5归纳性知识

虽然演绎知识的特点是精确的逻辑结果，但通过归纳获得知识涉及从一组给定的输入观察中概括模式，然后可以用于生成新颖但可能不精确的预测。例如，从大型数据图中通过地理和航班信息，我们可以观察到几乎所有省会城市的模式的国家有国际机场为它们服务，因此预测如果圣地亚哥是首都城市，它可能有一个国际机场服务它；然而，从这种模式得出的预测不确定，其中（例如）列支敦士登首都瓦杜兹没有（国际） 机场服务它。因此，预测通常与置信水平相关联；例如，可以说首都有187个国际机场 195个案例，提供了一个信心使用该模式进行的预测为 0.959。然后我们指的是归纳获得的知识作为归纳知识，它既包括用于编码模式的模型，也包括这些模型做出的预测。尽管容易出错，但归纳性知识可能非常有价值。

可以采用几种方法从图形中归纳地获取知识，基于有监督或无监督的方法。监督方法学习一个函数（又名模型）来映射一个给定一组标记输出的示例输入；然后可以将模型应用于未标记的输入来预测他们的输出。鉴于手动标记可能成本高昂，一些有监督的方法而是想办法从输入中自动生成输入-输出对，然后输入一个有监督的过程来学习一个模型；在这里，我们将这种方法称为自我监督。或者，无监督过程不需要标记的输入-输出对，而是应用一个预定义的函数（通常是统计性质的），用于将输入映射到输出。

在无监督方法的情况下，有大量关于图分析的工作，它使用众所周知的功能/算法来检测社区或集群，找到中心节点和边缘，在图形/网络中发现有趣的路径等。或者，知识图嵌入可用于学习（通常）映射的知识图谱的低维数值模型输入边缘到输出合理性分数，指示边缘为真的可能性；这样的嵌入可以使用正边缘和负边缘以自我监督的方式学习从知识图本身自动生成。而不是在平面中编码图形数值表示，图的结构可以直接用于监督学习，正如在图神经网络的上下文中所探索的那样，它学习应用的模型（可能递归地）在图的邻域上，通常为节点生成输出。最后，上述技术学习数值模型时，符号学习可以学习符号模型-即逻辑公式-从可以与演绎推理相结合的图表到作出预测。我们现在依次讨论上述四种技术中的每一种。

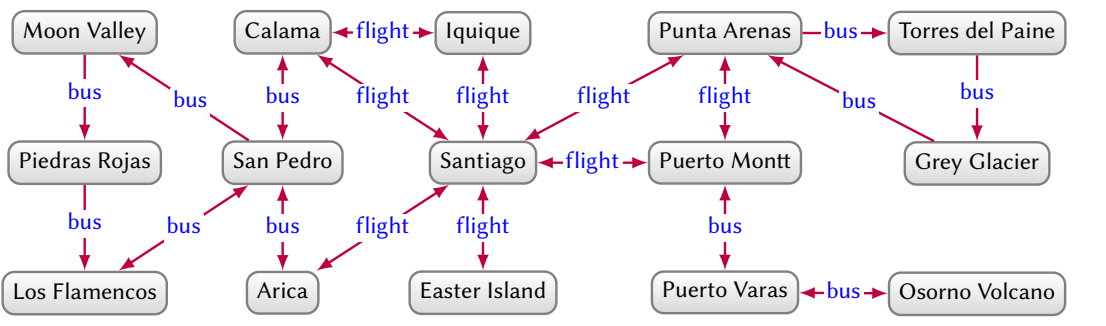


图 21 代表智利运输路线的数据图

5.1 图分析

分析是发现、解释和交流有意义的模式的过程 （通常是大型的）数据集合所固有的。 图分析是分析的应用 处理（通常是大的）图形数据。图的性质自然地适用于某些基于拓扑结构得出关于节点和边的结论的分析类型图，即图的节点如何连接。图形分析因此吸引了许多 图论和网络分析等相关领域的技术已被用于研究代表社交网络、网络、互联网路由、交通网络的图表，生态系统、蛋白质-蛋白质相互作用、语言共现等等[142]。

回到我们运行示例的领域,旅游局可以使用图形分析提取知识，例如：服务于许多旅游景点的主要交通枢纽（中心性）；同一游客参观的景点分组（社区检测）；景点在发生罢工或其他路由故障（连接）时可能无法访问，或彼此相似的景点对（节点相似性）。鉴于此类分析将需要一个复杂的大比例图，为了便于说明，在图 21 中我们展示了一个更智利一些针对热门旅游目的地的交通连接的简明示例目的地。我们首先介绍了一些可用于图形分析的关键技术。然后，我们讨论可用于在实践中计算此类分析的框架和语言。鉴于许多传统的图算法是为未标记图定义的，然后我们描述 在有向边标记图上应用分析的方式。最后我们讨论图分析与查询和推理之间的潜在联系。

5.1.1技术

可以将多种技术应用于图形分析。在下面的我们将列举一些主要的技术——例如，通过对 Iosup 的调查等[250] –可以在此设置中调用。

* 中心性：旨在识别图的最重要（也称为中心）节点或边。具体的节点中心性度量包括度、介数、接近度、特征向量、PageRank、HITS、卡茨等。中介中心性也可以应用于边缘。节点中心性措施将允许，例如，预测图 21 中的交通枢纽，而边缘中心性将使我们能够找到许多最短路线所依赖的边缘来预测交通。
* 社区检测：旨在识别图中的社区，即子图与图的其余部分相比，内部连接更密集。社区检测 算法，例如最小割算法、标签传播、Louvain模块化等。能够发现这样的社区，应用于图 21 的社区检测可能，对于例如，检测左侧（指智利北部）、右侧（指 到智利南部），也许还有中心（指有机场的城市）。
* 连接性：旨在估计图的连接程度，例如，揭示图元素的弹性和（不）可达性。具体技术包括测量图密度或k-连通性，检测强连通分量和弱连通分量组件、计算生成树或最小割等。在图 21 的上下文中，这样的分析可能会告诉我们，通往奥索尔诺火山和彼德拉斯罗哈斯的路线是最“脆弱”的，如果两条公交路线之一出现故障，则与主要集线器断开连接。
* 节点相似度：旨在通过它们的方式找到与其他节点相似的节点在他们的社区内相连。节点相似性度量可以使用结构等价、随机游走、扩散核等来计算。这些方法提供了一个了解连接节点的内容，以及之后它们在哪些方面相似。在里面在图 21 的上下文中，这样的分析可以告诉我们 Calama 和 Arica 是基于相似节点的都有返回圣地亚哥的航班和返回圣佩德罗的巴士。

虽然以前的技术只接受一个图作为输入18 其他形式的图分析可能进一步接受一个节点，一对节点等，以及图。

* 路径查找：旨在查找图中的路径，通常在作为输入的节点对之间。存在各种限制此类节点之间有效路径集的技术定义，包括不两次访问同一节点的简单路径、访问同一节点的最短路径 最少的边数，或者——如之前在第 2.2 节中讨论的——常规路径查询限制路径 [14] 可以遍历的边的标签。 我们可以使用这样的算法例如，查找图 21 中从 Torres del Paine 到 Moon Valley 的最短路径。

