

Algorithm Model of Optimization Path Base on Vector Hyper Graph

Pei Yan¹, Ya ng Guangming²

Software School, Northeastern University, Shenyang, PRC

1. E-mail: peiyan@hotmail.de, 2. E-mail: ygm@sict.ac.cn

Abstract: The article proposes a model of the optimization path based on the data stored in the form of vector hyper graph, compromise of the multi-aims and blindfold searching with the heuristic searching. It gives the definition of the vector hyper graph in the mathematics form, proposes an applicable and integrated definition of the vector hyper graph object. In the engineering practice, there is no the optimization path accord with the multi-aims, and the time used in the calculation process is also one of the aims, this problem is described in the mathematics layout difficultly. It proposes a method of multi-aims converted into one aim. This representation of model is optimizing through the engineering practice using the huge data.

Keywords: vector hyper graph; heuristic algorithms; multi-aims tradeoff; optimization path

一种基于矢量超图的最优路径算法模型

裴岩¹, 杨广明²

东北大学软件学院, 沈阳, 中国, 110004

1. E-mail: peiyan@hotmail.de, 2. E-mail: ygm@sict.ac.cn

【摘要】本文提出了一种基于矢量超图数据存储、多目标折中函数、盲目搜索结合启发式搜索的最优路径数学模型。利用超图理论,给出了一种适用、完整的矢量超图对象模型定义的数学描述。由于实际工程应用通常不存在同时满足多目标的最优路径,而且计算所用时间本身也是目标之一,这种复杂问题很难用数学模型描述,为此提出了一种折中多目标为单目标的方法。通过对海量数据的实际算例应用,模型取得了很好的效果。

【关键词】矢量超图; 启发式算法; 多目标折中; 最优路径

1 引言

最优路径问题^[1-4]是单源最短路径问题的一种概率近似,单源最短路径问题是许多领域内经常出现的一类问题,在现实生活中的许多实际问题经转化后都可以归结为单源最短路径问题。但是在现实生活应用中这类问题抽象的数学模型是:在简单有向图中每条边上的权值种类不仅仅为一个,这就形成了多目标的最优路径问题^[5]。文献[6][7]结合地理信息系统的特点,对最优路径问题做了有益的研究,但未应用到实际工程中去。

本文从超图理论的一般概念出发,给出一些基本的定义和结论,为本文论述的矢量超图的定义提供一些必要理论依据。最优路径问题在优化组合和图论中已经有了一些基本的算法,因此,笔者将首先引入超图同构的概念,为矢量超图对象模型的建模和模型同构性的证明提供支撑,并在此基础上说明矢量超图对

象模型在最优路径问题中,各个元素与实际问题域的抽象对应关系,继而给出基于这个模型的盲目搜索结合启发式搜索的最优路径问题搜索算法。最后给出一些在仿真实验中统计的结果,提出一些在工程应用中所应注意的问题。

2 矢量超图对象模型

超图概念是C.Berge^[8]于1970年提出的,文献[9]第一次系统地阐述了超图的基本概念,建立了无向超图理论,此后有关超图理论及其应用的发展也越发深入。为了将超图理论用于计算机科学的相关领域,在此基础上还发展了有向超图的理论。

2.1 矢量超图

矢量超图是基于超图理论中,简单一致超图概念建立起来的,其意义在于基于这种超图概念而建立的矢量超图对象模型可以更好地解决最优路径的排列

组合问题。

矢量超图的数学定义为：

$$G = \{V, \delta\} \tag{1}$$

其中 $V = \{v_1, v_2, \dots, v_n\}$ 是一个点的非空有限集合， δ 是关于 V 的各种关系的总的集合，

$$\delta = \{\varepsilon_k \mid k \in I\} \tag{2}$$

ε_k 是 δ 关系集合的一个覆盖， ε_k 是根据 V 中元素 v 的层次属性不同而组成的关系的集合，

$$\varepsilon_k = \{E_{ki} \mid k \in I, i \in I\} \tag{3}$$

E_{ki} 是 ε_k 关系的一种划分，

$$E_{ki} = \{R_{kij} \mid k \in I, i \in I, j \in I\} \tag{4}$$

其中 R_{kij} 是 E_{ki} 中的一种矢量关系。

对于给定的 k ， E_{ki} 是 V 的拓扑积子集，并且满足：

- (1) 当 $k=1$ 时， $E_{ki} \neq \phi$ ， $\bigcup E_{ki} = V$ ；
- (2) $k \geq 2$ 时， $E_{ki} \cap E_{(k+1)i} \neq \phi$ 。

