小置信范围下指示函数的沃尔什逼近*)

原峰山[1,2]　朱思铭[1]

(中山大学数学与计算科学学院　广州510275)[1]　(广州航海高等专科学校现代化教学中心　广州510725)[2]

摘　要　文章讨论了基于统计学习理论中满足较小置信范围时，采用沃尔什函数逼近机器学习响应函数，从而在结构风险最小化原则条件下使风险泛函最小，保证了小样本环境下机器学习的误差最小，并且获得较好的学习推广能力。

关键词　SRM原则，机器学习，沃尔什逼近

Walsh Approaximation of Indicator Function Under Small Trust Scope

YUAN Feng-Shan[1,2]　ZHU Si-Ming[1]

(School of Mathematics and Computing Science, Zhongshan University, Guangzhou 510275)[1]

(Guangzhou Maritime College, Guangzhou 510725)[2]

Abstract　This paper presents a new solution that risk functional becomes the least, under SRM principle when Walsh function is used to approximate responding function of learnig machine, which also leads error of the learning machine to be the least respecting to less trust scope with the small sample problems, and the better learning generalization ability can be obtained.

Keywords　SRM principle, Machine learning, Walsh approximation

1. 引言及预备知识

1.1　学习过程的SRM

1.1.1　*学习问题的表示*　根据Vapnik的统计学习理论[2]，设有定义在空间Z上的概率测度$F(z)$，存在函数集合$S=\{Q(z,\alpha),\ \alpha\in\Lambda\}$，学习的目标是使风险泛函最小化。即：

$$R(\alpha)=\int L(y,f(x,\alpha))dF(x,y)$$
$$=\int Q(z,\alpha)dF(z)\quad \alpha\in\Lambda \tag{1}$$

其中：$(z_1,\cdots,z_l)$为训练集，l为样本数。由于$F(x,y)$的未知性，(1)式是无法确定的，因此依据经验风险最小化原则(ERM)，通常以经验风险泛函$R_{emp}(\alpha)$去近似$R(\alpha)$，即：

$$R_{emp}(\alpha)=\frac{1}{l}\sum_{i=1}^{l}Q(z_i,\alpha) \tag{2}$$

1.1.2　*结构风险最小化原则及问题的提出*　Vapnik同时指出[2]：在机器学习过程中样本数是有限的，所以在l为有限的小样本情况下，若保证机器学习有良好的推广能力，应满足如下条件：

在损失函数集$S=\{Q(z,\alpha),\alpha\in\Lambda\}$上构造一个结构$S_1\subset S_2\subset S_3\cdots\subset S_n$，从中选一个合适的元素$S_k$及$Q(z,\alpha_l^k)\in S_k$，使：

$$R(\alpha_l^k)\leqslant R_{emp}(\alpha_l^k)+\Phi(\frac{l}{h_k}) \tag{3}$$

其中右端第一项为经验风险，第二项为置信范围。学习的问题就是使(3)式的不等式最小化，有两种方法可实现这一点。

第一种，确定(3)式的VC维h_k，并使置信范围较小，当$R_{emp}(\alpha_l^k)$最小时，则$R(\alpha_l^k)$最小。

第二种，保持$R_{emp}(\alpha)$值固定，采取最小化置信范围的方法使$R_{emp}(\alpha_l^k)$最小。

本文采用的是第一种方法。

2. 指示函数的沃尔什函数逼近

满足上述第一种方法，引入指示函数集合$f(w,x)$，$w\in R^n$，将：

$$R_{emp}(\alpha)=\frac{1}{l}\sum_{i=1}^{l}Q(Z_i,\alpha)=\frac{1}{l}\sum_{i=1}(y(x)-f(w,x))^2 \tag{4}$$

