

# 0 阶 L2 型 TSK 迁移学习模糊系统

蒋亦樟, 邓赵红, 王士同

(江南大学数字媒体学院, 江苏无锡 214122)

**摘 要:** 针对传统模糊系统在面对源场景存在某种程度的信息缺失或所采集的信息有限导致受训所得系统泛化能力较差之挑战, 探讨了具有迁移学习能力的模糊系统. 该类模糊系统不仅能充分利用当前场景的数据信息, 还可有效利用历史知识对当前源场景的建模过程进行辅助学习, 以达到弥补信息缺失之目的. 基于此思想以 0 阶 L2 型 TSK 模糊系统为研究对象构造了 0 阶 L2 型 TSK 迁移模糊系统(0-L2-TSK-TFS). 在模拟数据集以及真实数据集上的实验研究亦验证了该迁移模糊系统在应对存在信息缺失的场景时, 较之于传统模糊建模方法有着更好的适用性.

**关键词:** 迁移学习; 信息缺失; 历史知识; 0-L2-Takagi-Sugeno-Kang (0-L2-TSK) 模糊系统

**中图分类号:** TP391      **文献标识码:** A      **文章编号:** 0372-2112 (2013)05-0897-08

**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.3969/j.issn.0372-2112.2013.05.011

## 0-Order-L2-Norm-Takagi-Sugeno-Kang Type Transfer Learning Fuzzy System

JIANG Yi-zhang, DENG Zhao-hong, WANG Shi-tong

(School of Digital Media, Jiangnan University, Wuxi, Jiangsu 214122, China)

**Abstract:** When the information of partial data is missing, the classical fuzzy systems constructed based on this dataset will have the weak generalization abilities for this scene. In order to overcome this shortcoming, the fuzzy system with the transfer learning abilities, i. e., transfer fuzzy system, is proposed. In the learning procedure, the transfer fuzzy system can learn not only from the data information in the current scene, but also from the existing useful historical knowledge. Based on this idea, a transfer learning mechanism, a specified, and L2-norm penalty based 0-order-TSK-transfer fuzzy system(0-L2-TSK-TFS) was proposed, and a transfer learning mechanism was introduced. The proposed method was verified by experiments on simulation data and real data, and shows better adaptability than traditional fuzzy modeling methods in the scene with information missing.

**Key words:** transfer learning; information missing; history knowledge; 0-order-L2-norm-Takagi-Sugeno-Kang fuzzy system

### 1 引言

迁移学习理论自 1995 年被提及以来, 在机器学习领域产生了巨大的影响<sup>[1]</sup>. 大量的研究表明, 传统的机器学习方法都是假定源数据域所含数据包含的信息是充分的. 由于该假设条件的存在, 使得传统方法存在着一个如下的关键问题: 如果当前产生数据的数据源所处的领域为一新兴的领域, 则可能关于该领域的认识与采集到的信息均不充分, 那么, 在针对此类事物进行系统建模时, 若继续依照传统的建模方法, 则前期的工作只能停留在等待数据的采集与储备工作的完成, 只有当采集到的数据积累到建模所需的最低标准时, 方可进入下一步也是最为关键的阶段——系统建模阶段. 然而, 依

照此类方法所构之系统泛化性能常受限于所采集数据的充分程度. 当数据不充分时, 则所建系统泛化能力将急剧下降. 而数据的不充分性又常体现在新兴事物的发展初期, 为了适应这种社会发展的需要, 若继续使用传统的机器学习方法对不同的新场景建模已无法满足此类需求. 鉴于以上情景, 研究人员探讨了更为智能的机器学习方法, 即迁移学习方法<sup>[1]</sup>. 纵观近几年该理论的发展, 其已被成功应用于分类、聚类以及回归等方面. 其中又以分类领域的成果最为丰富, 一些具有代表性的工作如文献[2~7]. 而针对回归而言, 虽然将迁移学习理论应用于该方向是非常自然地, 但是目前在该方面的成果还极少, 代表性的工作有 Yang 等人提出的 HiRBF 算法, 该算法被用于处理非线性回归问题<sup>[8]</sup>. 在众多的回

归智能建模方法中,模糊系统方法由于其良好的可解释性以及其强大的学习能力得到了广泛的关注<sup>[9~12]</sup>.但目前针对模糊系统领域而言,基于迁移学习方法的模糊系统还几乎没有被探讨.

本研究将以模糊系统作为具体模型来重点探讨迁移学习在回归中的应用.具体地,本文提出了一种具有迁移学习能力的模糊系统建模方法.首先,鉴于 0 阶 L2 型 TSK 模型在面对中小规模数据建模时优越的性能,本文以其为研究对象来构建迁移模糊系统.其次,在此模型的基础上,通过引入迁移学习机制,构造出了新的目标函数,从而有效地弥补由于信息缺失造成的模糊系统泛化性能变差之问题.

## 2 Takagi-Sugeno-Kang 模糊系统

经典模糊系统模型可分为以下几类,如表 1 所述.

表 1 经典模糊系统对应之模糊规则

经典模糊系统	相关模型的模糊规则
TSK 型模糊系统 <sup>[13,14]</sup>	IF $x_1$ is $A_1^k \wedge x_2$ is $A_2^k \wedge \dots \wedge x_d$ is $A_d^k$ THEN $y^k$ is $f^k(x), k=1, \dots, K$ .
ML 型模糊系统 <sup>[15]</sup>	IF $x_1$ is $A_1^k \wedge x_2$ is $A_2^k \wedge \dots \wedge x_d$ is $A_d^k$ THEN $y^k$ is $B_k(b_k, v_k), k=1, \dots, K$ .
GFM 型模糊系统 <sup>[16]</sup>	IF $x_1$ is $A_1^k \wedge x_2$ is $A_2^k \wedge \dots \wedge x_d$ is $A_d^k$ THEN $y^k$ is $B_k(f^k(x), v^k), k=1, \dots, K$ .

在表 1 中,  $A_i^k$  表示输入向量  $\mathbf{x}$  第  $i$  维所对应的第  $k$  条规则的模糊子集,  $K$  表示为模糊规则数,  $\wedge$  则为模糊合取操作,  $f^k(x)$  为 TSK 模型中可自定义的模糊集,  $B^k(\cdot)$  为 ML 模型中第  $k$  条模糊规则对应的模糊集,  $b_k$  为质心和  $v_k$  为模糊系数.

