分类号 密级

U D C 编号

中 南 大 学

CENTRAL SOUTH UNIVERSITY

硕 士 学 位 论 文

论 文 题 目 基于SVM的氧化铝高压溶出苛性比值与溶出率软测量

学科、专业 控制科学与工程

研究生姓名 李 学 军

导师姓名及专业技术职称 桂 卫 华 教 授

部队导师姓名及军衔 王 维 大 校

MS THESIS

Soft Sensing for Ratio of Soda to Alumina and Leaching Rate in High Pressure Digestion process of alumina based on SVM

Specialty： Control Science and Engineering

Master Degree Candidate： Li Xuejun

Supervisor： Prof. Gui Weihua

Army supervisor: Senior Colonel . Wang wei

College of Information Science & Engineering

Central South University

ChangSha Hunan P.R.

摘 要

苛性比值与溶出率是氧化铝高压溶出过程中两个重要的经济技术指标。它们不仅决定了氧化铝溶出的效果及碱耗，而且对氧化铝的后续生产具有极大的影响。目前没有任何仪表能够实时直接测量这两个值，而只能通过化学分析获得，因此存在很大的时间延迟，严重影响了高压溶出过程的优化控制。为此，研究高压溶出过程中苛性比值与溶出率的软测量预测模型进行实时在线检测，具有十分重要的意义。本论文采用支持向量机（SVM）方法对苛性比值与溶出率进行预测。

首先，论文阐述了高压溶出的机理过程，对高压溶出过程的化学反应以及影响苛性比值与溶出率的主要因素进行了详尽的分析。

其次，对统计学与支持向量机原理进行了阐述。介绍了统计学理论的三个核心概念：VC 维、推广能力的界和结构风险最小化；阐述了支持向量机回归与核函数的方法和理论；比较了支持向量机与神经网络两种方法的优缺点以及最小二乘支持向量机（LS-SVM）与标准支持向量机（S-SVM）的不同。为建立基于特征提取的 LS-SVM 苛性比值与溶出率软测量模型奠定了基础。

最后，采用最小二乘支持向量机为苛性比值与溶出率的预测过程建立模型。详细探讨了数据预处理的方法，深入分析了模型参数对预测性能的影响，在此基础上，建立了基于特征提取的自适应参数 LS-SVM 苛性比值与溶出率软测量模型，利用 Matlab 平台进行仿真，取得了很好的效果。

关键词：苛性比值，溶出率，高压溶出，软测量，支持向量机

ABSTRACT

Ratio of Soda Alumina (RSA) and Leaching Rate(LR) are two very important economical and technical indices in the process of High-Pressure Digestion(HPD) of alumina. Not only they affect output of alumina and alkali consumption, but also they make great influence on successive production of alumina. However, until now there is no any instrument can be used to measure RSA and LR directly and immediately. RSA and LR can only be measured through chemical analysis, so large lag exists and optimal control for HPD process is influenced severely. So it is very significant to implement online measurement of RSA and LR by establishing prediction model using soft sensing technology. This paper applies Support Vector Machine(SVM) method to forecast RSA and LR.

Firstly, the mechanism process of high pressure digestion is described, then the primary chemical reaction in this process and the main factors affecting the ratio of soda to alumina and leaching rate are analyzed particularly.

Secondly, the theory of SVM and Statistical Learning Theory(STL) are represented. Three main concepts of STL including VC dimension, the bound of extending and structural risk minimization are introduced. The method and theory of support vector regression and kernel function are elaborated, then the advantage and disadvantage of support vector machine and neural network are compared, the differences of least square support vector machine(LS-SVM) and standard support vector(S-SVM) machine are compared too, establishing foundation for proposing the soft sensing model of RSA and LR based on LS-SVM with characters selected.

Finally, the prediction model of the RSA and LR are established by adopting LS-SVM. The data pre-processing method is described and the reason how the model parameters influence the prediction performance is analyzed, then the soft sensor model of the RSA and LR is set up based on LS-SVM with characters selected and parameters adopting. The simulation is realized through Matlab and the better consequence is obtained.

KEY WORDS: Ratio of Soda to Alumina (RSA), Leaching Rate(LR), High Pressure Digestion(HPD), Soft Sensing, Support Vector Machine(SVM)

目 录

第一章 绪论

金属铝的优良性质促使氧化铝生产在全世界得到了迅速的发展。在氧化铝生产工艺流程中，无论是烧结法、联合法以及混联法，高压溶出过程都对从铝土矿中提取氧化铝起着重要作用[1]。溶出过程中氧化铝溶出的好坏主要体现在苛性比值的高低与溶出率的大小。因此，严格控制高压溶出过程中的苛性比值与溶出率，尽可能使其达到最优就成了氧化铝生产企业亟待解决的问题[2]。

1.1 课题的来源、背景及研究意义

1.1.1 课题的来源、背景

随着全球经济迅猛发展，能源供应和需求之间的矛盾日渐突出，节约能源已成为全人类的主题，是每个国家发展经济的一项长远的战略方针[3]。研究工业生产过程，特别是高能耗行业的节能降耗对保证我国工业可持续发展具有十分重要的意义。

有色金属作为我国国民经济和国防军工发展的重要基础原料和战略物资，在国民经济发展中占有十分重要的地位[4]。有色金属工业是我国的高能耗行业，与工业发达国家相比，能源消耗方面存在着较大的差距[5]，主要冶炼工艺能耗比国际发达国家至少高出 20%以上，最高达到 142%，能源费用占到了生产成本的30-40%，导致我国有色冶金企业的生产成本远远高于工业发达国家，严重影响了我国有色冶金企业的国际竞争力，制约了我国有色工业的发展。

有色金属主要有 64 种元素，其中铝的产量几乎占有色金属总产量的一半。铝及其合金因特殊的物理化学特性，已成为需求量大、应用面广的基础材料，而且铝工业生产能耗高（吨铝综合能耗是铜的两倍多，是钢铁的十倍多，电解铝能耗已占全国总能耗的 2.6%，与钢铁、水泥成为国家宏观调控的行业，其能量综合利用率要比发达国家低 15%左右），工艺机理复杂、生产流程长、工序多，是具有代表性的典型有色冶金过程。为此，以铝生产过程为主要研究对象，研究面向节能降耗的过程控制的若干理论与方法，必将为其它有色冶金过程提供借鉴。

本课题系国家自然科学基金项目：面向节能降耗的有色冶金过程控制若干理论与方法研究（课题编号 60634020）的子课题“关键参数的软计算理论与方法”。

1.1.2 课题研究意义

高压溶出过程中苛性比值和溶出率是否稳定，对氧化铝生产有很大的影响。

苛性比值如果过大就会导致产量下降，严重影响生产正常进行；同样，苛性比值如果过低就会使指标下降，影响产品的质量。氧化铝溶出率是直接反映产率的生产指标，溶出液苛性比值和溶出率由企业根据氧化铝生产技术水平和经济效益决定，企业依据此指标指导配料流程中矿石配碱量。保持适当的苛性比值和一定溶出率，对整个氧化铝工业的稳定高产具有重要作用。为了有效地进行生产操作和控制，需要对溶出生产过程的苛性比值和溶出率进行在线实时控制，以便为生产操作提供依据。

目前，国内外氧化铝生产中，苛性比值与溶出率的检测基本上都是人工完成的，整个检测过程非常繁琐。化验室操作人员每隔两个小时从原矿浆及溶出矿浆中取样，对其进行液固分离，然后在化验室用传统的方法分别对液相和固相成分进行化学分析，根据检测结果以及苛性比值与溶出率的定义，计算出实际的苛性比值与溶出率。由于溶出矿浆是在高压溶出过程结束后取样的，因此苛性比值与溶出率的测量必须等到高压溶出过程结束后才能进行，别一方面，整个高压溶出过程大约持续一个半到两个小时，而液相分析需要两个小时，固相分析更是需要八个小时，因此苛性比值与溶出率的检测远远滞后于实际生产。

利用这些滞后的数据来指导生产，存在着较大的偏差，对氧化铝生产造成了十分不利的影响。因此，研究经济适用而且不需要大量增加工人劳动强度的软测量方法，是实现苛性比值与溶出率在线检测的最有效手段，也是节能降耗和自动化生产的一项重要任务。

1.2 软测量及支持向量机概述

1.2.1 软测量概述

在许多生产过程与生产设备中，有一些重要的参数需要实时检测和控制。但由于各种不同的原因，这些参数无法通过传感器进行直接实时检测。为此，人们提出了软测量技术的概念。软测量就是选择与被测参数有关的一组可测参数，构造某种以可测参数为输入、被测参数为输出的数学模型，用计算机软件进行计算从而得到被测参数的估计值，这种技术称为软测量技术或“软仪表”，它为许多难测对象和难测参数的测量提供了一种新方法。

（一）软测量的主要内容

软测量技术主要包括 4 个方面的内容，即辅助变量的选择、数据处理、软测量模型的建立和测量模型的修正。其基本结构如图 1-1 所示。

1. 辅助变量的选择

辅助变量的选择是建立软测量模型的基础，它对于实现软测量尤其重要。辅

助变量的选择包括变量类型、变量数量和检测点的选择。

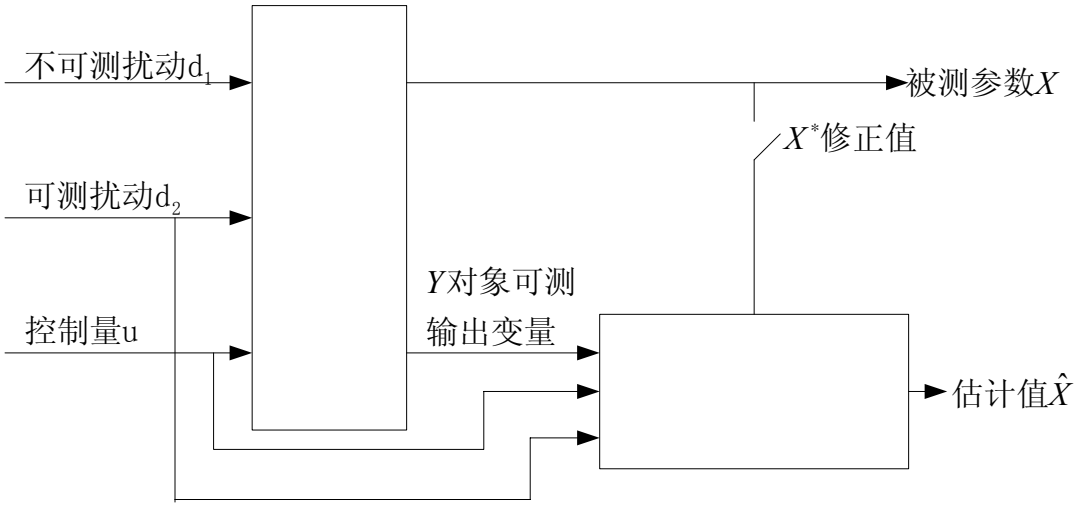


图 1-1 软测量系统的基本结构

（1）变量的选择原则

辅助变量的选择一般应从以下几个方面来考虑[6-8]：

a.灵敏性，即对对象输出能做出快速反应；

b.特异性，即对对象输出之外的干扰不敏感；

c.适用性，即可直接测量并具有一定的测量精度；

d.精确性，即构成的数学模型能满足精度要求；

e.鲁棒性，即构成的数学模型对误差不敏感。

（2）辅助变量个数的选择

辅助变量可选数量的下限是被估计的变量数，而最佳数目则与对象的自由度、测量噪声以及模型的不确定性有关。

（3）辅助变量检测点位置的选择

辅助变量检测点位置一般应选择最能代表对象特性的位置，检测点位置的不同将影响数学模型的结构和参数。因此，当模型确定后，检测点位置就不能随意变动。

2. 测量数据的处理

输入的辅助变量参数常因自身特点或外部干扰而引起误差，不能直接用于软测量。因此，输入的数据必须预先进行必要的处理。

（1）误差处理

直接采集的辅助变量一般都有测量误差，其中包括系统误差、随机误差和粗大误差，对于随机误差，由于它符合统计规律，因此，可以采用数字滤波算法进行处理，可收到较好的效果。对于粗大误差也有许多处理方法，如统计假设检验

法、广义似然比法和贝叶斯法等。

（2）数据变换

选择的各种辅助变量往往具有不同的工程单位，变量之间在数值上可能相差几个数量级。直接使用这些数据进行计算可能会影响数学模型的计算精度。因此，需要对有关的数据包括标度、转换和权函数进行必要的变换[9]。对于在数值上相差几个数量级的测量数据可利用合适的因子进行标度，以改善算法的精度和稳定性。采取直接转换或寻找新变量代替原变量进行转换，可有效地降低原对象的非线性特性(如进行对数转换)。权函数可实现对变量动态特性的补偿。

3. 软测量模型的建立

软测量模型注重的是通过辅助变量来获得对主导变量的最佳估计，本质上是要完成辅助变量构成的可测信息集到主导变量估计的映射。软测量模型的好坏，将直接影响到软测量的精度。

从图 1-1 可以看出，软测量模型是软测量技术的核心，通过软测量模型，可以由可测辅助参数计算出被测参数 X 的最优估计值 $\hat{X}$：

$$\hat{X} = f(d_2, y, u, X^*, t) \tag{1-1}$$

软测量建模的方法有很多，如机理建模方法、辨识建模方法、回归分析法、现代时间序列法、基于神经网络、基于支持向量机的建模方法等，实际中要根据现场的实际情况，选用合适的建模方法。

4. 软测量模型的在线修正

由于工业对象并非一成不变的，在长期运行中，对象特征会因操作条件变化、生产原料改变、装置改造等原因而发生变化，从而导致软仪表的精度下降。因此在软测量技术的应用中，必须能对软测量模型进行在线校正，以使软仪表能跟踪系统的变化，提高模型的精度和适应性。

通常对软测量模型的校正是根据离线测量(或长周期采样)数据 X^* 来修正模型参数。最简便的方法为常数项修正法，即取软测量模型为：

$$\hat{X} = f(d_2, u, y) + \Delta x$$
$$\Delta x = \beta(x^* - \hat{x}) \tag{1-2}$$

式中：β 为自适应因子，用以调节模型自校正的强度。

(二) 软测量的研究现状

软测量作为一个概括性的科学术语是在上世纪 80 年代中后期提出来的。由于工业过程的实际需要，软测量技术迎来了一个发展的黄金时期。1992 年，过程控制专家 Mcacvoy[10] 提出了 7 项技术前沿，将软测量技术列在首位。软测量已成为过程控制领域的研究热点和主要发展趋势之一。目前人们对软测量技术的研究主要集中在两个方面：一是建模理论和方法的研究；二是软测量技术在过程

控制中的应用研究[11]。

传统工业过程建模方法是机理建模和基于系统辨识的建模方法。在实际实用中，机理建模很大程度依赖于工程研发人员对工艺机理的了解程度，由于实际过程的复杂性和不确定性，建立严格的机理模型十分困难，甚至不可能。系统辨识的方法用于非线性工业过程建模的精度也不理想，不能满足实际生产的需要。随着人工智能的出现，人们在软测量建模理论和方法上取得了突破性的进展。作为人工智能领域最活跃的神经网络、模糊理论、专家系统等技术，越来越多地被用在软测量的建模方面，如文献[12][13][14]分别运用了神经网络、模糊理论和专家系统的方法进行建模，仿真结果表明了算法的有效。不论是传统模型、智能模型都有各自的优点，但同时也各有不足之处。比如，机理模型要求对工艺机理有很透彻的了解；神经网络对训练样本集的质量要求很高；模糊模型学习能力较差；专家系统获取知识困难等等。随着现代智能技术的发展，基于数据的机器学习问题越来越显示出其重要性，现有机器学习方法共同的重要理论是统计学。统计学习理论和支持向量机(SVM)正在成为继神经网络等方法之后新的研究热点，已经成为小样本估计和预测学习的最佳理论。SVM 在过程控制领域主要用来建立软测量模型，大量的研究结果表明，基于 SVM 技术的软测量模型在精度、泛化能力和样本数据依赖上都要好于用传统方法建立的软测量模型[15]。

经过近 20 年的发展，软测量技术不仅在建模理论研究上取得了辉煌的成绩，在实际应用中也获得了巨大的成功，并开发出了大量实用的软测量软件包，为社会和企业带来了巨大的经济效益。国外一些公司以商品化的软件形式推出了各自的软测量仪表。国内有关高等院校、科研院所和企业也自行开发了不少软测量技术并应用于实际生产中，取得了较好的效果，为企业创造了显著的经济效益。

1.2.2 支持向量机概述

支持向量机（Support Vector Machines 或 SVM）是在统计学习理论基础上发展起来的一种新的通用学习方法。统计学习理论建立在一套坚实的理论基础之上，为解决有限样本学习问题提供了一个统一的框架，它能将很多现有的方法纳入其中，有望帮助解决许多原来难以解决的问题，比如神经网络结构选择问题、局部极小点问题等。 1963 年，V.Vapnik 等人提出支持向量方法原型；1971 年，V.Vapnik 和 A.Cherbonenkis 在“The Necessary and Sufficient Conditions for the Uniforms Convergence of Averages to Expected Value”一文中，提出了 SVM 的重要理论基础-VC 维理论，Kimeldorf 则用 SV 的核空间解决非线性问题；1982 年，在“Estimations of Dependences Based on Empirical Data”一书中，V.Vapnik 进一步提出了具有划时代意义的结构风险最小化原理，堪称 SVM 算法的基石；1992

年，Boser, Guyon and Vapnik 在“A Training Algorithm for Optimal Margin Classifiers”一书中，提出了最优边界分类器；1993 年，Cortes 和 Vapnik 在“The Soft Margin Classifier”一书中，进一步探讨了非线性最优边界的分类问题；1995 年，Vapnik 在“The nature of Statistical Learing Theory”一书中完整地提出了 SVM 分类，系统描述了统计学习理论；1997 年 V.Vapnik、A.Smola 等在文献[16]中，详细介绍了基于 SVM 方法的回归算法和信号处理方法；2000 年，Scholkoph 和 Smola 提出了 V-SVM 方法[17]；2002 年，基于正则化思想在标准支持向量机的基础上提出了最小二乘支持向量机[18]。另外许多学者在研究一些新型的支持向量机，如加权支持向量机[19]及模糊、粗糙集支持向量机[20]等。

支持向量机与传统机器学习理论最大的不同在于它服从结构风险最小化原理而非经验风险最小化原理。关于 SVM 的深入研究近几年才开始，支持向量机是机器学习领域若干标准技术的集大成者。它集成了最大间隔超平面、Mercer 核、凸二次规划、稀疏解和松弛变量等多项技术，在若干挑战性的应用中，获得了目前为止最好的性能。美国科学杂志上，支持向量机以及核学习方法被认为是“机器学习领域非常流行的方法和成功例子，并是一个十分令人瞩目的发展方向”。它初步表现出很多优于已有方法的性能，一些学者认为，SVM 正在成为继神经网络研究之后新的研究热点，并将有力地推动机器学习理论和技术的发展。

由于支持向量机具有完备的统计学习理论基础和出色的学习性能，目前 SVM 的研究在国内外正处在热潮。V.Vapnik 对统计学习理论及 SVM 的研究进行了开拓性的工作，B.Scholkopf, J.Shawe-Taylor, D.A.Mcallester, R.Herbrich 等人在 SVM 及相关理论研究方面作了大量的工作[21-23]，T.Joachims, J.C.platt, S.S.Keerthi 等人在 SVM 的实现算法方面作了大量的研究[24-25]，国内研究也取得了长足的进展，清华大学在 SVM 的研究取得了众多的成果，台湾的林智仁博士在 SVM 实现算法的收敛性研究方面作了大量工作[26-27]。清华大学张学工博士将 V.Vapnik 的经典著作“The Nature of Statistical Learning Theory”译成中文，并在 2000 年出版，这极大地推动了国内统计学习理论及 SVM 的研究。

