中的医学实体，得到新的医学知识作为错误医学知识。

5. 根据权利要求2所述的方法，其特征在于，所述获取每一个所述医学

关系的关系矩阵的步骤之前，所述方法还包括：

获取每一个所述医学实体的独热码向量：

$$\mathbf{x}_e = (\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_N), \quad \mathbf{a}_j = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases}$$

N 为所述医学知识库中的所有医学实体的个数， $\mathbf{x}_e$  为  $N \times 1$  维的向量；

根据所述每一个所述医学实体的独热码向量，获取每一个所述医学实体的向量：

$$\mathbf{y}_e = \sigma(\mathbf{W}\mathbf{x}_e), \quad \sigma(x) = 1/(1 + e^{-x})$$

$\mathbf{y}_e$  为  $K \times 1$  维的向量，W 为映射矩阵，W 为  $K \times N$  维的向量，K 为预定的数值，K 小于 N。

6. 根据权利要求 5 所述的方法，其特征在于，所述打分函数为：

$$\text{Score}(r(e_1, e_2)) = \mathbf{y}_{e_1}^T \mathbf{M}_r \mathbf{y}_{e_2}$$

其中，r 为所述医学关系， $e_1, e_2$  为所述医学实体， $r(e_1, e_2)$  为所述医学知识，Score(·) 为打分函数， $\mathbf{M}_r$  为医学关系 r 的关系矩阵。

7. 根据权利要求 6 所述的方法，其特征在于，所述预设目标函数为：

$$L = \sum_{r(e_1, e_2) \in T} \sum_{r(e'_1, e'_2) \in T'} \max\{\text{Score}(r(e'_1, e'_2)) - \text{Score}(r(e_1, e_2)) + 1, 0\}$$

其中，L 为目标函数，T 为所述正确医学知识集合， $T'$  为所述错误医学知识集合， $\mathbf{M}_r$  为所述医学关系 r 的关系矩阵。

8. 根据权利要求 1 所述的方法，其特征在于，所述根据所述医学关系的关系矩阵，获取推理规则的步骤包括：

从所有医学关系中，选择多个医学关系组，构建多个待验证推理规则，其中，每一个待验证推理规则对应一个医学关系组，每一个医学关系组包括三个医学关系；

计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵的相似度；以及

根据所述相似度，选择部分待验证推理规则作为最终得到的推理规则。

9. 根据权利要求 8 所述的方法，其特征在于，所述根据所述相似度，选

择部分待验证推理规则作为最终得到的推理规则的步骤包括：

将所有所述医学关系组对应的相似度进行排序，得到相似度最大的预设数目个待验证推理规则；以及

将所述预设数目个待验证推理规则作为最终得到的推理规则。

10. 根据权利要求 8 或 9 所述的方法，其特征在于，每一个医学关系组中的三个医学关系满足如下条件：

$$(1) L_p \cap H_q \neq \emptyset;$$

$$(2) H_p \cap H_r \neq \emptyset;$$

$$(3) L_q \cap L_r \neq \emptyset;$$

其中， $H_p$ 、 $H_q$ 和 $H_r$ 分别是医学关系组中的三个医学关系 p、q 和 r 对应的三元组中的所有第一个医学实体的集合， $L_p$ 、 $L_q$ 和 $L_r$ 分别是医学关系组中的三个医学关系 p、q 和 r 对应的三元组中的所有第二个医学实体的集合。

11. 根据权利要求 8 或 9 所述的方法，其特征在于，所述计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵的相似度的步骤包括：

计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵 L2 范数，将所述 L2 范数作为相似度。

12. 一种推理规则自动发现系统，应用于医学知识库，所述医学知识库中包括多个医学知识，每一个所述医学知识为两个医学实体和一个医学关系的组合，其特征在于，所述系统包括：

关系矩阵学习器，用于获取每一个所述医学关系的关系矩阵，所述关系矩阵反映具有所述医学关系的医学实体的向量之间的关系；以及

推理规则发现器，用于根据多个所述医学关系的关系矩阵，获取推理规则，其中，获取的每一个推理规则中包括第一医学关系、第二医学关系和第三医学关系，所述推理规则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系。

13. 根据权利要求 12 所述的系统，其特征在于，所述关系矩阵学习器，

用于执行如下操作：针对每一个医学关系，构建一个初始的关系矩阵；获取与所述医学知识库中的所述医学关系对应的医学知识作为正确医学知识；更换所述正确医学知识中的医学实体以得到新的医学知识作为错误医学知识；采用打分函数为所述正确医学知识和错误医学知识分别打分；以及采用预设目标函数，对所述初始的关系矩阵进行学习，使得所述正确医学知识的打分高于所述错误医学知识的打分，得到所述医学关系的关系矩阵。

