

基于支持向量回归的三元溶液浓度 重构算法研究

魏 国, 刘 昕, 孙金玮, 孙圣和

(哈尔滨工业大学自动测试及控制系, 黑龙江哈尔滨 150001)

摘 要: 由氯化钠、蔗糖组成的三元水溶液在食品渗透脱水过程中有着广泛的应用. 本文采用一种新颖的多功能传感器, 在线敏感三元溶液的超声波传输速度、电导率以及温度三个参量. 基于获取的三个可测参量, 本文重点研究利用支持向量回归方法实现多功能传感器信号重构, 同时获得三元溶液中各成分的浓度估计值. 支持向量机是基于结构风险最小化准则的新型机器学习方法, 适用于小样本标定数据情况, 可有效抑制过拟合问题并改善泛化性能. 实验结果显示测试数据集氯化钠浓度和蔗糖浓度的平均绝对误差分别为 0.00615 mol/kg 和 0.00369 mol/kg , 表明所提出的重构算法具有较高的可靠性和准确度, 验证了方法的有效性.

关键词: 支持向量回归; 多功能传感器; 信号重构; 三元溶液; 浓度

中图分类号: TN98 文献标识码: A 文章编号: 0372-2112 (2008) 04-0695-06

Support Vector Regression Based Reconstruction Algorithm for Concentrations of Ternary Solution

WEI Guo, LIU Xin, SUN Jirwei, SUN Shenghe

(Department of Automatic Measurement and Control, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China)

Abstract: Ternary solution with NaCl and sucrose is widely employed in the osmotic dehydration process of food. In this paper, a novel multifunctional sensor was adopted to online sense temperature, ultrasonic velocity and electrical conductivity of ternary solution. With these three measurable parameters as basis, this study laid emphasis on the support vector regression (SVR) method used to implement multifunctional sensor signal reconstruction, where the concentrations of two components in ternary solution with NaCl and sucrose can be simultaneously estimated. Support vector machine (SVM) is a new machine learning method based on structural risk minimization, which is well adapted to small sample size problem of calibration data and can efficiently restrain over fitting and improve the generalization capability. The experimental results show that the mean absolute errors of the reconstructed concentrations of NaCl and sucrose for test data set are 0.00615 and 0.00369 mol/kg , respectively. It could demonstrate the high reliability and high accuracy of the proposed reconstruction algorithm and verify the feasibility.

Key words: support vector regression; multifunctional sensor; signal reconstruction; ternary solution; concentration

1 引言

在食品(主要是果品和蔬菜)的加工和处理过程中, 与传统的风干和冷冻等脱水处理方法相比, 渗透脱水处理技术无论从提高产品质量还是从经济角度看, 都具有很大的优越性和发展潜力^[1]. 可以肯定的是, 浓度是影响物质迁移速率和产品质量的重要物理参数(其它过程变量还包括渗透介质成分、密度、粘度、温度、压力、搅拌、样品形状尺寸、溶液样品体积比、浸入时间等). 研究

表明^[2], 相对于二元溶液(盐或糖的水溶液)而言, 三元溶液(盐和糖的混合水溶液, 通常为氯化钠、蔗糖混合水溶液)会脱出更多的水且伴随着更少量的固体物析出, 被广泛应用于食品的渗透脱水处理. 因而, 在线监测该三元溶液中各成分的浓度非常有意义, 有助于深入研究物质迁移规律, 保证处理过程正常进行, 进而提高产品质量. 此外, 一旦测量获得三元溶液各成分的浓度, 便可通过经验公式计算得到三元溶液的其它两个重要物理参数—粘度和密度^[3].

收稿日期: 2007-03-12; 修回日期: 2007-11-06

基金项目: 国家自然科学基金(No. 60772007, 60672008); 教育部留学回国人员科研启动基金(No. BAQQ24403602); 中国博士后科学基金(No. 20070410258)

现实情况是,对二元溶液中单一溶质浓度测量问题的研究非常充分并有很多商品化传感器出现.对三元或三元以上溶液各溶质浓度的测量问题,由于多被测量(本研究中表现为两种溶质的浓度和温度)同时交叉作用于传感器,使得传统测量方法难以应用或导致测量准确度降低.而一类多函数型多功能传感器具有多输入多输出特性,可同时敏感多个物理、化学参量,通过巧妙利用传感器的交叉敏感特性来实现多参数测量任务,提高测量准确度.多功能传感器具有体积小、功耗低、多敏感的特点,在环境参量和一般工业测量^[4,5]中已经得到了一定程度的开发应用,并在航空、航天、微机械制造以及食品加工等领域具有重要意义和广阔的应用前景.