同样如表 1 所述,可知由于 TSK 模型输出的简洁性,以及其所对应的训练过程通常可转化为线性回归问题或二次规划问题来求解,这使得该模型较之于其他模型而言训练过程更为快速有效.本文即针对 TSK 模型进行了迁移学习机制的探讨.具体地,本文将 0 阶 L2 型 TSK 模糊系统<sup>[13,17]</sup>为基础模型,探讨了具有迁移学习能力的 0 阶 L2 型 TSK 模糊系统的建模方法.

### 2.1 0 阶 Takagi-Sugeno-Kang (TSK) 模糊系统

参考文献[17]中 0 阶 TSK 型模糊系统的描述可知其模糊推理规则表达如下:

$$\begin{aligned} & \text{IF } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \dots \wedge x_d \text{ is } A_d^k \\ & \text{THEN } y^k = f^k(x) = p_0^k, \quad k=1, \dots, K. \end{aligned} \quad (1)$$

在经过一系列的操作及输出去模糊化处理之后,最终可得 0 阶 TSK 模糊系统的输出表达式如下:

$$y^0 = \sum_{k=1}^K \frac{\mu^k(\mathbf{x})}{\sum_{k=1}^K \mu^k(\mathbf{x})} f^k(\mathbf{x}) = \sum_{k=1}^K \tilde{\mu}^k(\mathbf{x}) f^k(\mathbf{x}) = \sum_{k=1}^K \tilde{\mu}^k(\mathbf{x}) p_0^k \quad (2)$$

其中,

$$\mu^k(\mathbf{x}) = \prod_{i=1}^d \mu_{A_i^k}(x_i) \quad (3)$$

$$\tilde{\mu}^k(\mathbf{x}) = \mu^k(\mathbf{x}) / \sum_{k=1}^K \mu^k(\mathbf{x}) \quad (4)$$

若引高斯函数作为隶属度函数,则式(3)中  $\mu_{A_i^k}(x_i)$  部分可具体表示为:

$$\mu_{A_i^k}(x_i) = \exp\left(\frac{-(x_i - c_i^k)^2}{2\delta_i^k}\right) \quad (5)$$

其中,参数  $c_i^k$  以及  $\delta_i^k$  都可通过经典的聚类算法或其它方法来获取.若采用 FCM 算法<sup>[18,19]</sup>,则  $c_i^k$  以及  $\delta_i^k$  可通过如下表达式估计:

$$c_i^k = \frac{\sum_{j=1}^N u_{jk} x_{ji}}{\sum_{j=1}^N u_{jk}} \quad (6)$$

$$\delta_i^k = h \cdot \sum_{j=1}^N u_{jk} (x_{ji} - c_i^k)^2 / \sum_{j=1}^N u_{jk} \quad (7)$$

上式中,  $u_{jk}$  表示输入向量  $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jd})^T$  隶属于第  $k$  类的隶属度.  $h$  则为人工可调的尺度参数.

参考文献[13]的策略,令

$$\mathbf{x}_g = (\tilde{\mu}^1(\mathbf{x}), \tilde{\mu}^2(\mathbf{x}), \dots, \tilde{\mu}^K(\mathbf{x}))^T$$

$$\mathbf{p}_g = (p_0^1, p_0^2, \dots, p_0^K)^T$$

此时,可将式(2)表示为:

$$y^0 = \mathbf{p}_g^T \mathbf{x}_g \quad (8)$$

通过上述描述,可知 0 阶 TSK 型模糊系统的参数学习问题最终可转化为线性回归问题<sup>[13,20]</sup>.这也使得利用 0 阶 TSK 模型进行系统建模时,可借鉴线性回归模型方面的技术进展来进行探讨.

### 2.2 基于 $\epsilon$ -不敏感损失函数的 0 阶 L2 型 TSK 模糊系统

根据文献[13,17]的原理,基于  $\epsilon$ -不敏感损失函数的 0 阶 L2 型 TSK 模型的优化目标函数可表示如下:

$$\min_{p_g} L = \sum_{i=1}^N |y_i^0 - y_i|_\epsilon = \sum_{i=1}^N |\mathbf{p}_g^T \mathbf{x}_{gi} - y_i|_\epsilon \quad (9)$$

在通过引入 L2 范式的惩罚项  $\xi_i^+$  以及  $\xi_i^-$  后,式(9)可等价的表示为:

$$\min_{p_g, \xi^+, \xi^-} L(\mathbf{p}_g, \xi^+, \xi^-) = \sum_{i=1}^N [(\xi_i^+)^2 + (\xi_i^-)^2] \quad (10)$$

$$\text{s.t.} \quad \begin{cases} y_i - \mathbf{p}_g^T \mathbf{x}_{gi} < \epsilon + \xi_i^+ \\ \mathbf{p}_g^T \mathbf{x}_{gi} - y_i < \epsilon + \xi_i^-, \forall i \end{cases}$$

类似 L2-SVR<sup>[21]</sup>的学习机理,文献[13]通过引入结构风险项,将式(10)进一步改进为式(11):

$$\min_{p_g, \xi^+, \xi^-, \epsilon} L(\mathbf{p}_g, \xi^+, \xi^-, \epsilon) = \frac{1}{N\tau} \sum_{i=1}^N [(\xi_i^+)^2 + (\xi_i^-)^2]$$

$$\begin{aligned}
& + \frac{1}{2}(\mathbf{p}_g^T \mathbf{p}_g) + \frac{2}{\tau} \cdot \epsilon \\
\text{s.t. } & \begin{cases} y_i - \mathbf{p}_g^T \mathbf{x}_{gi} < \epsilon + \xi_i^+ \\ \mathbf{p}_g^T \mathbf{x}_{gi} - y_i < \epsilon + \xi_i^-, \forall i \end{cases} \quad (11)
\end{aligned}$$

