

基于更贪心策略的快速正交核匹配追踪算法

付丽华¹, 李宏伟¹, 张 猛²

(1. 中国地质大学(武汉)数学与物理学院, 湖北武汉 430074; 2. 华中师范大学计算机学院, 湖北武汉 430079)

摘 要: 为了得到信号的稀疏表达结果, 正交核匹配追踪采用贪婪算法, 在逐步回归建模过程中, 每步只寻求当前最优原子, 这使得计算效率大大降低. 针对此局限性, 提出了一种新的更加贪心的策略: 在每次回归时, 选择超过阈值的一个或者多个原子. 为了更好的提高算法精度和稀疏度, 再利用原子相似度对所挑选的原子做进一步筛选. 实验结果和计算复杂度的分析说明: 较传统的方法, 新的基于更贪心策略的方法不仅能够提高计算效率, 而且所得到的模型具有稀疏性好, 泛化能力高等优点.

关键词: 更贪心策略; 正交; 核匹配追踪; 稀疏

中图分类号: TP301.6 **文献标识码:** A **文章编号:** 0372-2112 (2013) 08-1580-06

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2013.08.020

Fast Orthogonal Kernel Matching Pursuit Based on Greedier Strategy

FU Li-hua¹, LI Hong-wei¹, ZHANG Meng²

(1. School of Mathematics and Physics, China University of Geosciences, Wuhan, Hubei 430074, China;

2. Department of Computer, Central China Normal University, Wuhan, Hubei 430079, China)

Abstract: Orthogonal kernel matching pursuit (OKMP) for constructing sparse kernel models has been recently introduced, in which a greedy scheme is utilized to select a single element per iteration. To improve the efficiency and performance of the greedy-scheme-based OKMP, a greedier algorithm is considered. The main contribution is the development of a new selection strategy that effectively selects several elements in each iteration. The efficiency is achieved by reducing the regressor steps, thus the computation time of the orthogonalization that each newly selected regressors to all the selected terms before is saved. A pruned algorithm is proposed based on the similarity of the atoms to improve the accuracy of the approximation. Numerical results and computational complexity analysis show that this new scheme is capable of producing a much sparser regression model with better generalization than the conventional approaches.

Key words: greedier strategy; orthogonal; kernel matching pursuit; sparse

1 引言

核匹配追踪 (Kernel Matching Pursuit, KMP) 方法是在匹配追踪 (Matching Pursuit, MP) 的基础上发展起来的^[1], 其主要思想是利用核函数方法将训练样本映射到高维空间中, 形成一组基函数字典库, 然后利用贪婪算法在基函数字典库中寻找一组使目标函数最优的原子的线性组合. 与 MP 不同的是, KMP 利用核函数方法生成字典库. 此外, KMP 的优点还在于对核函数的要求很低, 不需要满足 Mercer 条件; 与匹配追踪以及支持向量机 (Support Vector Machine, SVM) 相比, 具有非常稀疏的解^[1~3]. 目前, 已经成功的应用到目标分类^[4]、图像分类^[5,6]、手写体识别^[7]、特征模式识别^[8]、目标识别^[9]等方面.

正交核匹配追踪 (Orthogonal Kernel Matching Pursuit, OKMP) 的特点是在每提取一个原子后都要对已提取的各原子集合进行一次 Gram-Schmidt 正交化运算. 显然, OKMP 算法进行稀疏表示的效果与事先建立的字典库息息相关. 传统的 OKMP 将所有的训练样本作为核函数中心, 核尺度则经过类似交叉验证等方法确定后固定不变. 但是, 当输入信号在不同区域震荡程度不一样的时候, 固定尺度的 OKMP 则需要更多的原子来表示信号, 势必破坏了表示的稀疏性. 关于字典库的构成, Chen S. 提出了正交最小二乘 (Orthogonal Least Square, OLS) 的方法^[10], 避免了冗余字典库的建立, 利用重复加权提升搜索算法搜索每次回归时核函数的最优参数; Billings 则提出了用样本点或样本点聚类中心作为核函数中心参数, 经验产生的多尺度作为核函数尺度参数^[11].