一般而言,多功能传感技术包含两方面内容:可敏感多个参量的多功能传感器设计和基于多功能传感器输出信号的被测量重构算法研究^[6].当被测量同时由三个或三个以上参量(表现为多功能传感器的多个输出信号)决定时,其函数关系通常呈高维非线性状态,导致被测参量的逆变求解空间都在三维或三维以上.在多功能传感器标定时,通过多次实验所获取的实验数据相对多功能传感器输入输出空间是一个小样本集合,如何实现小样本标定数据情况下的信息值分离和重构是多功能传感器信号处理亟待解决的关键问题.这方面的研究有助于合理减少标定样本,优化标定策略.目前,针对多功能传感器信号重构算法的深入研究已经开展^[7,8],但这些方法普遍采用的是经验风险最小化准则,该准则在样本数量足够大时满足经验风险趋近于真实风险;而当样本数量较少时,最小化经验风险并不能够保证最小化真实风险,因而易导致泛化性差和过拟合等问题.支持向量机方法^[9]是适用于小样本情况的新机器学习方法,它利用结构风险最小化准则来代替传统的经验风险最小化准则,该方法体现了在对给定数据逼近精度和逼近函数复杂度之间的折中思想,因而更适应小样本情况下的学习问题.

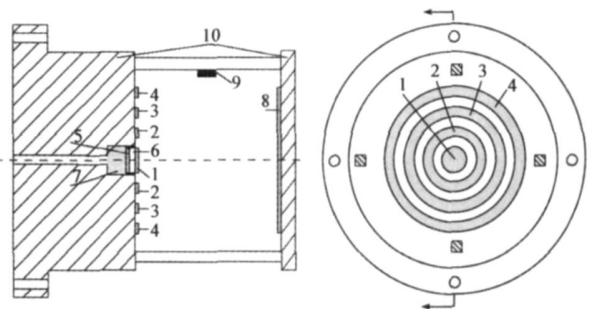
本文采用一种新颖的多功能传感器,可直接在线敏感三元溶液的超声波传输速度、电导率以及温度三个参量.文中重点讨论三元溶液中两种溶质浓度的重构算法.该方法基于多功能传感器获取的三个可测参量并利用支持向量回归构造非线性函数模型,实现小样本情况下的多功能传感器信号重构,可同时得到三元溶液中各成分的浓度估计值.

2 传感器原理、信号重构原理及算法简介

2.1 多功能传感器结构设计

三元溶液浓度测量多功能传感器的结构设计如图1所示^[6],其特点在于超声波敏感部件的上表面电极同时又用作四电极电导率敏感部件的一个电极.这一新颖

设计将超声波敏感功能和电导率敏感功能巧妙地结合起来,简化了传感器设计,使之在结构上更加紧凑.此外,考虑到三元溶液中超声波传输速度和溶液电导率受溶液温度影响较大,在传感器中嵌入一只热敏电阻用以检测溶液实际温度.超声波敏感部件工作于脉冲回波方式:超声波敏感部件经电脉冲激励后发出超声波,发射超声波经由被测三元溶液到达反射板产生回波信号,对回波信号激发超声波敏感部件产生的电信号进行检测即可获得超声波在溶液中的渡越时间,进而得到声速;电导率敏感部件工作于四电极方式:一对电势电极和一对电流电极的应用,可有效地克服由双电层效应引起的测量误差,同时可在很大程度上降低由于杂质淀积引起的测量误差,提高传感器的抗污染能力.



1: 银电极(电压电极1); 2: 不锈钢电极(电流电极1); 3: 不锈钢电极(电流电极2); 4: 不锈钢电极(电压电极2); 5: 银电极; 6: 压电陶瓷超声波敏感部件; 7: 吸声材料; 8: 不锈钢反射板; 9: 热敏电阻; 10: 丙烯酸树脂基底

图1 多功能传感器结构

上述多功能传感器可直接在线敏感三元溶液温度 T 和三元溶液的另外两个物理参量,即与超声波传输速度成反比的渡越时间(Time of flight, ToF) t 以及与溶液电导率成正比的输出电压(Output voltage) V_o . 对于由水、氯化钠和蔗糖所组成的三元溶液,渡越时间 t 和输出电压 V_o 与氯化钠浓度 C_N 、蔗糖浓度 C_S 以及温度 T 之间的非线性函数关系可表述为:

$$\begin{cases} t = f(C_N, C_S, T) \\ V_o = g(C_N, C_S, T) \end{cases} \quad (1)$$

此时,主要问题就是研究如何利用合适的算法获得氯化钠浓度和蔗糖浓度的估计值,这一过程也称作传感器信号重构.