由于式(11)引入了结构风险项,使得其继承了统计学习理论学习算法的一些优点,如对小样本集较鲁棒。

对于式(11),通过优化学习理论,其对偶问题可表示为如下的二次规划问题:

$$\begin{aligned}
\max_{\lambda^+, \lambda^-} & - \sum_{i=1}^N \sum_{j=1}^N (\lambda_i^+ - \lambda_i^-) (\lambda_j^+ - \lambda_j^-) \cdot \mathbf{x}_{gi}^T \mathbf{x}_{gj} - \frac{N\tau}{2} \sum_{i=1}^N (\lambda_i^+)^2 - \frac{N\tau}{2} \sum_{i=1}^N (\lambda_i^-)^2 + \sum_{i=1}^N \lambda_i^+ \cdot y_i \cdot \tau - \sum_{i=1}^N \lambda_i^- \cdot y_i \cdot \tau \\
\text{s.t. } & \sum_{i=1}^N (\lambda_i^+ + \lambda_i^-) = 1, \lambda_i^+ \geq 0, \lambda_i^- \geq 0, \forall i \quad (12)
\end{aligned}$$

令:

$$\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_{2N})^T = ((\boldsymbol{\lambda}^+)^T, (\boldsymbol{\lambda}^-)^T)^T \quad (13)$$

$$\mathbf{z}_i = \begin{cases} \mathbf{x}_{gi}, & i = 1, \dots, N \\ -\mathbf{x}_{g(i-N)}, & i = N+1, \dots, 2N \end{cases} \quad (14)$$

$$\boldsymbol{\beta} = (\boldsymbol{\tau}\mathbf{y}^T, -\boldsymbol{\tau}\mathbf{y}^T)^T, \mathbf{y} = (y_1, y_2, \dots, y_N)^T \quad (15)$$

通过式(13)~(15),可将式(12)简化为如下的标准二次规划<sup>[22]</sup>形式:

$$\begin{aligned}
& \arg \max_{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}^T \tilde{\mathbf{K}} \tilde{\boldsymbol{\alpha}} + \tilde{\boldsymbol{\alpha}}^T \boldsymbol{\beta} \\
\text{s.t. } & \tilde{\boldsymbol{\alpha}}^T \mathbf{1} = 1, \quad \alpha_i \geq 0, \quad \forall i \quad (16)
\end{aligned}$$

其中,  $\tilde{\mathbf{K}} = [\tilde{k}_{ij}]_{2N \times 2N}$ ,  $\tilde{k}_{ij} = \mathbf{z}_i^T \mathbf{z}_j + \frac{N\tau}{2} \delta_{ij}$  为核矩阵,  $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$ .

通过对式(16)的求解,最终可获取式(12)中参数  $\lambda^+$ ,  $\lambda^-$  的最优解,根据对偶理论,原优化问题式(11)中的各参数的最优解可分别表示为:

$$\mathbf{p}_g = \frac{2}{\tau} \sum_{i=1}^N (\lambda_i^+ - \lambda_i^-) \mathbf{x}_{gi} \quad (17)$$

$$\xi_i^+ = N\lambda_i^+ \quad (18)$$

$$\xi_i^- = N\lambda_i^- \quad (19)$$

$$\begin{aligned}
\epsilon = & \sum_{i=1}^N (\lambda_i^+ - \lambda_i^-) y_i - \frac{N}{2} \sum_{i=1}^N [(\lambda_i^+)^2 + (\lambda_i^-)^2] \\
& - \frac{1}{\tau} \sum_{i=1}^N \sum_{j=1}^N (\lambda_i^+ - \lambda_i^-) (\lambda_j^+ - \lambda_j^-) \cdot \mathbf{x}_{gi}^T \mathbf{x}_{gj} \quad (20)
\end{aligned}$$

式(17)所得解即为 0-L2-TSK-FS 训练得到的后件部分的参数最优解。

### 3 0 阶 L2 型 TSK 迁移模糊系统 (0-L2-TSK-TFS)

通过对 0 阶 L2 型 TSK 模糊系统的规则以及参数学

习策略的分析与认识,本文将以 0 阶 L2 型 TSK 模糊系统(0-L2-TSK-FS)为基础,进行相应的具有迁移学习能

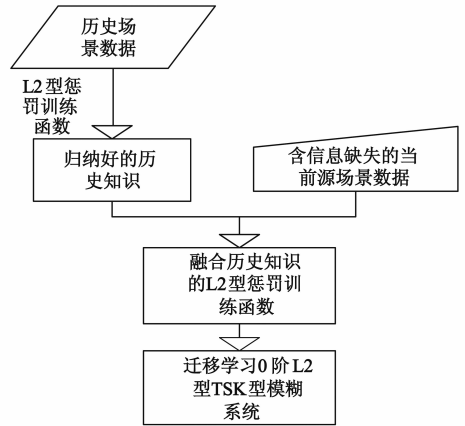


图1 0阶L2型TSK迁移模糊系统(0-L2-TSK-TFS)构造原理之示意图

力的模糊系统建模之探讨,具体如图 1。

#### 3.1 融合知识迁移的目标函数构造

基于 L2 型惩罚函数构造的 0 阶 TSK 型模糊系统,其中一类有用的知识可以描述为该系统对应的参数  $\mathbf{p}_g$ . 因而对于某历史场景受训得到的 0 阶 L2 型 TSK 模糊系统,可把其对应的  $\mathbf{p}_g$  作为已有的可用历史知识,记为  $\mathbf{p}_{g_0}$ . 为了从该历史知识进行有效地迁移学习,本文构造一个如下的引入迁移学习机制的新学习目标函数:

$$\begin{aligned}
\min_{\mathbf{p}_g, \xi^+, \xi^-, \epsilon} & \tilde{L}(\mathbf{p}_g, \xi^+, \xi^-, \epsilon) = \Theta_{\text{current}}(\mathbf{p}_g, \xi^+, \xi^-, \epsilon) \\
& + \lambda \Theta_{\text{history}}(\mathbf{p}_g, \mathbf{p}_{g_0}) \quad (21)
\end{aligned}$$

$$\text{s.t. } \begin{cases} y_i - \mathbf{p}_g^T \mathbf{x}_{gi} < \epsilon + \xi_i^+ \\ \mathbf{p}_g^T \mathbf{x}_{gi} - y_i < \epsilon + \xi_i^-, \forall i \end{cases}$$