大多数此类技术已被提出并研究用于简单图或有向图，而无需边缘标签。我们将讨论它们在更复杂的图形模型中的应用——以及它们如何与推理和查询等其他技术相结合——稍后在第 5.1.3 节中。

5.1.2框架

已经为大规模图分析提出了各种框架，通常在分布式（集群）环境中。其中，我们可以提到 Apache Spark (GraphX) [114, 529]，GraphLab [306]、Pregel [314]、Signal–Collect [478]、Shark [530]等。这些图并行框架应用基于有向图的收缩抽象[285]，其中节点是处理器，可以 沿边向其他节点发送消息。然后计算是迭代的，在每次迭代中，每个节点读取通过向内边缘接收的消息（可能还有它自己的先前状态），执行计算，然后根据结果通过向外边缘发送消息。然后，这些框架在数据图之上定义收缩计算抽象正在处理：数据图中的节点和边成为收缩图中的节点和边。我们参考附录 B.6.1 了解更多关于图并行框架的正式细节。

举个例子，假设我们希望计算最容易（或最不容易）到达的地方通过图 21 中显示的路线。衡量这一点的一个好方法是使用中心性，我们选择 PageRank [369]，它计算游客随机跟随的概率在给定数量的“跳跃”之后，图中显示的路线位于特定位置。我们可以使用图并行框架在大型图上实现 PageRank。在图22中，我们提供图21的说明性子图的 PageRank 迭代示例。节点是初始化为 1 |V | = 1 6 ，我们假设游客有平等的出发机会在任何时候。在消息阶段（Msg），每个节点v通过一个分数dRi（v）|E(v) |在其每一个传出边，我们用 d 表示用于确保收敛的恒定阻尼因子（通常 d = 0.85，表示游客随机“跳”到任何地方的概率），由 Ri(v) 得分迭代 i 中节点 v 的概率（第 i 跳后游客在节点 v 的概率），并且通过 |E(v)| v 的出边数。然后v的聚合阶段 (Agg) 将所有传入的边相加接收到的消息及其在阻尼因子中的恒定份额 ( 1−d |V | ) 来计算 Ri+1(v)。然后我们进入下一次迭代的消息阶段，一直持续到某个终止达到标准（例如，迭代次数或残差阈值等）并输出最终分数。

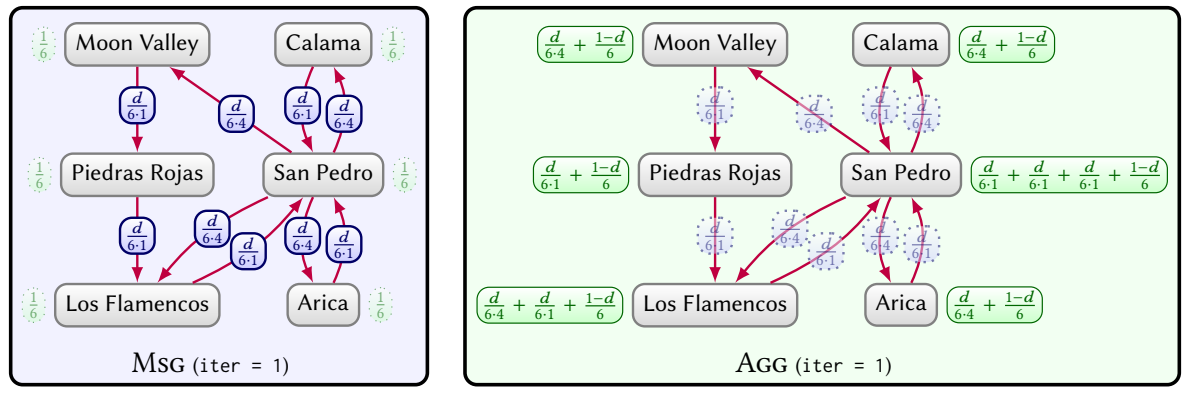


图 22 图 21 的示例子图的 PageRank 收缩迭代示例

虽然给定的示例是针对 PageRank，但收缩抽象足够通用，可以支持各种各样的图形分析，包括前面提到的那些。这里面的一个算法框架由在消息阶段 (Msg) 中计算消息值的函数组成，以及在聚合阶段（Agg）累积消息。该框架将负责分发、消息传递、容错等。然而，这样的框架——基于消息在邻居之间传递——有局限性：并非所有类型的分析都可以用这样的方式表达 frameworks [531].19 因此，框架可能允许附加功能，例如全局步骤在所有节点上执行全局计算，使结果可用于每个节点[314]；或一个允许在处理过程中添加或删除节点和边的突变步骤[314]。

5.1.3 数据图分析

如前所述，迄今为止提出的大多数分析是，在他们的 “本机”形式，适用于没有边元数据的无向图或有向图-即边标签或属性-值对——典型的图形数据模型。可以使用20多种策略应用于使数据图受到这种形式的分析：

* 投影涉及通过选择选择简单地“投影”无向图或有向图数据图的子图，所有边缘元数据都从该子图中删除；例如，图 21可能是提取由边缘标签总线和从更大的数据图中飞行，然后删除标签以创建有向图。
* 加权涉及根据一些将边缘元数据转换为数值功能。许多上述技术很容易适应加权的情况 （有向）图；例如，我们可以考虑图21中的权重，表示行程持续时间（或价格、交通等），然后计算最短路径加上持续时间21在没有外部权重的情况下，我们宁愿将边缘标签映射到权重， 根据某些标准，为所有飞行边缘、所有公共汽车边缘等分配相同的权重。

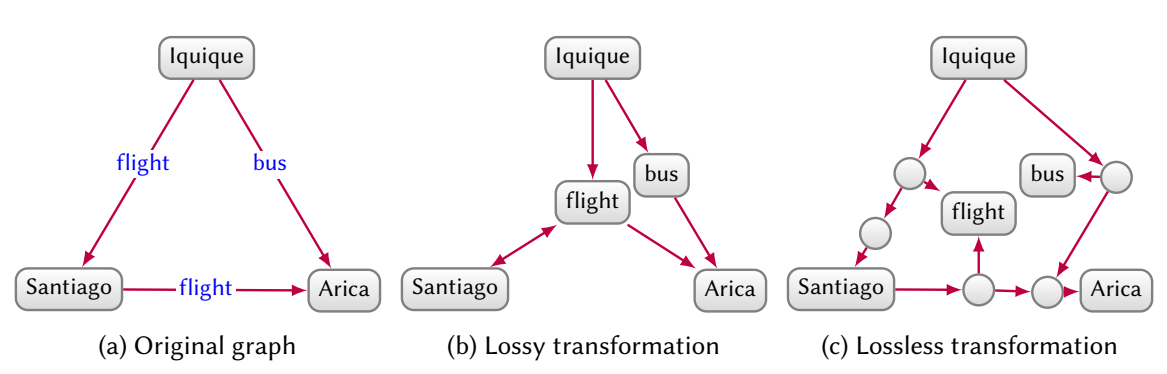


图 23 从有向边标记图到有向图的转换

* 转换涉及将图转换为较低元的模型。转型可能是有损的，意味着无法恢复原始图；或无损，这意味着可以恢复原始图。图 23 提供了从有向边标记图到有向图的有损和无损转换的示例。 在有损变换中，例如，我们无法判断原始图是否包含边缘 Iquique flight Santiago 或伊基克飞行阿里卡等。无损变换必须引入新节点（类似于 具体化）来维护有关有向标记边缘的信息。两个转换图进一步尝试保持原始图的方向性。
* 定制涉及更改分析程序以合并边缘元数据，例如基于路径表达式的路径查找。其他例子可能 包括节点相似性的结构测量，不仅考虑共同的邻居，但也有共同的邻居，由具有相同标签的边连接，或聚合中心性捕获按标签分组的边的重要性等的度量。

