

基于 PLS 和 GAs 的径向函数网络构造策略*

赵伟祥, 吴立德

(复旦大学 计算机科学与工程系, 上海 200433)

E-mail: weixiang_zhao@263.net

http://www.fudan.edu.cn

摘要: 鉴于传统径向函数网络(radial basis function network,简称 RBFN)构造策略的不足,提出了基于偏最小二乘法(partial least squares,简称 PLS)和遗传算法(genetic algorithms,简称 GAs)的 RBFN 构造策略和一种更有效的径向宽度取值方法.在这个集成构造策略中,PLS 克服了 K-Means 算法求取径向易陷入局部最优的弊病,并使合成径向比由正交算法获取的径向更具代表性;而所提出的径向宽度取值方法和 GAs 则为网络性能和结构的实质性改善与优化提供了保障.实验证实了基于 PLS 和 GAs 的 RBFN 构造策略及所提出的径向宽度取值方法的优越性、可靠性和有效性.

关键词: 径向函数网;聚类;正交算法;偏最小二乘回归;遗传算法

中图法分类号: TP183 文献标识码: A

在众多的神经网络中,径向函数网(radial basis function network,简称 RBFN)已在模式识别、时间序列预测及故障诊断等领域得到了广泛应用^[1].构造径向函数网的关键是确定径向个数、径向向量和每个径向的宽度.传统的构造方法是 K-Means 聚类算法和正交学习算法(orthogonal learning algorithm,简称 OLA)^[2,3].然而在 K-Means 算法中,径向数目是经验值,径向向量的求取又极易陷入局部最小;而 OLA 则无法保证所选径向能够反映训练样本集的全部信息.为此,有学者提出采用遗传算法(genetic algorithms,简称 GAs)来获得径向向量及径向宽度^[1],但这将耗费巨额的搜索时间.

本文提出了一种基于偏最小二乘法(partial least square,简称 PLS)和 GAs 的 RBFN 构造策略,以克服常规 RBFN 构造策略的不足,并把由此得到的网络称为偏径向函数网(PRBFN).另外,为了充分利用样本空间分布信息,本文提出了一种更有效的径向宽度取值方法,并通过它与 GAs 的结合有效地解决了网络构造中的两个耦合问题(给定网络结构,如何使网络精度达到最高;给定网络拟合误差上限,如何使网络结构趋于最简,并在这个最简结构的基础上使得网络精度达到最优),从而为 PRBFN 的最优设计提供了根本保障.本文最后通过一系列的实验证实了该 RBFN 集成构造策略以及所提出的径向宽度取值方式的优越性、有效性和可靠性,也为今后更广泛的应用提供了依据.

1 常规 RBFN 构造策略

标准 RBFN 通常是一种两层前传网^[2].网络输入层接收外来信号,输入层与隐层间的权向量 w_i 对应于径向向量 c_i ,输出层则对隐层输出加权求和.整个系统可用一个插值函数 $f(x_i)$ 来描述(以一维输出为例):

$$f(x) = \sum_{i=1}^m v_i \exp(-\|x - c_i\|^2 / \sigma_i^2). \quad (1)$$

* 收稿日期: 2001-05-28; 修改日期: 2002-01-10

基金项目: 国家自然科学基金资助项目(69935010)

作者简介: 赵伟祥(1973 -),男,浙江绍兴人,博士,主要研究领域为神经网络,模式识别,智能信息处理;吴立德(1937 -),男,江苏宜兴人,教授,博士生导师,主要研究领域为图像处理,模式识别,计算机视觉,文本处理.

式中 x 为输入样本向量(注:本文所指的样本向量以及由其产生的中心向量和径向向量均为行向量), \cdot 为欧氏范数, m 为径基数目, v_i 为隐层至输出层的权系数, σ_i^2 为宽度参数.在 OLA 策略中,径向向量 c_i 由正交算法选取所得^[2], σ_i^2 往往取相同的值,显然,我们无法保证被选径基能充分反映样本集信息;而在 K-Means 算法中,由聚类求取径向向量 c_i 又极易陷入局部最优,其 σ_i^2 一般按下式计算^[2]:

$$\sigma_i^2 = \frac{1}{M_i} \sum_{x \in \theta_i} (x - c_i)(x - c_i)^T \quad (2)$$

式中, M_i 是属于第 i 类的样本个数, c_i 是该类的中心向量, θ_i 是属于该类的样本子集.显见上述两种构造策略中的不足都会影响网络的最终性能.

