



(12) 发明专利

(10) 授权公告号 CN 113961724 B

(45) 授权公告日 2022. 04. 22

(21) 申请号 202111575407.3

(51) Int.Cl.

(22) 申请日 2021.12.22

G06F 16/36 (2019.01)

G06K 9/62 (2022.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 113961724 A

审查员 单娟

(43) 申请公布日 2022.01.21

(73) 专利权人 山东新希望六和集团有限公司

地址 266100 山东省青岛市崂山区九水东

路592-26号3号楼4楼

专利权人 四川新希望动物营养科技有限公  
司

新希望六和股份有限公司

(72) 发明人 杨帆 刘旭 马文彦 蒋智权

(74) 专利代理机构 广州三环专利商标代理有限  
公司 44202

代理人 郭浩辉 许羽冬

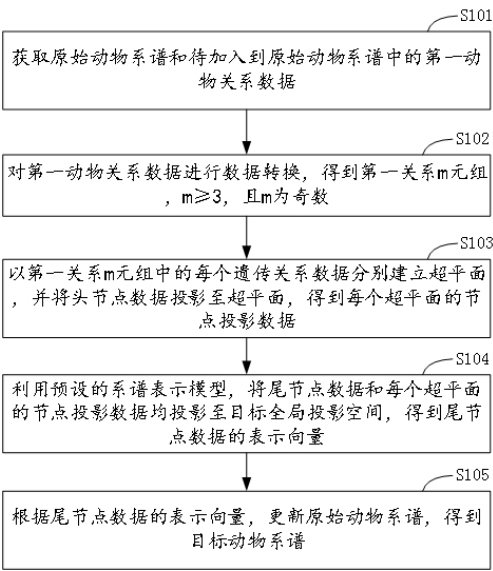
权利要求书3页 说明书11页 附图4页

(54) 发明名称

动物系谱的更新方法、装置及计算机设备

(57) 摘要

本申请公开了一种动物系谱的更新方法、装置及计算机设备,通过获取原始动物系谱和待加入到原始动物系谱中的第一动物关系数据,对第一动物关系数据进行数据转换,得到第一关系m元组;以第一关系m元组中的每个遗传关系数据分别建立超平面,并将头节点数据投影至超平面,得到每个超平面的节点投影数据;利用预设的系谱表示模型,将尾节点数据和每个超平面的节点投影数据均投影至目标全局投影空间,得到尾节点数据的表示向量;最后根据尾节点数据的表示向量,更新原始动物系谱,得到目标动物系谱,从而只需计算局部系谱关系,而无需重新计算所有系谱关系,提高了运算效率。



1. 一种动物系谱的更新方法,其特征在于,包括:

获取原始动物系谱和待加入到所述原始动物系谱中的第一动物关系数据,所述第一动物关系数据包括多个头节点数据、一个尾节点数据以及每个所述头节点数据与所述尾节点数据之间的遗传关系数据;

对所述第一动物关系数据进行数据转换,得到第一关系 $m$ 元组, $m \geq 3$ ,且 $m$ 为奇数;

以所述第一关系 $m$ 元组中的每个所述遗传关系数据分别建立超平面,并将所述头节点数据投影至所述超平面,得到每个所述超平面的节点投影数据;

利用预设的系谱表示模型,将所述尾节点数据和每个所述超平面的节点投影数据均投影至目标全局投影空间,得到所述尾节点数据的表示向量;

根据所述尾节点数据的表示向量,更新所述原始动物系谱,得到目标动物系谱。

2. 如权利要求1所述的动物系谱的更新方法,其特征在于,所述对所述第一动物关系数据进行数据转换,得到第一关系 $m$ 元组,包括:

根据所述头节点数据的数量,确定多元组数据转换模板;

基于所述多元组数据转换模板,对所述第一动物关系数据进行数据转换,得到第一关系 $m$ 元组。

3. 如权利要求2所述的动物系谱的更新方法,其特征在于,所述头节点数据的数量为2,则所述第一关系 $m$ 元组为 $(H_1, D_{r1}, H_2, D_{r2}, S)$ ,其中 $H_1$ 为第1个所述头节点数据, $H_2$ 为第2个所述头节点数据, $S$ 为所述尾节点数据, $D_{r1}$ 为头节点数据 $H_1$ 与所述尾节点数据 $S$ 之间的遗传关系数据, $D_{r2}$ 为头节点数据 $H_2$ 与所述尾节点数据 $S$ 之间的遗传关系数据。

4. 如权利要求1所述的动物系谱的更新方法,其特征在于,所述以所述第一关系 $m$ 元组中的每个所述遗传关系数据分别建立超平面,并将所述头节点数据投影至所述超平面,得到每个所述超平面的节点投影数据,包括:

以所述第一关系 $m$ 元组中的每个所述遗传关系数据分别建立超平面,得到超平面集合 $W_r = (W_{r1}, W_{r2}, \dots, W_{rn})$ ;

将所述头节点数据投影至所述超平面 $W_{rn}$ ,得到所述超平面 $W_{rn}$ 的头节点投影数据 $H_{n_1} = H_n - W_{rn}^T H_n W_{rn}$ , $H_n$ 为所述第一关系 $m$ 元组中的第 $n$ 个头节点数据, $W_{rn}^T$ 为 $H_n$ 与所述尾节点数据 $S$ 之间的遗传关系数据构建的超平面 $W_{rn}$ 的转置矩阵。

5. 如权利要求1所述的动物系谱的更新方法,其特征在于,所述利用预设的系谱表示模型,将所述尾节点数据和每个所述超平面的节点投影数据均投影至目标全局投影空间,得到所述尾节点数据的表示向量,包括:

利用所述系谱表示模型,确定每个所述超平面的投影关系 $d_{rn}$ 和所述目标全局投影空间的投影矩阵 $M$ ,以及所述尾节点数据在所述目标全局投影空间的尾节点投影数据 $S \times M$ ;

根据所述投影关系和所述投影矩阵,确定所述尾节点数据的表示向量,所述表示向量为:

$$V = S \times M = (H_{1_1} + d_{r1}) \times M + (H_{2_1} + d_{r2}) \times M + \dots + (H_{n_1} + d_{rn}) \times M;$$

$H_{n_1}$  为头节点投影数据。

6. 如权利要求1所述的动物系谱的更新方法,其特征在于,所述利用预设的系谱表示模型,将所述尾节点数据和每个所述超平面的节点投影数据均投影至目标全局投影空间,得到所述尾节点数据的表示向量之前,还包括:

获取多组第二动物关系数据;

对每组所述第二动物关系数据进行数据转换,得到多组第二关系m元组;

