



(21) 申请号 201880035469.5

(22) 申请日 2018.06.22

(65) 同一申请的已公布的文献号
申请公布号 CN 110692050 A

(43) 申请公布日 2020.01.14

(30) 优先权数据
15/632,564 2017.06.26 US
15/844,723 2017.12.18 US

(85) PCT国际申请进入国家阶段日
2019.11.28

(86) PCT国际申请的申请数据
PCT/IB2018/054610 2018.06.22

(87) PCT国际申请的公布数据
W02019/003069 EN 2019.01.03

(73) 专利权人 国际商业机器公司
地址 美国纽约阿芒克

(72) 发明人 A.莫汉 S.麦卡蒂尔
D.麦克洛斯基 M.索格林

(74) 专利代理机构 北京市柳沈律师事务所
11105

专利代理师 邸万奎

(51) Int.Cl.
G06F 16/36 (2019.01)

(56) 对比文件
CN 106339401 A,2017.01.18
JP 2016212838 A,2016.12.15
US 2008263038 A1,2008.10.23
US 2017024461 A1,2017.01.26
US 9336306 B2,2016.05.10
Ann Devitt et al.Sentiment polarity
identification in financial news: A
cohesion-based approach.《Proceedings of
the 45th annual meeting of the
association of computational
linguistics》.2007,

审查员 卢雅莹

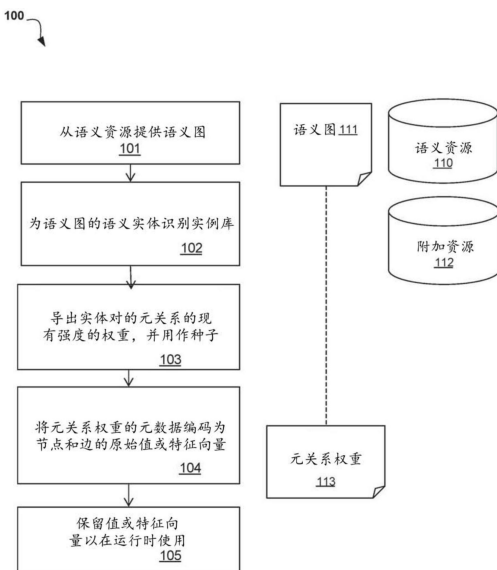
权利要求书2页 说明书18页 附图9页

(54) 发明名称

语义图中元关系的自适应评估

(57) 摘要

提供了用于语义图中的元关系的自适应评估的方法和系统。该方法包括基于知识库提供语义图,在其中以图节点形式的概念通过以图边形式的语义关系链接。元数据编码在语义图的边和节点中,用于度量元关系的权重,其中元关系应用于语义图的概念,并且独立于由语义图的边定义的语义关系。执行与语义图的一个或多个概念相关的输入上下文的图激活,其中通过语义图将权重应用于扩散激活信号以产生用于语义图的概念的子集的元关系的度量。



1. 一种用于语义图中的元关系的自适应评估的计算机实现的方法,所述方法包括:

基于知识库从自然语言源推导语义图,在其中以图节点形式的概念通过以图边形式的语义关系链接;

在语义图的边和节点的元数据中将用于度量元关系的权重编码为特征向量,其中元关系应用于语义图的概念并且独立于由语义图的边定义的语义关系,其中所述特征向量包括元关系的强度和强度的置信度分数,并且其中所述强度使用针对节点的运行时输入;以及

执行与语义图的一个或多个概念相关的输入上下文的图激活,其中通过语义图将作为特征向量的权重应用于扩散激活信号以产生用于语义图的概念的子集的元关系的度量;

响应于图激活为语义图的边和节点进行评分;

用从独立于语义图所基于的知识库的资源集中获得的用于度量元关系的种子权重来播种图;

响应于对独立于所述语义图所基于的知识库的资源集的添加,更新所述种子权重。

2. 如权利要求1所述的方法,其中执行用于所述输入上下文的图激活还包括:

发现输入上下文中的概念的实例;

激活与语义图中的概念相对应的节点;

向外遍历信号到相邻节点;

在对信号应用作为特征向量的权重的同时依次激活所述相邻节点;

确定具有最高结果激活信号的一个或多个焦点节点。

3. 如权利要求2所述的方法,还包括输出语义图的结果激活部分,其反映输入上下文中的元关系的度量。

4. 如权利要求1所述的方法,其中用于度量元关系的作为特征向量的权重是从资源集获得的并且响应于对资源集的添加被更新的原始值,其中原始值在图激活期间被应用。

5. 如权利要求1所述的方法,其中作为特征向量的权重指示用于元关系的不同方面和/或元关系的极性的多维度量。

6. 如权利要求1所述的方法,其中作为特征向量的权重是响应于输入上下文概念的实例的节点的运行时输入计算的。

7. 如权利要求6所述的方法,其中特征向量包括要应用于节点的运行时输入的相关性因素。

8. 如权利要求7所述的方法,其中相关性因素对于不同的节点是不同的。

9. 如权利要求6所述的方法,其中除了元关系之外,特征向量还包括输入上下文中的概念的实例的语义和词汇特征。

10. 如权利要求6所述的方法,其中特征向量定义作为特征向量的权重的置信度分数。

11. 如权利要求6所述的方法,其中特征向量使用统计技术定义元关系度量的聚合。

12. 如权利要求1所述的方法,其中元关系涉及以下组中的一种形式的现象:

情感分析、偏差评估、预测分析中的偏差、使用信息检索的查询扩展、风险评估、地理空间推断、以及包括临床试验匹配的治疗、使用或处理的适用性。

13. 一种用于语义图中的元关系的自适应评估的计算机系统,所述计算机系统包括:

处理器和存储器,被配置为向处理器提供计算机程序指令以执行部件的功能;

语义图组件,用于基于知识库从自然语言源推导语义图,在其中以图节点形式的概念

通过以图边形式的语义关系链接；

元关系组件，用于在语义图的边和节点的元数据中将用于度量元关系的权重编码为特征向量，其中元关系应用于语义图的概念并且独立于由语义图的边定义的语义关系，其中所述特征向量包括元关系的强度和强度的置信度分数，并且其中所述强度使用针对节点的运行时输入；以及

运行时组件，用于执行与语义图的一个或多个概念相关的输入上下文的图激活，其中通过语义图将作为特征向量的权重应用于扩散激活信号以产生用于语义图的概念的子集的元关系的度量；

评分组件，用于响应于图激活为语义图的边和节点进行评分；

权重添加组件，用于用从独立于语义图所基于的知识库的资源集中获得的用于度量元关系的种子权重来播种图；

权重更新组件，用于响应于对独立于所述语义图所基于的知识库的资源集的添加，更新所述种子权重。

14. 如权利要求13所述的计算机系统，其中运行时组件包括：

概念检测组件，用于发现输入上下文中的概念的实例；以及

图激活组件，用于激活与语义图中的概念相对应的节点，向外遍历信号到相邻节点，在对信号应用权重的同时依次激活这些节点，以及确定具有最高结果激活信号的一个或多个焦点节点。

15. 如权利要求14所述的计算机系统，还包括输出组件，用于输出语义图的结果激活部分，其反映输入上下文中的元关系的度量。

16. 如权利要求15所述的计算机系统，其中输出组件包括用于输出激活的子图的环境上下文组件，其中节点和边上的激活权重表示元关系的度量。

17. 如权利要求13所述的计算机系统，其中所述权重更新组件响应于对资源集或输入上下文的添加而更新权重。

18. 如权利要求13所述的计算机系统，其中所述运行时组件响应于输入上下文概念的实例的节点的运行时输入计算特征向量。

19. 如权利要求18所述的计算机系统，其中特征向量包括要应用于节点的运行时输入的相关性因素。

20. 如权利要求18所述的计算机系统，其中除了元关系之外，特征向量还包括输入上下文中的概念的实例的语义和/或词汇特征。

21. 如权利要求18所述的计算机系统，其中特征向量使用统计技术定义元关系度量的聚合。

22. 如权利要求13所述的计算机系统，其中元关系涉及以下组中的一种形式的现象：

情感分析、偏差评估、预测分析中的偏差、使用信息检索的查询扩展、风险评估、地理空间推断、以及包括临床试验匹配的治疗、使用或处理的适用性。

23. 一种用于语义图中的元关系的适应性评估的计算机可读存储介质，具有体现其中的程序指令，所述程序指令可由处理器执行以使处理器执行权利要求1至12中任一项的方法。

语义图中元关系的自适应评估

技术领域

[0001] 本发明涉及语义图，并且更具体地，涉及语义图中的元关系(meta-relationships)的自适应评估。

[0002] 从结构化知识导出的语义概念和关系的图对于高精度自然语言处理(NLP)系统是极其有价值的资源。

[0003] 在计算机科学中，本体被定义为“共享概念化的显式规范”，其中概念化可以是关于给定任务的要求的真实世界语义的某个子集。它可以包含对象的概念或类、对象属性和概念间关系，以及目标域中的这些的实例。这种结构化资源便于共享和重用领域知识，并且对于NLP应用是非常宝贵的。这种资源的主要例子是由国家医学图书馆(NLM, 2013)提供的统一医学语言系统(UMLS)。数据集由结合医学领域中的概念和概念间关系的本体的大词典组成，包括数百万个实例表面形式。

[0004] 首先在人类语义记忆的抽象模型中提出了扩散激活的理论，以便人为地表示人类大脑可以处理和理解自然语言的语义的手段。该模型被增强用于检索任务，并为许多其它相关领域的研究提供了灵感，尤其是从认知心理学到神经科学、到自然语言处理，等等。

[0005] 扩散激活的基本前提与人工智能中的连接主义(connectionism)的前提有关，连接主义使用神经网络的类似模型来反映电信号在人脑中的扇出效应。在神经网络的情况下，图中的顶点可以表示单个神经元，并且边可以表示突触。在信息检索和词义消歧中，通常顶点将表示词义，而边将表示这些词义之间的某种形式的关系，无论是词汇的还是语义的链接。

[0006] 扩散激活策略在2014年COLING会议论文集，第25届国际计算语言学会议：技术论文，第2237-2248页，都柏林，爱尔兰，2014年8月23日至29日，Ronan Mac an tSaoir撰写的论文“使用扩散激活来评估和改善本体”中进行了描述。该技术涉及处理文本文档，并且当发现语义图中的节点时，通过从这些节点传播信号来激活这些节点，然后该信号在图中传播，并且潜在地在文本中未出现的其它节点处累积。

[0007] 这个基本思想有各种不同的实现，但是由**IBM®** Galaxy实现的一种实现(**IBM®**和Galaxy是国际商业机器公司的商标)对于字义消歧和字义推断的目的是有用的。

[0008] NLP中有各种领域来分析与实体之间的关系有关的现象。这些领域的示例包括：情感分析、偏见检测、地理空间推断、上下文相关性和风险评估。在现有技术方法中需要跨实体的这种现象的准确评分，并且需要对这种现象的改进的分析。

发明内容

[0009] 根据本发明的一方面，提供了用于语义图中的元关系的自适应评估的计算机实现的方法，包括：提供基于知识库的语义图，在其中以图节点形式的概念通过以图边形式的语义关系链接；在语义图的边和节点的元数据中编码用于度量元关系的权重，其中元关系应用于语义图的概念并且独立于由语义图的边定义的语义关系；以及执行与语义图的一个或多个概念相关的输入上下文的图激活，其中通过语义图将权重应用于扩散激活信号以产生

用于语义图的概念的子集的元关系的度量。

[0010] 所描述的方法提供了使用图激活来跨语义图的子集中所定义的概念对元关系进行评分的优点。元关系的编码权重可针对输入上下文的元关系被激活,从而能够对相关实体的元关系进行复杂的评分和分析。

