



(12) 发明专利

(10) 授权公告号 CN 113946695 B

(45) 授权公告日 2022.04.22

(21) 申请号 202111558018.X

(51) Int.CI.

(22) 申请日 2021.12.20

G06F 16/36 (2019.01)

(65) 同一申请的已公布的文献号

G06F 16/28 (2019.01)

申请公布号 CN 113946695 A

G06K 9/62 (2022.01)

(43) 申请公布日 2022.01.18

(56) 对比文件

(73) 专利权人 山东新希望六和集团有限公司

CN 112231469 A, 2021.01.15

地址 266100 山东省青岛市崂山区九水东路592-26号3号楼4楼

CN 111680109 A, 2020.09.18

专利权人 四川新希望动物营养科技有限公司

Zhen Wang. Knowledge Graph Embedding by Translating on Hyperplanes.《AAAI》.2014, 第1112-1119页.

新希望六和股份有限公司

审查员 何蒙蒙

(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

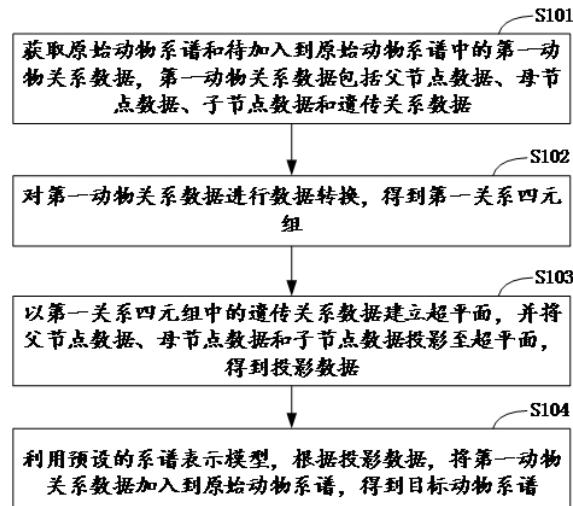
权利要求书3页 说明书9页 附图2页

(54) 发明名称

动物系谱的生成方法、装置及计算机设备

(57) 摘要

本申请公开了一种动物系谱的生成方法、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,得到第一关系四元组,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;再以第一关系四元组中的遗传关系数据建立超平面,并将父节点数据、母节点数据和子节点数据投影至超平面,得到投影数据,从而分解出子节点与父节点和母节点之间的线性关系;最后利用预设的系谱表示模型,根据投影数据,将第一动物关系数据加入到原始动物系谱,得到目标动物系谱,实现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. 一种动物系谱的生成方法,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父节点数据、母节点数据、子节点数据和遗传关系数据;

基于预设的四元组模板,对所述第一动物关系数据进行数据转换,得到第一关系四元组;

以所述第一关系四元组中的所述遗传关系数据建立超平面,并将所述父节点数据、所述母节点数据和所述子节点数据投影至所述超平面,得到投影数据;

利用预设的系谱表示模型,根据所述投影数据,将所述第一动物关系数据加入到原始动物系谱,得到目标动物系谱。

2. 如权利要求1所述的动物系谱的生成方法,其特征在于,所述第一关系四元组为 (F, M, S, r) ,其中 F 为所述父节点数据, M 为所述母节点数据, S 为所述子节点数据, r 为所述遗传关系数据。

3. 如权利要求1所述的动物系谱的生成方法,其特征在于,所述以所述第一关系四元组中的所述遗传关系数据建立超平面,并将所述父节点数据、所述母节点数据和所述子节点数据投影至所述超平面,得到投影数据,包括:

以所述第一关系四元组中的所述遗传关系数据建立超平面 W_r ;

将所述父节点数据投影至所述超平面 W_r ,得到父节点投影数据 $F_{\perp} = F - W_r^T F W_r$, F 为所述第一关系四元组中的父节点数据, W_r^T 为所述超平面 W_r 的转置矩阵;

将所述母节点数据投影至所述超平面 W_r ,得到母节点投影数据 $M_{\perp} = M - W_r^T M W_r$, M 为所述第一关系四元组中的母节点数据;

将所述子节点数据投影至所述超平面 W_r ,得到子节点投影数据 $S_{\perp} = S - W_r^T S W_r$, S 为所述第一关系四元组中的子节点数据。

4. 如权利要求3所述的动物系谱的生成方法,其特征在于,所述利用预设的系谱表示模型,根据所述投影数据,将所述第一动物关系数据加入到原始动物系谱,得到目标动物系谱,包括:

利用所述系谱表示模型,确定所述子节点投影数据与所述父节点投影数据和所述母节点投影数据之间的投影关系 d_r ;

根据所述投影关系 d_r ,确定所述子节点投影数据的表示向量,所述表示向量为 $\alpha \times F_{\perp} + (1 - \alpha)M_{\perp} + d_r$, α 为经验参数;

将所述子节点投影数据的表示向量输入到所述原始动物系谱,得到所述目标动物系谱。

5. 如权利要求1所述的动物系谱的生成方法,其特征在于,所述利用预设的系谱表示模型,根据所述投影数据,将所述第一动物关系数据加入到原始动物系谱,得到目标动物系谱之前,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,得到多组第二关系四元组;

