

## 一种基于图立方体的移动对象网络压缩算法

薛向前, 胡孔法, 李 珑, 王 俊, 陈 峻

(扬州大学 信息工程学院, 江苏 扬州 225009)

**摘 要:** 针对大规模移动对象网络在构建图立方体的过程中产生的大量浓缩图, 引入了图压缩的思想, 提出了进一步压缩浓缩图的算法 MC-compress, 用来合并浓缩图中顶点和相应的边. 通过将图中相邻的两个顶点进行合并, 再比较两条边合并后权重的最大差异度, 从而找出最优的合并顶点对, 最终产生构建压缩图的超级顶点和超级边. 通过压缩图结构, 加快了在图立方体上查询图结构的显示过程, 减少了构建图立方体过程中存储大量浓缩图的空间.

**关键词:** 移动对象网络; 图立方体; 图压缩; 浓缩图

**中图分类号:** TP39

**文献标识码:** A

**文章编号:** 1000-7180(2012)11-0022-03

## A Kind of Compression Algorithm of Moving Object Network Based on the Graph Cube

XUE Xiang-qian, HU Kong-fa, LI Long, WANG Jun, CHEN Ling

(College of Information Engineering, Yangzhou University, Yangzhou 225009, China)

**Abstract:** Aimed at the amount of aggregation graphs of moving objects network produced by building graph cube, introducing the idea of compressing graph, putting forward the algorithm of compressing aggregation graph MC-compress method used to merge vertices and edges in the aggregation graph. Comparing the largest differences of weight of two edges when merging two vertexes in the aggregation graph, and to find the optimal merging pairs of vertexes, finally, generating the super vertexes and super edges to construct compressed graph. The method speeds up the displaying after querying graph structure in graph cube and reducing the space used to store large amounts of aggregation graph in the process of building graph cube.

**Key words:** moving objects network; graph cube; compressing graph; aggregate graph

### 1 引言

近年来随着互联网的飞速发展, 各个领域都已实现网络化, 如传感器网络、生物网络以及社会网络. 这些网络都可以采用图结构的形式进行模型化, 如用图的顶点代表实体本身, 边用来描述实体之间的关系. 除了图本身所具有的拓扑结构之外, 图的节点往往具有多维属性, 就形成了多维网络. 移动对象除了具有多维属性外还具有很重要的时空特征, 那么将移动对象作为图的节点, 节点之间可以相互通信形成移动对象网络.

本文提出的压缩移动对象网络浓缩图的算法, 是通过近似化信息的方法对图进行压缩. 在对非权重图进行压缩的文章中也可以看到类似的处理过程. 例如, 文献[1-2]提出在顶点之间通过合并组成超级节点, 连接超级顶点之间的边是超级边, 非带权图的压缩技术已经被广泛应用在图的存储和处理, 例如, 文献[3]在压缩图的帮助下实现了生物学模型. 文献[4]通过一个拓扑 OLAP 框架的压缩图, 实现了对在线合作图的分析. 不管是否采用基于用户查询的方式<sup>[5-6]</sup>, web 图压缩的目的是尽可能用不同的方式简化图的表示.

**收稿日期:** 2012-05-08; **修回日期:** 2012-06-06

**基金项目:** 国家自然科学基金项目(61070047, 61003180); 江苏省“六大人才高峰”项目; 江苏省“333”项目

## 2 最优化超级边的权重与合并

### 2.1 超级边的权重

给定一个压缩图结构,很容易通过优化简单的距离度量值设置超级边的权重.源顶点的每个顶点对 $\{u, v\} \in V \times V$ 可以由超级顶点集合的一对顶点 $\{u', v'\} \in V \times V$ 表示,包括 $u=v$ 和 $u'=v'$ 的情况.当 $L=1$ 并且 $q(e)=w(e)$ 时,给出超级顶点 $V'$ ,需要最小化超级顶点集的每个顶点对的总和 $\sum_{\{u, v\} \in u' \times v'} (w(\{u, v\}) - w'(\{u', v'\}))^2$ .当超级边的权重就是源边的权重的时候,超级边的权重是最小的(包括那些没有连接任何边的“0权重”的顶点对):

$$w'(\{u', v'\}) = \frac{\sum_{\{u, v\} \in u' \times v'} w(\{u, v\})}{|u'| |v'|}$$

其中 $|x|$ 是超级顶点中源顶点的数量.