其实,如果将所有样本都作为径基(暂称其为样本径基,而所在的隐层则可称为样本径基层),就无须通过聚类形成径基,而且该径基层所含的信息是最充分的.但随之而来的问题是:(1) 网络模型的过拟合;(2) 如果有样本靠得很近,则会使最小二乘(least squares,简称 LS)求解输出层权系数的系数矩阵病态,甚至奇异.

为此,本文将采用多元统计方法来解决上述问题.将线性统计方法与径基函数相结合的研究已始有报道^[4,5],但这些研究的主要目的是拓宽线性统计方法的应用范围,而且也没有对模型关键参数的选取作深入探讨.因此,本文将分别采用两种常用的多元统计方法,主成分分析(principal component analysis,简称 PCA)和 PLS 对样本径基进行成分分析,并将性能较优的方法与 GAs 相结合,提出 RBFN 的集成构造策略,而后对径基宽度等关键参数的取值作深入探讨,以寻求一条确定参数的有效途径.

2 基于 PLS 和 GAs 的 RBFN 构造策略

2.1 PCA和PLS

PCA 和 PLS 都是非常有效的数据降维和特征提取手段,已有很多文献对它作了详细介绍^[6-8].PCA 或 PLS 成分矩阵 $T = (t_1, t_2, \dots, t_k)$ 与原自变量矩阵 X 的关系均可写为

$$T = XP \quad (3)$$

其中 P 为转换矩阵.但从与因变量的相关关系来看,在成分提取过程中,PCA 并没有考虑自变量与因变量的关系,而 PLS 却保留了较多的与因变量的相关性,显然,这对回归模型是有益的.在下文的实验中我们可以发现 PLS 较之 PCA 的优点.

2.2 PLS在网络构造的作用

PLS 对样本径基层的输出进行降维处理(PCA 作用相似,故不另加说明).我们将由此得到的合成径基(synthesized radial basis,简称 SRB)称为 PLS 径基,形成的网络称为 PRBFN,其结构如图 1 所示.现参照图 1 将 PRBFN 的结构及训练过程简述如下(假设有 m 个训练样本):

(1) 将 m 个训练样本构成由 m 个节点组成的隐层(样本径基层),其权向量 w_i 即样本向量 x_i .这样 m 个样本经过隐层就形成一个 $m \times m$ 的输出矩阵 A ,其元素 $a_{ji} (j, i = 1, 2, \dots, m)$ 可按式(4)求取.

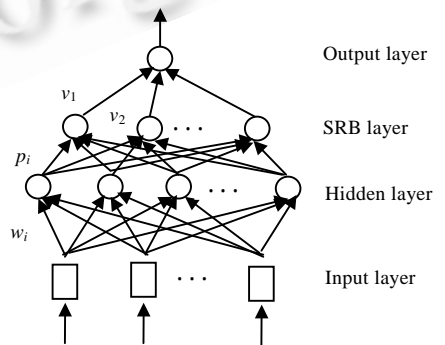
$$a_{ji} = \exp\left(-\frac{\|x_j - x_i\|^2}{\sigma_i^2}\right) \quad (4)$$

(2) 输出矩阵 A 经过 PLS 产生 k 个合成径基,并构成合成径基层.隐层和合成径基层间的权矩阵即为式(3)中的转换矩阵 P .

(3) 对 PLS 成分矩阵和目标 Y 进行 LS 回归,求取输出层的权系数 v_1, v_2, \dots, v_k .

确定合成径基(PLS 成分)数目 k 的常用方法简述如下,具体内容可参考文献[7,8].

- 以自变量矩阵或因变量矩阵的某种范数作为判别标准.



输入层, 隐层, 合成径基层, 输出层.

Fig.1 The structure of PRBFN

图 1 PRBFN 的结构

- 以从自变量矩阵和因变量矩阵提取的得分向量的相关性作为标准.

至此我们已经通过 PLS 确定了合成径基层的节点数和权向量.