利用多组所述第二关系m元组,对预设的知识图谱模型进行训练,直至所述知识图谱模型的损失函数达到预设收敛条件,停止训练,得到所述系谱表示模型,其中所述损失函数为:

$$Loss = \max(0, Difscore);$$

$$Difscore = margin + score(H_n', D_{rn}', S') - score(H_n'', D_{rn}'', S'');$$

其中  $Loss$  为损失函数值,  $margin$  为超参数,  $score(H_n', D_{rn}', S')$  为正样本打分函数,  $score(H_n'', D_{rn}'', S'')$  为负样本打分函数,  $H_n'$  为头节点正样本数据,  $D_{rn}'$  为遗传关系正样本数据,  $S'$  为尾节点正样本数据,  $H_n''$  为头节点负样本数据,  $D_{rn}''$  为遗传关系负样本数据,  $S''$  为尾节点负样本数据。

7. 如权利要求6所述的动物系谱的更新方法,其特征在于,所述正样本打分函数为:

$$\begin{aligned} score(H_n', D_{rn}', S') \\ = \|(H_{1_1}' + D_{r1_1}') \times M \\ + (H_{2_1}' + D_{r2_1}') \times M + \dots \\ + (H_{n_1}' + D_{rn_1}') \times M - S' \times M\|_2^2 \end{aligned};$$

其中  $H_n'$  为头节点正样本数据,  $D_{rn}'$  为遗传关系正样本数据,  $S'$  为尾节点正样本数据,  $d_{rn}'$  为基于  $D_{rn}'$  建立的超平面上的投影关系,  $M$  为所述目标全局投影空间的投影矩阵,  $H_{n_1}'$  为  $H_n'$  在基于  $D_{rn}'$  建立的超平面上的投影数据,  $S' \times M$  为  $S'$  在目标全局投影空间上的投影数据。

8. 一种动物系谱的更新装置,其特征在于,包括:

获取模块,用于获取原始动物系谱和待加入到所述原始动物系谱中的第一动物关系数据,所述第一动物关系数据包括多个头节点数据、一个尾节点数据以及每个所述头节点数据与所述尾节点数据之间的遗传关系数据;

转换模块,用于对所述第一动物关系数据进行数据转换,得到第一关系m元组,  $m \geq 3$ , 且  $m$  为奇数;

建立模块,用于以所述第一关系m元组中的每个所述遗传关系数据分别建立超平面,并将所述头节点数据投影至所述超平面,得到每个所述超平面的节点投影数据;

投影模块,用于利用预设的系谱表示模型,将所述尾节点数据和每个所述超平面的节点投影数据均投影至目标全局投影空间,得到所述尾节点数据的表示向量;

更新模块,用于根据所述尾节点数据的表示向量,更新所述原始动物系谱,得到目标动

物系谱。

9. 如权利要求8所述的动物系谱的更新装置,其特征在于,所述转换模块,包括:

第一确定单元,用于根据所述头节点数据的数量,确定多元组数据转换模板;

转换单元,用于基于所述多元组数据转换模板,对所述第一动物关系数据进行数据转换,得到第一关系 $m$ 元组。

10. 一种计算机设备,其特征在于,包括处理器和存储器,所述存储器用于存储计算机程序,所述计算机程序被所述处理器执行时实现如权利要求1至7任一项所述的动物系谱的更新方法。

## 动物系谱的更新方法、装置及计算机设备

### 技术领域

[0001] 本申请涉及动物系谱管理领域,尤其涉及一种动物系谱的更新方法、装置及计算机设备。

### 背景技术

[0002] 在动物育种过程中,为了改良动物种群的整体水平,需要分析动物种群的系谱关系,实现育种估算和亲子鉴定。因此,对于新生的动物幼崽,需要将幼崽加入到原动物系谱中,以更新系谱关系。

[0003] 目前,动物系谱的更新基本采用Blupf90模型实现。Blupf90模型属于单机模型,其在生成动物系谱或者更新动物系谱时,涉及大量迭代运算和拟合运算,导致运算效率极低。例如,对于新生的种猪,需要将种猪更新到动物系谱中,Blupf90模型需要对动物系谱中的系谱关系重新进行迭代运算和拟合运算,以得到种猪的表示向量,从而更新动物系谱。

### 发明内容

[0004] 本申请提供了一种动物系谱的更新方法、装置及计算机设备,以解决当前动物系谱的更新方法存在运算效率低的技术问题。

[0005] 为了解决上述技术问题,第一方面,本申请实施例提供了一种动物系谱的更新方法,包括:

[0006] 获取原始动物系谱和待加入到原始动物系谱中的第一动物关系数据,第一动物关系数据包括多个头节点数据、一个尾节点数据以及每个头节点数据与尾节点数据之间的遗传关系数据;

[0007] 对第一动物关系数据进行数据转换,得到第一关系 $m$ 元组, $m \geq 3$ ,且 $m$ 为奇数;

[0008] 以第一关系 $m$ 元组中的每个遗传关系数据分别建立超平面,并将头节点数据投影至超平面,得到每个超平面的节点投影数据;

[0009] 利用预设的系谱表示模型,将尾节点数据和每个超平面的节点投影数据均投影至目标全局投影空间,得到尾节点数据的表示向量;

[0010] 根据尾节点数据的表示向量,更新原始动物系谱,得到目标动物系谱。

[0011] 本实施例通过获取原始动物系谱和待加入到原始动物系谱中的第一动物关系数据,对第一动物关系数据进行数据转换,得到第一关系 $m$ 元组,以采用关系多元组描述尾节点与多个头节点之间的遗传关系,从而解决传统三元组描述节点之间遗传关系存在关系表示冲突的问题;再以第一关系 $m$ 元组中的每个遗传关系数据分别建立超平面,并将头节点数据投影至超平面,得到每个超平面的节点投影数据,从而分解出每个头节点之间的线性空间位置,以便于后续采用智能模型进行数据分析;然后利用预设的系谱表示模型,将尾节点数据和每个超平面的节点投影数据均投影至目标全局投影空间,得到尾节点数据的表示向量,以将多个超平面和尾节点数据在同一个维度上进行运算,从而利用多个超平面内的投影数据综合分析尾节点与多个头节点之间的关系向量,实现利用在原始动物系谱中的多个

头节点表示尾节点在原始动物系谱的表示向量；最后根据尾节点数据的表示向量，更新原始动物系谱，得到目标动物系谱，从而只需计算局部系谱关系，而无需重新计算所有系谱关系，提高了运算效率。

[0012] 在一实施例中，对第一动物关系数据进行数据转换，得到第一关系 $m$ 元组，包括：

[0013] 根据头节点数据的数量，确定多元组数据转换模板；

[0014] 基于多元组数据转换模板，对第一动物关系数据进行数据转换，得到第一关系 $m$ 元组。

[0015] 本实施例通过头节点数据的数量，以确定需要将第一动物数据转换几元组，从而能够适用于多个头节点数据时的多元组描述。

[0016] 优选地，头节点数据的数量为2，则第一关系 $m$ 元组为 $(H_1, D_{r1}, H_2, D_{r2}, S)$ ，其中 $H_1$ 为第1个头节点数据， $H_2$ 为第2个头节点数据， $S$ 为尾节点数据， $D_{r1}$ 为头节点数据 $H_1$ 与尾节点数据 $S$ 之间的遗传关系数据， $D_{r2}$ 为头节点数据 $H_2$ 与尾节点数据 $S$ 之间的遗传关系数据。