其中,

$$\Theta_{\text{current}} = \frac{1}{N\tau} \sum_{i=1}^N [(\xi_i^+)^2 + (\xi_i^-)^2] + \frac{1}{2}(\mathbf{p}_g^T \mathbf{p}_g) + \frac{2}{\tau} \cdot \epsilon$$

$$\Theta_{\text{history}} = (\mathbf{p}_g - \mathbf{p}_{g_0})^T (\mathbf{p}_g - \mathbf{p}_{g_0})$$

在式(21)中,  $\Theta_{\text{current}}$  部分表示为利用当前场景数据对模型训练学习,  $\Theta_{\text{history}}$  部分则表示利用历史知识进行迁移学习,该项可以有效地利用历史知识尽可能地对当前场景数据的信息缺失进行信息补偿. 上述两项的功能分别为使得学习得到的模型逼近当前场景采样数据对应的模型和历史场景已构建的模型. 参数  $\lambda$  使得最终学习得到的模型在两者之间达到一个平衡. 参数  $\lambda$  可人工设定,也可通过机器学习中常用的交叉验证策略<sup>[23]</sup>选择一个合适的值。

#### 3.2 0-L2-TSK-TFS 参数学习规则

为了有效地求解式(21)中的模糊系统学习目标函数,本文给出如下定理:

**定理 1** 若  $\sum_{k=1}^K \mu^k(\mathbf{x}) \neq 0$ , 则原始优化问题式(21)

的对偶问题是

$$\begin{aligned} \arg \max_{\alpha_i, \alpha_i^*} & -\frac{2\lambda}{1+2\lambda} \mathbf{p}_{g_0}^T \sum_{i=1}^N (\alpha_i - \alpha_i^*) \mathbf{x}_{g_i} \\ & -\frac{1}{2(1+2\lambda)} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \mathbf{x}_{g_i}^T \mathbf{x}_{g_j} \\ & -\frac{N\tau}{4} \sum_{i=1}^N [(\alpha_i)^2 + (\alpha_i^*)^2] + \sum_{i=1}^N (\alpha_i - \alpha_i^*) y_i \\ \text{s.t. } & \boldsymbol{\alpha} \geq 0, \boldsymbol{\alpha}^* \geq 0, \sum_{i=1}^N (\alpha_i + \alpha_i^*) = \frac{2}{\tau} \end{aligned} \quad (22)$$

若令  $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_N, \tilde{\alpha}_1^*, \dots, \tilde{\alpha}_N^*)^T = ((\boldsymbol{\alpha})^T, (\boldsymbol{\alpha}^*)^T)^T$

$$\mathbf{z}_i = \begin{cases} \mathbf{x}_{g_i}, & i = 1, \dots, N \\ -\mathbf{x}_{g(i-N)}, & i = N+1, \dots, 2N \end{cases}$$

$\boldsymbol{\beta} = (-\frac{2\lambda}{1+2\lambda} \mathbf{p}_{g_0}^T \mathbf{x}_g + \mathbf{y}^T, \frac{2\lambda}{1+2\lambda} \mathbf{p}_{g_0}^T \mathbf{x}_g - \mathbf{y}^T)^T, \mathbf{y} = (y_1, y_2, \dots, y_N)^T$  则式(22)可表示为如下的矩阵形式:

$$\begin{aligned} \arg \max_{\tilde{\boldsymbol{\alpha}}} & -\tilde{\boldsymbol{\alpha}}^T \tilde{\mathbf{K}} \tilde{\boldsymbol{\alpha}} + \tilde{\boldsymbol{\alpha}}^T \boldsymbol{\beta} \\ \text{s.t. } & \tilde{\boldsymbol{\alpha}}^T \mathbf{1} = \mathbf{1}, \quad \alpha_i \geq 0, \quad \forall i \end{aligned} \quad (23)$$

其中,

$$\tilde{\mathbf{K}} = [\tilde{k}_{ij}]_{2N \times 2N}, \tilde{k}_{ij} = \frac{2\lambda}{1+2\lambda} \mathbf{z}_i^T \mathbf{z}_j + \frac{N\tau}{4} \delta_{ij} \quad (24)$$

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (25)$$

**证明** 如果  $\sum_{k=1}^K \mu^k(\mathbf{x}) \neq 0$ , 最小值问题式(21)的拉格朗日函数为

$$\begin{aligned} L(\mathbf{p}_g, \boldsymbol{\xi}^+, \boldsymbol{\xi}^-, \boldsymbol{\alpha}, \boldsymbol{\alpha}^*) & = \frac{1}{\tau} \cdot \frac{1}{N} \sum_{i=1}^N [(\xi_i^+)^2 + (\xi_i^-)^2] \\ & + \frac{1}{2} (\mathbf{p}_g^T \mathbf{p}_g) + \frac{2}{\tau} \cdot \epsilon + \lambda (\mathbf{p}_g - \mathbf{p}_{g_0})^T (\mathbf{p}_g - \mathbf{p}_{g_0}) \\ & + \sum_{i=1}^N \alpha_i (y_i - \mathbf{p}_g^T \mathbf{x}_{g_i} - \epsilon - \xi_i^+) + \sum_{i=1}^N \alpha_i^* (\mathbf{p}_g^T \mathbf{x}_{g_i} - y_i - \epsilon - \xi_i^-) \end{aligned} \quad (26)$$

其中,  $\boldsymbol{\alpha}$  与  $\boldsymbol{\alpha}^*$  为拉格朗日乘子,  $\boldsymbol{\alpha} \geq 0, \boldsymbol{\alpha}^* \geq 0$ . 根据对偶理论可知, 式(26)对应于  $\mathbf{p}_g, \boldsymbol{\xi}^+, \boldsymbol{\xi}^-, \epsilon$  的最小值等于其对应于  $\boldsymbol{\alpha}, \boldsymbol{\alpha}^*$  的最大值, 令:

$$\frac{\partial L}{\partial \mathbf{p}_g} = (1+2\lambda) \mathbf{p}_g - 2\lambda \mathbf{p}_{g_0} - \sum_{i=1}^N (\alpha_i - \alpha_i^*) \mathbf{x}_{g_i} = 0 \quad (27)$$

$$\frac{\partial L}{\partial \xi_i^+} = \frac{2}{N\tau} \xi_i^+ - \alpha_i = 0 \quad (28)$$

$$\frac{\partial L}{\partial \xi_i^-} = \frac{2}{N\tau} \xi_i^- - \alpha_i^* = 0 \quad (29)$$

$$\frac{\partial L}{\partial \epsilon} = \frac{2}{\tau} - \sum_{i=1}^N \alpha_i - \sum_{i=1}^N \alpha_i^* = 0 \quad (30)$$

