关于最小测试集的线性规划松弛近似

崔 鹏[1] 刘红静[2]

(中国人民大学信息资源管理学院 北京 100871)[1] (保定市财贸学校 保定 071000)[2]

摘 要 目前最小测试集的最佳近似比是贪心算法的 $2\ln n+o(1)$。这个近似比能否改进是一个公开的问题。本文讨论了最小测试集的基于线性规划松弛的近似比证明方法的能力问题。我们证明最小测试集的整性间隙至少为 $0.72\ln n$,而且最小测试集整性间隙的系数可以与最小集合覆盖的整性间隙的系数一样大。另外,我们说明加权最小测试集的贪心算法的近似比不能通过对偶拟合方法改进超过一个常数。

关键词 最小测试集,贪心算法,整性间隙,对偶拟合

On LP-relaxation Approximation of Minimum Test Set

CUI Peng[1] LIU Hong-Jing[2]

(Computer Science and Technology Departmant, Peking Univercity, Beijing 100871)[1]

(School of Finance and Trade of Baoding, Baoding 071000)[2]

Abstract The up-to-date best approximation ratio of minimum test set is $2\ln n+o(1)$of the greedy algorithm. It is an open problem if this approximation ratio can be improved. This paper discusses the ability of the methods for proof of the approximation ratios based on the LP-relaxation. We prove the integrality gap of minimum test set is at least $0.72\ln n$ and the coefficient of the integrality gap of minimum test set can be as large as that of minimum set cover. Moreover, we show one cannot improve the approximation ratio of the greedy algorithm for weighted minimum test set by more than a constant.

Keywords Minimum test set, Greedy algorithm, Integrality gap, Dual fitting

1 引言

最小测试集(Minimum Test Set,简称 MTS)经常出现于故障诊断、软件测试、模式识别和生物信息学等领域。问题的输入是一个条目的集合,$S=\{a_1,\cdots,a_n\}$,和 S 的子集(称为 S 的测试)的一个集合,$\mathcal{T}=\{T_1,\cdots,T_m\}$。一个测试 T 区分条目对$\{i,j\}$,如果 i 和 j 恰好有一个属于 T,即$|T\cap\{i,j\}|=1$。测试的一个集合 $\mathcal{T}'\subseteq\mathcal{T}$是一个 S 的测试集,如果每一个条目对被 $\mathcal{T}'$中至少一个测试所区分。问题的度量是测试集的大小,目标是找最小的测试集。

MTS 是 NP 难的证明见文[1]。文[2]是一篇 MTS 的综述,介绍了 MTS 的贪心算法和分支限界算法。贪心算法是 MTS 的近似算法,可以按照分割标准或信息标准实现。按照信息标准实现的贪心算法至今只能证明平凡的近似比 $O(n)$,因此本文只关心按照分割标准实现的贪心算法。

MTS 可以自然地归约为最小集合覆盖(Minimum Set Cover,简称 MSC):底集是以条目集为顶点集的完全图的边集,每个测试 T 转换为 T 与 $S-T$ 之间的割集的边集。众所周知,MSC 的贪心算法有紧密的近似比 H_N[3],其中 N 是底集的大小,H_N 是第 N 个调和数。利用 MTS 到 MSC 的归约,可以得到 MTS 的贪心算法的近似比 $H_{n(n-1)/2}=2\ln n+O(1)$[4]。关于 MTS 的近似难度,文[5,6]的结果是:除非 $P=NP$,不能近似到 $o(\ln n)$;除非 $NP\subseteq DTIME(n^{\log\log n})$,对任何 $\varepsilon>0$ 不能近似到$(1-\varepsilon)\ln n$。

我们知道,MTS 按照可近似性属于类 $\log-APX$,即具有近似比 $O(\log_2|I|)$的问题的类。在目前最佳近似比与近似难度的结果中间,有一个因子 2 的间隙。文[7]猜测贪心算法的近似比可以是 $\ln n+o(1)$。用来证明 MTS 的对数近似比的两个方法:对偶拟合和随机取整,都是基于线性规划松弛的方法。基于这个原因,本文讨论了 MTS 的基于线性规划松弛的近似比证明方法的能力问题。我们指出 MTS 的整性间隙至少为 $0.72\ln n$,从而说明对偶拟合和随机取整得到的近似比无条件地不能优于这个整性间隙。我们还证明 MTS 的整性间隙的系数可以与 MSC 的整性间隙的系数一样大,说明 MTS 的线性规划松弛近似的能力限制可以跟 MSC 的能力限制一样大。另外,我们通过例子说明加权 MTS 的贪心算法不能通过对偶拟合方法改进超过一个常数。

