

核自适应滤波

Kernel Adaptive Filtering

【美】 Weifeng Liu José C. Príncipe

【加】 Simon Haykin 著

田亚娟 张峰 肖忠祥 程国建 译

WILEY |



國防工業出版社
National Defense Industry Press

核自适应滤波

Kernel Adaptive Filtering

[美] Weifeng Liu José C. Príncipe

[加] Simon Haykin 著

田亚娟 张峰 肖忠祥 程国建 译

国防工业出版社

·北京·

著作权合同登记 图字：军-2014-006 号

图书在版编目 (CIP) 数据

核自适应滤波/(美)刘伟峰,(美)普林西佩(Principe, J. C.),
(加)赫金(Haykin, S.)著;田亚娟等译. —北京:国防工业
出版社, 2014.9

书名原文: Kernel adaptive filtering

ISBN 978-7-118-09597-5

I. ①核… II. ①刘… ②普… ③赫… ④田… III. ①自适
应滤波 IV. ①TN713

中国版本图书馆 CIP 数据核字 (2014) 第 207954 号

Copyright © 2010 by John Wiley & Sons, Inc. All rights reserved.

All rights reserved. This translation published under license. Authorized translation from the English language edition entitled Kernel Adaptive Filtering, ISBN 978-0-470-44753-6, by Weifeng Liu, José C. Principe and Simon Haykin, Published by John Wiley & Sons. No part of this book may be reproduced in any form without the written permission of the original copyrights holder.

Copies of this book sold without a Wiley sticker on the cover are unauthorized and illegal.

本书简体中文版由 John Wiley & Sons, Inc. 授权国防工业出版社独家出版。
版权所有, 侵权必究。

※

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

国防工业出版社印刷厂印刷

新华书店经售

*

开本 880×1230 1/32 印张 7¼ 字数 198 千字

2014 年 9 月第 1 版第 1 次印刷 印数 1—2000 册 定价 68.00 元

(本书如有印装错误, 我社负责调换)

国防书店: (010) 88540777

发行邮购: (010) 88540776

发行传真: (010) 88540755

发行业务: (010) 88540717

前 言

本书首次对核自适应滤波器作了一个全面和统一的介绍。自适应信号处理理论基于三大支柱，分别是线性模型、均方代价 (Mean Square Cost) 和自适应最小均方学习算法 (Adaptive Least-square Learning Algorithm)。当必须采用非线性模型时，线性自适应滤波器的简洁性优点就会丧失，设计人员就只能采取函数逼近 (Function Approximation)、神经网络、局部最小、正则化以及其他措施。这是解决非线性问题的唯一道路吗？也许还有其他方法，这正是本书的关注重点。最基本的思路是在一个线性空间实现自适应滤波器，而这个线性空间与原始的输入空间是非线性相关的。如果这个思路可行，那么所有三大支柱以及关于线性模型的知识就仍然可以使用，最终就可以在输入空间实现非线性滤波器。

本书首先通过一个被称为再生核的正定函数推导出再生核希尔伯特空间 (Reproduction Kernel Hilbert Spaces, RKHS) 理论，该理论可以将输入数据非线性变换到高维特征空间。如果在 RKHS 空间将要进行的滤波和自适应操作可以表示为投影样本的内积，那么在输入空间这些操作就可以直接通过核评价公式 (Kernel Evaluations) 来计算。基于此方法，我们将进一步介绍 RKHS 空间的自适应滤波算法家族：

- 核最小均方算法；
- 核仿射投影算法；
- 核递归最小均方算法；
- 扩展核递归最小均方算法。

这些核学习算法在两个重要领域——自适应滤波和神经网络之间构建了一个紧密桥梁，并且将误差修正学习和记忆学习两大重要方法学完美地集于一身。将 RKHS 方法应用于非线性滤波设计的瓶颈在

于如何正则化、选择合适的核函数以及如何对滤波器生长过程进行削减。本书将采用严格的数学推导形式，指出这些问题，同时给出具体的解决方法，并且通过具体实例展示核自适应滤波方法的优越性能。

第1章首先介绍机器学习、线性自适应滤波和传统非线性方法的基本概念；接着介绍核自适应滤波的数学基础——再生核希尔伯特空间理论。本章着重强调核自适应滤波器是一个通用的函数逼近器，在自适应过程中无局部最小值以及仅需要合理的计算资源。

第2章研究核自适应滤波家族中最简单的核最小均方算法。本章将逐步介绍其算法实现，并对实际中如何选择核函数、步进值大小、稀疏化以及正则化方法做深入研究，最后，给出了两个计算机实验：一个是麦克-格拉斯（Mackey-Glass）混沌时间序列预测；另一个是非线性通道均衡。