[0011] 执行用于输入上下文的图激活包括:发现输入上下文中的概念的实例以及激活与语义图中的概念相对应的节点,向外遍历信号到相邻节点,在对信号应用权重的同时依次激活这些节点,以及确定具有最高结果激活信号的一个或多个焦点节点。该方法可以输出语义图的结果激活部分,其反映输入上下文中的元关系的度量。这提供了对输入上下文的概念的实例之间的元关系进行评分的优点。

[0012] 该方法可以用从独立于语义图所基于的知识库的资源集获得的元关系的度量的权重来播种图。这使得能够针对特定应用对权重进行编码。语义图可以用作元关系的度量可以被叠加的基础。

[0013] 权重可以指示用于元关系的不同方面和/或元关系的极性的多维度量。元关系可以具有可以被度量并编码为权重中的维度的不同方面。权重还可以具有极性,使得可以包括正值和负值。

[0014] 在一个实施例中,权重可以是资源集获得的原始值,并且可以响应于对资源集的添加而被更新,其中在图激活期间应用原始值。

[0015] 在另一实施例中,权重可以是响应于输入上下文概念的实例的节点的运行时输入计算的特征向量。这实现了基于用于正被评分的输入上下文的输入的权重的运行时自适应的优点。特征向量可以包括要应对于节点的运行时输入的相关性因素,并且相关性因素对于不同节点可以是不同的。

[0016] 另外,特征向量除了元数据关系特征之外还可以包括输入上下文中的概念的实例的语义和/或词汇特征。这使得语义图的语义关系能够被并入到分析中。或者,特征向量可以仅与元关系有关。

[0017] 特征向量可以使用统计技术来定义元关系度量的权重和聚合的置信度得分。

[0018] 元关系可以涉及以下面组中之一的形式的现象:情感分析、偏差评估、预测分析中的偏差、使用信息检索的查询扩展、风险评估、地理空间推断、以及包括临床试验匹配的治疗、使用或处理的适用性。这些现象可以普遍地应用于语义图的概念。

[0019] 根据本发明的另一方面,提供了用于语义图中的元关系的自适应评估的系统,包括:处理器和存储器,其被配置为向处理器提供计算机程序指令以执行部件的功能;语义图组件,用于基于知识库提供语义图,在其中以图节点形式的概念通过以图边形式的语义关系链接;元关系组件,用于在语义图的边和节点的元数据中编码用于度量元关系的权重,其中元关系应用于语义图的概念并且独立于由语义图的边定义的语义关系;以及运行时部件,用于执行与语义图的一个或多个概念相关的输入上下文的图激活,其中通过语义图将权重应用于扩散激活信号以产生用于语义图的概念的子集的元关系的度量。

[0020] 该系统可以提供以下优点:使得能够对元关系权重进行编码,其中,在为输入上下文的元关系进行复杂评分的情况下,可以将该元关系权重自动应用于输入上下文。

[0021] 运行时部件可以包括:概念检测组件,用于发现输入上下文中的概念的实例;以及图激活组件,用于激活与语义图中的概念相对应的节点,向外遍历信号到相邻节点,在对信

号应用权重的同时依次激活这些节点,以及确定具有最高结果激活信号的一个或多个焦点节点。

[0022] 该系统可以包括输出组件,用于输出反映输入上下文中元关系的度量的语义图的结果激活部分。输出组件可包括用于输出激活的子图的环境上下文组件,其中节点和边上的激活权重表示元关系的度量。

[0023] 该系统可包括权重添加组件,用于向该图播种从独立于语义图所基于的知识库的资源集获得的元关系的度量的权重。该系统还可包括用于响应于对资源集或输入上下文的添加来更新权重的权重更新组件。

[0024] 该系统可以被并入用于度量与以下面组中之一的形式的现象相关的元关系的系统中:情感分析、偏差评估、预测分析中的偏差、使用信息检索的查询扩展、风险评估、地理空间推断、以及包括临床试验匹配的治疗、使用或处理的适用性。

[0025] 根据本发明的另一方面,提供了用于语义图中的元关系的适应性评估的计算机程序产品,所述计算机程序产品包括具有体现其中的程序指令的计算机可读存储介质,所述程序指令可由处理器执行以使处理器:提供基于知识库的语义图,在其中以图节点形式的概念通过以图边形式的语义关系链接;在语义图的边和节点的元数据中编码用于度量元关系的权重,其中元关系应用于语义图的概念并且独立于由语义图的边定义的语义关系;以及执行与语义图的一个或多个概念相关的输入上下文的图激活,其中通过语义图将权重应用于扩散激活信号以产生用于语义图的概念的子集的元关系的度量。

附图说明

[0026] 在说明书的结论部分特别指出并清楚地要求保护本发明的主题。当结合附图阅读时,通过参考以下详细描述,可以最好地理解本发明的组织和操作方法,以及其目的、特征和优点。

[0027] 现在将参考以下附图仅通过示例的方式描述本发明的优选实施例,在附图中:

[0028] 图1A是根据本发明的方法的一个方面的示例实施例的流程图;

[0029] 图1B是根据本发明的另一方面的示例实施例的流程图;

[0030] 图2是根据本发明的系统的示例实施例的框图;

[0031] 图3是示出根据本发明的方法的第一示例实施例的示意图;

[0032] 图4A和4B是示出根据本发明的方法的第二示例实施例的示意图;

[0033] 图5是其中可以实现本发明的计算机系统或云服务器的实施例的框图;

[0034] 图6是其中可以实现本发明的云计算环境的示意图;以及

[0035] 图7是其中可以实现本发明的云计算环境的抽象模型层的图示。

[0036] 应当理解,为了说明的简单和清楚,图中所示的元件不一定按比例绘制。例如,为了清楚起见,一些元件的尺寸可能相对于其他元件被放大。此外,在认为适当的情况下,附图标记可以在附图中重复以指示对应或类似的特征。

具体实施方式

[0037] 所描述的方法在自然语言处理期间使用语义图中的扩散激活来量化领域中的实体之间的现象的“元关系”的强度。元关系的细节以及扩散信号应当如何产生这种现象的强

度的度量被编码在图形元数据中,并且根据任务的性质和所讨论的领域而改变。

[0038] 元关系描述了以属性或特性的形式普遍应用于图中的实体的现象。当附加的输入资源可用时,现象的强度可以在语义图中被播种,并且使用诸如扩散激活的图激活来自适应地度量。

[0039] 这提供了对在领域的输入上下文中跨实体对现象评分的问题的技术解决方案。该解决方案使用语义图的图激活,其使得能够在实体之间进行复杂的评分和度量。

[0040] 术语“元关系”用于描述涉及应用于实体的现象的元级关系或通用关系。“元”的解释是更高级的并且独立于现有的图形语义关系。元关系被叠加在任何特定图的现有语义上。

[0041] 元关系不是初始语义图的一部分,并且不一定与图中的边所表示的实体之间的关系相关。虽然语义图可能包含用于语义相关性的边,诸如在两个节点或概念之间的“写”,像“作者”和“文章”,但是所描述的方法不是集中在这个特定语义关系的性质或重要性上。相反,利用该结构来量化现象、属性或特性的一些更高级的元上下文,诸如“情感”、“偏见”、“地理空间相关性”或“风险暴露”,其可以整体地或普遍地但以不同的强度应用于图中的实体。

[0042] 传统的语义图形术语是指概念之间的语义关系的图形节点和边,其是自然语言处理中的语义图形的主要功能。所描述方法叠加了可以度量的附加方面或现象的权重。这种额外的现象不限于各个链接的节点对的细节。

[0043] 语义图以元关系的强度的度量作为种子,该度量来自使用附加资源的上下文到用于提供初始语义图的那些上下文,以便在自然语言处理期间实现对元关系的主题的分析。

[0044] 通过处理存储在图中的信息,元关系被暴露并被分配强度。除了语义图中的现有关系之外的元关系扩展需要除了语义图所基于的源的第一数据库之外的附加资源的第二数据库。

[0045] 语义图的默认解释通过所描述的方法用由元关系表示的任何数量的各种通用现象来扩充,所述元关系从根本上改变了在自然语言处理中可以如何使用相同的语义图。当用户通常通过任何语义图发送信号时,他们期望权重来告诉他们在语义上有多紧密相关的事情。这被重新连线以表示在上下文中优选的任何元关系。这可以是“地理上有多近的事物”或“事物彼此有多偏差”等。

[0046] 使用扩散激活权重来量化已经存在的实体之间的语义关系的强度。所描述的方法通过使用元关系强度将其扩展到量化除了简单上下文相关性之外的某种现象。

[0047] 语义图可以在节点和边级被定制,无论是通过将权重修改器归于特定节点和边类型,还是通过修改特定节点的起始权重。所描述的方法提供了在更高级的上下文方面的图结构和内容的领域定制扩散激活。对诸如语料库相关性、偏差或风险暴露等现象或上下文的元关系的知晓对于自然语言处理具有非常不同的考虑和后果。

[0048] 在自然语言处理领域内有许多主题,这些主题集中于检测特定语言现象的强度程度或量化一些分析主题的特定特征。所描述的方法提供了一种技术,该技术可以应用于许多这样的任务以便产生相同数量或现象的环境表示。诸如偏差或地理相关性之类的现象的环境表示本质上是激活的子图,其中节点和边上的激活权重表示现象的强度程度。这里的词“环境”反映了表示的网络或类似网络的结构。

[0049] 描述了自然语言处理的以下特定领域中的示例实施例：情感分析、偏差检测、地理空间推断、上下文相关性和风险评估。

[0050] 参考图1A，流程图100示出了所描述的将现象的元关系应用于语义图以使用图激活进行自适应评估的方法的一个方面的示例实施例。

[0051] 提供101语义图111，其可以是提供知识库的一个或多个语义资源110建立的或者可以是现有的图。用于该操作的资源可以是解析帧、本体、命名实体识别 (NER) 数据的共现等。在该描述的方法中，语义图111可以是现有的预建立图，其可以由所描述的方法使用，在这种情况下，不必再访原始资源110。

[0052] 该方法不是特定于任何特定风格的自然语言文本或领域术语。这里使用的语义图可以从任何自然语言源导出，在该自然语言源上可以应用命名实体识别 (NER) 和自然语言语法分析器。

[0053] 该方法可以标识102以针对感兴趣的现象要引用的附加资源112的形式的实例全集。对于要使用元关系来度量强度的特定现象，使用附加资源112形式的示例，用于图激活的初始元数据基于该示例。附加资源112是指在感兴趣现象的上下文中语义图111的实体。

[0054] 该方法可以导出103反映现象的所有实体对之间的现有强度和/或极性的元关系的权重，所述现象是可以用作种子的NLP任务的目标。这可以使用目标的先前分析来提供，诸如情感分析或偏差检测数据、地理距离数据、IR统计、风险因子分数等。该方法利用背景知识来根据任务和领域对语义图进行加权。

[0055] 该方法可以在元数据104中将元关系权重113编码为硬编码的原始值或特征向量。节点和边权重是输入数据特征的函数，并且可以被存储为原始值或特征向量以用于稍后的计算。

[0056] 用于现有语义图实现的元数据值或特征向量通常是反映在图中的节点或边处的扩散信号的累积的单个正值。在所描述的方法中，所提供的权重具有与被评估的元关系有关的新含义。

[0057] 特征向量是表示某个对象的数字特征的n维向量。特征向量等同于在统计过程中使用的解释变量的向量。特征向量常常与使用点积的权重组合，以便构建用于确定进行预测的分数的线性预测函数。