SVM理论研究取得了很大进展，其应用研究也正在兴起。随着贝尔实验室应用SVM对美国邮政手写数字库识别研究[28]，并取得了较大成功后，有关SVM的应用研究得到了各国研究者的广泛重视。目前，在人脸检测、语音识别、文字/手写体识别、图像处理及其他应用研究等方面取得了大量的成果，在精度上已经超过传统的学习算法或与之不相上下。

（1）在模式识别中的应用

Osuna[29]最早将SVM应用于人脸检测领域，并取得了较好的效果。马勇[30]等人提出了分别由一个线性和非线性SVM组成的层次型SVM分类器，不仅具有

较高的判别精度，而且具有较快的训练速度。文献[31]和[32]分别提出了扩展的SVM算法来处理人脸检测中的复杂问题。贝尔实验室通过对美国手写数字库进行的实验结果：专门针对该特定问题设计的五层神经网络错误率为5.1%，决策树方法识别错误率是16.2%，然而选择多项式、RBF和Sigmlid核函数的SVM算法的错误率分别为4.0%、4.1%和4.2%。另外，在图像处理方面SVM也取得了良好的效果[33−34]。

（2）在回归分析中的应用

1998年，Smola[35−36]等人在SVM基础上提出了支持向量回归机（SVR），并研究了时间序列预测问题，大大地促进了SVR的应用研究。文献[37]、[38]和[39]将交通系统与SVR的应用相结合，分别对交通流进行了建模、预测等研究。在社会经济领域，[40]和[41]分别研究了经济时间序列、商业交易市场的时间序列预测等问题。

由于SVM表现出了良好的性能优势，逐渐地被应用于信息处理的各个分支中，如信号处理[42]等。

1.3 论文的研究内容及结构

本论文基于复杂工业过程软测量技术的研究现状，针对氧化铝高压溶出过程中苛性比值与溶出率难以在线检测的问题，以工程应用为背景，着眼点在铝工业，面向整个有色冶金过程，其方法和结论也将为其它工业对象提供一定的参考价值。课题研究的内容包括：

（1）氧化铝高压溶出过程工艺和机理的研究；

（2）统计学和 SVM 理论的研究；

（3）输入数据预处理方法的研究；

（4）SVM 参数选择的研究；

（5）基于 SVM 软测量模型的建立。

根据课题的研究内容，全文共分为五章，其主要内容和章节安排如下：

第一章，绪论。介绍了课题的来源及研究意义，综合阐述了软测量及支持向量机的基本理论、研究现状，并就本文的研究内容和结构作了介绍。

第二章，氧化铝高压溶出机理分析。介绍了氧化铝高压溶出的工艺，对高压溶出过程的主要化学反应以及影响苛性比值与溶出率的主要因素进行了详尽的分析，为实现苛性比值与溶出率的在线预测打下了基础。

第三章，统计学与 SVM 理论研究。简要介绍了统计学理论的基本原理、重要概念，以及核函数方法与实施步骤；重点阐述了 SVM 回归的方法与理论；详细讨论了 SVM 与神经网络两种方法的优缺点，比较了最小二乘支持向量机与标

准支持向量机的不同。

第四章，苛性比值与溶出率软测量模型的建立。提出了系统软测量模型的整体框架，重点介绍了数据预处理的方法，详细探讨了 SVM 参数的选择，最后建立了基于自适应参数 LS-SVM 的苛性比值与溶出率软测量模型，并进行了仿真。

第五章，结论与展望。总结全文，提出下一步研究方向。

第二章 氧化铝高压溶出过程机理分析

高压溶出是一个极其复杂的生产工序，整个生产过程中，同时发生着各种各样的物理化学反应。为了能够建立合理的软测量模型，实现对苛性比值与溶出率的在线检测，首先必须对生产工艺有深刻的认识。

2.1 氧化铝高压溶出工艺概述

工业生产上高压溶出不是用纯的苛性钠溶液，而是用生产流程中的循环母液。循环母液主要成分是苛性钠和铝酸钠溶液，此外是碳酸钠、硫酸钠，以及少量铝硅酸盐等；铝土矿主要成分是氧化铝，另外还含有不少有害杂质，主要是氧化硅、氧化钛、氧化铁、碳酸盐、有机物及硫化物等。

高压溶出流程如图 2-1 所示[2]，生产流程主要包括以下工序：

（1）预脱硅：高压溶出过程中，铝土矿中的二氧化硅是主要危害杂质。一方面，二氧化硅以铝硅酸钠的形式与氧化铝和苛性钠反应进入赤泥，造成氧化铝和苛性钠的损失；另一方面，铝硅酸钠往往在溶出管道中形成结疤，降低溶出过程的热交换效率，甚至堵塞溶出管道，引起机组故障。所谓预脱硅过程，就是在原矿浆进入溶出管道前，使矿石中大部分二氧化硅预先反应并析出，从而降低硅结疤对机组正常运行的危害。具体操作方法是：配料车间配置好的原矿浆进入原矿浆槽进行搅拌，使在 90～100℃的条件下，停留 4～6 小时，以消除溶液中水合铝硅酸钠过饱和，既是铝土矿预脱硅。

（2）预热：经过预脱硅的原矿浆，利用油隔泵打入若干个串联双程预热器进行预热（预热至 160～170℃）。在溶出车间，对原矿浆中预热是利用溶出过程后期余热实现的。原矿浆在通过溶出器高温溶出后进入蒸发器进行冷却。冷却过程中料浆释放的乏汽被用作预热器的热源。预热过程是通过热交换器完成。其中，原矿浆吸收乏汽中热量获得温升，乏汽则变成高压冷凝水进入冷凝水自蒸发器，然后送往后续工序作为洗涤用水。利用乏汽对原矿浆进行预热，可节约大量新蒸汽，减少能耗。另外，在应用蒸汽直接加热原矿浆的高压溶出过程中，压煮器内通入的新蒸汽在加热矿浆的同时，变成水分稀释矿浆中苛性钠浓度，影响溶出速度；而预热过程却是通过热交换器间接加热，不存在这个缺陷。因此，预热温度的高低，对氧化铝高压溶出过程具有重要意义，在很大程度上影响到溶出质量的好坏。由于要在不影响流程工业连续运行前提下定期清理预热器管道结疤并进行故障检修，两个高压溶出机组均备有备用预热器。

（3）加热：预热后的原矿浆顺序进入两个加热溶出器（$1^{\#}$、$2^{\#}$）通以高压

新蒸汽（32Kg），直接加热矿浆到溶出温度（250～260℃）。

（4）溶出反应：经过加热溶出器后，原矿浆的温度已经加热到理想的溶出温度。原矿浆逐个流经 7 个反应溶出器（3#～9#），在高温高压条件下，原矿浆中的铝酸钠溶液将矿石中的氧化铝溶解出来。

（5）冷却：溶出以后的矿浆从 9# 溶出器顺序排入若干个串联自蒸发器进行冷却。从最后一个自蒸发器卸出的溶出矿浆通过溶出矿浆缓冲器送入稀释槽稀释并沉降。

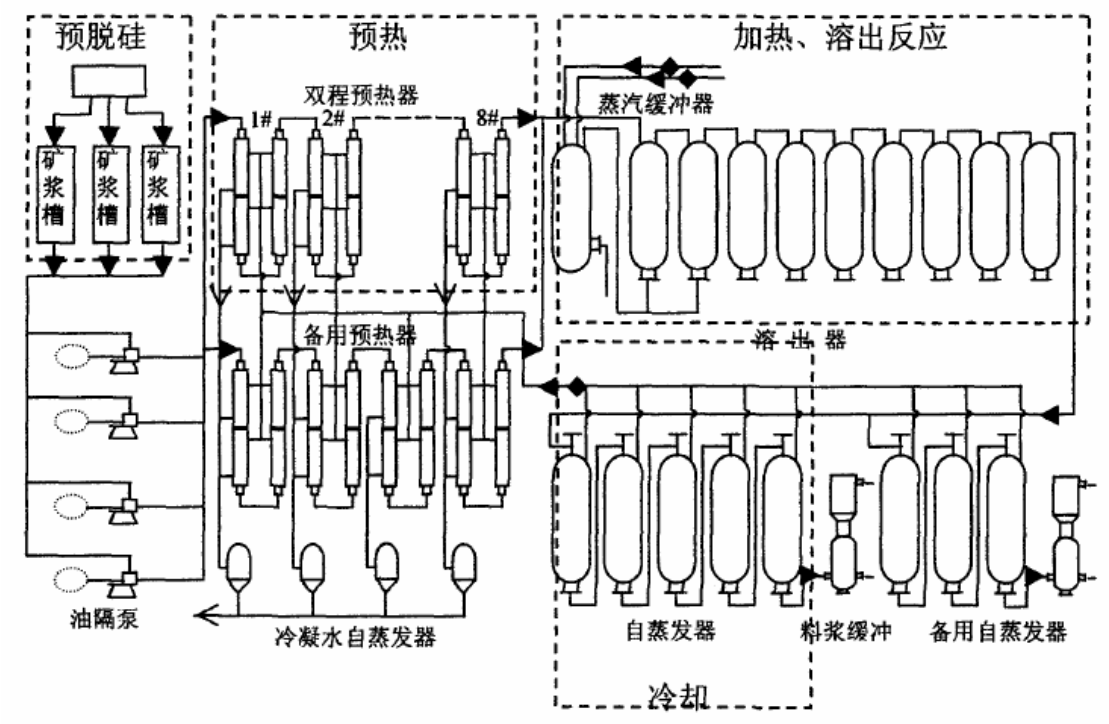


图 2-1 高压溶出流程图

2.2 苛性比值与溶出率

在氧化铝高压溶出工序中，溶出液的苛性比值以及铝土矿中氧化铝的溶出率是两个主要的衡量溶出效果的经济技术指标。它们的定义如下：

（1）苛性比值

苛性比值指的是溶出液中苛性氧化钠分子数与氧化铝分子数的比值，它实际反映了高压溶出过程中苛性钠的消耗量，其定义如下：

$$\alpha_k = 1.645 \times N_k / A \tag{2-1}$$

式中，1.645—— Al_2O_3 和 Na_2O 的分子量比值，即 102/62；

N_k ——溶出液中的苛性氧化钠浓度，（ 克/升）；

A ——溶出液中氧化铝浓度，（ 克/升）。

（2）溶出率

指的是铝土矿中氧化铝在高压溶出过程中被循环母液中苛性钠溶出的比例。溶出率按下式计算

$$\eta=\left(1-\frac{(A/S)_{赤泥}}{(A/S)_{矿石}}\right)\times 100\% \tag{2-2}$$

式中，η ——氧化铝的溶出率，（%）；

$(A/S)_{赤泥}$——赤泥中的氧化铝和氧化硅的重量比（铝硅比）；

$(A/S)_{矿石}$——铝土矿中的氧化铝和氧化硅的重量比（铝硅比）；

生产中，一般用原矿浆固相中的铝硅比和溶出率固相中的铝硅比来表示氧化铝的溶出率，即

$$\eta=\left(1-\frac{(A/S)_{溶出液}}{(A/S)_{原矿浆}}\right)\times 100\% \tag{2-3}$$

式中，$(A/S)_{溶出液}$——溶出液固相氧化铝和氧化硅的重量比（铝硅比）；

$(A/S)_{原矿浆}$——原矿浆中固相氧化铝和氧化硅的重量比（铝硅比）。

2.3 氧化铝高压溶出过程中的化学反应

溶出过程的化学反应十分复杂。一般来说，溶出过程的化学反应可分为两大类：氧化铝水合物的溶出反应（主反应），各种杂质在溶出过程中的化学反应（副反应）。虽然原矿浆中各种杂质对苛性比值及溶出率都会有影响，但是其中一些杂质含量极少，影响也很小。在正常情况下，与一些主要的杂质相比，它们对这两个指标影响可以忽略。理论分析及实践都表明，高压溶出过程中最主要的反应包括三方面[43]：

（1）溶出反应：

我国铝土矿主要成分是一水硬铝，其中每个氧化铝分子只含有一个分子的结晶水。分子式是a-ALOOH或写成$\alpha-Al_2O_3\cdot H_2O$。氧化铝的溶出，就是用苛性钠溶液把铝土矿中的氧化铝溶出来。其中，发生的主要的化学反应就是一水硬铝与苛性钠（NaOH）反应生成铝酸钠，其反应方程式为

$$ALOOH+NaOH+H_2O=NaAlO(OH)_2+H_2O$$

铝酸钠在一定的苛性钠浓度和温度下都可以在苛性钠水溶液中稳定存在，形成铝酸钠溶液。铝土矿的溶出过程可以分为下列几个步骤：

①循环母液湿润矿粒的表面；

②OH^-与氧化铝水合物的反应；

③形成$NaAl(OH)_4$或$NaAlO(OH)_2$的扩散层；

④ $AL(OH)_4^-$ 或 $ALO(OH)_2^-$ 从扩散层扩散出来，而 OH^- 则从溶液中扩散到固相接触面上。

对于铝土矿溶出来说，第二个步骤（化学反应）和第四个步骤（扩散）在一定条件下起主导作用。

（2）氧化硅的化学反应：

氧化硅在铝土矿中含量仅次于氧化铝，是主要的杂质。在高压溶出中，氧化硅发生的化学反应主要包括两个方面：

① 氧化硅与碱液的反应：氧化硅与 NaOH 溶液反应生成易溶于水的硅酸钠

$$SiO_2+2NaOH=Na_2SiO_3+H_2O$$

②生产硅渣的反应：硅酸钠与铝酸钠溶液反应，生成溶解度极小的铝硅酸钠沉淀进入赤泥，同时引起一定数量的碱和氧化铝的损失。另外，在高压溶出过程中，一般都会加入大量的石灰，这时铝土矿中的氧化硅除了生成钠硅渣（铝硅酸钠的水合物）外，还将生成钙硅渣（铝硅酸钙的水合物），其化学反应的方程式为

$$2Na_2SiO_3+2NaAlO(OH)_2+H_2O = Na_2O\cdot Al_2O_3\cdot 2SiO_2\cdot nH_2O+4NaOH+H_2O$$

$$Na_2O\cdot Al_2O_3\cdot 1.7SiO_2\cdot nH_2O + Ca(OH)_2 = CaO\ \cdot AlO_3\cdot 1.7SiO_2\cdot nH_2O + 2NaOH$$

（3）氧化钛与碱液及氧化钙的反应：

氧化钛是铝土矿中另外一种重要的杂质，对苛性比值与溶出率影响非常大。高压溶出铝土矿时，如果没有添加石灰，TiO_2 与苛性钠反应，生成不溶性的钛酸钠（$Na_2O\cdot 3TiO_2\cdot 2.5H_2O$）；当原矿浆中有足够的石灰时，则不生成钛酸钠，而与 CaO 反应生成不溶性的钛酸钙（$2CaO\cdot TiO_2\cdot 2H_2O$），其化学方程式为：

$$3TiO_2+2NaOH+H_2O=Na_2O\cdot 3TiO_2\cdot 2.5H_2O+H_2O$$

$$2CaO+TiO+2H_2O=2CaO\cdot TiO_2\cdot 2H_2O$$

2.4 溶出过程机理分析及参数影响

动力学分析：当溶出时间一定，氧化铝的溶出速度越快，其溶出率越高，设备产能也越大，其他个项也可以得到一定的改善。所以说提高氧化率溶出速度是提高溶出过程重要因素。

铝土矿的溶出与通常的液固相反应过程一样是多相反应过程，包括以下过程：

①含 Al_2O_3 矿物的表面被溶剂——含大量 NaOH 的循环母液所湿润；

②含 Al_2O_3 矿物与氢氧根离子相互作用生成 $NaAlO_2$；

③铝酸根离子通过在矿物表面生成的扩散层扩散到整个溶液当中，而氢氧

根离子通过扩散层扩散到矿物表面上去，使反应进行下去。

其中最慢的步骤决定整个溶出过程的反应速度。一水硬铝石溶出过程的研究表明，在高温下用高碱溶液溶出时，溶出过程属于一级反应，反应速度由扩散速度所决定，而有较低温度和碱度下，为二级反应，这是由于化学反应速度减慢所引起的。

铝土矿中 Al_2O_3 的溶出速度除了受到温度及苛性碱度等主要因素的影响之外，它还受到流动类型，相间界面状况以及矿石特征等因素影响。由于铝土矿组成、晶型结构和溶出条件的复杂性，所以目前还比较难建立完整的系统的分析溶出过程的理论模型。

参数影响：铝土矿高压溶出过程受以下因素影响，温度、苛性比值及循环母液碱浓度、搅拌强度和石灰添加量等。其中苛性比值及循环母液碱浓度是主要因素。

在一定温度下，循环母液碱浓度越高，苛性比值越高，溶液未饱和程度越大，氧化铝溶出速度和设备产能越大，溶出率也得到提高。但是过分地提高母液碱浓度和苛性比值又会为后续过程带来危害，主要是减低溶出率。

2.5 影响苛性比值与溶出率的因素分析

要建立苛性比值与溶出率的软测量模型，首先必须找出影响它们的主要因素。氧化铝高压溶出过程是一个非常复杂的物理化学反应过程，影响苛性比值和溶出率的因素非常多，包括原矿浆的各化学成分及溶出过程中的各个工艺参数。下面，对一些主要参数影响进行分析。

（1）原矿浆中的铝硅比：铝硅比指的是 Al_2O_3 对 SiO_2 的重量比。通过对高压溶出过程的分析知道铝土矿中氧化铝只有极少数溶解于碱溶液，绝大部分以铝土矿硅酸钠和铝硅酸钠钙沉淀进入赤泥。因此，如果原矿浆中的氧化硅含量越高（也就是铝硅比越小），则带入赤泥中的氧化铝也就越多，因此，溶出率也就越小，溶出液苛性比值也就越大。

（2）原矿浆中氧化钙和氧化钛含量：氧化钛也是铝土矿中一种主要的杂质，在高压溶出过程中，氧化钛成凝胶状把一水氧化铝包裹着，氧化钛与碱液作用生成坚实致密的钛酸钠薄膜，使 OH^- 不能穿透，因而大大增加了扩散阻力，减少苛性比值与溶出率。如果在原矿浆中加入足够的石灰，氧化钛就会优先与氧化钙反应，生成 $2CaO\cdot TiO_2\cdot 2H_2O$，这种化合物松脆多孔，$OH^-$ 及 $AlO(OH)_2^-$ 离子容易自由穿透，而且 $2CaO\cdot TiO_2\cdot 2H_2O$ 薄膜在搅拌时由于粒子互相磨擦极易脱落，因而扩散阻力在为减小，接触表面增加，使溶出速度大大提高。别外，氧化钙还会与氧化硅在铝酸钠溶液中反应生成铝硅酸钙沉淀，这样可以减少碱随同赤泥的

化学损失，这从另一方面影响了苛性比值。

（3）循环母液的苛性碱浓度及苛性比值：当其它外部条件均相同时，在一定浓度范围内，氧化铝的溶出率随着循环母液碱浓度的升高而迅速上升，与之对应的溶出液苛性比值则显著降低。

（4）原矿浆中固含及液固比：原矿浆的固体含量及液体与固体的质量比值直接反映了矿浆中原矿石的含量，并从一个侧面给出了矿浆的浓度。矿浆的浓度则会对氧化铝的溶出速度产生影响，从而影响溶出率及苛性比值。