第 2 节是背景介绍,第 3 节给出整性间隙的结果,第 4 节给出加权 MTS 的对偶拟合的能力的结果,最后是结束语。

2 背景

给定最小化问题 Ⅱ 的一个近似算法 A 和实例 I,用 $m_A(I)$表示算法 A 在 I 上返回解的度量,用 $m^*(I)$表示 I 的最优解的度量。称 $m_A(I)/m^*(I)$是 A 在 I 上的近似比。用$|I|$表示实例 I 的大小,对 MTS 和 MSC,$|I|$分别规定为 n 和 N。对一个定义在 Z^+上的非降函数 $r(|I|)$,我们称 $r(|I|)$是 A 的近似比,如果存在 $n_0\in Z^+$,对任何满足$|I|>n_0$ 的实例 I,都有 $m_A(I)/m^*(I)\leqslant r(|I|)$。如果存在常数 c,对任意大的 n_1,都存在满足$|I|>n_1$ 的实例 I,使得 $m_A(I)/m^*(I)\geqslant r(|I|)+c$,则称近似比 $r(|I|)$是紧密的。

MSC 是一个受到大量关注的难解问题。MSC 的实例由底集 U,U 的子集的一个集合 C 组成。$|U|=N$。U 的一个集

崔 鹏 博士,研究方向为政务信息管理。刘红静 讲师,研究方向为信息管理。

合覆盖是 C 的一个子集 C'，使得 U 的每个元素至少属于 C' 的一个成员（或称为 U 的这个元素被 C' 的这个成员覆盖）。问题的度量是集合覆盖的大小，目标是寻找最小的集合覆盖。

MSC 可以表示为一个整数规划问题。对每个 $c\in C$ 分配一个取 0/1 值的变量 x_c。x_c 设为 1 当且仅当在集合覆盖中取 c。显然，问题的约束是对每个 $e\in U$，我们要求至少有一个包含 e 的集合被选择。

$$\text{minimize } \sum_{c\in C} x_c$$
$$\text{subject to } \sum_{e\in c} x_c \geqslant 1, e\in U$$
$$x_c\in\{0,1\}, c\in C$$

如果我们让 x_c 的取值范围扩大到 $1\geqslant x_c\geqslant 0$，就得到这个整数规划的线性规划松弛。因为 x_c 的上界是多余的，我们得到下面的线性规划。这个线性规划的解叫做一个分数集合覆盖。

$$\text{minimize } \sum_{c\in C} x_c$$
$$\text{subject to } \sum_{e\in c} x_c \geqslant 1, e\in U$$
$$x_c\geqslant 0, c\in C$$

对每个 $e\in U$ 引进一个变量 y_e，得到下面的对偶规划。

$$\text{minimize } \sum_{e\in U} y_e$$
$$\text{subject to } \sum_{e\in c} y_e \leqslant 1, c\in C$$
$$y_e\geqslant 0, e\in U$$

MSC 的常用算法，贪心算法，可叙述为：每次选择一个子集，使得这个子集覆盖的底集的元素个数最大，直到底集所有的元素被覆盖。众所周知，MSC 的贪心算法可以用对偶拟合的方法证明近似比 H_N[8]。做法是贪心算法每次选择子集覆盖的元素个数的倒数（称为这个子集的价格）收缩一个因子，填到这个集合覆盖的元素上；然后证明贪心算法结束后，所有元素上填的值组成线性规划松弛的对偶规划的一个可行解，从而证明这个因子为近似比。

设 (S,T) 是 MTS 的一个实例，我们可以把它自然地归约为一个 MSC 的实例 (U,C)，其中 $U=\{\{a,b\}\mid a,b\in S, a\neq b\}$，$C=\{c(T)\mid T\in T\}$，$c(T)=\{\{a,b\}\mid a\in T, b\in S-T\}$。通过这个归约，我们立即可以得到 MTS 实例的整数规划，相应的线性规划松弛及其对偶规划。