2.2 用于回归的支持向量机

对于回归问题,即研究 n 维输入向量与标量输出之间的函数关系问题,可表述为:

$$y = f(x) + v \quad (2)$$

其中, y 为函数的标量输出, x 为 n 维输入向量, v 为统计独立于输入向量的噪声项,函数 $f(\cdot)$ 和噪声 v 的统计特性未知.此时回归问题可表示为,利用一组训练数据 $\{(x_i, y_i)\}_{i=1}^N$ (x_i 为 n 维输入向量的一个样本值, y_i 为相应的标量输出值)来估计标量输出与 n 维输入变

量之间的函数关系。

对于线性回归问题, 可用如下的线性回归超平面来估计方程(2): $f(x, w) = w^T x + b$ (3)

令 $X = [1, x]^T$ (x 为 $n+1$ 维输入向量), $W = [w_0, w]^T$ (W 为 $n+1$ 维权值向量), 则方程(3)可改写为:

$$f(X, W) = W^T X \quad (4)$$

其中, 权值 w_0 表示方程(3)中的偏置 b 。估计近似误差采用 ϵ 不敏感损失函数表示:

$$L(y, f(x, a)) = \begin{cases} 0, & \text{if } |y - f(x, a)| \leq \epsilon \\ |y - f(x, a)| - \epsilon, & \text{else} \end{cases}$$

此时, 支持向量回归问题就是在 ϵ -不敏感损失函数意义下最小化经验风险:

$$R_{\text{emp}} = \frac{1}{N} \sum_{i=1}^N |y_i - f(X_i, W)|_{\epsilon}$$

并满足如下约束条件:

$$\|W\|^2 = W^T W \leq c$$

其中 c 为常数。支持向量回归试图同时优化经验风险以及 $\|W\|^2$ 项, 它体现了对于给定数据点的逼近精度和逼近函数的复杂度之间的一种折中, 回归超平面可通过如下最优化问题得到:

$$\begin{aligned} \min_{W, \xi, \xi^*} J(W, \xi, \xi^*) &= \frac{1}{2} W^T W + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{s. t.} \quad &\begin{cases} y_i - W^T X_i \leq \epsilon + \xi_i \\ W^T X_i - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (5)$$

其中, ξ_i 和 ξ_i^* 为松弛变量 (slack variable), 它们可看作是回归误差对于 ϵ -不敏感区域的偏离程度的度量。正参数 C 控制对超出误差 ϵ 的样本的惩罚度, 由用户设定。通过构造 Lagrange 函数并应用对偶定理可得到如下对偶最优化问题:

$$\begin{aligned} \max_{\alpha, \alpha^*} Q(\alpha, \alpha^*) &= \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) \\ &\quad - \frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) X_i^T X_j \\ \text{s. t.} \quad &\sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0; 0 \leq \alpha_i, \alpha_i^* \leq C \end{aligned} \quad (6)$$

通过计算拉格朗日乘子 α_i 和 α_i^* , 可以得到回归超平面的最优权值:

$$W_{\text{opt}} = \sum_{i=1}^N (\alpha_i - \alpha_i^*) X_i \quad (7)$$

ϵ -不敏感损失函数的引入使得解具有稀疏性, 即最优化计算得到的 α_i 和 α_i^* 只有少数满足 $\alpha_i - \alpha_i^* \neq 0$ (多数为 0), 它们对应的样本点就是支持向量。此时最优回归超平面为:

$$f(X, W) = W^T X = \sum_{i=1}^N (\alpha_i - \alpha_i^*) X_i^T X_i \quad (8)$$

对于非线性回归, 支持向量回归通过以下步骤实现: 利用非线性映射 $\Phi(\cdot): R^n \rightarrow H$ 将输入 n 维空间的样本映射到高维(可能是无穷维)的特征空间中, 并在特征空间中构造最优回归超平面。通过观察方程(6)至(8)可知, 支持向量回归算法仅依赖于原向量空间中的点积, 因此, 如果能够找到一个函数使得 $K(X_i, X_j) = \Phi^T(X_i) \Phi(X_j)$, 则在高维空间实际上只需进行内积运算, 这种内积运算又可通过原空间中的函数运算实现而无须了解非线性映射的具体形式^[10]。此时, 非线性支持向量回归问题可重述如下:

$$\begin{aligned} \max_{\alpha, \alpha^*} Q(\alpha, \alpha^*) &= \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) \\ &\quad - \frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(X_i, X_j) \\ \text{s. t.} \quad &\sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0; 0 \leq \alpha_i, \alpha_i^* \leq C \end{aligned} \quad (9)$$