14. 根据权利要求 13 所述的系统，其特征在于，对所述初始的关系矩阵进行学习时，采用的所述正确医学知识和错误医学知识的个数差小于预设阈值。

15. 根据权利要求 13 所述的系统，其特征在于，所述关系矩阵学习器用于执行如下操作：采用除了包含在与所述医学关系对应的所有正确医学知识的医学实体之外的医学实体，更换所述正确医学知识中的医学实体，得到新的医学知识作为错误医学知识。

16. 根据权利要求 13 所述的系统，其特征在于，还包括：  
实体向量学习器，用于执行如下操作：  
获取每一个所述医学实体的独热码向量：

$$\mathbf{x}_e = (\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_N), \quad \mathbf{a}_j = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases}$$

N 为所述医学知识库中的所有医学实体的个数， $\mathbf{x}_e$  为  $N \times 1$  维的向量；  
根据所述每一个所述医学实体的独热码向量，获取每一所述医学实体的向量：

$$\mathbf{y}_e = \sigma(\mathbf{W}\mathbf{x}_e), \quad \sigma(\mathbf{x}) = 1/(1 + e^{-\mathbf{x}})$$

$\mathbf{y}_e$  为  $K \times 1$  维的向量，W 为映射矩阵，W 为  $K \times N$  维的向量，K 为预定的数值，K 小于 N。

17. 根据权利要求 16 所述的系统，其特征在于，所述打分函数为：

$$\text{Score}\left(r(\mathbf{e}_1, \mathbf{e}_2)\right) = \mathbf{y}_{\mathbf{e}_1}^T \mathbf{M}_r \mathbf{y}_{\mathbf{e}_2}$$

其中，r 为所述医学关系， $\mathbf{e}_1$ ， $\mathbf{e}_2$  为所述医学实体， $r(\mathbf{e}_1, \mathbf{e}_2)$  为所述医学知识，Score( $\cdot$ ) 为打分函数， $\mathbf{M}_r$  为所述医学关系 r 的关系矩阵。

18. 根据权利要求 17 所述的系统，其特征在于，所述预设目标函数为：

$$L = \sum_{r(e_1, e_2) \in T} \sum_{r(e_1, e_2) \in T'} \max\{\text{Score}(r(e_1', e_2')) - \text{Score}(r(e_1, e_2)) + 1, 0\}$$

其中，L 为目标函数，T 为所述正确医学知识集合，T' 为所述错误医学知识集合， $M_r$  为所述医学关系 r 的关系矩阵。

19. 根据权利要求 12 所述的系统，其特征在于，所述推理规则发现器，用于执行如下操作：从所有医学关系中，选择多个医学关系组，构建多个待验证推理规则，其中，每一个待验证推理规则对应一个医学关系组，每一个医学关系组包括三个医学关系；计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵的相似度；以及根据所述相似度，选择部分待验证推理规则作为最终得到的推理规则。

20. 根据权利要求 19 所述的系统，其特征在于，所述推理规则发现器，用于执行如下操作：将所有所述医学关系组对应的相似度进行排序，得到相似度最大的预设数目个待验证推理规则；以及将所述预设数目个待验证推理规则作为最终得到的推理规则。

21. 根据权利要求 19 或 20 所述的系统，其特征在于，每一个医学关系组中的三个医学关系满足如下条件：

- (1)  $L_p \cap H_q \neq \emptyset$ ；
- (2)  $H_p \cap H_r \neq \emptyset$ ；
- (3)  $L_q \cap L_r \neq \emptyset$ ；

其中， $H_p$ 、 $H_q$  和  $H_r$  分别是医学关系组中的三个医学关系 p、q 和 r 对应的三元组中的所有第一个医学实体的集合， $L_p$ 、 $L_q$  和  $L_r$  分别是医学关系组中的三个医学关系 p、q 和 r 对应的三元组中的所有第二个医学实体的集合。

22. 根据权利要求 19 或 20 所述的系统，其特征在于，所述推理规则发现器，用于执行如下操作：计算所述医学关系组中的两个医学关系的关系矩阵的乘积与另一个医学关系的关系矩阵 L2 范数，将所述 L2 范数作为相似度。