2.3 径基宽度参数的取值

对 PRBFN 而言,除了合成径基个数外,可调节参数只有样本径基宽度 σ_i^2 . 鉴于现有宽度取值方法的不足,我们提出一种充分利用样本空间分布信息的有效取值方法. 具体步骤见式(5)和式(6).

$$d_i^2 = \frac{1}{N} \sum_{x_j \in \theta_i} (x_i - x_j)(x_i - x_j)^T, \quad (5)$$

其中 d_i^2 代表第 i 个样本 x_i 在空间分布的离散度, θ_i 是与 x_i 最近的 N 个样本所组成的集合.

$$\sigma_i^2 = \mu(d_i^2)^\alpha, \quad (6)$$

式中 σ_i^2 为第 i 个样本径基的宽度, μ 是比例参数, α 是形状参数.

显见由 N , μ 和 α 所确定的样本径基宽度取值方法既反映了样本空间分布的信息,又增加了式(2)所示取值方法的多样性和自适应性,从而为研究如何提高网络精度和优化网络结构带来了方便.

2.4 GAs在网络构造中的作用

上文通过 PLS 确定了合成径基,同时给出了径基宽度的取值方法. 本节将通过 GAs 来解决 PRBFN 构造中的两类耦合问题,从而为网络结构的优化提供根本保障.

第 1 类耦合问题就是在给定合成径基个数后,如何获取一组 N , μ 和 α , 使网络拟合精度达到最佳. 本文提出将 PRBFN 与 GAs 结合,以确定最佳方案. GAs 在优化问题上的优越性能及其算法已有多处报道^[9,10],这里不加累述. 现将解决该优化问题的适应度函数定义为

$$f = 1/(1 + \bar{E}_{N,\mu,\alpha}\{x_i\}), \quad (7)$$

其中, $\bar{E}_{N,\mu,\alpha}\{x_i\}$ 表示给定合成径基个数后,网络在 N , μ 和 α 为某一状态时所有输入样本 x_i ($i = 1, 2, \dots, m$) 的实际输出与它们目标输出的平均相对误差. 很明显,适应度越大的染色体所对应的平均相对拟合误差就越小.

第 2 类耦合问题则是如何在拟合误差不高于设定值的情况下,使网络规模最小. 虽然如第 2.2 节所述,合成径基个数可以按统计方法来确定,但界限值的选取却是一个经验问题,为此本文将它与其他 3 个参数一并优化,以期找到一个最佳方案. 优化方法仍然采用 GAs,相应的适应度函数则为

$$f = \begin{cases} 1/(k + \bar{E}_{k,N,\mu,\alpha}\{x_i\}), & \text{if } \bar{E}_{k,N,\mu,\alpha}\{x_i\} < \delta, \\ 0.00001, & \text{else.} \end{cases} \quad (8)$$

其中, k 为合成径基个数,也就是该问题的优化目标, $\bar{E}_{k,N,\mu,\alpha}\{x_i\}$ 表示网络在 k , N , μ 和 α 为某一状态时所有输入样本 x_i ($i = 1, 2, \dots, m$) 的实际输出与它们目标输出的平均相对误差, δ 为设定的误差上限. 从式(8)可见,如果某个成员(染色体)所对应的网络平均相对误差大于设置的上限,那么该成员的适应度将明显小于任何能满足拟合精度要求的个体所对应的适应度. 同时对满足拟合精度要求的个体而言,由于 $\bar{E}_{k,N,\mu,\alpha}\{x_i\} < \delta < 1$, 所以只有具有相同 k 值的成员才具有可比性. 因此,适应度越大的个体越为我们所需求,而且最终优化结果在理论上能实现以下两个目标:(1) 与所有满足拟合精度要求的 PRBFN 相比,优化结果所构造的 PRBFN 具有最少的合成径基个数;(2) 该 PRBFN 的拟合误差在同结构 PRBFN 中又是最小的.

与传统 RBFN 构造策略相比,本文提出的网络构造策略首先克服了由聚类形成径基的随机性,其次通过统计方法构造合成径基层,既利用了所有样本的信息,又避免了将所有样本作为径基可能带来的副作用;而所提出的径基宽度取值方法和 GAs 又为提高网络性能、优化网络结构提供了保障. 下面我们将验证基于 PLS 和 GAs 的 RBFN 构造策略的性能,并对关键参数的取值作详细讨论.