最小化，其中$(x_1,y_1),\cdots,(x_l,y_l)$为训练集，$x_i$为向量，$y_i\in\{1,-1\}$，$i=1,2,\cdots,l$，$f(w,x)\in\{1,-1\}$，$w$为参数向量。将$f(w,x)$表达为沃尔什函数，令：

$$f(w,x)=\prod_{r=0}^{l-1}sgn[\cos(w\cdot 2^r\cdot\pi x)] \tag{5}$$

此式为沃尔什函数的余弦表达[2]，且：$w=\{w_1,w_2,\cdots,w_l\}$。令w_i为非负数，且其二进制表示式为：

$w_i=\sum_{r=0}^{p-1}w_{ir}2^r$，($w_{ir}$为0或1)，$p$为$w_i$的二进制表示时的位数。

$\therefore f(0,x_i)=sgn[\cos(0\cdot\pi x_i)]=1$

$f(1,x_i)=sgn[\cos(1\times 2^0\cdot\pi x_i)]=sgn[\cos(\pi x_i)]$

$f(2,x_i)=sgn[\cos(1\times 2^1\pi x_i)]\cdot sgn[\cos(0\cdot\pi x_i)]=sgn[\cos(2\pi x_i)]$

$f(3,x_i)=sgn[\cos(1\times 2^1\cdot\pi x_i)]\cdot sgn[\cos(1\times 2^0\cdot\pi x_i)]=f(2,x_i)\cdot f(1,x_i)$

同理：

$f(4,x_i)=sgn[\cos(4\pi x_i)]$

$f(5,x_i)=sgn[\cos(1\times 2^2\cdot\pi x_i)]\cdot sgn[\cos(1\times 2^0\cdot\pi x_i)]=f(4,x_i)\cdot f(1,x_i)$

$f(6,x_i)=f(4,x_i)\cdot f(2,x_i)$

*)基金项目：国家自然科学基金资助项目(编号：10071097)。**原峰山**　高级工程师，博士生，研究方向：人工智能与计算机网络。**朱思铭**　教授，博士生导师，研究方向：应用数学，人工智能与计算机网络。

$$f(7,x_i)=f(6,x_i)\cdot f(1,x_i)$$
$$f(8,x_i)=sgn[\cos(8\pi x_i)]$$
$$f(9,x_i)=f(8,x_i)\cdot f(1,x_i)$$
………

根据上述规律，注意到：

$$f(w_i,x_{i+1})=\prod_{r=0}^{l-1}sgn[\cos(w_{ir}\cdot 2^r\cdot\pi(x_i+1))]$$
$$=\prod_{r=0}^{l-1}sgn[\cos(w_{ir}\cdot 2^r\cdot\pi(x_i)]=f(w_i,x_i)$$

所以，(5)式是以1为周期的周期函数，并且随 w_i 的变化函数值呈明显的规律分布。

3. 逼近函数的完备正交及其正交函数集的表达

3.1 逼近函数集的完备正交

引理 若函数集 $F=\{f(n,t),t\in N,n\in M\}$，满足：

$$\sum_{t=0}^{N-1}f(m,t)\cdot f(n,t)=\begin{cases}N,n=m\\0,n\neq m\end{cases}$$

则 F 是完备正交的。

证明：∵ $f(w,x)=\prod_{r=0}^{l-1}sgn[\cos(w\cdot 2^r\cdot\pi x)]$，并且由于连乘积内的符号函数是二值函数，

∴可将 $f(w,x)$ 表达为 $N=2^l$ 个用二进制表达的一组函数的连乘积形式：

$$f(w,x)=F(\omega_{l-1},\omega_{l-2},\cdots,\omega_0;\ x_{l-1},\ x_{l-2},\cdots,x_0)$$
$$=\prod_{r=0}^{l-1}(-1)^{w_{l-1-r}\cdot(x_r+x_{r+1})} \tag{6}$$