分析过程的结果可能会发生巨大变化，具体取决于先前的哪一个选择策略来准备数据进行分析。这个选择可能是一个不平凡的选择先验的，可能需要经验验证。需要更多的研究才能更普遍地理解这些策略对不同分析技术结果的影响。

5.1.4带有查询的分析

如第2.2节所述，用于查询图的各种语言已提出[14]。可以考虑多种方式，其中查询语言和分析可以相互补充。首先，我们可以考虑使用查询语言来投影或变换适合特定分析任务的图，例如提取图 21 的图从更大的数据图中。查询语言，例如 SPARQL [207]、Cypher [159] 和 G-CORE [13] 允许输出图表，其中此类查询可用于选择子图表进行分析。这些语言还可以表达一些有限的（非递归）分析，其中可以使用聚合例如，计算度中心性；他们也可能有一些内置的分析支持，例如，Cypher [159] 允许找到最短路径。另一方面，分析可以在优化方面为查询过程做出贡献，例如，分析连通性可能会建议如何更好地将大型数据图分布在多台机器上使用例如最小切割 [6, 253] 进行查询。分析也被用于对查询结果进行排名在大图 [146, 516]上，选择最重要的结果呈现给用户。

在某些用例中，我们可能还希望交叉查询和分析过程。为了例如，从旅游局收集的完整数据图中，考虑即将发生的航空公司罢工董事会希望在无法到达的城市中找到罢工期间发生的事件圣地亚哥因罢工而乘坐公共交通工具。假设，我们可以使用查询来提取不包括航空公司航线的运输网络（假设根据图 3，航空公司信息可用），使用分析提取包含 Santiago 的强连接组件，以及最后使用查询来查找给定日期不在圣地亚哥组件中的城市中的事件。22而 可以使用命令式语言（例如 Gremlin [420]、GraphX [529] 或 R [491]，更多的声明性语言也正在探索中，以更容易地表达这些任务，建议包括扩展具有递归功能的图查询语言[43, 414],23 将线性代数与关系（查询）代数[244]相结合，等等。

5.1.5包含蕴含的分析

知识图通常与语义模式相关联或定义领域术语语义的本体，从而产生蕴涵（根据第 4 节）。在有或没有此类蕴涵的情况下应用分析——例如，在具体化之前或之后——可能产生截然不同的结果。例如，一旦逆公理托管 inv，观察边缘 Santa Lucía 托管 EID15 在语义上等同于边缘 EID15 场地 Santa Lucía。场地是被调用；然而，从分析技术的角度来看，这些边缘远非等效考虑边缘方向，其中包括一种类型的边缘，或另一种，或两者，可能有对最终结果产生重大影响。据我们所知，分析的结合和蕴涵没有得到很好的探索，留下了一些有趣的研究问题。沿着这些行，探索产生相同的语义不变分析可能很有趣在语义等价的图（即，相互关联的图）上得到结果，从而分析知识图谱的语义内容，而不仅仅是知识图谱的拓扑特征数据图；例如，语义不变的分析将在图上产生相同的结果包含逆公理主机 inv。场地和许多主机边，相同的图，但其中每个主机边缘都被一个反向场地边缘替换，并且两个图的并集。

5.2知识图谱嵌入

近年来，机器学习方法受到了广泛关注。在知识图谱的上下文中，机器学习可以用于细化知识图谱本身，这涉及预测新边和/或识别错误边（在第8节中进一步讨论）；或用于下游任务，其中知识图谱用于训练应用领域中的分类、推荐、回归等模型（第10节将进一步讨论）。然而，许多传统的机器学习技术假设密集的数字输入以向量的形式表示，这与图形通常的表达方式截然不同。那么，图谱——其节点、边等，如何被编码为数字向量呢？

使用向量表示图谱的第一种尝试是使用一个热编码，为每个长度为|L|·| V|的节点生成一个向量，其中|V|表示输入图谱中的节点数和|L|表示边标签数——在相应的索引处放置一以指示图形中各自边的存在，否则为零。然而，这种表示通常会导致大而稀疏的向量，这对大多数机器学习模型都是有害的。知识图嵌入技术的主要目标是在连续的低维向量空间中创建图谱的密集表示（即嵌入图谱），然后用于机器学习任务。嵌入的维数d是固定的，通常较低（通常为50≥ D≥ 1000). 通常，图谱嵌入由每个节点的实体嵌入组成：一个具有d维的向量，我们用e表示；每个边标签的关系嵌入：（通常）我们用r表示的d维向量。这些向量的总体目标是抽象和保留图中的潜在结构。有很多种方法可以实例化嵌入的概念。最常见的情况是，给定一个边s—>p—>o，一个特定的嵌入方法定义了一个评分函数，该函数接受es（节点s的实体嵌入）、rp（边缘标签p的实体嵌入）和eo（节点o的实体嵌入），并计算边缘的合理性：其可能性有多大。给定一个数据图，目标是根据给定的评分函数，计算尺寸d的嵌入，以最大化正边（通常是图中的边）的合理性，并最小化负示例（通常是图中的边，节点或边标签已更改，因此它们不再在图中）的合理性。由此产生的嵌入可以被视为通过自我监督学习的模型，对图形的（潜在）特征进行编码，将输入边映射到输出合理性分数。

然后，嵌入可以用于许多低级任务，这些任务涉及计算它们的图的节点和边标签。首先，我们可以使用似然性评分函数将置信度分配给可能（例如）从外部源中提取的边（后面将在第6节中讨论）。第二，为了链路预测的目的，合理性评分函数可用于完成缺少节点/边缘标签的边缘（后面将在第8节中讨论）；例如，在图21中，我们可能会问，图中的哪些节点可能会完成边Grey Glacie—>巴士—>？，除了已经给出的Punta Arenas，我们可能直觉地认为Torres del Paine是一个合理的候选人。第三，嵌入模型通常会将相似的向量分配给相似的节点和相似的边缘标签，因此它们可以用作相似性度量的基础，这可能有助于找到引用同一实体的重复节点，或提供建议（稍后在第10节中讨论）。

已经提出了一系列知识图嵌入技术[519]，其中我们的目标是对迄今为止提出的一些最流行的技术进行高级介绍。首先，我们讨论采用几何视角的转换模型，其中关系嵌入将低维空间中的主体实体转换为客体实体。然后，我们描述了张量分解模型，该模型提取了近似图结构的潜在因子。此后，我们讨论了使用神经网络训练嵌入的神经模型，这些嵌入可以提供准确的合理性分数。最后，我们讨论了利用现有单词嵌入技术的语言模型，提出了为预期（文本）输入生成类似图形的类比的方法。附录B.6.2中提供了对这些模型的更正式的处理。

5.2.1转换模型.