（5）原矿浆中铝土矿的磨细程度：高压溶出过程中，溶出速度与接触表面积成正比。对单位重量的矿石来说，颗粒愈小，表面积就愈大，所以矿石磨得越细，溶出速度就愈快。磨细还可以把杂质包围起来的氧化铝水合物表面暴露出来，而能与碱液接触。矿石经过研磨，裂缝增多，有利于溶出加速。铝土矿的磨细程度是用三种不同尺寸筛子的残留值来表示的。

（6）溶出温度：在一定的苛性钠浓度下提高溶出温度，可以提高氧化铝的溶解度，加快溶出速度，获得苛性比值较低的铝酸钠溶液。溶解温度与溶出率及溶出液苛性比值之间的关系如表 2-1 和表 2-2 所示。

表 2-1 溶出温度对溶出率的影响

溶出温度（℃）	238	240	242	244
溶出率（%）	81.3	84	85.8	86.9

表 2-2 溶出温度对苛性比值的影响

溶出温度（℃）	260	290	315
苛性比值	1.4	1.3	1.25

（7）溶出压力及压差：在高压溶出过程中，经常会出现这样的问题：①在循环母液碱浓度一定，溶出温度随着溶出器的压力的波动而波动；②在一定操作压力下，溶出温度随着循环母液苛性碱浓度的波动而变化，碱浓度升高，溶出温度也有所提高；③在一定操作条件下，温度升高到一定程度时，溶出器便开始震动；④在一定操作压力下溶出温度有时会显著下降。归纳起来就是溶出温度和溶出压力存在着一定的内在联系。因此，溶出器压力会影响溶出率和苛性比值。另外，不同溶出器之间的压差也会影响溶出过程，这是因为，压差会影响矿浆在溶出器中流动的速度，也就是影响溶出器的满罐率和整个溶出时间。如果压差过大，

会使得溶出器的满罐率过小，溶出时间过短，这样会降低溶出率并使得溶出率的苛性比值提高。

（8）原矿浆的流量：根据高压溶出过程中的物料平衡，进入溶出器原矿浆量应该等于出来的溶出矿浆量。如果原矿浆流量变化，则其在溶出器中的反应时间也会发生变化，也就会影响苛性比值和溶出率。

2.6 小结

本章主要阐述了氧化铝高压溶出的机理过程，分析了影响苛性比值与溶出率的主要因素及其之间的关系，为软测量模型辅助变量的选择奠定了基础。

第三章 统计学习理论与支持向量机

基于数据的机器学习问题是现代智能技术的重要方面，研究从观测数据（样本）出发寻找规律，利用这些规律对未来数据或无法观测的数据进行预测，包括模式识别、神经网络等在内，现有机器学习方法共同的重要理论是统计学。统计学习理论和支持向量机正在成为继神经网络之后新的研究热点，是目前小样本估计和预测学习的最佳理论，并将有力地推动机器学习理论和技术的发展。

3.1 统计学习理论

传统统计学研究的是样本数目趋于无穷大时的渐近理论，现有的学习方法也多是基于此假设。但在实际问题中，样本数往往是有限的，因此一些理论上很优秀的学习方法实际中表现却不尽人意。与传统统计学相比，统计学习理论（Statistical Learning Theory 或 SLT）是一种专门研究小样本情况下机器学习规律的理论。该理论针对小样本统计问题建立了一套新的理论体系，在这种体系下的统计推理规则不仅考虑了对渐近性能的要求，而且追求在现在有限的条件下得到的最优结果。Vapnik 等人从六、七十年代开始致力于此方面研究，到九十年代中期，随着其理论的不断发展和成熟，也由于神经网络等方法在理论上缺乏实质性进展，统计学习理论开始受到越来越广泛的重视。

统计学习理论是建立在一套较坚实的理论基础之上的，其主要内容包括以下四个方面[44]：

（1）关于统计推断一致性的充分必要条件的一系列概念；

（2）在这些概念基础上的反映学习机器推广能力的界；

（3）在这些界的基础上建立的针对小样本数的归纳推理原则；

（4）实现这些准则的实际方法（算法）。

这里的学习一致性就是指当训练样本数目趋于无穷大时，经验风险的最优值能够收敛到真实风险的最优值。以上这四条内容中，核心内容是：VC 维，推广能力的界，结构风险最小化。

3.1.1 VC 维

统计学习理论的一个核心概念就是 VC 维（Vapnik-Chervonerkis Dimension）概念，它是描述函数集或学习机器的复杂或者说是学习能力的一个重要指标，在些概念基础上发展出了一系列关于统计学习的一致性、收敛速度、泛化性能等的重要结论。

模式识别的分类方法中 VC 维的直观定义是：对于一个指示函数集，如果存在 l 个样本能够被函数集中的函数按所有可能的 2^l 种形式分开，则称函数集能够把 l 个样本打散，则函数集的 VC 维就是它能打散的最大样本数目 l。若对任意数目的样本都有函数能将它们打散，则函数集的 VC 维可以通过用一定的阈值将它转化为指示函数来定义。

VC 维反映了函数集的学习能力，VC 维越大则学习机器越复杂，因为 VC 维越大则容量越大。但在实际应用中，VC 维往往是难以计算的，甚至是无穷的（需要无限多的训练样本），目前还没有通用的关于任意函数集 VC 维计算的理论，只是对一些特殊的函数集知道其 VC 维，比如在 n 维实数空间中线性分类器和线性实函数的 VC 维是 $n+1$，对于一些较复杂的学习机器如神经网络，其 VC 维除了与函数集（神经网络结构）有关外，还受学习算法等的影响，其确定更加困难。那么对于给定的学习函数集，如何用理论或实验的方法来计算其 VC 维是当前统计学习理论中还有待研究的问题之一[45]。

3.1.2 推广能力的界

统计学习理论系统的研究了对于各种类型的函数集，经验风险和实际风险之间的关系即推广能力的界[44]。对指示函数集中所有函数，包括经验风险最小的函数，经验风险 $R_{\text{emp}}(w)$ 和实际风险 $R(w)$ 之间以至少 1-η 的概率满足如下关系[46]：

$$R(w) \le R_{\text{emp}}(w) + \sqrt{\frac{h(\ln(2l/h)+1) - \ln(\eta/4)}{l}} \tag{3-1}$$

其中 h 是函数集的 VC 维。l 是样本数。这一结论从理论上说明了学习机器的实际风险是由两部分组成的，一个是经验风险，另一个是置信范围。置信范围与学习机器的 VC 维以及训练样本数有关。它表明，在有限训练样本情况下，学习机器不仅要使经验风险最小，同时还要使 VC 尽量小，是为了达到缩小置信范围的目的，这样会使得对未来样本有较好的推广性。推广性的界是对于最坏情况的结论，所给出的界在很多情况下是较松的，文献[46]指出：当 h/l >0.37 时这个界肯定是松弛的，当 VC 维是无穷大时这个界就不再成立；图 3-1 表明：在 VC 维满足不是无穷大这个条件的同时，VC 维的值 h 与样本数目 l 之比大于 0.37，并且其比值单调上升时，VC 置信范围是单调变化的。而且，这种界只在对同一类学习函数进行比较时有效，可以指导我们从函数集中选择最优的函数，在不同函数集之间比较却不一定成立。

寻找更好地反映学习机器能力的参数和得到更紧的界是学习理论今后的研究方向之一[44]。

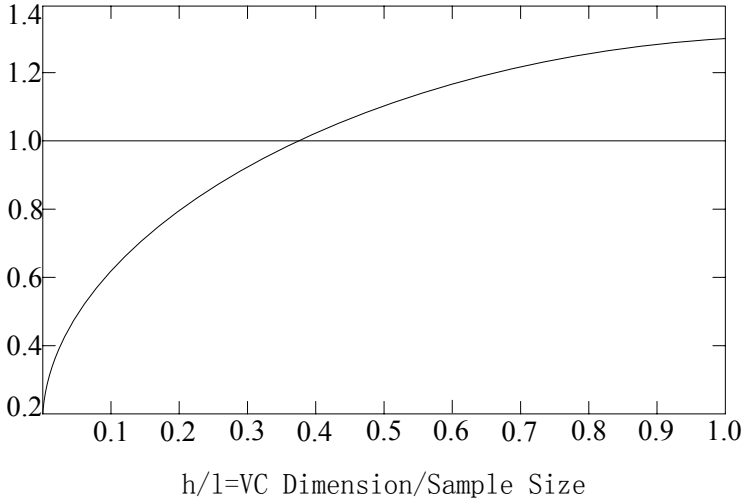


图 3-1 VC 置信范围与 h/l 的关系

3.1.3 经验风险最小化和结构风险最小化

（1）经验风险最小化（ERM）

学习问题被看作是利用有限数量的观测样本来寻找待求依赖关系的问题。给定的训练样本集合{（x_i， y_i）}，$i=1,2,\cdots$ $x_i \in R^n$，其中 x_i 是输入训练样本，已知变量 y 与 x 之间存在一定的未知依赖关系，即存在一个未知的联合分布 $F(x,y)$，我们的学习就是根据这 l 个独立同分布观测样本（x_1， y_1），…，（x_l，y_l），寻求一个最优函数 $f(x)$，使预测的期望风险最小化。

$$R(f)=\int L(x,y,f(x))dF(x,y) \tag{3-2}$$

其中 $L(x,y,f(x))$ 为由于用 $f(x)$ 对 y 进行预测而造成的损失函数。

学习的目标在于使期望风险最小化，虽然知道存在一相确定的概率分布 $F(x,y)$，但并不知道这个分布是什么，所以实际上期望风险无法计算，因此传统的学习方法中采用了经验最小化准则[45]

$$R_{\text{emp}}(f)=\frac{1}{l}\sum_{i=1}^{l}L(x_i,y_i,f(x_i)) \tag{3-3}$$

然而，用经验风险准则代替风险最小化并没有经过充分的理论论证，只是直观上假定当 $l\to\infty$ 时，经验风险趋近于期望风险，可很多问题中的样本数目并不满足这一条件。

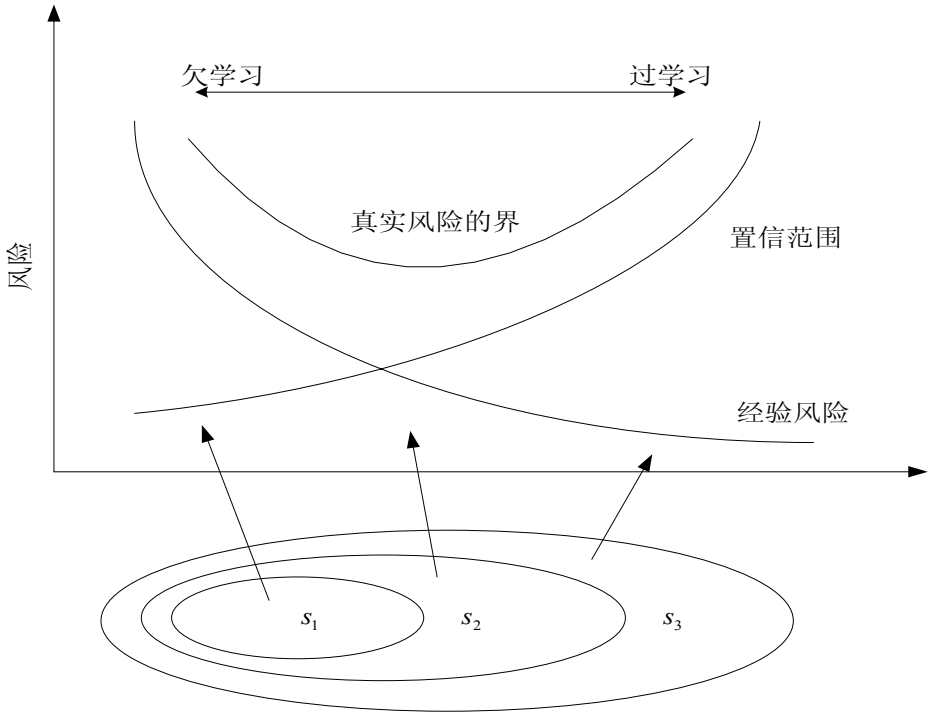
神经网络学习算法遵循的是经验风险最小化原则，开始很多注意力都集中在如何使 $R_{emp}(f)$ 变得更小，后来发现，训练误差小并不是总能得到好的预测效果，在某些情况下，训练误差过小反而会导致推广能力的下降，这就是“过学习问题”。

产生过学习的原因是试图用一个十分复杂的模型去拟合有限样本，这样就会丧失推广能力。

（2）结构风险最小化（SRM）

SRM 归纳原则旨在针对经验风险和置信范围这两项最小化风险泛函。风险的界是经验风险与置信范围之和。随着结构元素序号的增加，经验风险减小，而置信范围将增加。最小的风险上界是在结构的某个适当的元素上取得的。

统计学习理论提出了一种新的策略，即把函数集构造为一个函数集序列，使各个子集按照 VC 维的大小（亦即Φ的大小）排列；在每个子集寻找最小经验风险，在子集间折衷考虑经验风险和置信范围，使之达到实际风险的最小。实现 SRM 原则也可以有两种思路：一是在每个子集中求最小经验风险，然后选择使最小风险和置信范围之和最小的子集。显然这种方法比较费时，当子集数目很大甚至是无穷时不可行；因此第二种思路是设计函数集的某种结构使每个子集中都能取得最小的经验风险（如使训练误差为 0），然后只需选择适当的子集使置信范围最小，则这个子集中使经验风险最小的函数就是最优函数。支持向量机方法实际就是这种思想的具体实现。如图 3-2 所示。



函数集子集：$s_1 \in s_2 \in s_3$

VC维：$h_1 \le h_2 \le h_3$

图 3-2 结构风险最小化示意图

$$R(\alpha^k) \le R_{emp}(\alpha_l^k) + \phi(\frac{l}{h_k}) \tag{3-4}$$

在此式中，不等号右边的第一项是经验风险，第二项是置信范围，其中 $Q(Z,\alpha_i^k)\in S_k$，$S=Q(Z,\ \alpha),\ \alpha\in\Lambda$。在设计学习机器时确定一个 VC 维为某个 h^* 的容许函数集，对于给定数量 l 的训练数据，h^* 的值确定了机器的置信范围，如果对于一个给定数目的训练数据，设计了一个过于复杂的机器，置信范围将会很大，此时即使可以把经验风险最小化为零，在测量集上的错误数目仍可能很大，那么为了避免这种过学习或过适应，即为了得到小的置信范围，必须构造 VC 维小的学习机器，另一方面同时也要得到小的逼近误差，故必须选择合适的机器构造。根据在前面讨论的方法中，神经网络有过学习的现象，而支持向量机却很好地解决了这个问题，那么通过比较可知，神经网络采用的方法是保持置信范围固定（通过选择一个适当构造的机器）并最小化经验风险；支持向量机采用的方法是保持经验风险固定（比如等于零）并最小化置信范围[45]。

3.2 支持向量机回归

支持向量机最初主要用于分类问题。通过引入一个可选的损失函数，支持向量机也能用于回归问题，这个损失函数必须进行适当修改，使之包含一个距离的度量。

支持向量机回归的基本思想是：设 t 时刻有输入和输出样本集

$z_t=\{(x_1,y_1),\cdots,(x_t,y_t)\}$，$x\in R^n, y\in R^m$，式中，$x_i$ 为输入量，y_i 为输出量。通过支持向量机训练回归出一个函数 $f(x)$，使由该函数求出的每个输入样本的输出值和输入样本对应的目标值相差不超过误差 ε，同时使回归出的函数尽量的平滑。

支持向量机回归方法的整个过程可以描述为：首先，输入数据通过映射函数 ϕ 映射到高维空间中，然后计算 $\phi(x)$ 和 $\phi(x_i)$ 的点积，找到相应合适的核函数来代替这个点积，最后使用权重 $\alpha_i-\alpha_i*$ 的值乘以相应的 $K(x,x_i)$，再把这些积求和，加上偏置 b 就是最终的预测值。

根据统计学理论，要解决学习问题首先要定义一个风险函数，把学习看作是从函数集中估计最小化泛函的函数的过程。式(3-2)中的损失函数是用来衡量在一个实际的点 x 上的实际的函数值 y 和估计值 η 之间的差异，为了最小化风险函数，需要最小化损失函数的期望值，因此对于不同的问题需要选择不同的损失函数。支持向量机回归的引入主要是通过损失函数来实现的。

图 3-3 描述了几种可选的损失函数：其中，（a）对应于最小平方误差标准；（b）是一种拉普拉斯损失函数，对出格点（outlier）不如（a）敏感；当数据的分布未知时，Huber 提出的（c）具有最佳性能。然而上述三个损失函数并不能使得支持向量稀疏。为了解决这个问题，Vapnik 提出了（d），这是对 Huber 损失

函数的一个近似，能使得支持向量机变得很稀疏。

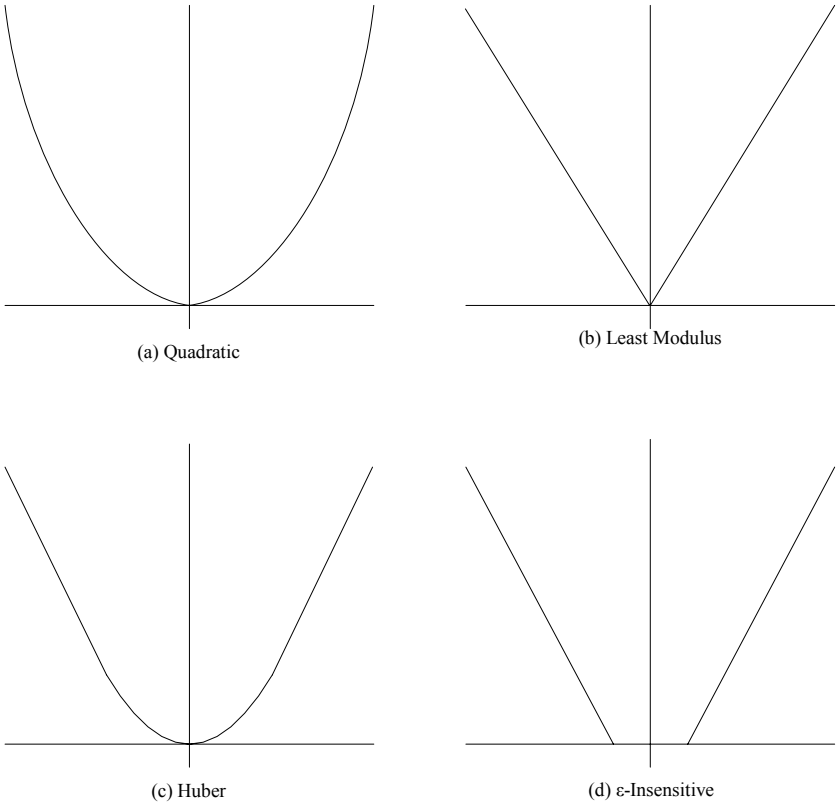


图 3-3 损失函数

3.2.1 线性回归

假设给出训练数据集合 $D=\{(x_1,y_1),\cdots(x_l,y_l)\},x\in R^n,y\in R$，有线性函数：

$$f(x)=(w\cdot x)+b\text{，其中 }w\in X\text{，}b\in R \tag{3-5}$$

最优化回归函数问题就归结为求下式的最小化的解：

$$\phi(w,\xi^*,\xi)=\frac{1}{2}\|w\|^2+C(\sum_{i=1}^{l}\xi_i+\sum_{i=1}^{l}\xi_i^*) \tag{3-6}$$