[0017] 本实施例采用五元组表示2个头节点数据、1个尾节点数据以及头节点数据分别与尾节点数据的遗传关系，解决了传统知识图谱采用三元组描述节点之间遗传关系存在关系表示冲突的问题，提高了遗传关系的表示准确性。

[0018] 在一实施例中，以第一关系 $m$ 元组中的每个遗传关系数据分别建立超平面，并将头节点数据投影至超平面，得到每个超平面的投影数据，包括：

[0019] 以第一关系 $m$ 元组中的每个遗传关系数据分别建立超平面，得到超平面集合 $W_r = (W_{r1}, W_{r2}, \dots, W_{rm})$ ；

[0020] 将头节点数据投影至超平面 $W_{rn}$ ，得到超平面 $W_{rn}$ 的头节点投影数据 $H_{n_1} = H_n - W_{rn}^T H_n W_{rn}$ ， $H_n$ 为第一关系 $m$ 元组中的第 $n$ 个头节点数据， $W_{rn}^T$ 为 $H_n$ 与尾节点数据 $S$ 之间的遗传关系数据构建的超平面 $W_{rn}$ 的转置矩阵。

[0021] 本实施例通过以各个遗传关系分别建立超平面，并将头节点投影到对应的超平面，从而分解出各个头节点的线性空间位置，实现数据的线性分析。

[0022] 在一实施例中，利用预设的系谱表示模型，将尾节点数据和每个超平面的节点投影数据均投影至目标全局投影空间，得到总投影数据，包括：

[0023] 利用系谱表示模型，确定每个超平面的投影关系 $d_{rn}$ 和目标全局投影空间的投影矩阵 $M$ ，以及尾节点数据在目标全局投影空间的尾节点投影数据 $S \times M$ ；

[0024] 根据投影关系和投影矩阵，确定尾节点数据的表示向量，表示向量为：

[0025]  $V = S \times M = (H_{1_1} + d_{r1}) \times M + (H_{2_1} + d_{r2}) \times M + \dots + (H_{n_1} + d_{rn}) \times M$ 。

[0026] 本实施例通过系谱表示模型学习到目标全局投影空间的投影矩阵，并分析出每个超平面的投影关系，以将多个超平面投影至同一个全局空间，从而在全局空间内实现对超平面的投影数据的线性运算，实现采用多个头节点在原始动物系谱中的表示向量综合表示尾节点在原始动物系谱的表示向量，进而便于后续无需重新对原始动物系谱进行运算，而直接将尾节点加入到原始动物系谱中。

[0027] 在一实施例中，利用预设的系谱表示模型，将尾节点数据和每个超平面的节点投

影数据均投影至目标全局投影空间,得到尾节点数据的表示向量之前,还包括:

[0028] 获取多组第二动物关系数据;

[0029] 对每组第二动物关系数据进行数据转换,得到多组第二关系m元组;

[0030] 利用多组第二关系m元组,对预设的知识图谱模型进行训练,直至知识图谱模型的损失函数达到预设收敛条件,停止训练,得到系谱表示模型,其中损失函数为:

[0031]  $Loss = \max(0, Difscore)$ ;

[0032]  $Difscore = margin + score(H_n', D_{rn}', S') - score(H_n'', D_{rn}'', S'')$ ;

[0033] 其中  $Loss$  为损失函数值,  $margin$  为超参数,  $score(H_n', D_{rn}', S')$  为正样本打分函数,  $score(H_n'', D_{rn}'', S'')$  为负样本打分函数。

[0034] 本实施例通过包括正样本和负样本的多组第二动物关系数据,对知识图谱模型进行训练,从而使模型学习到节点数据之间的关系特征,进而能够在应用节点实现尾节点的向量表示。

[0035] 在一实施例中,正样本打分函数为:

$$\begin{aligned}
 & score(H_n', D_{rn}', S') \\
 & = \|(H_{1\perp}' + D_{r1\perp}') \times M \\
 [0036] \quad & + (H_{2\perp}' + D_{r2\perp}') \times M + \dots \\
 & + (H_{n\perp}' + D_{rn\perp}') \times M - S' \times M\|_2^2
 \end{aligned}$$

[0037] 其中  $H_n'$  为头节点正样本数据,  $D_{rn}'$  为遗传关系正样本数据,  $S'$  为子节点正样本数据,  $d_{rn}'$  为基于  $D_{rn}'$  建立的超平面上的投影关系,  $M$  为目标全局投影空间的投影矩阵,  $H_{n\perp}'$  为  $H_n'$  在基于  $D_{rn}'$  建立的超平面上的投影数据,  $S' \times M$  为  $S'$  在目标全局投影空间上的投影数据。

[0038] 第二方面,本申请实施例提供一种动物系谱的更新装置,包括:

[0039] 获取模块,用于获取原始动物系谱和待加入到原始动物系谱中的第一动物关系数据,第一动物关系数据包括多个头节点数据、一个尾节点数据以及每个头节点数据与尾节点数据之间的遗传关系数据;

[0040] 转换模块,用于对第一动物关系数据进行数据转换,得到第一关系m元组,  $m \geq 3$ , 且  $m$  为奇数;

[0041] 建立模块,用于以第一关系m元组中的每个遗传关系数据分别建立超平面,并将头节点数据投影至超平面,得到每个超平面的节点投影数据;

[0042] 投影模块,用于利用预设的系谱表示模型,将尾节点数据和每个超平面的节点投影数据均投影至目标全局投影空间,得到尾节点数据的表示向量;

[0043] 更新模块,用于根据尾节点数据的表示向量,更新原始动物系谱,得到目标动物系谱。

[0044] 第三方面,本申请实施例提供一种计算机设备,包括处理器和存储器,存储器用于存储计算机程序,计算机程序被处理器执行时实现如第一方面任一项的动物系谱的更新方法。

[0045] 需要说明的是,上述第二方面至第三方面的有益效果参见第一方面的相关描述,在此不再赘述。

### 附图说明

[0046] 图1为本申请实施例提供的动物系谱的更新方法的流程示意图;

[0047] 图2为本申请实施例提供的三元组在超平面的投影示意图;

[0048] 图3为本申请实施例提供的父头实体在超平面的投影示意图;

[0049] 图4为本申请实施例提供的母头实体在超平面的投影示意图;

[0050] 图5为本申请实施例提供的尾节点数据和2个超平面的节点投影数据在目标全局投影空间的投影示意图;

[0051] 图6为本申请实施例提供的动物系谱的更新装置的结构示意图;

