

ML 型迁移学习模糊系统

蒋亦樟¹ 邓赵红¹ 王士同¹

摘要 经典模糊系统构建方法训练时通常仅考虑单一的场景, 其伴随的一个重要缺陷是: 如当前场景重要信息缺失, 则受训所得系统泛化能力较差. 针对此问题, 以 Mamdani-Larsen (ML) 型模糊系统为对象, 探讨了具有迁移学习能力的模糊系统, 即 ML 型迁移学习模糊系统. ML 型迁移学习模糊系统不仅能充分利用当前场景的数据信息, 而且能有效地利用历史知识来进行学习, 具有通过迁移历史场景知识来弥补当前场景信息缺失的能力. 具体地, 基于经典的压缩集密度估计 (Reduced set density estimator, RSDE) ML 型模糊系统构建方法, 通过引入迁移学习机制提出了一种基于密度估计的 ML 型迁移模糊系统构建方法. 在模拟数据和真实数据上的实验研究亦验证了该迁移模糊系统在信息缺失场景下较之于传统模糊系统建模方法的更好适应性.

关键词 迁移学习, 信息缺失, 压缩集密度估计, Mamdani-Larsen 模糊系统

引用格式 蒋亦樟, 邓赵红, 王士同. ML 型迁移学习模糊系统. 自动化学报, 2012, 38(9): 1393–1409

DOI 10.3724/SP.J.1004.2012.01393

Mamdani-Larsen Type Transfer Learning Fuzzy System

JIANG Yi-Zhang¹ DENG Zhao-Hong¹ WANG Shi-Tong¹

Abstract The classical fuzzy system modeling methods only consider the single scene, which may bring up the following weakness: if the information of partial data is missing, the fuzzy systems constructed based on this dataset will have the weak generalization abilities for this scene. In order to overcome this shortcoming, by focusing on the Mamdani-Larsen type fuzzy system (ML-FS) model, the fuzzy system with the transfer learning abilities, i.e. ML-transfer fuzzy system is proposed. The ML-transfer fuzzy system will not only make full use of the data information in the learning procedure, but also effectively learn from the existing useful historical knowledge, such as parameters of the fuzzy system obtained from the dataset of historical scene, in order to make up the information lack in the current scene. Based on this idea, a specified ML-transfer fuzzy system (ML-TFS) based on reduced set density estimation (RSDE) technology is proposed by introducing transfer learning mechanism. It has been verified by experiments on simulation data and real data that the ML-transfer fuzzy system has a better adaptability than the traditional fuzzy modeling method in the scene with information missing.

Key words Transfer learning, information missing, reduced set density estimator (RSDE), ML fuzzy inference systems

Citation Jiang Yi-Zhang, Deng Zhao-Hong, Wang Shi-Tong. Mamdani-Larsen type transfer learning fuzzy system. *Acta Automatica Sinica*, 2012, 38(9): 1393–1409

模糊系统理论是以模糊集和模糊推理理论为基础演变而来的. 它能对处理人类生产与实践过程中的思维、分析、推理以及决策过程建立一种与之对应的数学模型, 使得自然语言直接转译成计算机能够识别的机器语言^[1–4]. 其特色是: 不但具有高度的可解释性, 还具有强大的学习能力. 目前, 模糊系统

已广泛应用于各个领域, 如智能控制、信号处理、模式识别等方面^[5–9].

模糊系统的参数学习是模糊系统研究的重要内容. 对于经典的模糊系统而言, 其参数学习一般有基于专家经验的人为赋值法和基于可用数据集的训练学习法两类方法. 后一类方法由于其强大的学习能力在实践中应用地更为广泛. 但目前训练学习法普遍存在着同一问题: 只考虑单一的场景, 当前场景训练数据信息严重影响受训模糊建模之性能. 如该场景信息有较大的信息缺失, 则所得受训系统泛化能力将严重下降. 在实际生产过程中, 由于信息采集器、传感器等设备的稳定性以及抗外界干扰性不一定令人满意, 经常出现短路或者是其他的故障, 此时这些设备采集到的数据往往存在信息丢失或信息的不完整. 针对此种情况, 历史相关场景的信息 (如历史场景的数据信息或历史场景归纳出来的知识) 对当前场景的建模来说是一种有益的补充. 虽然历史场

收稿日期 2011-08-02 录用日期 2012-03-05

Manuscript received August 2, 2011; accepted March 5, 2012

国家自然科学基金 (60903100, 61170122), 江苏省自然科学基金 (BK2009067), 中央高校基本科研业务费专项资金 (JUSRP21128), 江苏省信息融合软件工程技术研究开发中心开放基金项目 (SR-2011-01) 资助

Supported by National Natural Science Foundation of China (60903100, 61170122), Natural Science Foundation of Jiangsu Province (BK2009067), Fundamental Research Funds for the Central Universities (JUSRP21128), and Opening Project of Jiangsu Engineering Research and Development Center for Information Fusion Software (SR-2011-01)

本文责任编辑 王聪

Recommended by Associate Editor WANG Cong

1. 江南大学数字媒体学院 无锡 214122

1. School of Digital Media, Jiangnan University, Wuxi 214122

景数据含有一定的有益信息,但利用历史数据时也会面临一些挑战:一方面,对于一些具有保密性的历史数据,普通研究人员往往无法获取,仅能获取一些历史知识(如模型参数).另一方面,直接利用历史场景数据时,如果历史场景和当前场景之间存在较大的漂移,此时大量历史数据的采用可能会带来一定的负面影响.根据如上分析可见,在如下情形下:1)当前场景的数据存在信息缺失;2)历史场景和当前场景之间存在漂移现象,此时仅利用当前场景数据或同时利用所有场景数据(如果可用)所构建出的模糊系统性能一般不能令人满意.由于这类现象在实际应用中普遍存在,因此研究出一种即不需大量相关历史场景数据又能够一定程度地承受当前场景部分数据信息缺失的模糊系统构建方法变得很有必要.

据我们所知,在模糊系统建模方面针对前面所述挑战的研究在文献中还几乎看不到.但在相近的研究领域,如针对小样本数据集的鲁棒模糊建模技术方面,目前已出现一些重要的研究成果,代表性的工作可见文献[10–13].在此类研究中,引入的一个重要技术是支撑向量学习技术.该类技术引入了结构风险来优化相应的系统,从而在小样本建模环境下显示出了较好的鲁棒性.文献[10]提出了一种基于自组织模糊聚类和支撑向量学习技术的TS型模糊神经网络分类方法.文献[11]基于模糊聚类和支撑学习技术提出了针对0阶TS型模糊推理规则的模糊系统分类算法并应用于肤色分割.文献[12]则基于支撑学习技术提出了针对回归任务的TS型模糊系统学习算法.文献[13]亦基于支撑向量技术探讨了模糊建模及其学习方法.上述模糊系统学习算法虽然在小样本环境下较之传统方法显示出了更好的鲁棒性,但针对训练数据集有重要信息缺失的场景,建模效果依然不理想.而且此类算法亦无有效利用历史场景有用信息来弥补当前场景信息缺失之能力.

值得指出的是,和本文所探讨问题密切相关的一个研究方向是迁移学习^[14–20].根据文献[14]迁移学习的理论可知,其常被用来解决交叉域问题,即当两个数据域的数据在特征空间或属性方面总体上近似,但仍存在一定差异性的情况下,可利用数据较为充足的一域去指导另一域的学习.当前迁移学习已经在分类、聚类以及回归等方面得到关注.纵观近几年来迁移学习领域的进展,可以发现其在分类领域的研究已较为丰富,代表性的工作有:2007年Raina等将迁移学习的理论应用到了未标记数据的分类问题上^[15];Yang等在文献[16]中提出了基于迁移学习理论的TPLSA算法并将其应用于文本分类方向,并在2009年的文献[17]中提出了一种半监督的域适应迁移成分分析(Transfer component analysis, TCA)方法,该算法也被广泛应用到分类

领域中;文献[18]中Glorot等进一步将域适应的迁移学习方法成功应用到了大规模情感数据的分类问题上;2012年Duan等成功地将迁移学习理论与多核学习融合并应用于分类领域^[19].在聚类方面,迁移学习也已被关注,2008年Yang等在文献[20]中将其应用于聚类的迁移学习,提出了一种自学习的聚类算法.除了分类和聚类之外,回归方面的迁移学习的研究同样值得关注,虽然迁移学习应用于回归方面是非常自然的,但是目前在该方面的成果还极少.本研究中,我们将以模糊系统为具体模型来重点探讨迁移学习在回归中的应用.