[0058] 特征向量可以包括但不限于以下：

[0059] i. 语料库中概念实例的语义和词汇特征；

[0060] ii. 任务的主要目标的元关系的强度；这也可以是多维的，例如，如在可以被极化的偏差或情绪的情况下；

[0061] iii. 所有上述内容的置信度得分；以及

[0062] iv. 使用统计技术，例如对相同向量的逻辑回归，聚集相关性和强度度量。

[0063] 特征向量中的元关系的强度可以使用针对节点的运行时输入，其可以与硬编码的相关性因素组合。这些对于每个节点或节点类型可以是不同的。

[0064] 节点之间的物理链路意义上的关系不被修改。相反，元关系反映在激活权重的数值和值(+/-)的符号中。元关系的标签(诸如偏差或风险暴露)不在图内以节点或边来编码。变化是扩散激活之后的信号强度具有新的意义。

[0065] 通过保存105原始值或特征向量以供在运行时使用，该方法还便于根据对系统的

未来输入的性质来任选地适配图形权重。

[0066] 不管意图如何,该方法都不需要显著的适应,无论是情感分析、观点偏差检测、地理空间相关性、语义相关性,风险暴露等。

[0067] 参考图1B,流程图150示出了使用图激活来自适应评估现象的元关系的所述方法的另一方面的示例实施例。

[0068] 当在运行时处理151以输入上下文114形式的未见文本时,该方法可以发现152实体引用并在那些开始点处激活语义图115。通常,这可以使用与在构建101语义图时使用的NER生成器相同的NER生成器。

[0069] 该方法可以通过根据在图1A的方法中添加的节点和边的元关系的权重传播信号来执行153语义图中的扩散激活。

[0070] 通过在输入上下文114中发现概念的实例,使用用于实例的唯一标识符的集合,可以在语义图中激活对应节点,从该语义图中信号向外穿过相邻节点,依次激活这些节点。

[0071] 在一个实施例中,输入上下文中的概念的发现的实例可以使用节点的权重的硬编码的原始值来激活对应的节点。在另一实施例中,来自概念实例的输入可以用于计算节点的特征向量。输入可以与硬编码的相关性因素组合,该相关性因素可以针对如在权重特征向量中定义的不同节点或节点类型而变化。

[0072] 随着信号从源节点扩散得更远,它减弱了在针对图中的节点和边的相关联的权重模型中指定的量。如果信号从多个附近的源节点扩散,则信号将组合,并且重叠点将被更大程度地激活。累积最多激活的节点被认为是输入上下文的焦点节点。所得到的图的激活部分将反映输入上下文中元关系现象的固有强度。

[0073] 输入上下文114可以由检测到的实体根据激活结束时的信号强度来评分154,并且该评分可以作为结果输出。

[0074] 可以将得分用于155环境上下文节点,以便在相同的环境中导出进一步的结果,其可以作为进一步的结果输出。例如,作为输入提供的特定法律的相关地理管辖权的推断。

[0075] 所描述的方法不取决于在所分析的上下文中实体的显式出现;附加外部概念的关联和评分经由图中的环境连接性是可能的。该方法允许间接相关实体的环境上下文影响对特定上下文中的关系的强度的评估,从而给出更全面的视图。

[0076] 该方法促进了详细的实体级关系和实体对之间的关系强度的分数,而不是整个上下文的简单聚合的“要点”分数。

[0077] 在若干实施例中,根据在处理了输入上下文114之后信号如何改变来更新156元关系权重116可能是有帮助的。这可以使用如图1A中所使用的导出103权重的方法。例如,作者和未列出主题之间的历史偏差现在可以由于使用输入上下文114的激活而不同。

[0078] 参照图2,框图以语义图评估系统210的形式示出了作为所描述的系统的示例实施例的系统200。

[0079] 语义图评估系统210包括至少一个处理器201、硬件模块或用于执行所描述的组件的功能的电路,其可以是在至少一个处理器上执行的软件单元。可以提供运行并行处理线程的多个处理器,使得能够并行处理组件的一些或所有功能。存储器202可以被配置为向至少一个处理器201提供计算机指令203以执行组件的功能。

[0080] 语义图评估系统210可以作为语义图创建系统(未示出)的一部分提供,或者作为

独立系统提供,用于通过基于语义资源110向语义图111添加要评估的现象的元关系权重113来自适应评估语义图中的元关系。语义图评估系统210可以使用与被评估的现象相关的附加资源112来为元关系播种权重值。可以在运行时提供输入上下文114,以便使用图激活来评估现象。

[0081] 语义图评估系统210可以包括语义图组件211,用于基于形成知识库的语义资源110来提供语义图,在知识库中,图节点形式的概念通过图边形式的语义关系来链接。

[0082] 语义图评估系统210可包括元关系组件220,用于在语义图111的边和节点的元数据中编码元关系的强度的权重。元关系可以应用于语义图111的概念,并且可以独立于由语义图的边定义的语义关系。

[0083] 元关系组件220可包括用于选择元关系权重可基于的附加资源112的附加资源组件221。元关系组件220还可包括种子权重导出组件222,其用于导出将用于从附加资源112导出的元关系的种子权重。元关系组件220还可包括权重添加组件223,用于以从独立于语义图111所基于的知识库的附加资源112的集合获得的元关系的强度的权重来播种语义图。

[0084] 元关系组件220可包括用于响应于对附加资源112或输入上下文114的添加来更新权重的权重更新组件224。

[0085] 语义图评估系统210可以包括用于执行与语义图的一个或多个概念相关的输入上下文114的图激活的运行时组件230。可以通过语义图111将权重应用于扩散激活信号,以产生语义图的概念的子集的元关系的强度的度量。

[0086] 运行时组件230可包括用于选择要评估的输入上下文114的输入上下文选择组件231。运行时组件230可包括用于发现输入上下文114中的概念的实例的概念检测组件232。

[0087] 运行时组件230可包括元关系的图激活组件233,用于激活对应于语义图111中的概念的节点,将信号向外遍历到相邻节点,从而依次激活这些节点,同时将权重应用于该信号,并确定具有最高结果激活信号的一个或多个焦点节点。

[0088] 运行时组件230可包括用于响应于图激活来对节点评分的评分组件234以及用于所激活的子图的环境上下文组件235,其中节点和边上的激活权重表示元关系的强度的程度。

[0089] 运行时组件230可包括输出组件237,用于输出反映输入上下文中元关系的强度的语义图的结果激活部分。

[0090] 在以下部分中描述了评估元关系的不同现象的各种示例。

[0091] 临床试验匹配

[0092] 在示例实施例中,可以从用于医学领域概念(诸如:“肿瘤”、“孕酮”或“淋巴结”等)的统一医学语言系统(UMLS)数据来构建语义图或者使用现有的图。

[0093] 结合该语义图,成功的试验候选的注释患者记录的语料库可以用于配置针对该任务的图激活的权重。或者,可以以类似的方式使用源自医学专家的用于试验中的验收标准的一组硬编码值或值范围。

[0094] 节点权重可以被配置用于节点与临床试验的相关性作为特征向量。特征向量将特定患者对节点的输入作为输入。

[0095] 在该示例中,最终分数编码患者与临床试验的要求匹配的程度。图中的节点和边可以是医学领域概念,诸如:“肿瘤”或“孕酮”或“淋巴结”等。这些概念中的每一个都可以具

有不同的相关性,这取决于被测试的药物,并且匹配的强度是该先前硬编码的相关性连同针对该概念所评估的患者正示出的值的函数。

[0096] 因此,特征向量对于每个节点可以是不同的。以下是一些示例:

[0097] a) “TumourSize”节点具有硬编码的相关因子0.9,因此当患者进入并具有该概念的相关值时,该向量可以包含归一化值(患者肿瘤大小*0.9);

[0098] b) 如果“孕酮水平”结点与该药物的相关性为0.8,则该向量含有(患者的孕酮水平*0.8);

[0099] c) “Ki67增殖指数”的节点可具有标准化的百分比($n\%*0.7$)。

[0100] 然后,使用不可用的特征的默认值,使用患者的可用值来激活该图,并且每个重要概念上的信号净累积反映了该患者的试验的一般适用性。

[0101] 情感分析

[0102] 情感分析涉及提取主观意见并将其分类为正面、负面或中性,或者扩展这种能力以提供某种数值表示。通常,在文档级别应用情感分析,并且将文档的聚集极性应用于讨论主题。短语级情感检测越来越多地被应用于主题的特定元素,被称为“层面情绪分析”。在这种情况下,针对整个主题的每个子元素,在文档或语料库中聚集单独的分数。该过程的最困难部分之一是识别单词的固有情感在上下文中何时改变,诸如在短语“阴性测试结果”对比“阴性体验”中。

[0103] 情感分析是可以使用语义图使用所描述的方法来应用和评估元关系的现象的示例。

[0104] 情感分析评价的例子:

[0105] 在一个示例实施例中,可以从统一医学语言系统(UMLS)本体数据来构建语义图或者使用现有的图。

[0106] 附加的医学文档可以用作与待分析的情感有关的附加资源。可以用语义图的相同概念来注释医学文档。命名实体短语可以与在不同上下文中具有固有情感极性的情感短语,诸如“强烈增加”,相关联。对于语料库或附加资源中的每个文档,使用原始值或向量来为情感配置节点权重。

[0107] 图3是从语义图中提取300的示意图,其示出了情感分析的元关系权重。语义图提取300示出了与UMLS概念相关的节点310,其中边320与节点310之间的语义关系相关。例如,“药物X”的节点311具有“产生效果”与“寿命增加”的节点312的关系321。

[0108] 情感的元关系权重的数值被示为节点权重330和边权重340。

[0109] 在运行时使用具有情感权重330、340的语义图300来检测输入情感短语的正确解释,类似于词义消歧(WSD)任务。

[0110] 示例输入上下文是:“将成分A加入到药物X中强烈地增加了患者的寿命”:与“当与药物Z一起服用时,药物Y的毒性强烈地增加”。概念实体加下划线,并且情感以粗体示出。

[0111] 输入上下文用于激活图,使用下划线的概念实体在那些开始节点处激活语义图300,并且每个重要概念上的信号的净累积反映了该短语的情感。

[0112] 预测分析中的偏差/公平

[0113] 随着情感分析技术的使用变得更加普遍,情感评估中的系统偏差可能危害基于自然语言文本的自动情感分析的决策的完整性。例如,当出版了关于公司或其竞争对手之一

的文章并且该文章表达了关于这些参与者及其未来财务业绩或关于特定竞争产品的正面或负面意见时,关键是暴露任何固有偏差以便适当地处理文章。然后,可以从决策过程中排除该文章,或者可以在产生分数时考虑偏差的强度。最可信的文章是不在一个或另一个方向倾斜的那些文章。许多系统假设嘈杂语料库中的平均情感得分是中性的。

[0114] 检测这种偏差是复杂的过程,其应当依赖于背景信息,诸如同作者的先前文章,但是许多解决方案仅依赖于忽略其他上下文的用户独立或本地化方法。通过所述方法解决了以有意义地表达两个实体之间的偏差强度的方式来聚集该偏差。

[0115] 所描述的方法旨在产生与单独极性不同的偏置强度的平衡表示。度量作者相对于给定主题的公正性是评估数据源的可信性的关键使能因素。当A出版社发布关于P公司的正面新闻报道时,这是从中立的角度来看的,还是有其他既得利益者?当A出版社发表一篇关于Q公司(P公司的竞争对手)的负面文章时,它能被信任吗?检测偏差传统上可能导致内容被拒绝;然而,如果可以度量该偏差的程度和极性,则可以通过考虑该偏差的程度来使用该内容。