[0052] 图7为本申请实施例提供的计算机设备的结构示意图。

### 具体实施方式

[0053] 下面将结合本申请实施例中的附图,对本申请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申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0054] 如相关技术记载,目前,动物系谱的更新基本采用Blupf90模型实现。Blupf90模型属于单机模型,其在生成动物系谱或者更新动物系谱时,涉及大量迭代运算和拟合运算,导致运算效率极低。例如,对于新生的种猪,需要将种猪更新到动物系谱中,Blupf90模型需要对动物系谱中的系谱关系重新进行迭代运算和拟合运算,以得到种猪的表示向量,从而更新动物系谱。

[0055] 由于新生幼崽与其父母分别有对应的遗传关系,即父子关系和母子关系,而采用关系四元组表示三者之间的遗传关系,无法表示父子关系和母子,其仅仅只能表示父母子关系,例如采用四元组表示为(父亲、母亲、幼崽、遗传关系),则该遗传关系无法表示到底是父子关系还是母子关系。

[0056] 为此,本申请实施例提供一种动物系谱的更新方法、装置及计算机设备,通过获取原始动物系谱和待加入到原始动物系谱中的第一动物关系数据,对第一动物关系数据进行数据转换,得到第一关系 $m$ 元组, $m \geq 3$ ,且 $m$ 为奇数,以采用关系多元组描述尾节点分别与每个头节点之间的遗传关系,从而解决传统三元组描述节点之间遗传关系存在关系表示冲突的问题,以及解决采用四元组无法明确尾节点与每个头节点之间的遗传关系的问题;再以第一关系 $m$ 元组中的每个遗传关系数据分别建立超平面,并将头节点数据投影至超平面,得到每个超平面的节点投影数据,从而分解出每个头节点之间的线性空间位置,以便于后续采用智能模型进行数据分析;然后利用预设的系谱表示模型,将尾节点数据和每个超平面的节点投影数据均投影至目标全局投影空间,得到尾节点数据的表示向量,以将多个超平面和尾节点数据在同一个维度上进行运算,从而利用多个超平面内的投影数据综合分析尾节点与多个头节点之间的关系向量,实现利用在原始动物系谱中的多个头节点表示尾节点在原始动物系谱的表示向量;最后根据尾节点数据的表示向量,更新原始动物系谱,得到目



标动物系谱,从而只需计算局部系谱关系,而无需重新计算所有系谱关系,提高了运算效率。

[0057] 参照图1,图1为本申请实施例提供的一种动物系谱的更新方法的流程示意图。本申请实施例的动物系谱的更新方法可应用于计算机设备,该计算机设备包括但不限于智能手机、平板电脑、笔记本电脑、桌上型计算机、物理服务器和云服务器等计算设备。如图1所示,动物系谱的更新方法包括步骤S101至S105,详述如下:

[0058] 步骤S101,获取原始动物系谱和待加入到所述原始动物系谱中的第一动物关系数据,所述第一动物关系数据包括多个头节点数据、一个尾节点数据以及每个所述头节点数据与所述尾节点数据之间的遗传关系数据。

[0059] 在本步骤中,头节点数据为上级节点数据,如父节点数据和母节点数据等,例如猪群中的公猪和母猪,尾节点数据为下级节点数据,如子节点数据,例如猪群中的猪仔。其中公猪与猪仔之间的遗传关系数据为父子关系,母猪与猪仔之间的遗传关系数据为母子关系。

[0060] 可以理解的是,原始动物系谱记录有多个头节点,尾节点为待加入到原始动物系谱的节点,以更新原始动物系谱。例如,公猪A和母猪B繁育出猪仔C,则需要将猪仔C加入到原始动物系谱中。

[0061] 需要说明的是,原始动物系谱采用数据向量表示各个节点,所以将尾节点数据加入到原始动物系谱之前,需要计算尾节点的数据向量,以往的方式是需要对所有遗传关系进行迭代和拟合运算,以得到尾节点向量。

[0062] 步骤S102,对所述第一动物关系数据进行数据转换,得到第一关系 $m$ 元组, $m \geq 3$ ,且 $m$ 为奇数。

[0063] 在本步骤中,传统的知识图谱采用三元组(子节点数据,父节点数据,母节点数据)进行数据表示,但是三元组的表示方式存在表示冲突问题。例如,(猪仔ID,公猪ID,母猪ID),当计算猪仔ID对应的数据向量时,可以采用公猪ID的数据向量进行计算,也可以采用母猪ID的数据向量进行计算,所以当基于公猪ID的数据向量进行计算得到猪仔向量 $X$ ,与基于母猪ID的数据向量进行计算得到的猪仔向量 $Y$ 不同时,子节点的数据向量就会存在表示冲突问题,从而存在无法加入到原始动物系谱或者加入错误的情况。本实施例通过将第一动物关系数据转换为第一关系 $m$ 元组,以采用多元组描述节点数据,从而便于后续基于多个头节点和遗传关系共同计算尾节点的数据向量,避免尾节点的数据向量表示冲突的问题。

[0064] 可选地,通过预先设置多个 $m$ 元组模板,再将第一动物关系数据中的各个数据代入到预设的 $m$ 元组模板,以实现数据转换,得到第一关系 $m$ 元组。

[0065] 步骤S103,以所述第一关系 $m$ 元组中的每个所述遗传关系数据分别建立超平面,并将所述头节点数据投影至所述超平面,得到每个所述超平面的节点投影数据。

[0066] 在本步骤中,将遗传关系抽象为一个向量空间中的超平面,将头节点数据映射对应的超平面上,得到每个头节点数据在对应超平面上的投影数据。

[0067] 示例性地,对于第一关系 $m$ 元组为三元组(头节点数据 $H$ ,遗传关系数据 $D_r$ ,尾节点数据 $S$ ),如图2示出的三元组在超平面的投影示意图。将遗传关系数据 $D_r$ 建立一个超平面 $W_r$ ,以及定义该超平面的平移向量(即投影关系) $d_r$ ,平移向量为未知值,需要通过系谱

表示模型进行推理。由于只有一个超平面,所以无需再进行空间投影,而是将头节点数据 $H$ 和尾节点数据 $S$ 投影到该超平面上,得到节点投影数据 $H_1 = H - W_r^T H W_r$ 和 $S_1 = S - W_r^T S W_r$ 。以及该三元组的打分函数为 $score(H, D_r, S) = \|H_1 + d_r - S_1\|_2^2$ 。

[0068] 可选地,对于第一关系 $m$ 元组为五元组或以上,则将每个遗传关系数据建立一个超平面,以及定义该超平面的平移向量,将每个头节点数据投影至对应的超平面。

[0069] 步骤S104,利用预设的系谱表示模型,将所述尾节点数据和每个所述超平面的节点投影数据均投影至目标全局投影空间,得到所述尾节点数据的表示向量。

[0070] 在本实施例中,由于存在多个超平面,不同超平面的投影数据无法类似四元组构建超平面的方式直接进行运算,所以本实施例定义所有超平面共同的目标全局投影空间,以在同一个向量空间内根据投影数据的空间几何关系线性函数,以实现多元组的数据分析。本实施例将尾节点数据和多个超平面的节点投影数据输入到系谱表示模型,通过系谱表示模型确定超平面的平移向量和目标全局投影空间的投影矩阵,完成投影数据之间的数据分析。