针对部分重要信息缺失环境下模糊系统建模面临的挑战,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,由于ML型模糊系统的简单性及受到广泛应用等优点,本文以其为对象来探讨迁移模糊系统的构建.其次,在此模型的基础上通过引入迁移学习机制构造新的目标函数,达到弥补信息缺失之目标.提出的ML型迁移模糊系统只需要历史场景所得的模型参数作为历史知识,而不需要历史场景的数据集,因而其对历史场景数据具有很好的隐私保护作用.同时,由于提出的ML型迁移模糊系统能有效地继承有用历史知识,其允许当前场景的数据集在某种程度上存在数据丢失或残缺的现象.因此,本文提出的ML型迁移学习模糊系统具有较好的适应性.特别是对于传感器大量存在的发酵生产过程等方面,本文方法将会有很好的潜在应用前景.

1 ML型模糊系统

经典模糊系统模型可分为以下几类:Mamdani-Larsen (ML)模型^[21–24],Takagi-Sugeno (TS)模型^[25]及Generalized fuzzy model (GFM)模型^[26].ML模型由于其简洁性是受到较多应用的一种模糊模型.本文以其为具体研究对象,探讨了具有迁移学习能力的模糊系统建模方法.

1.1 Mamdani-Larsen (ML)型模糊系统

ML型模糊系统的模糊推理规则可表达如下:

$$\begin{aligned} \text{IF } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \cdots \wedge x_d \text{ is } A_d^k \\ \text{THEN } y \text{ is } B^k(b_k, v_k) \end{aligned}$$

其中每一条规则都有与之对应的输入向量 $\mathbf{x} = [x_1, x_2, \dots, x_d]^T$,并把输入空间的模糊集 $A^k \subset \mathbf{R}^d$ 映射到输出空间的模糊集 $B^k \subset \mathbf{R}$.这里 A_i^k 表示第 i 维输入在第 k 条模糊规则中对应的模糊子集, \wedge 表示模糊合取操作.在ML模糊推理系统中, $B^k(b_k, v_k)$ 表示第 k 条模糊规则THEN部分中分别以 b_k 为质心和以 v_k 为模糊系数的模糊集.令 $\mu^k(\mathbf{x})$ 为模糊集 B^k 对应的模糊隶属度函数,第 k 条规则的隶属度

$\mu^k(\mathbf{x})$ 可由 IF 部分中各维对应的隶属度值通过合取操作 (\wedge) 获得, 即

$$\mu^k(\mathbf{x}) = \mu_1^k(x_1) \wedge \mu_2^k(x_2) \wedge \cdots \wedge \mu_d^k(x_d) \quad (1)$$

当采用乘算子、乘算子和加算子分别作为合取、蕴含和析取操作算子时, ML 模糊推理系统在去模糊化后, 最终输出可表示为

$$y^0 = \frac{\sum_{k=1}^K \mu^k(\mathbf{x}) \times v_k \times b_k}{\sum_{k'=1}^K \mu^{k'}(\mathbf{x}) \times v_{k'}} = \frac{\sum_{k=1}^K \prod_{j=1}^d \mu_j^k(x_j) \times v_k \times b_k}{\sum_{k'=1}^K \prod_{j=1}^d \mu_j^{k'}(x_j) \times v_{k'}} \quad (2)$$

1.2 基于 RSDE 密度估计的 ML 模糊系统构建方法

对于 ML 模糊系统, 一个重要的进展是基于密度估计技术的构建方法^[27]. 文献 [28] 提出了一种新的核密度估计技术, 即压缩集密度估计方法 (Reduced set density estimator, RSDE). 文献 [29] 探讨了 RSDE 和最小包含球技术之间存在的等价性, 基于核心集快速最小包含球逼近策略提出了快速的 RSDE 密度估计算法 FRSE (Fast RSDE). 文献 [27] 指出了 RSDE 和 ML 型模糊系统构建之间存在的等价性, 从而利用 RSDE^[28] 和 FRSE^[29] 提出了基于 RSDE 密度估计的 ML 型模糊系统训练算法. 在后面部分, 本文将在文献 [27] 提出的基于 RSDE 密度估计的 ML 型模糊系统构建方法基础上, 研究当存在有用历史场景知识可用时具有迁移学习能力的 ML 型模糊系统训练算法. 本文提出的具有迁移学习能力的 ML 型模糊系统训练算法本质上依然是一个密度估计过程. 和文献 [27] 的基于密度估计的构建方法相比, 其独特之处是在密度估计过程中具有了从历史知识中学习的能力, 即具有迁移学习能力.

为了便于理解, 这里对基于密度估计技术的 ML 模糊系统构建方法简介如下. 给定参考样本集 $S = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\} \in \mathbf{R}^d$, RSDE 核密度估计器的一般形式可表示为

$$\hat{p}(\mathbf{x}, h, \gamma) = \sum_{i=1}^C \gamma_i K_h(\mathbf{x}, \mathbf{x}_i) \quad (3a)$$

其中, C 是由包含 N 个样本的参考样本集得到的压缩集中的样本数目, \mathbf{x}_i 为压缩集中第 i 个样本^[28-29]. 在文献中, RSDE 密度估计器被提出用以实现核密

度估计. 对于给定的多输入单输出 (MISO) 回归数据集 $D = \{\mathbf{x}_i, y_i\}, i = 1, \cdots, N$, 其相应的 RSDE 可表示为

$$\hat{p}([\mathbf{x}; y], h, \gamma) = \sum_{i=1}^C \gamma_i K_h([\mathbf{x}; y], [\mathbf{x}_i, y_i]) \quad (3b)$$

式 (3a) 中的 (\mathbf{x}_i, y_i) 为得到的压缩集中的第 i 个输入输出数据对.

对于多输入单输出回归数据集对应的 RSDE 密度估计, 在给定输入 \mathbf{x} 时, 如果 RSDE 采用的核函数为高斯密度函数 (用符号 G 表示), 那么它所对应的期望输出可表示为

$$E[y|\mathbf{x}] = \frac{\sum_{i=1}^C \gamma_i \prod_{j=1}^d G(x_j, x_{ij}, h) y_i}{\sum_{i'=1}^C \gamma_{i'} \prod_{j=1}^d G(x_j, x_{i'j}, h)} \quad (4)$$

根据式 (4) 和式 (2), 文献 [27] 证明了 RSDE 与 ML 模糊系统之间存在的等价性. 该等价性可归纳为如下结论.

结论 1^[27]. 对于某回归数据集, 其对应的 RSDE 密度估计的条件期望输出和对应的某个 ML 模糊推理系统的输出具有相同的数学表达式, 因而可用回归数据集的 RSDE 密度估计来构造相应的 ML 模糊推理系统.

根据结论 1, 文献 [27] 提出了新的 ML 模糊系统构建方法及其在大规模数据环境下的可扩展算法. 这里对基于 RSDE 密度估计的 ML 模糊系统构建过程简述如下.

基于 RSDE 密度估计的 ML 模糊系统构建方法:

- 1) ML 模糊系统的规则数 K 等于压缩集密度估计器 RSDE 中的成员函数数目 C , 即 $K = C$;
- 2) ML 模糊系统每条规则结论部分的模糊系数 v_k 等于压缩集密度估计器 RSDE 中各成分的先验概率 γ_i , 即 $v_k = \gamma_i$;
- 3) ML 模糊系统中每条规则 THEN 部分模糊集的质心为压缩集密度估计器 RSDE 中压缩集的一个样本点的输出, $b_k = y_i$;
- 4) ML 模糊系统中 IF 部分的隶属度函数为 $\mu^k(x_j) = G(x_j, x_{ij}, h) = G(x_j, x_{ij}, \sigma^2), j = 1, 2, \cdots, d$;
- 5) ML 模糊系统的合取、蕴含和析取算子分别为乘算子、乘算子和加算子.