第3章介绍核仿射投影算法（Kernel Affine Projection Algorithm, KAPA），具体包括四个类似的算法；对滤波和自适应的数学方程以及有用的实现技术作了一个完整的推导和讨论。很多著名算法可以作为特例由KAPA推出。本章给出的三个详细应用展示了KAPA算法应用广泛、设计灵活的特点。

第4章介绍核递归最小均方算法和高斯过程回归理论。本章将讨论稀疏化方法——近似线性相关；借助于贝叶斯解释，本章介绍了一个强大的模型选择方法，即“最大边际似然法”；最后，通过两个计算机实验研究不同的稀疏方法的性能以及“最大边际似然法”用于确定核参数的有效性。

第5章在核递归最小均方算法基础上讨论扩展核递归最小均方算法。我们系统地研究了在RKHS空间通过一个有噪观测序列估计线性动态系统的状态，给出了几个重要定理及其证明过程以简述其重要意义及基本方法。本章给出了两个实例，即瑞利信道跟踪和洛伦兹（Lorenz）时间序列建模。

第6章介绍核自适应滤波器的主要瓶颈。我们介绍了一个被称为“惊奇”（Surprise）的主观信息测度，并给出了一个统一的稀疏方法来有效地削减核自适应滤波器的生长；最后，介绍了三个有趣的计算机仿真以说明这些理论。

本书适合对自适应滤波、神经网络和核方法感兴趣的工程师、计算机科学家以及研究生等使用。精心设计的 12 个计算机实验（表 1）贯穿于全书各章节，以加强读者对于书中概念的理解。实验源程序可从网址 <http://www.cnel.ufl.edu/~weifeng/publication.htm> 直接下载。为保证代码的可读性，我们在程序设计和实现过程中相对于性能更注重程序的简洁性。除提供程序外，对其应用不做任何保证。

表 1 书中所有计算机实验列表

计算机实验	名 称
2.1	KLMS 算法用于麦克-格拉斯时间序列预测
2.2	KLMS 算法用于非线性通道均衡
3.1	KAPA 算法用于麦克-格拉斯时间序列预测
3.2	KAPA 算法用于噪声消除
3.3	KAPA 算法用于非线性通道均衡
4.1	KRLS 算法用于麦克-格拉斯时间序列预测
4.2	最大边际似然法用于模型选择
5.1	EX-KRLS 用于瑞利信道跟踪
5.2	EX-KRLS 用于洛伦兹时间序列建模
6.1	“惊奇” SC 准则用于非线性回归
6.2	“惊奇” SC 准则用于麦克-格拉斯时间序列预测
6.3	“惊奇” SC 准则用于 CO ₂ 浓度预报

在本书中，我们一直努力反映这个新兴领域的所有最新进展。每一章节都包括研究现状和潜在的未来研究方向。本书可为拟采用非线性自适应滤波方法解决实际问题以及在相关领域寻找启发性研究灵感的读者提供有益指导。

本书获得了西安石油大学优秀著作出版基金资助出版，在此表示感谢。

致 谢

我们首先要感谢希捷科技有限责任公司 Puskal P. Pokharel 博士，佛罗里达大学的 Murali Rao 博士、Jay Gopalakrishnan 博士以及 Il Park 在核自适应滤波理论研究中给予的帮助。非常感激奥尔巴尼医学院和 Wadsworth 中心的 Aysegul Gunduz 博士、佛罗里达大学的 John Harris 博士以及西班牙坎塔布里亚大学的 Steven Van Vaerenbergh 博士对本书手稿早期版本许多有用的建议以及建设性的回复。

许多人士认真地阅读了本书的部分章节，他们分别是（按字母排序）：

佛罗里达州盖恩斯维尔市佛罗里达大学的 Erion Hasanbelliu

加利福尼亚州圣克拉拉市英特尔公司的 Kyu-hwa Jeong 博士

加利福尼亚州圣迭戈市加州大学的 Li Ruijiang 博士

犹他州盐湖城市犹他大学的 Antonio Paiva 博士

佛罗里达州盖恩斯维尔市佛罗里达大学的 Alexander Singh

香港科技大学的 Wang Yiwen 博士

伊利诺伊州芝加哥市芝加哥大学的 Xu Jianwu 博士

我们还想感谢(按字母排序)：加州大学伯克利分校的 Peter Bartlett 博士；日本 Riken 脑科学研究所的 Andrzej Cichocki 博士；谷歌实验室的 Corinna Cortes 博士；英国纽卡斯尔大学的 Graham C. Goodwin 博士；加州大学伯克利分校的 Michael Jordan 博士；斯坦福大学的 Thomas Kailath 博士；美国兰德公司的 Joel S. Kvitky 博士；纽约大学的 Yann LeCun 博士；伊利诺伊大学芝加哥分校的 Liu Derong 博士；英国剑桥大学的 David J. C. MacKay 博士；麻省理工学院的 Tomaso Poggio 博士；加州大学洛杉矶分校的 Ali H. Sayed 博士；德国马普生物控制所的 Bernhard Schölkopf 博士；希腊雅典大学的 Sergios

