

# 知识图谱划分算法研究综述

靳清涵 李菁菁 赵宇轩

(沈阳航空航天大学 辽宁 沈阳 110136)

**[摘要]** 伴随着国民经济的蓬勃发展,当前的科技也有所突破与创新,信息图谱这个范畴也得到了更多的关注。知识图谱分类是对知识图谱的分布式保存、检索以及推理与挖掘工作的重要基础与前提条件,所以知识图谱的分类对于知识图谱技术的应用与发展前程产生的影响意义重大。所以,本章着重对知识图谱划分算法课题展开了深入研究,谨供参考。

**[关键词]** 知识图谱;划分算法;研究方向

**[DOI]** 10.12252/j.issn.2096-6261.2021.10.2026

## 引言

随着时间的不断发展我国的科学技术水平也有了质的飞跃,科学技术的不断突破和革新造就了知识图谱的华丽诞生,知识图谱技术的广泛应用给知识图谱划分任务带来了巨大的挑战。基于此,本文专门就知识图谱划分算法课题研究进行了分析研讨,希望相关人士能够交流参考。

## 一、知识图谱介绍

知识图谱其英文名字为Knowledge graph,知识图谱作为一种可视化工具揭示以及显示了科学知识发展进程与结构关系。绘制知识图谱的绘制工具主要有CiteSpace、Ucinet以及Gephi和Bibexcel,所谓知识图谱是指客观世界中实体与概念之间关系利用绘图的形式所表现出来一种的知识体系,知识图谱的搜索引擎获取的信息和内容来源于用户搜索内容的输入,并且当用户增加搜索次数,知识图谱获取的信息和内容范围就越广。知识图谱在字串方面有所创新,不在是以往的简单的字串。同时,知识图谱技术具有连贯性的优势,促进了用户使用知识图谱搜索内容的流畅性。知识图谱体系提供给用户的信息准确度非常高。知识图谱的实际应用主要有三个方面,分别是语义搜索、智能回答问题和决策能力。其构成框架主要包括内容层和数据层。暂时中国尚不能建立知识图谱的划分尺度,但根据知识图谱的形成特征和依赖自动提取技术作为进行知识图谱划分的基础,则知识图谱可分为早期的本体知识图谱,以及在开放的网络环境中由自行提取实体和关系所构成的知识图谱。同时范围也可成为知识图谱划分的基本尺度,若根据范围对知识图谱加以划分,则知识图谱也可包括通用知识图谱和专业知识图谱。

## 二、知识图谱划分算法

### (一) 基本图划分算法

#### 1. 普划分算法

普划分算法的创始人为Donath和Fiedler。普划分算法又称普二分法。普划分算法是知识图谱系统中经常用到的一种划分算法。知识图谱中的基本图划分算法中的普划分算法的划分标准为计算图中的Laplace矩阵的Fiedler向量值,当Fiedler向量值大于2时,适合于运用基本图划分算法中的普划分算法。在普划分算法中计算Fiedler向量值的算法大多采用多级方法。

#### 2. 几何划分算法

几何划分算法是人类知识图谱中的基本图形划分算法中的第二大划分算法,而几何划分算法的划分数据主要取决于顶点的相对位置数据。几何划分算法中最经典的算法是坐标二分法,不过坐标二分法也有着很突出的缺陷,就是

坐标二分法太过度地依赖坐标系,并且因为坐标二分法对坐标系的超级依赖性,而导致的知识图谱系统即使在同一幅图中,在不同的坐标系中所产生的分类结果也不同,中国国内的知识画纸划分及相关科研工作者为克服坐标二分法的缺陷,已经成功地研发出了惯性分类技术,而通过惯性分类技术实现知识画纸中的几何划分算法划分,知识画纸就能够很好地克服了坐标系二分法划分的缺陷。

#### 3. 分支定界算法

分支定界算法是解决NP常用的手段方法,同时也是图划分常用的划分算法之一。所谓分支定界算法是指利用起始问题分解成多个子问题,并且采用递归方式逐一求解,最终锁定最有效地解决计划。在知识图谱的划分算法问题中,需要同时获得上界和下界的值,其中获取的方法不一,半定规划和二次规划是常用的方法。