2 ML 型迁移模糊系统

在进行具有迁移学习能力的 ML 型模糊系统构造时, 一个关键的工作是: 针对当前场景出现的信息

缺失, 如何有效利用已有的历史知识来进行相应的信息补偿, 进而得到较之于仅利用当前场景数据信息受训所得系统具有更好适应性的模糊系统. 本研究提出的 ML 型迁移模糊系统 (ML-TFS) 的核心思想可用图 1 来表示.

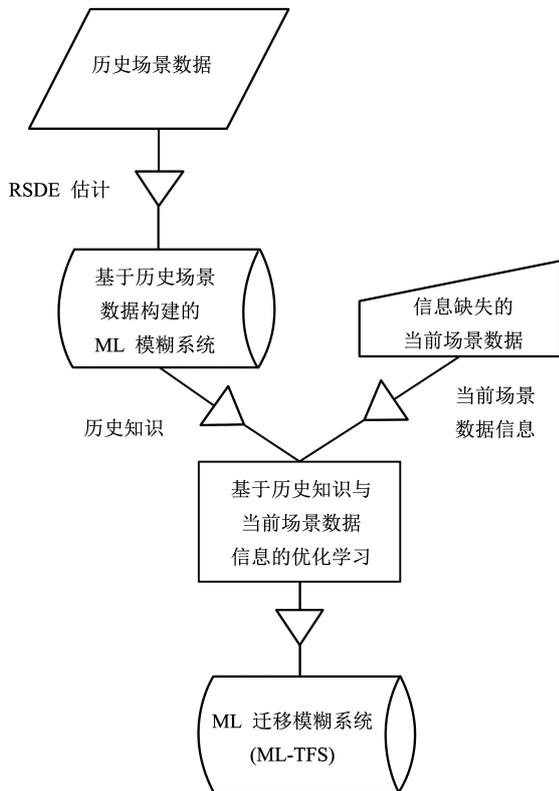


图 1 ML 型迁移模糊系统 (ML-TFS) 构造原理示意图
Fig. 1 The principle of ML-type transfer learning fuzzy system

2.1 融合知识迁移的目标函数构造

基于 RSDE 密度估计构造的模糊系统, 一类有用的知识可以描述为该系统相应的由 RSDE 估计的概率密度函数. 因而对于某历史场景受训得到的 ML 模糊系统, 可把其对应的概率密度函数作为已有的可用历史知识. 为了从该历史知识来有效地进行迁移学习, 我们将构造出一个具有迁移学习能力的新目标函数.

首先可把历史场景基于 RSDE 密度估计得到的 ML 模糊系统对应的概率密度分布函数表示如下:

$$\hat{p}(\mathbf{x}) = \sum_{i=1}^C p_i G(\mathbf{x}, \mathbf{u}_{p,i}, \sigma_{p,i}^2) \quad (5)$$

式 (5) 即为可利用的历史知识. 同时可把在当前场景下最终得到的模糊系统对应的概率密度分布函数表示为

$$\hat{q}(\mathbf{x}) = \sum_{i=1}^C q_i G(\mathbf{x}, \mathbf{u}_{q,i}, \sigma_{q,i}^2) \quad (6)$$

已有的基于 RSDE 密度估计的模糊系统构建方法, 在进行 RSDE 密度估计时仅考虑到当前场景数据的信息, 没有迁移学习的能力, 因而得到的模糊系统仅能从当前场景数据中进行学习. 这使得如果当前场景有重要的信息缺失时, 受训所得的模糊系统泛化能力将会变差. 针对此, 本文在密度估计的过程中引入可用历史知识, 使得该历史知识能有效地影响密度估计的结果, 进而使其对应的模糊系统也能有效地学习到历史场景的有益信息. 基于此点, 本研究给出如下的密度估计优化目标函数:

$$\begin{aligned} \arg \min_{q_j, \mathbf{u}_{q,j}, \sigma_{q,j}} J = & \\ \arg \min_{q_j, \mathbf{u}_{q,j}, \sigma_{q,j}} & \left(\int (\hat{p}(\mathbf{x}) - \hat{q}(\mathbf{x}))^2 d\mathbf{x} + \right. \\ & \left. \lambda \int (\hat{q}(\mathbf{x}) - q(\mathbf{x}))^2 d\mathbf{x} \right) \quad (7a) \end{aligned}$$

这里 $q(\mathbf{x})$ 表示当前场景数据对应的真实概率密度分布. 式 (7a) 中的目标函数包含两部分: 其中第一个部分为历史知识的传承, 其作用是使得学习得到的概率密度估计尽可能接近历史信息对应的密度估计; 第二个部分是利用当前场景数据对密度分布进行估计, 其目标是尽可能的使得估计出的密度分布逼近于真实的密度分布. 目标函数的第二部分直接继承于 RSDE 密度估计的目标函数, 关于其更详细的解释可参见文献 [28–29]. 式 (7a) 中的参数 λ 用来平衡两部分的影响, 其可用经典的交叉验证策略来估计其最优值. 通过对式 (7a) 分析可知, 该目标函数即能继承历史知识, 又能够通过当前场景数据信息来学习. 进一步地, 可将式 (7a) 展开得到:

$$\begin{aligned} \arg \min_{q_j, \mathbf{u}_{q,j}, \sigma_{q,j}} J = & \\ \arg \min_{q_j, \mathbf{u}_{q,j}, \sigma_{q,j}} & \left(\int \left(\frac{\hat{p}(\mathbf{x})^2 - 2\hat{p}(\mathbf{x})\hat{q}(\mathbf{x})}{\hat{q}(\mathbf{x})^2} + \right) d\mathbf{x} + \right. \\ & \left. \lambda \int (\hat{q}(\mathbf{x})^2 - 2\hat{q}(\mathbf{x})q(\mathbf{x}) + q(\mathbf{x})^2) d\mathbf{x} \right) \Rightarrow \\ \arg \min_{q_j, \mathbf{u}_{q,j}, \sigma_{q,j}} J = & \\ \arg \min_{q_j, \mathbf{u}_{q,j}, \sigma_{q,j}} & \left(\int \left(\frac{-2\hat{p}(\mathbf{x})\hat{q}(\mathbf{x})}{\hat{q}(\mathbf{x})^2} + \right) d\mathbf{x} + \right. \\ & \left. \lambda \int (\hat{q}(\mathbf{x})^2 - 2q(\mathbf{x})\hat{q}(\mathbf{x})) d\mathbf{x} \right) \Rightarrow \\ \arg \min_{q_j, \mathbf{u}_{q,j}, \sigma_{q,j}} J = \arg \min_{q_j, \mathbf{u}_{q,j}, \sigma_{q,j}} & \left((1 + \lambda) \int \hat{q}(\mathbf{x})^2 d\mathbf{x} - \right. \\ & \left. 2 \int \hat{p}(\mathbf{x})\hat{q}(\mathbf{x}) d\mathbf{x} - 2\lambda \int q(\mathbf{x})\hat{q}(\mathbf{x}) d\mathbf{x} \right) \quad (7b) \end{aligned}$$

由于 $\int G(\mathbf{x}, \mathbf{u}_{q,j}, \sigma_{p,i}^2)G(\mathbf{x}, \mathbf{u}_{q,j}, \sigma_{q,j}^2)d\mathbf{x} = G(\mathbf{u}_{p,i}, \mathbf{u}_{q,j}, \sigma_{p,i}^2 + \sigma_{q,j}^2)$, 利用式 (5) 和式 (6) 可得:

$$\int \widehat{p}(\mathbf{x})\widehat{q}(\mathbf{x})d\mathbf{x} = \sum_{i=1}^C \sum_{j=1}^C p_i q_j G(\mathbf{u}_{p,i}, \mathbf{u}_{q,j}, \sigma_{p,i}^2 + \sigma_{q,j}^2) \quad (8a)$$

$$\int \widehat{q}(\mathbf{x})^2 d\mathbf{x} = \sum_{i=1}^C \sum_{j=1}^C q_i q_j G(\mathbf{u}_{q,i}, \mathbf{u}_{q,j}, \sigma_{p,i}^2 + \sigma_{q,j}^2) \quad (8b)$$