相较于标准的 KMP, OKMP 容易有更稀疏的解,但也存在一定的局限性. 其一, 在每次回归过程中, OKMP 需要将备选原子与所有的已选原子做正交化, 过程较为耗时; 其二, 由于 OKMP 是基于贪婪算法的, 每次只选择一个最优的原子, 效率不高.

本文在此基础上, 提出一种新的基于更加贪心策略的正交核匹配追踪方法(GSOKMP). 在每次回归时, 不是仅仅选择一个使目标函数最优的原子, 而是选择大于阈值的一个或者多个原子, 由于挑选出来的原子可能存在较高的相关度, 再利用原子相似度对挑选出来的原子做进一步的筛选. 这样的做法, 不仅能够提高效率, 缩短迭代步数, 而且由于迭代步数的减少, 正交化的过程也大大减少了. 计算复杂度分析和实验验证了本方法不仅具有较好的精度和稀疏度, 而且效率大大提高.

2 正交核匹配追踪(OKMP)方法

2.1 正交匹配追踪方法(OMP)

给定 $\{\mathbf{x}(t), \mathbf{y}(t)\}_{t=1}^N$ 为 N 个训练样本. MP 的主要思想是在一个有限的具有 M 个原子的字典库 $\Omega = \{\boldsymbol{\varphi}_m\}_{m=1}^M$ 中将训练样本分解为库中一组成员的线性组合:

$$\mathbf{y} = \hat{\mathbf{f}} + \mathbf{e} = \sum_{m=1}^n w_m \boldsymbol{\varphi}_m + \mathbf{e} \quad (1)$$

$\mathbf{y} = [y(1), \dots, y(N)]^T$, $\hat{\mathbf{f}}$ 是对训练样本值 \mathbf{y} 的逼近, n 为最终表示的原子个数, \mathbf{e} 为误差.

OMP 则是在每次的迭代过程中从字典库 Ω 中挑选与当前残差最匹配的原子, 并在提取一个原子后要对其已提取的各原子集合进行一次正交化:

初始时令 $\mathbf{r}_0 = \mathbf{y}$, 第 1 次迭代时选择字典库里的第 l_1 个原子:

$$l_1 = \operatorname{argmax}_{1 \leq j \leq M} \{C(\mathbf{r}_0, \boldsymbol{\varphi}_j)\} \quad (2)$$

$C(\cdot, \cdot)$ 表示平方相关系数, 具体定义如下:

$$C(\mathbf{x}_1, \mathbf{x}_2) = \frac{(\mathbf{x}_1^T \mathbf{x}_2)^2}{(\mathbf{x}_1^T \mathbf{x}_1)(\mathbf{x}_2^T \mathbf{x}_2)},$$

第一个原子 $\boldsymbol{\alpha}_1$ 即为字典库 Ω 的第 l_1 个原子, $\boldsymbol{\alpha}_1 = \boldsymbol{\varphi}_{l_1}$, 正交基函数中第一个原子 $\mathbf{q}_1 = \boldsymbol{\varphi}_{l_1}$, 且第 1 次迭代后残差:

$$\mathbf{r}_1 = \mathbf{r}_0 - \frac{\mathbf{r}_0^T \mathbf{q}_1}{\mathbf{q}_1^T \mathbf{q}_1} \mathbf{q}_1 \quad (3)$$

设在第 $m-1$ 次提取后取得了 $m-1$ 个原子 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_{m-1}$, 并且经过 Gram-Schmidt 正交化过程计算得到了一组正交基 $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{m-1}$, 则第 m 步迭代时, 让字典库的原子先跟已选原子做一个正交化,

$$\boldsymbol{\varphi}_j^{(m)} = \boldsymbol{\varphi}_j - \sum_{k=1}^{m-1} \frac{\boldsymbol{\varphi}_j^T \mathbf{q}_k}{\mathbf{q}_k^T \mathbf{q}_k} \mathbf{q}_k \quad (4)$$