MTS 的贪心算法可以叙述为：每次选择一个测试，使得这个测试区分的条目对数最大，直到所有的条目对被区分。显然，(S,T) 与 (U,C) 的最优解的度量相等；MTS 的贪心算法在上返回解的度量，恰好与 MSC 的贪心算法在 (U,C) 上返回解的度量相等。因为 MSC 的贪心算法有近似比 H_N，加权 MTS 的贪心算法有近似比 $H_{(n(n-1)/2}=2\ln n+O(1)$。

3 整性间隙

当我们为一个最小化问题设计近似算法时，通常为这个问题的最优解寻找一个下界，在近似算法返回解的度量与最优解的下界之间建立联系，从而得到近似比。在许多情况下，线性规划提供了这样一个方案。

对 MSC，设 $OPT(I)$ 是实例 I 的最优解的度量，$OPT_f(I)$ 是实例 I 的线性规划松弛的最优解的目标函数值。定义实例 I 的整性间隙为

$$\frac{OPT(I)}{OPT_f(I)}$$

MSC 的整性间隙是

$$IP(N)=\sup_{|I|=N}\frac{OPT(I)}{OPT_f(I)}$$

其中 $\sup\limits_{|I|=N}$ 表示在所有大小为 N 的 I 上取上确界。

MSC 的对偶拟合方法和随机取整方法[9]都采用线性规划松弛作为最优解的下界。一方面，我们证明的近似比是整性间隙的上界，因此 MSC 的整性间隙不超过 $\ln N+o(1)$；另一方面，构造尽量大的整性间隙，可以推知我们能够证明近似比的极限。目前所知 MSC 的最大整性间隙来自文[10]给出的一个例子，它的整性间隙大于 $\log_2 N/2=0.72\cdots\ln N$，说明以上两种方法不能证明小于 $0.72\ln N$ 的近似比。

对 MTS，设 $OPT(I)$ 是实例 I 的最优解的度量，$OPT_f(I)$ 是实例 I 的线性规划松弛最优解的目标函数值。定义实例 I 的整性间隙为

$$\frac{OPT(I)}{OPT_f(I)}$$

MTS 的整性间隙是

其中 $\sup\limits_{|I|=n}$ 表示在所有大小为 n 的 I 上取上确界。

对 MTS 的一个实例 (S,T) 和一个测试集 $T'\subseteq T$，我们将证明 S 中至多有一个条目不属于 T' 中的任何测试。如果存在这样的条目，称为这个实例的孤立条目。

引理 1 S 中至多有一个条目不属于 T' 中的任何测试。

我们可以修改文[10]中的例子，即填加一个元素，得到一个 MTS 的实例。设 $n=2^k$，其中 k 是一个正整数，$S=\{a_0, a_1,\cdots,a_{n-1}$。对 $0\leqslant i\leqslant n-1$，把 i 写成二进制的形式，然后视为 $GF[2]$ 上的一个 k 维向量，用 i 表示。对 $1\leqslant i\leqslant n-1$，定义测试 $T_i=\{a_j\mid i\cdot j=1\}$，其中 $i\cdot j$ 表示两个向量的内积。令 $T=\{T_1,L,T_{n-1}\}$。

引理 2 每个测试的大小是 $n/2$。

引理 3 每个条目对恰好被 $n/2$ 个测试区分。

由引理 3，$x_i=2/n, 1\leqslant i\leqslant n-1$ 是这个例子的线性规划松弛的一个解，其目标函数值是 $2(n-1)/n$。因为 a_0 是孤立条目，根据引理 1，任何测试集一定覆盖所有 $a_i, 1\leqslant i\leqslant n-1$。根据[10]，我们知道这要求这个测试集至少有 k 个测试。所以，这个例子的整性间隙至少为

$$\frac{2(n-1)}{n}\cdot\log_2 n=0.72\mathrm{L}\ \ln n-o(1)。$$

这就得到了下面的命题。

命题 1 对充分大的 n，MTS 的整性间隙至少为 $0.72\ln n$。

给定一个 MSC 的实例 $I=(U,C)$，$U=\{e_1,\cdots,e_N\}$，$C=\{c_1,\cdots,c_m\}$，和正整数 b,p。设 $B=2^b$，正整数 K 满足 $B^{K-1}<N\leqslant B^K$。构造 MTS 的实例 $I'=(S,T)$ 如下。