同时, 对于 $\int \widehat{q}(\mathbf{x})q(\mathbf{x})d\mathbf{x}$, 可近似为^[28-29]

$$\int \widehat{q}(\mathbf{x})q(\mathbf{x})d\mathbf{x} = E(\widehat{q}(\mathbf{x})) = E\left(\sum_{j=1}^C G(\mathbf{x}, \mathbf{u}_{q,j}, \sigma_{q,j}^2)\right) \approx \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C q_j G(\mathbf{x}_i, \mathbf{u}_{q,j}, \sigma_{q,j}^2) \quad (8c)$$

把式 (8a), (8b), (8c) 代入式 (7b) 得:

$$\begin{aligned} \arg \min_{q_j, \mathbf{u}_{q,j}, \sigma_{q,j}} J = & \arg \min_{q_j, \mathbf{u}_{q,j}, \sigma_{q,j}} \left(-2 \sum_{i=1}^C \sum_{j=1}^C p_i q_j \times \right. \\ & G(\mathbf{u}_{p,i}, \mathbf{u}_{q,j}, \sigma_{p,i}^2 + \sigma_{q,j}^2) + \\ & (1 + \lambda) \sum_{i=1}^C \sum_{j=1}^C q_i q_j G(\mathbf{u}_{q,i}, \mathbf{u}_{q,j}, \sigma_{q,i}^2 + \sigma_{q,j}^2) - \\ & \left. 2\lambda \sum_{i=1}^N \sum_{j=1}^C \frac{1}{N} q_j G(\mathbf{x}_i, \mathbf{u}_{q,j}, \sigma_{q,j}^2) \right) \quad (9) \end{aligned}$$

基于式 (9), 本研究将给出具体的 ML 迁移模糊系统参数学习规则.

2.2 参数学习规则

在式 (9) 中, $q_j, \mathbf{u}_{q,j}$ 以及 $\sigma_{q,j}$ 均为需优化的变量. 直接对这些参数优化是比较困难的. 本研究采用了模糊聚类、模糊神经网络等技术中经常采用的交替迭代策略来对式 (9) 进行参数优化^[30-35]. 在该迭代过程中主要包含如下三个主要的步骤:

步骤 1. 固定 $\mathbf{u}_{q,j}, \sigma_{q,j}$ 不变, 利用式 (9) 优化 q_j . 当 $\mathbf{u}_{q,j}, \sigma_{q,j}$ 不变时, 式 (9) 对于 q_j 的优化为典型的二次规划 (Quadratic programming, QP) 问题^[36]. 此时对应的优化问题为

$$\arg \min_{q_j} J = \arg \min_{q_j} \left(\sum_{i=1}^C \sum_{j=1}^C q_i q_j h_{ij} + \sum_{j=1}^C q_j \beta_j \right) \quad (10a)$$

其中,

$$h_{ij} = (1 + \lambda)G(\mathbf{u}_{q,i}, \mathbf{u}_{q,j}, \sigma_{q,i}^2 + \sigma_{q,j}^2) \quad (10b)$$

$$\beta_j = \left[-2 \sum_{i=1}^C p_i G(\mathbf{u}_{p,i}, \mathbf{u}_{q,j}, \sigma_{p,i}^2 + \sigma_{q,j}^2) \right] + \left[-2\lambda \sum_{i=1}^N \frac{1}{N} G(\mathbf{x}_i, \mathbf{u}_{q,j}, \sigma_{q,j}^2) \right] \quad (10c)$$

此时采用经典的二次规划算法即可得到 q_j 在当前迭代的解.

步骤 2. 固定 $q_j, \sigma_{q,j}$ 不变, 利用式 (9) 求解 $\mathbf{u}_{q,j}$. 此时式 (9) 对应取极值的必要条件可表示为

$$\mathbf{u}_{q,j} = \frac{\mathbf{b}}{a} \quad (11a)$$

$$\begin{aligned} a = & - \sum_{i=1}^C p_i q_j G_{1ij} (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} + \\ & (1 + \lambda) \sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j (G_{2ij} (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1}) - \\ & \lambda \sum_{i=1}^N \frac{1}{N} q_j G_{3ij} \times \sigma_{q,j}^{-2} \quad (11b) \end{aligned}$$

$$\begin{aligned} \mathbf{b} = & - \sum_{i=1}^C p_i q_j G_{1ij} (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} \mathbf{u}_{p,i} + \\ & (1 + \lambda) \sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j (G_{2ij} (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1}) \mathbf{u}_{q,i} - \\ & \lambda \sum_{i=1}^N \frac{1}{N} q_j G_{3ij} \times \sigma_{q,j}^{-2} \mathbf{x}_i \quad (11c) \end{aligned}$$

这里 $G_{1ij} = G(\mathbf{u}_{p,i}, \mathbf{u}_{q,j}, \sigma_{p,i}^2 + \sigma_{q,j}^2)$, $G_{2ij} = G(\mathbf{u}_{q,i}, \mathbf{u}_{q,j}, \sigma_{q,i}^2 + \sigma_{q,j}^2)$, $G_{3ij} = G(\mathbf{x}_i, \mathbf{u}_{q,j}, \sigma_{q,j}^2)$, 类似于 FCM 聚类等技术, 式 (11a) 给出的必要条件即可作为参数 $\mathbf{u}_{q,j}$ 在交替迭代学习过程中的学习规则.

步骤 3. 固定 $q_j, \mathbf{u}_{q,j}$ 不变, 利用式 (9) 求解 $\sigma_{q,j}$. 此时式 (9) 对应的极值的必要条件可表示为

$$\frac{\partial J}{\partial \sigma_{q,j}} = 0 \quad (12a)$$

由于式 (12a) 得不到参数 $\sigma_{q,j}$ 的解析解, 我们给出如下的梯度学习规则来进行参数学习:

$$\sigma_{q,j}(s+1) = \sigma_{q,j}(s) - \eta \frac{\partial J}{\partial \sigma_{q,j}(t)} \quad (12b)$$

其中,

$$\begin{aligned} \frac{\partial J}{\partial \sigma_{q,j}} = & 2 \sum_{i=1}^C p_i q_j G_{1ij} \times \\ & \left[d (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} - (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-2} \times \right. \\ & \left. (\mathbf{u}_{p,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \right] \sigma_{q,j} + \\ & (1 + \lambda) \left\{ -2 \sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j G_{2ij} \times \right. \\ & \left[d (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} - (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-2} \times \right. \\ & \left. (\mathbf{u}_{q,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \right] \sigma_{q,j} - 2 q_j^2 G_{2ij} \times \\ & \left. d (2 \sigma_{q,j}^2)^{-1} \sigma_{q,j} \right\} + 2 \lambda \sum_{i=1}^N \frac{1}{N} q_j G_{3ij} \times \\ & \left[d \sigma_{q,j}^{-1} - \sigma_{q,j}^{-2} (\mathbf{x}_i - \mathbf{u}_{q,j})^T (\mathbf{x}_i - \mathbf{u}_{q,j}) \right] \sigma_{q,j} \end{aligned}$$

此式可作为内循环的学习规则来对参数 $\sigma_{q,j}$ 学习, 其对应于外部迭代中的一步学习. 式 (11a)~(11c) 及 (12b) 的具体推导过程, 可详见附录 A.

2.3 算法描述

根据前一节推导所得到的学习规则, 给出如下的 ML 型迁移模糊系统学习算法:

算法 1. ML 迁移模糊系统训练算法 (ML-type transfer learning fuzzy system training algorithm)

初始化

$t = 0, q_j(0) = p_j(0), \mathbf{u}_{q,j}(0) = \mathbf{u}_{p,j}(0), \sigma_{q,j} = \sigma_{p,i}$; 计算 $J(0)$; 设置迭代最大次数 $tmax, smax$, 平衡因子 λ 和误差阈值 ε .