第 m 次迭代所选原子为第 l_m 个:

$$l_m = \operatorname{argmax}_{j \neq l_1, l_2, \dots, l_{m-1}} \{C(\mathbf{r}_{m-1}, \boldsymbol{\varphi}_j^{(m)})\} \quad (5)$$

于是, $\boldsymbol{\alpha}_m = \boldsymbol{\varphi}_{l_m}$, $\mathbf{q}_m = \boldsymbol{\varphi}_{l_m}^{(m)}$, 第 m 次迭代后残差为:

$$\mathbf{r}_m = \mathbf{r}_{m-1} - \frac{\mathbf{r}_{m-1}^T \mathbf{q}_m}{\mathbf{q}_m^T \mathbf{q}_m} \mathbf{q}_m \quad (6)$$

当 $\|\mathbf{r}_m\|^2 < \eta$ 时, 迭代停止. 阈值 η 可以选择为噪声的方差, 但由于噪声方差一般都是未知的, 这里将采用交叉验证的方法确定 η .

经过 n 次迭代即可得到训练样本的稀疏表示结果:

$$\mathbf{y} = \sum_{m=1}^n \frac{\mathbf{r}_{m-1}^T \mathbf{q}_m}{\mathbf{q}_m^T \mathbf{q}_m} \mathbf{q}_m + \mathbf{r}_n \quad (7)$$

由以上过程可知, OMP 采用贪婪算法, 在逐步回归的过程中, 只在当前字典库中寻求与当前残差最为匹配的一个最优原子. 当迭代达到终止条件即结束信号的稀疏分解过程. “稀疏”的含义就是使所需原子数目最少(即 n 越小越好).

2.2 正交核匹配追踪(OKMP)方法

OKMP 的本质是采用核方法生成字典库, 首先将训练样本映射到高维空间, 并通过计算样本间核函数值来生成字典库 $\Omega = \{\boldsymbol{\varphi}_i = K(\cdot, \mathbf{x}_i)\}_{i=1}^N$, 然后与 OMP 过程一样寻找最“稀疏表示”的基函数的线性叠加. OKMP 的想法来源于 SVM, 但是却不需要满足 SVM 中要求核函数必须满足的 Mercer 条件^[12]. 常用的核函数有:

$$\text{高斯核: } K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right)$$

$$\text{多项式核: } K(\mathbf{x}, \mathbf{x}_i) = [(\mathbf{x} \cdot \mathbf{x}_i) + 1]^q$$

$$\text{S 形函数核: } K(\mathbf{x}, \mathbf{x}_i) = \tanh(v(\mathbf{x} \cdot \mathbf{x}_i) + c)$$

本文将采用高斯核函数. 传统的基于高斯核函数的 OKMP 将所有的训练样本(或者随机的选择训练样本的子空间)作为核函数中心, 采用固定不变的尺度参数形成字典库. 这样的处理方式较为简单, 但是表示结果的稀疏度不够高.

2007 年, Billings 提出多尺度径向基函数网络(Multi-scale Radial Basis Function, MSRBF)^[11]: 核函数采用的是高斯核, 当样本点长度不是很长时, 就用样本数据来作为核函数中心; 当样本长度较大时, 可以用 k 均值聚类的方法得到数据点中心, 即为核函数中心. 高斯核函数的尺度则通过经验给出 $(p+1)$ 个多尺度:

$$\sigma = \max_t \{y(t)\} - \min_t \{y(t)\},$$

$$\sigma^{(i)} = 2^{-i} \sigma, i = 0, 1, \dots, p \quad (8)$$

3 基于更贪心策略的正交核匹配追踪(GSOKMP)方法

3.1 基于阈值的一次选多个回归子

由于 OKMP 及在此基础上改进的 MSRBF 算法是基

于贪婪算法的:如公式(2)、(5)所示,每次迭代过程只选择与当前残差平方相关系数最大的一个原子作为当前回归子.这样,当面对大数据量样本时,迭代步数会增加,从而影响计算效率.这里,将考虑一次选择多个回归子:

在第 m 次迭代过程中,令

$$g_j^{(m)} = \{C(\mathbf{r}_{m-1}, \mathbf{q}_j^{(m)})\}_{j \neq l_k, 1 \leq k \leq m-1}$$

$g_j^{(m)}$ 为当前残差与字典库的第 j 个原子的平方相关系数结果.

$$I^{(m)} = \{i: |g_i^{(m)}| \geq \alpha \cdot \max_j |g_j^{(m)}|, j \neq l_k, 1 \leq k \leq m-1\} \quad (9)$$

α 为阈值参数 ($\alpha \leq 1$), $I^{(m)}$ 为第 m 次迭代过程中所有超过 $\alpha \cdot \max_j |g_j^{(m)}|$ 的原子下标的集合.很显然,当 $\alpha = 1$ 时, $I^{(m)}$ 就只包含当前最优的一个原子,即为经典的贪婪算法.这里,更贪心策略是指当 α 小于 1 时, $I^{(m)}$ 就有可能包含了不只最优的一个结果,从而在第 m 次迭代过程可能就挑选了不只 1 个最优的回归子,可以大大提高效率.

3.2 原子相似度进行进一步筛选

令 $card(I^{(m)})$ 表示指标集 $I^{(m)}$ 中元素的个数,即第 m 次迭代中挑选出原子的个数.很明显,更贪心策略中阈值参数 α 越小,那么每步迭代中选中的原子个数(即 $card(I^{(m)})$)就会较多,迭代步数较少,但是可能由于原子之间相关度较高会导致误选原子的概率较大,从而影响了回归的精度和稀疏度.

相反, α 越大,迭代步数会较大,每步迭代中选中的原子个数(即 $card(I^{(m)})$)就会较少,虽然误选原子的概率较小,但是由于迭代步数与正交化次数紧密相连,从而时间耗费较多.

如何又能减少迭代步数(等同于减少正交化次数,及减少计算时间),又能提高稀疏度和精度呢?这里,在第 m 次迭代中将采用原子相似度的方法对所选的 $card(I^{(m)})$ 个原子做进一步筛选.

原子相似度定义为:

$$\rho_{ij} = \frac{\langle \mathbf{a}_i, \mathbf{a}_j \rangle}{\|\mathbf{a}_i\| \|\mathbf{a}_j\|}$$

给定相似度阈值 ξ , $0 \leq \xi \leq 1$, 当挑选的原子与当前最优原子的相似度小于 ξ 时,该原子被保留.

3.3 计算复杂度分析

传统的 OKMP 算法中,设 N 为训练样本个数, M 为字典库里原子个数, n 为迭代步数,即最终表示为 n 个原子的线性叠加,则所需计算次数为:

$$(6N-1) \times M \times [1+2+\dots+(n-1)] +$$

$$6N \times M \times n + (6N-1) \times n$$

$$= \frac{(6N-1)}{2} \times M \times n^2 + \frac{(6N+1)}{2} \times M \times n + (6N-1) \times n$$

计算复杂度为 $O(N \times M \times n^2)$. 当样本点长度和字典库个数固定时,计算复杂度与迭代步数 n 的平方呈正比关系.事实上,在该算法中,正交化过程耗费计算量较大:当迭代步数 n 较大时,正交化过程也增多,从而导致计算复杂度也增大.

下面考虑 GSOKMP 算法的计算复杂度,算法分为两部分:基于阈值的核匹配追踪和计算原子相似度.假设经过 n_c 步迭代,新算法达到与 OKMP 相同的稀疏度(即最终也是 n 个原子的线性叠加).显然, $\sum_{m=1}^{n_c} card(I^{(m)}) = n$.

