

# 贝叶斯L2型TSK模糊系统

杭文龙<sup>1†</sup>, 梁爽<sup>2</sup>, 刘解放<sup>1</sup>, 王士同<sup>1</sup>

(1. 江南大学 数字媒体学院, 江苏 无锡 214122; 2. 中国科学院深圳先进技术研究院  
广东省机器视觉与虚拟现实技术重点实验室, 广东 深圳 518055)

**摘要:** 针对传统 Takagi-Sugeno-Kan(TSK)模糊系统处理大规模数据时间代价较高的问题,提出一种基于概率模型框架的 L2 型 TSK 模糊系统建模策略,建立具有处理大规模数据能力的贝叶斯 L2 型 TSK 模糊系统(B-TSK-FS). 具体地,基于 L2 型 TSK 模糊系统的输出误差概率化表示,对系统前后件参数联合学习,提高系统的泛化能力. 另外,引入狄利克雷先验分布函数,对模糊隶属度稀疏化表示,实现样本的压缩,降低运算时间. 在模拟和真实数据集上的实验结果验证了所提出模糊系统的优势.

**关键词:** 贝叶斯; L2 型 TSK 模糊系统; 大规模数据; 狄利克雷先验分布

**中图分类号:** TP181

**文献标志码:** A

## Bayesian L2-norm-Takagi-Sugeno-Kang fuzzy system

HANG Wen-long<sup>1†</sup>, LIANG Shuang<sup>2</sup>, LIU Jie-fang<sup>1</sup>, WANG Shi-tong<sup>1</sup>

(1. School of Digital Media, Jiangnan University, Wuxi 214122, China; 2. Guangdong Provincial Key Laboratory of Computer Vision and Virtual Reality Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

**Abstract:** Classical Takagi-Sugeno-Kang(TSK) fuzzy systems take too much training time on the large scale datasets. To overcome this difficulty, the new TSK fuzzy system in probability modeling framework, i.e., Bayesian L2-Norm-Takagi-Sugeno-Kang fuzzy system(B-TSK-FS), is proposed. The B-TSK-FS has the ability to handle large datasets. Specifically, based on the probabilistic output error of the L2-Norm TSK-FS, the B-TSK-FS has better generalization abilities through taking both antecedents and consequents of fuzzy rules into consideration simultaneously. In addition, the Dirichlet prior distribution is introduced to the B-TSK-FS, which makes the fuzzy memberships be represented sparsely, so as to condense the large scale datasets for reducing the running time. The experimental results on the synthetic and real-world datasets show the advantage of the proposed method.

**Keywords:** Bayesian; L2-norm TSK fuzzy system; large scale datasets; Dirichlet prior distribution

## 0 引言

模糊集理论<sup>[1-2]</sup>自从 1965 年被扎德教授等人提出以来,在控制论及其应用<sup>[3-5]</sup>等方面产生了深远的影响. 随着模糊集理论的发展以及模糊推理理论的出现促使了模糊系统的产生. 模糊系统具备将自然语言直接转译为计算机语言的能力<sup>[6-7]</sup>,能够模拟人类在生产与实践过程中的思维、分析、推理及决策过程,可以建立一套与之相对应的数学模型. 由于模糊系统具有高度的可解释性和强大的学习能力,已被广泛应用于数据挖掘、模式识别、信号处理等领域<sup>[8-11]</sup>.

经典的模糊系统模型可归纳为 Takagi-Sugeno-Kang(TSK)型模糊系统<sup>[8,12]</sup>, Mamdani-Larsen(ML)型

模糊系统<sup>[13]</sup>和 Generalized fuzzy(GF)型模糊系统. 在不确定非线性系统逼近方面,模糊系统作为一种非常有效的工具,能够通过一组 If-Then 规则及语义解释法则得到相应的学习模型. 特别地,TSK 模糊系统提供了一个较为完整的学习框架,可以将一个复杂的非线性系统分解为多个局部线性模型,再利用其相应的规则进行学习. TSK 模糊系统的研究与改进,对于解释控制领域中诸多复杂的非线性系统模型<sup>[15]</sup>以及推动其实用化进程具有重要意义. 因此,本文将以 L2 型 TSK 模糊系统为基础,对其建模方法进行讨论. 目前,针对 L2 型 TSK 模糊系统的建模策略及其改进方法被相继提出,大致可分为 3 大类:

收稿日期: 2016-07-03; 修回日期: 2016-09-20.

基金项目: 国家自然科学基金项目(61300151, 61305097); 深圳市基础学科布局项目(JCYJ20160429190300857).

作者简介: 杭文龙(1988—),男,博士,从事模糊系统、数据挖掘等研究;王士同(1964—),男,教授,博士生导师,从事模式识别、人工智能等研究.

<sup>†</sup>通讯作者. E-mail: hwl881018@163.com

1) 调节模糊规则的后件参数. 该建模策略需要事先确定规则库和由不同聚类算法得到的模糊规则前件, 然后通过经典的最小二乘误差<sup>[16]</sup>或反向传播神经网络学习方法<sup>[17]</sup>得到其后件参数.

2) 修剪模糊规则. 该建模策略需要固定前件隶属度函数和后件参数, 然后通过一些优化算法(诸如遗传算法等)选择最佳模糊规则子集<sup>[18]</sup>.

3) 同时学习模糊规则前件和后件. 该方法本质上可看作上述两种方法的结合, 正逐渐成为当下模糊系统建模研究的热点<sup>[19]</sup>. 其建模策略通过同时学习前、后件参数更易找到全局最优解, 保证模糊系统具有较高的识别精度.

目前, 利用神经网络学习算法实现模糊规则前件和后件参数的同时学习是一种切实有效的做法<sup>[20-26]</sup>. 然而, 神经网络参数的逐层学习方式以及大规模数据的出现, 导致该类TSK模糊系统收敛速度慢且模糊规则可解释性差.