3 实验

3.1 实验1

本实验将通过一个函数逼近的实例来验证基于 PLS 合成径基的 RBFN 构造策略的优点. 用于函数逼近的

数据来源于式(9),具体方法是在区间[0,4]和[-1,1]内分别产生 100 个随机数作为自变量 x 和 y 的取值,并产生 100 个相应输出.在这 100 组数据中选取 60 组作为训练样本,剩余 40 组作为预测样本.

$$z = \sin(x) + y^2. \tag{9}$$

我们分别观察基于 K-Means 算法,OLA,PLS 和 PCA 的 RBFN 构造策略的建模效果.为了便于比较,基于 OLA,PLS 和 PCA 构造策略的各样本半径宽度都取 1.具体结果见表 1 和表 2.

Table 1 The modeling effects of RBFNs based on K-Means algorithm and OLA

表 1 基于 K-Means 算法和 OLA 的 RBFN 建模效果

Number of radial bases	K-Means		OLA	
	Average relative error of self-test (%)	Average relative error of prediction (%)	Average relative error of self-test (%)	Average relative error of prediction (%)
8	123.70	68.83	82.96	18.94
12	103.05	58.18	17.00	9.67
20	16.10	22.19	4.90	5.60

径基个数, K-Means 算法, 正交学习算法, 平均自检相对误差, 平均预测相对误差.

Table 2 The modeling effects of PLS and PCA based RBFNs

表 2 基于 PLS 和 PCA 的 RBFN 建模效果

Number of SRB	Average relative error of self-test with PLS (%)	Average relative error of prediction with PLS (%)	Average relative error of self-test with PCA (%)	Average relative error of prediction with PCA (%)
8	7.13	6.29	7.40	6.57
12	3.59	4.94	8.01	5.86
18	0.33	0.2	0.4	2.04
24	0.05	0.2	0.15	0.87

合成径基数目, 采用 PLS 的平均自检相对误差, 采用 PLS 的平均预测相对误差, 采用 PCA 的平均自检相对误差, 采用 PCA 的平均预测相对误差.

从表 1 和表 2 可见,基于 K-Means 算法的 RBFN 构造策略的建模效果最不理想.另外,由合成径基(无论通过 PLS 还是 PCA 提取)所构成的 RBFN 的精度都优于具有相同径基数目的基于 OLA 的 RBFN.这个现象证实了合成径基所含信息的充足性以及基于合成径基的 RBFN 构造策略的有效性.

从表 2 还可以看出,提取相同数目的合成径基,采用 PLS 所建模型的精度要比采用 PCA 高.这正说明了在回归问题上 PLS 较之 PCA 的优点.所以,下文将 PLS 与 GAs 相结合,以确保网络性能的可靠.

另外,在实验中发现,合成径基并非越多越好.由于合成径基是由原样本径基作变换后所得,合成径基选取的越多就越接近原样本径基,所以如果用原样本径基直接求输出层权系数就会出现矩阵病态,那么用较大数目的合成径基来求输出层权系数同样会发生矩阵病态,从而影响建模效果.

3.2 实验2

上文证实了 PRBFN 在建模方面的优势,下面将探讨如何使网络性能和结构趋于最优化.为此,本节设计了两个优化问题,并通过 GAs 将本文所提出的网络构造策略的性能得以充分发挥.

问题 1.给定 PRBFN 的合成径基数目为 8,如何提高网络精度.我们分别对两种不同的径基宽度取值方式进行优化,试找出能获取最佳拟合精度的径基宽度(注:为方便讨论,我们把所有径基宽度都取相同值的方法称为“同一法”,而把本文提出的取值方法称为“各异法”).

由于 σ_i^2 反映了径基函数的响应区域,过大或过小的取值都会影响建模效果.为了保证优化结果的合理性(主要是防止过拟合现象发生),我们采取如下方法来设定 σ_i^2 的搜索范围(假定样本已经过标准化处理).首先按式(10)计算所有训练样本间的平均距离 d_{mm}^2 ,随后按式(11)分别求得 σ_i^2 的上下限.

$$d_{mm}^2 = \text{Mean}_{\text{for}(i)} \left\{ \text{Mean}_{\text{for}(j)} \left\{ \|x_i - x_j\|^2 \right\} \right\}, \tag{10}$$

$$\text{Low} = d_{mm}^2 / \rho \quad \text{and} \quad \text{Up} = d_{mm}^2 \cdot \rho. \tag{11}$$

式中, ρ 是比例系数(在本实验中,当 $\rho=20$ 时径基函数的响应区域已经充分小或充分大了).

