一个由接口路径求 Hamilton 回路的算法研究

刘 超 王文杰
(中国科学院研究生院(本部) 北京 100049)

摘 要 为了求简单图中的所有 Hamilton 回路,首先,提出了一种对集合幂集进行编码的算法,引入了接口路径的概念,将 Hamilton 回路的运算转换为等级接口路径矩阵的运算;其次,结合肖尔茨猜想的证明,对算法复杂性的上限进行了估算;最后,以中国旅行商问题为例,给出了求解 CTSP 的精确算法。

关键词 哈密顿回路,中国旅行商,组合优化,接口路径,无穷悖论

中图法分类号 TP301.6 **文献标识码** A

Research on Hamiltonian Cycle Based on Path with Interface

LIU Chao WANG Wen-jie
(Graduate University Chinese Academy of Sciences, Beijing 100049, China)

Abstract In order to find all Hamiltonian cycles in a digraph, to begin with, it presented an encoding method for the power set, which converts the problem of Hamilton circuits into the computation of hierarchical matrix. Secondly, it estimated the complexity of algorithm with the proof of Xiaerci guess. Finally, it gave the exact algorithm for CTSP.

Keywords Hamiltonian cycle, CTSP, Combination optimization, Interface path, Infinity paradox

判断图中是否存在哈密顿回路问题是 NP 完全问题,求哈密顿回路中最短路径的 TSP 问题是 NP 难问题[20]。随着对 Hamilton 回路问题的深入理解,使用精确算法求解的研究逐步减少,各种近似算法已形成主流。然而,文献[1,2]分别从通路性质与存储方式出发,试图探索求取图中所有 Hamilton 回路的精确算法,在获得一定优化效果的同时,也表明精确算法仍然有一定的研究空间。中国旅行商问题(Chinese Travelling Salesman Problem, CTSP)是以中国的 31 个城市为模型的最优 TSP 问题,由靳藩教授在文献[12]中提出。近 20 年来,中外学者进行了相关的研究工作,然而所有的研究基本上都局限于近似算法——蚁群算法、遗传算法,或其他一些有针对性的近似优化算法[4,5,7-10],目前没有使用精确算法给出最优解的文献发表——CTSP 问题在最初 10 年(1991—2000)中不断地有新的优秀解出现,文献[20]等在 2000 年左右发布其满意解为 15404 后很少有更好的解发布,但是,文献[9]2004 年 6 月发布其满意解为 15379,文献[4]2007 年 5 月发布其满意解依然为 15404,文献[8]2008 年 4 月发布其满意解为 15378,介于 CTSP 近似算法研究的现状问题,有必要对 CTSP 的最优解进行研究考证。另外,关于使用近似算法获得旅行商问题最优解的时间、性能分析,已经有作者进行了相关研究,文献[3]分析了循环 LK 算法瞬时得到最优解的概率,结论是对于大规模 TSP 问题,随着运行时间的增加,得到最优解的概率是降低的。无论文献[3]的结论在今后的实践中能否得到验证,关于 TSP 问题最优解的研究都是近似算法性能评价的基本依据。本文通过引入一种路径编码算法,讨论了在空间复杂性允许的情况下,如何降低 Hamilton 回路问题的时间复杂性,并探讨了最优解组成的复杂性,最后给出了 CTSP 问题的精确算法以及求解结果。

1 相关定义

图、边、顶点、邻接矩阵、导出子图、初级通路等概念参考文献[6]等相关文献。本文仅讨论简单有向图的初级通路问题。首先给出几个后文会用到的有关定义:

1. 初级通路的连接:若 ζ_1,ζ_2 是有向图 $D=\langle V,E\rangle$ 的两条初级通路,其中 $\zeta_1=V_{10},V_{11},V_{12},\cdots,V_{1k}$, $\zeta_2=V_{20},V_{21},V_{22},\cdots,V_{2t}$ 如果 $V_{10},V_{11},V_{12},\cdots,V_{1k},V_{20},V_{21},V_{22},\cdots,V_{2t}$ 中除 $V_{1k}=V_{20}$ 外,其余顶点互不相同,则称 ζ_1,ζ_2 能够连接,记 $\zeta_1\oplus\zeta_2=\zeta$,如果 ζ_1,ζ_2 不能连接,记 $\zeta_1\oplus\zeta_2=\emptyset$。

2. 接口路径:由连接定义可知,ζ_1,ζ_2 能否连接与 ζ_1,ζ_2 中间点的顺序无关,本文把以 V_{i0} 为起点 V_{ik} 为终点的初级通路集记作 $IP(\Gamma_{V_{i0}V_{ik}})$,并把这种记法称为初级通路的接口表示法。如果能用一个数值 x 描述该初级通路集经过的所有中间点序列,则可以进一步地表示为 $IP(\Gamma_{V_{i0}V_{ik}})=x$。本文把使用接口表示法表示的初级通路称为接口路径(由于初级回路不存在连接问题,因此不讨论回路的接口表示问题)。