将式(27)~(30)代入式(26), 即可得到式(21)的对偶问题:

$$\begin{aligned} \arg \max_{\alpha_i, \alpha_i^*} & \frac{\lambda}{1+2\lambda} \mathbf{p}_{g_0}^T \mathbf{p}_{g_0} - \frac{2\lambda}{1+2\lambda} \mathbf{p}_{g_0}^T \sum_{i=1}^N (\alpha_i - \alpha_i^*) \mathbf{x}_{g_i} \\ & - \frac{1}{2(1+2\lambda)} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \mathbf{x}_{g_i}^T \mathbf{x}_{g_j} \\ & - \frac{N\tau}{4} \sum_{i=1}^N [(\alpha_i)^2 + (\alpha_i^*)^2] + \sum_{i=1}^N (\alpha_i - \alpha_i^*) y_i \end{aligned} \quad (31)$$

上式中, 由于  $\frac{\lambda}{1+2\lambda} \mathbf{p}_{g_0}^T \mathbf{p}_{g_0}$  为常数项, 不影响对偶问题的极值求解, 因此式(31)可进一步化为式(22)或式(23). (证毕)

根据优化理论, 式(13)为典型的二次规划问题且仅当核函数  $K(\cdot)$  为 Mercer 核时, 才能确保其为二次凸规划, 使得求得的解为全局最优解. 为此, 本文给出如下定理:

**引理 1**<sup>[24]</sup> 令  $X$  是  $R^n$  上的一个紧集,  $\varphi(\mathbf{x}, \mathbf{z})$  是 Mercer 核当且仅当  $\varphi(\mathbf{x}, \mathbf{z})$  是  $X \times X$  上的连续对称函数且关于任意的  $x_1, \dots, x_n \in X$  的 Gram 矩阵半正定.

**定理 2** 式(24)所表达的核函数依旧是 Mercer 核.

**证明** 根据式(24)及式(25)的定义, 可知  $\tilde{\mathbf{K}}$  矩阵依旧是一个对称矩阵. 其次, 对任意的  $a_1, \dots, a_n \in R$ , 有

$$\begin{aligned} \sum_{i,j} a_i a_j \tilde{k}_{ij} & = \sum_{i,j} a_i a_j \left( \frac{2\lambda}{1+2\lambda} \mathbf{z}_i^T \mathbf{z}_j \right) + \sigma \\ & = \frac{2\lambda}{1+2\lambda} \left( \sum_i a_i \mathbf{z}_i \right)^2 + \sigma \end{aligned} \quad \text{其中 } \sigma = \frac{N\tau}{4} \sum_{i,j} a_i a_j \delta_{ij}$$

由于  $\frac{2\lambda}{1+2\lambda} \left( \sum_i a_i \mathbf{z}_i \right)^2 \geq 0$  且  $\mathbf{z}^T \mathbf{z}$  为  $X \times X$  上的连续对称函数, 因此  $\frac{2\lambda}{1+2\lambda} \mathbf{z}^T \mathbf{z}$  为半正定矩阵, 又因为  $\boldsymbol{\delta}$  为单位阵, 则  $\frac{N\tau}{4} \boldsymbol{\delta}$  为对角线为  $\frac{N\tau}{4}$  的对角阵. 所以有核  $\tilde{\mathbf{K}} = [\tilde{k}_{ij}]_{2N \times 2N}$  为半正定矩阵.

综上根据引理 1 的解释, 得证形如式(24)的核依旧是 Mercer 核. (证毕)

**引理 2**<sup>[24]</sup> 假设二次规划中的 Gram 矩阵为半正定矩阵, 则该二次规划为凸二次规划.

**引理 3**<sup>[24]</sup> 假设二次规划为凸二次规划, 则 KKT 条件也是充分条件, 因此得到的二次规划的解为全局最优解.

**定理 3** 设  $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_N, \tilde{\alpha}_1^*, \dots, \tilde{\alpha}_N^*)^T$  是对偶问题式(26)的解, 则原始问题式(21)对于  $\mathbf{p}_g$  的解为全局最优解, 并可表示为

$$\mathbf{p}_g = \frac{2\lambda}{1+2\lambda} \mathbf{p}_{g_0} + \frac{1}{1+2\lambda} \sum_{i=1}^N (\alpha_i - \alpha_i^*) \mathbf{x}_{g_i} \quad (32)$$

**证明** 根据引理 2 以及定理 2 的证明, 可知式(22)或式(23)为凸二次规划, 而又根据引理 3 的满足条件可知该二次规划的解为全局最优解.

因此, 若  $\tilde{\boldsymbol{\alpha}} = ((\alpha_i)^T, (\alpha_i^*)^T)^T$  为式(26)的解, 那么

根据式(27)可得:

$$\mathbf{p}_g = \frac{2\lambda}{1+2\lambda}\mathbf{p}_{g_0} + \frac{1}{1+2\lambda} \sum_{i=1}^N (\alpha_i - \alpha_i^*) \mathbf{x}_{gi}$$

其中  $\frac{2\lambda}{1+2\lambda}\mathbf{p}_{g_0}$  是常数项,因而由此求得的  $\mathbf{p}_g$  则为原始问题式(21)的全局最优解。(证毕)

这里值得指出的是式(32)给出的最优解,其中

$\frac{2\lambda}{1+2\lambda}\mathbf{p}_{g_0}$  部分为从历史中学习得到的知识,  $\frac{1}{1+2\lambda} \sum_{i=1}^N (\alpha_i - \alpha_i^*) \mathbf{x}_{gi}$  部分则为从当前数据中学习获取的知识。

### 3.3 算法描述

根据 3.2 的推导和结论,本小节给出算法 0-L2-TSK-TFS 具体训练步骤如下:

历史总结阶段:

Step1: 设置模糊规则数  $K_{his}$ , 正则参数  $\tau_{his}$ ;

Step2: 利用 FCM 算法得到历史场景的系统前件参数,进而得到对应于回归模型的数据集  $D_{his} = \{\mathbf{x}_{gi}, y_i\}$ ,  $i = 1, 2, \dots, N$ ;

Step3: 式(12)求解得到拉格朗日乘子  $\lambda_{his}^+$  以及  $\lambda_{his}^-$ ;

Step4: 根据 Step3 得到的  $\lambda_{his}^+$  以及  $\lambda_{his}^-$ , 并利用式(17)求得  $\mathbf{p}_{his}$ ;

迁移学习阶段:

Step5: 设置模糊规则数  $K = K_{his}$ , 正则参数  $\tau$  及平衡参数  $\lambda$ ,  $\mathbf{p}_{g_0} = \mathbf{p}_{his}$ ;

Step6: 继承历史模型的前件参数, 并利用历史场景系统的前件参数得到当前场景系统的回归数据集:

$$D_{current} = \{\mathbf{x}_{gi}, y_i\}, i = 1, 2, \dots, N;$$

Step7: 利用式(23)求解得到拉格朗日乘子  $\alpha$  及  $\alpha^*$ ;

0-L2-TSK-TFS 生成阶段:

Step8: 根据 Step7 所得  $\alpha$  及  $\alpha^*$ , 并利用式(32)可得  $\mathbf{p}_g$ ;

Step9: 根据学习得到的后件参数和继承的前件参数生成 0 阶 L2 型迁移学习 TSK 模糊系统。

## 4 实验研究

### 4.1 实验设置

为对本文所提之 0-L2-TSK 迁移模糊系统性能进行验证与评估, 本文将从两个方面进行实验比较:

(1) 本文方法与经典的 0 阶 L2 型 TSK 模糊系统(0-L2-TSK-FS)进行了性能比较;

(2) 本文方法与相关的具有迁移能力的回归方法

进行了比较。

为方便实验描述, 对实验中所涉及的有关符号给出相应的定义, 详见表 2。

表 2 本文所涉及的相关符号定义

相关符号	定义
D1	历史场景采样数据集
D2	当前场景采样训练数据集(含信息缺失)
D2_test	当前场景采样测试数据集
$r$	当前场景和历史场景相关系数
FS(D1)	经典 0-L2-TSK-FS 方法基于历史场景数据集构建的模糊系统
FS(D2)	经典 0-L2-TSK-FS 方法基于当前场景数据集构建的模糊系统
FS(D1 + D2)	经典 0-L2-TSK-FS 方法基于历史场景与当前场景数据集构建的模糊系统
TFS(D2 + 历史知识)	提出的迁移模糊系统基于当前场景数据和历史知识构造的模糊系统

为对本文的实验结果进行合理评价, 本文选用文献[13]中的评价指标如下:

$$J = \sqrt{\frac{\sum_{i=1}^N (y'_i - y_i)^2}{\sum_{i=1}^N (y'_i - \bar{y})^2}} \quad (33)$$

$J$  越小则表示泛化能力越好。

在实验中, 相关可人工设定的参数一律采用以下标准: (1) 有关模糊规则数  $K$ , 正则参数  $\tau$  以及平衡参数  $\lambda$  的取值均采用交叉验证的方式获取; (2) 本文采用 FCM 算法得到系统前件部分的空间划分。

实验环境: 实验硬件平台为 Intel Pentium × 2CPU, 其主频为 1.6GHz, 内存为 1GB。编程环境为 MATLAB 7.0。

### 4.2 本文方法与经典 0-L2-TSK-FS 性能比较

#### 4.2.1 合成数据集

为了体现本文方法之优势, 本文通过表 3 给出相关的历史数据以及当前场景数据的生成函数。

表 3 相关数据源生成函数定义

	定义函数	功能
历史数据源	$Y = f(x) = \cos(x) * x$	产生 D1
当前场景数据源	$y = r * Y = r * f(x)$	产生 D2 及 D2_test
$x$ 取值范围	$x \in [-10, 10]$	

根据表 3, 首先通过函数  $Y$  生成 400 个数据对作为 D1。其次, 利用函数  $y$  分别生成 142 个数据对和 200 个数据对用以构成 D2 和 D2\_test。为了构造合理的信息缺失场景, 本文在 D2 的基础上, 令数据对在输入采样区间  $[-7.2, -5.6]$  和  $[0, 2]$  内产生信息缺失。同时, 为了构造相似场景, 本实验设置  $r$  分别为 0.9、0.85、0.8、0.75 以及 0.7。图 2(a) 为  $r = 0.85$  时 D1 和 D2 数据集的生成函数, 图 2(b) 为  $r = 0.85$  时, D1 和 D2 的数据集的

具体构成.实验结果如表 4 和图 3 所示.

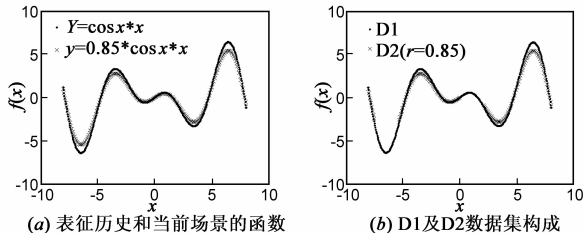


图2 相关系数为0.85时,各场景示意函数及相应的采样数据集

表 4 在当前场景测试集的泛化性能( $J_0$ )比较

$r$	0-L2-TSK-FS			0-L2-TSK-TFS
	FS(D1)	FS(D2)	FS(D1 + D2)	TFS(D2 + 历史知识)
0.9	0.1220	0.2927	0.0890	0.0383
0.85	0.1709	0.2987	0.1443	0.0577
0.8	0.2485	0.2923	0.2073	0.0827
0.75	0.3302	0.2925	0.2533	0.1421
0.7	0.4230	0.2911	0.3432	0.2183