Theodoridis 博士；谷歌实验室的 Yoram Singer 博士；雅虎研究的 Alexander J. Smola 博士；比利时鲁汶大学的 Johan Suykens 博士；美国国家科学基金会的 Paul Werbos 博士；斯坦福大学的 Bernard Widrow 博士和英国爱丁堡大学的 Chris Williams 博士。

另外还需要感谢 Wiley 出版社的工作人员：出版者 George Telecki、助理编辑 Lucy Hitz、制作编辑 Kris Parrish 以及 Toppan Best-Set Premedia 公司的项目经理 Stephanie Sakson，感谢他们在手稿准备阶段、封面选择以及印刷出版等过程中的全力支持和背后辛勤的工作。

最后，特别要感谢麦克马斯特大学的 Lola Brooks 为本书手稿中的许多章节所做的文字输入工作。

标 记 法

本书讨论了许多算法，包括了大量的数学公式。为了能够清晰地表述核自适应滤波理论的基本思想，有必要采用一种方便、统一的标记法。为了方便读者，我们认为有必要在本书的开始部分形成一个标记法说明，以总结和解释本书的标记规则。

书中主要有三类变量需要特别区分：标量、向量和矩阵变量。

下面是本书中用到的标记规则：

(1) 使用小写斜体表示标量变量。例如，滤波器的输出是一个标量变量，记作 y 。

(2) 使用大写斜体表示标量常数。例如，滤波器的阶是一个标量常数，记作 L 。

(3) 使用小写黑斜体表示向量。

(4) 使用大写黑斜体表示矩阵。

(5) 使用括号表示任意变量（标量、向量或矩阵）的时间关系。例如， $d(i)$ 表示标量 d 在时间（或迭代步） i 的值。 $\mathbf{u}(i)$ 表示向量 \mathbf{u} 在时间（或迭代步） i 的值。类似地， $\mathbf{G}(i)$ 表示矩阵 \mathbf{G} 在时间（或迭代步） i 的值。全书均遵循上述规则。 f_i 表示在时间（或迭代步） i 的输入-输出映射估计值，括号保留给输入变量，例如 $f_i(\mathbf{u})$ 。

(6) 使用上标 T 表示转置。例如，如果

$$\mathbf{d} = \begin{bmatrix} d(1) \\ d(2) \\ \vdots \\ d(N) \end{bmatrix}$$

那么

$$\mathbf{d}^\top = [d(1), d(2), \dots, d(N)]$$

(7) 本书中所讨论的所有变量均为实值，复数不在本书讨论之内。

(8) 本书中的所有向量如不声明均为列向量。

(9) 下标的使用主要有以下两种情况：①表示向量或矩阵的分量；②表示一个普通向量，它的索引与时间（或迭代步）无关。例如， \mathbf{c}_i 根据上下文可能为一组向量中的第 i 个向量或向量 \mathbf{c} 的第 i 个分量。

为方便读者，我们尽力在全书中保证符号一致统一。表 2 为典型的符号表示。

表 2 符号表示

需表示的量	描述	实例
标量	小写斜体	d
向量	小写黑斜体	$\mathbf{w}, \boldsymbol{\omega}, \mathbf{c}_i$
矩阵	大写黑斜体	$\mathbf{U}, \boldsymbol{\Phi}$
时间或迭代次数	括号中的索引	$u(i), d(i)$
向量或矩阵的分量	下标索引	$a_j, \mathbf{G}_{i,j}$
线性空间	大写 <code>mathbb</code> 字体	\mathbb{F}, \mathbb{H}
标量常数	大写斜体	L, N