平移模型将边缘标签解释为从主体节点（又名源节点或头部）到对象节点（又名目标节点或尾部）的变换；例如，在边 San Pedro—>bus—>Moon Valley中，边bus被视为将San Pedro转变为Moon Valley，同样地，对于其他bus也是如此。这个家族中最基本的方法是TransE[59]。在所有正边缘s—>p—>o上，TransE学习向量es、rp和eo，以使es+rp尽可能接近eo。相反，如果边缘是一个负面的例子，TransE试图学习一种使es+rp远离eo的表示法。为了说明这一点，图24提供了一个二维（d=2）实体和关系嵌入的玩具示例，由TransE计算。为了清晰起见，我们保持向量的方向与原始图形相似。对于原始图形中的任何边s—>p—>o，添加向量es+rp应近似于eo。在这个玩具示例中，向量精确地对应于添加Licantén（eL.）向量的位置（rwo）以西给出一个对应于库里科（eC.）的向量。我们可以使用这些嵌入来预测边缘（以及其他任务）；例如，为了预测图中哪个节点最有可能位于Antofagasta（A）以西，通过计算eA+rwo。我们发现结果向量（图24c中的虚线）最接近eT，因此预测Toconao（T.）是最合理的此类节点。

除了这个玩具的例子，TransE可能过于简单化；例如，在图21中，bus不仅将（San Pedro）转换为（Moon Valley），还将其转换为（Arica）、（Calama）等地。在这种情况下，TransE的目标是为所有此类目标位置提供类似的向量，考虑到其他边缘，这可能是不可行的。TransE也倾向于将循环关系指定为零向量，作为方向分量会相互抵消。为了解决这些问题，研究了TransE的许多变体。例如，其中TransH[521]使用不同的超平面表示不同的关系，其中对于边s—>p—>o，s首先投影到p的超平面上，然后再学习转换为o（不受带有s和o的其他标签的边的影响）。TransR[256]通过将s和o投影到特定于p的向量空间中，将s和o的实体嵌入乘以特定于p的投影矩阵，从而推广了这种方法。TransD[256]通过将实体和关系与第二个向量相关联，简化了TransR，其中这些二级向量用于将实体投影到特定于关系的向量空间中。关于其他翻译模型的讨论，我们参考了Wang等人的调查。[519]

图表, 箱线图

描述已自动生成

图24。TransE学习的二维关系和实体嵌入的玩具示例；实体嵌入使用缩写，并包括向量加法的示例，以预测Antofagasta以西的情况

5.2.2张量分解模型

推导图谱嵌入的第二种方法是应用基于张量分解的方法。张量是将标量（0阶张量）、向量（1阶张量）和矩阵（2阶张量）推广到任意维/阶的多维数值域。张量已经成为机器学习中广泛使用的抽象概念[402]。张量分解涉及将张量分解为更多“基本”张量（例如，低阶张量），从中可以通过固定的基本操作序列重新组合（或近似）原始张量。这些元素张量可以被视为捕捉原始张量中包含的信息背后的潜在因素。张量分解有很多方法，现在我们将简要介绍秩分解背后的主要思想[402]。暂时撇开图形，考虑（a-b）矩阵（即，二阶张量）C，其中A是智利城市的数量，B是一年中的月份数，并且每个元素（Cij）表示第i个城市j第个月的的平均温度。注意到智利是一个又长又薄的国家——从南部的亚极地气候到北部的沙漠气候——我们可能会发现C分解为两个代表潜在因素的向量——特别是x（含m元素），对于纬度较低的城市给出较低的值，而y（含n元素），在温度较低的月份给出较低的值，这样计算两个向量的外积24可以相当好地近似于C:x⊗ Y≈ C.在（不太可能）存在向量x和y的情况下，C正好是两个向量（x）的外积⊗ y=C）我们称C为秩1矩阵；然后我们可以使用m+n值而不是m×n值精确地编码C。然而，大多数情况下，为了得到精确的C，我们需要对多个秩-1矩阵求和，其中C的秩r是需要求和以得到精确的C的秩-1矩阵的最小数目，

x1⊗ y1+…+xr⊗ yr=C。在温度示例中，x2⊗ y2可能对应于高度修正值x3⊗ y3表示更高的温度变化进一步向南等。矩阵的（低）秩分解然后在秩上设置限制d，并计算向量（x1，y1，…，xd，yd），以便x1⊗ y1+…+除息的⊗ yd给出了C的最佳d秩近似。注意到要生成n阶张量，我们需要计算n个向量的外积，我们可以将这个想法推广到张量的低秩分解；这种方法被称为标准多聚体（CP）分解[225]。例如，我们可能有一个三阶张量C，包含智利城市一天中四个不同时间的月温度，可以近似为x1⊗ y1⊗ z1+。除息的⊗ yd⊗ zd（例如，x1可能是纬度因子，y1可能是月变化因子，z1可能是日变化因子，等等）。然后存在各种算法来计算（近似）CP分解，包括交替最小二乘法、Jennrich算法和张量幂法[402]。

电脑屏幕的照片上有字

中度可信度描述已自动生成

图25图24a张量的CP-d秩分解的抽象图示

回到图谱，可以使用类似的原理将图谱分解为向量，从而产生嵌入。特别地，一个图谱可以被编码为带有|V |×|L |×|V |元素的一个3阶张量G，其中，如果第i个节点链接到第k个节点，且其边具有第j个标签，则元素（G）ijk被设置为一，否则为零。如前所述，这样的张量通常非常大且稀疏，因此秩分解是适用的。CP分解[225]将计算向量序列（x1，y1，z1，…，xd，yd，zd），使得x1⊗ y1⊗ z1+…+ xd ⊗ yd⊗ zd≈ G.我们在图25中说明了这个方案。让X，Y，Z表示由？x1···xd？y1·yd？z1··zd？，每个向量分别构成对应矩阵的一列，然后我们可以提取Y的第i行作为第i个关系的嵌入，提取X和Z的第j行作为第j个实体的两个嵌入。然而，知识图嵌入通常旨在为每个实体分配一个向量。DistMult[533]是一种基于秩分解计算知识图嵌入的开创性方法，其中每个实体和关系都与维度d的向量相关联，因此对于边s—>p—>o，定义了似然性评分函数˝di=1（es）i（rp）i（eo）i，其中（es）i（rp）i（eo）i分别表示向量es、rp、eo的第i个元素。然后，目标是学习每个节点和边标签的向量，以最大化正边的合理性，并最小化负边的合理性。这种方法相当于图张量G的CP分解，但其中实体有一个使用两次的向量：x1⊗ y1⊗ x1+…+ xd ⊗ yd⊗ zd≈ G.该方法的一个缺点是，根据评分函数，s—>p—>o的合理性始终等于o—>p—>s的合理性；换句话说，DistMult不考虑边缘方向。