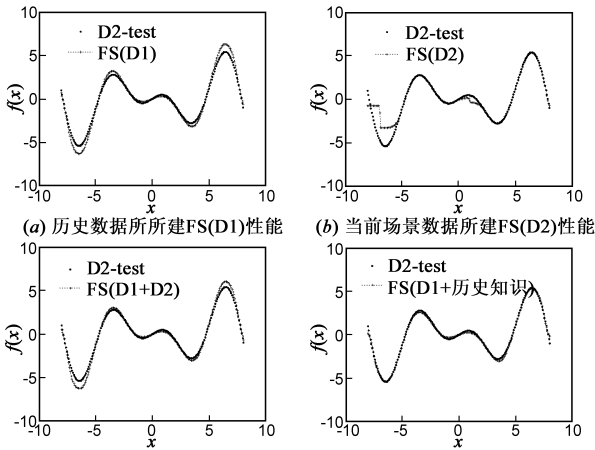


图3 在当前场景测试集(D2\_test)上的各模糊系统泛化性能比较

根据表 4 和图 3 所示的合成数据集上的实验结果可得到如下结论:

本文之方法在面对含信息缺失的场景时由于有效地利用了历史知识对当前场景的信息缺失部分进行了信息弥补.这使得该方法较之其余三种基于不同数据集构成的非迁移 0 阶 L2 型 TSK 模糊系统有着更好的拟合能力.此外,本文方法因只需历史知识(历史模型参数)以及当前场景的数据即可建模,而并不需要历史数据作为训练数据,因此,其在隐私保护方面也体现了较大的优势.

#### 4.2.2 真实数据集

本部分利用发酵过程建模仿真实验来验证本文方法较之于经典 0 阶 TSK 模糊系统的性能比较.通过对采集到的工厂发酵数据进行筛选和分析后,本文选择其中具有代表性并且覆盖较为广泛的 21 批共 294 组数据样

本,取其中 1~16 批共 224 组数据作为历史生产过程中得到的经验数据(D1),并构造出历史模型 FS(D1).其中剩余的 17~21 批共 70 组数据中有 42 组数据(17~19 批)在 6~12 时刻存在数据丢失现象,本次实验将此 42 组数据作为采集到的当前场景训练集(D2)使用,并利用其构建出系统 FS(D2).在这 70 组数据中,剩余的 28 组数据(20,21 批)信息完整,因此本文利用这 28 组数据作为测试集使用(D2-test).其中:(1)输入参数分别为:发酵过程的采样时间( $h$ ),葡萄糖浓度(Glucose Concentration, GC),菌体浓度(Thalli Concentration, TC)以及谷氨酸浓度(Glutamic Acid Concentration, GAC);(2)输出参数分别为:下一时刻的葡萄糖浓度( $h+1$ )、菌体浓度( $h+1$ )以及谷氨酸浓度( $h+1$ ).仿真实验结果如表 5 所示.

表 5 真实数据(发酵数据)上的泛化性能( $J_0$ )比较

输出属性	0-L2-TSK-FS			0-L2-TSK-TFS
	FS(D1)	FS(D2)	FS(D1 + D2)	TFS(D2 + 历史知识)
GC	0.4085	0.4716	0.3871	0.2699
TC	0.8244	0.9593	0.7405	0.4252
GAC	0.3660	0.4828	0.3085	0.1898

从表 5 的实验结果可以发现,其实验结果与上节中合成数据的观察结果基本是一致的即本文的方法在信息弥补方面较之其他建模方法有着更大的优势.

#### 4.3 本文方法与相关迁移回归算法性能比较

在回归领域内,迁移学习的研究还不多见.为了进一步评估本文建模方法的性能,本节将采用一个代表性的迁移回归方法,即 HiRBF 算法<sup>[8]</sup>与本文方法进行性能比较.本章节分别通过合成数据集实验以及真实数据实验对两种方法作性能比较.

##### 4.3.1 合成数据集

本小节依旧采用 4.2.1 小节所设计的模拟数据集及实验方案对 HiRBF 算法及本文方法作出评价.表 6 给出了相应的实验结果.

表 6 在当前场景模拟数据测试集下各模型之泛化性能( $J_0$ )比较

信息缺失区间	$r$	HiRBF		0-L2-TSK-TFS
		D2	D1 + D2	TFS(D2 + 历史知识)
[-7.2, -5.6]	0.9	0.2556	0.1098	0.0383
	0.85	0.2602	0.1534	0.0577
[0, 2]	0.7	0.2608	0.3577	0.2183

虽然在文献[8]中,算法 HiRBF 描述为一种有效地处理非线性回归的迁移学习方法,但是从表 6 的实验结果表明,不管是仅利用当前场景数据(D2)还是同时利用当前和历史场景数据(D1 + D2)进行建模,HiRBF 方法在应对当前场景存在信息缺失的情况时,依旧会造成系统某种程度的泛化性能下降的现象,究其原因在于该算法仅是借用了历史场景的数据信息,而未能有效地利用更高级的历史知识来应对信息缺失的问题.

### 4.3.2 真实数据集

本节将续用 4.2.2 小节所提供的发酵数据集对本文方法与 HiRBF 算法再次作出评价. 结果如表 7 所示.

表 7 各算法于真实数据(发酵数据)上的性能( $J_0$ )比较

缺失时刻( $h$ )	输出属性	HiRBF		0-L2-TSK-TFS
		D2	D1 + D2	TFS(D2 + 历史知识)
6, 8, 10, 12	GC	0.3646	0.3510	0.2699
	TC	0.8582	0.7026	0.4252
	GAC	0.4489	0.4117	0.1898

从表 7 的结果显示,在真实的发酵数据集上,HiRBF 算法的性能与模拟数据集上的性能表现基本是一致的.

## 5 结论

本文通过采用知识迁移的思想,针对多场景相关且某一场景信息缺失带来的模糊系统建模之挑战,提出了具有知识迁移能力的 0-L2-TSK 迁移模糊系统(0-L2-TSK-TFS)建模方法. 本文方法有效地解决了当前场景下采集到的新数据在一定范围内存在信息缺失造成所构之模糊系统泛化性能下降的缺陷. 另外,由于本文方法不需要大量历史数据的支持,仅是继承简洁的历史知识,使得本方法与生俱来就拥有了对历史数据的隐私保护能力. 通过对合成数据及发酵过程的仿真实验,所获得结果均明显的表示出本文方法较之其它传统的建模技术具有更好的适用性.