GSOKMP 算法的计算复杂度总计为:

$$(6N-1) \times M \times [card(I^{(1)}) + \sum_{m=1}^2 card(I^{(m)}) + \dots + \sum_{m=1}^{n_c} card(I^{(m)})] + 6N \times M \times n_c + (6N-1) \times n + (6N-3) \times (n - n_c)$$

事实上,两个算法中复杂度最大的都在正交化过程, OKMP 中正交化所需计算复杂度为:

$$(6N-1) \times M \times [1+2+\dots+(n-1)];$$

GSOKMP 则为:

$$(6N-1) \times M \times [card(I^{(1)}) + \sum_{m=1}^2 card(I^{(m)}) + \dots + \sum_{m=1}^{n_c} card(I^{(m)})],$$

由于

$$\{card(I^{(1)}), \sum_{m=1}^2 card(I^{(m)}), \dots, \sum_{m=1}^{n_c} card(I^{(m)})\} \subset \{1, 2, \dots, n\}$$

因此, GSOKMP 算法将在正交化过程中少耗时.当训练样本为大尺度数据时, GSOKMP 算法在计算复杂度方面的优势将更加明显.

4 算法

GSOKMP 的具体算法如下:

Step1 初始化: $\mathbf{r}_0 = \mathbf{y}$;

Step2 利用训练样本的子集作为核函数的中心,尺度则利用式(8)生成具有 M 个原子的多尺度核字典 $\Phi = \{\boldsymbol{\varphi}_m\}_{m=1}^M$;

Step3 for $m=1; m=m+1$ 直到满足迭代终止条件.

$$(1) \mathbf{q}_j^{(m)} = \boldsymbol{\varphi}_j - \sum_{k=1}^{m-1} \frac{\boldsymbol{\varphi}_j^T \mathbf{q}_k}{\mathbf{q}_k^T \mathbf{q}_k} \mathbf{q}_k, 1 \leq j \leq M;$$

(2) 计算:

$$g_j^{(m)} = \{C(\mathbf{r}_{m-1}, \mathbf{q}_j^{(m)})\}_{j \neq l_k, 1 \leq k \leq m-1};$$

(3) 基于阈值选择第 m 步迭代的原子下标集合:

$$I^{(m)} = \{i: |g_i^{(m)}| \geq \alpha \cdot \max_j |g_j^{(m)}|, j \neq l_k, 1 \leq k \leq m-1\}$$

(4) 利用原子相似度做进一步筛选:当挑选的原子与 $\max_j |g_j^{(m)}|$ 的相似度小于 ξ 时,该原子被保留.

(5) 更新残差:

$$r_m = r_{m-1} - \sum_{i=1}^{cont(I^{(m)})} \frac{r_{m-1}^T q_i^{(m)}}{(q_i^{(m)})^T q_i^{(m)}} q_i^{(m)}$$

5 实验

5.1 参数设置

本文所提的 GSOKMP 中阈值参数 α 和原子相似度阈值参数 ξ ,将运用交叉验证与网格搜索的方法得到.其中, $\alpha \in \{0.5, 0.6, \dots, 1.0\}$, $\xi \in \{0.5, 0.55, 0.6, 0.65, \dots, 0.95\}$.

5.2 仿真数据实验

考虑 sinc 函数的拟合问题:

$$f(x) = \frac{\sin x}{x}, -10 \leq x \leq 10$$

从 $y = f(x) + \epsilon$ 带噪声数据中产生 400 个等间隔数据点, ϵ 服从正态分布 $N(0, 0.04)$. 200 个偶数点用于训练,余下 200 个奇数点做测试.为了更好的与经典的 MSRBF^[11]相比,核函数选择为高斯核函数,中心选择为训练样本的子集(采样间隔为 10),尺度则选择与 MSRBF 实验中 sinc 相同的经验多尺度集合: $\{2.0, 2.5, 3.0, 3.5, 4\}$. 分别用 MSRBF 和新提出来的 GSOKMP 算法对所给数据进行拟合.本实验中 $\alpha = 0.5, \xi = 0.8$.