3. 初级通路以及接口路径间的关系表示:如果初级通路 ζ 在接口路径 $IP(\Gamma_{V_{i0}V_{ik}})$ 中,记 $\zeta\in IP(\Gamma_{V_{i0}V_{ik}})$;如果 $\forall\zeta\in IP(\Gamma_{V_sV_t})\rightarrow\zeta\in IP(\Gamma_{V_iV_j})$,记 $IP(\Gamma_{V_sV_t})\subseteq IP(\Gamma_{V_iV_j})$。

2 基本定理

引理 1 若 V 是含有 n 个元素的有限集,$P(v)$ 是 V 的幂集,则存在一个从 $P(v)$ 到 B 的双射函数 f,其中 $B=\{x|0\leqslant$

到稿日期:2009-10-20 返修日期:2010-01-28

刘 超(1975—),男,硕士生,高级工程师,主要研究方向为计算机应用,E-mail:liu_chao@yahoo.cn;**王文杰**(1964—),男,博士,副教授,主要研究方向为智能信息处理。

$x \leqslant 2^n-1, x \in \mathrm{N}\}$，N 是自然数集合(后文讨论的自然数集合均包括零，不再说明)。

证明：设 $V=\{V_0, V_1, V_2, \cdots, V_{n-1}\}$，则 $P(V)=\{\emptyset, \{V_0\}, \{V_1\}, \{V_0V_1\} \cdots V\}$

$$令\ f(A)=\begin{cases}0, \text{if } A=\emptyset \\ 2^{x1}+2^{x2}+2^{x3}+\cdots+2^{xk}, \text{if } A=\{V_{x1}, V_{x2}, \\ V_{x3}, \cdots, V_{xk}\}, 0 \leqslant xi \leqslant n-1\end{cases}$$

其中，$A \in P(v)$。

先给 $f(A)$ 提供一个解释：令 $f(A)$ 为一个 n 比特(bit)的二进制数，把集合 V 的 n 个顶点编号为 $0,1,2,\cdots,n-1$，若集合 A 中含有编号为 k 的顶点，则相应 $f(A)$ 的第 k 比特为 1，否则为 0。

由二、十进制数的对应关系可知：

(1) $\forall A \in P(v)$，有 $0 \leqslant f(A) \leqslant 2^n-1$，即存在且仅存在一个 $b=f(A) \in B$。

(2) $\forall b \in B$，存在且仅存在一个 $A \in P(v)$，满足 $f(A)=b$。

综合(1)、(2)可知，f 是 $P(v)$ 到 B 的双射函数。

含有 n 个元素的有限集 V，其幂集 $P(v)$ 的元素个数是 2^n，这是集合论中已经得到证明的定理，本引理的主要意义在于：

1. 给出幂集元素的编码方案(为推论 1 提供理论基础)。

2. 构造一个一一对应函数(为推论 2 提供理论原型)。

推论 1 若 $D=\langle V,E\rangle$ 是含有 n 个顶点的简单有向图，则存在一个 $n \times n$ 全接口路径矩阵 $IP(D)=(a_{ij})_{n\times n}$，满足 $a_{ij}=IP(\Gamma_{V_iV_j})$。

证明：构造一个长度为 2^n 比特的二进制数 x，将顶点集 V 所有子集一一映射到 x 的每一位(如定理 1 证明中编码方案，分别映射到 $0 \sim 2^n-1$ 比特)。若 V_i 到 V_j 没有通路，记 $a_{ij}=0$，若 V_i 到 V_j 有通路，则把所有通路中间点集合对应 bit 位置 1。

这样，就得到了 $IP(D)$ 的表示。

将定理 1 中的 V 推广到可数无限集合(如自然数集)，可得到如下结论。

推论 2 自然数集合的任一子集，均可对应于一个自然数编码。

证明：依据定理 1 中的编码规则，我们给任一自然数集的子集 A 指定一个自然数编号：

$$f(A)=\begin{cases}0, \text{if } A=\emptyset \\ 2^{x1}+2^{x2}+2^{x3}+\cdots+2^{xk}, \text{if } A=\{x1, x2, x3, \cdots, xk\}\end{cases}$$

结合文献[6]中关于集合等势的定义以及康托尔定理的证明，可得到实无穷系统中一个悖论：

(1)由文献[6]可知：自然数集合的势小于其幂集的势。

(2)由推论 2 可知，自然数集合与其幂集是等势的。

文献[13-15]均对实无穷的概念提出了质疑，文献[16-18]在未明确否认实无穷的基础上，对潜无穷观点进行了相关的哲学分析，指出了实无穷理论中的一些逻辑局限。本文以实无穷逻辑系统中一一对应的概念为前提，构造了一个新悖论，从而表明康托尔关于无穷集合等势的说明以及连续统假设的相关基础是存疑的[1]，这为文献[13-18]提供了直接的证据(由于实际应用中只考虑有限集，该悖论仅具有理论意义)。