近年来, 概率论在模糊系统领域的应用与日益增, 取得了一系列研究成果, 代表性的工作有模糊贝叶斯统计推断<sup>[22]</sup>、贝叶斯模糊聚类<sup>[23]</sup>等. 然而, 目前概率论方法在TSK模糊系统中鲜有涉及. 本文通过深入探讨概率论与L2型TSK模糊系统之间的关系, 提出一种贝叶斯L2型TSK模糊系统, 利用贝叶斯决策方法实现模糊规则前件与后件参数的同时学习, 规避了由于使用神经网络方法导致的收敛速度慢、模糊规则可解释性差等问题. 另外, 本文通过引入狄利克雷先验分布函数, 使得所建立的贝叶斯L2型TSK模糊系统兼具处理大规模数据的能力, 解决了传统的L2型TSK模糊系统在处理大规模数据时引起的训练时间过长的问题.

## 1 L2型TSK模糊系统

L2型TSK模糊系统模糊推理规则 $R^k$ 表示如下:

$$\begin{aligned} & \text{If } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \cdots \wedge x_d \text{ is } A_d^k, \\ & \text{Then } y \text{ is } f^k(\mathbf{x}). \end{aligned} \quad (1)$$

其中 $f^k(\mathbf{x})$ 可以线性表示为

$$f^k(\mathbf{x}) = p_0^k + p_1^k x_1 + \cdots + p_d^k x_d, \quad k = 1, 2, \dots, K. \quad (2)$$

式(1)中, 每条规则都是基于输入样本 $\mathbf{x} = [x_1, x_2, \dots, x_d]$ , 且将输入空间 $A^k \subset R^d$ 中的模糊子集映射到由式(2)表达的不同单例模式中,  $A_i^k$ 对应输入变量 $x_i$ 的第 $k$ 条模糊子集,  $K$ 为规则数,  $\wedge$ 为模糊连接运算符. 经过一系列的运算及去模糊化处理后, 最终得到L2型TSK模糊系统的输出表达式为

$$y^o = \frac{\sum_{k=1}^K \frac{\mu^k(\mathbf{x})}{\sum_{k'=1}^K \mu^{k'}(\mathbf{x})} f^k(\mathbf{x})}{\sum_{k=1}^K \tilde{\mu}^k(\mathbf{x}) f^k(\mathbf{x})}. \quad (3)$$

其中 $\mu^k(\mathbf{x})$ 和 $\tilde{\mu}^k(\mathbf{x})$ 分别对应模糊集 $A^k$ 的模糊隶属度函数及其标准化形式, 表示如下:

$$\mu^k(\mathbf{x}) = \prod_{i=1}^d \mu_{A_i^k}(x_i), \quad (4)$$

$$\tilde{\mu}^k(\mathbf{x}) = \frac{\mu^k(\mathbf{x})}{\sum_{k'=1}^K \mu^{k'}(\mathbf{x})}. \quad (5)$$

若将高斯函数作为隶属度函数, 则式(4)中的 $\mu_{A_i^k}(x_i)$ 可表示为

$$\mu_{A_i^k}(x_i) = \exp\left(\frac{-(x_i - c_i^k)^2}{2\delta_i^k}\right), \quad (6)$$

其中参数 $c_i^k$ 和 $\delta_i^k$ 通过聚类划分方法得到. 若采用模糊C均值聚类算法<sup>[24]</sup>, 则

$$c_i^k = \frac{\sum_{j=1}^N u_{jk} x_{ji}}{\sum_{j=1}^N u_{jk}}, \quad (7)$$

$$\delta_i^k = h \frac{\sum_{j=1}^N u_{jk} (x_{ji} - c_i^k)^2}{\sum_{j=1}^N u_{jk}}. \quad (8)$$

其中:  $u_{jk}$ 为输入样本 $\mathbf{x} = [x_1, x_2, \dots, x_d]$ 隶属于第 $k$ 条规则的隶属度,  $h$ 为人工可调节尺度参数.

如文献[8]所述, 大多数TSK模糊系统的建模策略在完成对前件参数求解后, 对后件参数的学习可以转化为如下线性回归问题<sup>[8]</sup>:

$$y^o = \mathbf{p}_g^T \mathbf{x}_g. \quad (9)$$

其中

$$\begin{aligned} \mathbf{x}_g &= ((\tilde{\mathbf{x}}^1)^T, (\tilde{\mathbf{x}}^2)^T, \dots, (\tilde{\mathbf{x}}^K)^T)^T, \\ \mathbf{p}_g &= ((\mathbf{p}^1)^T, (\mathbf{p}^2)^T, \dots, (\mathbf{p}^K)^T)^T, \\ \mathbf{p}^k &= (p_0^k, p_1^k, \dots, p_d^k)^T, \\ \tilde{\mathbf{x}}^k &= \tilde{\mu}^k(\mathbf{x}) \mathbf{x}_e, \\ \mathbf{x}_e &= (1, \mathbf{x}^T)^T. \end{aligned}$$

## 2 贝叶斯L2型TSK模糊系统

基于传统L2型TSK模糊系统定义的模糊隶属度函数(4)和(5), 式(4)可进一步表示为

$$\tilde{\mu}^k(\mathbf{x}) = \frac{\prod_{i=1}^d \mu_{A_i^k}(\mathbf{x}_i)}{\sum_{k'=1}^K \prod_{i=1}^d \mu_{A_i^{k'}}(\mathbf{x}_i)}. \quad (10)$$

令  $\tilde{\mu}^k(\mathbf{x}) = v^k(\mathbf{x})$ , 则有如下关系成立:

$$\sum_{k=1}^K v^k(\mathbf{x}) = 1. \quad (11)$$

式(3)中的L2型TSK模糊系统模型可重新表示为

$$y^o(\mathbf{x}) = \sum_{k=1}^K v^k(\mathbf{x}) f^k(\mathbf{x}).$$

## 2.1 贝叶斯L2型TSK模糊系统目标函数构造

**定义1** 已知  $\bar{y}(\mathbf{x})$ ,  $y(\mathbf{x})$ ,  $y^k(\mathbf{x})$  分别表示样本  $\mathbf{x}$  的真实输出, 模糊系统输出以及对应第  $k$  条规则下的输出, 则贝叶斯L2型TSK模糊系统模型为

$$\begin{aligned} J(\mathbf{x}, \mathbf{v}, \mathbf{y}) = & \frac{1}{C} \exp \left( - \sum_{n=1}^N \frac{(\bar{y}(\mathbf{x}_n) - y(\mathbf{x}_n))^2}{2} \cdot \sigma^2 \right) = \\ & \frac{1}{C'} \exp \left[ - \sum_{n=1}^N \left[ \bar{y}(\mathbf{x}_n) \sum_{k=1}^K v^k(\mathbf{x}_n) - \right. \right. \\ & \left. \left. \sum_{k=1}^K v^k(\mathbf{x}_n) y^k(\mathbf{x}_n) \right]^2 \right] = \\ & \frac{1}{C'} \exp \left[ - \sum_{n=1}^N \left[ \sum_{k=1}^K v^k(\mathbf{x}_n) \bar{y}(\mathbf{x}_n) - y^k(\mathbf{x}_n) \right]^2 \right]. \end{aligned} \quad (12)$$

其中: 常数  $C = (\sqrt{2\pi}\sigma)^d$ ,  $C' = (\sqrt{2\pi}\sigma)^d \exp(-1/2\sigma^2)$ .