基中 C 是一个预先给定的值，即正则参数，w 为权值向量，ξ 和 ξ^* 是松驰变量，并且分别是系统输出结果的约束条件的上界和下界。

下面给出不同损失函数的具体算法：

(1) ε-不敏感损失函数

使用如图 3-3（d）描述的 ε-不敏感损失函数

$$L(y,f(x,\alpha))=L(|y-f(x,\alpha)|_\varepsilon) \tag{3-7}$$

$$\text{其中，}|y-f(x,\alpha)|_\varepsilon=\begin{cases}0 & for\,|y-f(x,\alpha)|\le\varepsilon\\ |y-f(x,\alpha)|-\varepsilon & otherwise\end{cases} \tag{3-8}$$

其对偶二次型为：

$$\max_{\alpha,\alpha^*} W(\alpha,\alpha^*) = \max_{\alpha,\alpha^*} \left\{ \begin{array}{l} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)(x_i\cdot x_j) \\ +\sum_{i=1}^{l}\alpha_i(y_i-\varepsilon)-\alpha_i^*(y_i+\varepsilon) \end{array} \right\} \tag{3-9}$$

约束如下：

$$\begin{aligned} &0 \le \alpha_i \le C, i=1,\cdots l \\ &0 \le \alpha_i^* \le C, i=1,\cdots l \\ &\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0 \end{aligned} \tag{3-10}$$

其中 α_i、α_i^* 是拉格朗日乘子，通过由(3-9)式求出 α_i、α_i^* 并代入下式，求出 $\overline{w}$，$\overline{b}$：

$$\begin{aligned} &\overline{w}=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)x_i \\ &\overline{b}=-\frac{1}{2}\overline{w}\cdot[x_r+x_s] \end{aligned} \tag{3-11}$$

再代入 $f(x)=(\overline{w},x)+\overline{b}$，求出回归值。

(2)二次损失函数

使用图 3-3（a）所示的二次损失函数：

$$L(f(x)-y)=(f(x)-y)^2 \tag{3-12}$$

其对偶二次型为：

$$\max_{\alpha,\alpha^*} W(\alpha,\alpha^*) = \max_{\alpha,\alpha^*} \left\{ \begin{array}{l} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)(x_i\cdot x_j) \\ +\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)y_i-\frac{1}{2C}\sum_{i=1}^{l}(\alpha_i^{*2}+\alpha_i^2) \end{array} \right\} \tag{3-13}$$

对应的优化问题可以使用 KKT 条件而得到简化，这意味着 $\beta i^{'}=|\beta i|$，这样优化问题变成：

$$\min_{\beta}\left[\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\beta_i\beta_j(x_i\cdot x_j)-\sum_{i=i}^{l}\beta_i y_i+\frac{1}{2C}\sum_{i=1}^{l}\beta_i^2\right] \tag{3-14}$$

约束为：

$$\sum_{i=1}^{l}\beta_i=0 \tag{3-15}$$

求出 β_i 后，代入下式求出 $\overline{w}$，$\overline{b}$：

$$\overline{w}=\sum_{i=1}^{l}\beta_i x_i$$

$$\overline{b}=-\frac{1}{2}\overline{w}\cdot[x_r+x_s] \tag{3-16}$$

再把 $\overline{w}$，$\overline{b}$ 代入 $f(x)=(\overline{w},x)+\overline{b}$，即可求出回归函数值。

（3）Huber 损失函数

使用图 3-3(c)所示的 Huber 损失函数：

$$L(f(x)\text{-}y)=\begin{cases}\dfrac{1}{2}(f(x)-y)^2 & for|f(x)-y|<\mu\\ \mu|f(x)-y|-\dfrac{\mu^2}{2} & otherwise\end{cases} \tag{3-17}$$

相应的对偶化二次型为：

$$\max_{\alpha,\alpha^*}W(\alpha,\alpha^*)=\max_{\alpha,\alpha^*}\left\{\begin{array}{l}-\dfrac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)(x_i\cdot x_j)\\ +\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)y_i-\dfrac{1}{2C}\sum_{i=1}^{l}(\alpha_i^{*2}+\alpha_i^2)\mu\end{array}\right\} \tag{3-18}$$

其中，α，α^* 为拉格朗日乘子，采用二次优化方法，令 $\beta_i=\alpha_i^*-\alpha_i, i=1,\cdots,l$，代入（3-18）式得：

$$\min_{\beta}\left[\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\beta_i\beta_j(x_i\cdot x_j)-\sum_{i=i}^{l}\beta_i y_i+\frac{1}{2C}\sum_{i=1}^{l}\beta_i^2\mu\right] \tag{3-19}$$

约束式为：

$$\text{-}C\le\beta_i\le C, i=1,\cdots l$$

$$\sum_{i=1}^{l}\beta_i=0 \tag{3-20}$$

由 Vapnik 提出的 ε 不敏感损失函数是 Huber 损失函数的近似值，并且在这几个损失函数中，除了 ε 不敏感损失函数之外，其余的损失函数不满足支持向量稀疏性特点。

在采用 ε 不敏感损失函数的估计回归函数中有三个重要的自由参数：ε 不敏感性值、正则化参数 C 值及核参数（多项式核的阶数，径向基核函数的宽度参数、样条生成核的样条阶数等等）。

3.2.2 非线性回归

和分类问题类似，非线性模式一般需要适当的模型数据。和支持向量分类器的方法相似，用一个非线性映射，把数据映射到高维空间，在这个空间里使用线性的方法解决。核函数同样被引入，用来解决高的维数带来的麻烦。使用图 3-3(d)描述的ε-不敏感损失函数，给出非线性回归问题的对偶形式：

$$\max_{\alpha,\alpha^*} W(\alpha,\alpha^*) = \max_{\alpha,\alpha^*} \left\{ \begin{array}{l} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(x_i\cdot x_j) \\ \quad +\sum_{i=1}^{l}\alpha_i^*(y_i-\varepsilon)-\alpha_i(y_i+\varepsilon) \end{array} \right\} \tag{3-21}$$

约束如下：

$$0 \le \alpha_i \le C, i = 1,\cdots l$$
$$0 \le \alpha_i^* \le C, i = 1,\cdots l$$

$$\sum_{i=1}^{l}(\alpha_i-\alpha_i^*) = 0 \tag{3-22}$$

解(3-22)约束的(3-21)式得到拉格朗日乘数α_i、α_i^*，回归函数由下式给出：

$$f(x) = \sum_{sv}(\overline{\alpha_i}-\overline{\alpha_i}^*)K(x_i,x)+\bar{b} \tag{3-23}$$

其中：$\bar{w}\cdot x = \sum_{i=1}^{l}(\overline{\alpha_i}-\overline{\alpha_i^*})K(x_i,x)$

$$\bar{b} = -\frac{1}{2}\sum_{i=1}^{l}(\overline{\alpha_i}-\overline{\alpha_i^*})[K(x_r,x_i)+K(x_s,x_i)] \tag{3-24}$$

其他的损失函数，其优化过程都是类似的，只要用核函数替换高维空间的点积就行了。ε-不敏感损失函数是很吸引人的，他和二次损失函数及 Huber 损失函数不一样，后两者的所有样本点都将是支持向量，而他的支持向量是稀疏的。在回归方法中，有必要选择一个有代表性的损失函数，以及对噪声的分布有掌握。如果缺乏这些了解，那么图(c)所示的 Huber 损失函数就是最好的选择。Vapnik 开发的ε-不敏感损失函数是一个折中，他具有 Huber 函数的部分特性，而且使得所产生的支持向量稀疏。但是实际计算量上更庞大，而且ε-不敏感区存在缺陷。

3.3 核函数

核技巧是支持向量机的重要组成部分，支持向量机是应用核技巧的一个典型例子，核函数是核技巧的基础[47]。不论是非线性分类情况还是非线性回归情况，

SVM 都通过核函数的办法巧妙地解决了“维数灾难”问题，从而提高了 SVM 的推广性[48-49]。

核函数就是满足 Mercer 条件的任意对称函数。当遇到线性不可分时就要考虑转化为非线性进行求解，那么就要有输入向量映射到高维特征空间后，在高维特征空间进行线性可分，这时若根据要求构造不同类型的 SVM，则需要选择满足 Mercer 条件的不同的核函数来代替原式中输入向量的内积。

定理 3.1 (Mercer 条件) 对于任一的对称函数$K(x,x')$，它是某个特征空间中的内积运算的充分必要条件是，对于任一的$\phi(x)\neq 0$且$\int \phi^2(x)dx<\infty$，有

$$\iint K(x,x')\phi(x)\phi(x')dxdx'>0 \tag{3-25}$$

3.3.1 常用的核函数

（1）多项式核函数

多项式映射是非线性模型的最普通方法，

$$K(x,x_i)=(x\cdot x_i)^d \qquad \text{d=1,2}\cdots\text{N} \tag{3-26}$$

$$K(x,x_i)=((x\cdot x_i)+1)^d \qquad \text{d=1,2}\cdots\text{N} \tag{3-27}$$

第二个核函数即(3-26)式一般在避免 Hessian 矩阵为 0 的情况下采用。

（2）高斯径向基函数（Gaussian RBF）

$$K(x,x_i)=\exp(-\frac{\|x-x_i\|^2}{2\sigma^2}) \tag{3-28}$$

这个核函数由于具有某些良好的性质，表现出较强的学习能力，而受到大家的亲睐。其基本点如下：

①当参数$\sigma\to 0$时，$\alpha_l>0$即全部样本点都是支持向量。

②对于任意的样本集Φ，只要$\sigma>0$充分小，高斯核函数支持向量机必定可对其正确分类。

③当$\sigma\to\infty$时，高斯核函数支持向量机的判别函数为一常数，即把所有样本点判为一类。

④σ过大，其分类能力较差；但σ过小也会造成“过拟合”现象而降低对新样本的正确分类能力。

（3）指数径向基函数（Exponential RBF）

$$K(x,x_i)=\exp(-\frac{\|x-x_i\|}{2\sigma^2}) \tag{3-29}$$

（4）sigmoid 核函数

$$K(x,x_i)=\tanh(\nu(x,x_i)+c),(\nu,c\in R) \tag{3-30}$$

（5）样条(spline)核函数

$$K(x,x_i)=1+(x,x_i)+\frac{1}{2}(x,x_i)\min(x,x_i)-\frac{1}{6}\min(x,x_i)^3 \tag{3-31}$$

核函数的选择需要一定的先验知识，目前还没有一般性的结论，Scholkopf等在文献[50]中就核函数的选择和构造作了讨论。

3.3.2 核函数方法的实施步骤

核函数方法是一种模块化方法，它可以分为核函数设计和算法两个部分[51]，具体如图3-4所示。

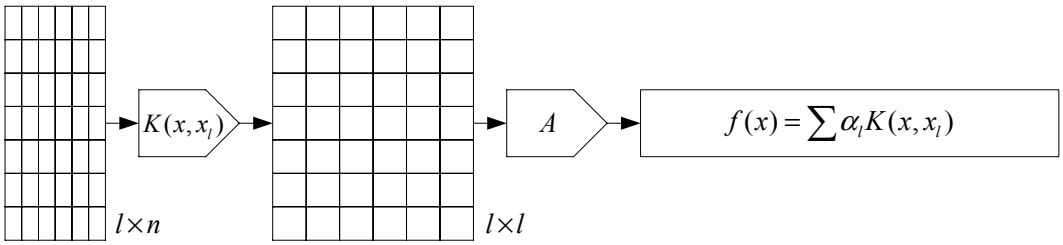


图 3-4 核函数方法的实施步骤

其实施步骤，可以具体的描述为：

1、收集和整理样本，并进行标准化；

2、选择或构造核函数；

3、用核函数将样本变换成为核函数矩阵，这一步相当于将输入数据通过非线性函数映射到高维特征空间；

4、在特征空间对核函数矩阵实施各种线性算法；

5、得到输入空间中的非线性模型。

显然，将样本数据核化成核函数矩阵是核函数方法中的关键。注意到核函数矩阵是$l\times l$的对称矩阵，其中l为样本数。

3.4 支持向量机与神经网络

人工神经网络被广泛认为是较强的非线性估计器之一。理论上，神经网络可以逼近定义在R^n中紧集上的任意连续函数。它在许多领域，尤其是在非线性预测方面得到了广泛的应用。但是神经网络存在一些应用上的难点，如需要确定网络结构、过学习与欠学习、局部极小点问题等，引起这些问题的最根本原因是神

经网络方法缺乏理论上的突破。

支持向量机方法和神经网络相比有较成熟的理论基础。虽然小样本统计学习理论还需要进一步完善和发展，但支持向量机方法已经表现出许多特有的优势，在小样本学习、非线性、高维问题以及泛化能力方面表现非凡。

3.4.1 相似点

由上过可以看出支持向量机和神经网络都有较强的非线性学习能力。从模型结构角度分析，支持向量机理论上的模型结构和神经网络的实际模型结构也类似。

(1) 非线性学习能力

神经网络的每一个处理单元或网络节点，仅具有简单的非线性处理能力，但是通过它们之间的相互连接、相互作用，便能表现出复杂的智能处理能力和非线性处理能力。神经网络所提供的非线性静态映射关系是它在各个领域成功应用的基础。

支持向量机则是通过变换把非线性的输入映射到高维特征空间，然后在此空间中进行最优解的求取，因此它可以有效处理非线性的问题。

对于复杂的预测对象，影响因素很多，诸多影响因素很难全部掌握，所能考虑的只能是一些主要因素，而这些主要因素与目标之间很少是线性关系，因此支持向量机和神经网络的非线性能力在预测领域是很重要的。

(2) 模型结构特点

由于 BP 网络具有很强的非线性映射能力，模型结构简单，是目前应用最广泛的一种神经网络。以三层 BP 网络为例，网络由一个输入层、一个隐含层和一个输出层组成。如图 3-5 所示，输入层节点只是传递输入信号到隐含层，隐含层激励函数 f 通常取为 S 型函数。

$$f(x)=\frac{1}{1+e^{-x}} \tag{3-32}$$

$$\text{其中}\ u_j=f\left(\sum_i^n w_{ji}x_i-\theta_j\right),\ \ j=1,2,\cdots N_h \tag{3-33}$$

其中 u_j 是第 j 个隐含层节点的输出，x_i 是输入样本 $X=(x_1,x_2,\cdots x_n)^T$ 的第 i 维，w_{ji} 为输入层第 i 节点到隐含层第 j 个节点的权值，θ_j 为隐含层第 j 个节点的阈值，N_h 是隐含层节点数。输出层节点通常是简单的线性函数，为隐含层节点输出的线性组合

$$y_i=\sum_{j=1}^{N_h} w_{ij}u_j-\theta_i=W_i{}^TU\ ,\quad i=1,2,\cdots m \tag{3-34}$$

其中，$W_i = (w_{i1}, w_{i2}, \cdots w_{iNh}, -\theta_i)$，

$$U = (u_1, u_2, \cdots u_{Nh}, 1)^T$$

W_i 为隐含层节点到输出节点 i 的权值矩阵，θ_i 为相应的阈值。这样的 BP 网络通过正向计算误差，反向修正权值和阈值最小化目标函数（误差函数）来确定网络参数，从而实现了一个 $R^n \rightarrow R^m$ 的非线性函数映射关系。

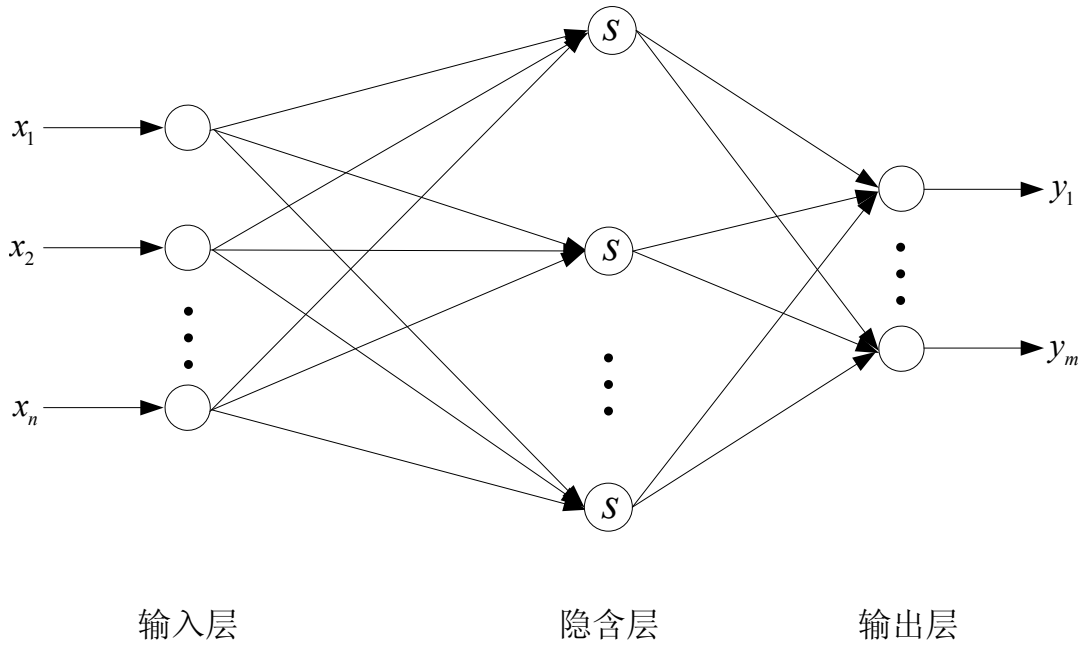


图 3-5 BPNN 网络模型结构图

支持向量机在理论上有与神经网络极其相似的结构。如图 3-6 所示，支持

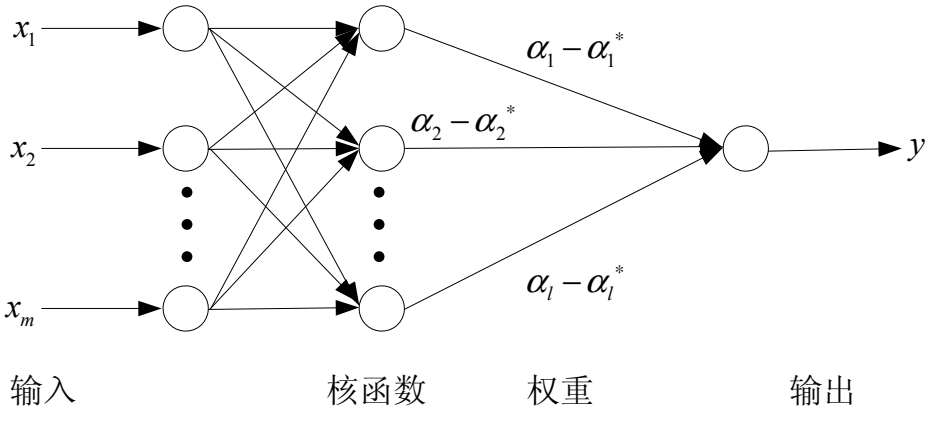


图 3-6 支持向量机的结构图

向量机相当于具有一个隐含层的三层神经网络，它的输入与神经网络的输入相对应，支持向量机中的支持向量对应于隐含层，隐含层节点数 L 相当于支持向量

的个数 L，它的网络权重为 $\alpha_i - \alpha_i^*$。输入层实现 $x \rightarrow$ 核函数 $K(x_i, x_j)$ 的非线性映射，即通过核函数把输入映射到线性的特征空间，实现非线性映射。输出层实现核函数 $K(x_i, x_j) \rightarrow y$ 的线性映射，即

$$y = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) K(x_i, x) + b \tag{3-35}$$