推论 2 给出了任意一个由自然数组成的集合的编码公式，将该公式应用于推论 1，可得到与顶点数无关的、具有形式上意义的 $IP(D)$(后文讨论除特殊说明外，均采用该规则描述接口路径)，为了计算上的方便，把长度为 K 的接口路径放在一个矩阵中，记作 K 级接口路径矩阵 $IP_k(D)=(a_{ij})_{n\times n}$，并称之为等级接口路径矩阵，简称接口矩阵。显然，1 级接口矩阵与邻接矩阵在数值完全相同。同时，把矩阵元素与接口矩阵的关系记作：$(IP_m)_{ir} \in IP_m(D)$。

下面定义接口路径的相关运算规则(常数需要依赖于确定的最大编号 n)：

1. 常量 $S(p)$：所有包含顶点 p 的接口路径值。

2. 常量 $SP(m)$：所有中间点个数为 m 的接口路径值。

3. 运算 $ADD(\Gamma,p)$：为接口路径中每一个元素增加一个编号为 p 的中间点，若接口路径 $IP(\Gamma_{V_{i0}V_{ik}})=x$，则有 $ADD(\Gamma,p)=(x\&(\sim S(p)))<<2^p$，其中～表示按位取反运算，& 表示按位取与运算，<<表示对值做左移运算(原来接口路径中含有 p 点的通路将被删除)。

4. 运算 $ADD(\Gamma,A)$，$A=\{a_1,a_2,\cdots,a_k\}$：为接口路径中每一个元素增加一个中间点集合 A，通过递归调用 $ADD(\Gamma,p)$ 可以得到 $ADD(\Gamma,A)$，对应类 C++代码如下：

```
for(int i=1;i<=k;i++)
{
  IP(Γ)=ADD(Γ,a_i);
}
```

5. 运算 $(IP_m)_{ir} \oplus (IP_n)_{rt}$：接口路径的连接运算。若 $(IP_m)_{ir}=x$，$(IP_n)_{rt}=y$，对应类 C++代码如下：

```
s=0;
t=0;
for(int i=1;i<=k;i++)
{
  if(bit(y,i))
  {//the bit i of y is 1
   t=x<<i;
   s|=t;
  }
}
s=s&(~S(r))<<2^r;//delete point r
s=s&SP(m+n-1);
```

其中，k 为数值 y 的二进制位数，$bit(y,i)$ 表示 y 的第 $i(1\sim k)$ 位为 1，s 为运算结果。限于篇幅，证明从略，同时，该运算类似于二进制乘法运算，可以通过相关硬件技术来优化，不进一步讨论。

6. 运算 $IP_m(D) \oplus IP_n(D)$：令 $IP_m(D) \oplus IP_n(D)=IP_s(D)$，则有：$(IP_s)_{ij}=\sum_{r=1,r\neq i,r\neq j}^{n}(IP_m)_{ir} \oplus (IP_n)_{rj}$，$(i \neq j)$，其中，求和符号 $\sum_{r=1,r\neq i,r\neq j}^{n}$ 表示对得到的接口路径值作位或运算。

定理 2 若 $D=\langle V,E\rangle$ 是一顶点数为 N 的有向图，则有 $IP_m(D) \oplus IP_n(D)=IP_{m+n}(D)$；$(m+n)<N$。

证明：令 $IP_m(D) \oplus IP_n(D)=IP_s(D)$，

(1) $\forall IP(\Gamma)=(IP_s)_{ij}$，由路径连接运算定义可知，满足 $d_{(IP(\Gamma))}=m+n$，所以有 $IP(\Gamma) \in IP_{m+n}(D)$

(2) $\forall IP(\Gamma)=(IP_{m+n})_{ij} \in IP_{m+n}(D)$，由路径定义可知，$\forall \zeta=(V_{i0}V_{i1}V_{i2},\cdots,V_{i(m+n)}) \in IP(\Gamma)$，$\exists V_t$ 使得 ζ 可以分解为两条路径：$\zeta_1=(V_{i0}V_{x1}V_{x2},\cdots,V_t)$，$\zeta_2=(V_tV_{y1}V_{y2},\cdots,$

[1] 相应地，文献[6]中关于实数集合与自然数集合不等势的证明，也可引出另外一个悖论，限于篇幅，本文不进一步讨论。

$V_{i(m+n)}$)，满足 $d_{(\zeta_1)}=m, d_{(\zeta_2)}=n$

所以 $\exists IP(\Gamma_1)\in IP_m(D), IP(\Gamma_2)\in IP_n(D)$，满足 $\zeta_1\in IP(\Gamma_1), \zeta_2\in IP(\Gamma_2)$，即有 $\zeta\in IP(\Gamma_1)\oplus IP(\Gamma_2)$，所以有 $IP(\Gamma)\in(IP_m(D)\oplus IP_n(D))=IP_s(D)$