利用多组所述第二关系四元组,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:

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

$$Difscore = margin + score(F', M', S', r') - score(F'', M'', S'', r'');$$

其中 $Loss$ 为损失函数值, $margin$ 为超参数, $score(F', M', S', r')$ 为正样本打分函数, $score(F'', M'', S'', r'')$ 为负样本打分函数。

6. 如权利要求5所述的动物系谱的生成方法,其特征在于,所述正样本打分函数为:

$$score(F', M', S', r') = \|\alpha \times F'_\perp + (1 - \alpha)M'_\perp + d'_r - S'_\perp\|_2^2;$$

其中 F' 为父节点正样本数据, M' 为母节点正样本数据, S' 为子节点正样本数据, r' 为遗传关系正样本数据, α 为经验参数, F'_\perp 为 F' 在基于 r' 建立的超平面上的投影数据, M'_\perp 为 M' 在基于 r' 建立的超平面上的投影数据, S'_\perp 为 S' 在基于 r' 建立的超平面上的投影数据, d'_r 为基于 r' 建立的超平面上的关系向量。

7. 一种动物系谱的生成装置,其特征在于,包括:

第一获取模块,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父节点数据、母节点数据、子节点数据和遗传关系数据;

第一转换模块,用于基于预设的四元组模板,对所述第一动物关系数据进行数据转换,得到第一关系四元组;

投影模块,用于以所述第一关系四元组中的所述遗传关系数据建立超平面,并将所述父节点数据、所述母节点数据和所述子节点数据投影至所述超平面,得到投影数据;

加入模块,用于利用预设的系谱表示模型,根据所述投影数据,将所述第一动物关系数据加入到原始动物系谱,得到目标动物系谱。

8. 如权利要求7所述的动物系谱的生成装置,其特征在于,所述投影模块,包括:

建立单元,用于所述第一关系四元组中的所述遗传关系数据建立超平面 W_r ;

第一投影单元,用于将所述父节点数据投影至所述超平面 W_r ,得到父节点投影数据 $F_\perp = F - W_r^T F W_r$, F 为所述第一关系四元组中的父节点数据, W_r^T 为所述超平面 W_r 的转置矩阵;

第二投影单元,用于将所述母节点数据投影至所述超平面 W_r ,得到母节点投影数据 $M_\perp = M - W_r^T M W_r$, M 为所述第一关系四元组中的母节点数据;

第三投影单元,用于将所述子节点数据投影至所述超平面 W_r ,得到子节点投影数据 $S_\perp = S - W_r^T S W_r$, S 为所述第一关系四元组中的子节点数据。

9. 如权利要求7所述的动物系谱的生成装置,其特征在于,所述生成装置还包括:

第二获取模块,用于获取多组第二动物关系数据;

第二转换模块,用于对每组所述第二动物关系数据进行数据转换,得到多组第二关系四元组;

训练模块,用于利用多组所述第二关系四元组,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:

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

$$Difscore = margin + score(F', M', S', r') - score(F'', M'', S'', r'');$$

其中 $Loss$ 为损失函数值, $margin$ 为超参数, $score(F', M', S', r')$ 为正样本打分函数, $score(F'', M'', S'', r'')$ 为负样本打分函数。

10. 一种计算机设备,其特征在于,包括存储器及处理器,所述存储器用于存储计算机程序,所述处理器运行所述计算机程序以使所述计算机设备执行如权利要求1至6中任一项所述的动物系谱的生成方法的步骤。

动物系谱的生成方法、装置及计算机设备

技术领域

[0001] 本申请涉及动物养殖技术领域,尤其涉及一种动物系谱的生成方法、装置及计算机设备。

背景技术

[0002] 动物育种作为动物养殖中的重要环节,通过不断挑选优秀个体进行繁殖,以改良动物种群的整体水平。其中动物亲缘相关系数、近交系数等数据被作为种猪优良水平的重要指标。为此,采用计算机技术分析动物育种过程中动物系谱关系,实现育种估算和亲子鉴定等。

[0003] 目前,基于物种系谱的计算模型主要有Blupf90模型。但是Blupf90模型在计算各指标时涉及大量的迭代和拟合运算,若动物系谱需要更新,则需要重新计算全部系谱关系,导致运算效率极低。例如从猪群繁殖的子代中选育出新的种猪,需要将种猪更新到动物系谱中,则此时需要将动物系谱中的系谱关系重新进行迭代和拟合运算。同时,Blupf90模型作为单机模型,当进入模型的数据量随着种群数量的增长而不断扩大时,其需要的计算资源和存储资源会呈指数级别扩增,非常不利于动物系谱的延续性。

发明内容

[0004] 本申请提供了一种动物系谱的生成方法、装置及计算机设备,以解决目前动物系谱的生成方式存在运算效率低的技术问题。

[0005] 为了解决上述技术问题,本申请实施例提供了一种动物系谱的生成方法,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父节点数据、母节点数据、子节点数据和遗传关系数据;