TSK模糊系统的第  $n$  个样本的输出误差为  $(\bar{y}(\mathbf{x}_n) - y(\mathbf{x}_n))$ , 可进一步表示为

$$e_n = \bar{y}(\mathbf{x}_n) - \sum_{k=1}^K v^k(\mathbf{x}_n) y^k(\mathbf{x}_n), \quad (13)$$

其中真实输出  $y(\mathbf{x}_n)$  为常量. 由于数据样本独立同分布, 其在第  $k$  条规则下的输出误差同样满足离散型分布<sup>[26]</sup>, 即具有有限的均值和方差. 中心极限定理<sup>[25-26]</sup>表明, 若大量的独立同分布随机变量存在有限的均值和方差, 则其和可近似于高斯分布. 因此, 当样本数量足够多时, 误差  $e_n$  满足高斯分布. 可以看出, 定义1为传统的L2型TSK模糊系统模型与概率模型之间建立了联系. 上述贝叶斯L2型TSK模糊系统模型由以下几个部分组成.

**定义2** 贝叶斯L2型TSK模糊系统模型样本似然分布(TFDL)可表示如下:

$$p(\mathbf{x}|\mathbf{v}, \mathbf{y}) = \prod_{n=1}^N \text{TFDL}(\mathbf{x}_n|\mathbf{v}(\mathbf{x}_n), \mathbf{y}) =$$

$$\prod_{n=1}^N \frac{1}{C'} \cdot \exp \left[ - \left[ \sum_{k=1}^K v^k(\mathbf{x}_n) [\bar{y}(\mathbf{x}_n) - y^k(\mathbf{x}_n)] \right]^2 \right], \quad (14)$$

其中常量  $C' = (\sqrt{2\pi}\sigma)^d \exp(-1/\sigma^2)$ .

**定义3** 贝叶斯L2型TSK模糊系统隶属度先验分布(TFP)定义如下:

$$\hat{p}(\mathbf{v}|\mathbf{y}) = \prod_{n=1}^N \text{TFP}(\mathbf{v}(\mathbf{x}_n)|\mathbf{y}) = \prod_{n=1}^N C' \cdot \prod_{k=1}^K \text{Dirichlet}(v^k(\mathbf{x}_n)|\alpha^k). \quad (15)$$

将式(15)分解为  $F_2 = \prod_{k=1}^K \text{Dirichlet}(v^k(\mathbf{x}_n)|\alpha^k)$  和  $F_1 = C'$  两项. 第1项的作用是对数据似然分布的反平衡, 即  $F_1$  可以消除TFDL中的规则化常数项. 第2项是如下含参数  $\alpha$  的狄利克雷函数:

$$\text{Dirichlet}(\mathbf{x}|\alpha) = \frac{\Gamma\left(\sum_{k=1}^K \alpha^k\right)}{\prod_{k=1}^K \Gamma(\alpha^k)} \prod_{k=1}^K \chi_k^{\alpha^k-1}. \quad (16)$$

其中:  $\chi_k \geq 0$ ,  $k = 1, 2, \dots, K$ ,  $\sum_{k=1}^K \chi_k = 1$ . 狄利克雷函数的值域为  $\{0, 1\}$  且函数值和为1, 其产生的值符合模糊隶属度函数的要求<sup>[23]</sup>. 因此, 可将其引入模糊系统中, 使得贝叶斯L2型TSK模糊系统隶属度函数自动满足模糊隶属度和为1的约束条件.

**定义4** 贝叶斯L2型TSK模糊系统模型输出原型表示为如下高斯先验分布模型:

$$p(\mathbf{y}) = \prod_{n=1}^N \prod_{k=1}^K \exp((-y^k(\mathbf{x}_n))^2). \quad (17)$$

结合TFDL(14), TFP(15)和输出原型(17), 可得到样本及参数的联合似然估计如下:

$$\begin{aligned} p(\mathbf{x}, \mathbf{v}, \mathbf{y}) = & p(\mathbf{x}|\mathbf{v}, \mathbf{y}) \hat{p}(\mathbf{v}|\mathbf{y}) p(\mathbf{y}) \propto \\ & \exp \left\{ - \frac{1}{2} \sum_{n=1}^N \left[ \sum_{k=1}^K v^k(\mathbf{x}_n) [\bar{y}(\mathbf{x}_n) - y^k(\mathbf{x}_n)] \right]^2 \right\} \times \\ & \prod_{n=1}^N \prod_{k=1}^K v^k(\mathbf{x}_n)^{\alpha^k-1} \exp \left( - \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^K (y^k(\mathbf{x}_n))^2 \right). \end{aligned} \quad (18)$$

因此, 目标函数可以改写为上述联合似然估计负对数形式, 即

$$J(\mathbf{x}, \mathbf{v}, \mathbf{y}) = \sum_{n=1}^N \left[ \sum_{k=1}^K v^k(\mathbf{x}_n) (\bar{y}(\mathbf{x}_n) - y^k(\mathbf{x}_n)) \right]^2 -$$

$$2 \sum_{n=1}^N \sum_{k=1}^K (\alpha^k - 1) \log(v^k(\mathbf{x}_n)) + \sum_{n=1}^N \sum_{k=1}^K (y^k(\mathbf{x}_n))^2. \quad (19)$$

## 2.2 参数学习规则

在本节中,将对式(19)进行参数优化,即求解B-TSK-FS前、后件参数. 本文基于最大后验概率参数方法,迭代求解B-TSK-FS中各样本点隶属度值及后件参数.