综合(1)、(2)可知：

$IP_m(D)\oplus IP_n(D)=IP_{m+n}(D)$

推论 3 若 $D=\langle V,E\rangle$ 是含有 n 个顶点的简单有向图，设 $D'=\langle V',E'\rangle$ 是 D 关于顶点 V' 的导出子图，满足 $V'=\{u\mid u\in V, u\neq V_0\}$，$V_0\in V$，即 D' 有 $n-1$ 个顶点，则 D 的 Hamilton 回路问题可转换为 D' 的 $(n-2)$ 级接口矩阵问题。

证明：令 D 中以 V_0 为起点的边的终点集为 Ω_s，以 V_0 为终点的边的起点集为 Ω_e，由 Hamilton 回路定义可知：

(1) D 中任一 Hamilton 回路都可表示为 $\zeta(V_0V_s)\oplus\zeta(V_sV_{x0}\cdots V_{x(n-3)}V_e)\oplus\zeta(V_eV_0)$，其中 $V_s\in\Omega_s, V_e\in\Omega_e$，且有 $d_{(\zeta(V_sV_{x0}\cdots V_{x(n-3)}V_e))}=n-2$，由等级接口矩阵的定义可知，$\forall d_{(\zeta(V_sV_{x0}\cdots V_{x(n-3)}V_e))}=n-2$，$\exists(IP_{(n-2)}(D'))_{ij}$，且满足 $\zeta(V_sV_{x0}\cdots V_{x(n-3)}V_e)\in(\mathrm{IP}_{(n-2)}(D'))_{ij}$。

(2) D' 中满足 $V_s\in\Omega_s, V_e\in\Omega_e$ 的任一个 $IP(\Gamma_{V_sV_e})$，有 $\forall\zeta\in IP(\Gamma_{V_sV_e})_{d(\zeta)}=n-2$，所以 $d(\zeta_{V_0V_s}\oplus\zeta\oplus\zeta_{V_eV_0})=n$，即 $\zeta_{V_0V_s}\oplus\zeta\oplus\zeta_{V_eV_0}$ 一定是 D 中的 Hamilton 回路。

综合(1)、(2)可知：D 的 Hamilton 回路问题可转换为 D' 的 $(n-2)$ 级接口矩阵问题。

推论 4 若 $D=\langle V,E\rangle$ 是含有 n 个顶点的简单有向图，则存在一个以 $IP_1(D)$（值等同于邻接矩阵）起始，总数不超过 m 的等级矩阵序列：$IP_1(D), IP_2(D), \cdots, IP_{n-1}(D)$，且满足：任何一个 $k(k>1)$ 级接口矩阵均可以由前面的两个等级接口矩阵相连接得到。其中 $m=2\times\lceil\log_2^n\rceil$，$\lceil\log_2^n\rceil$ 表示不小于 $\log_2^n$ 的最小整数。

证明：设十进制数 $(n-1)$ 的二进制表示为：$P_{k-1}P_{k-2}\cdots P_1P_0$，由二、十进制进制转换规则可知：$k=\lceil\log_2^n\rceil$，现在构造一个数列 $1=x_0, x_1, x_2, \cdots x_{k-1}, \cdots x_{2k-2}, x_{2k-1}=n-1$，其中

$$x_i=\begin{cases}2^i, & \text{if } i\leqslant k-1\\ x_{i-1}, & \text{if } i\geqslant k \text{ and } P_{i-k}=0\\ x_{i-1}+2^{i-k}, & \text{if } i\geqslant k \text{ and } P_{i-k}=1\end{cases}$$

把上述序列中值相同的数字仅保留一个，得到新序列

$1=y_0, y_1, y_2, \cdots, y_p=n-1, p\leqslant 2k-1$，由 x_i 的定义可知：

$$y_i=\begin{cases}y_{i-1}+y_{i-1}, & \text{if } i\leqslant k-1\\ y_{i-1}+y_\sigma, & \text{if } i\geqslant k; y_\sigma\in\{y_m\mid y_0\leqslant y_m\leqslant y_{k-1}\}\end{cases}$$

由定理 2 知：$s+t=m\rightarrow IP_s(D)\oplus IP_t(D)=IP_m(D)$，所以，序列 $y_0, y_1, y_2, \cdots, y_p$ 对应的等级接口矩阵 $IP_{y_0}(D)$，$IP_{y_1}(D), \cdots, IP_{y_p}(D)$ 即为所求等级矩阵序列。