#### 4. KL算法

普划分算法、几何划分算法以及分支定界算法三者都是在原始图上进行划分的,而KL算法与三者有所不同的是,KL算法划分是指通过创建新的划分计划,以此来改进原始划分计划,不断的重复循环该过程,当计划划分方案不能改进便选择终止。KL算法的创始人Kernighan和Lin.Fiduccia-Mattheyses,FM算法是KL算法的改进算法,其中顶点移动是两者之间最显著的差别,由于两者的顶点移动不同,因此FM算法在划分效率和划分质量方面具有明显的优势。

### (二) 多级图划分算法

#### 1. 基于匹配的算法

基于匹配的算法是多级图划分算法的第一种算法。基于匹配的算法通常包括粗糙化算法、多级图划分算法以及分布式多级图划分算法。粗糙化算法需要得到最大匹配值,而深度优先算法与随机算法是粗糙化算法最大匹配值获取的最佳途径,并且其获取最大匹配值的时效非常快。多级图划分算法又称METIS,其操作过程中运用到了三元组查询负载感知器,多级图划分算法通过三元组查询负载感知器实现了查询问题的高效处理。分布式多级图划分算法又称KaPPa, KaFFPa算法和KaFFPaE算法是KaPPa算法的改进算法。

#### 2. 基于聚合的算法

基于聚合的算法主要是指权重聚合的方法,权重聚合的方法来源于代数中的多重阿格法,代数中的多重阿格法又叫AMG,权重聚合方法同样包含了粗糙化方法,权重聚合中的粗糙化方法包含了三个阶段,分别为选取聚合的种子、通过计算规则得到非聚合种子的权重以及计算求出粗糙化每个顶点之间的连接强度,同时为了保证知识图谱的负载平衡,因此

在选取AMG方法的时候要谨慎。

### （三）流试图划分算法

#### 1. Hash算法

在流试图划分算法中Hash算法是最简单的一种算法，因此Hash算法具有划分算法简单的优点，但是其划分质量比较差，并且跨越分区的顶点数目比较多。Hash算法又称预映射算法，Hash算法利用散列的映射函数来完成对知识图谱的划分计算，1990年是Hash算法的问世时间，DBH算法是Hash算法的改进算法，DBH算法又叫边划分算法，DBH算法的主要优势是能够优先切割更高度数的顶点。

#### 2. 贪心算法

贪心算法又被人们称作贪婪算法，其英文名为greedy algorithm，依据题意选取一种度量标准是贪心算法的核心，贪心算法属于一种比较规则的划分算法，并且贪心算法比较适用于动态知识图谱的划分。其适合于动态知识图谱划分的原因在于贪心算法可以做到同时完成处理输入流和维护度信息。HDRF算法是贪心算法的改进算法，并且其算法复杂程度较高。

#### 3. Fennel算法

Fennel算法是知识图谱中的流试图划分算法的第三大算法。此算法在知识图谱中的流划分算法中较为普遍，并且由于Fennel算法能够实现同时考虑相邻顶点数和非相邻顶点数，使得Fennel算法的流分配质量比Hash算法和贪心算法都高。对于最大化邻近顶点与最小不邻近顶点插值，是Fennel算法的核心内容和关键问题。而Ginger算法则是对Fennel算法的改进算法，Fennel算法比较适合于度数较低的知识图谱划分算法，我们实际生活中的道路交通网络的划分计算就是采用了流试图划分算法中的Fennel算法。并且Ginger算法更适用于有长尾分布的社交网络知识图谱的换分计算。

### （四）分布式图划分算法

#### 1. Kappa算法

Kappa算法的创始人是Holtgrewe，Kappa算法的过程是指以知识谱图的原始划分入手，边切割数作为划分标准，对其开展多次的筛选，从而得出最佳划分结果。提前存储顶点和边是Kappa算法必不可少的准备工作，并且Kappa算法又引入了分布式匹配算法，因此Kzppa算法能够划分大规模的知识图谱。