图 1 是经典的 MSRBF 结果,共需要 8 项回归子,得到此模型的训练误差是 0.0384,测试误差为 0.0396.图 2 为新提出来的算法 GSOKMP 产生的结果图,结果显示 GSOKMP 仅需要经过 2 步迭代,最终表示为 6 项回归子

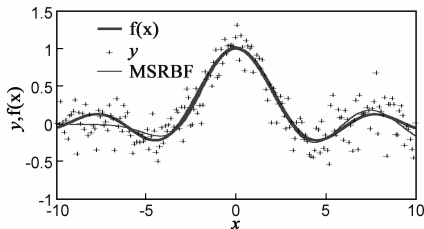


图1 MSRBF拟合结果

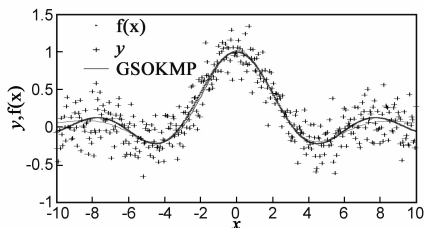


图2 GSOKMP拟合结果

的线性组合.表 1 为新方法所得模型的中心、尺度以及权值参数.训练误差是 0.0371,测试误差为 0.0378.

表 1 仿真数据实验中 GSOKMP 得到的模型参数

Centre	Scale	Weight	迭代步数
0.0251	2.5000	9.2743	1
1.0276	3.0000	-7.6959	1
-0.9774	3.0000	-8.0058	1
1.0276	3.5000	3.6786	1
-0.9774	3.5000	3.9763	1
0.0251	2.0000	-0.6972	2

将此实验独立重复 30 次,表 2 列出了平均结果.结果显示在这 30 组模拟实验中,经典的 MSRBF 需要 8.5 项来拟合 sinc 函数,而 $\alpha = 0.5, \xi = 0.8$ 时 GSOKMP 仅需要 2 步迭代,6.5 项来拟合,且所需时间最短.另外两组不同的参数下($\alpha = 0.7, \xi = 0.8$ 和 $\alpha = 0.9, \xi = 0.9$)实验结果同见表 2.实验结果说明,通过交叉验证选取阈值参数时,新的 GSOKMP 稀疏性较好,计算速度也较快.

表 2 仿真数据实验中 30 次实验的平均结果

	MSRBF	GSOKMP ($\alpha = 0.5,$ $\xi = 0.8$)	GSOKMP ($\alpha = 0.7,$ $\xi = 0.8$)	GSOKMP ($\alpha = 0.9,$ $\xi = 0.9$)
训练误差	0.0401	0.0386	0.0387	0.0391
测试误差	0.0412	0.0392	0.0409	0.0412
项数	8.5	6.5	7.5	8
迭代步数	8.5	2	5	7
时间	0.3949	0.2844	0.3107	0.3772

5.3 真实 ECG 数据实验

ECG 数据记录了人体的心电信号波形.心电信号的稀疏分解能够得到稀疏的表示形式.从 MIT-BIH 心电数据库,选择第 100 号记录的前 1000 个数据点.同样的,奇数点用于训练,偶数点用于测试.核函数中心选择为样本子集(采样间隔为 2),尺度则由式(8)生成经验多尺度.这样,字典库里共有原子 2500 个.分别用 MSRBF 和 GSOKMP 算法对所给数据进行拟合.

本实验中 $\alpha = 0.95, \xi = 0.6$.当 α 越大时,每次迭代所选原子个数就越少,这样,迭代步数会增加,从而计算效率会大大降低.反之,当 α 越小时,迭代步数会因每次迭代所选原子个数增加而减少.所以,在利用网格搜索法寻找最佳参数 α, ξ 时,可以尝试先让迭代步数尽可能的少,即先确定 α 的值,然后再调节 ξ .图 3 和图 4 中“.”表示心电信号原始数据,可以看出,心电信号的峰值很尖锐,变化很陡峭.图 3 是 MSRBF 拟合出的结果,其中共用到了 51 项回归子,迭代步数为 51 次,所耗时间为 463.8325s,其中训练误差为 0.0008,测试误差为 0.0007.