相应的非线性回归函数为:

$$f(X, W) = W^T X = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(X, X_i) \quad (10)$$

其中 $K(X, X_i)$ 应为满足 Mercer 定理定义的内积核函数, 常用的主要有^[11]: 多项式核函数、径向基核函数和感知器核函数等。

2.3 最优参数选择

由优化问题(9)可知, 支持向量回归的泛化性能主要依赖于三个自由参数的初始设置, 即参数 C 、 ϵ 和核参数。其中, 正则化参数 C 体现了模型复杂度和经验风险之间的折中, 当 C 过大时(如无穷), 由式(5)可知回归问题趋向于最小化经验风险而不考虑模型复杂度; 参数 ϵ 控制用于计算经验风险的 ϵ -不敏感区域的宽度, 它的选取影响构成回归函数式(10)中的支持向量的个数, 较大的 ϵ 会导致较少的支持向量; 核参数则控制模型的复杂度, 如对于多项式, 当指数 p 增长时, 其模型复杂度也随之增加^[12]。由于这三个参数都对支持向量回归的泛化性能有重要的影响, 因此需同时调整。针对训练数据集, 给定一个参数集合 $\{C, \epsilon, \sigma\}$ (以径向基为例), 对于集合中的每组参数, 利用交叉验证方法求取回归函数式(10), 并通过比较不同参数设置下的泛化性能选取最佳参数设置。交叉验证流程为: 对给定数据集 D , 将其随机分为 k 个不同的子集 $\{E_1, \dots, E_k\}$, 利用 $k-1$ 个子集训练回归函数, 并用余下的子集作为测试, 这一过程重复 k 次, 泛化性能利用所有子集上的均方差评价, 交叉验证的一个特例即为留一法 ($k = N$), 它适用于小数据集情况。

2.4 算法简介

对于方程(1)中的函数关系, 可认为每一组被测参量 $[C_N(i), C_S(i), T(i)]$ 都有且只有一组传感器输出信

号 $[t(i), V_o(i)]$ 与之相对应(若不然,则无法实现被测量的测量).此时,方程(1)所示的函数关系可改写为:

$$\begin{cases} C_N = f^*(t, V_o, T) \\ C_S = g^*(t, V_o, T) \end{cases} \quad (11)$$

对此多功能传感器系统,经过 N 次标定实验可得 N 组数据集合:

$$(t(k), V_o(k), T(k), C_N(k), C_S(k)) \quad k = 1, \dots, N$$

此时,多功能传感器信号重构问题转化为对非线性方程(11)的近似估计问题.由前文所述的支持向量回归方法,可基于这 N 组数据构成式(9)所示的最优化问题,通过求解该最优化问题即可得到相应的拉格朗日乘子,进而可分别得到形如式(10)所示的氯化钠浓度及蔗糖浓度近似估计方程:

$$\begin{cases} C_N(k) = f_{sv}^*(t(k), V_o(k), T(k)) \\ C_S(k) = g_{sv}^*(t(k), V_o(k), T(k)) \end{cases} \quad (12)$$

此时,将测量得到的渡越时间 t 、输出电压 V_o 和温度 T 带入到上式即可得到相应的氯化钠浓度 C_N 及蔗糖浓度 C_S 的重构值.