缩写和符号

为了方便读者更准确和更方便地进一步深入研究，下面将书中主要的术语、符号以及简写等一并按中英对照形式列出。

$(\cdot)^T$	vector or matrix transposition	向量或矩阵转置
A^{-1}	inverse of matrix A	逆矩阵
$E[\cdot]$	expected value of a random variable	随机变量的期望值
$m(\cdot)$	the mean of a random variable	随机变量的平均值
$\sigma^2(\cdot)$	the variance of a random variable	随机变量的方差
$\langle \cdot, \cdot \rangle$	inner product	内积
$\ \cdot\ $	norm of a vector; square root of the inner product with itself	向量的范数; 与自身内积的平方根
$ \cdot $	absolute value of a real number or determinant of a matrix	实数的绝对值或矩阵的行列式
\propto	proportional to	正比于
\sim	distributed according to	以……分布
∇	gradient	梯度
$\mathbf{0}$	zero vector or matrix	零向量或零矩阵
β	forgetting factor	遗忘因子
$C(i)$	dictionary or center set at iteration i	第 i 次迭代时的字典或中心集
$d(i)$	desired output at time or iteration i (a real scalar)	i 时刻或次迭代的期望输出
$\text{diag}\{a, b\}$	a diagonal matrix with diagonal entries a and b	对角单元为 a 和 b 的对角矩阵

δ_1	distance threshold in novelty criterion	新奇准则中的距离阈值
δ_2	prediction error threshold in novelty criterion	新奇准则中的预测误差阈值
δ_3	threshold in approximate linear dependency test	近似线性相关测试的阈值
δ_{ij}	Kronecker delta	Kronecker 增量
$\Delta \mathbf{w}(i)$	weight adjustment at time or iteration i (a column vector in an Euclidean space)	i 时刻或次迭代的权重调整
\mathcal{D}	data set	数据集
$e(i)$	output estimation error at time or iteration i	i 时刻或次迭代的输出估计
\mathbb{F}	feature space induced by the kernel mapping	核映射诱发的特征空间
\mathbf{G}	Gram matrix of (transformed) input data	变换后的输入数据组成的 Gram 矩阵
\mathbb{H}	reproducing kernel Hilbert space	再生核希尔伯特空间
\mathbf{I}	identity matrix	单位矩阵
$J(i)$	error cost at time or iteration i	i 时刻或次迭代的误差代价
$\mathbf{k}(i)$	Kalman gain (or gain vector) at time or iteration i	i 时刻或次迭代的卡尔曼增益 (或增益向量)
$K(\mathbf{A})$	condition number of a matrix \mathbf{A}	矩阵 \mathbf{A} 的条件数
$\kappa(\mathbf{u}, \mathbf{u}')$	kernel (or covariance) function evaluated at \mathbf{u} and \mathbf{u}'	对 \mathbf{u} 和 \mathbf{u}' 的核函数评价 (方差)
L	dimensionality of the input space	输入空间的维度
λ	regularization parameter	正则化参数
M	dimensionality of the feature space	特征空间的维度
\mathcal{M}	misadjustment of the least-mean-square algorithm	最

	小均方算法的失调量
$\mathbf{n}(i)$	additive noise in the state space at time or iteration i 状态空间在 i 时刻或次迭代的加性噪声
N	number of training data 训练数据的个数
η	step-size parameter 步进大小参数
$O(\cdot)$	of the order of a number 数量级
\mathbf{P}	state-error correlation matrix 状态-误差相关矩阵
$\varphi(\cdot)$	a mapping induced by a reproducing kernel 再生核诱发的映射
$\varphi(i)$	transformed filter input at time or iteration i (a column vector in a feature space) 在 i 时刻或次迭代的变换后的滤波输入 (特征空间的列向量)
Φ	transformed input data matrix 变换后的输入数据矩阵
\mathbf{R}	covariance matrix of (transformed) input data 输入数据 (变换后) 的方差矩阵
\mathbb{R}	the set of real numbers 实数集
\mathbb{R}^L	L -dimensional real Euclidean space L 维实数欧几里得空间
ζ_{\max}	the maximum eigenvalue 最大特征值
$\text{tr}(\mathbf{A})$	trace of matrix \mathbf{A} 矩阵 \mathbf{A} 的迹
T_1	abnormality threshold in surprise criterion 惊奇准则的异常阈值
T_2	redundancy threshold in surprise criterion 惊奇准则的冗余阈值
$\mathbf{u}(i)$	filter input at time or iteration i (a column vector in an Euclidean space) 在 i 时刻或次迭代的滤波输入 (欧几里得空间的列向量)
\mathbb{U}	input domain 输入域
\mathbf{U}	input data matrix 输入数据矩阵
$\mathbf{v}(i)$	additive noise in the output at time or iteration i 在 i 时