[0116] 所描述的方法提供了一种不仅检测偏差而且量化实体之间的关系的极性的强度的技术。这对于NLP系统是非常有用的。

[0117] 组织在其性能概要和展望的版本方面通常是乐观的。当计划的性能报告被发布时,它们应该被检查偏差。同样,当竞争对手或既得利益者表达对相同公司或相关主题的意见时,读者应当同样警惕。各种新闻机构和投资机构也公布其对组织的季度/年度报告的评估,并且其中许多不同于组织的自我评估。在这样的情况下,投资者或观察者能够得出真实的或加权的平均结论是非常有帮助的。为了实现这一点,所描述的方法考虑了由组织和机构之间的关系、所有权或历史交易等引起的偏差,并且能够量化该偏差。

[0118] 对于自动股票研究系统,偏差检测和分析是重要的任务。选择用于上下文的相关文章(例如,新药、金融产品、规则等)。在聚焦项目中的置信度是在那些文章的上下文中意见/语义和检测到的偏差的函数。例如,专利过期或政府规章的改变可能对生态系统中的不同实体有用或起反作用,这取决于它们的商业地位。

[0119] 在医疗领域中,有价值的专利的到期对于拥有的制药公司是商业损失;然而,对于想要制造这种配方仿制药的竞争对手以及需要更便宜的成本的患者来说,这是一种收益。同样,政府的贸易保护主义关税制度可以使当地生产者受益,并损害进口商。新闻媒体或新闻社的观点可以突出显示一侧而不是另一侧,或者保持中立,而股权投资者则需要工具来获得对该事件的平衡了解并做出购买/出售/持有决定。

[0120] 预测分析中的偏差是现象的示例,对于该现象,可以使用所描述的方法使用语义图来应用和评估元关系。

[0121] 偏差评价的例子:

[0122] 在一个示例实施例中,语义图可以使用命名实体识别(NER)数据并解析诸如以下数据的数据上的帧来从金融领域语料库中构建:

[0123] a) 具有数字细节和文本概要的年度/季度报告以及短期未来的前景;

[0124] b) 金融新闻报道;

[0125] c) 金融博客;

[0126] d) 所有上述内容的元数据,包括文章作者、出版细节等。

[0127] 可以对语料库中的每个句子执行短语级情感分析,并且为文本中出现的图中的每对实体产生总得分。这构成了该领域中的当前公平前景。实体之间的边权重可以用于描述来自先前关联的极性(正和负)的程度。在动态处理系统中,这些边权重可以保留该极性的最新整体状态,作为公平的概率。

[0128] 由于偏差可以是正的、负的或中性的,因此这里的特征向量可以包括一个或几个带符号的值(-/+)以表示两个实体之间的净极性和个体极性分数。

[0129] 可以通过检测输入上的相同实体和元数据,并使用权重在图中执行扩散激活,来处理新文章或上下文。信号将根据边权重的强度而遍历该图,并且它们的极性将对使用特定起始节点的每个节点上的总体总和有贡献,这对于该激活是特别的。

[0130] 每个节点上产生的信号累积值反映了影响这些实体的断言是否可信/公正,或者这些实体上的结论和断言是否可信。

[0131] 图4A和4B是来自语义图400的摘录的示意图,其示出了偏差强度和极性的元关系权重。

[0132] 图400包括实体的节点410,其中边420用于实体之间的语义关系。元关系权重430的数值被显示在边上。

[0133] 图4A中的节点411、412、413反映了新的未见过的文档的内容,其中[作者A]411是作者,主题是[公司P报告]413和[公司Q]412。

[0134] 以下是反映用于为图400的激活播种的值的场景。

[0135] 作者A的文章陈述了以下内容:

[0136] “公司P的有关财务前景新报告充满了不能反映现实的不精确性和前景不明显的预测。与CEO决定在最新技术中投资良好的公司Q不同,公司P不能声称在正确的方向发展,并且该报告简单地看起来像试图掩盖其日益糟糕的业绩中的缺陷。”

[0137] 从以上这篇文章的情感分析得到的输入是:

[0138] 作者A:权重+1,其被添加为节点权重421;

[0139] 公司P报告:权重-1,其作为节点权重423被添加;以及

[0140] 公司Q:权重+1,其作为节点权重422被添加。

[0141] 在图400中,作者A的先前意见正在影响他/她所说的解释。净结果表明对某些主题的意见不应被信任。固有偏差450的极性在现有的边权重451、452、453中示出。所表示的偏差460的极性在边权重461、462、463中示出。

[0142] 某些关系表现出固有的偏差。关于同一公司的公司创作的文章具有固有的偏差,因为该公司不太可能将其自身呈现在负面报道下。出于工作安全的原因,出版物所使用的作者具有对该出版物的一般情况和观点的固有偏差。公司还具有关于其竞争者的固有偏差。

[0143] 表达的偏差是对任何极性的情感的一些发布实例对主题或实体的反映,诸如发布文章的示例场景。

[0144] 在输入(例如,以上文章)之后,信号在加权的边上扩散。一旦完成了该过程,就可以分析激活以查看信号如何在输入节点处累积以及图中的其他节点如何碰巧被加权,以推断出在输入中发生的节点与图中的任何其他节点之间的偏差。

[0145] 当前文章的图激活计算扩散激活值471、472、473、474、475,并且导致以下偏差结

果:

[0146] 在[公司P]的[作者A] = $(-0.9) + (-1.3) + (-1.45) = (-3.65)$, 表示为偏差结果474。

[0147] 在[公司Q]的[作者A] = $(+1) + (0.7) = (+1.7)$, 显示为偏差结果473。

[0148] 在单个边上, 进入节点, 每个值是来自起始节点的值乘以边上的值的结果, 其中边存储在先偏差(和极性), 并且节点权重是来自移动通过图的信号的净累积值。

[0149] 对上述实施例的一些进一步的评论:

[0150] 表示作者和出版物的实体可以具有先前文章和出版社等的历史, 其可以具有可以针对内容中的概念之间的断言的极性而分析的内容。实例包括:

[0151] “A公司将破产”, 或

[0152] “B公司正在为人工智能开辟道路”, 或者

[0153] “对A公司的财务预测显然更糟”。

[0154] 可以使用诸如语气分析、情感分析、社交网络元数据(先前雇佣、客户历史等)以及常见主题和话题的标记之类的技术来从这样的内容提取附加特征。

[0155] 虽然情感和意见或偏差通常可以用数字术语来量化, 但是有趣的是, 将公正的概念瞄准在看似真实的并且没有任何固有极性的陈述周围, 但是基于上下文中这些实体之间的环境关系, 这些陈述实际上是有问题的。例如, 当公司与公司P签订合同并以可能潜在地影响公司P的竞争者的方式发布涉及公司P的竞争者的文章时, 能够检测并质疑这种断言的公正性是极其重要的, 语义图的环境性质实际上鼓励了这种公正性。

[0156] 当包含与图中的实体相关联的引用或元数据的新数据被释放时, 可以实现跟踪直接链接的实体对之间的偏差程度的后端图监视器。检测哪些边应当被更新可以以各种方式完成, 包括:

[0157] a) 不同范围内的共同出现的实体, 诸如句子、名词短语、段落、文档或甚至完整的出版物;

[0158] b) 由依赖关系解析结构链接的实体, 例如父节点->子节点, 或者作为相同动词的参数(arguments); 以及

[0159] c) 通过在文本中具有独立引用的特定关系链接的实体, 或者直接链接解析树中的领域和范围实体, 如在谓词(subj, obj)中。

[0160] 作为两个实体之间的边权重而存储的极性公式可以是所有先前正和负偏差的函数, 但像总和这样简单的东西就行了。考虑到特定功能的实用性超过其他功能的统计技术将有助于提供更强大且对域敏感的权重。

[0161] 在例如通过传统搜索/IR、权益研究焦点等识别了答案语料库的结果集之后, 相应地忽略具有太多偏差或重新加权的文档可能是适当的。

[0162] 从一个实体到另一个实体的偏差是正或负(其中中性偏差与该情形无关)。当分析两个实体之间的偏差时, 例如作者和主题之间的偏差、先前偏差的极性和文档内容的极性, 可以根据该数据的对齐做出不同的决定。

[0163] 下面是演示这可能如何发生的示例表格, 其中应将相同的符号偏差和情感标记为“有趣”或“不可信任”, 并应重新加权相反的符号偏差和情感。

[0164]		偏		
		差	正	负
	情			
	感			
	正		0*	正**
	负		负	0

[0165] 例如,如果对主题具有正偏差的作者说某物是正的,则它不是那么值得信任,但是如果他们说某物是负的,则它可能比无偏差作者的观点更值得注意(并且对于极性的反转,反之亦然)。

[0166] 因此,在上表中:

[0167] *0应被认为是不值得信任的,并且可以忽略;

[0168] **正/负是不期望的,并且因此具有附加值;以及

[0169] 中性值可以被认为等效的。

[0170] AlchemyLangage(国际商业机器公司的商标)是通过NLP提供文本分析的应用编程接口的集合。进一步的实现可以对AlchemyLangage对实体的意见进行情感分析。

[0171] 使用信息检索统计的查询扩展

[0172] 除了可能或多或少与语义图的内容相关的语义类别和实体的硬编码权重之外,还有许多其他方式可以将语义相关性编码为语义图中的扩散信号。所描述的方法的实施例可以被提供为使用来自域特定语料库的信息检索(IR)频率来对语义相关性的扩散激活进行自动配置,以对图形节点和链接进行加权。

[0173] 另一种查看语义相关性的方式是与特定内容主体的关联强度。因此,通过关联强度的倒数来重新加权语义图将促进较不常见的那些概念。这在针对新的域适配语义图时可以是非常有帮助的。在用于构造原始语义网络的初始域中,用于激活的默认权重可能是足够的。然而,关于特定概念的相关性,在该网络中存在固有的偏差。即使使用诸如术语频率和共现频率等传统IR统计也可帮助适应该图中的扩散信号,以便在新的域中产生更有意义的输出。

[0174] 附加语义评估的示例:

[0175] 在示例实施例中,可以使用现有语义图;例如,UMLS-医学域图或DBPedia-开放域图。

[0176] 新域中的自然语言文本的语料库可以用作元关系的附加资源,并且可以收集该域中的实体和关系的语料库频率。

[0177] 值可以被聚合并作为节点和边的元关系的权重而存储在语义图中:

[0178] a) 节点频率(Fn)反映术语的一般性,因此将使用倒数1/Fn;

[0179] b) 边频率(Fe)与术语的相关性成正比,并且使用值Fe;

[0180] c) 在用于传播信号的递归递减的扩散激活中,被按比例缩小以适合0和1之间或

0.5和1之间的值是非常常见的；

[0181] d) 上述其它变化将包括使用节点和边的tf/idf分数。

[0182] 给定输入上下文,实体的实例可以在语义图中识别,图在那些点处被激活,并且在特定NLP任务的上下文中执行扩散激活,诸如字义消歧(WSD)、字义归纳(WSI)或查询扩展。

[0183] 输出与域的相关性比其他情况要大得多。

[0184] 查询扩展示例:

[0185] 查询扩展的过程可被认为是理解输入短语或查询的上下文,并找到应帮助找到原始查询的正确文档的替换或附加短语。一个例子是“公司A在爱尔兰的办公室在哪里?”将“爱尔兰”细化为公司A在其有办公室的“科克”、“都柏林”和“高威”,将潜在地提高找到正确答案的机会。查询扩展通常集中在输入的各方面,并试图外推以提高匹配的可能性。

[0186] 可以证明在查询扩展中应用该技术的效果。中心节点可以是查询扩展候选,并且周围节点可以来自输入上下文。在配置扩散激活以考虑信息检索统计之后,更一般概念的相关性分数变得足够低以将它们排除为与上下文无关。这种情况的影响在于,查询扩展的候选的等级也改变。当按相互相关的相关输入的数目进行排名时,可以获得新的排名最高的查询扩展候选者。