3.4.2 不同点

支持向量机和神经网络相比，成功地解决了如何合理确定网络结构、过学习与欠学习、局部极小等棘手的问题。

(1) 理论依据方面

传统的神经网络的学习算法缺乏机理完备的理论依据，而 SVM 是基于统计学习理论和对偶定理（Duality Theory），具有完备的数学理论背景。

(2) 原理方面

神经网络的算法采用经验风险最小化原理，存在过学习现象，从而影响了它的泛化能力。而 SVM 采用结构风险最小化原理，提高了对未来的样本的泛化能力。

(3) 模型结构方面

支持向量机理论上的模型结构和三层神经网络很相似，它们的区别是支持向量机的支持向量个数即为隐含节点数，由 SVM 算法自动产生; 而神经网络的整个网络结构包括隐含层个数、节点数都是事先人为确定的。表现在训练过程上就是支持向量机自动产生所有的模型结构，而神经网络仅产生权重。这实际上说明了支持向量机具有自适应的模型结构，而神经网络是预先确定模型结构。

(4) 是否全局最优方面

在给定参数的情况下，支持向量机能够保证满足凸二次优化的 KKT 条件，可到得局部最优解，而神经网络可能会陷入局部极小点。不能保证达到全局最优。BP 算法收敛得到的解很可能是局部最优解，因为一般神经网络采用的是梯度下降法，训练是从某一起点沿误差函数的下降方向逐渐达到最小值。对于复杂的网络，误差函数为多维空间的曲面，曲面表面凹凸不平，训练过程中常会陷入这样的极小点。采用一些办法如更多层网络和较多的节点会改善解的效果，然而却增加了网络的复杂性和训练时间。

概括与总结 SVM 与 ANN 两种方法，表 3-4 列出了比较结果。

表 3-4 SVM 与 ANN 的比较

方法	原 理	模型结构	学习能力	泛化能力	收敛速度	全局最优
SVM	结构风险最小化	自适应确定	强	强	快	是
ANN	经验风险最小化	事先确定	较强	不强	慢	否

3.5 最小二乘支持向量机

我们将 Vapnik 提出的支持向量机称之为标准支持向量机(S-SVM)，最小二乘支持向量机（LS-SVM）则是由 J.A.K.Suykens 提出的对于 S-SVM 的一个重要的变种。它在保持了 S-SVM 优点的基础上，通过适当的变换，简化了运算算法，显著降低了计算成本。

3.5.1 LS-SVM 回归算法

和 S-SVM 一样，LS-SVM 最开始也是针对分类问题提出的，然后再引进了回归估计和控制领域[52]。本文，我们仅研究回归建模。

给定一个有 l 个样本的训练集合 $(x_i, y_i), x \in R^n, y \in R, i = 1,2,\cdots l$，其中 x_i 作为输入样本，y_i 作为输出样本。构造最优线性函数

$$f(x) = w\phi(x) + b \tag{3-36}$$

其中：$\phi(\cdot)$：$R^n \to R^{nh}$ 为将输入样本数据映射到高维特征空间的函数；$w \in R^{nh}$ 为权值向量。根据结构风险最小化原则, 函数拟合问题可描述为以下最优化问题：

$$\begin{cases} \min J(w,\xi) = \dfrac{1}{2} w^T w + \dfrac{1}{2} C \sum_{i=1}^{l} \xi_i^2 \\ y_i = w^T \phi(x_i) + b + \xi_i, i = 1,\cdots l \end{cases} \tag{3-37}$$

其中，$\phi(\cdot)$：$R^n \to R^{nh}$，和在 S-SVM 中一样，是将数据从原始空间映射到高维 Hilbert 特征空间的非线性函数；w 为权值向量；ξ_i 为误差变量（相当于 S-SVM 下的松弛因子 ξ）；C 是调整参数因子（通常叫正则参数或惩罚因子），$C > 0$，能够使训练误差和模型的复杂度之间取一个折中，以便使所求得的函数有较好的泛化能力。C 值越大，则模型的回归误差越小，但是，在训练数据噪声较大时，C 应取较小的值。

对于(3-37)的求解就是对应着 LS-SVM 模型。同 S-SVM 相比，LS-SVM 主要有两个地方改变：

1) 约束条件发生了变化，由原来的不等式约束变为了等式约束；

2) 损失函数发生了变化，由误差ξ_i的平方项ξ_i^2作为损失函数取代了ε不敏感函数。

同 S-SVM 一样，上述最优化问题(4-2)的求解，可以引入 Lagrange 乘子加以解决，Lagrange 函数为：

$$L(w,b,\xi,\alpha)=J(w,\xi)-\sum_{i=1}^{l}\alpha_i\{w^T\phi(x_i)+b+\xi_i-y_i\} \tag{3-38}$$

其中，α_i为 Lagrange 函数乘子。由于现在的约束为等式约束，所以 Lagrange 乘子α_i的值可正可负，而不是像 S-SVM 那样必须是非负值。

于是，对 Lagrange 函数的各变量求偏导，并令偏导数为零，得到如下方程组

$$\begin{cases}\dfrac{\partial L}{\partial w}=0\rightarrow w=\sum_{i=1}^{l}\alpha_i y_i\phi(x_i)\\ \dfrac{\partial L}{\partial b}=0\rightarrow\sum_{i=1}^{l}\alpha_i y_i=0\\ \dfrac{\partial L}{\partial \xi_i}=0\rightarrow\alpha_i=C\xi_i,i=1,\cdots l\\ \dfrac{\partial L}{\partial \alpha_i}=0\rightarrow w^T\phi(x_i)+b+\xi_i-y_i,i=1,\cdots l\end{cases} \tag{3-39}$$

式(3-39)的求解过程与在 S-SVM 中很相似，但是有一个条件变化最大，就是$\alpha_i=C\xi_i$。式$\alpha_i=C\xi_i$表明，每一个数据点对于回归估计函数都是有贡献的，也就是说所有的数据点都是支持向量。

在消去变量 w 和 ξ 后，就可以得到对偶问题的一个线性 Karush-Kuhn-Tucker(KKT)系统。

$$\begin{bmatrix}0 & 1_l^T\\ 1_l & \Omega+I/C\end{bmatrix}\begin{bmatrix}b\\ \alpha\end{bmatrix}=\begin{bmatrix}0\\ y\end{bmatrix} \tag{3-40}$$

其中$1_l^T=[1;\cdots 1]$，$y=[y_1;\cdots;y_l]$，$\alpha=[\alpha_1;\cdots;\alpha_l]$，$\xi=[\xi_1;\cdots;\xi_l]$，

$$\Omega_{ij}=\phi(x_i)^T\phi(x_j)=K(x_i,x_j),i、\ j=1,\cdots l$$

如果式(3-40)的左端$\Phi=\begin{bmatrix}0 & 1_l^T\\ 1_l & \Omega+I/C\end{bmatrix}$是可逆的，那么就可以通过解析的方

法来求得参数α和b，即$\begin{bmatrix} b \\ \alpha \end{bmatrix} = \Phi^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix}$。

上述式(3-40)的求解过程，与求解最小乘（LS）的方法类似，所以这个 S-SVM 的变换形式称为 LS-SVM。

而Ω中同样用到了核函数的技巧，形式如下

$$\Omega_{ij} = \phi(x_i)^T \phi(x_j) = K(x_i, x_j), i、\ j = 1, \cdots l \tag{3-41}$$

其中，$K(x_i, x_j)$为满足 Mercer 条件的核函数。

这样 LS-SVM 的函数估计表达式（即拟合模型）为

$$y(x) = \sum_{i=1}^{l} \alpha_i K(x_i, x) + b \tag{3-42}$$

3.5.2 LS-SVM 与 S-SVM 的比较

在 S-SVM 求解过程中的凸优化问题需要通过二次规划方法来解决，求解二次规划问题需要计算核函数矩阵，其计算量与训练样本数的立方成正比，当处理大规模数据回归问题时，计算量比较大，内存占用量较多[53]。利用 LS-SVM 可以在一定程度上解决 S-SVM 的计算复杂性的问题。它将 SVM 优化问题的不等式约束变为等式约束。获得 LS-SVM 模型仅仅需要求解一个线性方程组，从而演化成为简单的矩阵逆运算。

LS-SVM 和 S-SVM 主要不同之处在于，LS-SVM 将误差的二次平方作为损失函数而不是将ε不敏感函数作为损失函数。这样便可以将不等式约束条件转变成等式约束，从而用线性运算的方法就可以进行运算，详细算法参见文献[54−55]。

由 3.5.1 节的分析可以看出，LS-SVM 将一个凸二次规划问题的系数规划问题转变为一组线性方程的求解问题，从而同 S-SVM 相比，大大简化了计算复杂度。如果用l代表训练样本数，那么线性方程组的系数矩阵大小为$(l+1)\times(l+1)$，而在 S-SVM 中，凸二次规划问题的系数矩阵也就是海森（Hessian）矩阵的大小为训练样本点的平方，即$l^2 \times l^2$。所以 LS-SVM 不仅使计算的复杂度降低，还使得存储空间的要求大大减少，从而降低了计算成本。

3.6 小结

本章在简单介绍统计学理论的基础上，深入阐述了 SVM 以及 LS-SVM 回归的方法，详细比较了 SVM 与 ANN 相似点与区别，通过比较验证了 SVM 方法的优越，另外，本章也对 SVM 方法的重要组成部分核函数方法进行了介绍，并简单比较了 LS-SVM 与 S-SVM 的不同。

第四章 基于 SVM 的苛性比值与溶出率软测量模型

4.1 苛性比值与溶出率软测量模型的整体框架

经过多年的探索，人们已经总结出了多种针对不同对象的建模方法。回归建模因其良好的工程实用性，在软测量领域得到广泛的应用。回归分析方法作为一种经典的建模方法，不需对生产过程的机理有较多的分析，只要采集足够的过程变量数据和有关的分析数据，运用统计方法进行数据处理，从大量的可测得的过程输入输出数据中寻找过程变量之间的回归关系或相互关系，从而建立生产过程的数学模型。

氧化铝高压溶出工序中，苛性比值与溶出率不仅决定了产品的产量和碱耗，而且对氧化铝的后续生产有极大的影响。在实际生产中，由于条件限制只能定时取样离线分析原矿浆与溶出液化学成分，再计算得到苛性比值与溶出率，一般存在 2 个小时的滞后，难以实时优化控制。利用软测量技术建立数学模型实现对苛性比值与溶出率的在线实时预测对于提高自动化水平、优化控制非常必要。软测量模型整体框架如图 4-1 所示。

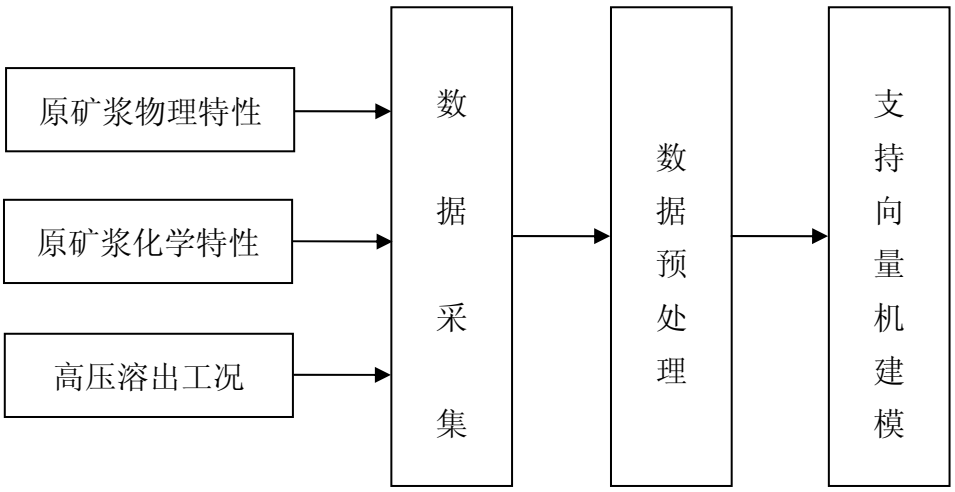


图 4-1　软测量模型整体框架图

从第二章可知，影响苛性比值与溶出率的因素主要包括三个方面，即原矿浆的物理特性、化学特性以及高压溶出工况。具体来讲，图 4-1 中，影响苛性比值与溶出率的原矿浆的物理特性包括：原矿浆的固体含量及液体与固体的质量比值、原矿浆中铝土矿的磨细程度；影响苛性比值与溶出率的原矿浆的化学特性包括：原矿浆中的铝硅比、原矿浆中氧化钙和氧化钛含量、循环母液的苛性碱浓度

及苛性比值；影响苛性比值与溶出率的高压溶出工况包括：溶出温度、溶出压力及压差、原矿浆的进口流量。

在高压溶出过程机理分析的基础上，从生产现场和实验室采集大量样本数据，并对所采集的数据进行预处理，最大限度地保留有用信息，滤去冗余信息，消除相关性和冗余性。最后，对高压溶出过程中苛性比值与溶出率进行回归建模。

4.2 数据采集

通过对工艺机理进行分析，在初步选择影响过程主导变量的辅助变量的基础上，如何采集过程数据也是一个非常重要的问题，关系到软测量建模的成败。数据样本的采集应结合以下原则：

(1) 均匀性

首先要详细研究待建模的生产过程的各种操作工况，确定要研究问题的各辅助变量的可能范围，采集的样本空间要尽量覆盖整个操作范围。即对于一个过程变量来说，采集的数据不仅要充满整个操作范围，而且在每个特征点的局部小范围内也应有一个合适的均匀分布的数据。每个特征点选取的样本量要适中，数据要均匀，不能在一个特征点上取大量重复数据，而在边缘点上只有零星几个数据。

(2) 精简性

采集的数据样本空间还要保证“精简”。这里的精简是一个相对的概念，如果一个特征点附近采集的数据量太多，一方面不利于建模，导致网络的学习困难；另一方面采集的样本量大，很难保证采集的数据相互之间保持一致性。

(3) 有代表性

由于生产过程的复杂性，采集的数据点有时会因为过程的噪声或干扰影响而使其在局部范围内具有随机性，例如在一个特征点附近采集的大量数据中有一小部分样本与其余样本的变化趋势相违背，这时采集的数据就缺乏代表性，影响软测量模型的训练过程。这些互相矛盾的数据使得建立的模型即使具有良好的拟合精度，也往往泛化结果不好。

采集的数据满足上述均匀性、精简性、有代表性的原则时，理论上，采集的样本越多，越能更好的反映过程的特性。根据这一理论，选取了一部分过程数据和实验室人工分析数据作为初选的训练样本。

4.3 数据预处理

氧化铝高压溶出是一个极其复杂的生产过程，整个高压溶出过程中同时发生多种物理化学变化。从第二章对高压溶出的机理分析可以知道，影响苛性比值与溶出率的因素主要包括三个方面，即原矿浆的物理特性、原矿浆的化学特性以及

高压溶出工况。为了从这些可以直接检测的数据中尽可能多的获得关于苛性比值与溶出率的信息，在条件允许的前提下，应该尽可能全面的考虑所有这些因素对苛性比值与溶出率的影响。综合考虑实际生产中的检测情况，总共选取了 25 个辅助变量，详见表 4-1。

表 4-1 选取的辅助变量列表

序号	辅助变量	序号	辅助变量	序号	辅助变量	序号	辅助变量	序号	辅助变量
1	Si^1	6	流量	11	Na^2	16	$3^{\#}T$	21	$7^{\#}P$
2	Ca^1	7	Si^2	12	$1^{\#}T$	17	$3^{\#}P$	22	$8^{\#}T$
3	Al^1	8	Ca^2	13	$1^{\#}P$	18	$5^{\#}T$	23	$8^{\#}P$
4	Ti^1	9	Al^2	14	$2^{\#}T$	19	$5^{\#}P$	24	$9^{\#}T$
5	Na^1	10	Ti^2	15	$2^{\#}P$	20	$7^{\#}T$	25	$9^{\#}P$

上表中：X^1表示原矿浆中 X 含量，X^2表示溶出液中 X 含量，$Y^{\#}T$表示 Y 号溶出器温度，$Y^{\#}P$表示 Y 号溶出器压力。

然而，如果将这 25 个辅助变量都直接作为模型的输入，会使得模型极其复杂，降低模型的泛化能力[56]。另一方面，这些变量并不是相互独立的，它们具有一定的相关性（如各溶出器的温度、压力等）。经验告诉我们，任何模型都能容纳一定程度的相关性，但是太多的相关信息会使得模型过于复杂，泛化能力差，从而影响软测量模型的精度和可信度。因此，在建立软测量模型之前，必须对输入样本进行适当的预处理，从众多影响因素中找出若干个公共的支配因子，最大限度地保留有用信息，滤去冗余信息，降低输入数据集的维数，这样就能大大简化模型结构，提高模型的泛化能力。

4.3.1 数据处理方法

（一）因素分析法

因素分析法[57]的概念是英美心理统计学者们最早提出的。因素分析法从试验所得的$m\times n$个数据样本出发，确定其中各因素对某一指标的影响程度，是统计学和数据处理中应用较为广泛的一种方法。因素分析的数学模型为：

$$X = P\cdot F + S \tag{4-1}$$

式中$X=[x_1,x_2,\cdots,x_m]^T$为可观测的m个随机向量，任一分量x_i可看作带随机性的时间序列变量$x_i=[x_{i1},x_{i2},\cdots,x_{in}]^T$，即第$i$个变量的$n$次样本值；$F=[f_1,f_2,\cdots,f_q]^T$为$q$个公共因素变量，其中$f_1=[f_{i1},f_{i2},\cdots,f_{in}]^T$；$P=[p_{ik}]_{m\times q}$为因素负荷矩阵；$S=[S_{ik}]_{m\times n}$为噪音矩阵。$X=[x_1,x_{2\ldots},x_m]^T$

为计算方便，常将X标准化，第i个分量的第j个样本标准化为：

$$z_{ij}=\frac{x_{i1}-u_i}{\sigma_i^{\ 2}}(i=1,2,\cdots m;\ j=1,2,\cdots n) \tag{4-2}$$

式中$u_i=\frac{1}{n}\sum_{j=1}^{n}x_{ij}$是第$i$个变量的样本均值；$\sigma_i^{\ 2}=\frac{1}{n}\sum_{j=1}^{n}(x_{if}-u_i)^2$是第$i$个变量的样本方差。从而$z_i$为均值、方差为1的随机变量，经过标准化的数学模型可表示如下：

$$z_{ij}=\sum_{k=1}^{q}p_{ik}\cdot f_{ik}+s_{if}(i=1,2,\cdots,m;\ j=1,2,\cdots,n) \tag{4-3}$$

且满足$E(f)=0$，$E(s)=0$，$E(ff)^T=I$，$E(ss^T)=\mathrm{diag}(\phi_1,\phi_2,\cdots,\phi_m)=\Phi$，其中$\Phi$称为个性方差阵。$m$维随机向量$Z$的协方差阵$\mathrm{cov}(Z,Z)$与相关系数阵$R_{mm}$有如下关系：

$$\mathrm{cov}(Z,Z)=\mathrm{PP}^{\mathrm{T}}+\Phi \tag{4-4}$$

因此，因素负荷矩阵P的统计意义为：P的行元素$h_i^{\ 2}=\sum_{j=1}^{q}p_{ij}(i=1,2,\cdots,m)$代表公共因素$F$对变量$z_i$的方差所做的贡献，反映了变量$z_i$对公共因素的依赖程度；$P$的列元素平方和$g_k^{\ 2}=\sum_{i=1}^{m}p^2{}_{ik}(k=1,2,\cdots,q)$表示第$k$个公共因素$f_k$对向量$Z$的所有分量的影响，反映了随机向量$Z$对$f_k$的依赖程度，是衡量公共因素$f_k$相对重要性的一个尺度。因此可以根据$g_k^{\ 2}$的大小将公共因素向量$F$分解为两个向量$F_1$和$F_2$，其中$F_1$表示对$z$的影响较大的一组公共因素，$F_2$表示对$z$的影响较小的一组公共因素。