3 实验结果及讨论

3.1 实验数据产生

用于标定多功能传感器的样本溶液由蒸馏水、氯化钠和蔗糖配比构成,配比精度为 $\pm 0.0001\text{g}$,其中氯化钠

Calibration solutions in molality(NaCl,Sucrose: mol/kg, mol/lg):

- -0.6,0.3; ▲ -0.6,0.6; ■ -0.6,0.9; ▼ -0.6,1.2; ◆ -0.6,1.5
- -1.2,0.3; ▲ -1.2,0.6; ■ -1.2,0.9; ▼ -1.2,1.2; ◆ -1.2,1.5
- -1.8,0.3; ▲ -1.8,0.6; ■ -1.8,0.9; ▼ -1.8,1.2; ◆ -1.8,1.5
- -2.4,0.3; ▲ -2.4,0.6; ■ -2.4,0.9; ▼ -2.4,1.2; ◆ -2.4,1.5
- -3.0,0.3; ▲ -3.0,0.6; ■ -3.0,0.9; ▼ -3.0,1.2; ◆ -3.0,1.5

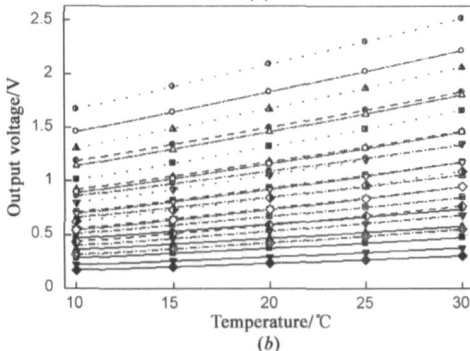
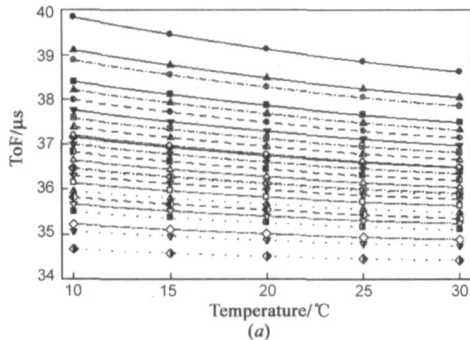


图2 多功能传感器输出与溶液温度关系

浓度可用集合 $A = \{0.6 \times i(\text{mol/kg}) \mid i = 1, \dots, 5\}$ 表示,蔗糖浓度可用集合 $B = \{0.3 \times i(\text{mol/kg}) \mid i = 1, \dots, 5\}$ 表述,共计得到 25 个标定试样.标定实验在温度 $T = \{5 + 5 \times i(\text{°C}) \mid i = 1, \dots, 5\}$ 下分别进行,因此可测得共 125 组实验数据.实验数据如图 2 所示,其中图 2(a) 为三元溶液的声速相关渡越时间-温度关系曲线,图 2(b) 为三元溶液的电导率相关输出电压-温度关系曲线.

由图 2 可见,随着温度增加,渡越时间 t 减小而输出电压 V_o 增大.当蔗糖浓度固定时,渡越时间 t 随氯化钠浓度增加而增加,当氯化钠浓度固定时,渡越时间 t 随蔗糖浓度增加而增加;此现象可从氯化钠和蔗糖分子与水发生氢和作用时会导致更加刚性致密结构方面得到解释.当蔗糖浓度固定而氯化钠浓度增加时,输出电压 V_o 非线性增加;此现象易于从 Na^+ 与 Cl^- 离子对溶液提供导电能力途径理解.当氯化钠浓度固定而蔗糖浓度增加时,输出电压 V_o 非线性减小;此现象是由溶液粘度增加和介电常数减小共同造成的.

3.2 信号重构结果及讨论

在所有的 125 组实验数据中随机选取 100 组作为训练数据,余下的 25 组数据作为测试数据,并将所有数据经预处理归一化到 $[-1, 1]$ 区间(求得重构数据须经相应后处理).对于训练数据采用支持向量回归方法及交叉验证训练可得如(12)所示的近似回归方程,参数集合为 $C = \{1, 10, 100, 1000\}$ 、 $-\log_{10}(\epsilon) = \{1, \dots, 6\}$ 以及 $\sigma = \{0.1 \times i \mid i = 1, \dots, 30\}$ 三者的直积,共计 720 组参数.再将测试数据代入所求得的支持向量回归方程,则可得对应被测参量 $\{C_N(j), C_S(j)\}_{j=1}^{25}$ 的估计值 $\{C_N^*(j), C_S^*(j)\}_{j=1}^{25}$,参数集合中的最佳参数选择可通过比较平均绝对误差(Mean absolute error, MAE)、平均相对误差(Mean relative error, MRE) 和均方差(Mean square error, MSE) 进行选取.以氯化钠浓度重构为例,不同 C 值下的氯化钠浓度的训练数据重构均方误差如图 3 所