[0071] 步骤S105,根据所述尾节点数据的表示向量,更新所述原始动物系谱,得到目标动物系谱。

[0072] 在本实施例中,尾节点数据的表示向量根据在原始动物系谱中的多个头节点数据的数据向量共同表示,使得尾节点数据能够直接加入到原始动物系谱,从而更新原始动物系谱,得到目标动物系谱。

[0073] 在一实施例中,在图1所示实施例的基础上,步骤S102,包括:

[0074] 根据所述头节点数据的数量,确定多元组数据转换模板;

[0075] 基于所述多元组数据转换模板,对所述第一动物关系数据进行数据转换,得到第一关系 $m$ 元组。

[0076] 在本实施例中,多元组数据转换模板包括但不限于三元组数据转换模板、五元组数据转换模板和七元组数据转换模板等。可以理解的是,第一动物关系数据又多个头节点数据、一个尾节点数据和对应数量的遗传关系数据。示例性地,头节点数据的数量为1,则第一动物关系数据包括一个头节点数据、一个尾节点数据和一个遗传关系数据,即多元组数据转换模板为三元组模板;头节点数据的数量为2,则第一动物关系数据包括两个头节点数据、一个尾节点数据和两个遗传关系数据,即多元组数据转换模板为五元组模板。

[0077] 可选地,将第一动物关系数据代入到多元组数据转换模板,得到第一关系 $m$ 元组。

[0078] 可选地,头节点数据的数量为1,则所述第一关系 $m$ 元组为 $(H, D_r, S)$ ,  $H$ 为头节点数据,  $S$ 为尾节点数据,  $D_r$ 为头节点数据与尾节点数据 $S$ 之间的遗传关系数据。

[0079] 优选地,头节点数据的数量为2,则所述第一关系 $m$ 元组为 $(H_1, D_{r1}, H_2, D_{r2}, S)$ , 其中 $H_1$ 为第1个所述头节点数据,  $H_2$ 为第2个所述头节点数据,  $S$ 为所述尾节点数据,  $D_{r1}$ 为头节点数据 $H_1$ 与所述尾节点数据 $S$ 之间的遗传关系数据,  $D_{r2}$ 为头节点数据 $H_2$ 与所述尾节点数据 $S$ 之间的遗传关系数据。

[0080] 在一实施例中,在图1所示实施例的基础上,步骤S103,包括:

[0081] 以所述第一关系 $m$ 元组中的每个所述遗传关系数据分别建立超平面,得到超平面

集合  $W_r = (W_{r1}, W_{r2}, \dots, W_{rn})$ ;

[0082] 将所述头节点数据投影至所述超平面  $W_{rn}$ , 得到所述超平面  $W_{rn}$  的头节点投影数据  $H_{n1} = H_n - W_{rn}^T H_n W_{rn}$ ,  $H_n$  为所述第一关系  $m$  元组中的第  $n$  个头节点数据,  $W_{rn}^T$  为  $H_n$  与所述尾节点数据  $S$  之间的遗传关系数据构建的超平面  $W_{rn}$  的转置矩阵。

[0083] 在本实施例中, 每个超平面对应又一个平移向量  $d_{rn}$ , 用于表示头节点数据在超平面上的投影关系, 其为待确定向量。示例性地, 头节点数据的数量为2, 包括父头节点数据和母头节点数据, 如图3示出的父头实体在超平面的投影示意图, 父头节点投影数据为  $H_{11} = H_1 - W_{r1}^T H_1 W_{r1}$ , 该父头节点投影数据的投影平移向量为  $H1 = H_{11} + d_{r1}$ 。如图4示出的母头实体在超平面的投影示意图, 母节点投影数据为  $H_{21} = H_2 - W_{r2}^T H_2 W_{r2}$ , 该母头节点投影数据的投影平移向量为  $H2 = H_{21} + d_{r2}$ 。

[0084] 在一实施例中, 在图1所示实施例的基础上, 步骤S104, 包括:

[0085] 利用所述系谱表示模型, 确定每个所述超平面的投影关系  $d_{rn}$  和所述目标全局投影空间的投影矩阵  $M$ , 以及所述尾节点数据在所述目标全局投影空间的尾节点投影数据  $S \times M$ ;

[0086] 根据所述投影关系和所述投影矩阵, 确定所述尾节点数据的表示向量, 所述表示向量为:

[0087]  $V = S \times M = (H_{11} + d_{r1}) \times M + (H_{21} + d_{r2}) \times M + \dots + (H_{n1} + d_{rn}) \times M$ 。

[0088] 在本实施例中, 示例性地, 如图5示出的尾节点数据和2个超平面的节点投影数据在目标全局投影空间的投影示意图。通过确定投影矩阵  $M$ , 以使超平面的节点投影数据在目标全局投影空间存在加性关系, 且投影平移向量与尾节点投影数据满足:

[0089]  $(H_{11} + d_{r1}) \times M + (H_{21} + d_{r2}) \times M + \dots + (H_{n1} + d_{rn}) \times M \approx S \times M$ ;

[0090] 故可以将尾节点数据的表示向量确定为:

[0091]  $V = S \times M = (H_{11} + d_{r1}) \times M + (H_{21} + d_{r2}) \times M + \dots + (H_{n1} + d_{rn}) \times M$ 。

[0092] 在一实施例中, 在图1所示实施例的基础上, 步骤S102之前, 还包括:

[0093] 获取多组第二动物关系数据;

[0094] 对每组所述第二动物关系数据进行数据转换, 得到多组第二关系  $m$  元组;

[0095] 利用多组所述第二关系  $m$  元组, 对预设的知识图谱模型进行训练, 直至所述知识图谱模型的损失函数达到预设收敛条件, 停止训练, 得到所述系谱表示模型, 其中所述损失函数为:

[0096]  $Loss = \max(0, Difscore)$ ;

[0097]  $Difscore = margin + score(H_n', D_{rn}', S') - score(H_n'', D_{rn}'', S'')$ ;

[0098] 其中  $Loss$  为损失函数值,  $margin$  为超参数,  $score(H_n', D_{rn}', S')$  为正样本打分函数,  $score(H_n'', D_{rn}'', S'')$  为负样本打分函数。

[0099] 在本实施例中, 知识图谱模型包括但不限于 transE、transR 和 transH 等。可以理解的是, 本申请的知识图谱模型是针对多元组进行改进后的模型。第二关系  $m$  元组包括正样本

和负样本,其中负样本可以通过对正样本对应的多元组中的元素进行随机替换得到。例如,第二关系m元组为五元组,正样本对应的五元组为(公猪A,父子关系,母猪B,母子关系,猪仔C),通过随机替换,得到负样本对应的五元组为(公猪D,父子关系,母猪B,母子关系,猪仔C)、(公猪A,父子关系,母猪E,母子关系,猪仔C)、(公猪A,父子关系,母猪B,母子关系,猪仔F)或者(公猪A,母子关系,母猪B,母子关系,猪仔C)等,在此不再赘述。