RESCAL[364]没有使用向量作为关系嵌入，而是使用了一个矩阵，它允许在所有维度上组合es和eo的值，因此可以捕捉（例如）边缘方向。然而，RESCAL在空间和时间方面比DistMult成本更高。HolE[363]使用向量进行关系和实体嵌入，但建议使用循环相关算子它沿着两个向量的外积的对角线求和，将它们组合起来。这个算子不是交换的，因此可以考虑边缘方向。另一方面，ComplEx[499]使用复数向量（即包含复数的向量）作为关系嵌入，这同样允许打破上述DistMult评分函数的对称性，同时保持较低的参数数量。SimplE[266]提出计算标准CP分解，计算X和Z中实体的两个初始向量，然后对X、Y和Z中的项进行平均，以计算最终的合理性得分。TuckER[27]采用了一种不同类型的分解——称为TuckER分解[501]，它计算较小的“核心”张量T和三个矩阵A、B和C的序列，从而G≈ T⊗ A.⊗ B⊗ C实体嵌入来自A和C，而关系嵌入来自B。在这些方法中，TuckER[27]目前提供了标准基准的最新结果

5.2.3神经模型

前面讨论的方法的一个局限性是，它们假设嵌入上的线性（保留加法和标量乘法）或双线性（例如矩阵乘法）运算来计算合理性分数。许多方法使用神经网络来学习具有非线性评分函数的嵌入，以获得合理性。最早提出的神经模型之一是语义匹配能量（SME）[185]，它学习两个函数的参数（也称为权重：w，w′）——fw（es，rp）和аw′（eo，rp）——这样两个函数的结果的点积——fw（es，rp）·аw′（eo，rp）——就给出了合理性得分。提出了fw和аw′的线性和双线性变体。另一个早期的建议是神经张量网络（NTN）[463]，它更倾向于保持内部权重的张量W，从而通过结合外部乘积es的复杂函数计算似然性得分⊗ W⊗ eo在es和eo上有一个标准的神经层，然后再与rp结合，生成一个似是而非的分数。张量W的使用导致大量参数，这限制了可伸缩性[519]。多层感知器（MLP）[126]是一个更简单的模型，其中es、rp和eo被连接并馈送到一个隐藏层，以计算合理性得分。

最近有许多方法提出在模型中使用卷积核。Conv[122]提出通过将每个向量“包装”到几行上，并将两个矩阵连接起来，从es和rp生成一个矩阵。串联矩阵用作一组（2D）卷积层的输入，该卷积层返回一个特征映射张量。使用参数化线性变换将特征映射张量矢量化并投影到d维。然后根据该向量与eo的点积计算似然性得分。ConvE的一个缺点是，通过将向量包装到矩阵中，它在嵌入上施加了一种人工的二维结构。HypER[26]是一个使用卷积的类似模型，但避免了将向量包装成矩阵的需要。相反，一个完全连接的层（称为“超网络”）应用于rp，并用于生成特定于关系的卷积滤波器矩阵。这些过滤器直接应用于es，以提供一个矢量化的特征图。然后应用与ConvE中相同的过程：将生成的向量投影到d维，并将点积与eo一起应用，以生成合理性得分。结果表明，该模型在标准基准上的表现优于Conv[26]。

所提出的方法在表达能力和需要训练的参数数量方面取得了不同的平衡。虽然更具表现力的模型，如NTN，可以通过使用更多隐藏参数，更好地在低维嵌入中拟合更复杂的似然函数，更简单的模型，如Dong等人[126]提出的模型，以及卷积网络[26，122]，通过在矩阵的不同区域应用相同（通常较小）的核来实现参数共享，总体而言，需要处理的参数更少，可扩展性更强。

5.2.4语言模型

嵌入技术最初是作为一种在机器学习框架内表示自然语言的方法而被探索的，word2vec[330]和GloVe[383]是两种开创性的方法。这两种方法都是基于文本40的大型语料库来计算单词的嵌入，以便在类似上下文中使用的单词（例如，“青蛙”、“蟾蜍”）具有相似的向量。Word2vec使用经过训练的神经网络，可以从周围的单词（连续的单词包）预测当前单词，也可以预测给定当前单词的周围单词（连续的跳过语法）。GloVe在词对的共现概率矩阵上应用回归模型。这两种方法生成的嵌入被广泛应用于自然语言处理任务中。

因此，图形嵌入的另一种方法是利用经过验证的语言嵌入方法。然而，当一个图由三个术语的无序序列（即一组边）组成时，自然语言中的文本由任意长度的术语序列（即单词的句子）组成。沿着这些线，RDF2Vec[416]在图上执行（偏置[91]）随机游动，并将路径（遍历的节点和边标签序列）记录为“句子”，然后作为word2vec[330]模型。从图21中提取的这类路径的一个例子可能是，例如，San Pedro-bus-> Calama- flight ->Iquique-flight->Santiago，在这里，论文用500条路径进行实验，每个实体的长度为8。RDF2Vec还提出了第二种模式，即从规范标记的子树中为节点生成序列，这些子树是其根节点，本文在这种模式下对深度为1和2的子树进行了实验。相反，KGloVe[92]基于手套模型。就像最初的手套模型[383]认为文本窗口中频繁出现的单词更相关一样，KGloVe使用个性化页面RANK25来确定与给定节点最相关的节点，然后将其结果输入GloVe模型。

5.2.5蕴含感知模型

到目前为止，嵌入仅考虑数据图。但如果提供了一个本体或一组规则呢？这种演绎知识可以用来改进嵌入。一种方法是使用约束规则来细化嵌入所做的预测；例如，Wang等人[520]使用函数和逆函数定义作为约束（在UNA下），例如，如果我们定义一个事件最多可以有一个场地值，这将用于降低将多个场地分配给一个事件的边的合理性。

最近的方法，而提出联合嵌入考虑数据图和规则时，计算嵌入。羽衣甘蓝（200）使用平移模型（特别是RANSE）计算实体和关系嵌入，适用于使用T-Fuzzy模糊逻辑进一步考虑规则。参考图21，考虑一个简单的规则?x-bus->y ⇒ ?x -connects to->?y。我们可以使用嵌入为新的边分配合理性分数，比如e1:Piedras Rojas-bus->Moon Valleybus。我们可以进一步应用之前的规则，从预测的边缘e1生成新的边缘e2:Piedras Rojas -connects to->Moon Valley连接。但我们应该赋予这第二条优势什么样的合理性呢？假设p1和p2是e1和e2的当前合理性分数（使用标准嵌入进行初始化），那么t范数模糊逻辑建议将合理性更新为p1p2− p1+1。然后，对嵌入进行训练，将更大的似然性分数分配给正面例子，而不是边缘和基本规则的负面例子。基于图21的积极基本原则的一个例子是Arica-bus->San Pedro⇒Arica -connects to->San Pedr。消极的基本规则随机替换规则头部的关系；例如，Arica-bus->San Pedro⇏Arica-flight->San Pedro。郭等人[201]后来提出了RUGE，它使用一个基于基本规则的联合模型（可能是带有置信度分数的软规则）和合理性分数来对齐这两种形式的未看到边缘的评分。制定基本规则可能代价高昂。另一种称为FSL[120]的方法观察到，在一个简单规则的情况下，例如?x-bus->y ⇒ ?x -connects to->?y，关系嵌入总线应始终返回比连接到更低的合理性。因此，对于所有这些规则，FSL建议训练关系嵌入，同时避免违反这些不平等。虽然相对简单，但FSL只支持简单的规则，而KALE也支持更复杂的规则。