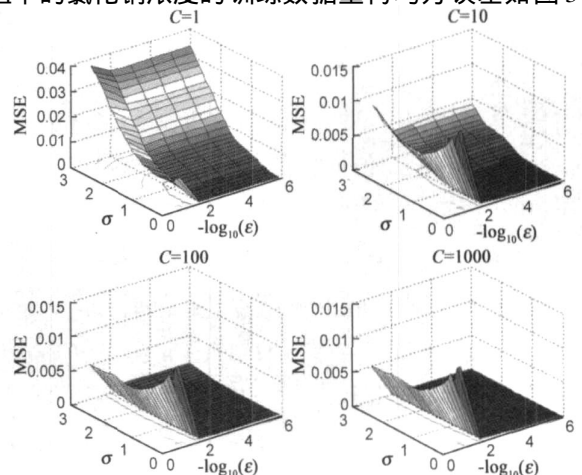


图3 不同 C 值下氯化钠重构均方差分布图

示, 图 4 为其相应的等高线填充图。

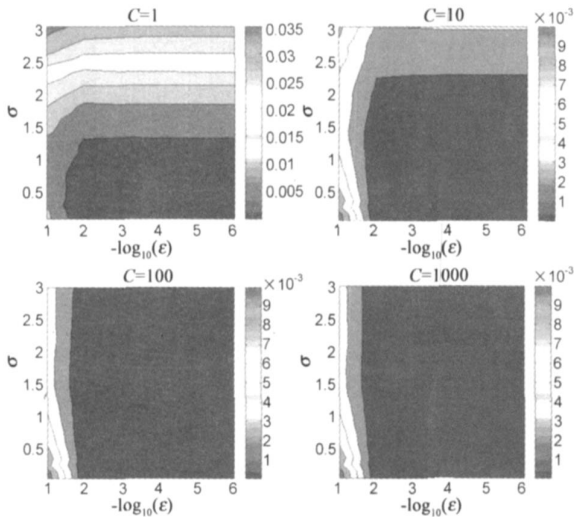


图 4 不同 C 值下氯化钠重构均方差等高线填充图

由图 3 可知, 当参数 C 显著增长时, 均方差有缓慢下降的趋势, 因此可认为参数 C 对于性能的影响较为不敏感; 由图 4 可知, 当 C 显著增长时, 最优化参数所在区域面积相应增大, 其中参数 σ 随之增长, 而参数

表 1 不同参数设置下氯化钠浓度重构误差指标

| 参数设置 | | | 训练数据 | | | 测试数据 | | |
|----------|--------------------|----------------|------------|------------|------------|------------|------------|------------|
| | | | MAE | MRE | MSE | MAE | MRE | MSE |
| C = 1 | $\epsilon = 0.001$ | $\sigma = 0.4$ | 8.49 e- 03 | 5.82 e- 03 | 3.86 e- 04 | 6.48 e- 02 | 3.86 e- 02 | 8.86 e- 03 |
| C = 10 | $\epsilon = 0.001$ | $\sigma = 1.0$ | 6.75 e- 03 | 5.22 e- 03 | 1.17 e- 04 | 1.65 e- 02 | 1.10 e- 02 | 4.46 e- 04 |
| C = 100 | $\epsilon = 0.001$ | $\sigma = 1.1$ | 2.96 e- 03 | 2.41 e- 03 | 2.43e- 05 | 8.64 e- 03 | 6.62 e- 03 | 1.07 e- 04 |
| C = 1000 | $\epsilon = 0.001$ | $\sigma = 1.4$ | 2.57 e- 03 | 2.13 e- 03 | 1.56e- 05 | 6.15 e- 03 | 4.86 e- 03 | 5.29e- 05 |

表 2 不同参数设置下蔗糖浓度重构误差指标

| 参数设置 | | | 训练数据 | | | 测试数据 | | |
|----------|--------------------|----------------|------------|------------|------------|------------|------------|------------|
| | | | MAE | MRE | MSE | MAE | MRE | MSE |
| C = 1 | $\epsilon = 0.001$ | $\sigma = 0.4$ | 9.05 e- 03 | 9.06 e- 03 | 4.58 e- 04 | 5.66 e- 02 | 5.18 e- 02 | 5.43 e- 03 |
| C = 10 | $\epsilon = 0.001$ | $\sigma = 1.0$ | 4.62 e- 03 | 4.00 e- 03 | 6.46 e- 05 | 1.22 e- 02 | 1.15 e- 02 | 2.72 e- 04 |
| C = 100 | $\epsilon = 0.001$ | $\sigma = 1.1$ | 1.78 e- 03 | 1.39 e- 03 | 9.41 e- 06 | 4.20 e- 03 | 4.11 e- 03 | 2.88e- 05 |
| C = 1000 | $\epsilon = 0.001$ | $\sigma = 1.4$ | 1.40 e- 03 | 1.00 e- 03 | 5.56e- 06 | 3.69 e- 03 | 4.19 e- 03 | 1.97e- 05 |