$$S=\bigcup_{t=1}^{l}(\{e_1^t,\cdots,e_N^t\}\cup\{f_1^t,f_N^t\}\cup\{g_{N+1}^t,\cdots,g_{B_k}^t\}),$$
$l=2^p$，$|S|=n=l(N+B^K)$。

$$T=\bigcup_{t=1}^{l}\{T_1^t,\cdots,T_m^t\}\cup\bigcup_{i=1}^{lB^K-1}\{T_i\},$$

其中，对任意 $1\leqslant s\leqslant m, 1\leqslant t\leqslant l$，$T_s^t=\{e_r^t\mid e_r\in c_s\}$。对 $0\leqslant i\leqslant lB^K-1$，把 i 写成二进制的形式，然后视为 $GF[2]$ 上的一个 $bK+p$ 维向量，用 i 表示。对 $1\leqslant i\leqslant lB^K-1$，定义测试

$$T_i=\{e_r^t,f_r^t\mid i\cdot j=1, j=(t-1)B^K+r, r \bmod B^K<N\}。$$
$$\cup\{g_r^t\mid i\cdot j=1, j=(t-1)B^K+r, r \bmod B^K\geqslant N\}$$

对任意 $1\leqslant r\leqslant N, 1\leqslant s\leqslant m, 1\leqslant t\leqslant l$，条目对 $\{e_r^t,f_r^t\}$ 被某个测试 T_s^t 所区分当且仅当 $e_r\in c_s$，而且不被任何 $T_i, 1\leqslant i\leqslant lB^K-1$ 所区分。由 T_i 的定义和引理 3，任意其他的条目对恰好被 $lB^K/2$ 个测试 T_i 所区分。

设 $x_{c_1},x_{c_2},\cdots,x_{c_m}$ 是 I 的线性规划松弛的一个最优解。对任意 $1\leqslant s\leqslant m, 1\leqslant t\leqslant l$，令 $x_{T_s^t}=x_{c_s}$。对任意 $1\leqslant i\leqslant lB^K-1$，令 $x_{T_i}=2/(lB^K)$。则

$x_{T_1^1}, x_{T_2^1}, \cdots, x_{T_m^1}, x_{T_1^2}, x_{T_2^2}, \cdots, x_{T_m^2}, \cdots,$

$x_{T_1^l}, x_{T_2^l}, \cdots, x_{T_m^l}, x_{T_1}, x_{T_2}, \cdots, x_{T_{lB^K-1}}$

是 I' 的线性规划松弛的一个解，其目标函数值是 $lOPT_f(I)+2(lB^K-1)/(lB^K)$，所以 $OPT_f(I')\leqslant lOPT_f(I)+2(lB^K-1)/(lB^K)$。

另一方面，设 $T'\subseteq T$ 是 S 的一个测试集，则由 T_i 的定义和命题 1 中的测试集至少有 k 个测试这个事实，

$$T'\cap\bigcup_{i=1}^{lB^K-1}\{T_i\}|\geqslant log_2(lB^K)。$$

对 $1\leqslant t\leqslant l$，设 $T'\cap\{T_1^t,\cdots,T_m^t\}$ 中测试的下标集合是 χ_t，则 $\{c_s|s\in\chi_t\}$ 一定是 U 的一个集合覆盖，从而 $T'\cap\{T_1^t,\cdots,T_m^t\}|=|\chi_t|\geqslant OPT(I)$。合起来有 $OPT(I')\geqslant|T'|\geqslant lOPT(I)+\log_2(lB^K)$。

所以，实例 I' 的整性间隙。

$$\frac{OPT(I')}{OPT_f(I')}\geqslant\frac{OPT(I)+\log_2(lB^K)/l}{OPT_f(I)+2(lB^K-1)/(l^2B^K)}。$$