定理 2 及其推论给出了一个把 n 顶点有向图 Hamilton 回路计算转换为 $n-1$ 顶点有向图的 $n-2$ 级接口矩阵计算的算法。然而，我们不能证明这种转换是最优的，一般而言，这和 n 的具体数值及初始关联矩阵值相关。进一步地，推论 2 中给出的 $IP_1(D), IP_2(D), \cdots, IP_{n-1}(D)$ 序列也不是最优列，下面引入一个与此问题相关的猜想——肖尔茨-布劳尔猜想[11]。

已知正整数 x，求 x 的 n 次方 x^n 的算法中，两个数相乘一次，记一步。由 x 计算 x^n 的各种算法中，最少的步数记作 $l(n)$。肖尔茨—布劳尔猜想是：对所有正整数，有 $l(2^n-1)\leqslant n-1+l(n)$。

由 x 求 x^n 的算法中可以看出，$l(n)$ 对应的路线恰好是一个最优的等级矩阵序列。河南张英杰先生曾经在奇迹网上给出一个肖尔茨猜想的证明框架[1]，然而在几个关键步骤上缺乏严格证明。下文根据张英杰先生的思路给出一个完整的证明。

证明：任取正整数 n，满足 $n\geqslant 2$ 且有 $l(n)=k$，令 $l(n)$ 经过的最佳路径点为：$a_0=1, a_1=2, a_2, a_3, \cdots, a_k=n$；相应地，其对应的 k 个递增数值为：$b_0=a_1-a_0, b_1=a_2-a_1, b_2\cdots, b_{k-1}=a_k-a_{k-1}$；根据定义，对于 b_i，有 $b_i\in\{a_0, a_1, a_2, a_3, \cdots, a_i\}$，现在构造如下所示 $k+1$ 个数：

$1=p_0=2^{a_0}-1$

$p_1=2^{a_1}-1$

$p_2=2^{a_2}-1$

$p_3, \cdots,$

$p_k=2^{a_k}-1=2^n-1$

先讨论已知 $p_0, p_1, p_3, \sim p_i$ 生成 p_{i+1} 的过程：由于

$$\begin{aligned}p_i\times(2^{b_i})+(2^{b_i}-1)&=p_i\times(2^{a_{i+1}-a_i})+(2^{a_{i+1}-a_i}-1)\\&=(2^{a_i}-1)\times(2^{a_{i+1}-a_i})+(2^{a_{i+1}-a_i}-1)\\&=2^{a_{i+1}}-1\\&=p_{i+1}\end{aligned}\qquad(*)$$

再令

$x_0=x^{p_0}=x$

$x_1=x^{p_1}$

$x_2=x^{p_2}$

$x_3, \cdots,$

$x_k=x^{p_k}=x^{(2^n-1)}$

由式(*)可以看出：$x_i=x^{p_i}$，经过 b_i 次平方运算，再与 $\sigma=x^{(2^{b_i}-1)}$ 相乘，可以得到 $x_{i+1}=2^{p(i+1)}$，因为 $b_i\in\{a_0, a_1, a_2, a_3\cdots a_i\}$，所以 $(2^{b_i}-1)\in\{p_0, p_1, p_2, \cdots p_i\}$，即 $\sigma\in\{x_0, x_1, x_2, \cdots x_i\}$ 所以从 $p_0, p_1, p_3, \sim p_i$ 生成 p_{i+1} 是一条合法的路径。

所以，上述构造的算法从 x_0 到 x_k 经过的总计算步数 θ 为：$b_0+b_1+b_2+\cdots+b_{k-1}$ 次平方运算＋K 次与 σ 的乘法运算。即

$$\begin{aligned}\theta&=(a_1-a_0)+(a_2-a_1)+\cdots+(a_k-a_{k-1})+k\\&=a_k-a_0+k=n-1+k=n-1+l(n)\end{aligned}$$

所以 $l(2^n-1)\leqslant n-1+l(n)$ 成立。

3 复杂性分析

因为 n 个顶点图的 Hamilton 回路问题可转换为 $n-1$ 个顶点图 $(n-2)$ 级接口矩阵问题，为了方便描述，下面直接讨论 n 个顶点的 $(n-1)$ 级接口矩阵的复杂性问题。

3.1 时间复杂性

由于所有的等级接口矩阵均为 $n\times n$ 矩阵，计算 $(n-1)$ 级等级接口矩阵的时间复杂性正比于矩阵的连接次数。虽然肖尔茨猜想给出了某些顶点数的一个较好的分解方法（比如顶点数为 (2^n-1) 的图），以 31 个点为例：

定理 2 的推论所给出的路径分解过程是：1—>2—>4—>8—>16—>17—>19—>23—>31。

前述肖尔茨猜想证明过程中给出的路径解是：1—>2—>3—>6—>12—>15—>30—>31。

[1] 张英杰先生给出的证明网址为：http://www.qiji.cn/eprint/abs/936.html

显然，后一个分解的连接次数小于前一个分解，然而，对于一般的任意整数 n，关于 $l(n)$ 的计算仍然没有进一步的进展，因而，此处只能以定理 2 的推论作为上限进行分析，由前述运算 6 的分析可知，矩阵乘法的运算复杂性为 $O(n^3)$，所以 $T(n)=2\times\lceil \log_2^n \rceil * n^3<n^4=O(n^4)$ 。