图4为GSOKMP拟合的结果,共用26步迭代的结果,得到了40项回归子,具体模型参数结果如表3所示.所耗时间为234.8957s,训练误差为0.0007,测试误差为0.0007.结果可见,由于GSOKMP节省了更多的迭代次数,一次选择多个回归子,所以具有比MSRBF方法更快的衰减速度(如图5所示),而且模型表示更为稀疏.

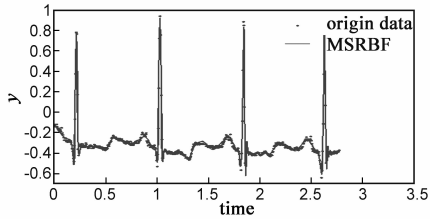


图3 MSRBF拟合ECG MIT100结果

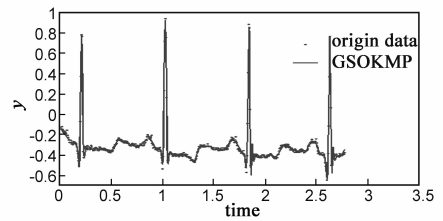


图4 GSOKMP拟合ECG MIT100结果

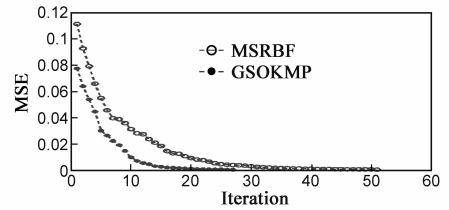


图5 MSRBF和GSOKMP拟合ECG MIT100的逐步误差图

表3 ECG数据实验中GSOKMP得到的模型参数

Centre	Scale	Weight	迭代步数	Centre	Scale	Weight	迭代步数	Centre	Scale	Weight	迭代步数
1.2556	0.1000	-0.4112	1	0.2111	0.0100	1.6470	7	2.4889	0.0300	0.1195	20
2.0111	0.1000	-22797.7	2	1.0222	0.0100	0.8967	8	1.8667	0.0100	0.7623	20
2.0778	0.1000	-23135.7	2	1.0333	0.0700	-0.7455	9	1.5222	0.0600	0.2413	21
2.0222	0.1000	133580.2	2	1.8444	0.0100	1.9367	10	0.2000	0.0100	-0.2459	21
2.0667	0.1000	134932.1	2	1.0333	0.0100	0.7895	11	1.0444	0.0100	-0.4286	22
2.0556	0.1000	-331298.1	2	0.4333	0.0900	1.7535	12	1.0333	0.0500	0.5136	23
2.0333	0.1000	-329609.1	2	1.8556	0.0100	-1.4953	13	2.3889	0.0200	-0.0957	24
2.0444	0.1000	438327.3	2	2.6444	0.0100	-1.2547	14	1.3111	0.0200	-0.0871	24
0.4333	0.1000	-2.0914	2	0.2111	0.0200	-0.4831	15	2.6667	0.0100	-0.1903	25
2.6222	0.1000	-3.3934	3	2.6222	0.0900	3.0415	16	0.5667	0.0200	0.0753	25
0.8111	0.1000	-1.1173	4	0.8222	0.0900	0.8575	17	1.8222	0.0100	-0.1860	25
1.5333	0.0900	-0.5692	5	2.6556	0.0100	0.5779	18	1.8778	0.0100	-0.2702	26
2.6333	0.0100	1.6657	6	2.6000	0.0100	-0.2806	19				
0.1222	0.1000	-0.2505	6	1.0111	0.0100	-0.4301	20				

6 总结

针对OKMP方法中一次迭代只选择一个最优原子,从而导致正交化步数较多使得计算效率不高的问题,本文提出了更贪心策略的GSOKMP算法,通过一次选择多项回归子,节省了算法迭代次数.实验结果表明该方法与经典的MSRBF方法相比,稀疏度有所提升,更重要的是,当面对大数据量数据时,计算效率有明显提高.