### 2.2 节点之间距离的界定

令 $d_{\max}(u, v; G_s)$ 是任何俩条边合并权重的最大差异度,做为 $u$ 和 $v$ 合并的结果:

$$d_{\max}(u, v; G_s) = \max\left\{ \begin{aligned} &\max_{x: \{u, x\} \text{ or } \{v, x\} \in E} (|Q_L(u, x; G_s) - Q_L(v, x; G_s)|), \\ &|Q_L(u, u; G_s) - Q_L(v, v; G_s)|, \\ &|Q_L(u, u; G_s) - Q_L(u, v; G_s)|, \\ &|Q_L(v, v; G_s) - Q_L(u, v; G_s)| \end{aligned} \right\}.$$

第一个元素是和顶点 $x$ 相邻的所有边的最大值,其余的是合并到自己边之间的相异度.

当 $L=1$ 时,很明显有边界

$$d_{\max}(G, G_{s1}) \leq d_{\max}(u, v; G_s)$$

## 3 算法描述

**算法1** 移动对象网络浓缩图的生成算法:

输入:一个移动对象网络 $N=(V, E, A)$ ,属性集合 $A'$

输出:浓缩图 $G'=(V', E', W_v, W_e)$

方法:

- (1) 初始化哈希结构 $H: A' \rightarrow V'$
- (2) for  $u \in v$  do
- (3) if  $H(A'(u)) = \text{NULL}$  then
- (4) 创建一个浓缩顶点 $u' \in V'$ ,标记为
- (5)  $u'.weight \leftarrow 0$
- (6)  $H(A'(u)) \leftarrow H(A'(u)).weight + 1$
- (7) for  $e(u, v) \in E$  do
- (8)  $u' \leftarrow H(A'(u))$
- (9)  $v' \leftarrow H(A'(v))$

(10) if  $e'(u', v') \notin E'$  then

(11) 创建浓缩边 $e'(u', v') \in E'$

(12)  $e'.weight \leftarrow 0$

(13)  $e'.weight \leftarrow e'.weight + 1$

(14) 返回浓缩图

**算法2** 基于移动对象浓缩图的压缩算法:MC-compress

输入:移动对象浓缩图 $G'=(V', E', W_v, W_e)$ ,压缩率 $r(0 < r < 1)$ ,最大路径长度 $L \in N$ .

输出:压缩图 $G_s(V'', E'', W_v'', W_e'')$ ,  $r(G_s) \leq r$ 并且 $d:st(G, dec(G''))$ 的值最小.

方法:

- (1)  $G'' \leftarrow G'$
- (2) while  $r(G'') > r$  do
- (3) 随机选择图 $G''$ 一个顶点 $v \in V''$
- (4) for 所有顶点 $u \in V''$  do  $\{(*)\}$
- (5)  $d_u \leftarrow d_{\max}(v, u; G'')$
- (6)  $z \leftarrow \{u \cup v\}$
- (7)  $V'' \leftarrow V'' \setminus \{u, v\} \cup \{z\}$
- (8) for 所有的 $x \in V'$ ,  $u \neq x \neq v$ 并且 $\{u, x\}$ 或者 $\{v, x\} \in E$  do
- (9)  $w'(\{z, x\}) = \frac{|u| Q_L(\{u, x\}; G'') + |v| Q_L(v, x; G'')}{|u| + |v|}$
- (10)  $w'(\{z, x\}) = \frac{W\{u, u\} + W(v, v) + W(u, v)}{|z|(|z| - 1)/2}$
- (11) 返回 $G''$
- (12) 函数 $W(x, y)$
- (13) if  $x \neq y$  then
- (14) 返回 $Q_L(\{x, y\}; G'') |x| |y|$
- (15) else
- (16) 返回 $Q_L(\{x, y\}; G'') |x| (|x| - 1)/2$

## 4 实验与结果分析

比较两种算法的运行时间,通过对同一移动对象数据库下不同的维度来比较两种算法的效果,实验如图1所示.实验的移动数据库对应的网络的边数为1000.维度数为1-3.