引理 4 对 MSC 的实例 I 和满足

$B=2^b, B^{K-1}<N\leqslant B^K$ 的正整数 b 和 K，

$$\frac{OPT(I)+(b+1)K/2^K}{OPT_f(I)+2/2^K}=\frac{OPT(I)}{OPT_f(I)}-O(bKB/2^K)。$$

命题 2 如果对任意 N_0，存在 MSC 的实例 I，满足 $|I|>N_0$，其整性间隙

$$\frac{OPT(I)}{OPT_f(I)}\geqslant a\ln N,$$

其中 a 是一个满足 $0<a\leqslant 1$ 的与 $|I|$ 无关的常数，则对任意 $\varepsilon>0$ 和任意 n_0，存在 MTS 的实例 I'，满足 $|I'|>n_0$，其整性间隙

$$\frac{OPT(I')}{OPT_f(I')}>(a-\varepsilon)\ln n。$$

证明：在构造 I' 时，令 $l=2^K$，则由 $B^{K-1}<N\leqslant B^K$ 和 $n=2^K(N+B^K)$ 得

$$a\ln N\geqslant a\ln n/(1+\frac{1}{K-1})(1+\frac{1}{b}))-O(1)。$$

如果 $\frac{OPT(I)}{OPT_f(I)}\geqslant a\ln N$，则由引理 4，

$\frac{OPT(I')}{OPT_f(I')}\geqslant a\ln n/((1+\frac{1}{K-1})(1+\frac{1}{b}))-R$，其中 $R=O(1)+O(bK/2^K)$。

显然，可以按照足够大的 b 和 N 构造 I'，使得实例 I' 满足 $|I'|>n_0$，其整性间隙

$$\frac{OPT(I')}{OPT_f(I')}>(a-\varepsilon)\ln n。$$ □

4 对偶拟合对加权 MTS 的能力

加权 MSC 是 MSC 的推广：每个 $c\in C$ 有一个权 $w(c)\in Q^+$。问题的度量是集合覆盖的成员的权的总和。这个问题的整数规划是

$$\text{minimize } \sum_{c\in C}w(c)x_c$$
$$\text{subject to } \sum_{c,e\in C}x_c\geqslant 1, e\in U$$
$$x_c\geqslant o, x_c\in Z, c\in C$$

它的贪心算法是：每一步在没有选择子集中选择价格最小的子集，一个子集的价格定义为这个子集的权与当前它覆盖的元素个数之比。实际上，文[8]用对偶拟合方法证明了加权 MSC 的贪心算法有近似比 H_N。

类似地，加权 MTS 是 MTS 的推广。每个 $T\in T$ 有一个权 $w(T)\in Q^+$。问题的度量是测试集的成员的权的总和。它的贪心算法是：在每一步，在没有选择测试中选择价格最小的测试，一个测试的价格定义为这个测试的权与当前它区分的条目对个数之比。

利用第 2 节的自然规约可以证明加权 MTS 的贪心算法有近似比 $H_{n(n-1)/2}=2\ln n+O(1)$。

命题 3 加权 MTS 的贪心算法的近似比不能通过对偶拟合方法改进超过一个常数。

证明：令 $S=\{a_1,a_2,\cdots,a_n\}$，$n=2n'$，$n'\geqslant 3$，$\mathrm{T}=\{T_0,T_1,\cdots,T_n\}$，$T_0=\{a_1,a_2,\cdots,a_{n'}\}$，

$T_i=\{a_i\}$，$1\leqslant i\leqslant n$，$w(T_0)=1$，

$$w(T_i)=\frac{2(n'-i+1)-1}{(n'-i+1)^2}, 1\leqslant i\leqslant n',$$

$$w(T_i)=\frac{2}{2n'-i+1}, n'+1\leqslant i\leqslant 2n'-1,$$

$w(T_n)=2$。

不难证明贪心算法可能按照 $T_1, T_{n'+1}, T_2, T_{n'+2}, T_3, T_{n'+3}$ 直到 $T_{n'-1}, T_{n'+(n'-1)}$，最后 $T_{n'}$ 的顺序选择测试得到解 $\mathrm{T}'=\{T_1, T_2, \cdots, T_{n-1}\}$。