[0007] 对第一动物关系数据进行数据转换,得到第一关系四元组;

[0008] 以第一关系四元组中的遗传关系数据建立超平面,并将父节点数据、母节点数据和子节点数据投影至超平面,得到投影数据;

[0009] 利用预设的系谱表示模型,根据投影数据,将第一动物关系数据加入到原始动物系谱,得到目标动物系谱。

[0010] 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,得到第一关系四元组,以采用关系四元组描述子节点与父节点和母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;再以第一关系四元组中的遗传关系数据建立超平面,并将父节点数据、母节点数据和子节点数据投影至超平面,得到投影数据,从而分解出子节点与父节点和母节点之间的线性关系,以便于后续采用智能模型进行数据分析;最后利用预设的系谱表示模型,根据投影数据,将第一动物关系数据加入到原始动物系谱,得到目标动物系谱,以采用系谱表示模型分析子节点与父节点和母节点之间的关系向量,实现采用父节点与母节点对子节点的向量表示,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。

[0011] 可选地,所述第一关系四元组为(F, M, S, r),其中 F 为所述父节点数据, M 为所述母节点数据, S 为所述子节点数据, r 为所述遗传关系数据。本实施例采用四元组表示子节点与父节点和母节点之间的遗传关系,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。

[0012] 在一实施例中,以第一关系四元组中的遗传关系数据建立超平面,并将父节点数据、母节点数据和子节点数据投影至超平面,得到投影数据,包括:

[0013] 以所述第一关系四元组中的所述遗传关系数据建立超平面 W_r ;

[0014] 将所述父节点数据投影至所述超平面 W_r ,得到父节点投影数据 $F_{\perp} = F - W_r^T F W_r$, F 为所述第一关系四元组中的父节点数据, W_r^T 为所述超平面 W_r 的转置矩阵;

[0015] 将所述母节点数据投影至所述超平面 W_r ,得到母节点投影数据 $M_{\perp} = M - W_r^T M W_r$, M 为所述第一关系四元组中的母节点数据;

[0016] 将所述子节点数据投影至所述超平面 W_r ,得到子节点投影数据 $S_{\perp} = S - W_r^T S W_r$, S 为所述第一关系四元组中的子节点数据。

[0017] 本实施例通过以遗传关系建立超平面,并将各个节点投影至该超平面,从而分解出节点之间的线性关系,以转换为线性问题。

[0018] 在一实施例中,利用预设的系谱表示模型,根据投影数据,将第一动物关系数据加入到原始动物系谱,得到目标动物系谱,包括:

[0019] 利用所述系谱表示模型,确定所述子节点投影数据与所述父节点投影数据和所述母节点投影数据之间的投影关系 d_r ;

[0020] 根据所述投影关系 d_r ,确定所述子节点投影数据的表示向量,所述表示向量为 $\alpha \times F_{\perp} + (1 - \alpha)M_{\perp} + d_r$, α 为经验参数;

[0021] 将所述子节点投影数据的表示向量输入到所述原始动物系谱,得到所述目标动物系谱。

[0022] 本实施例通过系谱表示模型,确定节点数据之间的投影关系,得到超平面的关系向量,实现采用父节点和母节点表示子节点,从而在原始动物系谱中能够以父节点和母节点计算得到子节点,进而将子节点加入到原始动物系谱中。

[0023] 在一实施例中,利用预设的系谱表示模型,根据投影数据,将第一动物关系数据加入到原始动物系谱,得到目标动物系谱之前,还包括:

[0024] 获取多组第二动物关系数据;

[0025] 对每组第二动物关系数据进行数据转换,得到多组第二关系四元组;

[0026] 利用多组第二关系四元组,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:

[0027] $Loss = \max(0, Difscore)$;

[0028] $Difscore = margin + score(F', M', S', r') - score(F'', M'', S'', r'')$;

[0029] 其中 $Loss$ 为损失函数值, $margin$ 为超参数, $score(F', M', S', r')$ 为正样本打分函数, $score(F'', M'', S'', r'')$ 为负样本打分函数。

[0030] 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, 从而学习节点数据之间的关系特征, 进而能够在应用阶段实现子节点的向量表示。

[0031] 在一实施例中, 正样本打分函数为:

[0032] $score(F', M', S', r') = \|\alpha \times F'_\perp + (1 - \alpha)M'_\perp + d'_r - S'_\perp\|_2^2$;

[0033] 其中 F' 为父节点正样本数据, M' 为母节点正样本数据, S' 为子节点正样本数据, r' 为遗传关系正样本数据, α 为经验参数, F'_\perp 为 F' 在基于 r' 建立的超平面上的投影数据, M'_\perp 为 M' 在基于 r' 建立的超平面上的投影数据, S'_\perp 为 S' 在基于 r' 建立的超平面上的投影数据, d'_r 为基于 r' 建立的超平面上的关系向量。

[0034] 第二方面, 本申请实施例提供一种动物系谱的生成装置, 包括:

[0035] 第一获取模块, 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父节点数据、母节点数据、子节点数据和遗传关系数据;

[0036] 第一转换模块, 用于对第一动物关系数据进行数据转换, 得到第一关系四元组;

[0037] 投影模块, 用于以第一关系四元组中的遗传关系数据建立超平面, 并将父节点数据、母节点数据和子节点数据投影至超平面, 得到投影数据;

[0038] 加入模块, 用于利用预设的系谱表示模型, 根据投影数据, 将第一动物关系数据加入到原始动物系谱, 得到目标动物系谱。

[0039] 在一实施例中, 投影模块, 包括:

[0040] 建立单元, 用于所述第一关系四元组中的所述遗传关系数据建立超平面 W_r ;

[0041] 第一投影单元, 用于将所述父节点数据投影至所述超平面 W_r , 得到父节点投影数据 $F_\perp = F - W_r^T F W_r$, F 为所述第一关系四元组中的父节点数据, W_r^T 为所述超平面 W_r 的转置矩阵;

[0042] 第二投影单元, 用于将所述母节点数据投影至所述超平面 W_r , 得到母节点投影数据 $M_\perp = M - W_r^T M W_r$, M 为所述第一关系四元组中的母节点数据;

[0043] 第三投影单元, 用于将所述子节点数据投影至所述超平面 W_r , 得到子节点投影数据 $S_\perp = S - W_r^T S W_r$, S 为所述第一关系四元组中的子节点数据。

[0044] 在一实施例中, 生成装置还包括:

[0045] 第二获取模块, 用于获取多组第二动物关系数据;

[0046] 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, 得到多组第二关系四元组;

[0047] 训练模块, 用于利用多组第二关系四元组, 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:

[0048] $Loss = \max(0, Difscore)$;

[0049] $Difscore = margin + score(F', M', S', r') - score(F'', M'', S'', r'')$;

[0050] 其中 $Loss$ 为损失函数值, $margin$ 为超参数, $score(F', M', S', r')$ 为正样本打分函数, $score(F'', M'', S'', r'')$ 为负样本打分函数。

[0051] 第三方面,本申请实施例提供一种计算机设备,包括存储器及处理器,存储器用于存储计算机程序,处理器运行计算机程序以使计算机设备执行如上述第一方面中任一项的动物系谱的生成方法。

[0052] 需要说明的是,上述第二方面至第三方面的有益效果请参见第一方面的相关描述,在此不再赘述。

附图说明

[0053] 图1为本申请实施例提供的动物系谱的生成方法的流程示意图;

[0054] 图2为本申请实施例提供的超平面投影示意图;

[0055] 图3为本申请实施例提供的动物系谱的生成装置的结构示意图;

[0056] 图4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

[0057] 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。

[0058] 如相关技术记载,Blupf90模型在计算各指标时涉及大量的迭代和拟合运算,若动物系谱需要更新,则需要重新计算全部系谱关系,导致运算效率极低。例如从猪群繁殖的子代中选育出新的种猪,需要将种猪更新到动物系谱中,则此时需要将动物系谱中的系谱关系重新进行迭代和拟合运算。同时,Blupf90模型作为单机模型,当进入模型的数据量随着种群数量的增长而不断扩大时,其需要的计算资源和存储资源会呈指级别扩增,非常不利于动物系谱的延续性。

[0059] 为此,本申请实施例提供一种动物系谱的生成方法、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,得到第一关系四元组,以采用关系四元组描述子节点与父节点和母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;再以第一关系四元组中的遗传关系数据建立超平面,并将父节点数据、母节点数据和子节点数据投影至超平面,得到投影数据,从而分解出子节点与父节点和母节点之间的线性关系,以便于后续采用智能模型进行数据分析;最后利用预设的系谱表示模型,根据投影数据,将第一动物关系数据加入到原始动物系谱,得到目标动物系谱,以采用系谱表示模型分析子节点与父节点和母节点之间的关系向量,实现采用父节点与母节点对子节点的向量表示,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。

[0060] 请参照图1,图1为本申请实施例提供的一种动物系谱的生成方法的流程示意图。本实施例的动物系谱的生成方法可应用于计算机设备,计算机设备包括但不限于智能手机、平板电脑、笔记本电脑、桌上型电脑、物理服务器和云服务器等计算设备。如图1所示,本动物系谱的生成方法包括步骤S101至步骤S104,详述如下:

[0061] 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父节点数据、母节点数据、子节点数据和遗传关系数据。

[0062] 在本步骤中,每组第一动物关系数据包括一个父节点数据、一个母节点数据、一个节点数据和一个遗传关系,例如,猪群中的公猪、母猪、猪仔和一级亲属关系,其中一级亲属关系为血缘关系中的父母关系。可以理解的是,原始动物系谱记录有父节点和母节点,子节点为待加入到原始动物系谱的节点。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。

[0063] 需要说明的是,原始动物系谱采用数据向量表示各个节点,所以将子节点加入到原始动物系谱之前,需要计算子节点的数据向量,以往需要对所有遗传关系进行迭代和拟合运算,得到子节点向量。