23. 一种医学知识库的数据检索方法，包括：获取每一个医学关系的关

系矩阵，其中，所述关系矩阵反映具有所述医学关系的两个医学实体的向量之间的关系，所述医学知识库包括多个医学知识，每一个所述医学知识为两个医学实体以及所述两个医学实体的医学关系的组合；

根据多个所述医学关系的关系矩阵，获取推理规则，其中，获取的每一个推理规则包括第一医学关系、第二医学关系和第三医学关系，所述推理规则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系；

输入检索词或检索式进行检索；

根据所述检索词或检索式基于所述推理规则进行检索获得与该检索词或检索式相关的医学知识；以及

输出与该检索词或检索式相关的医学知识。

24. 根据权利要求23所述的检索方法，其特征在于，所述输出与该检索词或检索式相关的医学知识包括显示与该检索词或检索式相关的医学知识。

25. 一种医学知识数据库，包括：

数据输入器，用于输入医学知识数据；

关系矩阵学习器，用于获取每一个所述医学关系的关系矩阵，其中，所述关系矩阵反映具有所述医学关系的医学实体的向量之间的关系，所述医学知识库包括多个医学知识，每一个所述医学知识为两个医学实体以及所述两个医学实体的医学关系的组合；

推理规则发现器，用于根据多个所述医学关系的关系矩阵，获取推理规则，其中，获取的每一个推理规则中包括第一医学关系、第二医学关系和第三医学关系，所述推理规则用于表示由所述第一医学关系和所述第二医学关系能够推理出所述第三医学关系；

检索词或检索式输入器，用于输入检索词或检索式；

检索器，用于根据所述检索词或检索式基于所述推理规则进行检索以获得与该检索词或检索式相关的医学知识；以及

输出器，用于输出与该检索词或检索式相关的医学知识。

26. 根据权利要求25所述的数据库，其特征在于，所述输出器用于显示与该检索词或检索式相关的医学知识。

27. 根据权利要求 25 所述的数据库，其特征在于，还包括：

实体向量学习器，用于执行如下操作：

获取每一个所述医学实体的独热码向量：

$$\mathbf{x}_e = (a_1, \dots, a_i, \dots, a_N), \quad a_j = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases}$$

$N$  为所述医学知识库中的所有医学实体的个数， $\mathbf{x}_e$  为  $N \times 1$  维的向量；

根据所述每一个所述医学实体的独热码向量，获取每一所述医学实体的向量：

$$\mathbf{y}_e = \sigma(W\mathbf{x}_e), \quad \sigma(x) = 1/(1 + e^{-x})$$

$\mathbf{y}_e$  为  $K \times 1$  维的向量， $W$  为映射矩阵， $W$  为  $K \times N$  维的向量， $K$  为预定的数值， $K$  小于  $N$ 。