考虑算法结束后，在 T_0 区分的 n'^2 条边上填的价格。选择 T_1 时，每个边 $\{a_1,a_j\}$，$n'+1\leqslant j\leqslant 2n'$ 上填的价格是 $\frac{1}{n'^2}$，和是 $\frac{1}{n'}$；选择 $T_{n'+1}$ 时，每个边 $\{a_{n'+1},a_j\}$，$2\leqslant j\leqslant n'$ 填的价格是 $\frac{1}{n'(n'-1)}$，和是 $\frac{1}{n'}$。一般地，对 $1\leqslant i\leqslant n'-1$，选择 T_i 时，每个边 $\{a_i,a_j\}$，$n'+i\leqslant j\leqslant 2n'$ 上填的价格是 $\frac{1}{(n'-i)^2}$，和是 $\frac{1}{n'-i}$；选择 $T_{n'+i}$ 时，每个边 $\{a_{n'+1},a_j\}$，$i+1\leqslant j\leqslant n'$ 上填的价格是 $\frac{1}{(n'-i)(n'-i-1)}$ 和是 $\frac{1}{n'-i}$。最后，选择 $T_{n'}$ 时，边 $\{a_{n'},a_{2n'}\}$ 上填的价格是 1。这样，T_0 区分的 n'^2 条边上填的价格总和是 $2H_{n'}-1=2\ln n+O(1)$。为了使得这个总和缩小到 $w(T_0)=1$，收缩的因子至少为 $2\ln n+O(1)$。

结束语 目前，MTS 的最佳近似比是贪心算法的 $2\ln n+O(1)$。这个近似比直接由 MSC 的贪心算法的近似比得到，因而是一个平凡的结果。MTS 的近似比能否改进，是算法理论界的一个公开的问题。而用来证明 MTS 的对数近似比的两个方法：对偶拟合和随机取整，都是基于线性规划松弛的方法。因此，讨论 MTS 的基于线性规划松弛的近似比证明方法的能力，就是一个有重要意义的问题。

本文的贡献有两个方面。一个是讨论了 MTS 的整性间隙，给出 MTS 的整性间隙的一个下界，并且指出 MTS 的整性间隙与 MSC 的整性间隙的关系。另一个是对加权 MTS，指出了对偶拟合方法的局限性。就作者所知，这是文献中首次对加权 MTS 给出的理论分析。

参 考 文 献

1 Garey M R, Johnson D S. Computers and Intractability: A Guide to the Theory of NP-completeness. W. H. Freeman, San Francisco, 1979

2 Moret B M E, Shipiro H D. On minimizing a set of tests. SAIM Journal on Scientific and Statistical Computing, 1985, 6: 983～1003

3 Johnson D S. Approximation algorithm for combinatorial problems. Journal of Computer System and Sciences, 1974, 9: 256～278

4 Kann V. On the Approximability of NP-complete Optimization Problems: [PhD thesis]. Royal Institute of Technology, Stockholm, Sweden, 1992

（下转第 166 页）

表1 第t年GDP与第t－k年经济指标的相关系数

与第t年GDP的相关系数		第$t-k$年经济指标数据				
		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
经济指标名称	固定资产投资	0.5889	0.5483	0.1775	－0.3132	－0.3850
	实际利用外资	0.7664	0.2519	－0.2602	－0.3822	－0.1750
	外贸出口总额	0.6514	0.2252	0.0703	－0.1630	－0.1807
	所在省GDP	0.6808	0.4247	－0.0526	－0.4767	－0.2602
	国家GDP	0.3567	0.0645	－0.2535	－0.5098	－0.4574
	自身GDP	0.6038	0.1725	－0.2267	－0.4340	－0.1548
	财政支出	0.5596	0.2034	－0.1523	－0.2837	－0.2487
	社会消费品零售总额	0.5010	－0.0317	－0.2032	－0.5375	－0.2395

从表1可得出如下结论：固定资产投资对GDP的稳定增长综合贡献最大，且可在两年内产生效益，其余的各项指标对GDP增长的贡献则主要集中在一年；引进外资和外贸出口对GDP的增长作用最直接；就环境而言，GDP与广东省的GDP相关性最大，与国家GDP增长关系最小。