[0064] 步骤S102,对所述第一动物关系数据进行数据转换,得到第一关系四元组。

[0065] 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。

[0066] 本申请实施例通过预先设置四元组模板,再将第一动物关系数据中的各个数据代入到预设的四元组模板,以实现数据转换,得到第一关系四元组,以便于后续基于父节点、母节点和遗传关系共同计算子节点的数据向量,避免子节点的数据向量表示冲突的问题。

[0067] 可选地,所述第一关系四元组为(F, M, S, r),其中 F 为所述父节点数据, M 为所述母节点数据, S 为所述子节点数据, r 为所述遗传关系数据。可以理解的是,上述第一关系四元组中的各个节点数据还可以是其他排列方式,如(S, F, M, r)等。

[0068] 步骤S103,以所述第一关系四元组中的所述遗传关系数据建立超平面,并将所述父节点数据、所述母节点数据和所述子节点数据投影至所述超平面,得到投影数据。

[0069] 在本步骤中,将遗传关系抽象为一个向量空间中的超平面,将父(头)节点数据、母(头)节点数据和子(尾)节点数据分别映射到该超平面上,得到每个节点对应的投影数据。

[0070] 步骤S104,利用预设的系谱表示模型,根据所述投影数据,将所述第一动物关系数据加入到原始动物系谱,得到目标动物系谱。

[0071] 在本步骤中,由于超平面上的投影数据属于平面数据,所以各个节点之间的投影数据存在数学几何关系,从而可以基于投影数据之间的数据几何关系建立线性函数,以实现第一关系四元组的数据分析。本实施例将投影数据输入到系谱表示模型,通过系谱表示模型根据超平面上的投影数据和平移向量(关系向量)计算头尾节点的差值,完成投影数据之间的数据分析,以得到子节点的表示向量,再将子节点的表示向量输入到原始动物系谱,得到目标动物系谱。

[0072] 在一实施例中,如图2所示的超平面投影示意图。以所述第一关系四元组中的所述遗传关系数据建立超平面 W_r ;

[0073] 将所述父节点数据投影至所述超平面 W_r ,得到父节点投影数据

$F_{\perp} = F - W_r^T F W_r$, F 为所述第一关系四元组中的父节点数据, W_r^T 为所述超平面 W_r 的转置矩阵;

[0074] 将所述母节点数据投影至所述超平面 W_r , 得到母节点投影数据 $M_{\perp} = M - W_r^T M W_r$, M 为所述第一关系四元组中的母节点数据;

[0075] 将所述子节点数据投影至所述超平面 W_r , 得到子节点投影数据 $S_{\perp} = S - W_r^T S W_r$, S 为所述第一关系四元组中的子节点数据。

[0076] 在本实施例中, 超平面 W_r 上有关系向量 d_r , 用于表示子节点投影数据 S_{\perp} 与父节点投影数据 F_{\perp} 和母节点投影数据 M_{\perp} 之间投影关系, 其为待确定向量。

[0077] 在一实施例中, 如图2所示的超平面投影示意图。上述步骤S104包括:

[0078] 利用所述系谱表示模型, 确定所述子节点投影数据与所述父节点投影数据和所述母节点投影数据之间的投影关系 d_r ;

[0079] 根据所述投影关系 d_r , 确定所述子节点投影数据的表示向量, 所述表示向量为 $\alpha \times F_{\perp} + (1 - \alpha)M_{\perp} + d_r$, α 为经验参数;

[0080] 将所述子节点投影数据的表示向量输入到所述原始动物系谱, 得到所述目标动物系谱。

[0081] 在本实施例中, α 为经验参数, 用于表征父子遗传的影响力, $1 - \alpha$ 用于表征母子遗传的影响力, $\alpha \in (0,1)$ 。 α 可以是用户根据经验的预设值, 也可以通过模型训练得到。优选地, 采用评价函数对模型训练过程时的 α 的评价, 并基于评价结果调整 α 的取值, 直至满足评价函数。

[0082] 在一实施例中, 在图1所示实施例的基础上, 上述步骤S104之前, 还包括:

[0083] 获取多组第二动物关系数据;

[0084] 对每组所述第二动物关系数据进行数据转换, 得到多组第二关系四元组;

[0085] 利用多组所述第二关系四元组, 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:

[0086] $Loss = \max(0, Difscore)$;

[0087] $Difscore = margin + score(F', M', S', r') - score(F'', M'', S'', r'')$;

[0088] 其中 $Loss$ 为损失函数值, $margin$ 为超参数, $score(F', M', S', r')$ 为正样本打分函数, $score(F'', M'', S'', r'')$ 为负样本打分函数。

[0089] 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四元组进行改进后的模型。第二关系四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四元组中的元素进行随机替换得到, 例如, 正样本对应的四元组为(公猪A, 母猪B, 猪仔C, 一级亲属关系), 通过随机替换, 得到负样本对应的四元组为(公猪D, 母猪B, 猪仔C, 一级亲属关系)、(公猪A, 母猪E, 猪仔C, 一级亲属关系)、(公猪A, 母猪B, 猪仔F, 一级亲属关系)或者(公猪A, 母猪B, 猪仔C, 二级亲属关系)等, 在此不再赘述。