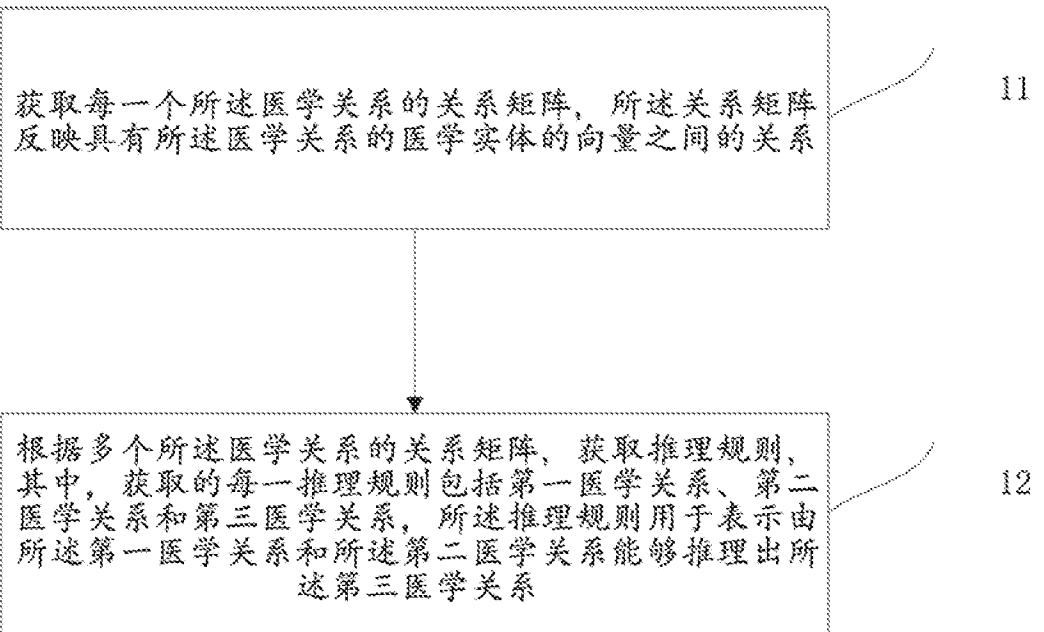


图 1

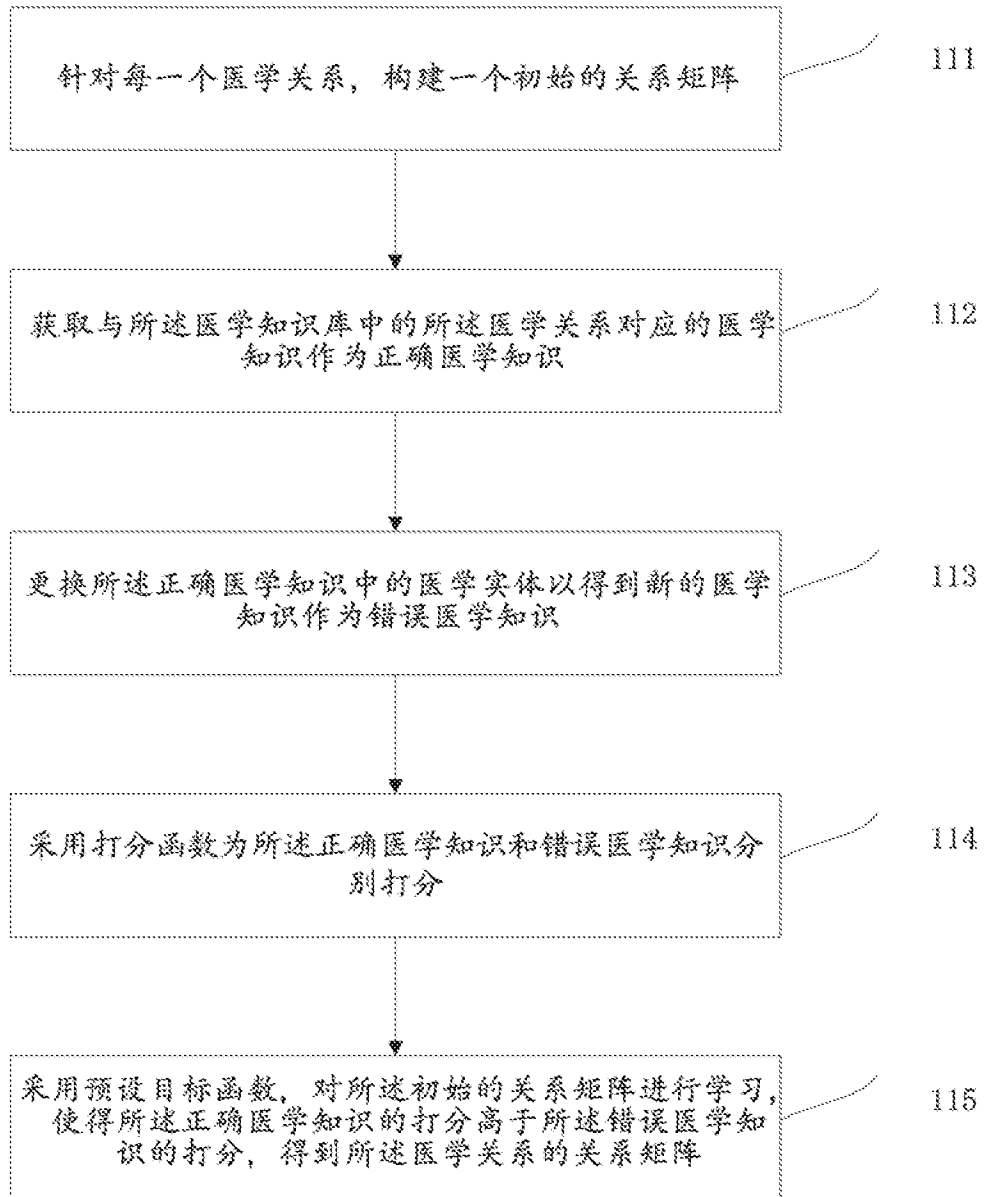


图 2

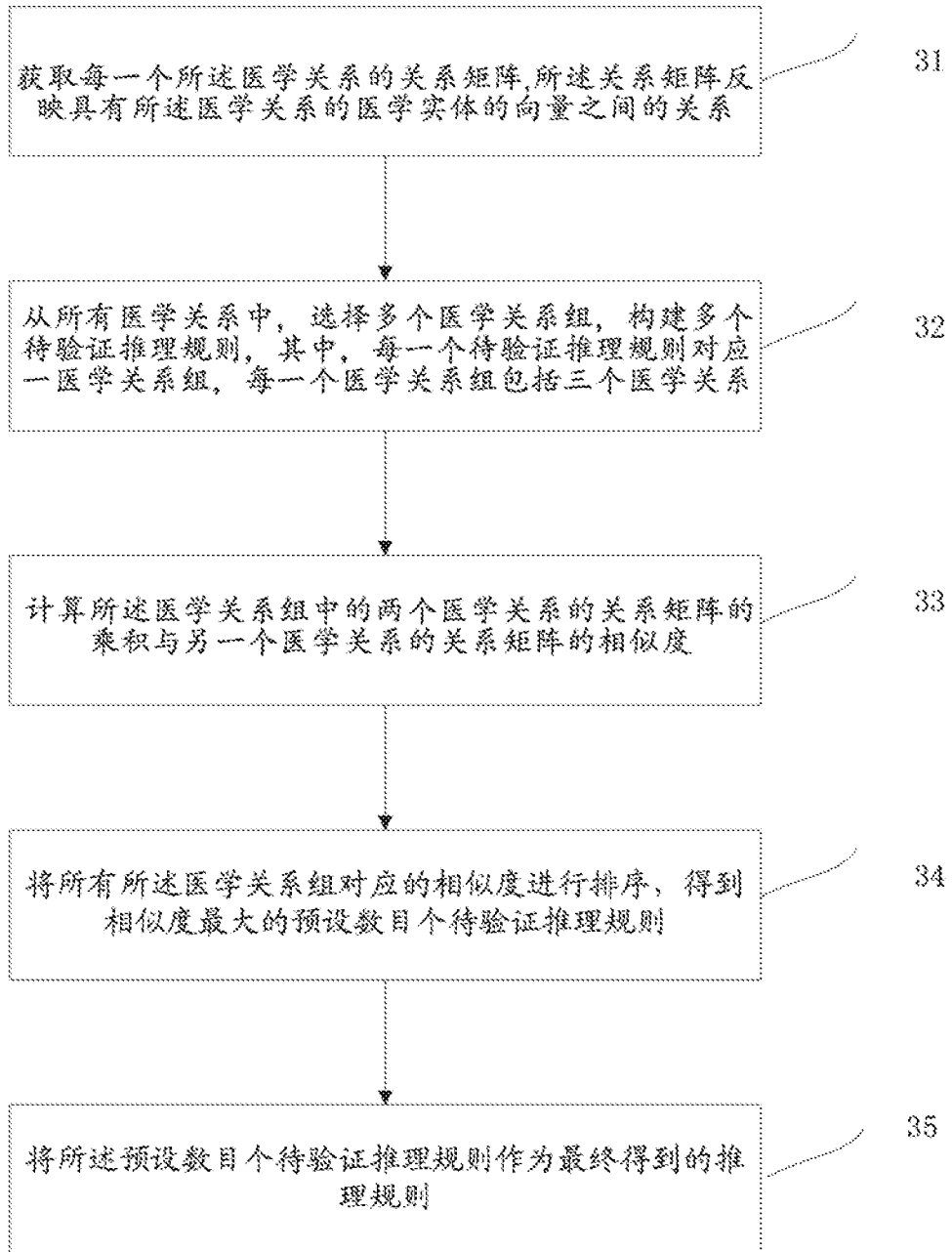


图 3

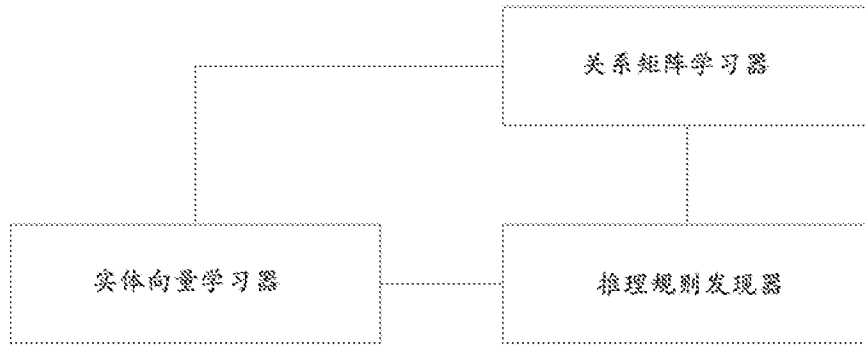


图 4

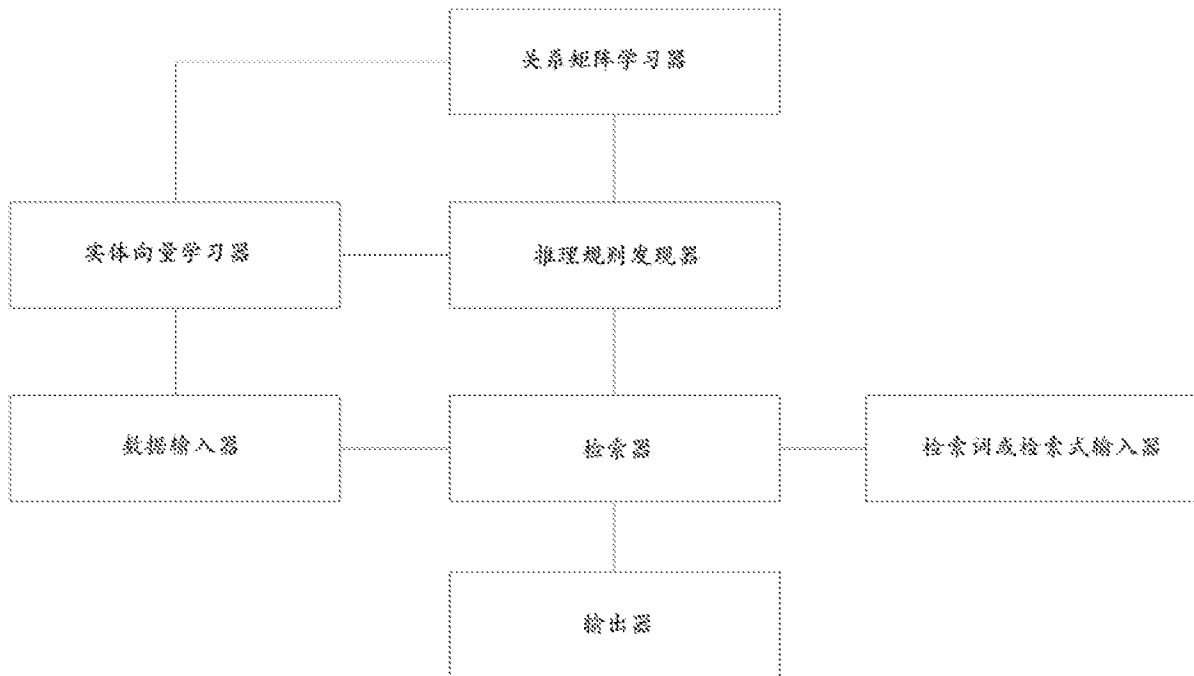


图 5

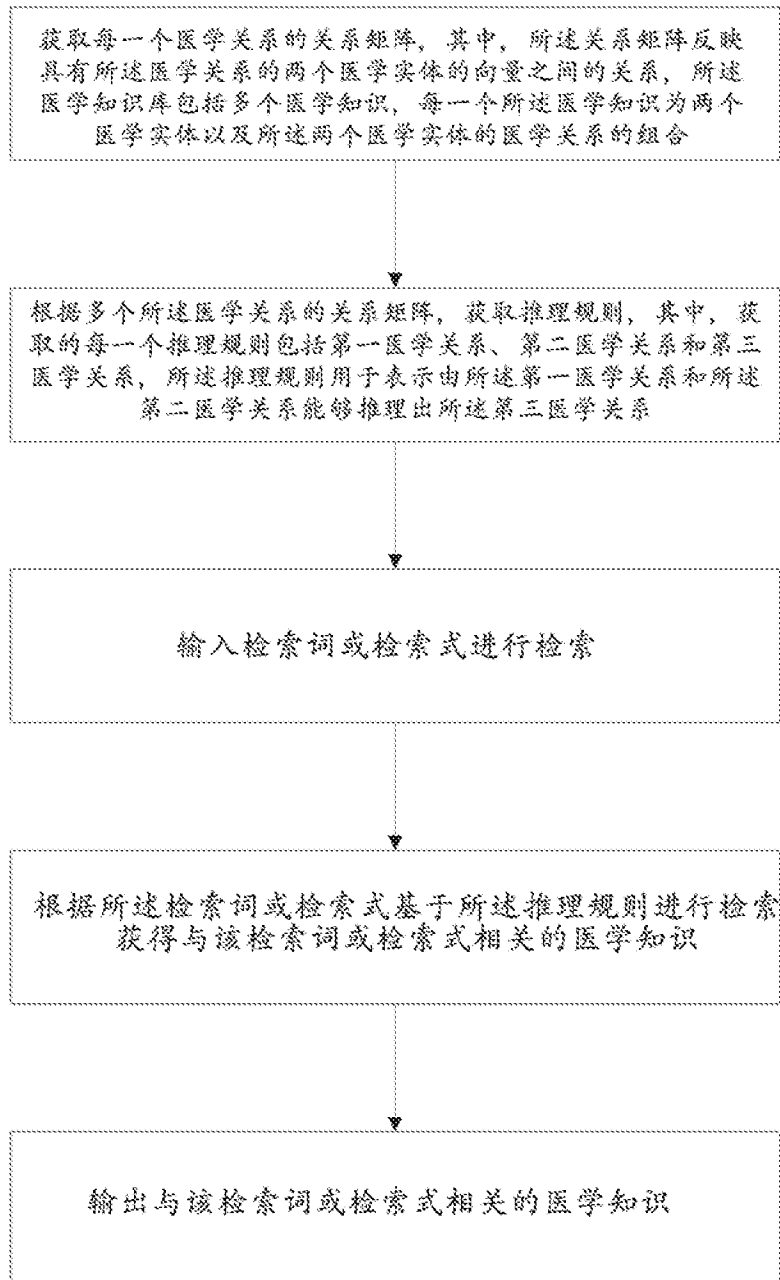


图 6

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/CN2018/073004

## A. CLASSIFICATION OF SUBJECT MATTER

G06F 19/00 (2018.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC, IEEE: 京东方, 张振中, 推理, 推测, 推导, 推断, 估测, 估计, 预测, 医学, 医疗, 医药, 关联, 关系, 矩阵, 向量, 矢量, 知识, 自动, 动态, 智能, 自适应, 图谱, inferenc+, reason???, estima+, knowledge, medical, iatrology, relat+, matrix, vector, auto