因素分析一般采用连锁替代法，它利用各个因素的实际数与标准数的连续替代来计算各因素脱离标准所造成的影响。

（二）主元分析法

主元分析(PCA)是输入数据降维处理的主要方法之一。主元分析又称主成分分析[58]，是一种将多个变量转化为少数变量的多元统计方法。主元分析的目的是对多变量数据表进行综合简化。如果在原数据表中有n个变量，主元分析将考虑对这个数据表中的信息重新调整组合，从中提取k个综合变量(主成分)，使这个综合变量能最多地概括原数据表中的信息，在力保数据信息损失最少的原则下，对高维变量空间进行降维处理。根据数据变化的方差大小来确定变化方向的主次地位，按主次顺序得到各主元素，这些主元素彼此之间是无关的。借助这一工具，可提炼变化信息，减轻数据分析的复杂程度。

PCA 思想是对高维的统计特征进行变换，从 n 维变量中提取贡献最大的 k 维[59]($k\leq n$)变量。其实质是对过程操作变量协方差的特征向量分解。

给定生产过程数据样本集 $X \in R^{m\times n}$ 已按列零均值标准化，以消除变量量纲不同和变化幅度不同带来的影响，其中 m 为样本个数，n 为输入变量个数。定义 X 的协方差矩阵为

$$Cov(X)=\frac{X^TX}{m-1} \tag{4-5}$$

对式（4-5）进行正交分解，得

$$Cov(X)=PDP^T \tag{4-6}$$

式中，$D=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$ 为协方差矩阵的特征根矩阵，$\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_n$；$P=[p_1,p_2,\cdots,p_n]$为特征向量，称为负荷矩阵。定义前 k 个主元的信息占有量为

$$\eta(k)=\sum_{i=1}^{k}\lambda_1/\sum_{i=1}^{n}\lambda_i \tag{4-7}$$

通常选取$\eta(k)\ge 95\%$，得 k 值。确定主元个数后，可得主元模型

$$X=TP^T \tag{4-8}$$

式中，T 为得分矩阵，P 为负荷矩阵。将 X 重构为

$$X=\hat{X}+\tilde{X} \tag{4-9}$$

式中，$\tilde{X}=T_kP_k^T$，$\tilde{X}$ 为 X 的残差。$\tilde{X}$ 主要是由测量噪声引起的，将其忽略不会造成数据中有用信息的明显损失，故有

$$X\approx\tilde{X} \tag{4-10}$$

这样，消除了数据样本集的共线性和冗余度，达到特征提取和降低变量维数的目的。从而，为回归模型提供了样本简洁、信息全面的输入。

（三）LS-SVM 特征提取法

数据集特征提取通常是对数据集的样本重新进行维数组合（线性或非线性）时，将分布在原有特征中的分类信息集中到较少数量的特征中来，这样就能达到降低样本维数，而又尽可能多地保持原有分类信息的目的。一般认为相似的输入很有可能属于同一类，输入变量的方差越大，其所具有的聚类能力相对也越好[60]。以线性特征提取为例，即通过对输入的样本数据集 $X=\{x_i\}$的特征重新线性组合，寻找单位列向量 w 使得线性组合数列 wx_i 具有最大的方差。

首先对样本数据集 X 进行零均值化：

$$x_i^{'}=x_i-\overline{X} \tag{4-11}$$

置输出值 $Y=\{y_i\}=0$，由(3-36)式 (即 $f(x)=w\phi(x)+b$)进行线性回归，得回归误差为：

$$\xi_i=f(x_i^{'})-y_i=wx_i^{'}+b \tag{4-12}$$

有：

$$\bar{\xi}=\frac{1}{l}\sum_{i=1}^{l}\xi_i=\frac{1}{l}\sum_{i=1}^{l}(wx_i^{'}+b)=\frac{w}{l}\sum_{i=1}^{l}x_i^{'}+b=b \tag{4-13}$$

其方差为：

$$D(\xi)=\frac{1}{l}\sum_{i=1}^{l}(\xi_i-\bar{\xi})^2=\frac{1}{l}\sum_{i=1}^{l}(wx_i^{'})^2=D(wx_i^{'}) \tag{4-14}$$

这样，求解线性组合数列wx_i具有最大的方差问题就转化为求$\max D(\xi)$问题。取式(3-38)中$C^{'}=-C$，那么 LS-SVM 的回归目标$\min D(\xi)$则转化为$\max D(\xi)$。因为$Y=0$，则式(3-41)可以简化为：

$$\left|\Omega-\frac{I}{C^{'}}\right|\alpha=0 \tag{4-15}$$

若$\alpha=0$，则$w=0$，那么就不存在支持向量，并与特征提取的实际意义相违背，故应求α的非零解。定义$\lambda=\frac{1}{C^{'}}$，则式（4-15）可以转化为对称矩阵特征值问题：

$$\Omega\alpha=\lambda\alpha \tag{4-16}$$

求解式（4-15）矩阵方程，得l个特征值$\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_l$，$\alpha_i$为其对应的特征向量。从而求得线性特征组合向量:

$$w_k=\sum_{i=1}^{l}\alpha_{ki}x_i^{'T} \tag{4-17}$$

则最佳分类特征为：

$$z_k(x)=w_kX^{'}=\sum_{i=1}^{l}\alpha_{ki}x_i^{'T}X^{'} \tag{4-18}$$

因此就生成了新的特征，将原特征从n维降低到$k(k<n)$维，实现特征的线性优化。

而对于非线性特征提取，可遵循 SVM 方法，将数据集$X^{'}$从原始特征空间映射到高维特征空间得$\phi(X^{'})$，并在高维空间中寻找使得组合数列$w\phi(x^{'})$具有最大的方差列向量w。实际计算时用原空间的核函数代替高维特征空间的点积运算，非线性特征提取的最佳分类特征表达为：

$$z_k(x)=w_k\varphi(x_i^{'})=\sum_{i=1}^{l}\alpha_{ki}K(x_i^{'},X^{'}) \tag{4-19}$$

这样，通过对输入样本集进行特征提取，就可以得到维数少、信息完善的新样本集，从而减少了软测量模型的输入量，加快了模型的计算速度，有效的避免了输出过拟合的现象[61]。

4.3.2 数据处理结果

由氧化铝高压溶出生产过程可知，影响苛性比值和溶出率的因素很多。包括原矿浆（铝土矿、循环母液及石灰石按一定比例配成）和溶出液的化学、物理特性、高压溶出器的温度、压力等 25 个因素，为避免由于软测量模型输入量的个数太多，引起过拟合，降低模型精度和计算速度，同时又能更好地反映整体特性，满足精度要求，需要对数据进行预处理。

在进行数据处理过程中，我们发现因素分析法编程较主元分析法和 LS-SVM 特征提取法麻烦，并且运算时间长。基于此，下面，我们仅比较主元分析法和 LS-SVM 特征提取法的数据处理结果。

（一）主元分析法数据处理结果

选取主元分析法中 $\eta(k)$ 大于 0.95。对这 25 个因素进行主元分析，结果如表 4-2 所示。

表 4-2　主元分析结果

序号	主元	λ	η	序号	主元	λ	η	序号	主元	λ	η
1	Si^1	0.3947	2.1%	10	Ti^2	0.0013	0.01%	19	$5^{\#}P$	0.0123	0.01%
2	Ca^1	0.0721	0.41%	11	Na^2	0.3413	1.8%	20	$7^{\#}T$	0.0927	0.50%
3	Al^1	0.1208	0.74%	12	$1^{\#}T$	1.2787	6.8%	21	$7^{\#}P$	0.0793	0.43%
4	Ti^1	0.4135	2.3%	13	$1^{\#}P$	0.9123	4.9%	22	$8^{\#}T$	0.0161	0.09%
5	Na^1	1.9083	10.8%	14	$2^{\#}T$	4.0809	22.1%	23	$8^{\#}P$	0.0931	0.49%
6	流量	0.6445	3.3%	15	$2^{\#}P$	1.7283	9.4%	24	$9^{\#}T$	2.4088	13.0%
7	Si^2	0.0073	0.04%	16	$3^{\#}T$	0.9391	5.5%	25	$9^{\#}P$	1.3164	7.06%
8	Ca^2	0.3161	1.8%	17	$3^{\#}P$	0.6919	3.75%				
9	Al^2	0.5087	2.7%	18	$5^{\#}T$	0.0344	0.02%				

上表中：X^1 表示原矿浆中 X 含量，X^2 表示溶出液中 X 含量，$Y^{\#}T$ 表示 Y 号溶出器温度，$Y^{\#}P$ 表示 Y 号溶出器压力；序号 1、4、5、6、8、9、11、12、13、14、15、16、17、24、25 对应的 λ 值之和大于所有 λ 值之和的 95%，对应的主元即为特征提取后的主元。

分析表 4-1 可知

1）原矿浆中 Si 和 Ti 的含量是影响溶出率和苛性比值的重要因素，因为两

者以氧化物形式存在，与碱液反应生产不溶性物质进入赤泥，同时引起一定量的碱和氧化铝的损耗。而溶出液中 Si 和 Ti 的含量很少，不再是影响溶出率和苛性比值的重要因素；

2）原矿浆中 Na 的含量是影响溶出率和苛性比值的重要因素，因为它直接反映加入苛性钠的质量和溶液的碱性；

3) 原矿浆的流量是影响溶出率和苛性比值的重要因素，因为它直接影响脱硅脱钛的反应程度及氧化铝溶出的程度；

4）溶出液中 Ca 的含量是影响溶出率和苛性比值的重要因素，因为它直接影响原矿浆的配比；

5）溶出液中 Al 的含量是影响溶出率和苛性比值的重要因素，因为它直接反映氧化铝溶出程度；

6）$1^{\#}$、$2^{\#}$、$3^{\#}$、$9^{\#}$高压溶出器的温度和压力是影响溶出率和苛性比值的重要因素，因为原矿浆经 $1^{\#}$、$2^{\#}$加热溶出器预热后不再加热、加压，显然这 9 个高压溶出器的温度、压力存在一定的相关性，而 $3^{\#}$、$9^{\#}$高压溶出器分别是第一个、最后一个溶出器，直接影响和反映氧化铝反应生成铝酸钠的程度。

利用 PCA 提取 25 个因素中的主元因素，得到 15 维样本，剔除了部分次要因素的影响，降低了样本维数，符合实际工艺。

（二）LS-SVM 特征提取法数据处理结果

LS-SVM 特征提取法具体步骤如下：

Step 1：根据式（4-11）对样本数据集 X 进行零均值化，得到数学期望为 0，方差为 1 的新数据集，消除不同量纲的影响；

Step 2：根据式（4-16）求取 λ（由大到小排列）及对应的特征向量，其中：$\lambda_1 = 2.7532$，$\lambda_2 = 1.2087$，$\lambda_3 = 0.7144$，$\lambda_4 = 0.1701$，$\lambda_5 = 0.1536$，$\lambda_6 = 0.0902$，$\lambda_7 = 0.0302$（在不影响模型输出精度的情况下，舍去相对较小的 λ 值，提高模型的运算速度）；

Step 3：根据式（4-19）求取最佳分类特征（本文采用的工况数据具有严重的非线性），核函数 $K(\cdot,\cdot)$，为线性函数($K(x_i, x_j) = x_i^T x_j$,)，这样就将输入样本数据集由 25 维降低为 7 维，大大简化了模型的输入信息。

这样，经 LS-SVM 特征提取得到重组后的输入样本数据集 $\widetilde{X}$，作为回归模型的输入，取软测量模型的输出值 $\widetilde{Y}$ 为实际输出值，建立软测量模型，即可实际离线测量得到的所求值。

由数据的处理结果，可以看出 LS-SVM 特征提取法较主元分析法的数据处理能力强，能为回归模型提供维数更少、冗余度更低、信息全面的输入数据，在不影响模型精度的情况下，提高了模型的运算速度。

4.4 回归模型的建立

应用支持向量机技术，并利用上述处理得到的数据便可建立苛性比值和溶出率的软测量回归模型。

4.4.1 核函数的选择

核函数选择对支持向量机建模的影响显著。从已有数据中选取100组数据，据3.3.1节选择不同常用核函数进行软测量建模，实验结果如表4-3。其中，Poly表示多项式机，RBF表示RBF机，Sigmoid表示神经网络机，Spline表示样条机。均方误差(MSE)定义为 $\{\frac{1}{100}\sum_{i=1}^{100}[y_i - f(x_i)]^2\}^{\frac{1}{2}}$。

从表4-3可以看出，不同的核函数对模型性能的影响非常大，无论是速度还是拟合精度，都具有决定性的影响。根据上述实验数据，本文选择RBF核函数。

表4-3 不同核函数对建模的影响

核函数	参数	支持向量	训练时间（s）	预测误差MSE
Poly	$d=3$	78	26.3	1.1506
RBF	$\sigma=5$	49	35.7	1.0725
Sigmoid	$v=0.0001$	86	31.4	1.6407
Spline	—	64	25.6	1.1970

4.4.2 参数的选择

通常支持向量机中的正则参数C和管道参数ε（也叫不敏感参数）对建模性能影响较大，为了得到最优的参数值，计算次数往往很大。由文献[62]可知，若正则参数C的值较小，则模型将欠拟合，反之，模型将过拟合。支持向量的个数与管道大小ε有关，若ε较大，则会减少支持向量的个数，使得解矩阵呈现稀疏性，而不影响支持向量机的性能。但ε对模型的泛化误差有一定的影响，这也是大多数文献没有考虑到的，因此对ε的取值不能按常规方法处理。

采用RBF核函数，并采用不同的正则参数、管道参数ε、宽度系数σ，对得到的训练误差和预测误差统计成表4-4。

结合氧化铝高压溶出工序中数据获取的特性，依据检测到的生产数据在时间顺序上对过程建模的重要性，提出自适应参数支持向量回归方法，重点探讨正则参数和管道值对模型性能的影响。

（一）正则参数 C 的选取

表4-4 参数对建模的影响

正则参数C	管道参数ε	宽度系数σ	训练误差	预测误差
10	0.1	0.05	2.0966	1.7082
10	0.01	0.05	2.0828	1.6687
10	0.001	0.05	2.0769	1.6636
100	0.1	0.1	1.4981	1.3077
100	0.01	0.1	1.5033	1.3428
100	0.001	0.1	1.5077	1.3739
1000	0.1	0.05	1.0088	1.0739
1000	0.01	0.05	1.0330	1.0771
1000	0.001	0.05	1.0521	1.0725
∞	0.1	0.1	1.0711	1.1802
∞	0.01	0.1	1.0844	1.1693
∞	0.001	0.1	1.1005	1.1616

支持向量机的正则风险函数包括正则项和经验误差，正则参数 C 决定了两者的折中。当正则参数为一定值时，支持向量机将为所有的ε不敏感误差分配相同的权重。标准支持向量机（S-SVM）中的经验误差函数为

$$E_{SVM} = C\sum_{i=1}^{k}(\xi_i + \xi_i^*) \tag{4-20}$$

根据高压溶出工序中取样的生产数据具有严重的大滞后特性，若以当前时刻为参考，较近时刻的训练数据相对于较远时刻的训练数据更重要，更能反映最近一段时间的运行情况。因此给较近时刻的训练数据的ε不敏感误差赋以较大的权值，而给较远时刻的训练数据的ε不敏感误差赋以较小的权值。根据这个特性，正则参数 C 采用 sigmoid 型函数进行调整：

$$E_{\mathrm{ASVM}} = \sum_{i=1}^{k} C_i(\xi_i + \xi_i^*) \tag{4-21}$$

$$C_i = C\frac{1}{1+\exp(m-2m\cdot i/k)} \tag{4-22}$$

式中 i 表示数据序列，当 i=k 时，表示最近时刻的训练数据序列；当 i=1 时，表示最远时刻的训练数据序列。m 是控制增长率的参数，C_i是递增正则常数，其值随着由最远的训练数据序列递增到最近的训练数据序列而不断增加。

当$m \to 0$时，则有$\lim_{m\to 0} C_i = C$，此时，所有的训练数据都被赋以相同的权值，且

$E_{\mathrm{ASVM}} = E_{\mathrm{SVM}}$；

1)当$m \to \infty$，则

$$\lim_{m\to\infty} C_i = \begin{cases} 0, & i < k/2 \\ 0.5C, i = k/2 \\ C, & i > k/2 \end{cases}$$

此时赋给前面一半的训练数据，即较远时刻数据的权值为0，而赋给后面一半的训练数据，即较近时刻数据的权值为1。

2)当$m \in [0,\infty]$，随着 m 不断增大，前面一半的训练数据的权值不断减小，而后面一半的训练数据的权值则不断增大。

（二） 管道参数ε的选取

为了使支持向量机的解矩阵更稀疏，ε采用如下指数型函数调整：

$$\varepsilon_i = \varepsilon \frac{1+\exp(n-2n\cdot i/k)}{2} \tag{4-23}$$

式中 i 表示数据序列，n 是控制减小率的参数，ε_i是递减的管道值，其值随着由最远的训练数据序列递增到最近的训练数据序列而不断减少。

从建模精度和解矩阵的特性来看，相对于较远时刻的训练数据，所提出的自适应参数ε应赋给较近时刻的训练数据更多的权重。在支持向量机中，较小的ε对应于较大的松弛变量ξ_i或ξ_i^*及较高的辨识精度。若松弛变量较大，则经验误差对正则项的影响就会很大。因此，具有较小ε值的最近时刻的训练数据将会得到比具有较大ε值的最远时刻的训练数据更多的惩罚。在回归问题中，支持向量就是那些落在决策函数ε边界上或边界外的训练数据点，当ε值增大时，支持向量的个数就会逐渐减少。那些ε值小的较近时刻的数据点将以更大的概率收敛到支持向量。

关于ε的表达式(4-23)可以分为如下三种情况：

1)当$n \to 0$，则有$\lim_{n\to 0} \varepsilon_i = \varepsilon$，此时，所有的训练数据的权值等于1；

2)当$n \to \infty$，则

$$\lim_{n\to\infty} \varepsilon_i = \begin{cases} \infty, & i < k/2 \\ \varepsilon, & i = k/2 \\ 0.5\varepsilon, i > k/2 \end{cases}$$

此时前面一半的训练数据的权值趋向于无穷大，而后面一半的训练数据的权值则等于0.5。

3)当$n \in [0,\infty]$，随着 n 不断增大，前面一半的训练数据的权值不断增大，而后面一半的训练数据的权值则不断减小。

因此，自适应支持向量回归中的正则风险函数由下式计算。

$$\min \frac{1}{2}\|w\|^2 + \sum_{i=1}^{k} C_i(\xi_i + \xi_i^*) \tag{4-24}$$

约束为：

$$y_i - w\cdot\phi(x_i) - b \le \varepsilon_i + \xi_i$$
$$w\cdot\phi(x_i) + b - y_i \le \varepsilon_i + \xi_i^*,\ \ i = 1,\cdots,k$$