对隶属度函数的条件概率分布采样使用如下狄利克雷函数:

$$\mathbf{v}'(\mathbf{x}_n) \sim \text{Dirichlet}(\boldsymbol{\alpha} = \mathbf{1}_K). \quad (20)$$

通过式(20)可得到样本隶属度值. 在目标函数 $J(\mathbf{x}, \mathbf{v}, \mathbf{y})$ 中,固定样本 $\mathbf{x}_n$ 和输出 $\mathbf{y}$ 对隶属度向量 $\mathbf{v}(\mathbf{x}_n)$ 进行最大似然估计,可得到如下后验概率模型:

$$\begin{aligned} p(\mathbf{x}_n, \mathbf{v}(\mathbf{x}_n) | \mathbf{y}) = \\ p(\mathbf{x}_n | \mathbf{v}(\mathbf{x}_n), \mathbf{y}) \hat{p}(\mathbf{v}(\mathbf{x}_n) | \mathbf{y}) = \\ \exp \left\{ -\frac{1}{2} \sum_{n=1}^N \left[ \sum_{k=1}^K v^k(\mathbf{x}_n) [\bar{y}(\mathbf{x}_n) - \mathbf{p}^k \mathbf{x}_n] \right]^2 \right\} \times \\ \prod_{k=1}^K \text{Dirichlet}(v^k(\mathbf{x}_n) | \boldsymbol{\alpha}) \propto \\ \exp \left\{ -\frac{1}{2} \sum_{n=1}^N \left[ \sum_{k=1}^K v^k(\mathbf{x}_n) [\bar{y}(\mathbf{x}_n) \mathbf{p}^k \mathbf{x}_n] \right]^2 \right\} \times \\ \prod_{k=1}^K v^k(\mathbf{x}_n)^{\alpha^k - 1}. \end{aligned} \quad (21)$$

按照下式的概率比值:

$$a_v = \min \left\{ 1, \frac{p(\mathbf{x}_n, \mathbf{v}'(\mathbf{x}_n) | \mathbf{y})}{p(\mathbf{x}_n, \mathbf{v}(\mathbf{x}_n) | \mathbf{y})} \right\} \quad (22)$$

确定是否用新一轮迭代产生的隶属度向量 $\mathbf{v}'(\mathbf{x}_n)$ 替换隶属度向量 $\mathbf{v}(\mathbf{x}_n)$ .

可以看出,已知隶属度向量 $\mathbf{v}(\mathbf{x}_n)$ ,后件参数 $\mathbf{p}^k = (p_0^k, p_1^k, \dots, p_d^k)^T$ 的求解则相对容易. 对式(19)关于 $\mathbf{p}^k$ 求导,可得

$$\begin{aligned} \frac{\partial J(\mathbf{x}, \mathbf{v}, \mathbf{y})}{\partial \mathbf{p}^k} = \\ \frac{\partial \sum_{n=1}^N \left[ \sum_{k=1}^K v^k(\mathbf{x}_n) (\bar{y}(\mathbf{x}_n) - \mathbf{p}^k \mathbf{x}_n) \right]^2}{\partial \mathbf{p}^k} - \\ \frac{\partial \sum_{n=1}^N \sum_{k=1}^K (p^k \mathbf{x}_n)^2}{\partial \mathbf{p}^k} + \frac{\partial g}{\partial \mathbf{p}^k}, \end{aligned} \quad (23)$$

$g$ 等价于 $2 \sum_{n=1}^N \sum_{k=1}^K (\alpha^k - 1) \log(v^k(\mathbf{x}_n))$ . 通过一系列数学推导,可以得到 $\mathbf{p}^k$ 的解析解如下:

$$\begin{aligned} \mathbf{p}^k = \\ \left[ \sum_{n=1}^N \bar{y}(\mathbf{x}_n) v^k(\mathbf{x}_n) \cdot \mathbf{x}_n^T - \sum_{n=1}^N \sum_{k'=1, k' \neq k}^K (v^k(\mathbf{x}_n) v^{k'}(\mathbf{x}_n) \mathbf{p}^{k'} \mathbf{x}_n \mathbf{x}_n^T) \right] M^{-1}, \quad (24) \\ M \text{ 等价于 } \left( \sum_{n=1}^N [v^k(\mathbf{x}_n)]^2 + 1 \right) \mathbf{x}_n \mathbf{x}_n^T. \end{aligned}$$

综上,可以得到贝叶斯L2型TSK模糊系统参数学习方法.

### 算法1 贝叶斯L2型TSK模糊系统.

Step 1: 设置规则数 $K$ ,迭代次数为 $N_{\text{iter}}$ ,  $\mathbf{v}_n \sim \text{Dirichlet}(\boldsymbol{\alpha} = \mathbf{1}_K)$ ,  $n = 1, 2, \dots, N$ ,根据式(23)计算 $\mathbf{p}^k, \mathbf{v}_n^* \leftarrow \mathbf{v}_n, \mathbf{p}_*^k \leftarrow \mathbf{p}^k$ ;

Step 2: 利用式(20)产生新的隶属度向量 $\mathbf{v}_n^+, n = 1, 2, \dots, N$ ;

Step 3: 利用式(21)判断 $p(\mathbf{x}_n, \mathbf{v}_n^+ | \mathbf{Y}^*) > p(\mathbf{x}_n, \mathbf{v}_n | \mathbf{Y}^*)$ ,如若成立,则 $\mathbf{v}_n^* \leftarrow \mathbf{v}_n^+, n = 1, 2, \dots, N$ ;

Step 4: 根据式(22),以概率 $a_v$ 接受 $(\mathbf{v}_n \leftarrow \mathbf{v}_n^+)$ ,  $n = 1, 2, \dots, N$ ;

Step 5: 根据式(24)计算 $\mathbf{p}^k, \mathbf{p}_*^k \leftarrow \mathbf{p}^k$ ;

Step 6: 判断 $p(\mathbf{X}, \mathbf{V} | \mathbf{Y}^*) > p(\mathbf{X}, \mathbf{V}^* | \mathbf{Y}^*)$ ,如若成立,则 $\mathbf{V}^* \leftarrow \mathbf{V}$ ;

Step 7: 重复Step 2 ~ Step 6,直至最大迭代次数 $N_{\text{iter}}$ .