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 106874695 A (BEIJING DASHU YIDA TECHNOLOGY CO., LTD.), 20 June 2017 (20.06.2017), description, paragraphs [0026]-[0042]	1-27
A	CN 103824115 A (INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES), 28 May 2014 (28.05.2014), entire document	1-27
A	CN 106528609 A (XIAMEN UNIVERSITY OF TECHNOLOGY), 22 March 2017 (22.03.2017), entire document	1-27
A	CN 105718726 A (SHENYANG UNIVERSITY OF TECHNOLOGY), 29 June 2016 (29.06.2016), entire document	1-27
A	CN 104239385 A (INTERNATIONAL BUSINESS MACHINES CORPORATION), 24 December 2014 (24.12.2014), entire document	1-27
A	US 2014229161 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION), 14 August 2014 (14.08.2014), entire document	1-27

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>“&amp;” document member of the same patent family</p>
---	---

<p>Date of the actual completion of the international search</p> <p style="text-align: center;">08 April 2018</p>	<p>Date of mailing of the international search report</p> <p style="text-align: center;">20 April 2018</p>
<p>Name and mailing address of the ISA</p> <p>State Intellectual Property Office of the P. R. China</p> <p>No. 6, Xitucheng Road, Jimenqiao</p> <p>Haidian District, Beijing 100088, China</p> <p>Facsimile No. (86-10) 62019451</p>	<p>Authorized officer</p> <p style="text-align: center;">LIU, Shen</p> <p>Telephone No. (86-10) 010-53961537</p>

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.  
PCT/CN2018/073004

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 106874695 A	20 June 2017	None	
CN 103824115 A	28 May 2014	None	
CN 106528609 A	22 March 2017	None	
CN 105718726 A	29 June 2016	None	
CN 104239385 A	24 December 2014	US 2014365504 A1	11 December 2014
		US 2014365503 A1	11 December 2014
		US 9542503 B2	10 January 2017
		US 9483580 B2	01 November 2016
US 2014229161 A1	14 August 2014	US 9020810 B2	28 April 2015
		US 2014229163 A1	14 August 2014
		WO 2014126657 A1	21 August 2014
		US 9135240 B2	15 September 2015