[0187] 获得用于输入“公共记录”、“法律”、“信息”、“政治细分”等的附加有用查询术语的初始尝试可以给出中心节点的最佳候选作为“执法机构”。在根据域语料库中的节点的频率对图进行加权之后,原始输入中的若干输入可能不再被认为是合适的输入,并且术语“机密信息”可以被替代地提议为中心节点。假定原始上下文在“数据隐私”的领域中,这是一个大的改进。

[0188] 自动化风险评估

[0189] 自动风险评估领域主要依赖于基于关键字的方法,其中在给定关键字和特定风险暴露水平之间存在直接链接。机器学习技术通常用于基于这些关键词及其相关联的风险因子来给出结果的更稳健的识别。可以推断风险暴露并使用知识量化风险暴露的方法(其可能不直接存在于上下文中)将对现有技术具有显著的附加值。

[0190] 暴露于风险的程度或强度是可以以与情感分析或地理空间推断类似的方式相对于特定风险因子量化的某种程度。利用相同技术来执行风险评估的方法的另一实施例被预期。使用NLP技术对该问题的现有解决方案涉及将特定关键词或短语链接到风险,类似于基于词典的情感分析。能够以更接近环境的方式推断风险暴露的系统在产生针对上下文中的风险暴露的强度度量方面将是极其有价值的。

[0191] 以医学领域为例,用于评估患者10年心血管疾病发展风险的Framingham风险评估工具的细节可以编码成图结构。当在图中执行扩散激活时,在每个相关命名实体上的信号传播强度与对患者的风险水平成正比。

[0192] 地理空间推断

[0193] 同样,可以根据历史情感极性的强度来量化偏差,可以根据历史关联和地理距离来评估实体或主题与地理地点列表的相关性。用于地理空间推断的现有方法限于文本中的直接提及,但是当文本中存在多个位置时,可以通过域语料库中的通用关联来确定相关性的方法更有价值。

[0194] 地理相关性的元关系权重的特征向量可以包括以下样本特征的组合:

[0195] a) 两个实体之间的标准化地理距离；

[0196] b) 语义关联相关性(例如,来自原始扩散激活实现的权重)；

[0197] c) 对b)的一种替代,其中只有地理空间语义类别被遍历并用于权重累加。

[0198] 例如,考虑加利福尼亚法令中的法律,该法律使用CA立法固有的许多词汇,但是引用其它州的法律或者出于讨论法律的目的(例如基于美国和欧盟的公司之间的客户数据传输)而通过名称提及不在加利福尼亚的特定地方。能够推断加利福尼亚在文本中明确提及的那些位置上的相关性的环境技术对于法律领域NLP任务将是极其有用的。

[0199] 现在参考图5,示出了计算机系统或服务器形式的系统500的示例的示意图。

[0200] 计算机系统/服务器512,其可与众多其它通用或专用计算系统环境或配置一起操作。众所周知,适于与计算机系统/服务器512一起操作的计算系统、环境和/或配置的例子包括但不限于:个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人电脑、小型计算机系统、大型计算机系统和包括上述任意系统的分布式云计算技术环境,等等。

[0201] 计算机系统/服务器512可以在由计算机系统执行的计算机系统可执行指令(诸如程序模块)的一般语境下描述。通常,程序模块可以包括执行特定的任务或者实现特定的抽象数据类型的例程、程序、目标程序、组件、逻辑、数据结构等。计算机系统/服务器512可以在通过通信网络链接的远程处理设备执行任务的分布式云计算环境中实施。在分布式云计算环境中,程序模块可以位于包括存储设备的本地或远程计算系统存储介质上。

[0202] 如图5所示,计算机系统/服务器512以通用计算设备的形式表现。计算机系统/服务器512的组件可以包括但不限于:一个或者多个处理器或者处理单元516,系统存储器528,连接不同系统组件(包括系统存储器528和处理单元516)的总线518。

[0203] 总线518表示几类总线结构中的一种或多种,包括存储器总线或者存储器控制器,外围总线,图形加速端口,处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说,这些体系结构包括但不限于工业标准体系结构(ISA)总线,微通道体系结构(MAC)总线,增强型ISA总线、视频电子标准协会(VESA)局域总线以及外围组件互连(PCI)总线。

[0204] 计算机系统/服务器512典型地包括多种计算机系统可读介质。这些介质可以是能够被计算机系统/服务器512访问的任意可获得的介质,包括易失性和非易失性介质,可移动的和不可移动的介质。

[0205] 系统存储器528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) 530和/或高速缓存存储器532。计算机系统/服务器512可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例,存储系统534可以用于读写不可移动的、非易失性磁介质(图1未显示,通常称为“硬盘驱动器”)。尽管图1中未示出,可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器,以及对可移动非易失性光盘(例如CD-ROM,DVD-ROM或其它光介质)读写的光盘驱动器。在这些情况下,每个驱动器可以通过一个或者多个数据介质接口与总线518相连。存储器528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)程序模块,这些程序模块被配置以执行本发明各实施例的功能。

[0206] 具有一组(至少一个)程序模块542的程序/实用工具540,可以存储在存储器528中,这样的程序模块42包括但不限于操作系统、一个或者多个应用程序、其它程序模块以及

程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块542通常执行本发明所描述的实施例中的功能和/或方法。

[0207] 计算机系统/服务器512也可以与一个或多个外部设备514(例如键盘、指向设备、显示器524等)通信,还可与一个或者多个使得用户能与该计算机系统/服务器512交互的设备通信,和/或与使得该计算机系统/服务器512能与一个或多个其它计算设备进行通信的任何设备(例如网卡,调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口522进行。并且,计算机系统/服务器512还可以通过网络适配器520与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图所示,网络适配器520通过总线518与计算机系统/服务器512的其它模块通信。应当明白,尽管图中未示出,其它硬件和/或软件模块可以与计算机系统/服务器512一起操作,包括但不限于:微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0208] 在任何可能的技术细节结合层面,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。

[0209] 计算机可读存储介质可以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——但不限于——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。

[0210] 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。

[0211] 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。

[0212] 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。

[0213] 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造品,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。

[0214] 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。

[0215] 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。

[0216] 云计算

[0217] 应当理解,尽管本公开包括关于云计算的详细描述,但是本文所陈述的教导的实现不限于云计算环境。相反,本发明的实施例能够结合现在已知或以后开发的任何其它类型的计算环境来实现。

[0218] 云计算是一种服务递送模型,用于实现对可配置计算资源(例如,网络、网络带宽、服务器、处理、存储器、存储、应用、虚拟机和服务)的共享池的方便的按需网络访问,所述可配置计算资源可以以最小的管理努力或与服务的提供者的交互来快速供应和释放。该云模型可以包括至少五个特性、至少三个服务模型和至少四个部署模型。

[0219] 特征如下:

[0220] 按需自助服务:云消费者可以单方面地自动地根据需要提供计算能力,诸如服务器时间和网络存储,而不需要与服务的提供者进行人工交互。

[0221] 广域网接入:能力可在网络上获得,并且通过促进由异构的薄或厚客户端平台(例如,移动电话、膝上型计算机和PDA)使用的标准机制来访问。

[0222] 资源池化:供应商的计算资源被集中以使用多租户模型来服务多个消费者,其中

不同的物理和虚拟资源根据需求被动态地分配和重新分配。存在位置无关性的意义,因为消费者通常不控制或不知道所提供的资源的确切位置,但可能能够在较高抽象级(例如,国家、州或数据中心)指定位置。

[0223] 快速弹性:在一些情况下,可以快速且弹性地提供快速向外扩展的能力,以及快速向内扩展的能力。对于消费者,可用于提供的能力通常表现为无限的,并且可以在任何时间以任何数量购买。

[0224] 度量服务:云系统通过利用在适合于服务类型(例如,存储、处理、带宽和活动用户账户)的某一抽象级别的计量能力来自动地控制和优化资源使用。可以监视、控制和报告资源使用,从而为所利用服务的提供者和消费者两者提供透明性。

[0225] 服务模型如下:

[0226] 软件即服务(SaaS):提供给消费者的能力是使用在云基础设施上运行的提供者的应用。应用程序可通过诸如web浏览器(例如,基于web的电子邮件)等瘦客户机界面从各种客户机设备访问。消费者不管理或控制底层云基础结构,包括网络、服务器、操作系统、存储、或甚至单独的应用能力,可能的例外是有限的用户专用应用配置设置。

[0227] 平台即服务(PaaS):提供给消费者的能力是将消费者创建或获取的应用部署到云基础设施上,该消费者创建或获取的应用是使用由提供者支持的编程语言和工具创建的。消费者不管理或控制包括网络、服务器、操作系统或存储的底层云基础设施,但具有对部署的应用和可能的应用托管环境配置的控制。

[0228] 基础设施即服务(IaaS):提供给消费者的能力是提供处理、存储、网络和消费者能够部署和运行任意软件的其它基本计算资源,所述软件可以包括操作系统和应用。消费者不管理或控制底层云基础设施,但具有对操作系统、存储、部署的应用的控制,以及可能对选择的联网组件(例如,主机防火墙)的有限控制。

[0229] 部署模型如下:

[0230] 私有云:云基础设施仅为组织操作。它可以由组织或第三方管理,并且可以存在于场所内或场所外。

[0231] 社区云:云基础设施由若干组织共享,并且支持具有共享关注(例如,任务、安全要求、策略和合规性考虑)的特定社区。它可以由组织或第三方管理,并且可以存在于场所内或场所外。

[0232] 公有云:云基础设施可用于一般公众或大型产业群,并且由销售云服务的组织拥有。

[0233] 混合云:云基础设施是两个或更多个云(私有、共同体或公共)的组合,所述云保持唯一实体但通过使数据和应用能够移植的标准化或私有技术(例如,用于云之间的负载平衡的云突发)绑定在一起。

[0234] 云计算环境是面向服务的,其焦点在于无状态、低耦合、模块性和语义互操作性。云计算的核心是包括互连节点的网络的基础设施。

[0235] 现在参考图6,描绘了说明性云计算环境650。如图所示,云计算环境650包括云消费者使用的本地计算设备可以与其通信的一个或多个云计算节点610,本地计算设备诸如例如个人数字助理(PDA)或蜂窝电话654A、台式计算机654B、膝上型计算机654C和/或汽车计算机系统654N。节点610可以彼此通信。它们可以被物理地或虚拟地分组(未示出)在一个

或多个网络中,诸如上文描述的私有云、社区云、公共云或混合云或其组合。这允许云计算环境650提供基础设施、平台和/或软件作为服务,云消费者不需要为其维护本地计算设备上的资源。应当理解,图6中所示的计算设备654A-N的类型仅旨在说明,并且计算节点610和云计算环境650可以通过任何类型的网络 and/或网络可寻址连接(例如,使用web浏览器)与任何类型的计算机化设备通信。

[0236] 现在参考图7,示出了由云计算环境650(图6)提供的一组功能抽象层。应当预先理解,图7中所示的组件、层和功能仅旨在说明,并且本发明的实施例不限于此。如所描绘的,提供了以下层和相应的功能:

[0237] 硬件和软件层760包括硬件和软件组件。硬件组件的示例包括:主机761;基于RISC(精简指令集计算机)架构的服务器762;服务器763;刀片服务器764;存储装置765;以及网络和联网组件766。在一些实施例中,软件组件包括网络应用服务器软件767和数据库软件768。

[0238] 虚拟化层770提供抽象层,从该抽象层可以提供虚拟实体的以下示例:虚拟服务器771;虚拟存储772;虚拟网络773,包括虚拟专用网络;虚拟应用和操作系统774;以及虚拟客户机775。

