

第一章 基于双哈希的图精确查询

由于传统算法在利用哈希表存储索引中多采用简单哈希，容易产生冲突问题，导致构建索引效率较低。本章在路径索引的基础上，探究了不同哈希方法对算法速度的影响，并提出一种基于双哈希的精确查询算法，来减少图查询中的过滤阶段的耗时，以提高查询速度。本章将先介绍下现有的哈希算法，然后详细介绍基于路径的查询方法包括作为此算法验证方法用的子图同构算法 *ULLMANN*^[2]。最后是实验结果与分析。

1.1 常用哈希方法

本节将介绍几种常用的哈希方法及字符串哈希函数。

补充

1.1.1 链表法

1.1.2 双哈希法

1.1.3 常用字符串哈希函数

1.2 基于路径的查询算法

本方法同 *GraphGrep* 算法^[2]，都是基于路径的精确子图查询算法。算法基本流程如下：(1) 遍历图数据库中图的路径，(2) 利用双哈希构建索引，(3) 遍历查询路径，(4) 利用基本索引特征做先验剪枝，(5) 路径合成进一步筛选候选集，(6) 子图同构确定最终结果。下面我们将分小节详细说明这些步骤。

1.2.1 数据库路径遍历

首先，对于每个数据库我们设定一个路径长度的上限 l_p ， l_p 越大意味着可记录的路径越长，索引集合也会相应增加。随后对于数据库中的每幅图，我们用深度优先搜索遍历每个节点，遍历最大深度为 l_p ，并记录遍历过程中经过的每一条路径，存成一个列表，用于下一步构建索引。

1.2.2 双哈希索引构建

如算法 xxx 所示，效果 table

1.2.3 查询图路径遍历

1.2.4 先验剪枝

1.2.4.1 包含逻辑规则

1.2.5 路径合成

1.2.6 子图同构

1.2.6.1 UHMANN 算法

1.3 实验结果与分析

1.3.1 实验环境

1.3.2 实验数据分析