Repeat

$t = t+1$;

Step 1

利用式 (10a)~(10c) 更新 $q_j(t)$;

Step 2

利用式 (11a)~(11c) 更新 $\mathbf{u}_{q,j}(t)$;

Step 3

$s=0$;

$\tilde{\sigma}_{q,j}(t, s) = \sigma_{q,j}(t-1)$;

Repeat

$s = s + 1$;

利用式 (12a) 和式 (12b) 求 $\tilde{\sigma}_{q,j}(t, s)$;

Until $\|\tilde{\sigma}_{q,j}(t, s) - \tilde{\sigma}_{q,j}(t, s-1)\| \leq \varepsilon$ or

$s \geq smax$;

$\sigma_{q,j}(t) = \tilde{\sigma}_{q,j}(t, s)$;

计算 $J(t)$;

Until $\|J(t) - J(t-1)\| \leq \varepsilon$ or $t \geq tmax$.

对于所给出的 ML 迁移模糊系统训练算法, 以第 $t+1$ 步的迭代学习为例对其收敛性作如下分析:

1) 在第 $t+1$ 步, 固定 $\mathbf{u}_{q,j}(t), \sigma_{q,j}(t)$ 不变, 可得式 (10a), 即

$$\begin{aligned} \arg \min_{q_j} J(q_j, \mathbf{u}_{q,j}(t), \sigma_{q,j}(t)) = \\ \arg \min_{q_j} \left(\sum_{i=1}^C \sum_{j=1}^C q_i q_j h_{ij} + \sum_{j=1}^C q_j \beta_j \right) \end{aligned}$$

进而求得关于 q_j 的一阶偏导数: $\frac{\partial J}{\partial q_j} = 2 \sum_{i=1, i \neq j}^C q_i h_{ij} + 2 q_j h_{jj} + \beta_j$, 二阶偏导数: $\frac{\partial^2 J}{\partial q_j^2} = 2 h_{jj} = 2 + 2\lambda$. 由于 $\lambda > 0$, 可知 $\frac{\partial^2 J}{\partial q_j^2} > 0$, 因此 $J(q_j, \mathbf{u}_{q,j}(t), \sigma_{q,j}(t))$ 是变量 q_j 的严格凸函数. 由于 $J(q_j, \mathbf{u}_{q,j}(t), \sigma_{q,j}(t))$ 是 q_j 的凸函数, 可知式 (10a) 所得的 $q_j(t+1)$ 是 $J(q_j, \mathbf{u}_{q,j}(t), \sigma_{q,j}(t))$ 的全局最优解, 从而有 $J(q_j(t+1), \mathbf{u}_{q,j}(t), \sigma_{q,j}(t)) \leq J(q_j(t), \mathbf{u}_{q,j}(t), \sigma_{q,j}(t))$.

2) 在第 $t+1$ 步, 固定 $q_j(t+1), \sigma_{q,j}(t)$ 不变, 式 (11a) 所得的 $\mathbf{u}_{q,j}(t+1)$ 仅是 $J(q_j(t+1), \mathbf{u}_{q,j}, \sigma_{q,j}(t))$ 取极值时的必要而非充分条件, 此时 $\mathbf{u}_{q,j}(t+1)$ 是 $J(q_j(t+1), \mathbf{u}_{q,j}, \sigma_{q,j}(t))$ 的极值或鞍点. 故变量 $\mathbf{u}_{q,j}$ 的更新规则不能严格保证 $J(q_j(t+1), \mathbf{u}_{q,j}(t+1), \sigma_{q,j}(t)) \leq J(q_j(t+1), \mathbf{u}_{q,j}(t), \sigma_{q,j}(t))$ 一定满足.

3) 在第 $t+1$ 步, 固定 $q_j(t+1), \mathbf{u}_{q,j}(t+1)$ 不变, 由式 (12b) 对 $J(q_j(t+1), \mathbf{u}_{q,j}(t+1), \sigma_{q,j})$ 进行迭代优化. 根据梯度下降算法的性质可知, 所得到的解 $\sigma_{q,j}(t+1)$ 是 $J(q_j(t+1), \mathbf{u}_{q,j}(t+1), \sigma_{q,j})$ 的某个局部最优解的近似解, 即 $J(q_j(t+1), \mathbf{u}_{q,j}(t+1), \sigma_{q,j}(t+1)) \leq J(q_j(t+1), \mathbf{u}_{q,j}(t+1), \sigma_{q,j}(t))$.

据上述分析可知, 提出的 ML 型迁移模糊系统训练算法并不能保证严格的收敛. 这里所采用的优化技术借鉴了模糊聚类、模糊建模和核特征抽取等智能方法中常用的交替迭代优化技术^[21, 30-35, 37]. 正如经典的 FCM 等聚类算法, 其通常能收敛于某个局部最优解或鞍点. 虽然此类算法目前在理论上尚未被证明是严格收敛的或是有条件收敛的, 但已有的采用交替迭代优化技术的迭代算法表明此优化技术在大多数场合是非常简单而有效的^[21, 30-35, 37]. 后面的实验部分给出了本文算法在部分数据集上运行时的收敛曲线, 以便对其收敛性做一个直观的了解.

必须指出的是: 本文算法的收敛性是一个有待进一步深入探讨的开放性问题, 未来在该方面还有很多工作值得深入探讨. 另外, 由于初始化等不同因素的影响, 许多交叉迭代算法多次执行后最终可能收敛于不同的局部最优解. 这个问题也是交叉迭代

算法本身无法完全克服的. 针对此问题, 更可行的解决方案是探讨新的优化算法来求解给定的优化目标函数. 例如, 近年来具有更好寻优能力的进化计算优化技术 (遗传算法、粒子群算法等) 受到较多关注并被尝试应用于优化不同的建模模型.

3 实验研究

3.1 实验设置

本文在实验研究部分将分别利用合成数据以及谷氨酸发酵过程中采集的真实数据对提出的 ML 型迁移模糊系统构建方法进行验证和评估. 其中合成数据以及谷氨酸真实数据的具体描述将分别于第 3.2 节和第 3.3 节中给出. 在第 3.2 节和第 3.3 节, 对经典的基于 RSDE 密度估计的 ML 系统建模方法 (表示为 RSDE-MLFIS) 和本文提出的 ML 型迁移模糊系统建模方法 (表示为 ML-TFS) 性能进行了比较. 为了能够突显本文算法的更好适应性, 本文利用 RSDE-MLFIS 方法构造出三个模糊系统, 这三个模糊系统分别为: 1) 直接利用历史数据构建的模糊系统 FS (D1); 2) 直接利用含信息缺失的当前场景数据构建的模糊系统 FS (D2); 3) 利用历史数据和当前场景数据构造的 FS (D1+D2). 然后, 利用本文方法基于当前场景数据以及历史知识构建的 ML 型迁移模糊系统与以上三种模糊系统进行了性能比较. 在第 3.4 节, 将进一步对本文方法和两种经典的针对小样本数据集的模糊系统建模方法进行性能比较.

性能比较时采用了式 (13) 定义的评价指标^[38]:

$$J_0 = \frac{\sum_{n=1}^N e^2(n)}{N \times y_r} \quad (13)$$

其中, N 为测试样本集尺寸; $e(n) = y_{\text{target}}(n) - y_{\text{model}}(n)$; $y_r = [\max(y_{\text{target}}) - \min(y_{\text{target}})]^2$. 这里 $y_{\text{target}}(n)$ 为第 n 个测试样本输出, $y_{\text{model}}(n)$ 为第 n 个测试样本对应的模糊系统输出. J 越小表示泛化能力越好.

实验中, 算法中的外部最大迭代次数 ($tmax$) 设置为 50, 内部最大迭代次数 ($smax$) 设为 100, 阈值设为 $\varepsilon = 1E-7$.