[0090] 可选地,所述正样本打分函数为:

$$[0091] \text{score}(\mathbf{F}', \mathbf{M}', \mathbf{S}', \mathbf{r}') = \|\alpha \times \mathbf{F}'_{\perp} + (1 - \alpha) \mathbf{M}'_{\perp} + \mathbf{d}'_{\mathbf{r}} - \mathbf{S}'_{\perp}\|_2^2;$$

[0092] 其中 \mathbf{F}' 为父节点正样本数据, \mathbf{M}' 为母节点正样本数据, \mathbf{S}' 为子节点正样本数据, \mathbf{r}' 为遗传关系正样本数据, α 为经验参数, \mathbf{F}'_{\perp} 为 \mathbf{F}' 在基于 \mathbf{r}' 建立的超平面上的投影数据, \mathbf{M}'_{\perp} 为 \mathbf{M}' 在基于 \mathbf{r}' 建立的超平面上的投影数据, \mathbf{S}'_{\perp} 为 \mathbf{S}' 在基于 \mathbf{r}' 建立的超平面上的投影数据, $\mathbf{d}'_{\mathbf{r}}$ 为基于 \mathbf{r}' 建立的超平面上的关系向量。

[0093] 可选地,所述负样本打分函数为:

$$[0094] \text{score}(\mathbf{F}'', \mathbf{M}'', \mathbf{S}'', \mathbf{r}'') = \|\alpha \times \mathbf{F}''_{\perp} + (1 - \alpha) \mathbf{M}''_{\perp} + \mathbf{d}''_{\mathbf{r}} - \mathbf{S}''_{\perp}\|_2^2;$$

[0095] 其 \mathbf{F}'' 为父节点负样本数据, \mathbf{M}'' 为母节点负样本数据, \mathbf{S}'' 为子节点负样本数据, \mathbf{r}'' 为遗传关系负样本数据, \mathbf{F}''_{\perp} 为 \mathbf{F}'' 在基于 \mathbf{r}'' 建立的超平面上的投影数据, \mathbf{M}''_{\perp} 为 \mathbf{M}'' 在基于 \mathbf{r}'' 建立的超平面上的投影数据, \mathbf{S}''_{\perp} 为 \mathbf{S}'' 在基于 \mathbf{r}'' 建立的超平面上的投影数据, $\mathbf{d}''_{\mathbf{r}}$ 为基于 \mathbf{r}'' 建立的超平面上的关系向量。

[0096] 为了执行上述方法实施例对应的动物系谱的生成方法,以实现相应功能和技术效果。参见图3,图3示出了本申请实施例提供的一种动物系谱的生成装置的结构框图。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生成装置,包括:

[0097] 第一获取模块301,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父节点数据、母节点数据、子节点数据和遗传关系数据;

[0098] 第一转换模块302,用于对所述第一动物关系数据进行数据转换,得到第一关系四元组;

[0099] 投影模块303,用于以所述第一关系四元组中的所述遗传关系数据建立超平面,并

将所述父节点数据、所述母节点数据和所述子节点数据投影至所述超平面,得到投影数据;

[0100] 加入模块304,用于利用预设的系谱表示模型,根据所述投影数据,将所述第一动物关系数据加入到原始动物系谱,得到目标动物系谱。

[0101] 可选地,所述第一关系四元组为 $(\mathbf{F}, \mathbf{M}, \mathbf{S}, \mathbf{r})$,其中 \mathbf{F} 为所述父节点数据, \mathbf{M} 为所述母节点数据, \mathbf{S} 为所述子节点数据, \mathbf{r} 为所述遗传关系数据。

[0102] 在一实施例中,投影模块303,包括:

[0103] 建立单元,用于以所述第一关系四元组中的所述遗传关系数据建立超平面 \mathbf{W}_r ;

[0104] 第一投影单元,用于将所述父节点数据投影至所述超平面 \mathbf{W}_r ,得到父节点投影数据 $\mathbf{F}_{\perp} = \mathbf{F} - \mathbf{W}_r^T \mathbf{F} \mathbf{W}_r$, \mathbf{F} 为所述第一关系四元组中的父节点数据, \mathbf{W}_r^T 为所述超平面 \mathbf{W}_r 的转置矩阵;

[0105] 第二投影单元,用于将所述母节点数据投影至所述超平面 \mathbf{W}_r ,得到母节点投影数据 $\mathbf{M}_{\perp} = \mathbf{M} - \mathbf{W}_r^T \mathbf{M} \mathbf{W}_r$, \mathbf{M} 为所述第一关系四元组中的母节点数据;

[0106] 第三投影单元,用于将所述子节点数据投影至所述超平面 \mathbf{W}_r ,得到子节点投影数据 $\mathbf{S}_{\perp} = \mathbf{S} - \mathbf{W}_r^T \mathbf{S} \mathbf{W}_r$, \mathbf{S} 为所述第一关系四元组中的子节点数据。