[0239] 在一个示例中,管理层780可以提供以下描述的功能。资源供应781提供对被用来执行云计算环境内的任务的计算资源和其他资源的动态采购。计量和定价782提供了在云计算环境中利用资源时的成本跟踪,以及用于消耗这些资源的记帐或发票。在一个示例中,这些资源可以包括应用软件许可证。安全性为云消费者和任务提供身份验证,以及为数据和其他资源提供保护。用户门户783为消费者和系统管理员提供对云计算环境的访问。服务级别管理784提供云计算资源分配和管理,使得满足所需的服务级别。服务水平协议(SLA)规划和履行785提供对云计算资源的预安排和采购,其中根据SLA预期未来需求。

[0240] 工作负载层790提供了云计算环境可以被用于的功能的示例。可以从该层提供的工作负载和功能的示例包括:映射和导航791;软件开发和生命周期管理792;虚拟教室教育传送793;数据分析处理794;事务处理795;以及使用自然语言处理796对语义图中的元关系进行自适应评估。

[0241] 以上已经描述了本发明的各实施例,上述说明是示例性的,并非穷尽性的,并且也不限于所披露的各实施例。在不偏离所说明的各实施例的范围和精神的情况下,对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择,旨在最好地解释各实施例的原理、实际应用或对市场中的技术的技术改进,或者使本技术领域的其它普通技术人员能理解本文披露的各实施例。