	刻或次迭代的输出中的加性噪声
$w(i)$	weight estimate at time or iteration i (a column vector in an Euclidean space) 在 i 时刻或次迭代的权重估计 (欧几里得空间的列向量)
$\omega(i)$	weight estimate at time or iteration i (a column vector in a feature space) 在 i 时刻或次迭代的权重估计 (特征空间的列向量)
z^{-1}	unit delay operator 单位延迟算子
AIC	Akaike information criterion Akaike 信息准则
ALD	approximate linear dependency 近似线性相关
APA	affine projection algorithm 仿射投影算法
BIC	Bayesian information criterion 贝叶斯信息准则
CC	coherence criterion 一致性准则
CV	cross-validation 交叉验证
ENC	enhanced novelty criterion 增强型新奇准则
EX-RLS	extended recursive least squares algorithm 扩展的递归最小二乘算法
EX-KRLS	extended kernel recursive least squares algorithm 扩展的核递归最小二乘算法
GPR	Gaussian process regression 高斯过程回归
LMS	least-mean-square algorithm 最小均方算法
LOOCV	leave-one-out cross-validation 留一交叉验证
LS	least squares 最小均方
MAP	maximum a posterior 最大后验
MDL	minimum description length 最小描述长度
MSE	mean square error 均方误差
MML	maximum marginal likelihood 最大边际似然
NC	novelty criterion 新奇准则
NLMS	normalized least-mean-square algorithm 标准最小均方算法

KA	kernel ADALINE 核 ADALINE 算法
KAPA	kernel affine projection algorithm 核仿射投影算法
KLMS	kernel least-mean-square algorithm 核最小均方算法
KRLS	kernel recursive least-squares algorithm 核递归最小二乘算法
PCA	principal components analysis 主成分分析
PDF	probability density function 概率密度函数
RAN	resource allocating network 资源配置网络
RBF	radial-basis function 径向基函数
RKHS	reproducing kernel Hilbert space 再生核希尔伯特空间
RLS	recursive least-squares algorithm 递归最小二乘算法
RN	regularization network 正则化网络
RNN	recurrent neural network 并发神经网络
SC	surprise criterion 惊奇准则
SNR	signal-to-noise ratio 信噪比
SVD	singular value decomposition 奇异值分解
SVM	support vector machine 支持向量机
SW-KRLS	sliding window kernel recursive least-squares algorithm 滑动窗核递归最小二乘算法

目 录

第 1 章 背景与展望	1
1.1 监督、序贯和主动学习	1
1.2 线性自适应滤波器	3
1.2.1 最小均方算法	4
1.2.2 递归最小二乘算法	5
1.2.3 扩展递归最小二乘算法	7
1.3 非线性自适应滤波器	10
1.4 再生核希尔伯特空间	11
1.5 核自适应滤波器	15
1.6 总结	20
注释	21
第 2 章 核最小均方算法	27
2.1 最小均方算法	28
2.1.1 LMS 算法的收敛性	29
2.1.2 LMS 算法的失调	30
2.1.3 学习曲线	30
2.2 核最小均方算法	31
2.3 核及参数选择	34
2.4 步进参数	37
2.5 新奇准则	38
2.6 KLMS 算法的自正则化特性	40
2.6.1 解范式约束	40
2.6.2 奇异值分析	42
2.6.3 单位下三角线性系统	47

2.7	泄漏核最小均方算法	48
2.8	标准核最小均方算法	48
2.9	核 ADALINE 算法	49
2.10	资源配置网络	54
2.11	计算机实验	55
2.11.1	KLMS 算法应用于 Mackey-Glass 时间序列预测	55
2.11.2	KLMS 应用于非线性信道均衡	63
2.12	结论	65
	注释	66
第 3 章	核仿射投影算法	70
3.1	仿射投影算法	70
3.2	核仿射投影算法	74
3.2.1	KAPA-1 (简单 KAPA)	74
3.2.2	KAPA-2 (标准化 KAPA)	77
3.2.3	KAPA-3 (泄漏 KAPA)	77
3.2.4	KAPA-4 (基于牛顿递归的泄漏 KAPA)	78
3.3	误差重用	79
3.4	滑动窗口 Gram 矩阵逆	80
3.5	相关算法之间的关系	80
3.5.1	KLMS 算法	80
3.5.2	NORMA 算法	81
3.5.3	核 ADALINE 算法	81
3.5.4	滑动窗核递归最小二乘算法	82
3.5.5	正则化网络	82
3.6	计算机实验	83
3.6.1	KAPA 应用于 M-G 时间序列预测	83
3.6.2	KAPA 应用于消除噪声	85
3.6.3	KAPA 应用于非线性信道均衡	89
3.7	结论	92
	注释	94