为了获得最大后验概率的全局最优参数,采用蒙特卡罗方法<sup>[23]</sup>. 其中,Step 2利用狄利克雷函数产生新的隶属度向量;Step 3以最大后验概率的方式分别判断是否接受样本新产生的隶属度向量;Step 4以概率方式接受新产生的隶属度向量;Step 5根据接受的隶属度向量计算后件参数;最终,Step 6以最大后验概率的方式判别接受Step 3或Step 4产生的隶属度.

## 2.3 大规模数据学习法则

由式(20)可以看出,当使用狄利克雷函数产生隶属度向量时,不仅能够自动满足隶属度和为1的约束,而且参数 $\alpha$ 值可以控制生成的隶属度值的稀疏性. 如表1所示,通过一组实例来描述狄利克雷函数生成的隶属度值的特性. 可以看出,当 $\alpha \rightarrow 0$ 时(如表中 $\boldsymbol{\alpha} = 0.001 \times \mathbf{1}_K$ ,其中 $\mathbf{1}_K$ 表示 $K$ 维单位向量),模糊隶属度 $\mathbf{v}$ 的值较稀疏;当 $\boldsymbol{\alpha} = 0.01 \times \mathbf{1}_K$ 时,非0隶属度值样本增多;当 $\boldsymbol{\alpha} = 0.1 \times \mathbf{1}_K$ 时,得到更多的非0隶属度值. 以此类推,随着 $\alpha$ 值的增大直至非0隶属度值不再出现,目标函数的计算时间也不断增加. 很显然,隶属度值为0的样本不会对模糊系统建模造成影响,可以将其去掉,从而降低式(21)和(24)的计算复杂度.

表1 狄利克雷分布不同 $\alpha$ 值产生的不同采样特性

| 参数 $\alpha$ 的值              | $v_j$   |       |         |         |         |
|-----------------------------|---------|-------|---------|---------|---------|
|                             | $v_1$   | $v_2$ | $v_3$   | $v_4$   | $v_5$   |
| $0.001 \times \mathbf{1}_K$ | 0       | 0     | 0       | 0       | 1       |
|                             | 0       | 0     | 1       | 0       | 0       |
|                             | 0       | 1     | 0       | 0       | 0       |
|                             | 1       | 0     | 0       | 0       | 0       |
|                             | 0       | 0     | 0       | 1       | 0       |
| $0.01 \times \mathbf{1}_K$  | 0       | 0     | 0       | 0       | 0.000 2 |
|                             | 0       | 0     | 0       | 1       | 0       |
|                             | 0       | 1     | 1       | 0       | 0.999 8 |
|                             | 1       | 0     | 0       | 0       | 0       |
|                             | 0       | 0     | 0       | 0       | 0       |
| $0.1 \times \mathbf{1}_K$   | 0.017 7 | 0     | 0.000 2 | 0.351 3 | 0.537 6 |
|                             | 0.310 5 | 0     | 0.430 6 | 0.001 4 | 0       |
|                             | 0       | 1     | 0       | 0.000 1 | 0.428 8 |
|                             | 0.671 8 | 0     | 0.569 1 | 0.009 7 | 0       |
|                             | 0       | 0     | 0.000 2 | 0.637 5 | 0.033 6 |

通过调节狄利克雷分布中的 $\alpha$ 值,使得贝叶斯L2型TSK模糊系统兼具处理大规模数据的能力.在实际处理过程中,一般会根据样本的大小来设置 $\alpha$ 的值,以达到简化计算的目的.当 $N \leq 10^3$ 时,取 $\alpha = 1 \times \mathbf{1}_K$ ;当 $10^3 < N \leq 10^4$ 时,取 $\alpha = 0.1 \times \mathbf{1}_K$ ;当 $10^4 < N \leq 10^5$ 时,取 $\alpha = 0.01 \times \mathbf{1}_K$ ;当样本个数 $N > 10^5$ 时,取 $\alpha = 0.001 \times \mathbf{1}_K$ .实验表明, $\alpha$ 的4种不同设置分别对应的采样率约为100%,70%,10%,1%.

## 2.4 时间复杂度分析

贝叶斯L2型TSK模糊系统计算模型的时间复杂度主要由3方面确定:对模糊隶属度 $\mathbf{V}^*$ 的采样,对后件参数 $\mathbf{P}^*$ 的求解以及对最大后验概率的计算,因此,贝叶斯L2型TSK模糊系统计算模型的时间复杂度为 $O(T(2N^2K(d+1) + NK^2(d+1)^4))$ .其中: $O(N^2K(d+1))$ 为式(21)中计算隶属度 $\mathbf{V}^*$ 的时间复杂度; $O(NK^2(d+1)^4)$ 为式(24)中计算后件参数 $\mathbf{p}^k(k=1,2,\dots,K)$ 的时间复杂度;计算最大后验概率的时间复杂度为 $O(N^2K(d+1))$ , $T$ 为迭代次数.实验表明,当样本维数较低时( $d \leq 6$ ),计算迭代100次基本满足精度要求,当样本维数 $d > 6$ 时,迭代次数达到1000次能够基本满足要求.

## 3 实验研究

本节将从人工模拟数据集以及真实数据集两个方面,与传统模糊建模方法进行比较,对本文所提出的B-TSK-FS性能进行验证与评估.

## 3.1 实验设置

本文所涉及的对比较算法如下:

- 1) 基于自适应网络的模糊推理系统(ANFIS)<sup>[20]</sup>;
- 2) L1型 $\varepsilon$ 不敏感TSK模糊系统(LSSLI)<sup>[27]</sup>;
- 3) L2型 $\varepsilon$ 不敏感TSK模糊系统(L2-TSK-FS)<sup>[9]</sup>;
- 4) 本文提出的贝叶斯L2型TSK模糊系统(B-TSK-FS).

在实验参数设置中,所有对比算法的模糊规则数均采用5折交叉验证来选择.其中:LSSLI、L2-TSK-FS和B-TSK-FS的规则数的选择范围为{4, 6, 8, 16, 32, 64, 128}.对于ANFIS,其规则数的选择可参考文献[20],即 $2^M$ , $M=2,3,4,5,6,7,8,9,10$ .另外,L2-TSK-FS以及LSSLI采用FCM算法对前件部分空间进行划分.