[0242] 在不脱离本发明的范围的情况下,可以对前述内容进行改进和修改。

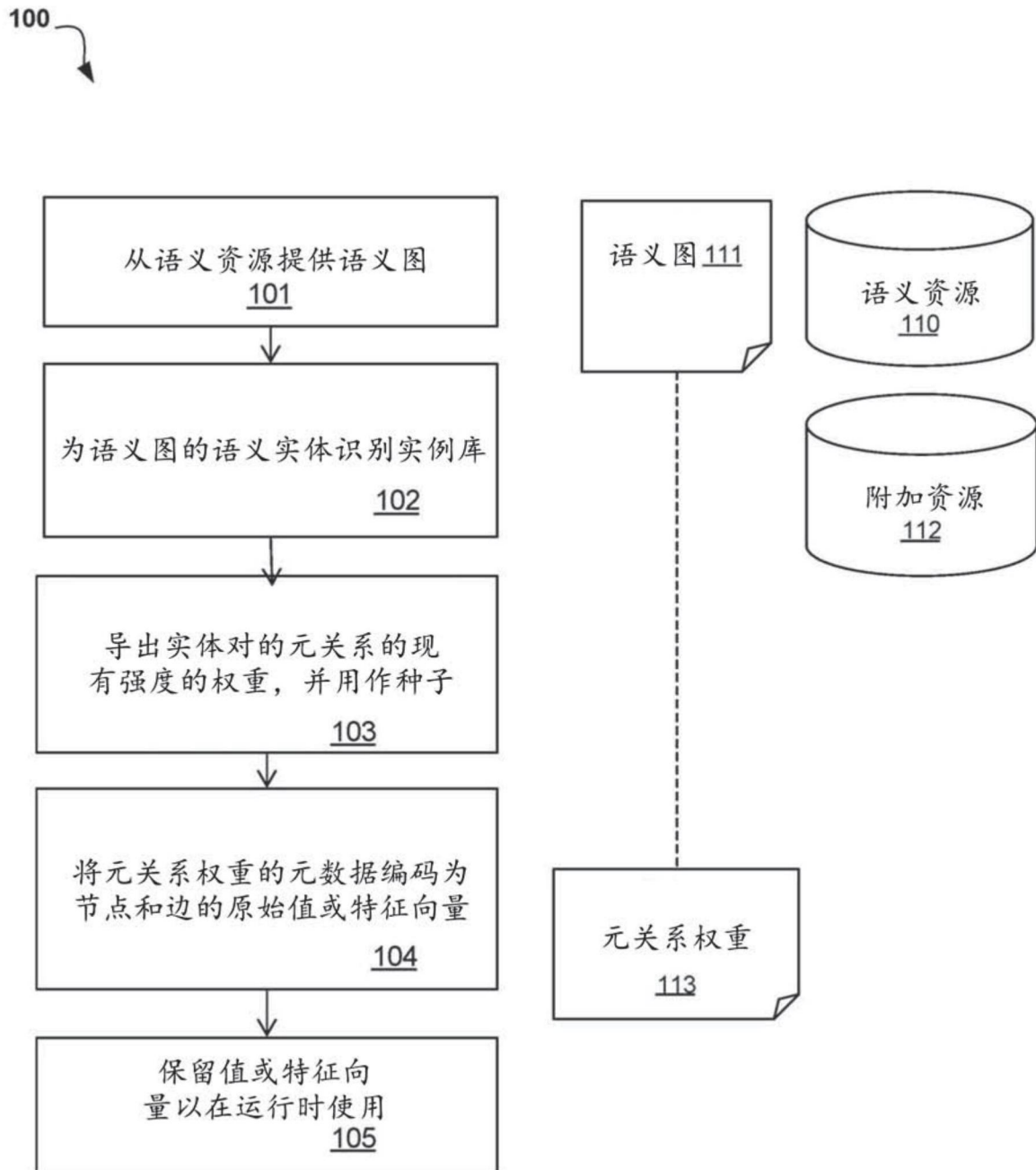


图1A

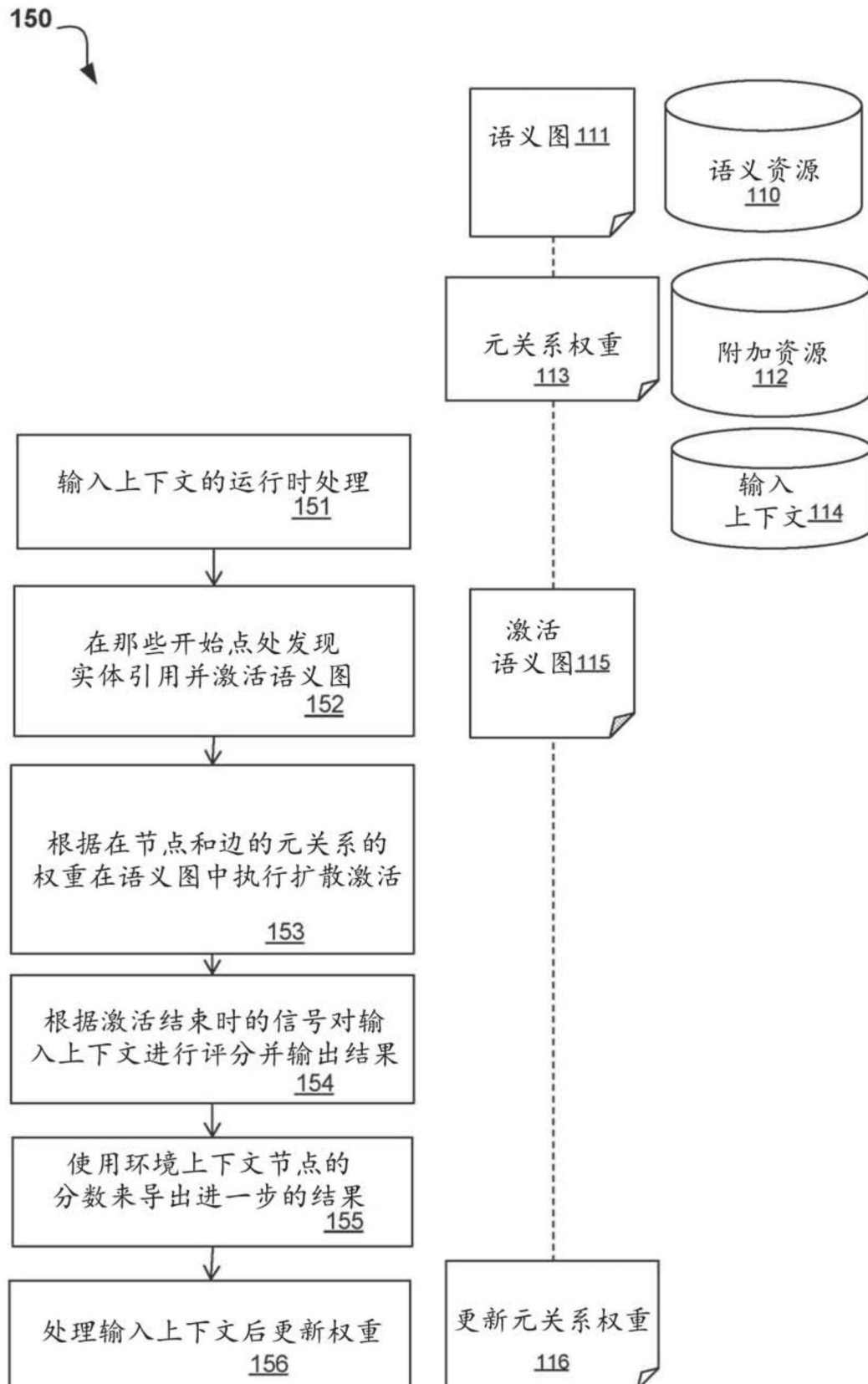


图1B

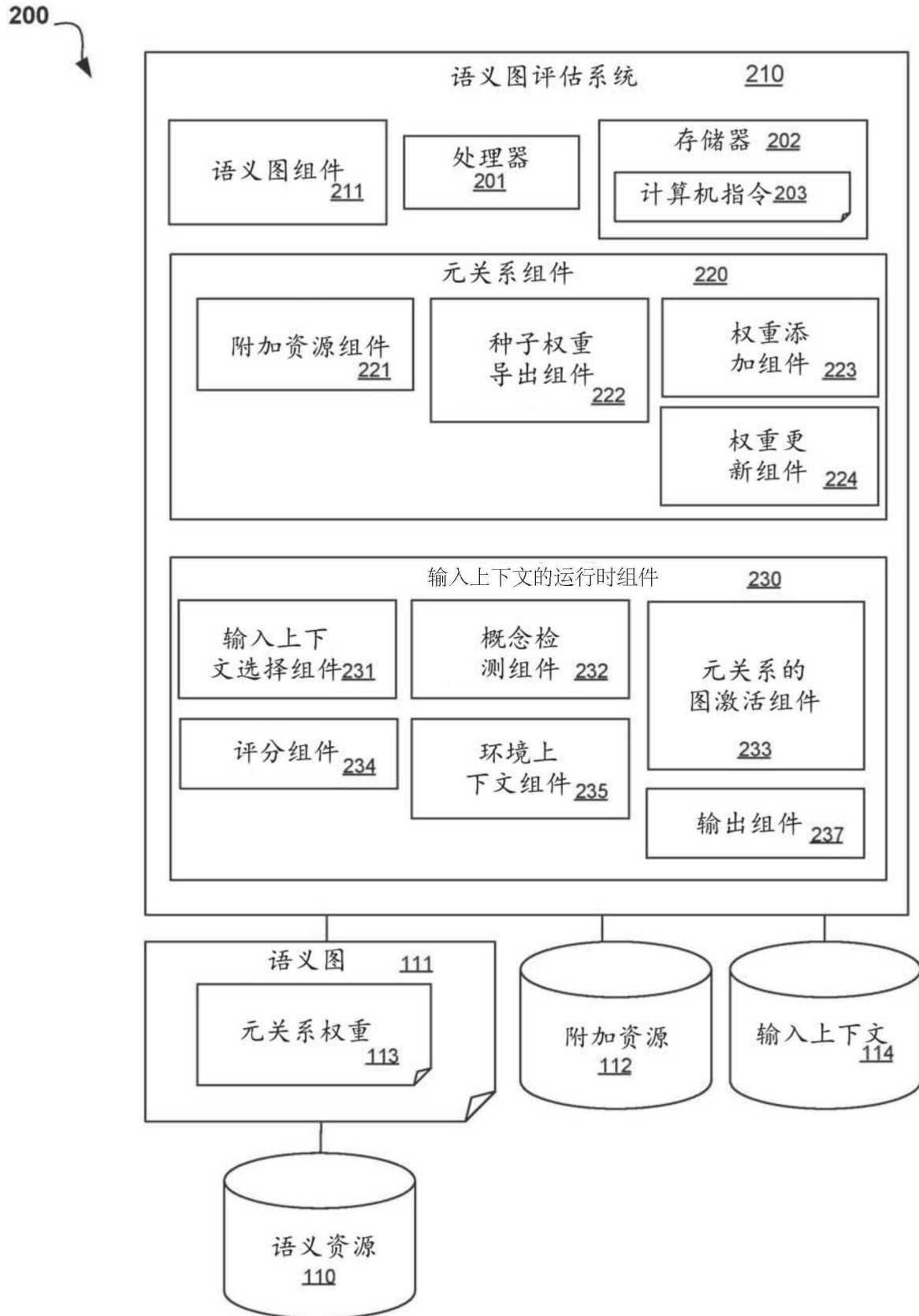


图2

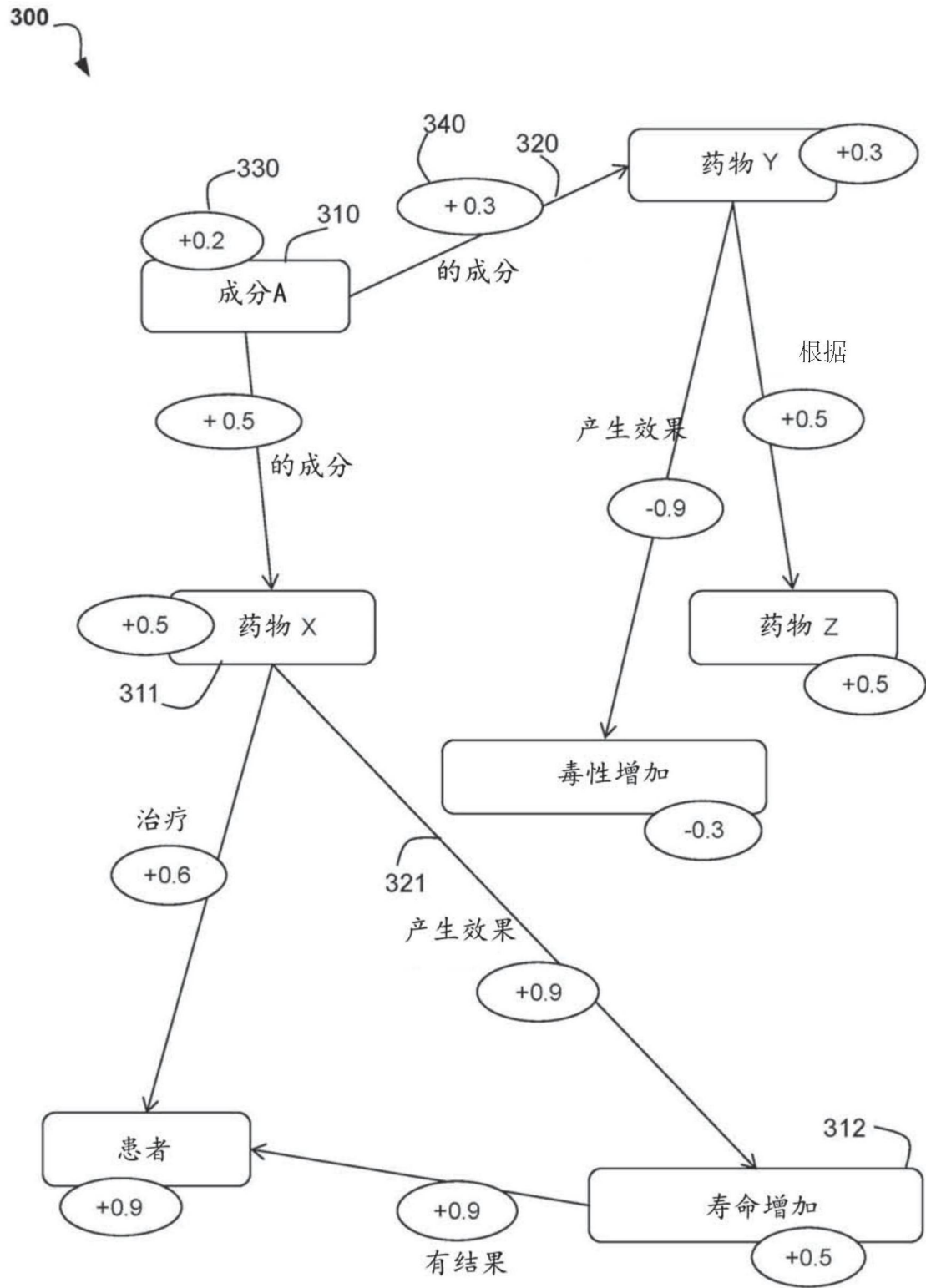


图3