这些作品是演绎和归纳知识形式——在本例中是规则和嵌入——如何相互作用和互补的有趣例子。

5.3 图神经网络

嵌入的目的是提供适于在现有机器学习模型中使用的图形的密集数字表示， 另一种方法是建立适合于图形结构数据的自定义机器学习模型。 大多数图形的自定义学习模型都是基于(人工)神经网络[527]，利用了两者之间的自然对应关系:神经网络已经对应于一个加权有向图，其中节点充当人工神经元，边充当加权连接(轴突)。然而，传统神经网络的典型拓扑结构，更具体地说，全连接前馈神经网络——是相当同构的，它是按照节点的顺序层来定义的，其中一层中的每个节点都与下一层中的所有节点相连接。相反，数据图的拓扑结构是非常异构的，由它的边所代表的实体之间的关系决定。

图神经网络(GNN)[437]基于数据图的拓扑结构构建神经网络;也就是说，每个数据图中的节点都连接到它们的邻居。通常情况下，然后学习模型以监督的方式映射节点的输入特征到输出特征；例如节点的输出特征可以手动标记，也可以从知识图中获取。

现在，我们将讨论GNN的两种关键形式——递归GNN和卷积GNN的主要思想，其中我们参考附录B.6.3以获得与GNN相关的更正式的定义。

5.3.1 递归图神经网络。

递归图神经网络(RecGNNs)是图神经网络的开创性方法[437,468]。该方法在概念上类似于图22所示的收缩抽象，在图22中，消息在邻居之间传递，以递归地计算某些结果。然而，我们不是定义用来决定要传递的消息的函数，而是标记节点训练集的输出，让框架学习生成预期输出的函数，然后应用它们来标记其他示例。

在一篇开创性的论文中，Scarselli等人[437]提出了他们通常称为图神经网络(GNN)的方法，该方法将一个有向图作为输入，其中节点和边与特征向量相关联，这些特征向量可以捕获节点和边的标签、权值等。这些特征向量在整个过程中保持不变。图中的每个节点还与一个状态向量相关联，该状态向量根据节点邻居的信息（即，相邻节点的特征和状态向量，以及从它们延伸到或者起始的边的特征向量）使用一个参数函数进行递归更新，称为传播函数。第二个参数函数称为输出函数，用于根据节点自身的特征和状态向量计算节点的最终输出。这些函数递归地应用到一个定点。两个参数函数都可以使用神经网络来实现，在神经网络中，给定图中的部分监督节点集，即，节点标记为过渡和输出函数的期望输出参数，可以学习到最接近监督输出。因此，结果可以看作是递归神经网络体系结构。为了确保收敛到一个固定点，应用了某些限制，即转换函数是一个承包商，这意味着每次应用该函数时，数值空间中的点会更紧密地结合在一起（直观地说，在这种情况下，每个应用程序的数字空间都会“缩小”，从而确保一个唯一的固定点）。

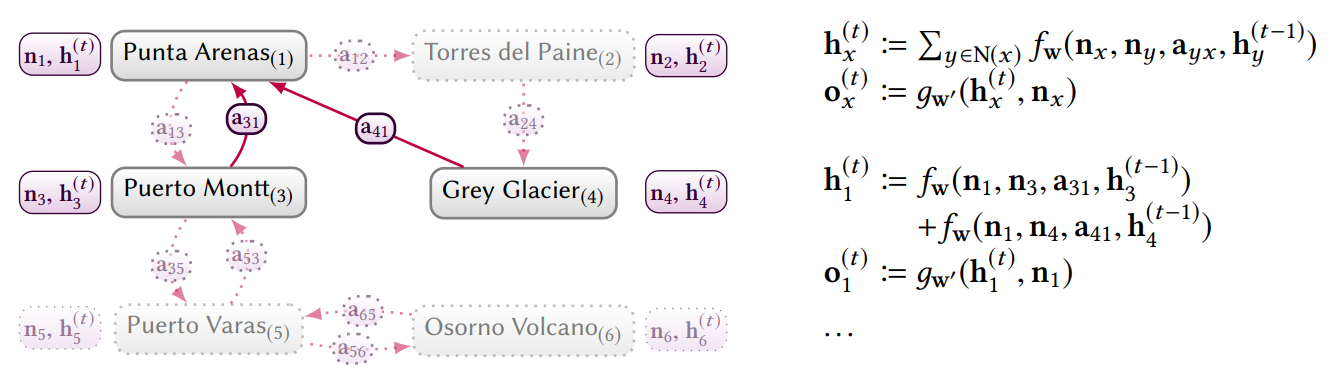


图26 左图为图21中突出显示Punta Arenas邻域的子图，其中节点用特征向量(nx)和步长t(h(t)x)的隐藏状态进行标注，边用特征向量(axy)进行标注;右侧为Scarselli等人提出的GNN转移和输出函数[437]，以Punta Arenas (x = 1)为例，其中N (x)表示x的邻近节点，fw(·)表示参数w的转移函数，gw '(·)表示参数w '的输出函数。

举例来说，考虑一下，我们希望找到建立新的旅游信息办公室的优先地点。一个好的策略是将它们安装在许多游客前往热门目的地的枢纽。沿着这些思路，在图26中，我们说明了Scarselli等人提出的GNN架构。[437]在图21的子图中，我们突出显示了蓬塔阿里纳斯附近地区。在这个图中，节点用特征向量(nx)和步长t(h(t)x)的隐藏状态进行标注，边用特征向量(axy)进行标注。例如，节点的特征向量可以对节点类型(City、Attraction等)进行one-hot编码，直接对每年的游客数量等统计数据进行编码。例如，节点的特征向量可以对边缘标签(运输类型)进行一次热编码，直接对统计数据进行编码，如每年售出的车票的距离或数量等。隐藏状态可以被随机初始化。图26的右侧为GNN传播和输出函数，其中fw(·)为参数w的传播函数，gw '(·)为参数w '的输出函数。Punta Arenas (x = 1)也提供了一个例子，其中N(x)表示x的邻近节点。这些函数将被递归应用，直到达到一个不动点。为了训练这个网络，我们可以列出已经(或应该)设立旅游局的地方和没有设立(或不应该)旅游局的地方。这些标签可以从知识图中获取，也可以手工添加。然后，GNN可以学习参数w和w '，这些参数给出了标记示例的预期输出，随后可以用于标记其他节点。

这种GNN模型是灵活的，可以以各种方式进行调整[437]:我们可以以不同的方式定义相邻节点，例如包括包括用于输出边的节点，或一跳或两跳之外的节点；我们可以允许节点对通过具有不同向量的多条边连接；我们可以考虑每个节点具有不同参数的转换和输出函数;我们可以为边添加状态和输出;我们可以把和换成另一个聚合函数;等。

5.3.2卷积图神经网络。

卷积神经网络(Convolutional neural networks, CNNs)已经得到了广泛的关注，特别是在涉及图像的机器学习任务方面[283]。图像设置的核心思想是使用卷积算子在图像的局部区域上应用小内核(也就是过滤器)来提取该局部区域的特征。当应用到所有局部区域时，卷积输出图像的特征映射。通常会应用多个核，形成多个卷积层。只要给出足够的带标签的例子，这些内核是可以学习的。