3.2 空间复杂性

本算法空间复杂性较大，在不做任何优化的情况下，每个矩阵元素需要 2^n bit，采用定理 2 的推论作为上限分析，故空间复杂性为：

$$S(n)=m\times(n^2/8)\times 2^n<n^3\times 2^{n-3}=O(n^3 2^{n-3})$$

其中，$m=\lceil \log_2^n \rceil$。

4 应用举例

本文采用 CTSP 做实验说明，实验条件如下。

CPU：AMD4200＋（单核 2100）；内存：2G；硬盘：320G；OS：Windows XP，SP3；编程工具：Visual C＋＋6；算法：单线程串行算法。在查阅 CTSP 相关文献的过程中，笔者发现两个版本的城市间距离表数据，一是文献[10]所附的距离表（数据比较旧），二是文献[9]使用的经纬度表（数据较新，本文数据及距离计算公式均来自文献[9]），经纬度数据也可参考 http://astro.sohu.com/20060517/n226130617.shtml。本文对两个版本数据分别进行了实验，结果参见表 1，下文重点讨论两个核心问题。

4.1 路径编码

前文给出的编码仅适应于理论分析，实际用时空间复杂性过大，考虑到等级接口矩阵中仅需区分具有相同中间点数的路径，故可进一步优化，结合哈密顿回路与旅行商问题的差异，本文处理方式如下：

在每一个等级矩阵中，按照从 $\{1,2,\cdots,n\}$ 中取出 r 个数的所有组合的字典数序编码（需保证每个组合都是升序序列），令 $f(i_1,i_2,\cdots,i_r,n)$：从 $\{1,2,\cdots,n\}$ 中的组合 $i_1,i_2,\cdots,i_r$ 的字典序编码（基数为 1），有递推公式：

$$f(i_1,i_2\cdots i_r,n)=\begin{cases} f(i_2-1,i_3-1,\cdots i_r-1,n-1), & \text{if } i_1=1 \\ C_{n-1}^{r-1}+C_{n-2}^{r-1}+\cdots+C_{n-i_1+1}^{r-1}+ & \\ f((i_2-i_1),(i_3-i_1)\cdots(i_r-i_1),n-i_1), & \text{if } i_1>1 \end{cases}$$

令 $w(r,n,i)=C_n^r-C_{n-i+1}^r$

多次代入递推公式可得：

$$f(i_1,i_2,\cdots,i_r,n)=C_n^r-\sum_{k=1}^{r}C_{n-i_k}^{r+1-k}$$

规定 $C_m^n=0$，if$(m<n)$；$C_0^0=0$。

程序中存储 C_m^n 矩阵序列，这样 k 级矩阵的元素编码问题就转换为$(k+1)$次加减法运算。

4.2 数据存储

据前文讨论，对于 31 个城市的 CTSP，可转换为 29 级接口矩阵问题，本文采用分段动态数据替代静态数据，即把 29 级接口矩阵看作如下动态连接：

$$IP_{29}(D)=IP_4(D)\oplus IP_4(D)\oplus IP_4(D)\oplus IP_4(D)\oplus IP_4(D)\oplus IP_9(D)$$

文章使用 CTSP31，CTSP30，CTSP26，CTSP22，CTSP18，CTSP14，CTSP10 七个类完成相关的运算，对应关系如下：

$$\begin{cases} \text{CTSP31：负责 } IP_{29}(D)\text{到哈密顿回路的转换} \\ \text{CTSP30：}IP_{29}(D) \\ \text{CTSP26：}IP_4(D)\oplus IP_4(D)\oplus IP_4(D)\oplus IP_4(D)\oplus IP_9(D) \\ \text{CTSP22：}IP_4(D)\oplus IP_4(D)\oplus IP_4(D)\oplus IP_9(D) \\ \text{CTSP18：}IP_4(D)\oplus IP_4(D)\oplus IP_9(D) \\ \text{CTSP14：}IP_4(D)\oplus IP_9(D)\text{；} \\ \text{CTSP10：}IP_9(D) \end{cases}$$

显然，将全部矩阵数据放在内存中是不现实的，为此，将小矩阵数据（$IP_1(D)$，$IP_2(D)$，$IP_4(D)$）直接存于内存数组中，大矩阵数据（$IP_9(D)$）采用 windows 内存映射技术存储，使用内存映射的相关步骤如下：

第一，创建内存映射文件对象，并设置其文件句柄为 INVALID_HANDLE_VALUE，函数原型 HANDLE CreateFileMapping（HANDLE hFile，LPSECURITY_ATTRIBUTES lpAttributes，DWORD flProtect，DWORD dwMaximumSizeHigh，DWORD dwMaximumSizeLow，LPCTSTR lpName）；