并且为了满足简单一致性，矢量超图中，对于任意给定的 $k \in I$ ，有： $E_{ki} \subset E_{kj} \Rightarrow i = j \quad i \in I, j \in I$ 。

ε_k 中， E_{ki} 所包含的点元素基本是相同。这样规定是为了在实际的工程应用中， E_{ki} 可以被同样的数据关系结构所表示。矢量超图是一种简单一致超图，在矢量超图中，边集的关系被定义为 m 个 n 元组的形式，即：对于任意给定的 $k \in I$ ， E_{ki} 中包含 m 个 $R_{kij} = \langle v_{kij1}, v_{kij2}, \dots, v_{kijn} \rangle$ ，记为：

$$R_{kij} = \langle v_{kij1}, v_{kij2}, \dots, v_{kijn} \rangle \tag{5}$$

这样定义使得在每种数据关系与实际情况的网络或者道路的拓扑情况相一致，可以很好的满足逻辑抽象与现实世界的一致性。

2.2 矢量超图对象模型

根据上述定义的矢量超图的概念，定义矢量超图对象模型。在矢量超图对象模型中，有限集 V 是实际道路数据中的节点，节点关系 E 是 n 个有拓扑关系的道路数据关系， ε 定义了数据节点和数据关系的层次结构，即在 $\langle V, \varepsilon_k \rangle$ 和 $\langle V, \varepsilon_{k+1} \rangle$ 之间存在着如下的同构关系，定义 φ 是 $\varepsilon_k \xrightarrow{\varphi} \varepsilon_{k+1}$ 的映射关系，则：

$$\varphi(\varepsilon_k) = \begin{cases} 1 & \exists E_{k+1} \in \varepsilon_{k+1}, v_{\varepsilon_k} \in E_{k+1} \\ 0 & o.w \end{cases} \tag{6}$$

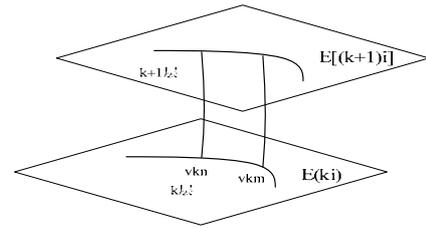


图 1 矢量超图对象模型层次关系的同构性

在 $\varepsilon_k = \{E_{ki} \mid k \in I, i \in I\}$ 上定义置换关系 π 为：

$$E_{\pi(ki)} = \pi(E_{ki})$$

$$\pi(E_{ki}) = \begin{cases} 1 & \exists v_{km} \in V, v_{kn} \in V, \exists E_{ki} \in \varepsilon_k, v_{km} \in E_{ki}, v_{kn} \in E_{ki} \\ 0 & o.w \end{cases} \tag{7}$$

特别地，如果 $v_{kn} \notin V$ 或者 $v_{km} \notin V$ ，则

$$\pi(v_{km}, v_{kn}) = \Delta \circ$$

下面证明矢量超图对象模型关系是同构的。

设存在两个代数系统 $\langle V, \varepsilon_k \rangle$ 和 $\langle V, \varepsilon_{k+1} \rangle$ ，有 $v_{kn} \in V$ 和 $v_{km} \in V$ ，根据上述的矢量超图对象模型的定义， v_{kn} 和 v_{km} 有如下关系：如果 v_{kn} 和 v_{km} 在一个关系 R_{kij} 中，并且 $R_{kij} \in E_{ki}$ ，则： $\varphi(E_{ki})=1$ 。又因为有 $v_{kn} \in V$ 和 $v_{km} \in V$ ，所以在 $\langle V, \varepsilon_{k+1} \rangle$ 一定存在一个关系 $R_{(k+1)ij}$ 中同时存在 v_{kn} 和 v_{km} ，则： $\pi(E_{(k+1)i})=1$ 即： $\varphi(E_{ki}) = \pi(E_{(k+1)i}) = E_{[\pi(k+1)i]}$ 。