对于本文提出的“各异法”,可以将各个径基宽度的平均值(M-Width)作为一种表征径基宽度整体状况的方

式.在 GAs 的优化过程中,确保每一代个体所产生的 M-Width 不超越式(11)所定义的上下限.

最后的优化结果列于表 3.从中可见,经过优化“各异法”在 $N=15, \mu=2.6017, \alpha=0.4546$ 时,网络的平均自检相对误差降至最低值 4.96%,明显低于“同一法” $\sigma_i^2=1$ 时的结果.更令人信服的是,“同一法”经过寻优后在 $\sigma_i^2=1.2620$ 时达到最优状态,但它所对应的最低平均自检相对误差 6.96% 仍然高于“各异法”的结果 4.96%.另外与“同一法”最优状态 $\sigma_i^2=1.2620$ 相比,“各异法”最优状态时的最大、最小和平均径基宽度(1.1639,0.5410 和 0.7692)在数量级上并没有很大的差异,从而充分说明了样本径基宽度的各异性对网络建模的益处.

Table 3 The effects of various determination methods for radial basis width

表 3 不同径基宽度取值方法的效果

Width determination method	Average relative error of self-test (%)	Average relative error of prediction (%)
$\sigma_i^2=1$ (based on the same width) $N=15 \mu=2.6017 \alpha=0.4546$	7.13	6.29
(based on the different widths)	4.96	8.36

宽度确定方法, 平均自检相对误差, 平均预测相对误差, 同一法, 各异法.

问题 2.限定平均自检相对误差的上限为 1%,如何使网络结构最简.根据式(8)所定义的适应度函数,最终的优化结果为 $k=13, N=9, \mu=2.1130$ 和 $\alpha=0.1973$,与之对应的平均自检和预测相对误差分别是 0.83%和 1.22%.可见,我们只需 13 个合成径基就能使网络的相对拟合误差小于 1%,而且该相对拟合误差应该是同结构 PRBFN 中最小的.显见,式(8)所定义的适应度函数为优化 PRBFN 的结构提供了一条有效途径.

3.3 实验3

本实验将通过一个时间序列预测实例来进一步验证所提出的径基宽度取值方法的优越性.所采用的时间序列模型是式(12)所示的 May 方程.当方程系数 v 等于 2.9 时,模型呈现出混沌现象¹.实验数据由该方程迭代 100 次所得($v=2.9$,取初值为 0.1).由于模型是一阶滞后预测模型,所以 100 个数据点能构造 99 个样本,我们将前 70 样本用于训练,后 29 个用于检测.

$$p_{\text{new}} = p + v \cdot p \cdot (1 - p). \quad (12)$$

我们同样设计了两个网络构造中的耦合问题.现将结果简述如下.

(1) 给定 PRBFN 的合成径基数目为 2,从表 4 可见“各异法”所能达到的最佳效果明显优于“同一法”的最优效果;(2) 限定平均自检相对误差的上限为 5%,那么只需 6 个合成径基就能满足该要求,并且当 $N=35, \mu=0.0870$ 和 $\alpha=0.3189$ 时,PRBFN 对这组混沌序列数据的平均自检和预测相对误差分别是 3.98%和 2.77%.这些结果再次证实了本文所提出的径基宽度取值方法的优势以及网络结构优化方法的有效性.

Table 4 The effects various determination methods for radial basis width

表 4 不同径基宽度取值方法的效果

Width determination method	Average relative error of self-test	Average relative error of prediction
$\sigma_i^2=0.0067$ (based on the same width) $N=6 \mu=0.0972 \alpha=0.3214$	17.85	15.68
(based on the different widths)	13.08	10.90

宽度确定方法, 平均自检相对误差, 平均预测相对误差, 同一法, 各异法.

综观上述实验可见,实验 1 证实了基于 PLS 的 RBFN 构造策略较之传统策略的优越性,实验 2、实验 3 则证实了本文所提出的径基宽度取值方式在提高网络性能上的效果,并验证了 GAs 在解决网络构造中两类耦合问题上的有效性,从而使 PRBFN 的优势得到保障和扩充.因此,它们的结合将在根本上改善传统 RBFN 的建模效果.