- [0107] 在一实施例中,加入模块304,包括:
- [0108] 第一确定单元,用于利用所述系谱表示模型,确定所述子节点投影数据与所述父节点投影数据和所述母节点投影数据之间的投影关系 d_r ;
- [0109] 第二确定单元,用于根据所述投影关系 d_r ,确定所述子节点投影数据的表示向量,所述表示向量为 $\alpha \times F_{\perp} + (1 - \alpha)M_{\perp} + d_r$, α 为经验参数;
- [0110] 输入单元,用于将所述子节点投影数据的表示向量输入到所述原始动物系谱,得到所述目标动物系谱。
- [0111] 在一实施例中,生成装置还包括:
- [0112] 第二获取模块,用于获取多组第二动物关系数据;
- [0113] 第二转换模块,用于对每组所述第二动物关系数据进行数据转换,得到多组第二关系四元组;
- [0114] 训练模块,用于利用多组所述第二关系四元组,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:
- [0115] $Loss = \max(0, Difscore)$;
- [0116] $Difscore = margin + score(F', M', S', r') - score(F'', M'', S'', r'')$;
- [0117] 其中 $Loss$ 为损失函数值, $margin$ 为超参数, $score(F', M', S', r')$ 为正样本打分函数, $score(F'', M'', S'', r'')$ 为负样本打分函数。
- [0118] 在一实施例中,所述正样本打分函数为:
- [0119] $score(F', M', S', r') = \|\alpha \times F'_{\perp} + (1 - \alpha)M'_{\perp} + d'_r - S'_{\perp}\|_2^2$;
- [0120] 其中 F' 为父节点正样本数据, M' 为母节点正样本数据, S' 为子节点正样本数据, r' 为遗传关系正样本数据, α 为经验参数, F'_{\perp} 为 F' 在基于 r' 建立的超平面上的投影数据, M'_{\perp} 为 M' 在基于 r' 建立的超平面上的投影数据, S'_{\perp} 为 S' 在基于 r' 建立的超平面上的投影数据, d'_r 为基于 r' 建立的超平面上的关系向量。
- [0121] 上述的动物系谱的生成装置可实施上述方法实施例的动物系谱的生成方法。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。
- [0122] 图4为本申请一实施例提供的计算机设备的结构示意图。如图4所示,该实施例的计算机设备4包括:至少一个处理器40(图4中仅示出一个)处理器、存储器41以及存储在所述存储器41中并可在所述至少一个处理器40上运行的计算机程序42,所述处理器40执行所述计算机程序42时实现上述任意方法实施例中的步骤。
- [0123] 所述计算机设备4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仅限于处理器40、存储器41。本领域技术人员可以理解,图4仅仅是计算机设备4的举例,并不构成对计算机设备4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。
- [0124] 所称处理器40可以是中央处理单元(Central Processing Unit, CPU),该处理器

40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。

[0125] 所述存储器41在一些实施例中可以是所述计算机设备4的内部存储单元,例如计算机设备4的硬盘或内存。所述存储器41在另一些实施例中也可以是所述计算机设备4的外部存储设备,例如所述计算机设备4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41还可以既包括所述计算机设备4的内部存储单元也包括外部存储设备。所述存储器41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41还可以用于暂时地存储已经输出或者将要输出的数据。

[0126] 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。

[0127] 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。

[0128] 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时,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。