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

3.2 模拟数据集实验

为了利用历史知识弥补当前场景下由于信息缺失造成受训系统泛化能力下降之缺陷, 合成数据集的构造需遵循如下的两个原则: 1) 历史场景和当前场景有较大的相关性, 它们之间既有很大的相似性,

又存在着一定的区别. 2) 当前场景采样的数据集存在一定的信息缺失.

为了表征如上的两个原则, 采用函数 $Y = f(x) = \sin(x) \times x$, $x \in [-10, 10]$ 来表示历史场景, 其用于产生历史场景数据集 (D1). 另一方面, 利用 $y = r \times Y = r \times f(x)$ 来表征当前场景, 其用来产生当前场景的训练数据集 (D2) 和测试集 (D2-test), 并且在构造训练集时, 在某些区间人为设置信息缺失. 为便于理解, 表 1 给出了相关符号的定义.

表 1 所涉及符号的定义

Table 1 Definition of the related notations

符号	定义
d	数据维数 (输入 + 输出)
D1	历史场景采样数据集
D2	当前场景采样训练数据集
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

根据上述设置, 利用函数生成 400 个输入输出数据对来构成用于表征历史场景的数据集 (D1), 另外通过函数分别生成 149 个数据对和 200 个数据对构成当前场景的训练集 (D2) 和测试集 (D2-test). 对于当前场景的训练集 (D2), 令数据对在输入采样区间 $[-6, -4]$ 和 $[0, 4]$ 内产生了信息缺失. 同时, 本实验设置历史场景和当前场景的相关系数分别为 0.9、0.85、0.8、0.75 和 0.7, 进而分别进行了实验研究. 图 2 (a) 给出了相关系数为 0.85 时表征历史场景和当前场景的两个数据生成函数, 图 2 (b) 示出了相关系数为 0.85 时, 两函数采样得到的历史场景数据集和当前场景训练数据集. 实验结果如表 2 和图 3 所示.

根据表 2 和图 3, 可给出如下的观察结论:

1) 从表 2 可以看出, 本文提出的 ML 迁移模糊系统取得较之于其他模糊系统更好的泛化性能.

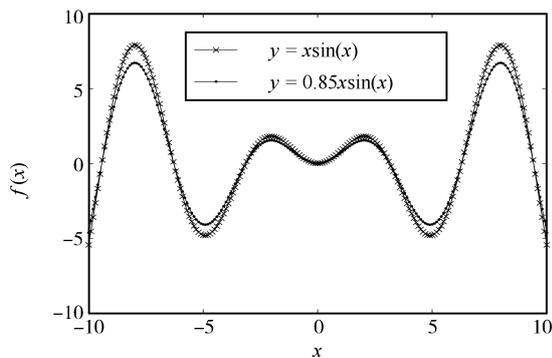
2) 对图 3 (a) 进行分析可知, 若直接使用历史模型对现有的场景进行性能测试, 其效果并不理想, 其原因在于现有场景与历史场景之间已存在一定的偏移, 也正是这种偏移导致了如果继续使用历史场景所得模型直接预测, 其性能将比较差, 无法达到与当前场景逼近的效果.

3) 对图 3 (b) 进行分析可知, 从图 3 (b) 可以明显的发现, 此模型在包含信息缺失的区间 $[-6, -4]$ 以及 $[0, 4]$ 上存在明显的性能恶化现象, 产生这种不足的原因源于两点: a) 我们在当前场景下采集的数据较之于历史场景数据而言是比较少的; b) 在信息缺失的空间上, 已有的方法只能达到同当前采样训练集逼近而并没能涉及到信息弥补, 因此其在信息缺失部分必然存在缺陷而最终导致整个系统性能的下降.

表 2 各种算法在模拟数据集上的泛化性能 (J0) 比较

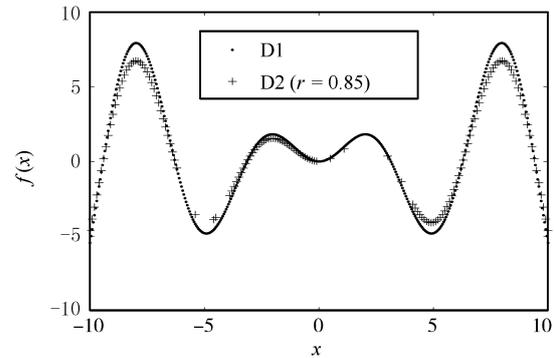
Table 2 Comparison of generation abilities (J0) of several algorithms on the synthetic datasets

当前场景信息缺失区间	r	FS(D1) 性能	FS(D2) 性能	FS(D1 + D2) 性能	ML-TFS 性能	
					FS (D2 + 历史知识)	λ 的取值 (利用交叉验证确定)
[-6, -4] and [0, 4]	0.9	8.937128E-004	6.508267E-003	4.721078E-004	1.141051E-004	100
	0.85	2.323151E-003	6.508821E-003	1.237628E-003	3.344995E-004	10 000
	0.8	4.743919E-003	6.508027E-003	2.483341E-003	7.133429E-004	20 000
	0.75	8.527649E-003	6.508873E-003	4.470552E-003	1.302578E-003	100 000
	0.7	1.420522E-002	6.508376E-003	7.400936E-003	1.135181E-003	100 000



(a) 表征历史和当前场景的函数

(a) Two functions to model the historical scene and the current scene



(b) 历史场景采样数据集 D1 和当前场景采样训练数据集 D2

(b) Sampling dataset of historical scene D1 and the sampling dataset of the current scene for training D2

图 2 相关系数为 0.85 时, 历史和当前场景示意函数及相应的采样数据集

Fig. 2 Two functions used to model the historical scene and the current scene with the related coefficient of 0.85 and the corresponding sampling datasets

4) 对图 3(c) 进行分析可知, 当使用历史场景数据与当前场景数据结合之后生成的模型进行测试时, 其结果好于以上两种方法, 但其逼近效果仍不是很理想. 造成这种结果的因素主要在于: 一方面历史场景和当前场景之间存在一定的漂移和偏差; 另一方面, 历史场景数据较之于当前场景下收集到的数据而言, 数据量较大, 因此在模型训练时, 其所占的比重较大, 得到的受训模型本身最终更加偏向于历史数据所得模型. 该方法的另一缺点是需要历史场景数据的支持. 但是一些高度机密的历史数据通常是很难获取的. 如果从历史场景仅能得到一些归纳出来的知识, 如对应的模糊系统的参数, 此时这种方法就变得不再可行.

5) 从图 3(d), 即本文方法结果可知: 首先, 本文的方法较之于图 3(a) 的方法有着更好的逼近效果; 其次, 与图 3(b) 的方法相比, 本文的方法在信息缺失区间 [-6, -4] 以及 [0, 4] 内利用历史知识较好地弥补了当前场景信息的不足, 使得系统的性能并

没有因为信息的缺失而显著下降; 最后, 本文方法与图 3(c) 的方法相比, 其不仅在逼近程度方面有着明显的改进, 而且值得指出的是: 本文的方法只需要历史知识 (历史模型参数) 以及当前场景的数据, 而不需要历史场景数据作为训练数据, 因而其在隐私保护方面也体现了较大的优势.

为了直观地观察本文所提方法的收敛情况, 图 4 给出了本文方法在合成数据集上某次运行时的收敛曲线情况. 从图 4 可以看出, 虽然该曲线不能保证相邻两步之间严格递减, 但总的趋势是逐渐递减的, 并最终在一个很小的变化范围内震荡, 此类现象在许多迭代算法中也是经常出现的^[37].