表 3 标定试样外三元溶液样本浓度重构及误差指标

| 溶液温度 ($^{\circ}\text{C}$) | 实际浓度(mol/kg) | | 重构浓度及误差(mol/kg) | | | |
|--------------------------------|--------------|------|-----------------|--------|--------|--------|
| | 氯化钠 | 蔗糖 | 氯化钠 | 误差 | 蔗糖 | 误差 |
| 12.5 | 0.90 | 1.35 | 0.8931 | 0.0069 | 1.3487 | 0.0013 |
| 17.5 | 0.90 | 1.35 | 0.8903 | 0.0097 | 1.3490 | 0.0010 |
| 22.5 | 0.90 | 1.35 | 0.8921 | 0.0079 | 1.3484 | 0.0016 |
| 27.5 | 0.90 | 1.35 | 0.8894 | 0.0106 | 1.3475 | 0.0025 |
| 12.5 | 2.70 | 0.45 | 2.6954 | 0.0046 | 0.4558 | 0.0058 |
| 17.5 | 2.70 | 0.45 | 2.6972 | 0.0028 | 0.4556 | 0.0056 |
| 22.5 | 2.70 | 0.45 | 2.7046 | 0.0046 | 0.4522 | 0.0022 |
| 27.5 | 2.70 | 0.45 | 2.7082 | 0.0082 | 0.4548 | 0.0048 |

为验证支持向量回归方法对于训练、测试数据集之外的数据点重构性能, 表 3 给出两组不同浓度配比的三元溶液在多个温度下的重构信号及绝对误差, 其中回归函数利用表 1 和表 2 中的最后一组参数设置, 由表

ϵ 所在区间则变化不大 ($-\log_{10}(\epsilon) \in [2, 6]$), 因而参数 σ 对于均方差的影响更大. 此外由于参数 ϵ 直接影响最终支持向量的个数, 因此当样本数量较大时, 可通过选择较大的参数 ϵ 来控制支持向量的个数.

对于氯化钠浓度重构信号和蔗糖浓度重构信号, 不同 C 值下的优化参数设置以及相应的误差指标分别如表 1 和表 2 所示. 由表 1 和表 2 可见, 当参数 C 增大时, 最优化的 ϵ 参数都为 0.001 而参数 σ 随之增长, 同时各项误差指标也随之下降. 因此, 在训练支持向量机时, 参数 C 不应设置过小, 过小的 C 影响最终回归函数的泛化性能; 同时参数 σ 随着 C 的增大而增大, 反之亦然. 通常参数 ϵ 与数据点的噪声水平相关^[13], 由前面讨论结果可知, 当参数 ϵ 小于一定数值时对泛化性能影响较小. 此外, 由表 1、表 2 可知, 利用支持向量回归方法得到的氯化钠和蔗糖浓度的平均训练数据绝对误差分别为 0.00257 mol/kg 和 0.00140 mol/kg, 而平均测试数据绝对误差分别为 0.00615 mol/kg 和 0.00369 mol/kg, 训练数据与测试数据的误差指标均较小且量级相当, 因而该方法具有较高的信号重构准确度.

3 可知, 温度与溶液的浓度配比与训练、测试数据中的数据全不相同, 氯化钠浓度的重构绝对误差最大为 0.0106 mol/kg, 而蔗糖浓度的重构绝对误差最大为 0.0058 mol/kg, 与测试数据集得到的误差指标在相当量级水平, 再次证明了支持向量回归方法具有较高的重构可靠性和准确度.

4 结论

本文采用一个新型的多功能传感器, 可同时敏感三元溶液的温度、超声波传输速度以及电导率三个参量, 通常溶液中各组成成分的浓度可表示为这几个物理参量的函数, 通过函数估计方法即可实现溶液浓度值的重构. 但该函数关系通常呈高维、非线性状态, 因

而实际观测数据相对于多功能传感器输入输出空间而言是一个小样本. 传统的经验风险最小化准则通常是以处理大样本问题出发的, 在小样本情况下易导致泛化性差和过拟合等问题. 因此采用基于结构风险最小化准则的支持向量机方法实现多功能传感器信号重构更为合理. 本文同时利用交叉验证方法分析了支持向量机的参数优化设置问题, 并讨论了不同参数设置对最终的泛化性能的影响. 理论分析和实验结果表明, 支持向量回归方法适用于小样本情况下的多功能传感器信号重构问题, 并且具有较高的信号重构可靠性和准确性.