则(4-24)式的对偶函数为

$$\max\left[\sum_{i=1}^{k} y_i(a_i - a_i^*) - \sum_{i=1}^{k}\varepsilon_i(a_i + a_i^*) - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}(a_i - a_i^*)(a_j - a_j^*)K(x_i, x_j)\right] \tag{4-25}$$

约束为：

$$\sum_{i=1}^{k}(a_i - a_i^*) = 0,\ \ 0 \le a_i, a_i^* \le C_i,\ \ i = 1,\cdots,k$$

对于本文所提出的自适应参数支持向量机仍然采用序列最小优化方法求解，此外应注意到每一个训练数据点的上界值C_i是不同的。

4.4.3 基于LS-SVM的软测量模型

在氧化铝高压溶出工艺过程中，苛性比值是非常重要的经济技术指标。根据实际的苛性比值，才能实时的调整原矿浆配料及高压溶出工况，从而实现苛性比值的优化控制、达到节能降耗的目的。由于工况条件限制，无法在线测量苛性比值，所以实际生产中引入了软测量技术，以实现对苛性比值的在线测量，满足实际生产的需求。李勇刚采用基于PCA的多神经网络软测量模型[59]和分布式SVM建立苛性比值的软测量模型[63]，达到了较好的预测效果，但模型结构复杂、训练时间长。本文基于LS-SVM的思想，结合了LS-SVM的特征提取[61]和非线性回归能力，建立了LS-SVM的软测量模型。

首先，利用 LS-SVM 对输入样本数据集进行特征提取，得到样本简洁、信息全面的输入；然后采用回归LS-SVM逼近实际对象。软测量模型结构图如下：

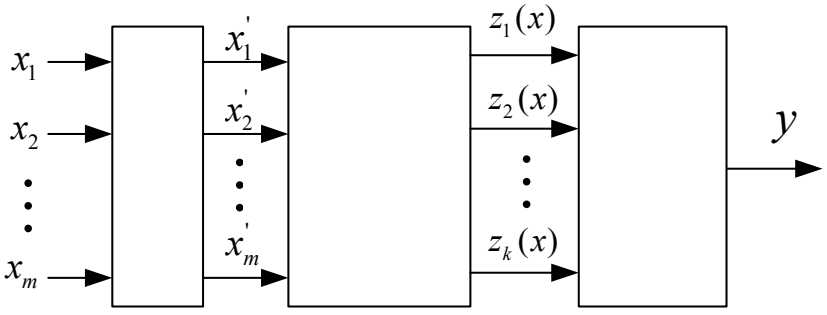


图 4-2 基于LS-SVM的软测量模型

基于LS-SVM的软测量模型算法实现的具体步骤为：

Step 1：根据式（4-11）对样本数据集X进行零均值化，消除不同量纲的影响；

Step 2：根据式（4-16）求取λ（由大到小排列）及对应的特征向量；

Step 3：根据式（4-19）求取最佳分类特征，这样就将输入样本数据集由25维降低为7维，大大简化了模型的输入信息；

Step 4：经特征提取得到重组后的输入样本数据集$\widetilde{X}$，作为回归LS-SVM的输入，取软测量模型的输出值$\overline{Y}$为实际输出值，即实际离线测量得到的苛性比值，建立软测量模型；

Step 5：选取管道参数ε，管道参数ε反映了LS-SVM对数据噪声水平的预测，噪声较大时，应该选用较大的ε，噪声较小时，应该选用较小的ε。但管道参数ε过小，参与回归的支持向量将增多，可能导致过拟合，ε过大, 则可能造成欠拟合；

Step 6：选取惩罚因子C，训练误差一般都会随着惩罚因子C的增大而单调下降，当C增至一定值后，这种下降幅度会变得很小，甚至趋于零；

Step 7：选取核函数$K(\cdot,\cdot)$为RBF函数，并选取$\sigma=0.075$，当宽度系数σ很小时，支持向量之间的联系比较松弛（距离小于σ的支持向量之间才存在联系），学习机器相对复杂，泛化推广能力较差。反之，σ太大，支持向量间的影响过强，回归模型难以达到足够的精度，容易产生欠拟合；

Step 8：根据LS-SVM拟合模型式（3-42）对输出进行非线性回归，得到软测量模型的预测输出，并进行比较。

4.5 仿真结果分析

在氧化铝实际生产过程中，每隔2小时对原矿浆及溶出矿浆进行1次化学分析，测得实际苛性比值。从氧化铝厂收集了一年多来实际运行数据共3000余条，为保证数据的可靠性，通过对这些数据认真分析和处理，剔除一些不合理的样本（由于某些仪表及人为误差，其中一些数据与实际值存在很大误差）。从中选取了900条数据样本（严格依照时间顺序），其中800条用于建立数学模型，另外100条用于检验模型精度。

利用不同的方法进行仿真见图4-3、图4-4、图4-5、图4-6（其中RSA表示苛性比值，无量纲；LR表示溶出率），结果如下：

（1） BP神经网络软测量模型仿真结果

由仿真图可以看出，传统的BP神经网络软测量模型预测结果有尖点，训练过程中陷入了局部极小点，不能保证达到全局最优，这样也在一定程度上影响了

预测的精度，预测效果相对较差，而且训练时间也不理想，整个训练过程耗时117.23S，训练时间比较长。

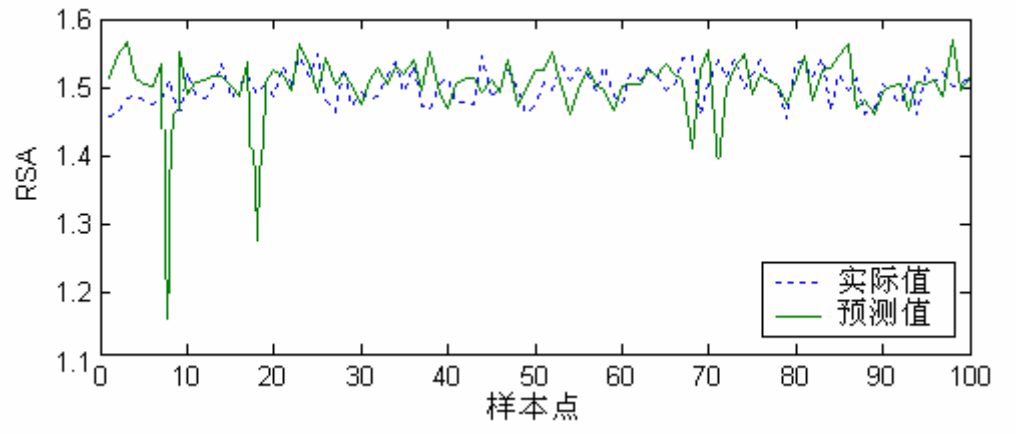


图 4-3（a） BP 神经网络模型 RSA 预测结果

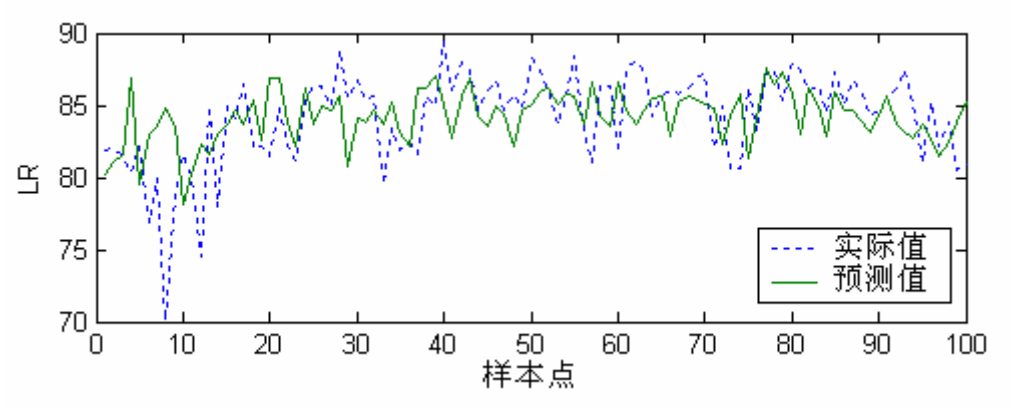


图 4-3（b） BP 神经网络模型 LR 预测结果

（2）基于 PCA 的 LS-SVM 回归模型仿真结果

数据预处理部分采用主元分析的方法，回归建模部分采用 LS-SVM 的方法建立新的模型，经主元分析后，提取 25 个因素中的主元因素，得到 15 维样本，降低了样本维数。同时，根据 4.4 节，选取正则参数（也叫惩罚因子）$C=100$，管道参数ε=0.01，核函数$K(\cdot,\cdot)$为 RBF 函数，并选取$\sigma=0.075$。

由仿真结果可以看出，基于 PCA 的 LS-SVM 预测模型精度较高，预测误差控制在-0.03 与+0.02 之间，预测过程中没有出现过尖点，整个训练过程没有陷入局部极小点，较好地实现了全局最优，从而也证实了 3.4.2 节的结论，体现了 LS-SVM 方法的优越。和 BP 神经网络软测量预测模型相比，不仅提高了预测精度，而且也降低了训练时间，整个训练耗时为 32.80S。

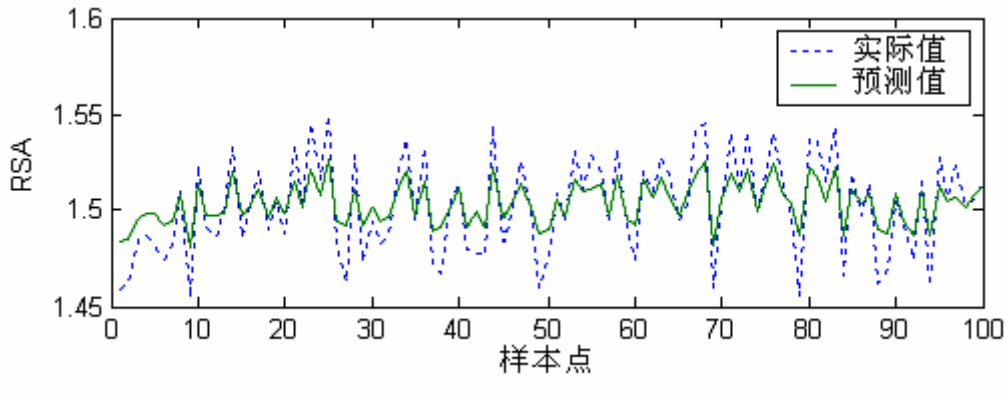


图 4-4（a） 基于 PCA 的 LS-SVM 回归模型 RSA 预测结果

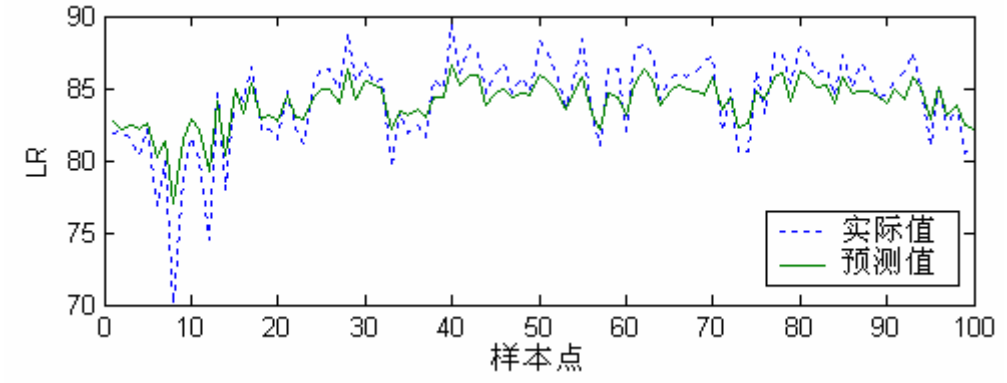


图 4-4（b） 基于 PCA 的 LS-SVM 回归模型 LR 预测结果

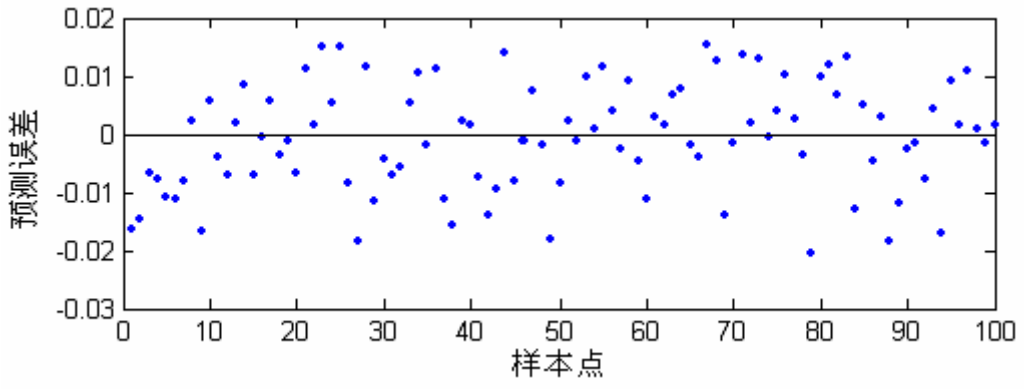


图 4-4（c） 基于 PCA 的 LS-SVM 回归模型预测误差

（3）基于 LS-SVM 特征提取的 LS-SVM 回归模型仿真结果

数据预处理部分采用 LS-SVM 特征提取的方法，回归建模部分采用 LS-SVM 回归的方法建立新的模型，经 LS-SVM 特征提取后，将输入样本由 25 维降低为 7 维，同时，根据 4.4 节，选择正则参数（也叫惩罚因子）$C=100$，管道参数 ε=0.01，核函数 $K(\cdot,\cdot)$ 为 RBF 函数，并选取 $\sigma=0.075$。

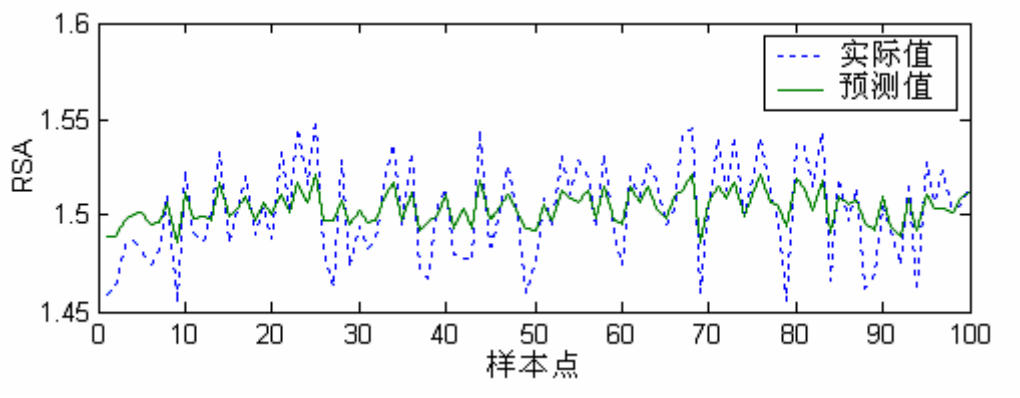


图 4-5（a） 基于特征提取的 LS-SVM 回归模型 RSA 预测结果

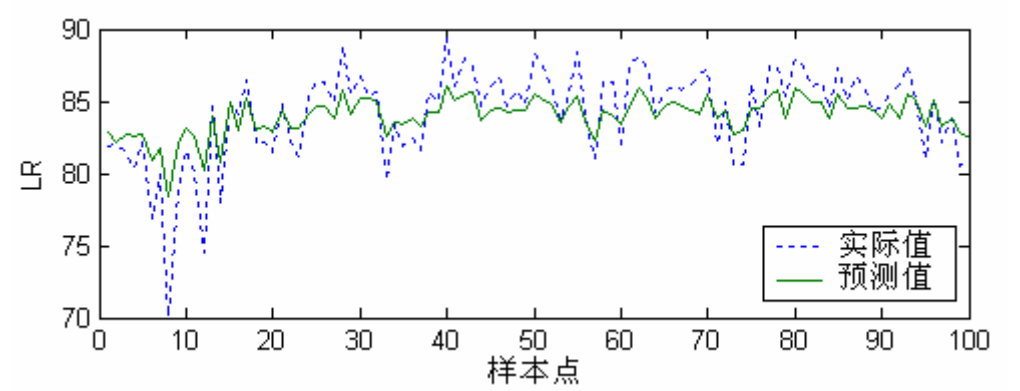


图 4-5（b）基于特征提取的 LS-SVM 回归模型 LR 预测结果

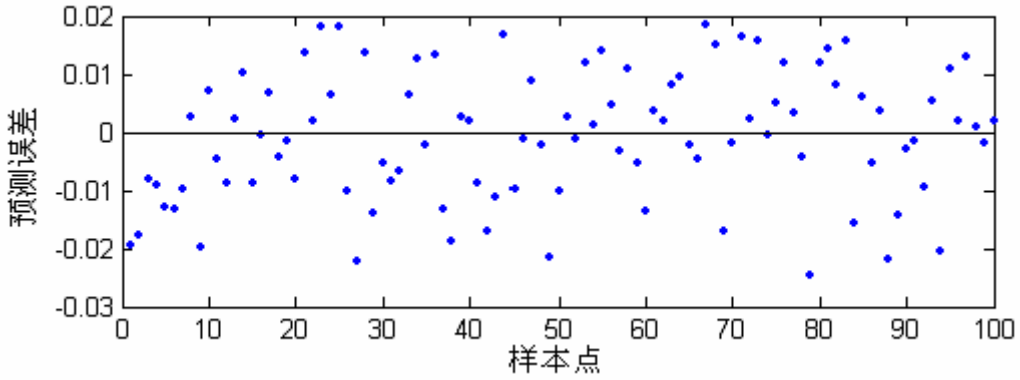


图 4-5（c） 基于特征提取的 LS-SVM 回归模型预测误差

由仿真结果可以看出，基于 LS-SVM 特征提取的 LS-SVM 预测模型精度较高，预测误差也控制在-0.03 与+0.02 之间，同基于 PCA 的 LS-SVM 预测模型精度基本相当，但从苛性比值的预测误差仿真图上可以看出，一些样本点更加接近误差上下边界，说明尽管两种预测模型的精度基本相当，但基于 LS-SVM 特征提取的 LS-SVM 预测模型精度还是要稍逊一些，但是，却较大地提高了训练速度，整个训练过程耗时为 18.32S。

（4）基于特征提取的自适应参数 LS-SVM 回归模型仿真结果

数据预处理部分采用 LS-SVM 特征提取的方法，回归建模部分采用自适应参数的 LS-SVM 回归方法建立新的预测模型，经 LS-SVM 特征提取后，将输入样本由 25 维降低为 7 维，选核函数 $K(\cdot,\cdot)$ 为 RBF 函数，并选取 $\sigma = 0.075$。同样根据 4.4 节，采用自适应的方法确定正则参数（也叫惩罚因子）C，管道参数 ε。