3.3 真实数据

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其余三种方法在真实数据集上的性能比较. 通过对采集到的工厂发酵数据进行筛选和分析后, 本文选取其中具有代表性且覆盖较为广泛的 21

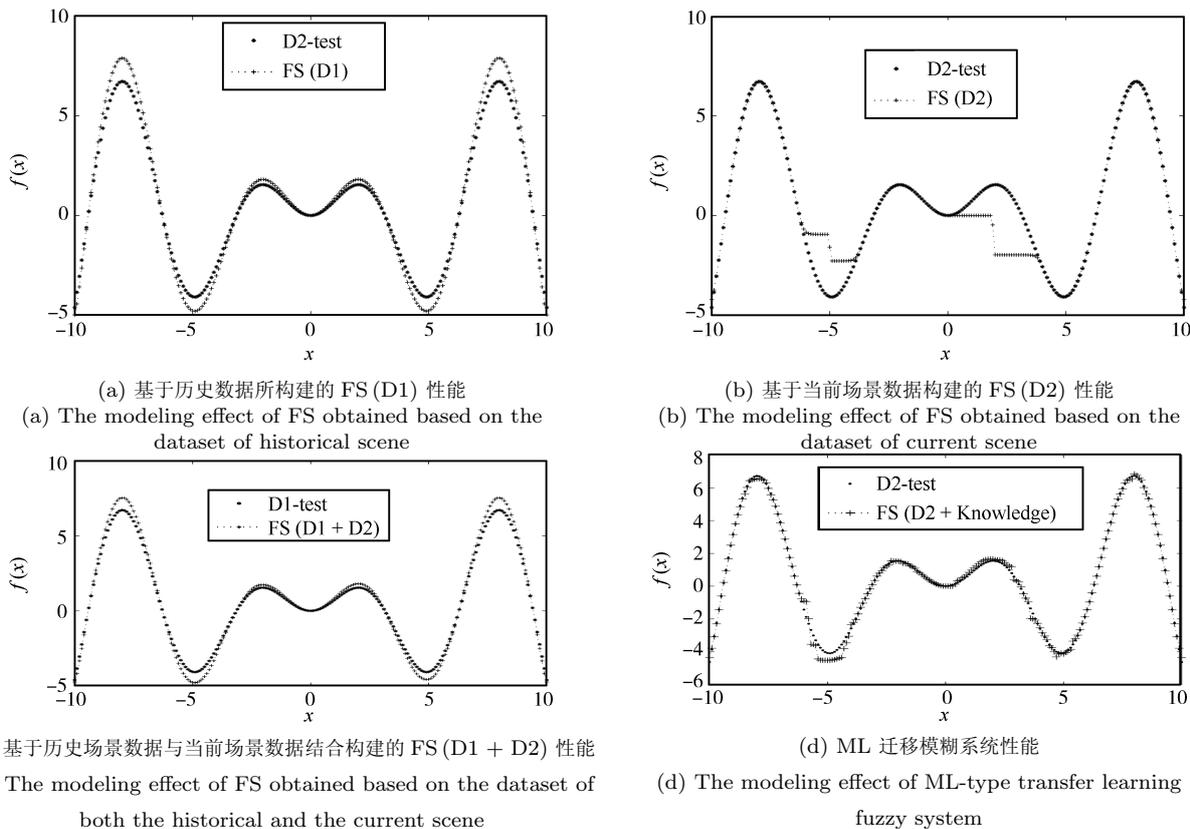


图 3 对当前场景测试集 (D2-test) 测试时各模糊系统建模效果比较

Fig.3 Comparison of modeling effect of several algorithms on the synthetic test dataset (D2-test)

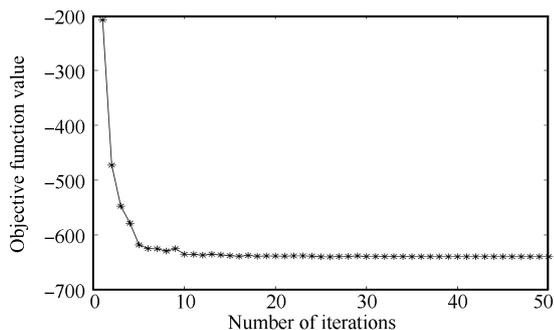


图 4 ML 迁移学习模糊系统算法在合成数据集上某次运行时的收敛曲线

Fig.4 The convergence curve of ML-type transfer learning fuzzy system obtained on the synthetic dataset with a certain time running

批共 294 组数据对, 取出其中 1~16 批共 224 组数据作为历史生产过程中得到的数据集 (D1), 并基于 D1 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) 在第 6 时刻至第 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), 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 批和 21 批) 信息完整, 因此本文利用这 28 组数据作为测试集 (D2-test). 该系统建模任

务是实现一个多输入多输出系统, 实验中把其分解为多个单输出系统来处理. 该系统输入参数分别为: 发酵过程的采样时间 (h), 葡萄糖浓度 (Glucose concentration), 菌体浓度 (Thalli concentration) 以及谷氨酸浓度 (Glutamic acid concentration); 其输出参数分别为: 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 3 和图 5 所示.

从表 3 的实验结果以及图 5 可以发现, 在该发酵过程仿真实验与前一节合成数据建模实验上所观察的结果基本是一致的. 当采用发酵生产中历史场景积累的历史知识进行迁移学习, 并通过当前场景采集到的含信息缺失的数据进行数据驱动模型训练, 最终能得到相应的 ML 型迁移模糊系统. 所得系统较之于直接利用历史场景数据构建的 FS (D1)、直接利用当前场景数据构建的 FS (D2) 和同时采用历史场景与当前场景数据组合构建的 FS (D1+D2), 都显示出了更好的泛化性能. 因此利用本方法构建的模糊系统可以更加准确地监控发酵过程, 并进行相应的预测.

同样地, 为了直观地观察本文方法的收敛情况, 图 6 给出了本文方法在真实数据集上对不同输出建模时在某次运行时得到的收敛曲线.