4 结 论

为了克服传统 RBFN 构造策略的弊病,本文首先提出将 PLS 引入 RBFN 的构造策略,并把所构成的网络称为 PRBFN.实验结果证实了 PRBFN 具有以下优点:(1) PLS 合成径基充分反映了训练样本集的信息,使模型表达

能力优于基于 OLA 的 RBFN 构造策略;(2) 克服了基于 K-Means 算法构造策略的随机性.本文还提出了一种更为有效的径向宽度取值方法,并将其与 GAs 结合,解决了网络构造的两个耦合问题,使得 PRBFN 的结构和性能趋于最优.实验结果证实了本文提出的网络构造策略和径向宽度取值方法在提高模型逼近效果上的优越性,也验证了本文提出的网络结构优化方法在 PRBFN 最优结构设计上的有效性.总之,PLS、新的径向宽度取值方法与 GAs 三者相结合为 RBFN 的构造提供了一条新颖且更可靠、更有效的途径.

References:

- [1] Whitehead, B.A., Choate, T.D. Cooperative-Competitive genetic evolution of radial basis function centers and widths for time series prediction. *IEEE Transactions on Neural Networks*, 1996,7(4):869~880.
- [2] Xu, Bing-zheng, Zhang, Bai-ling, Wei, Gang. *Theory and Application of Neural Networks*. Guangzhou: South China University of Technology Press, 1994 (in Chinese).
- [3] Chen, S., Cowan, C.F.N., Grant, P.M. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks*, 1991,2(2):302~309.
- [4] Walczak, B., Massar, D.L. The radial basis functions——partial least squares approach as a flexible non-linear regression technique. *Analytica Chimica Acta*, 1996,331:177~185.
- [5] Jia, F., Martin, E.B., Morris, A.J. Non-Linear principal components analysis for process fault detection. *Computers and Chemical Engineering*, 1998,22 (Suppl.):5851~5854.
- [6] Yang, Tai-ning, Wang, Sheng-de. Robust algorithms for principal component analysis. *Pattern Recognition Letters*, 1999,20:927~933.
- [7] Chen, De-zhao. *Multivariable Data Processing*. Beijing: Chemical Industry Press, 1998 (in Chinese).
- [8] Geladi, Paul. Notes on the history and nature of partial least squares (PLS) modelling. *Journal of Chemometrics*, 1988,2:231~246.
- [9] Goldberg, D.E. *Genetic Algorithms in Searching, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley, 1989. 1~25.
- [10] Michalewicz, Z., Janikow, C.Z., Krawczyk, J.B. A modified genetic algorithm for optimal control problem. *Computers Mathematics Application*, 1992,23(12):83~94.

附中文参考文献:

- [2] 徐秉铮,张百灵,韦岗. *神经网络理论与应用*.广州:华南理工大学出版社,1994.
- [7] 陈德钊. *多元数据处理*.北京:化学工业出版社,1998.

RBFN Structure Determination Strategy Based on PLS and GAs*

ZHAO Wei-xiang, WU Li-de

(Department of Computer Science and Engineering, Fudan University, Shanghai 200433, China)

E-mail: weixiang_zhao@263.net

<http://www.fudan.edu.cn>

Abstract: In view of the drawbacks of the conventional structure determination strategies for RBFN (radial basis function network), a RBFN structure determination strategy based on PLS (partial least squares) and Gas (genetic algorithms), and a more effective determination method for radial basis width are proposed in this paper. In this structure determination strategy, PLS avoids the local optima occurring in the process of calculating radial basis with K-Means algorithm and makes the synthesized radial bases more representative than the radial basis got by orthogonal algorithm. Furthermore the proposed determination method for radial basis width and GAs guarantee the substantial improvement and optimization on the networks performance and structure. Finally, the experiments demonstrate the superiority, the reliability, and the effectiveness of the proposed RBFN structure determination strategy based on PLS and GAs, and the proposed determination method for radial basis width.

Key words: radial basis function network; cluster; orthogonal algorithm; partial least square; genetic algorithm

* Received May 28, 2001; accepted January 10, 2002

Supported by the National Natural Science Foundation of China under Grant No.69935010