人们可能会注意到前面讨论过的GNN和应用于图像的CNN之间的相似性：

在这两种情况下，运算符都应用于输入数据的局部区域。在GNNs的情况下，传播函数应用于应用于图中的节点及其邻居。在CNN的情况下，卷积应用于图像中的像素及其邻域。根据这种直觉，人们提出了许多卷积图神经网络(卷积图神经网络，convgnn)[68, 271, 527]，其中的传播函数是通过卷积实现的。对于convgnn，一个关键的考虑因素是如何定义图的区域。与图像的像素不同，图中的节点可能有不同数量的邻居。这带来了一个挑战:cnn的一个优点是可以在图像的所有区域上应用相同的内核，但对于convgnn来说，这需要更仔细的考虑，因为不同节点的邻域可能是不同的。解决这些挑战的方法包括使用图的光谱(例如[68,271])或空间(例如[336])表示，从图中归纳出更规则的结构。另一种方法是使用注意机制[507]来学习对当前节点最重要的节点。

除了架构方面的考虑，recgnn和convgnn之间还有两个主要区别。首先，recgnn递归地将邻域的信息聚合到一个固定点，而convgnn通常应用固定数量的卷积层。其次，recgnn通常在统一的步骤中使用相同的函数/参数，而ConvGNN的不同卷积层可以在每个不同的步骤中应用不同的核/权值。

5.4符号学习

到目前为止讨论的监督技术——即知识图嵌入和图神经网络——通过图来学习数值模型。然而，这种模型通常很难解释或理解。例如,图27的图,图嵌入知识可能预测边缘是非常合理的,但是他们不会提供一个可判断的模式来帮助理解为什么会出现这种情况:结果可能的原因在于一个矩阵的参数学习适应训练数据的合理性评分。这些方法还存在词汇表外的问题，它们无法为包含以前不可见节点或边的边提供结果;例如，如果我们添加一条边，其中在图中是新的，那么知识图嵌入将不具有的实体嵌入，并且需要重新训练以估计一条边的可信性。

另一种(有时是互补的)方法是采用符号学习，以一种符号(逻辑)语言来学习假设，“解释”一组给定的正负边。这些边通常是由知识图自动生成的(类似于知识图嵌入的情况)。然后，假设作为可解释的模型，可用于进一步的演绎推理。例如，在图27的图中，我们可以通过观察到飞行路线往往是返回的路线，来学习的规则。或者，我们可以学习一个描述逻辑（DL）公理，说明机场要么是国内的，要么是国际的，或者两者都是:。这些规则和公理可以用于演绎推理，并为所包含/预测的新知识提供一个可解释的模型;例如，从上述往返航班的规则，可以解释为什么会预测到一个新的边缘。这进一步为领域专家提供了验证模型的机会——例如，规则和公理——由这些过程推导而来。最后,对规则/公理进行量化(所有航班往返飞行,所有国内或国际机场,等等),所以他们可以应用于有未登录词的情况(例如，使用上述规则,我们可以从一个新的边缘与看不见的节点推出)。

在本节中，我们将讨论符号学习的两种形式:规则挖掘，从知识图中学习规则;公理挖掘，学习其他形式的逻辑公理。关于这两个任务的更正式说明，请参阅附录B.6.4。

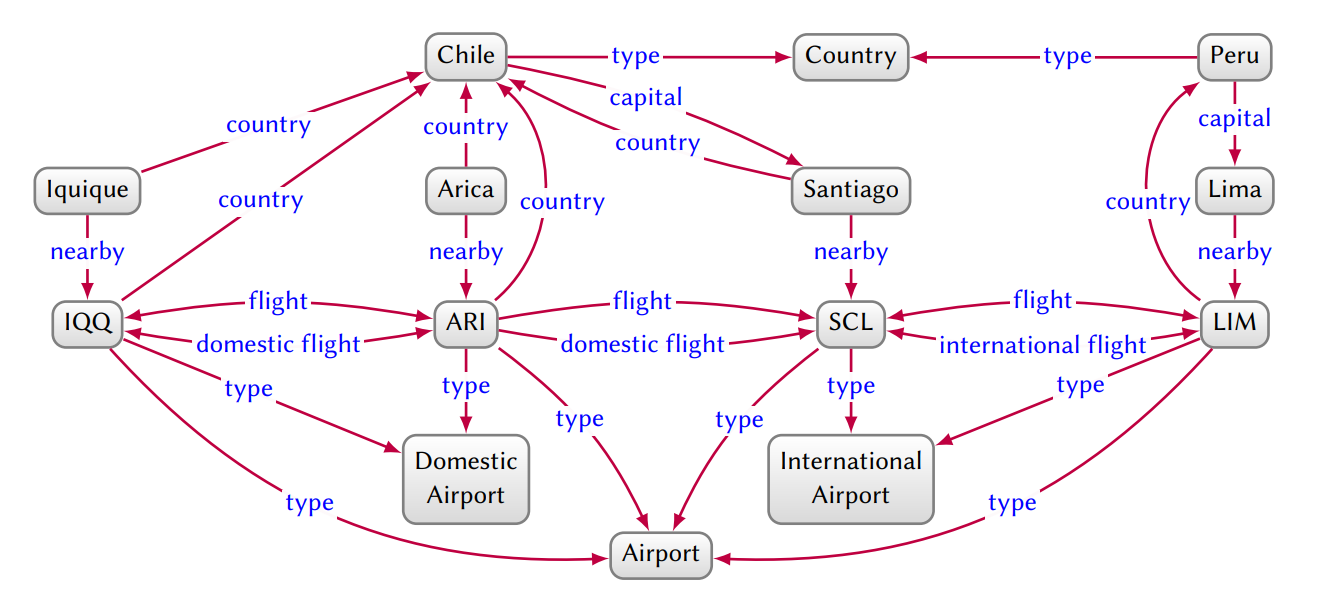


图27所示 描述机场之间航班的一种不完整的有向边标记图

5.4.1 规则挖掘

一般意义上的规则挖掘是指从大量的背景知识中发现以规则形式存在的有意义的模式。在知识图的上下文中，我们假定一组已知的正边和负边。典型的正边是观察到的边(即知识图中给定或包含的边)，而负边是根据给定的完整性假设定义的(稍后讨论)。规则挖掘的目标是识别出从其他正边中产生正边的比率较高，而从正边中产生负边的比率较低的新规则。所考虑的规则类型可能不同于更简单的情况，如前面所提到的,更复杂的规则,如，表示首都附近的机场倾向于国际机场;或，表示同一国家内的航班为国内航班(见章节4.3.1)。

根据国际机场的例子，规则并不是假定在所有情况下都适用，而是与它们在多大程度上符合正边和负边有关。更详细地说，我们称规则所包含的边和正边的集合(不包括所包含的边本身)为该规则的正包含。具有positive含义的数量称为规则的支持度，而具有positive含义的规则的比值称为规则的置信度[483]。因此，支持度和置信度分别表示“确认”规则的隐含数和比例为真，其中的目标是识别具有高支持度和高置信度的规则。事实上，关系设置中的规则挖掘技术在归纳逻辑编程(ILP)的背景下已经被探索了很长时间[405]。然而，由于数据的规模和频繁假设的不完整数据(OWA)，知识图呈现了新的挑战，而专门的技术已经被提出来解决这些问题[164]。