式中，ω 和 x 是用二进制码表示的两个变量。现将两个上述函数乘积之和表示为如下的二进制之和：

$$\sum_{x_{l-1}}^{1}\sum_{x_{l-2}}^{1}\cdots\sum_{x_0=0}^{1}F(m_{l-1},m_{l-2},\cdots,m_0;x_{l-1},x_{l-2},\cdots x_0)\times F(n_{l-1},n_{l-2},\cdots;x_{l-1},x_{l-2},\cdots x_0) \tag{7}$$

将(6)式代入，得：

$$\sum_{x_{l-1}}^{1}\sum_{x_{l-2}}^{1}\cdots\sum_{x_0=0}^{1}\prod_{r=0}^{l-1}(-1)^{(n_{l-1-r}+m_{l-1-r})(x_r+x_{r+1})}$$
$$=\prod_{r=0}^{l-1}\sum_{x_r=0}^{1}(-1)^{(n_{l-1-r}+m_{l-1-r})(x_r+x_{r+1})}$$
$$=\prod_{r=0}^{l-1}\{1+(-1)^{(n_{l-1-r}+m_{l-1-r})}\} \tag{8}$$

式中 n_{l-1-r} 和 m_{l-1-r} 只取0或1，∴当 $n_{l-1-r}=m_{l-1-r}$ 时，式(8)为

$$\prod_{r=0}^{l-1}(1+1)=2^l=N$$

当式(8)至少有一个 $n_{l-1-r}\neq m_{l-1-r}$ 时，连乘积至少有一项为0，∴整个乘积为0。因此，若用十进制表示，表达式为：

$$\sum_{x=0}^{N-1}f(m,x)\cdot f(n,x)=\begin{cases}N & n=m\\0 & n\neq m\end{cases} \tag{9}$$

论题得证。因此(5)式是完备正交的。

3.2 沃尔什函数表达的最小经验泛函

由于(5)式是完备正交的，因此可以有如下的表达式：

$$F(\omega,x)=\sum_{\omega=0}^{\infty}a_\omega f(\omega,x) \tag{10}$$

作为近似，

$$y(x)\approx\sum_{\omega=0}^{l-1}W(\omega,x) \tag{11}$$

其中：$a_\omega=\frac{1}{l}\sum_{i=0}^{l-1}x_i\cdot W_{kl}$，$x_i$ 为样本，W_{kl} 为 l 个样本的沃尔什函数表达的最小经验泛函，即：

$$R_{emp}(\alpha)=\frac{1}{l}\sum_{i=1}^{l}[y(x)-F(\omega,x)]^2$$
$$=\frac{1}{l}\sum_{i=1}^{l}[y(x)-\sum_{\omega=0}^{\infty}a_\omega W(\omega,x)]^2 \tag{12}$$

结论 (1)利用沃尔什函数逼近满足经验风险泛函的函数 $y(x)$，由于沃尔什函数的正交特征，当样本数 $l\to\infty$ 时可以完全逼近所描述的函数。

(2)在有限样本数 l 个内，可以通过沃尔什函数变换表达的函数近似逼近 $y(x)$。

(3)由于沃尔什函数的特性，进行有关运算时可以简化运算，如可将相乘转为模2加法，利用对称关系进行变量序数量移转。

(4)将沃尔什函数表达应用于SRM原则意义上的小样本学习，由于其函数的特征及良好的完备正交性质，可以较好地成为描述指示函数的应用函数。可以在置信范围一定的前提下使得风险泛函最小。

参考文献

1 郑君理，应启珩，杨为理．信号与系统[M]．高等教育出版社，2000

2 Vapnik V N. The Nature of Statistical Learning Theory [M]. Springer-Verlag New York, Inc. 1995

3 常迥译．沃尔什函数及其应用[M]．科学出版社，1980

4 Vapnik V N. Statistical Learning Theory [M]. New York: John Wiley and Sons, Inc. 1998

5 梅建新，等．支持向量机在小样本识别中的作用[J]．武汉大学学报(理学版)，2002(12)