### 参考文献

- [1] Sinno Pan J L, Yang Q. A survey on transfer learning[J]. IEEE Trans on Knowledge and Data Engineering, 2009, 22(10): 1345 - 1359.
- [2] Raina R, Battle A, Lee H, Packer B, Ng Y A. Self-taught learning: Transfer learning from unlabeled data[A]. Proc 24th ICML [C]. New York: ACM Press, 2007. 759 - 766.
- [3] Xue G R, Dai W Y, Yang Q, Yu Y. Topic-bridged PLSA for cross-domain text classification [A]. Proc 31th International ACM SIGIR Conference on Research and Development in Information Retrieval [C]. New York: ACM Press, 2008. 627 - 634.
- [4] Sinno Pan J L, Tsang I W, T. Kwok J, Yang Q. Domain adaptation via transfer component analysis [A]. Proc 21th IJCAI [C]. San Francisco: ACM Press, 2009. 1187 - 1192.
- [5] Glorot X, Bordes A, Bengio Y. Domain adaptation for large-scale sentiment classification: A deep learning approach [A]. Proc 28th ICML [C]. Bellevue: ACM Press, 2011. 513 - 520.
- [6] Duan L X, Tsang I W, Xu D. Domain transfer multiple kernel learning [J]. IEEE Trans on PAMI, 2012, 34(3): 465 - 479.
- [7] 于重重, 田蕊, 谭励, 等. 非平衡样本分类的集成迁移学习

算法[J]. 电子学报, 2012, 40(7): 1358 - 1363.

- YU Chong-chong, TIAN Rui, TAN Li, et al. Integrated transfer learning algorithmic for unbalanced samples classification [J]. Acta Electronica Sinica, 2012, 40(7): 1358 - 1363. (in Chinese)
- [8] Yang P, Tan Q, Ding Y. Bayesian task-level transfer learning for non-linear regression [A]. Proc CSSE [C]. Wu Han: IEEE Press, 2008. 62 - 65.
- [9] Astrom K J, McAvoy T J. Intelligent control [J]. Journal of Process Control, 1993, 2(3): 115 - 127.
- [10] Chung F L, Deng Z H, Wang S T. An adaptive fuzzy-inference-rule-based flexible model for automatic elastic image registration [J]. IEEE Trans on Fuzzy Systems, 2009, 17(5): 995 - 1010.
- [11] 袁学海, 李洪兴, 孙凯彪. 基于参数单点模糊化方法的模糊系统及其逼近能力 [J]. 电子学报, 2011, 39(10): 2372 - 2377.  
YUAN Xue-hai, LI Hong-xing, SUN Kai-biao. Fuzzy systems and their approximation capability based on parameter singleton fuzzifier methods [J]. Acta Electronica Sinica, 2011, 39(10): 2372 - 2377. (in Chinese)
- [12] 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 Society, 2002, 40(8): 1849 - 1864.
- [13] Deng Z H, Choi K S, Chung F L, Wang S T. Scalable TSK fuzzy modeling for very large datasets using minimal-enclosing-ball approximation [J]. IEEE Trans on Fuzzy Systems, 2011, 19(2): 210 - 226.
- [14] Takagi T, Sugeno M. Fuzzy identification of systems and its application to modeling and control [J]. IEEE Trans on Systems Man and Cybernetics, 1985, 15(1): 116 - 132.
- [15] Mamdani E H. Application of fuzzy logic to approximate reasoning using linguistic systems [J]. IEEE Trans on Computer, 1977, 26(12): 1182 - 1191.
- [16] 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.
- [17] Juang C F, Lo C. Zero-order TSK-type fuzzy system learning using a two-phase swarm intelligence algorithm [J]. Fuzzy Sets and System, 2008, 159(21): 2910 - 2926.
- [18] Hall L O, Goldgof D B. Convergence of the single-pass and online fuzzy c-means algorithms [J]. IEEE Trans on Fuzzy Systems, 2011, 19(4): 792 - 794.
- [19] Zhu L, Chung F L, Wang S T. Generalized fuzzy c-means clustering algorithm with improved fuzzy partitions [J]. IEEE Trans on SMC-B, 2009, 39(3): 578 - 591.
- [20] Leski J. TSK-fuzzy modeling based on  $\epsilon$ -insensitive learning [J]. IEEE Trans on Fuzzy Systems, 2005, 13(2): 181 - 193.

- [21] Tsang I W, Kwok J T, Zurada J M. Generalized core vector machines[J]. IEEE Trans on Neural Networks, 2006, 17(5): 1126 – 1140.
- [22] Fan R E, Chan P H, Lin C J. Working set selection using second order information for training support vector machines [J]. Journal of Machine Learning Research, 2005, 6: 1889 – 1918.
- [23] Ito K, Nakano R. Optimizing support vector regression hyper-parameters based on cross-validation [A]. Proc IJCNN [C]. Illinois: IEEE Press, 2003. 2077 – 2082.
- [24] 邓乃杨, 田英杰. 数据挖掘的新方法——支持向量机 [M]. 北京: 科学出版社, 2004. 5 – 16.  
DENG Nai-yang, TIAN Ying-jie. New Method in Data Mining: Support Vector Machine [M]. Beijing, China: Science Press, 2004. 5 – 16. (in Chinese)

#### 作者简介



**蒋亦樟** 男, 1988 年生于江苏无锡. 2006 年、2012 年分别在南京理工大学、江南大学获得工学学士、硕士学位, 2012 年进入江南大学攻读博士学位, 主要从事模式识别、人工智能、模糊系统等方面的研究.

E-mail: jyz0512@163.com



**邓赵红** 男, 1981 年生于安徽蒙城. 副教授、硕士生导师. 主要从事人工智能、模式识别、模糊系统、医学图像处理等方面的研究工作.

E-mail: dzh666828@yahoo.com.cn



**王士同** 男, 1964 年生于江苏扬州. 教授、博士生导师、中国计算机学会高级会员. 1984 年、1987 年在南京航空航天大学获得工学学士、硕士学位. 主要从事人工智能、模式识别、模糊系统、医学图像处理和生物信息学等方面的研究工作.

E-mail: wxwangst@yahoo.com.cn