参考文献

- [1] Pascal V, Yoshua B. Kernel matching pursuit [J]. Machine Learning, 2002, 48(1-3): 165 - 187.
- [2] Cevher V, Krause A. Greedy dictionary selection for sparse representation [J]. IEEE Journal of Selected Topics Signal Pro-

cessing, 2011, 5(5): 979 - 988.

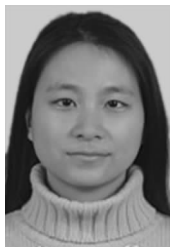
- [3] Sun P, Yao X. Sparse approximation through boosting for learning large scale kernel machines [J]. IEEE Transaction on Neural Networks, 2010, 21(6): 883 - 894.
- [4] Liao X J, Li H, Krishnapuram B. An M-ary kernel matching pursuit classifier for multi aspect target classification [A]. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) [C]. Montreal, Quebec, Canada, 2004. 61 - 64.
- [5] 缙水平, 焦李成. 基于多尺度几何分析与核匹配追踪的图象识别[J]. 模式识别与人工智能, 2007, 20(6): 776 - 781. Gou Shuiping, Jiao Licheng. Image recognition based on multi-scale geometric analysis and kernel matching pursuit [J]. Pattern Recognition and Artificial Intelligence, 2007, 20(6): 776 - 781. (in Chinese).
- [6] 张志龙, 张焱, 沈振康. 基于特征谱的高分辨率遥感图像

港口识别方法[J].电子学报,2010,38(9):2184-2188.

Zhang Zhilong, Zhang Yan, Shen Zhenkang. Port recognition in high resolution remote sensing images based on feature spectrum [J]. Acta Electronica Sinica, 2010, 38(9): 2184-2188. (in Chinese)

- [7] Davis G, Mallat S, Zhang Z. Adaptive time frequency decompositions [J]. Optical Engineering, 1994, 33(7): 2183-2191.
- [8] 李青, 焦李成, 周伟达. 基于模糊核匹配追踪的特征模式识别[J]. 计算机学报, 2009, 32(8): 1687-1694.
- Li Qing, Jiao Licheng, Zhou Weida. Pattern recognition based on the fuzzy kernel matching pursuit [J]. Chinese Journal of Computers, 2009, 32(8): 1687-1694. (in Chinese)
- [9] 雷阳, 雷英杰等. 基于直觉模糊核匹配追踪的目标识别方法[J]. 电子学报, 2011, 39(6): 1441-1446.
- Lei Yang, Lei Yingjie, et al. Techniques for target recognition based on intuitionistic fuzzy kernel matching pursuit [J]. Acta Electronica Sinica, 2011, 39(6): 1441-1446. (in Chinese)
- [10] Chen S, Wang X X, Brown D J. Orthogonal least squares regression with tunable kernels [J]. Electron Lett, 2005, 41(8): 484-486.
- [11] Billings S A, Wei H, Balikhin M. Generalized multi-scale radial basis function networks [J]. Neural Networks, 2007, 20: 1081-1094.
- [12] Scholkopf B, Smola A. Learning with Kernels [M]. Cambridge, MA: MIT Press, 1999.

作者简介



付丽华 女, 1979 年生于湖北枝江, 博士, 中国地质大学数理学院(武汉)副教授. 研究方向为模式识别、信号处理等.



李宏伟(通信作者) 男, 1965 年生于湖南汨罗, 博士, 中国地质大学数理学院(武汉)教授, 博士生导师. 研究方向为信号处理、机器学习等.

E-mail: hwli@cug.edu.cn



张猛 男, 1977 年出生于湖北武汉, 博士, 华中师范大学副教授, 硕士生导师. 研究方向为数据挖掘、模式识别等.