[0100] 可选地,预设收敛条件可以为知识图谱模型的损失函数小于预设阈值,也可以为模型迭代次数达到预设次数。

[0101] 可选地,所述正样本打分函数为:

$$\begin{aligned} & score(H_n', D_{rn}', S') \\ & = \|(H_{1\perp}' + D_{r1\perp}') \times M \\ [0102] & + (H_{2\perp}' + D_{r2\perp}') \times M + \dots ; \\ & + (H_{n\perp}' + D_{rn\perp}') \times M - S' \times M\|_2^2 \end{aligned}$$

[0103] 其中 $H_n'$ 为头节点正样本数据, $D_{rn}'$ 为遗传关系正样本数据, $S'$ 为子节点正样本数据, $d_{rn}'$ 为基于 $D_{rn}'$ 建立的超平面上的投影关系, $M$ 为所述目标全局投影空间的投影矩阵, $H_{n\perp}'$ 为 $H_n'$ 在基于 $D_{rn}'$ 建立的超平面上的投影数据, $S' \times M$ 为 $S'$ 在目标全局投影空间上的投影数据。

[0104] 可选地,负样本打分函数为:

$$\begin{aligned} & score(H_n'', D_{rn}'', S'') \\ & = \|(H_{1\perp}'' + d_{r1}'') \times M \\ [0105] & + (H_{2\perp}'' + d_{r2}'') \times M + \dots ; \\ & + (H_{n\perp}'' + d_{rn}'') \times M - S'' \times M\|_2^2 \end{aligned}$$

[0106] 其中 $H_n''$ 为头节点负样本数据, $D_{rn}''$ 为遗传关系负样本数据, $S''$ 为子节点负样本数据, $d_{rn}''$ 为基于 $D_{rn}''$ 建立的超平面上的投影关系, $M$ 为所述目标全局投影空间的投影矩阵, $H_{n\perp}''$ 为 $H_n''$ 在基于 $D_{rn}''$ 建立的超平面上的投影数据, $S_{\perp}''$ 为 $S''$ 在目标全局投影空间上的投影数据。

[0107] 为了执行上述方法实施例对应的动物系谱的更新方法,以实现相应的功能和技术效果。参见图6,图6示出了本申请实施例提供的一种动物系谱的更新装置的结构框图。为了便于说明,仅示出了与本实施例相关的部分,本申请实施例提供的动物系谱的更新装置,包括:

[0108] 获取模块601,用于获取原始动物系谱和待加入到所述原始动物系谱中的第一动物关系数据,所述第一动物关系数据包括多个头节点数据、一个尾节点数据以及每个所述头节点数据与所述尾节点数据之间的遗传关系数据;

[0109] 转换模块602,用于对所述第一动物关系数据进行数据转换,得到第一关系m元组,

$m \geq 3$ , 且  $m$  为奇数;

[0110] 建立模块603, 用于以所述第一关系  $m$  元组中的每个所述遗传关系数据分别建立超平面, 并将所述头节点数据投影至所述超平面, 得到每个所述超平面的节点投影数据;

[0111] 投影模块604, 用于利用预设的系谱表示模型, 将尾节点数据和每个所述超平面的节点投影数据均投影至目标全局投影空间, 得到所述尾节点数据的表示向量;

[0112] 更新模块605, 用于根据所述尾节点数据的表示向量, 更新所述原始动物系谱, 得到目标动物系谱。

[0113] 在一实施例中, 转换模块602, 包括:

[0114] 第一确定单元, 用于根据头节点数据的数量, 确定多元组数据转换模板;

[0115] 转换单元, 用于基于多元组数据转换模板, 对第一动物关系数据进行数据转换, 得到第一关系  $m$  元组。

[0116] 优选地, 头节点数据的数量为2, 则所述第一关系  $m$  元组为  $(H_1, D_{r1}, H_2, D_{r2}, S)$ , 其中  $H_1$  为第1个所述头节点数据,  $H_2$  为第2个所述头节点数据,  $S$  为所述尾节点数据,  $D_{r1}$  为头节点数据  $H_1$  与所述尾节点数据  $S$  之间的遗传关系数据,  $D_{r2}$  为头节点数据  $H_2$  与所述尾节点数据  $S$  之间的遗传关系数据。

[0117] 在一实施例中, 建立模块603, 包括:

[0118] 建立单元, 用于以所述第一关系  $m$  元组中的每个所述遗传关系数据分别建立超平面, 得到超平面集合  $W_r = (W_{r1}, W_{r2}, \dots, W_{rn})$ ;

[0119] 投影单元, 用于将所述头节点数据投影至所述超平面  $W_{rn}$ , 得到所述超平面  $W_{rn}$  的头节点投影数据  $H_{n_1} = H_n - W_{rn}^T H_n W_{rn}$ ,  $H_n$  为所述第一关系  $m$  元组中的第  $n$  个头节点数据,  $W_{rn}^T$  为  $H_n$  与所述尾节点数据  $S$  之间的遗传关系数据构建的超平面  $W_{rn}$  的转置矩阵。

[0120] 在一实施例中, 投影模块604, 包括:

[0121] 第二确定单元, 用于利用所述系谱表示模型, 确定每个所述超平面的投影关系  $d_{rn}$  和所述目标全局投影空间的投影矩阵  $M$ , 以及所述尾节点数据在所述目标全局投影空间的尾节点投影数据  $S \times M$ ;

[0122] 第三确定单元, 用于根据所述投影关系和所述投影矩阵, 确定所述尾节点数据的表示向量, 所述表示向量为:

[0123]  $V = S \times M = (H_{1_1} + d_{r1}) \times M + (H_{2_1} + d_{r2}) \times M + \dots + (H_{n_1} + d_{rn}) \times M$ 。

[0124] 在一实施例中, 上述更新装置, 还包括:

[0125] 第二获取模块, 用于获取多组第二动物关系数据;

[0126] 第二转换模块, 用于对每组第二动物关系数据进行数据转换, 得到多组第二关系  $m$  元组;

[0127] 训练模块, 用于利用多组第二关系  $m$  元组, 对预设的知识图谱模型进行训练, 直至知识图谱模型的损失函数达到预设收敛条件, 停止训练, 得到系谱表示模型, 其中损失函数为:

[0128]  $Loss = \max(0, Difscore);$

[0129]  $Difscore = margin + score(H_n', D_{rn}', S') - score(H_n'', D_{rn}'', S'');$

[0130] 其中  $Loss$  为损失函数值,  $margin$  为超参数,  $score(H_n', D_{rn}', S')$  为正样本打分函数,  $score(H_n'', D_{rn}'', S'')$  为负样本打分函数。

[0131] 在一实施例中, 正样本打分函数为:

$$score(H_n', D_{rn}', S') = \|(H_{1\perp}' + D_{r1\perp}') \times M + (H_{2\perp}' + D_{r2\perp}') \times M + \dots + (H_{n\perp}' + D_{rn\perp}') \times M - S' \times M\|_2^2$$