采用如下评价指标对本文的实验结果进行合理评估<sup>[8]</sup>:

$$J = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2 / N}{\sum_{i=1}^N (y_i - \bar{y})^2 / N}}. \quad (25)$$

$J$ 越小,表明算法的泛化能力越好.式(25)中, $N$ 为样本个数, $y_i$ 为第 $i$ 个样本的真实输出, $\hat{y}_i$ 为模糊系统输出,且 $\bar{y} = \left(\frac{1}{N}\right) \sum_{i=1}^N y_i$ .

实验环境:实验硬件平台为Windows 32位4 Intel Core i3,内存为2 GB.编程环境为Matlab 2012.

## 3.2 实验结果分析

### 3.2.1 模拟数据集

本节将在人工模拟数据集上验证所提出算法的性能,采用混沌时间序列对数据集模拟.具体地,通过如下Mackey-Glass混沌时滞微分方程<sup>[28]</sup>产生:

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t). \quad (26)$$

参数设置参照文献[28],取 $\tau=30$ .在实验过程中,分别生成5个不同大小的时序数据集,样本个数分别为数据集1:500;数据集2:1000;数据集3:2000;数据集4:4000;数据集5:8000以及数据集6:16000.从每组数据集中随机抽取40%的数据用于训练,剩下的数据用于测试.

本文以数据集1为例,给出在该数据集上4种不同算法的逼近效果,如图1所示.在其余的数据集上各算法的运行时间及泛化性能比较如表2所示,其中“/”表示运行时间较长,本文中不予给出.

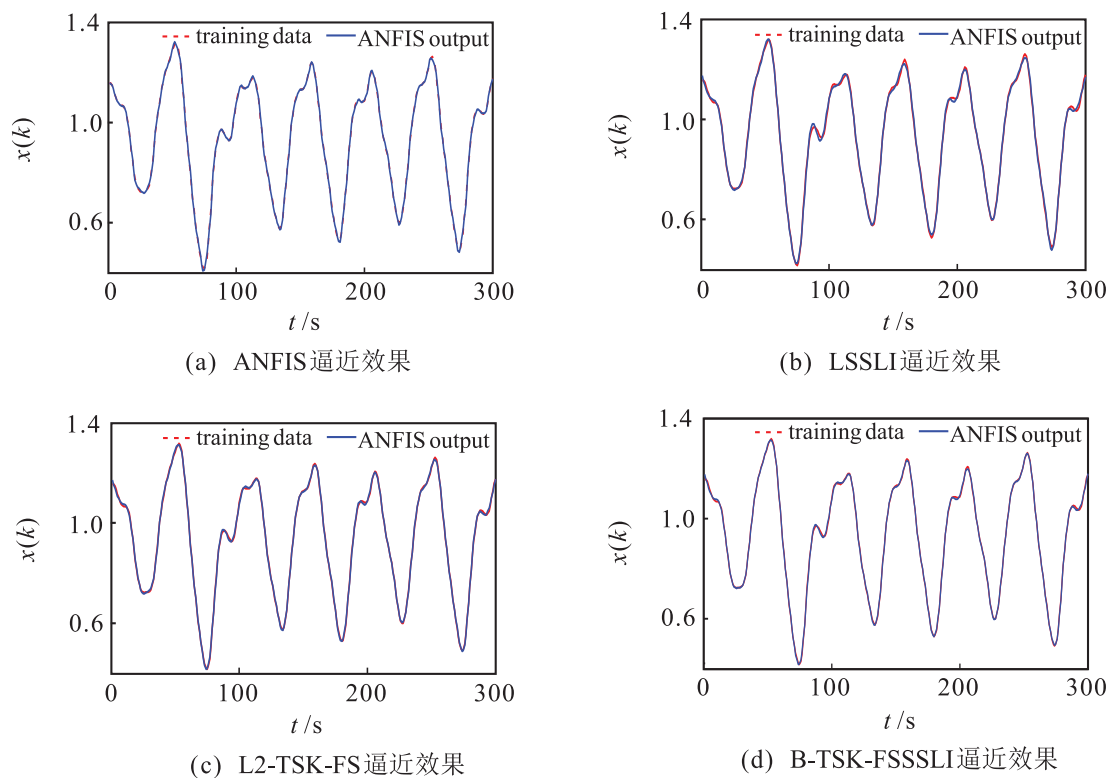


图1 4种模糊系统建模算法在数据集1上的逼近效果

表2 4种模糊系统建模算法在模拟数据集上的运行时间及泛化性能比较

| Datasets | Performance                                                    | ANFIS   | LSSLI   | L2-TSK-FS | B-TSK-FS |
|----------|----------------------------------------------------------------|---------|---------|-----------|----------|
|          | Running Time ( $M = 2^6$ ) ( $M = 8$ ) ( $M = 8$ ) ( $M = 8$ ) |         |         |           |          |
| 1        | $J$                                                            | 0.021 6 | 0.024 2 | 0.023 1   | 0.021 2  |
|          | $t/s$                                                          | 9.182 5 | 0.312 4 | 3.870 1   | 1.219    |
| 2        | $J$                                                            | 0.025 8 | 0.027   | 0.027 9   | 0.023 5  |
|          | $t/s$                                                          | 23.093  | 1.081   | 10.234    | 3.562 5  |
| 3        | $J$                                                            | 0.070 7 | 0.074 1 | 0.073 7   | 0.071 4  |
|          | $t/s$                                                          | 52.25   | 21.003  | 55.921    | 7.375    |
| 4        | $J$                                                            | 0.273   | 0.276 1 | 0.280 1   | 0.278 6  |
|          | $t/s$                                                          | 120.59  | 52.629  | 547.53    | 15.671   |
| 5        | $J$                                                            | /       | 0.361 7 | /         | 0.373 7  |
|          | $t/s$                                                          | /       | 129.07  | /         | 36.921   |
| 6        | $J$                                                            | /       | /       | /         | 0.425 6  |
|          | $t/s$                                                          | /       | /       | /         | 167.14   |

对图1和表2中的实验结果进行分析,得到如下结论:

1) 对比ANFIS、LSSLI和L2-TSK-FS算法,本文所提出的B-TSK-FS算法,在样本量较小的情况下(即在数据集1及数据集2上)无需压缩样本即可得到与3种对比算法相近甚至更佳的泛化性能。

2) 随着样本量的不断增加,所有算法的运行时间也都相应的增加。当样本量超过4 000时,ANFIS算法由于规则数较多,运行时间较长;同样地,L2-TSK-FS算法需要求解二次规划问题,导致运行时间呈指数级增长。而LSSLI算法在样本数超过8 000时,也由于运行时间过长而终止。相比之下,B-TSK-FS算法通过设置狄利克雷分布函数中不同的 $\alpha$ 值对样本进行压缩,在保证逼近效果达到可接受范围的同时,大大降低了运行时间,使该算法可以运用于大规模数据。

3.2.2 真实数据集

本节在真实数据集Census<sup>[29]</sup>及CartExamples<sup>[30]</sup>上验证算法性能,表3给出了两组真实数据集的详细描述。在实验设置中,随机选取不同比例的数据以控制数据集的大小,即5%,10%,20%,40%,80%,100%。对于每组数据集,随机抽取40%的数据用于训练,剩下的用于测试,实验结果如表4和表5所示。对比各个算法性能,可以看出,与3.2.1节中人造数据效果基本保持一致,且有如下结论:

1) 从指标 $J$ 来看,B-TSK-FS算法与其他3种模糊建模算法的泛化性能是可比较的,且大多数情况下相

表3 真实数据集属性描述

| Datasets     | 样本个数   | 样本维数 |
|--------------|--------|------|
| Census       | 22 784 | 8    |
| CartExamples | 40 768 | 10   |

表 4 4 种模糊系统建模方法在 Census 数据集上的运行时间及泛化性能比较

| Datasets / % | Performance  | ANFIS         | LSSLI       | L2-TSK-FS   | B-TSK-FS    |
|--------------|--------------|---------------|-------------|-------------|-------------|
|              | Running Time | ( $M = 2^8$ ) | ( $M = 6$ ) | ( $M = 6$ ) | ( $M = 6$ ) |
| 1            | $J$          | 0.741 6       | 0.721 2     | 0.717 9     | 0.720 4     |
|              | $t/s$        | 6 003.5       | 1.867 2     | 7.048 3     | 35.289      |
| 2            | $J$          | 0.710 8       | 0.694 7     | 0.710 6     | 0.697 3     |
|              | $t/s$        | 9 519.1       | 7.163 6     | 41.472      | 67.776      |
| 3            | $J$          | /             | 0.657 4     | 0.659 3     | 0.658       |
|              | $t/s$        | /             | 15.88       | 263.69      | 136.48      |
| 4            | $J$          | /             | /           | /           | 0.664 4     |
|              | $t/s$        | /             | /           | /           | 260.73      |
| 5            | $J$          | /             | /           | /           | 0.650 5     |
|              | $t/s$        | /             | /           | /           | 516.12      |
| 6            | $J$          | /             | /           | /           | 0.651 6     |
|              | $t/s$        | /             | /           | /           | 608.15      |

表 5 4 种模糊系统建模方法在 CartExamples 数据集上的运行时间及泛化性能比较

| Datasets / % | Performance  | ANFIS            | LSSLI       | L2-TSK-FS   | B-TSK-FS    |
|--------------|--------------|------------------|-------------|-------------|-------------|
|              | Running Time | ( $M = 2^{10}$ ) | ( $M = 6$ ) | ( $M = 6$ ) | ( $M = 6$ ) |
| 1            | $J$          | /                | 0.221 7     | 0.224 9     | 0.223       |
|              | $t/s$        | /                | 5.089 4     | 31.525      | 52.092      |
| 2            | $J$          | /                | 0.206 2     | 0.219 1     | 0.216 5     |
|              | $t/s$        | /                | 27.837      | 224.91      | 110.09      |
| 3            | $J$          | /                | /           | /           | 0.209 5     |
|              | $t/s$        | /                | /           | /           | 230.76      |
| 4            | $J$          | /                | /           | /           | 0.200 9     |
|              | $t/s$        | /                | /           | /           | 443.24      |
| 5            | $J$          | /                | /           | /           | 0.194       |
|              | $t/s$        | /                | /           | /           | 892.28      |
| 6            | $J$          | /                | /           | /           | 0.190 3     |
|              | $t/s$        | /                | /           | /           | 1 005.7     |

近甚至具有更佳的泛化性能. 值得注意的是,随着 B-TSK-FS 方法迭代次数的增加,势必会在泛化性能上超过其他对比算法,但是需要以增加一定的计算时间为代价.

2) 随着样本量的增加,ANFIS、LSSLI 和 L2-TSK-FS 算法因运行时间较长而不再适用. 具体地,在样本量达到 4 000 以上时,LSSLI 和 L2-TSK-FS 算法的指数级运行时间增长速度,使得两种算法对数据集的大

小较为敏感. ANFIS 算法由于样本维数及样本量的增大,导致交叉验证所选的规则数增多,同样不适用于大规模数据. 然而,得益于狄利克雷分布样本压缩策略,对不同数据规模采用不同的样本压缩率,使得本文所提出的 B-TSK-FS 算法可以高效地处理大规模数据.

4 结 论

本文采用概率论思想,针对传统 TSK 模糊系统建模策略,特别是在解决大规模数据处理问题上,提出了贝叶斯 L2 型 TSK 模糊系统. 该方法采用 TSK 模糊系统建模策略中同时处理前、后件的方式,保证了模糊系统具有较高的泛化性能. 利用贝叶斯决策方法实现了模糊规则前件与后件参数的同时学习,规避了目前采用神经网络方法所导致的收敛速度较慢、模糊规则可解释性差等问题. 另外,通过引入狄利克雷先验分布函数,可以有效地对隶属度函数进行稀疏化处理,达到数据压缩的功能,用于处理大规模数据. 人工数据以及真实数据实验结果亦验证了本文方法的有效性.

然而,考虑到 L2 型损失函数的简洁性,本文方法仅局限于对 L2 型 TSK 模糊系统的建模. 为了克服以上局限性,在今后工作中将致力在概率论以及模糊系统之间建立联系,进一步从概率论的角度研究其他类型的模糊系统.