在处理不完全知识图时，如何定义负边并不是很清楚。一种常见的启发式方法——也用于知识图的嵌入——是采用部分完整性假设, (PCA)[164]，该方法认为正边的集合是包含在数据图中的边，负例的集合是所有不在图中的边的集合但存在一个节点使得在图中。以图27为例，PCA下的负边为(考虑到的存在);相反，()既不正也不负。PCA的置信度是支持度与正集合或负集合中所有蕴涵的比值[164]。例如，规则的支持度为2(因为图中它包含了和)，而置信度是2/2 = 1(注意()虽然包含，但既不正也不负，因此在度量时被忽略)。对规则的支持度近似为4，而置信度为4/5 = 0.8(注意为负)。

我们的目标是找到满足给定支持和置信阈值的规则。AMIE是图形规则挖掘系统中一个很有影响力的系统[163,164]，它采用PCA的置信度度量，并以自上而下的方式[483]构建规则，规则头为。对于这种形式的每个规则头(每个边标签一个)，考虑三种类型的改进（refinements），每一种都向规则主体添加一条新边。这条新边从图中获取一个边标签，并可能使用之前没有出现在规则中的新变量、已经出现在规则中的现有变量或图中的节点。这三种改进可以:

1. 添加一个现有变量和一个新变量的边;例如，改进前面提到的规则头部可能会给出:;
2. 从图中添加一个已有变量和一个节点的边;例如，改进上面的规则可以给出:;
3. 添加一条带有两个现有变量的边;例如，改进上面的规则可能会给出:。

这些细化可以任意组合，从而产生一个潜在的指数搜索空间，在这个空间中，满足给定的支持和置信度阈值的规则将得到维护。为了提高效率，可以对搜索空间进行修剪;例如，这三种细化总是会降低支持度，因此如果规则不满足支持阈值，就不需要探索它的细化。对生成的规则类型施加了进一步的限制。首先，只考虑一定大小的规则。其次，规则必须是封闭的，这意味着每个变量至少出现在规则的两个边缘，这确保了规则是安全的，这意味着头部的每个变量都出现在身体中;例如，之前由第一次和第二次改进产生的规则既不是封闭的(变量出现一次)，也不是安全的(变量只出现在头部)。为了确保规则是封闭的，将应用第三次细化，直到规则关闭为止。进一步讨论基于剪枝和索引的可能优化，可参考AMIE+[163]的论文。

后来的工作建立在这些技术的基础上，从知识图中挖掘规则。Gad-Elrab等人[162]提出了一种学习非单调规则的方法——规则体中带有负边的规则——以捕获基本规则的异常;例如，该方法可以学习一条规则,表明航班在同一个国家除了(离职)机场是国际时,异常虚线所示,我们使用¬否定优势。规则系统[230]——它也能够学习非单调规则——提出通过扩展置信度来考虑知识图中没有出现的隐含边的知识图嵌入的可信性得分，从而减轻PCA启发式的局限性。在可能的情况下，明确说明知识图的完整性(例如以形状表示;(参见3.1.2节)可以代替PCA来识别负边。沿着这些思路，利用关于关系基数的额外知识来完善负面示例集和候选规则的置信度度量。或者，在可用的情况下，可以使用本体论，通过例如不相交公理，推导出OWA下的逻辑确定负边。d 'Amato等人提出的系统[109,110]利用本体论上包含的负边来确定通过进化算法生成的规则的置信度。

虽然之前的工作涉及到候选规则的离散扩展，并应用了固定的置信度评分函数，但另一项研究是关于一种称为可微规则挖掘的技术[419,430,534]，它允许端到端规则学习。其核心思想是，规则体中的连接可以表示为矩阵乘法。更具体地说，我们可以用邻接矩阵Ap(大小为|V|×|V|)表示边标p的关系，这样，如果从第i个实体到第j个实体有一条标记为p的边，则第i行第j列的值为1;否则值为0。现在我们可以将规则体中的连接表示为矩阵乘法;例如，给定，我们可以用矩阵乘法Adf.Ac.表示出该对象，它给出了一个邻接矩阵来表示限定的国家边缘，在这里我们应该期望Adf.Ac.中的1。被头部的邻接矩阵Ac覆盖。因为我们得到了所有边缘标签的邻接矩阵，所以我们需要学习单个规则的置信度分数，并学习具有阈值置信度的规则(长度不等)。沿着这些思路，NeuralLP[534]使用了一种注意机制，为形式为的类路径规则选择一个变长边缘标签序列，同样学习了这些规则的置信区间。DRUM[430]也学习类路径规则，其中，观察到一些边缘标签更可能/更不可能遵循规则中的其他标签——例如，在图21的图中，资本不会跟随飞行，因为连接将是空的——系统使用双向循环神经网络(一种流行的学习顺序数据的技术)来学习规则的关系序列及其置信度。然而，这些可微分的规则挖掘技术目前仅限于学习路径类规则。

5.4.2 公理挖掘。

除了规则之外，还可以从知识图中挖掘出更一般的公理形式——用逻辑语言(如DLs)表示(见4.3.2节)。我们可以将这些方法分为两类：那些挖掘特定公理的方法和更一般的公理的方法。

在挖掘特定类型公理的系统中，脱节公理是一个常见的目标;例如，脱节公理认为这两个类的交集等同于空类，或者更简单地说，没有节点可以同时是和类型。[512]提出的系统基于(负)关联规则挖掘[4]提取脱节公理，即在知识图中找到类对，其中每个类在知识图中有很多实例，但两个类的实例相对较少(或没有)。Töpper等人[497]更愿意为余弦相似度低于固定阈值的类对提取脱节。为了计算余弦相似度，类向量是使用TF-IDF类比来计算的，其中每个类的“文档”都是由它的所有实例构造而来的，这个文档的“术语”是在类的实例上使用的属性(保持多样性)。前两种方法发现了命名类之间的脱节约束(例如，城市与机场之间的脱节)，而Rizzo等人[418]提出了一种方法，可以捕获类描述之间的脱节约束(例如，附近没有机场的城市与一个国家的首都城市脱节)。该方法首先对知识库中的相似节点进行聚类。接下来，提取术语集群树，其中每个叶节点表示之前提取的集群，每个内部(非叶节点)节点是一个类定义(例如，城市)，其中左子节点是一个拥有该类中所有节点或子类描述的集群(例如，没有机场的城市)，右边的子类要么是一个类中没有节点的集群，要么是一个不连接的类描述(例如，有事件的非城市)。最后，对于树中不要求有子类关系的类描述对，提出了候选脱节公理。

其他系统则提出了学习更一般公理的方法。其中一个突出的系统是DLLearner[70]，它基于类学习(又称概念学习)的算法，给定一组正节点和负节点，目标是找到一种逻辑类描述来划分正集和负集。例如，给定{, }为正集，{}为负集，我们可以学习一个(DL)类描述，表示机场附近的实体不是大写的，其中所有正节点为实例，没有负节点为实例。学习此类类描述的方式与前面提到的系统(如AMIE)学习规则的方式类似，使用细化操作符从更一般的类转移到更具体的类(反之亦然)，使用置信评分函数和搜索策略。系统进一步支持通过一个记分函数学习更多的一般公理，该计分函数使用计数查询来确定在图中确实找到的预期边（公理为真时将包含的边）的比例；例如，为公理打分。在图27 ，我们可以使用一个图形查询数有多少节点从国内机场航班(3)传入,和有多少个节点是国际机场且从国内机场传入的航班(1)，这两方面的区别越大，公理的证据越弱。