[0132]

[0133] 其中  $H_n'$  为头节点正样本数据,  $D_{rn}'$  为遗传关系正样本数据,  $S'$  为子节点正样本数据,  $d_{rn}'$  为基于  $D_{rn}'$  建立的超平面上的投影关系,  $M$  为所述目标全局投影空间的投影矩阵,  $H_{n\perp}'$  为  $H_n'$  在基于  $D_{rn}'$  建立的超平面上的投影数据,  $S' \times M$  为  $S'$  在目标全局投影空间上的投影数据。

[0134] 上述的动物系谱的更新装置可实施上述方法实施例的动物系谱的更新方法。上述方法实施例中的可选项也适用于本实施例, 这里不再详述。本申请实施例的其余内容可参照上述方法实施例的内容, 在本实施例中, 不再进行赘述。

[0135] 图7为本申请一实施例提供的计算机设备的结构示意图。如图7所示, 该实施例的计算机设备7包括: 至少一个处理器70(图7中仅示出一个)处理器、存储器71以及存储在所述存储器71中并可在所述至少一个处理器70上运行的计算机程序72, 所述处理器70执行所述计算机程序72时实现上述任意方法实施例中的步骤。

[0136] 所述计算机设备7可以是智能手机、平板电脑、桌上型计算机和云端服务器等计算设备。该计算机设备可包括但不限于处理器70、存储器71。本领域技术人员可以理解, 图7仅仅是计算机设备7的举例, 并不构成对计算机设备7的限定, 可以包括比图示更多或更少的部件, 或者组合某些部件, 或者不同的部件, 例如还可以包括输入输出设备、网络接入设备等。

[0137] 所称处理器70可以是中央处理单元 (Central Processing Unit, CPU), 该处理器70还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现成可编程门阵列 (Field-Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0138] 所述存储器71在一些实施例中可以是所述计算机设备7的内部存储单元, 例如计算机设备7的硬盘或内存。所述存储器71在另一些实施例中也可以是所述计算机设备7的外部存储设备, 例如所述计算机设备7上配备的插接式硬盘, 智能存储卡 (Smart Media Card, SMC), 安全数字 (Secure Digital, SD) 卡, 闪存卡 (Flash Card) 等。进一步地, 所述存储器71还可以既包括所述计算机设备7的内部存储单元也包括外部存储设备。所述存储器71用

于存储操作系统、应用程序、引导装载程序(BootLoader)、数据以及其他程序等,例如所述计算机程序的程序代码等。所述存储器71还可以用于暂时地存储已经输出或者将要输出的数据。

[0139] 本申请实施例提供了一种计算机程序产品,当计算机程序产品在计算机设备上运行时,使得计算机设备执行时实现上述各个方法实施例中的步骤。

[0140] 在本申请所提供的几个实施例中,可以理解的是,流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分,所述模块、程序段或代码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。也应当注意的是,在有些作为替换的实现方式中,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。

[0141] 所述功能如果以软件功能模块的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0142] 以上所述的具体实施例,对本申请的目的、技术方案和有益效果进行了进一步的详细说明,应当理解,以上所述仅为本申请的具体实施例而已,并不用于限定本申请的保护范围。特别指出,对于本领域技术人员来说,凡在本申请的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