以相关系数大于0.5为界，对影响GDP增长的指标进行预测，最终选取如下8个指标作为区域经济发展的预测模型的输入：江门市GDP($t-1$)、广东省GDP($t-1$)、外贸出口总额($t-1$)、财政支出($t-1$)、社会消费品零售总额($t-1$)、固定资产投资($t-1$)、固定资产投资($t-2$)以及实际利用外资($t-1$)。

3.2 基于COMR模型的江门市经济发展预测

考虑到经济系统的动态性，为了体现对近期数据的重视，首先对样本数据进行等比加权操作，对靠近预测年份的样本赋予较大的权值。

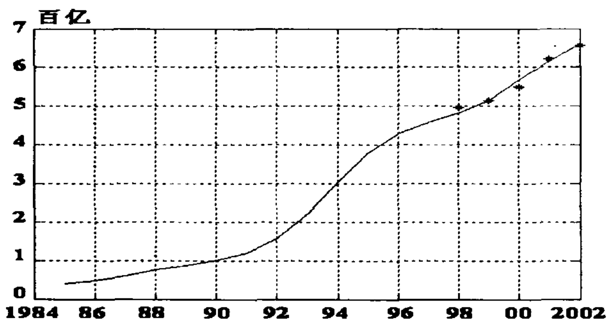

图3 COMR预测值与实际值的比较

为了说明方法的有效性，作者采用1985～1997年的数据作为学习样本，1998～2002年的数据为测试样本，采用式(13)所示的COMR模型进行检验，所得结果如图3和表2所示。

图3中实线为各年度的GDP实际值，“＊”对应相关年份的COMR模型预测值。

表2 预测GDP与实际GDP的比较

年份	预测GDP	实际GDP	误差％
1998	4963300	4818800	2.9990
1999	5139100	5146900	－0.1527
2000	5495800	5675100	－3.1604
2001	6220700	6151600	1.1229
2002	6565300	6608200	－0.6500

图3和表2表明，考虑到实际经济系统的复杂性，由CMOR模型得到的预测结果还是较为理想的。实际上，文献表明，同类研究的预测误差往往在5％以上。

结论 统计学习理论的提出，为复杂经济数据的处理提供了全新的研究途径。本文作者结合区域经济发展的特点，将统计学习理论与线性回归模型相结合，提出了带约束的多元回归模型。由于该模型是建立在统计学习理论的基础上，因而具有更好的泛化能力。同时，该模型有机地结合了经济发展预测的特点，也就保证了该模型能获得比一般计量经济学模型更好的预测精度。

参考文献

1 Hendry D, Ericsson N. Understanding Economic Forecasts. MIT Press, 2001
2 邓宏钟，迟妍，谭跃进．经济系统中的非线性建模与仿真．计算机工程与应用，2001，37(18)：7～9
3 王维，贺京同，张建勋，等．人工神经网络在非线性经济预测中的应用．系统工程学报，2000，15(2)：202～207
4 邵惠鹤．支持向量机理论及其应用．见：自动化博览二十周年纪念文集，2003．90～95
5 Vapnik V. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995
6 Gunn S. Support Vector Machines for Classification and Regression: [Technical Report]. University of Southamptom, 1998
7 Smola A J, Scholkopf B. A tutorial on support vector regression [R]: [NeuroCOLT TR NC-TR-98-030]. Royal Holloway College University of London, UK, 1998

(上接第159页)

5 Lakshmanan K B, Rosenkrantz DJ, Ravi S S. Alarm placement in systems with fault propagation. Theoretical Computer Science, 2000, 243: 269～288
6 De Bontridder K M J, Halldorsson B V. Halldorsson M M, et al. Approximation algorithm for the minimum test set problem: [Alcom-FT Technical Report Series ALCOMFT-TR-02-80]. Available at: http://www.brics.dk/ALCOM-FT/, 2002.
7 Halldorsson B V. Algorithms for Biological Sequence Problems: [PhD thesis]. Department of Mathematical Sciences. Carnegie Mellon University, USA, 2001
8 Chvatal V. A greedy heuristic for the set covering problem. Mathematics of Operations Research, 1979, 4: 233～235
9 Srivasan A. Improved approximations of packing and covering problems. In: Proc. 27th ACM Symposium on the Theory of Computing, 1995. 268～276
10 Lovasz L. On the ratio of optimal integral and fractional covers. Discrete Mathematics, 1975, 13: 383～390