由图1可见,当维度较大时,MC-compress算法的优势比较明显,因为在维度较大时图的结构比较复杂,压缩的效果更加明显,粗糙表算法因为查询时从磁盘中读取整个数据库的数据,其随着维度的增加对每条元组中需要比较的属性数增多,增加

了查询处理的时间,而采用压缩图部分概化的方法,可以从已经概化的图中直接获得.

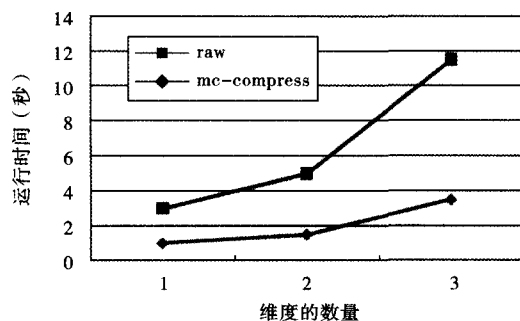


图1 不同的维度下的算法性能比较

## 5 结束语

本文对移动对象结构数据的查询,采用了一种新的对概化结构压缩的方法,来加快查询的速度.提出的新的压缩算法叫做 MC-compress,可以有效压缩图的结构.我们先通过算法 1 将结构数据在指定维度上进行概化,再通过算法 2 将对算法 1 中概化结构比较复杂的结构做进一步的压缩,形成新的压缩结构,压缩结构中的顶点可能包含多个概化结构的顶点,压缩结构的边的权重是顶点之间权重的相对近似值.从而大幅度压缩了结构比较复杂的多维结构数据.

将来的工作是将 MC-compress 应用到更多现实的数据集,并在不同的系统条件下测试它的性能.除此之外,我们也可以将 MC-compress 方法用于更多的应用方面,像频繁模式挖掘、图的聚类等等.我们的目的是解决物联网环境下针对移动对象

网络查询处理图结构数据的问题,为企业提供更多的效益和服务质量.

## 参考文献:

- [1] Navlakha S, Rastogi R, Shrivastava N. Graph summarization with bounded error [C] // Proceedings of the 2008 ACM SIGMOD international conference on Management of data (SIGMOD '08). New York, NY, USA: ACM, 2008: 419-432.
- [2] Tian Y, Hankins R, Patel J. Efficient aggregation for graph summarization [C] // Proceedings of the 2008 ACM SIGMOD international conference on Management of data (SIGMOD '08). New York, NY, USA: ACM, 2008: 567-580.
- [3] Navlakha S, Schatz M, Kingsford C. Revealing biological modules via graph summarization [C] // Presented at the RECOMB Systems Biology Satellite Conference. USA: Boston, 2009: 253-264.
- [4] Chen C, Yan X, Zhu F, et al. Graph OLAP: Towards online analytical processing on graphs [C] // Proceedings of the 2008 Eighth IEEE International Conference on Data Mining. Washington, DC, USA: IEEE Computer Society, 2008: 103-112.
- [5] Faloutsos C, McCurley K S, Tomkins A. Fast discovery of connection subgraphs [C] // Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: ACM, 2004: 118-127.
- [6] Hauguel S, Zhai C, Han J. Parallel pathFinder algorithms for mining structures from graphs [C] // 2009 Ninth IEEE International Conference on Data Mining, IEEE. USA: Miami, 2009: 812-817.