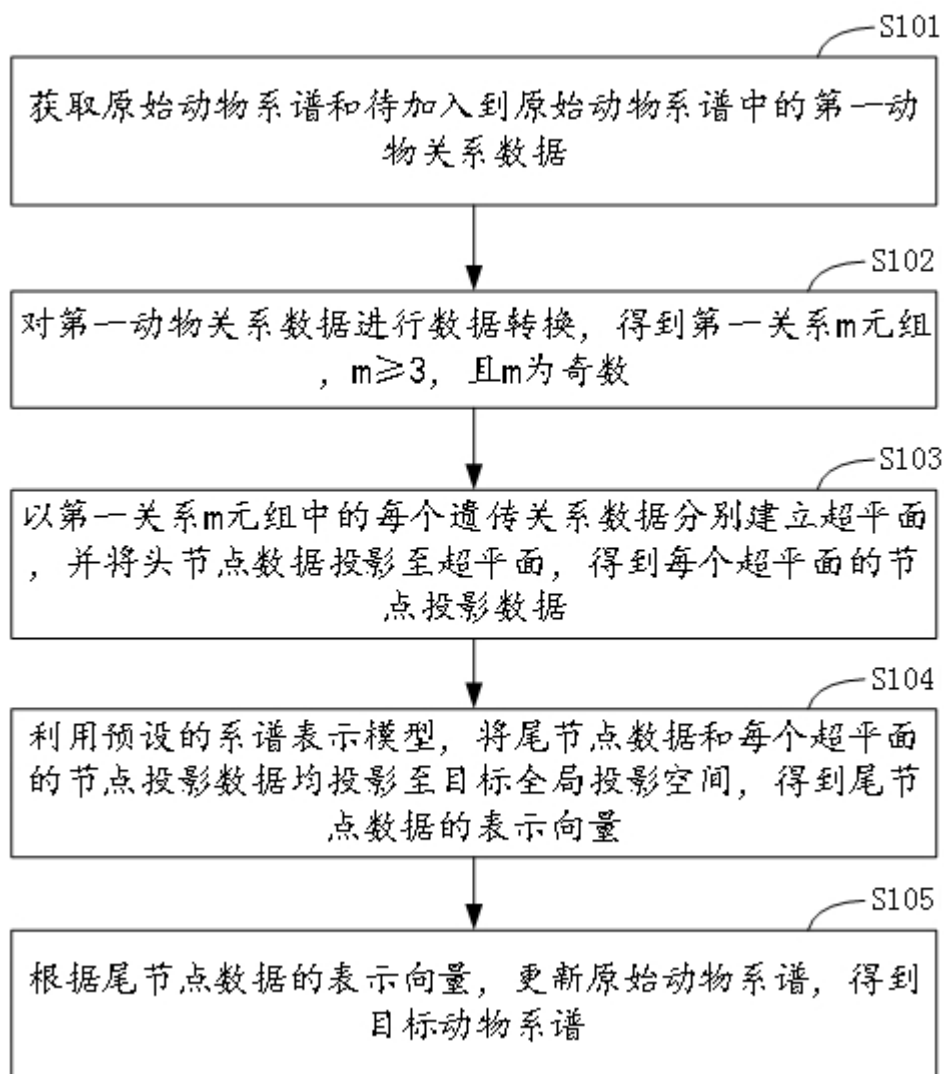


图1

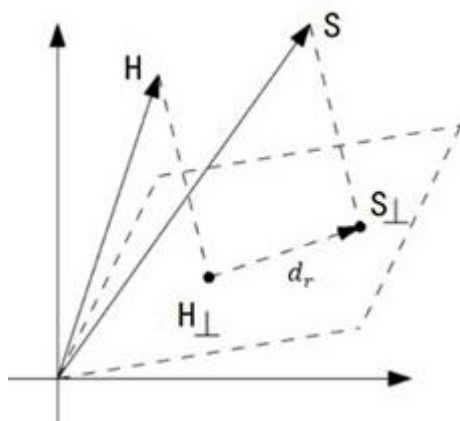


图2



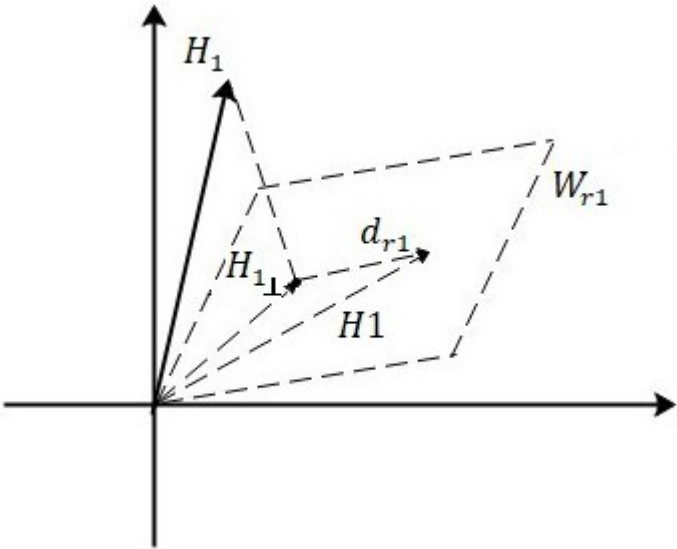


图3

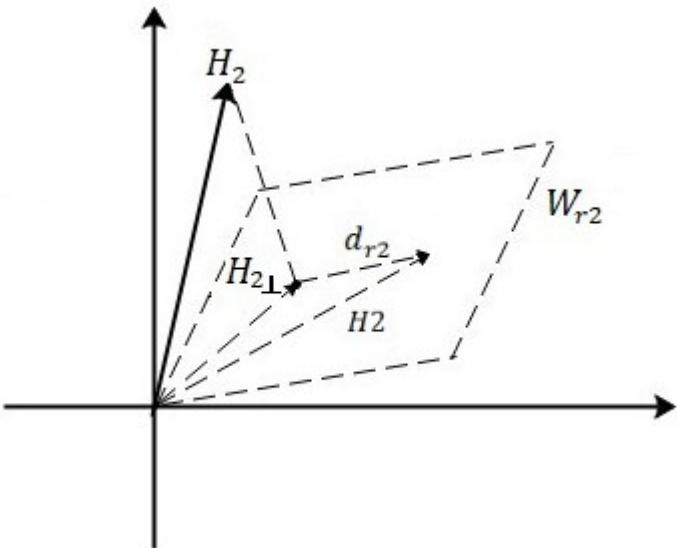


图4

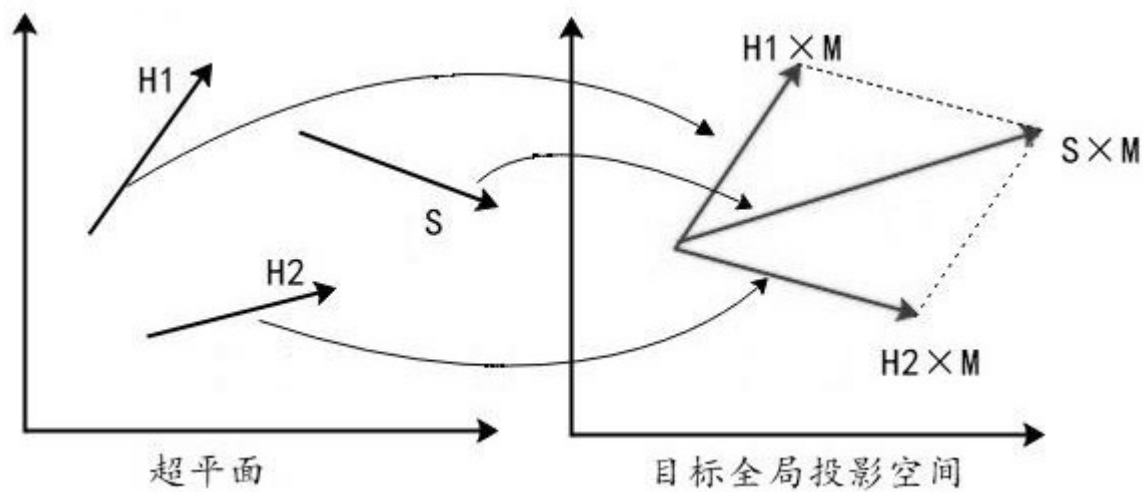


图5

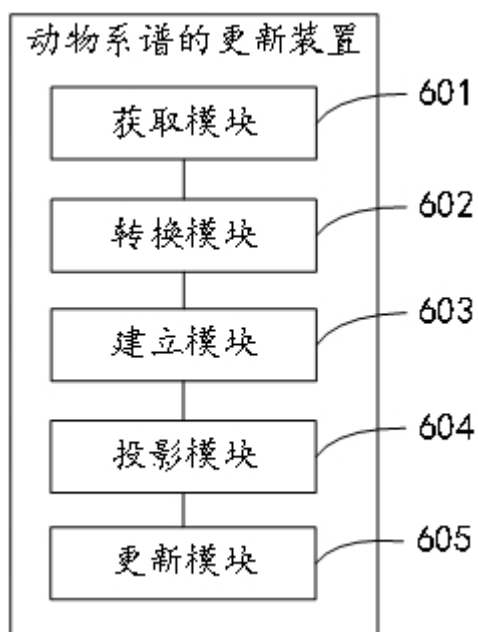


图6

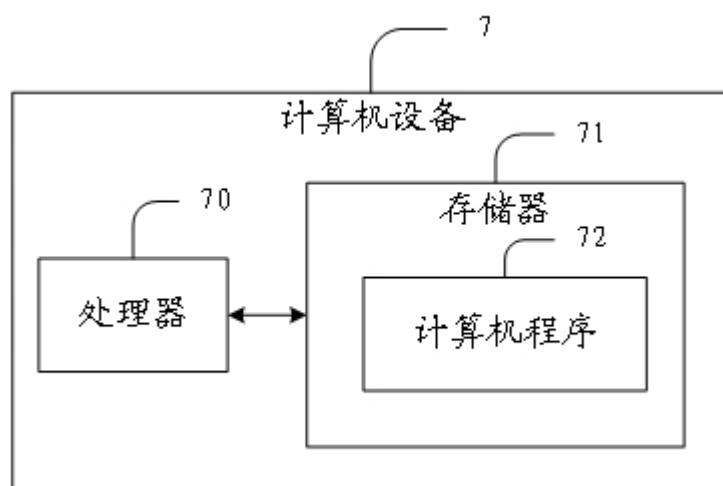


图7