第二，从文件中取一内存视图，函数原型 LPVOID MapViewOfFile（HANDLEhFileMappingObject，DWORD dwDesiredAccess，DWORDdwFileOffsetHigh，DWORD dwFileOffsetLow，SIZE_T dwNumberOfBytesToMap）；

第三，对得到的视图指针进行数据读写操作；

第四，释放第三步得到的视图指针，函数原型 BOOL UnmapViewOfFile（LPCVOID lpBaseAddress）。

通过内存映射技术，在第三步得到的指针和普通指针在操作上是相同的，不需做特别处理。

另外，为了进一步优化时间，在 $IP_9(D)$ 的存储计算中，我们使用贪心法等技术进行路径估算，裁剪了部分矩阵元素的计算，即有选择地填充 $IP_9(D)$ 元素，仅在该元素实际被使用的时候才真正计算。因此，此处不对时间复杂性做进一步的分析，而对于空间复杂性，也仅考虑 $IP_9(D)$ 的存储问题，其至少需要硬盘空间：$4*C_{30}^8$ Byte（内存映射需要硬盘空间，9 级接口矩阵需要 8 个中间点，每条路径值需要 4 个字节）。表 1 给出了相关的实验结果。

表 1 CTSP 运行结果

	CTSP30 调用次数	CTSP26 调用次数	CTSP22 调用次数	CTSP18 调用次数	CTSP14 调用次数	CTSP10 调用次数	最优解 (km)	整体搜索时间(s)
理论值	435	8906625	94641797250	4585395 07676250	$\approx 8.34*10^{17}$	$\approx 4.13*10^{20}$		
距离表程序分析	146	27767	415050	1266476	1232892	79648	15404	22467.281
经纬度表程序分析	140	24177	355980	1352329	2171643	225784	15290	27834.765

距离表版本对应的最优路径与文献[4,5]等的猜测是一致的。经纬度表版本对应的最优路径是：乌鲁木齐-西宁-兰州-银川-西安-郑州-太原-呼和浩特-石家庄-北京-天津-哈尔滨-长春-沈阳-济南-合肥-南京-上海-杭州-台北-福州-南昌-武汉-长沙-广州-海口-南宁-贵阳-昆明-成都-拉萨-乌鲁木齐，这个路径结果目前没有近似算法得到，也未见相关文献发表，比文献[9]猜测的最好结果（实为 15355）少 65km。

结束语 动态规划法是研究优化问题的一种常用算法，

满足最优性原理是使用动态规划法的前提条件，因此，许多文献用最优性原理的证明过程来分析最优解的求解过程，本质上说，文献[19]是这样思考 Hamilton 回路的，文献[20]也是这样讨论 TSP 的。本文的研究结果表明，最优性原理本身并没有提供最优的解结构，具体而言，对于 TSP 问题，虽然一条长度为 n 的最短路径一定是一长度为 1 的路径和一长度为 $n-1$ 的最短路径组合，然而，这种组合未必是求解长度为 n 的路径的最优过程，结合肖尔茨猜想的证明，本文探讨了最优解组成的复杂性，为进一步研究提供了理论基础。从工程应用而言，本文提供的算法具有如下特点：

1. 计算不依赖于初始解与经验值。

2. 没有欧氏空间的平面约束条件，有利于推广到其他相关非平面的应用领域。

3. 作为一种精确算法，其解可以作为近似算法的评价依据。

4. 在降低时间复杂度的同时，为以后优化留下了很大的研究空间——数据压缩与矩阵并行运算的相关研究将是该算法有潜力的发展方向。

参 考 文 献

[1] 文中华，姜云飞. 用分层关联方法求有向图中所有 Hamilton 回路的算法[J]. 计算机研究与发展，2005，42(10)：1809-1814

[2] 王彦祺. 逐点循环递归法求哈密顿回路[J]. 哈尔滨工业大学学报，2004，36(1)

[3] 邹鹏，周智，等. 求解旅行商问题的循环局部搜索算法的运行时间和性能分布分析[J]. 计算机学报，2006，29(1)

[4] 刘新，刘任任，侯经. 求解旅行商问题的整体优先算法[J]. 计算机应用，2007，27(5)

[5] 贺一，刘光远. 禁忌搜索算法求解旅行商问题研究[J]. 西南师范大学学报：自然科学版，2002，27(3)

[6] 耿素云，屈婉玲. 离散数学(修订版)[M]. 北京：高等教育出版社，2004

[7] 高尚，韩斌，吴小俊，等. 求解旅行商问题的混合粒子群优化算法[J]. 控制与决策，2004，19(11)

[8] 柴世红. 两段式遗传算法求解 CTSP[J]. 大众科技，2008(4)