参考文献:

- [1] A L Raoult Wack. Recent advances in the osmotic dehydration of foods[J]. Trends in Food Science and Technology, 1994, 5(8): 255- 260.
- [2] A M Sereno, et al. Mass transfer coefficients during osmotic dehydration of apple in single and combined aqueous solutions of sugar and salt[J]. Journal of Food Engineering, 2001, 47(1): 43- 49.
- [3] F Chenlo, et al. Viscosities of aqueous solutions of sucrose and sodium chloride of interest in osmotic dehydration processes [J]. Journal of Food Engineering, 2000, 54(4): 347- 352.
- [4] J Sun, K Shida. Multilayer sensing and aggregation approach to environmental perception with one multifunctional sensor[J]. IEEE Sensors Journal, 2002, 2(2): 62- 72.
- [5] T A Eftimov, WJ Bock. A simple multifunctional fiber optic level/moisture/vapor sensor using large core quartz polymer fiber pairs[J]. IEEE Transactions on Instrumentation and Measurement, 2006, 55(6): 2080- 2087.
- [6] G Wei, K Shida. Estimation of concentrations of ternary solution with NaCl and sucrose based on multifunctional sensing technique[J]. IEEE Transactions on Instrumentation and Measurement, 2006, 55(2): 675- 681.
- [7] F Alessandra, et al. Application of an optimal look up table to sensor data processing[J]. IEEE Transactions on Instrumentation and Measurement, 1999, 48(4): 813- 816.
- [8] 孙金玮, 刘昕, 孙圣和. 基于总体最小二乘的多功能传感器信号重构方法研究[J]. 电子学报, 2004, 32(3): 391- 394.
Sun jin wei, Liu Xin, Sun Sheng he. TLS algorithm based study on multi functional sensor data reconstruction [J]. Acta Electronica Sinica, 2004, 32(3): 391- 394. (in Chinese)
- [9] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32- 42.
Zhang Xue gong. Introduction to statistical learning theory and support vector machines[J]. Acta Automatic Sinica, 2000, 26(1): 32- 42. (in Chinese)
- [10] A J Smola, B Schölkopf. A tutorial on support vector regression[J]. Statistics and Computing, 2004, 14(3): 199- 222.
- [11] S Haykin. Neural networks: a comprehensive foundation, 2nd edition[M]. Prentice Hall, 1999.
叶世伟, 史忠植译. 神经网络原理[M]. 北京: 机械工业出版社, 2004.
- [12] C J C Burges. A tutorial on support vector machines for pattern recognition[J]. Knowledge Discovery and Data Mining, 1998, 2(2): 121- 167.
- [13] V Cherkassky, Y Ma. Multiple model regression estimation [J]. IEEE Transactions on neural networks, 2005, 16(4): 785 - 798.

作者简介:



魏 国 男, 1966 年 3 月生于辽宁省开原市. 2003 年 9 月毕业于日本佐贺大学理工学部, 获系统生产科学专攻博士学位. 现为哈尔滨工业大学副教授. 主要研究方向为测控系统、多功能敏感技术和信号处理.

E-mail: wg_ weiguo@yahoo.com.cn



孙金玮 男, 1964 年 10 月生于黑龙江省哈尔滨市, 2000 年 9 月毕业于日本佐贺大学理工学部, 获系统生产科学专攻博士学位. 现为哈尔滨工业大学教授、博士生导师. 主要研究方向为传感技术及其信号处理、测试仪器及系统.



刘 昕 男, 1980 年 2 月生于黑龙江省哈尔滨市. 2004 年 7 月毕业于哈尔滨工业大学自动化测试与控制系, 获仪器科学与技术学科硕士学位. 现为哈尔滨工业大学博士研究生. 研究方向为多功能传感器及其信号重构.



孙圣和 男, 1937 年 10 月生于山东省, 1961 年毕业于哈尔滨工业大学电磁测量专业, 1962 年毕业于同专业研究生班, 现为哈尔滨工业大学自动化测试与控制研究所所长、教授、博士生导师. 主要研究领域为自动化测试、信号与信息处理和光纤传感技术等.