<p><b>A. 主题的分类</b> G06F 19/00 (2018.01) i</p> <p>按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类</p>																							
<p><b>B. 检索领域</b></p> <p>检索的最低限度文献 (标明分类系统和分类号) G06F</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNPAT, CNKI, WPI, EPODOC, IEEE: 京东方, 张振中, 推理, 推测, 推导, 推断, 估测, 估计, 预测, 医学, 医疗, 医药, 关联, 关系, 矩阵, 向量, 矢量, 知识, 自动, 动态, 智能, 自适应, 图谱, inferenc+, reason???, estima+, knowledge, medical, iatrology, relat+, matrix, vector, auto</p>																							
<p><b>C. 相关文件</b></p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 106874695 A (北京大数医疗科技有限公司) 2017年 6月 20日 (2017 - 06 - 20) 说明书第[0026]-[0042]段</td> <td>1-27</td> </tr> <tr> <td>A</td> <td>CN 103824115 A (中国科学院计算技术研究所) 2014年 5月 28日 (2014 - 05 - 28) 全文</td> <td>1-27</td> </tr> <tr> <td>A</td> <td>CN 106528609 A (厦门理工学院) 2017年 3月 22日 (2017 - 03 - 22) 全文</td> <td>1-27</td> </tr> <tr> <td>A</td> <td>CN 105718726 A (沈阳工业大学) 2016年 6月 29日 (2016 - 06 - 29) 全文</td> <td>1-27</td> </tr> <tr> <td>A</td> <td>CN 104239385 A (国际商业机器公司) 2014年 12月 24日 (2014 - 12 - 24) 全文</td> <td>1-27</td> </tr> <tr> <td>A</td> <td>US 2014229161 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2014年 8月 14日 (2014 - 08 - 14) 全文</td> <td>1-27</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 106874695 A (北京大数医疗科技有限公司) 2017年 6月 20日 (2017 - 06 - 20) 说明书第[0026]-[0042]段	1-27	A	CN 103824115 A (中国科学院计算技术研究所) 2014年 5月 28日 (2014 - 05 - 28) 全文	1-27	A	CN 106528609 A (厦门理工学院) 2017年 3月 22日 (2017 - 03 - 22) 全文	1-27	A	CN 105718726 A (沈阳工业大学) 2016年 6月 29日 (2016 - 06 - 29) 全文	1-27	A	CN 104239385 A (国际商业机器公司) 2014年 12月 24日 (2014 - 12 - 24) 全文	1-27	A	US 2014229161 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2014年 8月 14日 (2014 - 08 - 14) 全文	1-27
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
A	CN 106874695 A (北京大数医疗科技有限公司) 2017年 6月 20日 (2017 - 06 - 20) 说明书第[0026]-[0042]段	1-27																					
A	CN 103824115 A (中国科学院计算技术研究所) 2014年 5月 28日 (2014 - 05 - 28) 全文	1-27																					
A	CN 106528609 A (厦门理工学院) 2017年 3月 22日 (2017 - 03 - 22) 全文	1-27																					
A	CN 105718726 A (沈阳工业大学) 2016年 6月 29日 (2016 - 06 - 29) 全文	1-27																					
A	CN 104239385 A (国际商业机器公司) 2014年 12月 24日 (2014 - 12 - 24) 全文	1-27																					
A	US 2014229161 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2014年 8月 14日 (2014 - 08 - 14) 全文	1-27																					
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p>																							
<p>* 引用文件的具体类型:                      “A” 认为不特别相关的表示了现有技术一般状态的文件                      “E” 在国际申请日的当天或之后公布的在先申请或专利                      “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)                      “O” 涉及口头公开、使用、展览或其他方式公开的文件                      “P” 公布日先于国际申请日但迟于所要求的优先权日的文件                      “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件                      “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性                      “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性                      “&amp;” 同族专利的文件</p>																							
国际检索实际完成的日期	国际检索报告邮寄日期																						
2018年 4月 8日	2018年 4月 20日																						
ISA/CN的名称和邮寄地址	受权官员																						
中华人民共和国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	刘申																						
传真号 (86-10) 62019451	电话号码 (86-10) 010-53961537																						

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2018/073004

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	106874695	A	2017年 6月 20日	无			
CN	103824115	A	2014年 5月 28日	无			
CN	106528609	A	2017年 3月 22日	无			
CN	105718726	A	2016年 6月 29日	无			
CN	104239385	A	2014年 12月 24日	US	2014365504	A1	2014年 12月 11日
				US	2014365503	A1	2014年 12月 11日
				US	9542503	B2	2017年 1月 10日
				US	9483580	B2	2016年 11月 1日
US	2014229161	A1	2014年 8月 14日	US	9020810	B2	2015年 4月 28日
				US	2014229163	A1	2014年 8月 14日
				WO	2014126657	A1	2014年 8月 21日
				US	9135240	B2	2015年 9月 15日