[9] 顾大权，侯太平，左莉，等. 用数据搅动算法求解 TSP 问题[J]. 计算机应用，2004(24)

[10] 杨忠，鲍明，张阿舟. 求解中国旅行商问题的新结果[J]. 数据采集与处理，1993，8(3)

[11] 胡作玄. 本世纪未解的数学难题(下)[J]. 百科知识，1999(5)

[12] 靳蕃，范俊波，谭永东. 神经网络与神经计算机[M]. 成都：西南交通大学出版社，1991

[13] 杨启宇. 关于无穷大基数的矛盾[J]. 成都大学学报：自然科学版，2004，23(4)

[14] 温邦彦. 自然数和偶数的个数一样多吗？[J]. 重庆工学院学报：自然科学版，2008，22(11)

[15] 欧阳耿. 无定义的实无穷概念是第三次数学危机的真正根源[J]. 宜春学院学报，2005，27(4)

[16] 杜国平. 潜无穷、实无穷探析[J]. 自然辩证法通讯，2009，31(3)

[17] 朱梧槚. 数学与无穷观的逻辑基础[M]. 大连：大连理工大学出版社，2008

[18] 柯华庆. 直觉主义数学哲学的两个阶段[J]. 学术研究，2005(2)

[19] 文中华，陈志红. 一个求简单图中所有 Hamilton 回路的算法[J]. 湘潭大学：自然科学学报，2005，27(4)

[20] 徐伯庆，宣国荣，柴佩琪. 中国旅行商问题的二叉树描述及其求解[J]. 模式识别与人工智能，2000，13(2)

[21] 王红梅. 算法设计与分析[M]. 北京：清华大学出版社，2006

[22] 郭文强，秦志光，冯昊. 一种基于归零矩阵的 TSP 求解算法[J]. 计算机科学，2009，36(8)：254-257

(上接第 213 页)

点数。然而，当 $5<z_{out}\leqslant 8$ 时，暗示着这 4 个社团的顶点越来越难以检测，基于模块 Q 谱分算法的曲线急速下滑，意味着成功检测已知社团顶点的偏离较大。特别地，当通过谱分方法或半指导方法初始化我们的核 k-means 算法时，这个核方法仍然能正确分类出 82%以上的顶点，比基于 Q 谱分算法准确性高。换句话说，随着 z_{out} 的变化，我们的基于核 k-means 算法比基于 Q 谱分算法更能正确检测出这 4 个已知的社团结构。

结束语 本文根据最近提出的用来评估与发现社团的模块密度 D 函数与核 k-means 之间的等价性，提出了一种新的社团检测核 k-means 算法。从计算机产生的人工网络的实验结果中可以看出，我们提出的核 k-means 方法比基于 Q 算法有较好的准确性。

由于核 k-means 方法对聚类开始的划分比较灵活，因此可以通过谱分方法、半指导或指导方法来初始化核 k-means 方法，以尽可能地提升社团检查结果。

参 考 文 献

[1] Fortunato S, Latora V, Marchiori M. Method to find community structure based on information centrality[J]. Phy Rev E, 2004, 70(13): 056105

[2] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proc Natl Acad Sci USA, 2002, 99(12): 7821

[3] Palla G, Derényi I, Farkas I, et al. Uncovering the overlapping community structure of complex network in nature and society [J]. Nature, 2005, 435: 814-818

[4] Newman M E J. The structure and function of complex net - works[J]. Proc Natl Acad Sci USA, 2006, 103: 8577-82

[5] Kernighan B W, Un S. An efficient heuristic procedure for partitioning graph[J]. Bell Syst Tech J, 1970, 49: 292-307

[6] Pothen A, Simon H D, Uou K P. Partitioning sparse matrices with eigenvectors of graphs[J]. Siam J Matrix Anal Appl, 1990, 11(3): 430-452

[7] Fiedler M. Albegraic connectivity of graphs[J]. Czech math J, 1973, 23: 298-305

[8] Newman M E J. Finding community structure in networks using the eigenvectors of matrices[J]. Phys Rev E, 2006, 74(3): 036104

[9] Newman M E J. Modularity and community structure in networks[J]. Proc Natl Acad Sci USA, 2006, 103: 8577-8582

[10] Fortunato S, Barthélemy M. Resolution limit in community detection[J]. Proc Natl Acad Sci USA, 2007, 104(1): 36-41

[11] Li Z P, Zhang S H, Wang R S, et al. Quantitative function for community detection[J]. Phys Rev E, 2008, 77(3): 036109

[12] Dhillon I S, Guan Y, Kulis B. Kernel K-means, Spectral Clusterng and Normalized Cuts[C]//Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, 2004: 551-556

[13] Dhillon I S, Guan Y, Kulis B. A unified view of kernel K-means, spectral clustering and graph cuts[R]. University of Texas at Austin, 2004