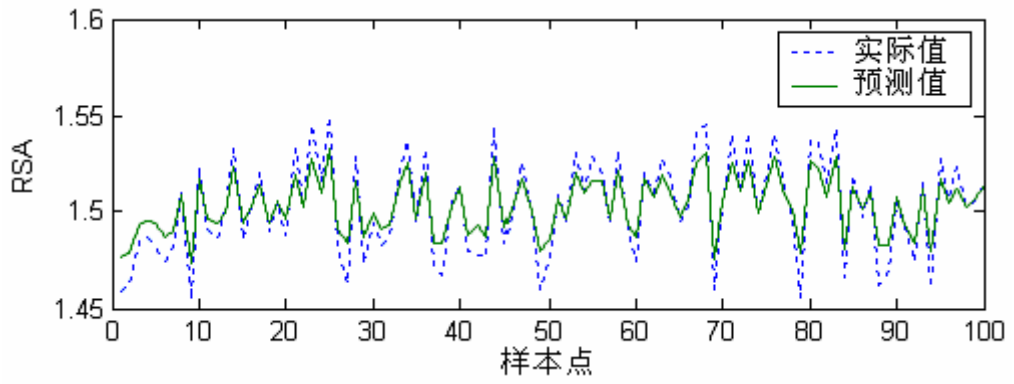


图 4-6（a） 基于特征提取的自适应参数 LS-SVM 回归模型 RSA 预测结果

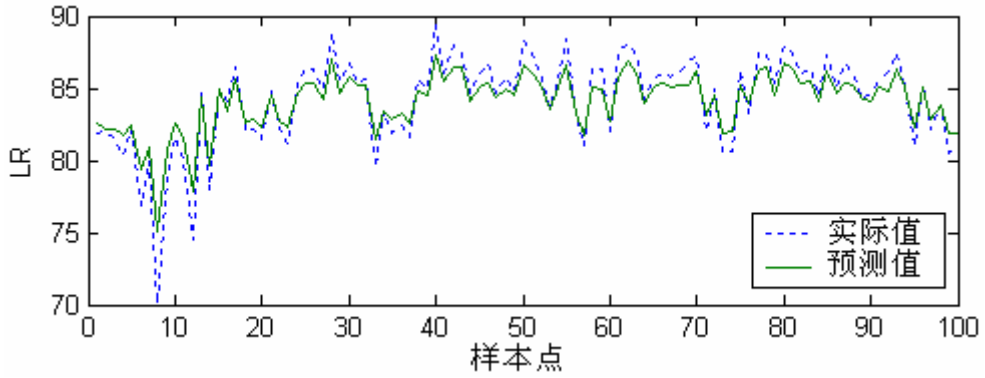


图 4-6（b） 基于特征提取的自适应参数 LS-SVM 回归模型 LR 预测结果

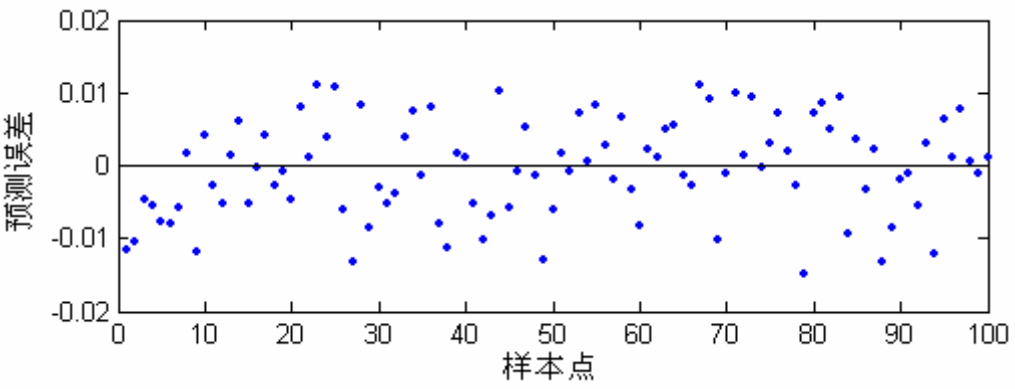


图 4-6（c） 基于特征提取的自适应参数 LS-SVM 回归模型预测误差

由仿真结果可以看出，基于 LS-SVM 特征提取的自适应参数 LS-SVM 预测

模型精度较高，预测误差控制在-0.02 与+0.02 之间，预测精度优于前面的几种预测模型，其训练时间也较短，介于 PCA 的 LS-SVM 回归模型与 LS-SVM 特征提取的 LS-SVM 回归模型之间，训练时间比较合理，整个训练过程耗时为 23.58S。

（5）不同模型的比较(见表 4-5)

表 4-5 不同模型的比较

模　型	预测精度	标准方差	训练时间
BP 神经网络	91.78%	0.0708	117.23
基于 PCA 的 LS-SVM	97.86%	0.0339	32.80
基于特征提取的 LS-SVM	97.51%	0.0351	18.32
基于特征提取的自适应参数 LS-SVM	98.53%	0.0297	23.58

通过比较可知，传统的 BP 神经网络软测量模型预测结果有尖点，和回归 LS-SVM 模型相比，精度差，训练时间长； LS-SVM 特征提取法较主元分析法数据处理能力更强，因此基于特征提取的 LS-SVM 回归模型在预测精度相当的情况下，训练时间较基于 PCA 的 LS-SVM 回归模型时间短；基于特征提取的自适应参数 LS-SVM 回归模型，在训练时间相当的情况下，确保了较高的预测精度。

4.6 小结

LS-SVM 具有很强特征提取能力，能有效的降低样本维数，为模型提供样本简洁、信息全面的输入；回归 LS-SVM 具有良好的非线性函数逼近能力。本文提出的软测量建模方法结合了两者的优点，同时，在模型的建立过程中，针对正则参数 C 和管道参数ε对软测量建模性能影响较大的实际，提出了自适应参数的方法，工业实例仿真表明，基于特征提取的自适应参数 LS-SVM 软测量模型具有良好的预测效果。

第五章 结论与展望

高压溶出是氧化铝生产中极为重要的一个环节。作为其中两个重要经济技术指标，苛性比值与溶出率的在线检测是实现高压溶出过程优化控制的必要条件。然而，在氧化铝生产中，由于反应机理复杂，非线性度高、耦合严重、干扰大，无法实现苛性比值与溶出率的在线检测；目前苛性比值与溶出率的检测是通过化学分析实现的，存在很大的滞后，严重影响了其优化控制。因此，根据实际生产情况，采用合理建模手段，建立适用的软测量模型非常关键。

论文通过分析影响苛性比值与溶出率的相关因素及各因素之间的关系，在进行数据预处理，消除数据冗余性及相关性、降低模型复杂度的基础上，采用新型的机器学习方法——SVM 来建立苛性比值与溶出率的数学模型，仿真结果表明，具有很好的效果。

SVM 是由 Vapnik 等人在 SLT 的基础上提出来的，不但有很好的理论基础，同时具有通用性、鲁棒性和有效性的特点，是现代人工智能中最活跃的领域之一。

本文把 SVM 的一个重要变形 LS-SVM 引入氧化铝高压溶出苛性比值与溶出率的建模之中，得到了较满意的结果。通过本文的研究，得到以下的结论：

1) 把 SVM 用于氧化铝高压溶出苛性比值与溶出率的建模，扩展了 SVM 的应用范围，为解决苛性比值与溶出率建模乃至整个有色冶金过程控制建模问题提供了一个有效的思路。

2) 探讨了数据预处理特别是数据降维的方法，比较了因素分析法、主元分析法、LS-SVM 特征提取法的处理结果，证实了 LS-SVM 特征提取法较强的数据处理能力，为解决数据降维问题做了有意义的尝试，拓宽了数据降维的方法与手段。

3) 详细的论述了参数对建模的影响，深入的讨论了参数的选择问题，充分的结合了 LS-SVM 较强的特征提取和非线性回归能力，综合的提出了基于特征提取的自适应参数 LS-SVM 回归模型。

由于氧化铝高压溶出是一个复杂的工业过程，其过程检测技术并不成熟，把 LS-SVM 用于苛性比值与溶出率的建模，在 Matlab 平台上仿真取得了很好的效果，但是，要把建模应用于实际工业过程，还需要进一步做大量的工作：

1) 在优化 LS-SVM 算法、克服 LS-SVM 缺点上下功夫。如 LS-SVM 对噪声点和异常点比较敏感，模糊 LS-SVM 算法就能有效的提高抗噪能力；LS-SVM 尽管算法简练、结构简单、计算速度快，但却丧失了标准 SVM 的松散性和鲁棒性，加权 LS-SVM 就能较好地克服奇异点对 LS-SVM 造成的影响，提高 LS-SVM 的

鲁棒性。

2) 本着通用、开放和可移植的特点开发系统软件。该软件不仅能够实现对溶出过程苛性比值和溶出率的预测与优化控制，还可适用于整个有色冶炼行业以至复杂工业过程。

3) 软测量技术应用在氧化铝高压溶出过程的目的，就是为了实现对生产过程的自动优化控制，提高氧化铝高压溶出过程的经济效益，这将是下一步研究的重点。

参 考 文 献

[1] H.N. 叶列明等著，王延明，扬重愚等译. 氧化铝生产过程与设备[M]. 北京：冶金工业出版社，1987，4：178-187

[2] 《联合法生产氧化铝》编写组. 联合法生产氧化铝—基础知识[M]. 北京：冶金工业出版社，1975，6：56-80

[3] 国家发展和改革委员会.节能中长期专项规划[J].有色冶金节能，2005,22(2):1-8

[4] 国家发展和改革委员会.能源节约与资源综合利用“十五”规划[J].有色冶金节能,2005,22(3):1-4

[5] 中国有色金属工业行业协会，中国节能技术政策大纲，2004 年 10 月

[6] 汤 伟，施颂椒. 软测量技术及其在抄纸过程中的应用[J].自动化仪表，2000，21(10)：1-3

[7] 徐敏，俞铸.软测量技术[J].石油化工自动化,1998,12(2):1-4

[8] 于静江，周春晖.过程控制中的软测量技术[J].控制理论与应用，1996,13(2):138-142.

[9] 杨欣荣，凌玉华. 软测量技术及其在铝电解槽温度测量中的应用. 中南工业大学学报（自然科学版），2003，34(5)：551-554.

[10] Mcacvoy TJ. Contemplative stance for chemical process control. Automatica，1992，28(2)：441-442.

[11] 李勇刚. 基于智能集成模型的苛性比值与溶出率软测量及应用研究：[博士学位论文].长沙：中南大学，2004

[12] 韩璞，王东风，瞿永杰. 基于神经网络的火电厂烟气含氧量软测量[J]. 信息与控制，2001，30(2)：189-192

[13] 王宏伟，贺汉根，黄柯棣. 基于模糊竞争学习的非线性系统自适应模糊建模方法[J]. 控制理论与应用，2003，20(2):249-252

[14] Shuixiang Li, Minggao Qiao. A hybrid expert system for finite element modeling of fuselage frames. Expert Systems with Applications, 2003, 24(1):87-93

[15] 陆荣秀. 支持向量机技术及其应用[J]. SCI-TECH INFORMATION DEVELOPMENT & ECONOMY，2006，16(14)：159~160

[16] V.Vapnik,S.Golowich,A.Smola.Support vector method for function approximation regression estimation and signal peocessing[J], NIPS ' 96,1996,1(1):57-71

[17] Scholkoph B, Smola A J, Bartlett P L. New support vector algorithms[J]. Neural

computation，2000, (12):1207-1245

[18] 孙德山，吴今培. LS-SVM 在混沌时间序列预测中的应用[J].微机发展,2004,1(1):21-24

[19] Tay FEH, Cao LJ. Modified support vector machines in financial time series forecasting[J]. Neurocomputing, 2002, (48)847-861

[20] 袁小芳，王耀南. 一种模糊支持向量机控制器的研究[J]. 控制与决策，2005, 20(5):537-540

[21] B.Scholkopf. Support Vector Learning[M]. Munich: R.Oldenbourg Verlag, 1997:51-54

[22] J.Shawe-Taylor et al. Structural risk minimization over data-dependent hierarchies. IEEE Trans. Information Theory, 1998,44(5):1926-1940

[23] R.llerbrich and T.Graepel. A PAC-Bayesian margin Bound for linear Classifiers:Why SVMs work[J]. In Advances in Neural Information Processing System ，2001，13(1):224-230

[24] John C. Platt. Using Analytic QP and sparseness to speed training of support vector machines[J]. In Advances in Neural Information Processing Systems，2001, 11(1):0550-0557

[25] Tjoachims. Making large-Scale SVM learning practice. In B.Scholkof, et ,al Editors, Advances in kernel Methods-Support Vector Learning. MIT Press, 1998.

[26] Chih-Jen Lin, etc. The analysis of decomposition methods for support vector machines In Proceeding of IJCAI99[J].SVM Workshop, 1999,1(1):1-12

[27] Chih-Jen Lin, etc. on the convergence of the decomposition method for support vector machines[J]. IEEE Trans Neural Networks, 2001, 4(3):1288-1298

[28] 边肇祺，张学工. 模式识别[M]. 北京：清华大学出版社，2000，4：17-33

[29] E.Osuna, R.Freund, Training support Vector Machines:An Application to Face Detection, Proc.of Computer Vision and Pattern Recognition, San Juan, Puerto Rico, IEEE Computer soc, 1997:130-136

[30] 马勇，丁晓青.基于层次型支持向量机的人脸检测[J].清华大学学报（自然科学版,2003,43(1):35-38

[31] 叶航军，白雪生，徐光祐.基于支持向量机的人脸姿态判定[J].清华大学学报,2004,43(1):67-70

[32] 凌旭峰，杨杰，叶晨州.基于支持向量机的人脸识别技术[J].红外与激光工程,2001,30(5):318-322

[33] 段立娟，崔国勤，高文等.多层次特定类型图像过滤方法[J].计算机辅助设计

与图形学学报,2002,14(5):1-6

[34] 张磊，林福宗，张钹.基于支持向量机的相关反馈图像检索算法[J].清华大学学报,2002,42(1):80-83

[35] A.Smola, B.Scholkopf. A Tutorial on Support Vector Regression, 1998, Technical Report Neuro COLT NC-TR-98-030

[36] K.R.Muller, A.Smola, V.Vapnik, etc. Predicting Time Series with support Vector Machines[J]. ICANN, 1997,9(7):999-1004

[37] Fei Luo, Yuge Xu, JianZhong Cao.Elevator Traffic Flow Prediction with Least S quares Support Vector Machines[J].Proceedings of the 4 th International Conference on Machine Learning and Cybernetics, 2005,8(4):4266-4270

[38] Ailing Ding, Xingmo Zhao, Licheng Jiao. Traffic Flow Time Series Prediction Based on Statistics Learning Theory[J]. the IEEE 5 th International Conference on Intelligent Transportation Systems, 2002,9(5):727-730.

[39] Lelitha Vanajakshi, Laurence R.Rilett, A Comparison of the Performance of Artificial Neural Networks and Support Vector Machines for the Prediction of Traffic Speed, IEEE Intelligeng Vehicles Symposium, University of Parma, Param, Italy, 2004:194-198

[40] T.B.Trafalis, H.Ince. Support Vector Machines for Regression and Applications to Financial Forecasting[J]. IJCNN,2000,8(5):348-353

[41] Dingzhou Cao, Sulin Pang,etc. Forecasting Exchange Rate Using Support Vector Machines[J].Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, 2005,9(4):3448-3452.

[42] V.Vapnik, S.Golowich, A.Smola. Support Vector Methord for Function Approximation, Regression Estimation, and Signal Processing, in Advances in Neural Information Processing Systems 9, Cambridge, MA:MIT Press, 1997:281-287.

[43]《联合法生产氧化铝》编写组编，联合法生产氧化铝：高压溶出[M]. 北京：冶金工业出版社，1974：37-110

[44] 张学工译. 统计学习理论的本质[M]. 北京:清华大学出版社，2000:5-60

[45] Vapnik V, Levin E, Le Cun Y. Measuring the VC-dimension of a learning machine. Neural Computation, 1994, (6):851-876

[46] Usama Fayyad. A Tutorial on support vector Machines for Pattern Recognition. CHRISTOPHER J.C BURGES, Bell Laboratories, Lucent Technologies. Data Mining and Knowledge Discovery, 2,121-167(1998)

[47] 邓乃扬, 田英杰. 数据挖掘中的最优化方法-支持向量机[M]. 北京：科学出版社，2004：31-75

[48] Alexander J.Smola, Learning with Kernels[M].Berlin,1998

[49] Bernhard S.SVM and keneral methods[M], Tubingen(Germany): available from www.learning-with-kernel.org,2002:1-3

[50] Scholkopf B, Simard P, Smola A J. Prior Knowledge in Support Vector Kernels. [C]. Advanced in Neural Information Processings Systems. MA:MIT Press, 1998.640-646

[51] Cristianini N, Shawe-Taylor J. Kernel Methods for Pattern Recognition[M]. Cambridge: Cambridge University Press, 2004:121-124

[52] Suykens J.A.K., Vandewalle J.De Moor B.Optimal control by least squares support vector machines [J].Neural Networks, 2001,5(14):23-35

[53] 张猛，付丽华，张维.一种新的最小二乘支持向量机算法[J].Computer Engineering and Applications,2007,43(11):33-35

[54] Suykens J.A.K., Vandewalle J.Least syuares support vector machine classifiers[J].Neural Processing Letters, 1999,9(3):293-300.

[55] Suykens J.A.K., De Brabanter J., L., Vandewalle J. Weighted least squares support vector machines:robustness and sparse approximation. Neurocomputing, 2002,48(1-4):85-100.

[56] 俞金寿.先进控制系统[J].化学世界,2000,13(1)：25-30

[57] 李勇刚，桂卫华，陈峰. 基于因素分析法的复合神经网络及其在软测量中的应用[J]. 信息与控制，2004，33(2):141-144

[58] 王京茹，李平康，郑宏伟. 主元分析法在火电厂过程控制中的应用. 仪器仪表学报，2004，25(4)：1016~1017

[59] 李勇刚，桂卫华，胡燕瑜. 基于 PCA 的多神经网络软测量模型及其在工业中的应用[J]. 小型微型计算机系统，2004，25(10):1781-1784

[60] WOLD S, ESBENSEN K, GELASI P. Principl Components Analysis[J]. Chemometrics and Intelligent Laboratory Systems,1987,12(2):37-46

[61] 吴德会. 一种基于 LS-SVM 的特征提取新方法及其在智能质量控制中的应用[J]. 计算机应用，2006，26(10):2446-2449

[62] V.Cherkarsky, Y.Ma.Practical selection of SVM parameters and noise estimation for SVM regression[J]. Neural Networks, 2004,17(1): 113-126

[63] 桂卫华，李勇刚，阳春华，陈志盛. 基于改进聚类算法的分布式SVM及其应用[J]. 控制与决策，2004，19(8):852-856

致 谢

值此论文完成之际，我谨向众多给我指导、帮助、关心和鼓励的老师、同学、朋友和家人表示衷心的感谢。

首先感谢导师桂卫华教授，整个课题的研究和论文的撰写都得到了导师的悉心指导。导师严谨的治学态度、执着饱满的工作热情、追求创新和卓越的精神时刻激励我奋发进取，并将是我永远学习的榜样。

感谢阳春华教授、喻寿益教授、唐朝晖副教授、谢永芳副教授和王雅琳副教授的指导和帮助，他们给我提出了很多宝贵的意见，给我以很大的启迪。

特别感谢李勇刚博士后为课题打下了坚实的基础；特别感谢王凌云博士、张艳存硕士在论文撰写过程中对我的具体指导和极大帮助。同时感谢科研室的彭晓波、鄢锋、黄灯等同门师兄弟的帮助。

最后，衷心感谢我原单位的领导、同事以及我的家人对我求学的支持、鼓励和关心。

李 学 军

2007 年 11 月于中南大学

攻读硕士学位期间主要研究成果

参加项目：

国家自然科学基金重点项目：“面向节能降耗的有色冶金过程控制若干理论与方法研究”（60634020）

发表论文：

[1] 李学军，桂卫华，张艳存，王凌云。基于LS-SVM的软测量模型及其工业应用。计算机测量与控制，2008，3(已录用)