#### 2. JA-BE-JA算法

JA-BE-JA算法的创始人是Rahimian，JA-BE-JA算法在对知识图谱进行划分计算的过程中运用了局搜索和模拟退火技术。JA-BE-JA算法具有计算简单以及局部性较强的优点，因此不容易陷入局部最优。JA-BE-JA-VC算法是JA-BE-JA算法的改进算法，其该算法的具体流程主要包括知识图谱中的任意一组边随机分配到不同的分区以及进行每个边和每个分区的能量函数定义从而得出是否交换边的结论。

#### 3. 轻量级重划分算法

轻量级重划分算法是针对动态图设计的一种化分计算方法，所谓轻量级重划分算法是指为了符合动态图的划分计算，对已有划分进行再次划分。轻量级重划分算法又可以分为Hermes算法和PLanar算法。由于轻量级重划分算法具有内

存较小的优势，因此轻量级重划分算法可以随时更新顶点。

### （五）标签传播算法

标签传播算法简称LPA，标签传播算法比较适用于静态的知识图谱划分计算，并且在采用标签传播算法来进行知识图谱的划分计算的前提和基础是必须将知识图谱转为无向图谱，但是标签传播算法无法保证负荷平衡，因此标签传播算法无法单独使用，必须与其他的划分算法一起使用。同时标签传播具有计算简单直接以及划分后边切割数较少的优点。

### 三、知识图谱划分算法的未来研究方向

#### （一）知识图谱划分标准与算法效率如何权衡

目前为止我国的知识图谱划分算法绝大多数都是针对一般图划分计算，但是我们实际生活中的知识图谱较丰富与复杂，给我国的知识图谱的划分计算工作增加了难度。顶点和边的切割数目较少、负载平衡以及具有较小的时间复杂度作为以前传统的知识图谱划分标准，传统的知识图谱划分计算标准已经无法适应知识图谱技术的快速发展，因此知识图谱划分算法的相关研究人员需要制定新的既能有利于支持知识图谱查询的快速执行和降低划分算法的复杂度的知识图谱划分标准。

#### （二）面向知识图谱高层语义如何进行知识图谱划分

对认知本体的描述作为认知图谱的重要组成部分，特别重要。比如OWL图，是语义Web开发的知识图谱中的一个，在中国目前为止已掌握了对OWL图图层的语义分类的具体算法，不过中国当前尚没有把OWL的本体语言作为启发信息分类的具体算法，所以通过对知识图谱高层语义信息的分类学习图谱，将是中国未来学习图像分类的开发重点之一。

#### （三）面向知识图谱应用需求如何划分

知识图谱的发展在中国已经获得了初期的成绩，而且在中国的经济建设发展中的运用范围也非常广阔，在各个领域中都有知识图谱发展的身影，比如社会网络、道路交通管理以及电商等。不过鉴于当前的知识图谱方法研发工作还处在初级阶段，所以针对不同专业应用中的具体知识图谱方法分类结果往往还缺乏具体的质量评估准则，从而根据不同的实际使用要求指导分类就是相当关键的研发目标所在。

### 结束语

总而言之，知识图谱是我国科学技术发展的必然趋势，知识图谱通过精确的语言描述了现实世界中的实体拉近了现实世界中的实体彼此之间的关系，是我国的企业发展壮大的重要基石。并且知识图谱划分作为知识图谱分布式处理的核心部分，其作用意义重大，希望知识图谱划分相关工作人员能根据知识图谱划分算法的未来发展方向，完善知识图谱划分工作，以此促进知识图谱的可持续性发展。

### 参考文献

- [1]王鑫,陈蔚雪,杨雅君,张小旺,冯志勇.知识图谱划分算法研究综述[J].计算机学报,2021,44(01):235-260.
- [2]王庆.基于投资图谱的信息挖掘研究[D].南京信息工程大学,2020.
- [3]张翔.大规模语义数据存储优化方案研究与实现[D].西安电子科技大学,2019.