参考文献(References)

[1] Zadeh L A. Fuzzy sets[J]. Information and control, 1965, 8(3): 338-353.

[2] Zadeh L A. Discussion: Probability theory and fuzzy logic are complementary rather than competitive[J]. Technometrics, 1995, 37(3): 271-276.

[3] Ding B. Dynamic output feedback predictive control for nonlinear systems represented by a Takagi-Sugeno model[J]. IEEE Trans on Fuzzy Systems, 2011, 19(5): 831-843.

[4] Khanesar M A, Kaynak O, Teshnehlab M. Direct model reference Takagi-Sugeno fuzzy control of SISO nonlinear systems[J]. IEEE Trans on Fuzzy Systems, 2011, 19(5): 914-924.

[5] Jiang Y, Chung F L, Ishibuchi H, et al. Multitask TSK fuzzy system modeling by mining intertask common hidden structure[J]. IEEE Trans on Cybernetics, 2015, 45(3): 534-547.

[6] Dubois D, Prade H. Fuzzy sets and probability: Misunderstandings, bridges and gaps[C]. IEEE Int Conf on Fuzzy Systems. San Francisco: IEEE, 1993: 1059-1068.

- [7] Mendel J M. Uncertain rule-based fuzzy logic systems: Introduction and new directions[M]. Upper Saddle River: Prentice Hall PTR, 2001: 131-184.
- [8] Deng Z, Choi K S, Chung F L, et al. Scalable TSK fuzzy modeling for very large datasets using minimal-enclosing-ball approximation[J]. IEEE Trans on Fuzzy Systems, 2011, 19(2): 210-226.
- [9] Deng Z, Cao L, Jiang Y, et al. Minimax probability TSK fuzzy system classifier: A more transparent and highly interpretable classification model[J]. IEEE Trans on Fuzzy Systems, 2015, 23(4): 813-826.
- [10] Le Moigne J, Campbell W J, Crompt R F. An automated parallel image registration technique based on the correlation of wavelet features[J]. IEEE Trans on Geoscience and Remote Sensing, 2002, 40(8): 1849-1864.
- [11] Dai X, Khorrarn S. A feature-based image registration algorithm using improved chain-code representation combined with invariant moments[J]. IEEE Trans on Geoscience and Remote Sensing, 1999, 37(5): 2351-2362.
- [12] 邓赵红, 张江滨, 蒋亦樟, 等. 基于模糊子空间聚类的0阶岭回归TSK模糊系统[J]. 控制与决策, 2016, 31(5): 882-888.  
(Deng Z H, Zhang J B, Jiang Y Z, et al. Fuzzy subspace clustering based 0-order ridge regression TSK fuzzy system[J]. Control and Decision, 2016, 31(5): 882-888.
- [13] Mamdani E H. Application of fuzzy logic to approximate reasoning using linguistic synthesis[J]. IEEE Trans on Computers, 1977, 100(12): 1182-1191.
- [14] Azeem M F, Hanmandlu M, Ahmad N. Generalization of adaptive neuro-fuzzy inference systems[J]. IEEE Trans on Neural Networks, 2000, 11(6): 1332-1346.
- [15] Wang L, Langari R. Building Sugeno-type models using fuzzy discretization and orthogonal parameter estimation techniques[J]. IEEE Trans on Fuzzy Systems, 1995, 3(4): 454-458.
- [16] de Jesús Rubio J. SOFMLS: Online self-organizing fuzzy modified least-squares network[J]. IEEE Trans on Fuzzy Systems, 2009, 17(6): 1296-1309.
- [17] Wang X Z, Zhang T, He L. Application of fuzzy adaptive back-propagation neural network in thermal conductivity gas analyzer[J]. Neurocomputing, 2010, 73(4): 679-683.
- [18] Ishibuchi H, Nozaki K, Yamamoto N, et al. Selecting fuzzy if-then rules for classification problems using genetic algorithms[J]. IEEE Trans on Fuzzy Systems, 1995, 3(3): 260-270.
- [19] Juang C F, Tsao Y W. A self-evolving interval type-2 fuzzy neural network with online structure and parameter learning[J]. IEEE Trans on Fuzzy Systems, 2008, 16(6): 1411-1424.
- [20] Jang J S R. ANFIS: adaptive-network-based fuzzy inference system[J]. IEEE Trans on Systems, Man, and Cybernetics, 1993, 23(3): 665-685.
- [21] Deng Z, Choi K S, Jiang Y, et al. Generalized hidden-mapping ridge regression, knowledge-leveraged inductive transfer learning for neural networks, fuzzy systems and kernel methods[J]. IEEE Trans on Cybernetics, 2014, 44(12): 2585-2599.
- [22] Yang C C. Fuzzy bayesian inference[C]. IEEE Int Conf on Computational Cybernetics and Simulation. Orlando: IEEE, 1997, 3: 2707-2712.
- [23] Glenn T, Zare A, Gader P. Bayesian fuzzy clustering[J]. IEEE Trans on Fuzzy Systems, 2015, 23(5): 1545-1561.
- [24] Zhu L, Chung F L, Wang S. Generalized fuzzy c-means clustering algorithm with improved fuzzy partitions[J]. IEEE Trans on Systems, Man, and Cybernetics, Part B, 2009, 39(3): 578-591.
- [25] Rosenblatt M. A central limit theorem and a strong mixing condition[J]. Proc of the National Academy of Sciences, 1956, 42(1): 43-47.
- [26] Shen C, Li H. On the dual formulation of boosting algorithms[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2010, 32(12): 2216-2231.
- [27] Leski J M. TSK-fuzzy modeling based on  $\varepsilon$ -insensitive learning[J]. IEEE Trans on Fuzzy Systems, 2005, 13(2): 181-193.
- [28] Zhang Q, Chung F, Wang S. Transformation between type-2 TSK fuzzy systems and an uncertain Gaussian mixture model[J]. Soft Computing, 2010, 14(7): 701-711.
- [29] Tsang I W H, Kwok J T Y, Zurada J M. Generalized core vector machines[J]. IEEE Trans on Neural Networks, 2006, 17(5): 1126-1140.
- [30] Jorge A M, Azevedo P J. Optimal leverage association rules with numerical interval conditions[J]. Intelligent Data Analysis, 2012, 16(1): 25-47.

(责任编辑: 孙艺红)