[0129] 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。

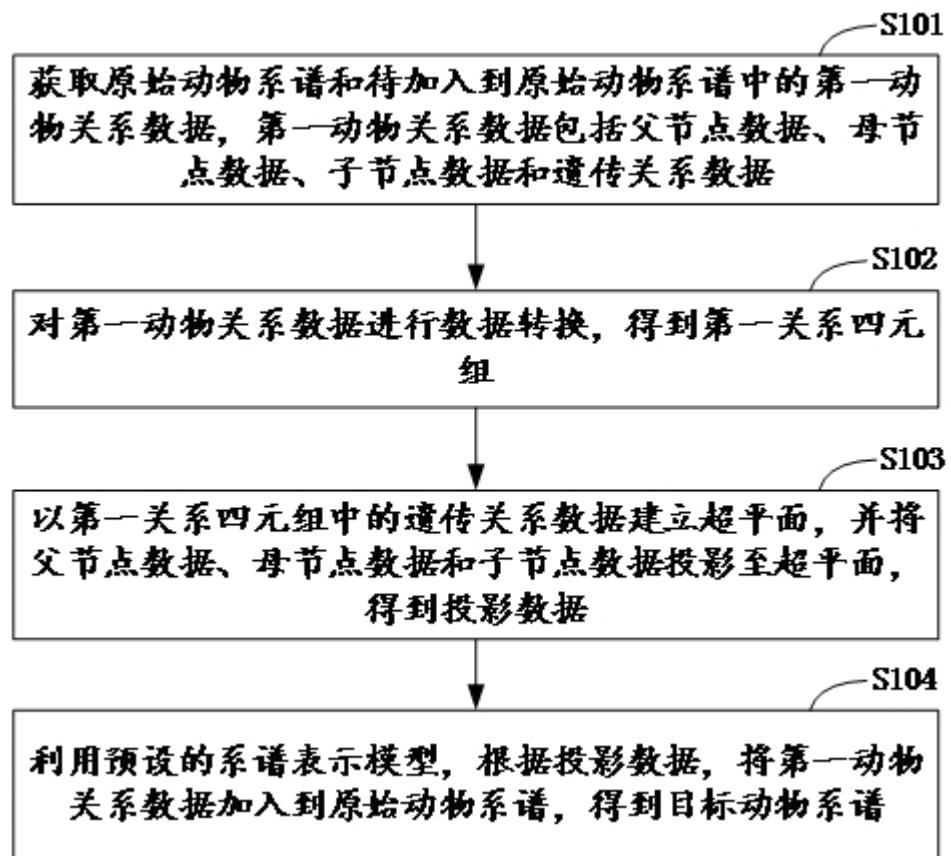


图1

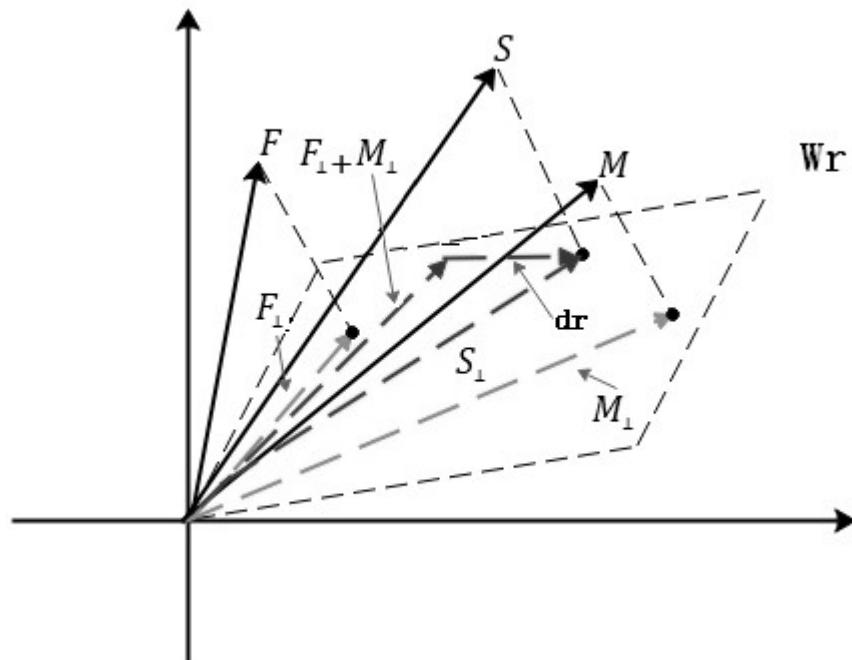


图2

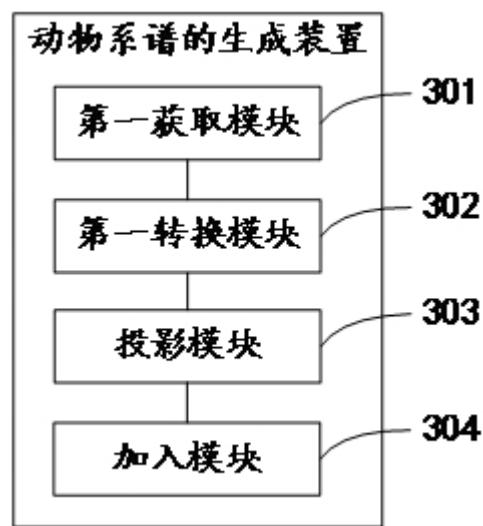


图3

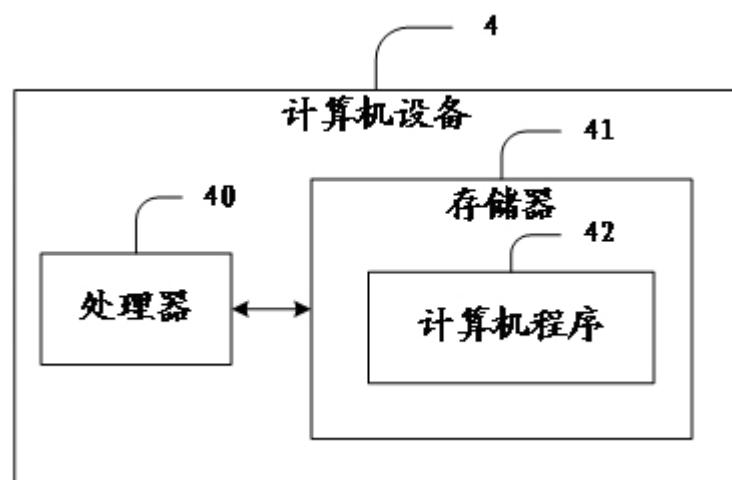


图4