成立，根据文献[9]所述的两个超图同构的定义，可得矢量超图对象模型关系是同构的，参见图 1。

3 多目标最优路径的算法模型

3.1 算法的主要思路

在实际工程应用中，选择最优路径的评价函数所考虑的参数往往不是一个，所以通常不存在同时满足多个目标的最优路径，而且算法所基于的道路数据是海量的，在有限的时空内，要得到在资源有限的条件下的最优路径，不得不考虑多个目标的折中和算法执行时空的有效利用。基于这种考虑，多目标最优路径的算法在进行目标搜索的过程中，在 k 比较小的时候，一般定义阈值为 k_λ ，当 $k \leq k_\lambda$ 时，执行传统的 *Dijkstra* 算法，收集足够用于扩张的拓扑关系 R_{kij} 。在算法执行到特定的时空时，即在 $k > k_\lambda$ 时，确定唯一的 $k = k^* (k^* > k_\lambda)$ ，用启发式的 A^* 算法扩张探索。当执行算法到一定程度时，由算法终止的判定条件决定算法终止。在算法执行的整个过程中，每次扩张探索都需要多目标折中函数的控制，所以多目标折中函数的设计在一定程度上，影响着算法执行与收敛的质量。

3.2 k_λ 的确定

在算法执行的过程中,收集到足够用于扩张的拓扑关系 R_{kij} 是决定算法继续执行和收敛速度的关键因素,这里给出一种基于时间考虑的 k_λ 参数确定方法。设 $P_1(x), P_2(x), \dots, P_{k_\lambda}(x)$ 是 k 分别取 $k \in [1, k_\lambda]$ 时的由概率计算统计出来的时间消耗, x 是所要寻优的两点之间距离的变量,暂设 $P_i(x) \sim N(\frac{d_{i\max}}{2}, (\frac{d_{i\max}}{3})^2)$ 的正态分布, 设用于执行 *Dijkstra* 算法的时间阈值为 t_{\max} , 则有:

$$\sum_{i=1}^{k_j} \int_{t_i}^{t_{i+1}} P_i(x) dt < t_{\max} \quad (8)$$

k_λ 由上式确定, $k_\lambda = \max\{k_j\}$ 。

在 $k = k^*$ 时,执行 A^* 算法进行目标搜索可以选用单方向搜索和双方向搜索不同的搜索策略。在理论上,双方向搜索策略在搜索空间上约减少 50%,其运行时间比单向搜索要短,时间复杂度为 $O(n/2)$, 理论上的搜索范围见图 2。

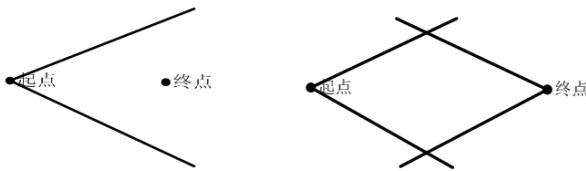


图 2 理论上单双向探索的拓扑覆盖情况 (左: 单方向 右: 双向)

在算法的设计上,可以设计为用单一线程执行单一方向的搜索,这样省去了执行双向探索的线程切换的时间消耗,拓扑数据情况见图 3。也可以设计为切换两个搜索线程分别执行一方的搜索,但是这样实际系统在线程切换时会消耗一些时间,时间复杂度要比理论上高一些,仿真试验结果中的拓扑数据情况见图 3。在搜索过程中参照 3.4 节所述的多目标函数控制搜索方向。



图 3 实际单双向探索的拓扑情况 (左: 单方向, 右: 双向)

3.3 k^* 的确定

算法在矢量超图对象模型中确定 k^* 直接影响了算法成功与否, k^* 值太大, 矢量超图对象模型中, 没有相应的拓扑关系, 在 A^* 算法进行目标搜索的过程中必然失败; k^* 值太小, 需要遍历许多拓扑关系和关键节点, 算法执行效率低。所以正确、合理地选择 k^* 也是算法要关注的问题之一, 这里给出一种基于权衡时间和空间因素的 k^* 参数确定方法。

设 $Q_{k_{i+1}}(x), Q_{k_{i+2}}(x), \dots, Q_{k^*}(x)$ 是 k 分别取 $k \in [k_\lambda, k_{\max}]$ 时的由概率计算统计出来的时间消耗, x 是所要寻优的两点之间距离的变量, 同上暂设 $Q_i(x) \sim N(\frac{d_{i\max}}{2}, (\frac{d_{i\max}}{3})^2)$ 的正态分布, 则 k^* 应满足如下关系式:

$$k^* \leq \min\{\int_{t_i}^{t_c} Q_k(x) dt\} \quad (9)$$

定义 D_i 是在 $k = i, i \in I$ 时的连通距离, 即: $\exists d$, 当 $d \leq D_i$ 时, 在 $k \leq i$ 时, 可以找到从两点直线距离小于等于 d 的连通枝, D_i 在构造矢量超图对象模型时确定。则 k^* 还应满足: $k^* \leq i$ 。综上,

$$k^* = \min\{i, \int_{t_i}^{t_c} Q_k(x) dt\} \quad (10)$$

3.4 多目标折中函数的确定

将多目标折中转化为单目标, 需要考虑各目标在接近目标理想点的权重, 本文给出一种以距离函数为主要考虑因素, 结合算法执行时间的数学描述模型。设 R 是目标是实际值, R^* 是目标的理想值, R 一般被定义为维数为 n 的矢量空间, $n-2$ 为目标个数, 则定义多目标折中函数为:

$$M(r) = \|R - R^*\| + \int Q_k(x) dt \quad (11)$$

设 z 和 z^* 分别是 R 和 R^* 中第 j 个目标, w_j 是第 j 个目标的权重, 则:

$$\|R - R^*\| = \sqrt{\sum_{j=1}^n w_j |z_j - z_j^*|^2} \quad (12)$$

当 $k = k^*$ 时, 执行 A^* 算法, 将最接近理想点的解作为下一步扩张探索的起点。

3.5 算法的终止和收敛判定条件

在工程应用上述多目标最优路径的算法模型时, 在判定算法终止和收敛时存在着不同的标准, 为了权衡算法的效率和算法所提供路径的质量, 有些时候不得不在二者之间做出妥协。效率优先时, 只要算法在 $k = k^*$ 时, 用 A^* 算法找到一条两点之间的路径即认为

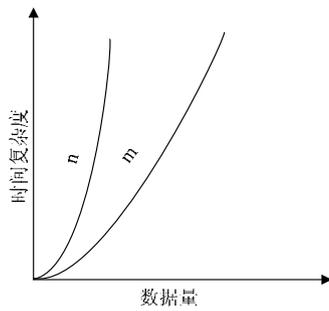


图 4 算法的时间复杂度

最优，算法收敛终止。质量优先时，用A*算法找到n条两点之间的最优路径之后，算法继续执行到第n+m条路径之后，所有路径的质量与理想质量之间满足：

$$\min\{(\sum_i \|R - R_p^*\|) - R_p^*\} < \omega$$

$$i = 1, 2, \dots, n, \dots, n + m$$

R_p^* 为算法理想点， ω 为算法可接受阈值，算法收敛终止，满足上式的路径*i*即为最终的最优路径。

3.6 算法时空复杂度分析

设算法所基于的地图数据 $k = \{1, 2, 3, 4\}$ ，其中 $k_\lambda = \{1, 2, 3\}$ ， $k^* = 4$ 。在 $k = i$ ， $i = \{1, 2, 3, 4\}$ 时的数据量为 $\frac{n}{2^{i-1}}$ ，假设应用 3.4 所述的多目标折中函数，每一步确定的下一步扩张探索的起点即为理想点解。

设整个最优路径每层的拓扑数据量为 m_i ， $i = \{1, 2, 3, 4\}$ ，即整个最优路的拓扑数据量为 $\sum_{i=1}^4 m_i$ ，在最好情况下每次选择的数据即为最优路径的数据，其时间复杂度为 $o(\sum_{i=1}^3 m_i^2 + m_4)$ ；最坏情况下，需要遍历每个数据才能确定，其时间复杂度为：

$$o(\sum_{i=1}^3 (\frac{n}{2^{i-1}})^2 + \frac{n}{2^3}) \tag{13}$$

整个算法在时间复杂度

$$o(\sum_{i=1}^3 m_i^2 + m_4) \text{ 到 } o(\sum_{i=1}^3 (\frac{n}{2^{i-1}})^2 + \frac{n}{2^3}) \tag{14}$$

见图 4，在曲线 m 和 n 在之间的部分。

4 仿真实验结果

本文所述的最优路径算法模型在 SH4 嵌入式 CPU，32M 内存的硬件平台和 WINCE 5.0 操作系统的实验环境下，对于现存的海量实验数据，选取了现实地理长度为 3000 多公里的起点和终点，进行仿真实验，并对于实验结果进行了分析。

表 1 算法实验耗时(Unit: ms)