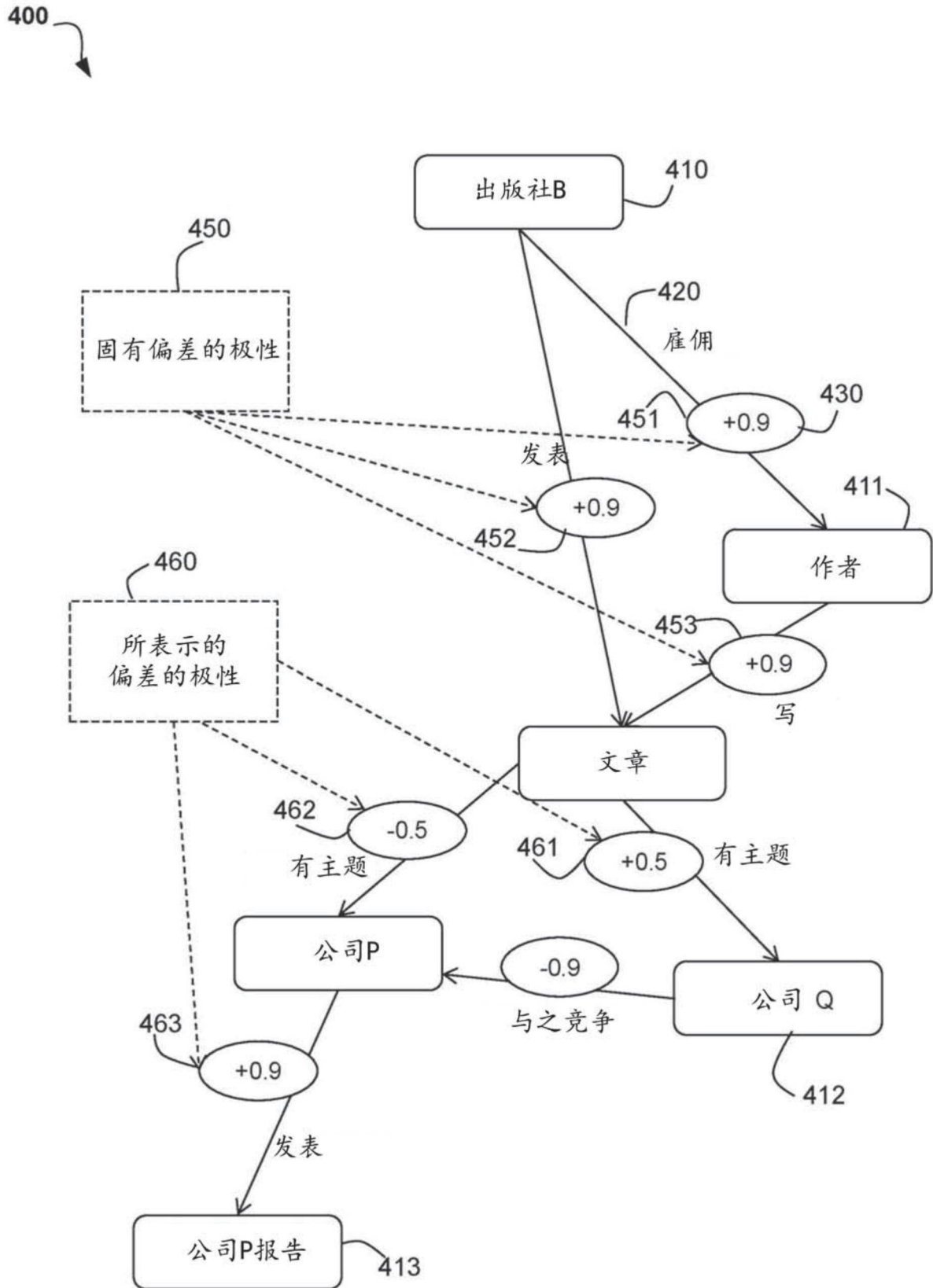


图4A

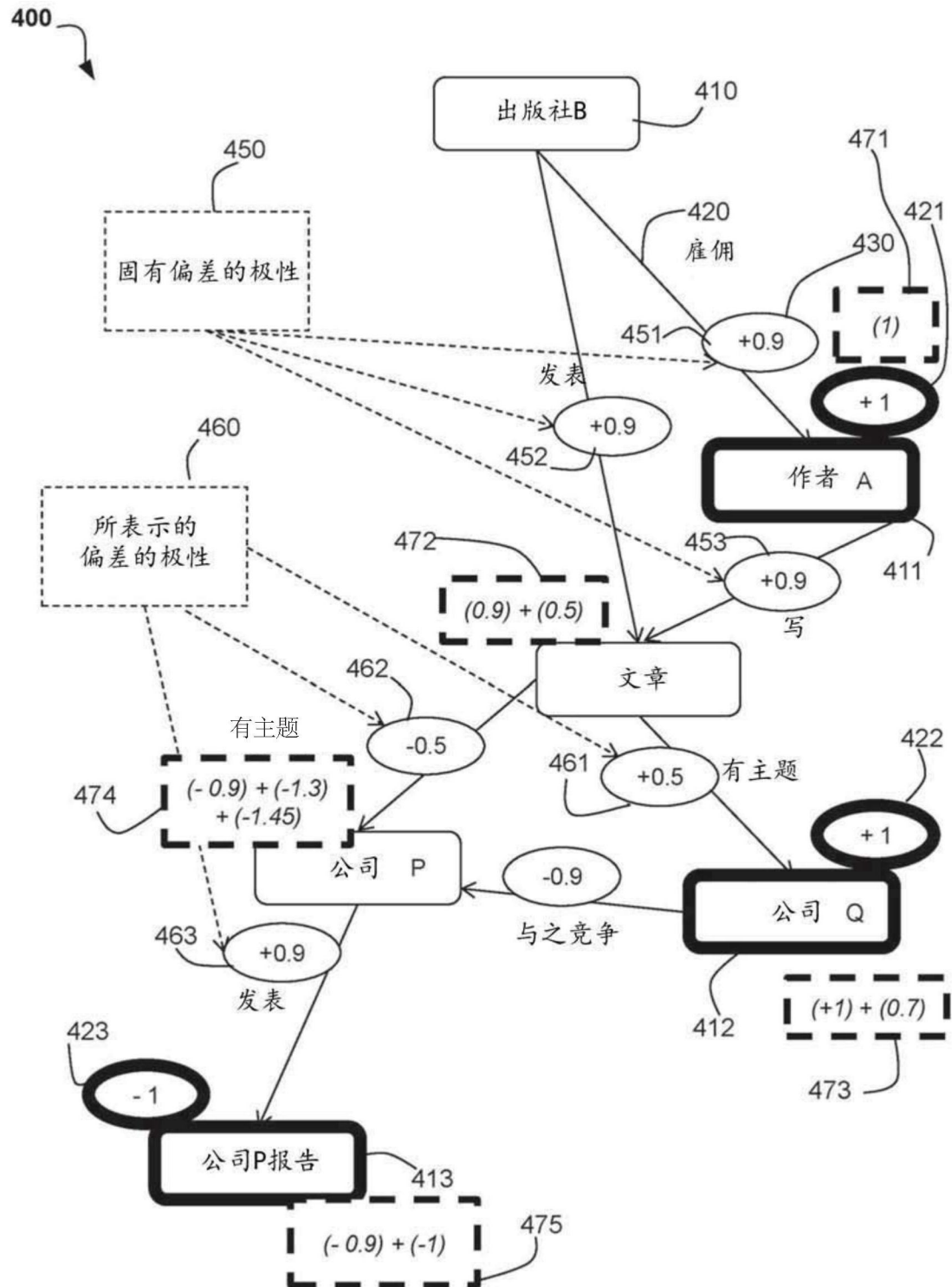


图4B

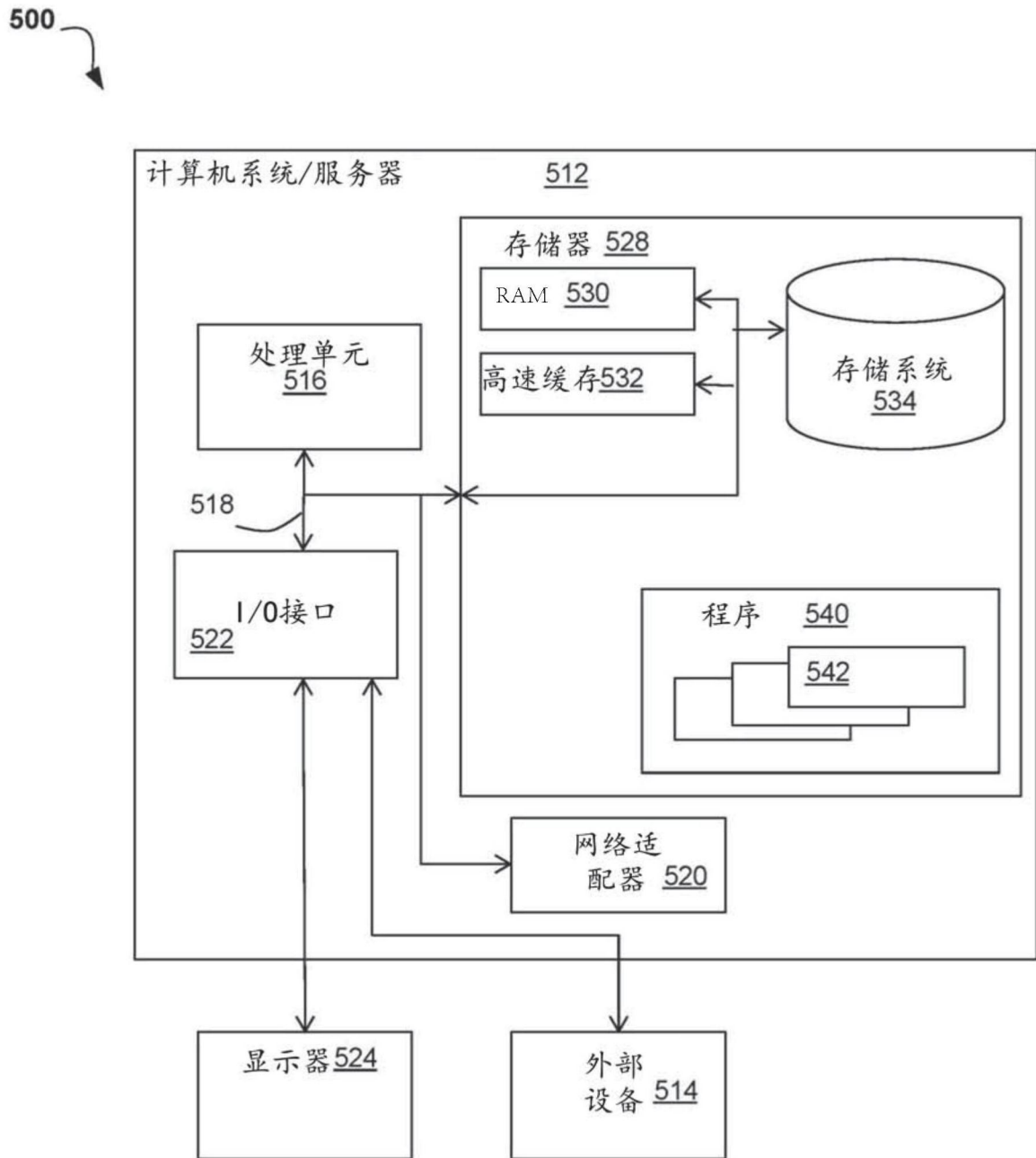


图5

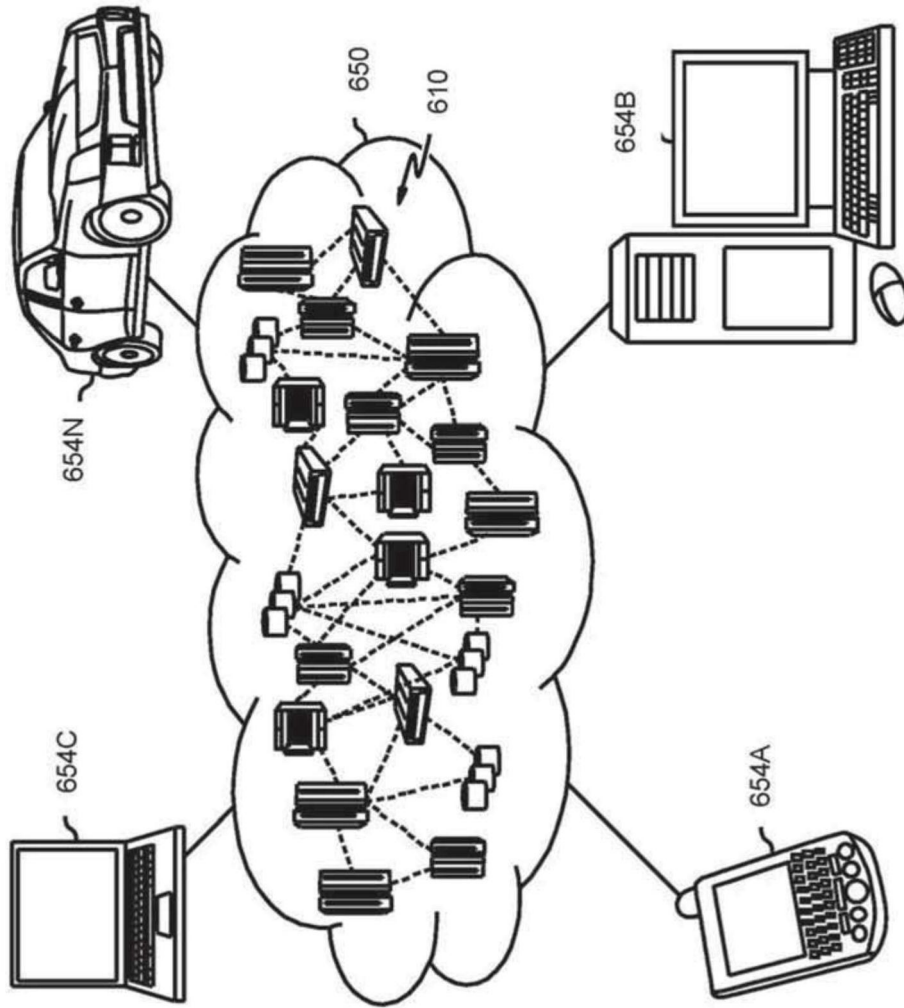


图6

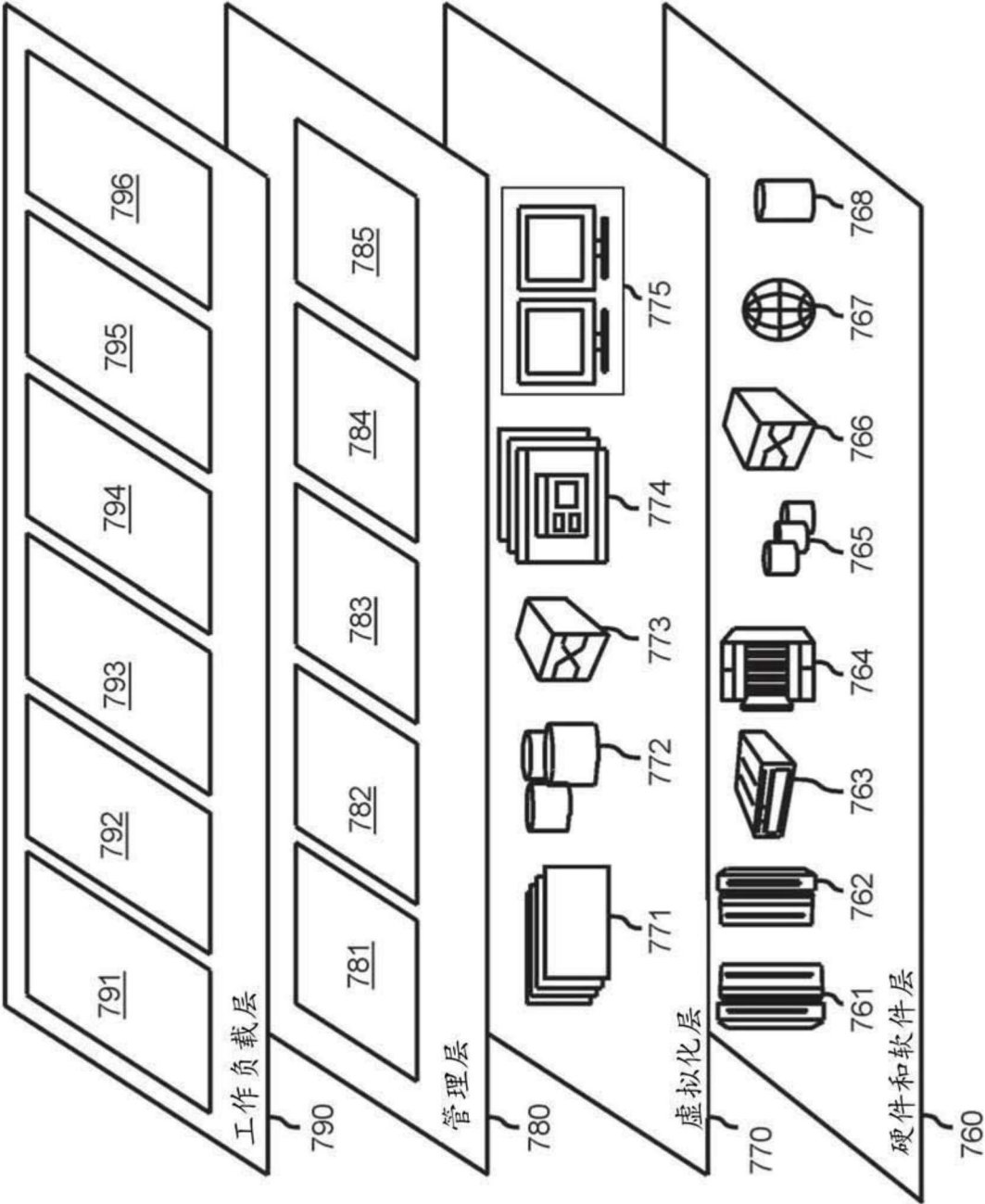


图7