Item	1 次	2 次	3 次	4 次
起点 $k_\lambda = 1$	30	40	38	25
终点 $k_\lambda = 1$	49	45	225	48
起点 $k_\lambda = 2$	28	31	23	26
终点 $k_\lambda = 2$	16	16	17	16
起点 $k_\lambda = 3$	358	348	233	358
终点 $k_\lambda = 3$	293	279	251	277
$k^* = 4$	1521	1537	1506	1664
总计	2295	2296	2293	2414

实验过程中，在 1 到 3 层数据进行双方向探索，在 4 层数据进行单方向探索，在每层数据探索阶段，统计每个过程的系统时间差，得到算法的时间耗时，参见表 1。

表 2 算法数据量(Unit: 56byte)

Item	数据量
起点 $k_\lambda = 1$	1077
终点 $k_\lambda = 1$	317
起点 $k_\lambda = 2$	1400
终点 $k_\lambda = 2$	1400
起点 $k_\lambda = 3$	2500
终点 $k_\lambda = 3$	2500
$k^* = 4$	1786
总计	10980

算法过程中，对于每一层数据，算法确定的收敛条件为：在第 1 层要找到一定数量上层映射关系后才能进行 2 层探索；在第 2 和 3 层设定固定的探索关系，即 2 层 1400 个拓扑关系，3 层 2500 个拓扑关系，对应于程序就是代表拓扑关系的结构体 (struct)，在 4 层按照 3.5 节所述的算法的终止和收敛判定条件中提到的连接即收敛的原则，结束算法。对于同一个地点，算法探索所用的拓扑关系 R_{kij} 统计，即数据量见表 2。

进行多次仿真试验结果表明，在现有数据条件下，如果数据连通性良好，算法是在资源有限的嵌入式硬件平台条件下收敛的。

5 结束语

本文在超图理论的基础上，给出了一种适合最优路径算法建模的矢量超图的数学描述，在此基础上定义了矢量超图对象模型，推广了超图理论在最优路径算法领域的建模应用，拓展了相关概念，这

是本文的一点创新。在矢量超图对象模型的基础上,给出了多目标最优路径算法模型的执行步骤和关键参数的确定方法,讨论了 k_1 参数、 k^* 参数的确定和多目标折中函数的构造和应用问题。实际工程应用中,给出了一种权衡系统时空性能和最优解的算法收敛方法。

但是算法应用的关键还在于海量数据的制作和连通性的设定,这也成为算法能够大量实际应用的瓶颈。通过实际系统对于现存的海量数据的算例应用,模型取得了很好的效果。在理论与实践上,均证明这种基于矢量超图的最优路径算法模型定义是完备的,算法是可以收敛的。

References (参考文献)

- [1] DeoN, PangCY. Shortest-path algorithms: Taxonomy and annotation [J]. Networks, 1984, 14: 275-323.
- [2] Zhan B F. Three fastest shortest path algorithms on real road networks [J]. J Geogr Inf Decis Anal, 1995, 1(1): 69-82.
- [3] Dreyfus S E. An appraisal of some shortest-path algorithms [J]. OperatRes, 1969, 17: 395-412.
- [4] Cormen T H, Leiserson C E, Rivest R L, et al. Introduction to Algorithms (2nd Edition) [M]. Boston: TheMITPress, 2001.
- [5] 戴树贵. 求带多个限制条件的单源多权最短路径算法[J], 计算机应用与软件, 2004(12)
DaiShugui, An Algorithm for the Multiply Weights Shortest-Path Problem with Multiply Constraints[J], Computer Application and Software, 2004(12)
- [6] 唐文武, 施晓东, 朱大奎. GIS 中使用改进的 Dijkstra 算法实现最短路径的计算. 中国图像图形学报, 2000, 5A(12): 1019
TANG Wen wu; SHI Xiao dong; ZHU Da kui The Calculation of the Shortest Path Using Modified Dijkstra Algorithm in GIS Journal of Image and Graphics, 2000, 5A(12):1019
- [7] Agrawal R, Jagadish H V. Materialization and incremental update of path information Proceedings of the 5th International Conference on Data Engineering. Los Angeles, 1989: 374
- [8] 许小满. 超图理论及其应用[J]. 电子学报. 1994, 22(8): 65-72. Xu Xianman. Hyper graph Theory with Applications[J]. ACTA ELECTRONICA SINICA. 1994, 22(8):65-72.
- [9] C. Rerge. Graph and Hyper graph[M]. Amsterdam: North-Holland, 1973.