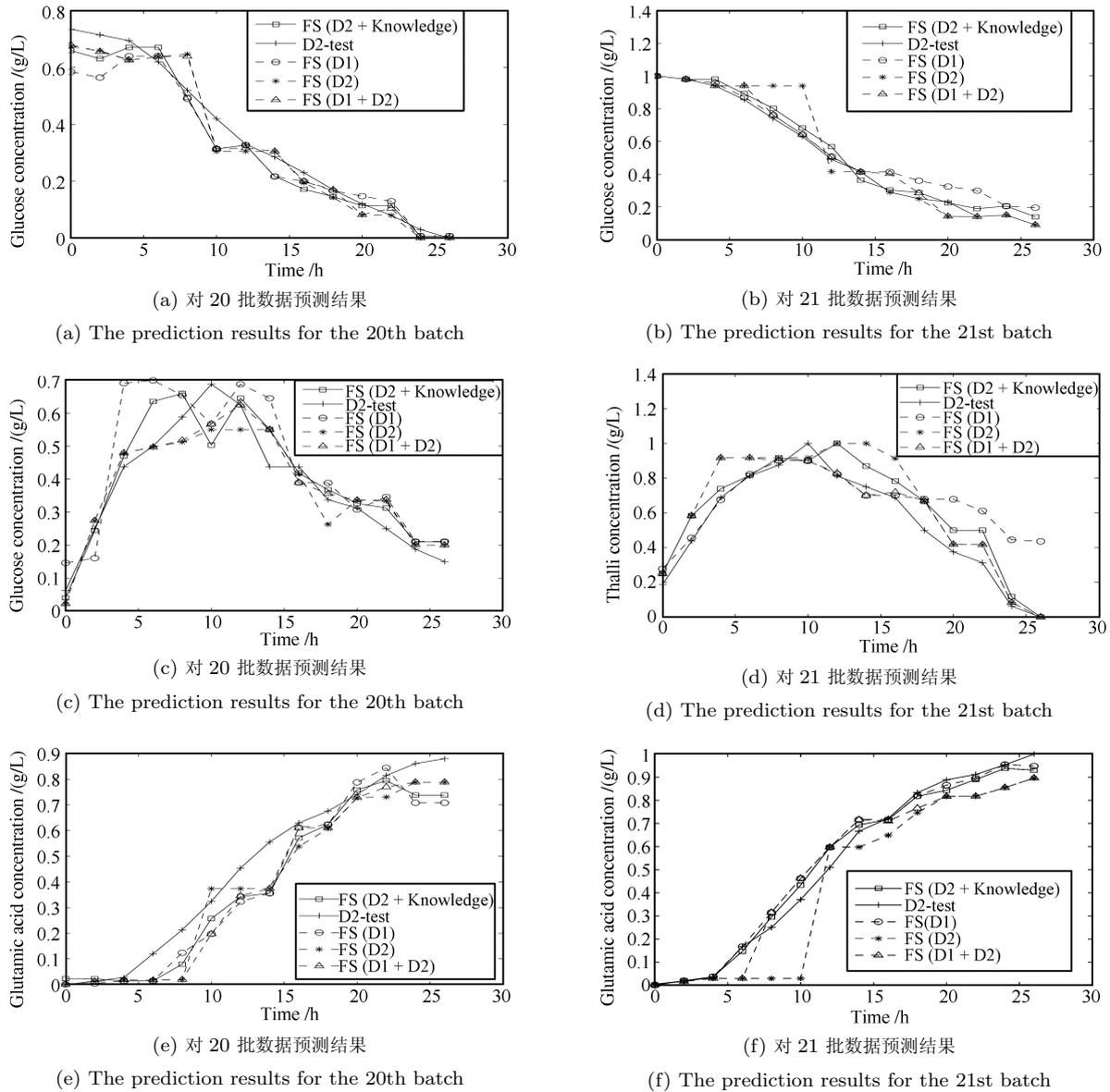


图 5 各种方法对发酵建模泛化性能比较 ((a), (b) 为各模糊系统对葡萄糖浓度预测泛化性能比较; (c), (d) 为各模糊系统对菌体浓度预测泛化性能比较果; (e), (f) 为各模糊系统对谷氨酸浓度预测泛化性能比较)

Fig. 5 The modeling effect of several algorithms for fermentation process modeling ((a) and (b) are the prediction results of glucose concentration; (c) and (d) are the prediction results of thalli concentration; (e) and (f) are the prediction results of glutamic acid concentration.)

表 3 各算法于真实数据 (发酵数据) 上的性能 (J0) 比较
Table 3 Comparison of generation abilities (J0) of several algorithms on the real-world (fermentation process) datasets

当前发酵过程信息 缺失时刻 (h)	输出属性	FS(D1) 性能	FS(D2) 性能	FS(D1 + D2) 性能	ML-TFS 性能	
					FS(D2 + 历史知识)	λ 的取值 (利用交叉验证确定)
6, 8, 10, 12	葡萄糖浓度	5.254659E-003	7.263440E-003	4.578475E-003	2.353124E-003	10
	菌体浓度	2.763054E-002	1.250041E-002	1.106049E-002	9.332362E-003	1 000
	谷氨酸浓度	6.366049E-003	1.318121E-002	7.523369E-003	5.396974E-003	1 000

3.4 与经典相关算法的性能比较

为了进一步评估本文建模方法的性能, 本节对提出的算法和两种经典的针对小样本数据集的模糊建模算法性能进行了比较. 所采用的两种相关算法分别是基于模糊聚类和支撑向量学习技术的 TSFS-SVR 方法^[10, 12] 和 FS-FCSVM 方法^[11].

3.4.1 模拟数据集

本节依旧采用第 3.2 节中产生的模拟数据集进行训练与预测. 对两种经典的相关算法主要针对两种情况来考量: 1) 在历史场景数据 (D1) 未知的情况下, 仅利用存在信息缺失的当前场景数据 (D2) 建模得到受训 FS (D2); 2) 在历史数据 (D1) 已知的情

况下, 利用历史数据 (D1) 以及当前场景数据 (D2), 共同建模得到受训 FS (D1 + D2). 最终在这两种情况下得到的系统和本文方法所得系统进行了性能比较. 表 4 和图 7 给出了相应的实验结果比较.

虽然相关文献表明 TSFS-SVR 和 FS-FCSVM 方法得到的模糊系统较之于传统模糊系统建模方法所得系统在小样本环境下显示出了更好的鲁棒性. 表 4 的实验结果表明, 不管是仅利用当前场景数据还是同时利用当前和历史场景数据进行建模, TSFS-SVR 和 FS-FCSVM 方法在当前场景存在信息缺失时, 由于没有迁移学习能力, 其泛化性能都逊于本文提出的迁移模糊系统建模算法. 图 7 所示的建模效果图亦能直观地印证这一点.

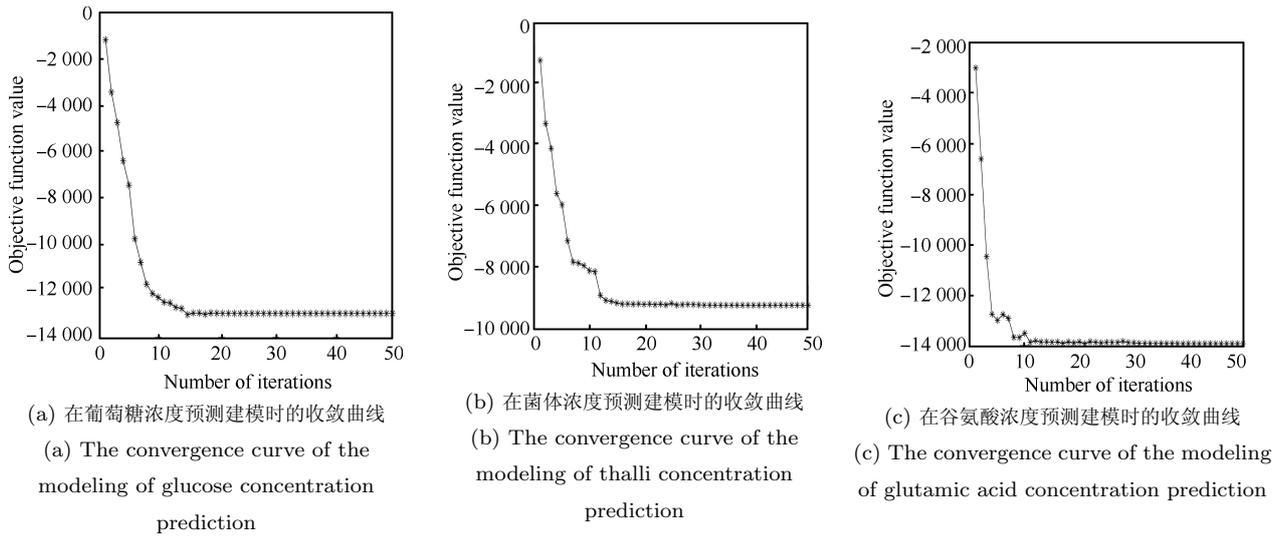


图 6 ML 迁移学习模糊系统算法在发酵数据集上某次运行时得到的收敛曲线

Fig. 6 The convergence curves of ML-type transfer learning fuzzy system obtained on the fermentation dataset with a certain time running

表 4 各种算法在模拟数据集上的泛化性能 (J0) 比较

Table 4 Comparison of generation abilities (J0) of several algorithms on the synthetic datasets

当前场景信息	r	TSFS-SVR 性能		FS-FCSVM 性能		ML-TFS 性能
缺失区间		D2	D1 + D2	D2	D1 + D2	D2 + 历史知识
$[-6, -4]$	0.9	1.848150E-003	8.007008E-004	2.683312E-003	1.066735E-003	1.141051E-004
and	0.85	1.821935E-003	2.155936E-003	2.663271E-003	2.336151E-003	3.344995E-004
$[0, 4]$	0.7	1.822882E-003	7.716933E-003	2.624874E-003	7.888074E-003	1.135181E-003

表 5 各算法于真实数据 (发酵数据) 上的性能 (J0) 比较

Table 5 Comparison of generation abilities (J0) of several algorithms on the real-world (fermentation process) datasets

当前发酵过程信息	输出属性	TSFS-SVR 性能		FS-FCSVM 性能		ML-TFS 性能
缺失时刻 (h)		D2	D1 + D2	D2	D1 + D2	D2 + 历史知识
	葡萄糖浓度	6.130038E-003	4.786755E-003	5.727892E-003	3.337321E-003	2.353124E-003
6, 8,	菌体浓度	1.353173E-002	1.178146E-002	1.546183E-002	1.363673E-002	9.332362E-003
10, 12	谷氨酸浓度	1.331485E-002	9.002180E-003	1.305347E-002	7.286164E-003	5.396974E-003

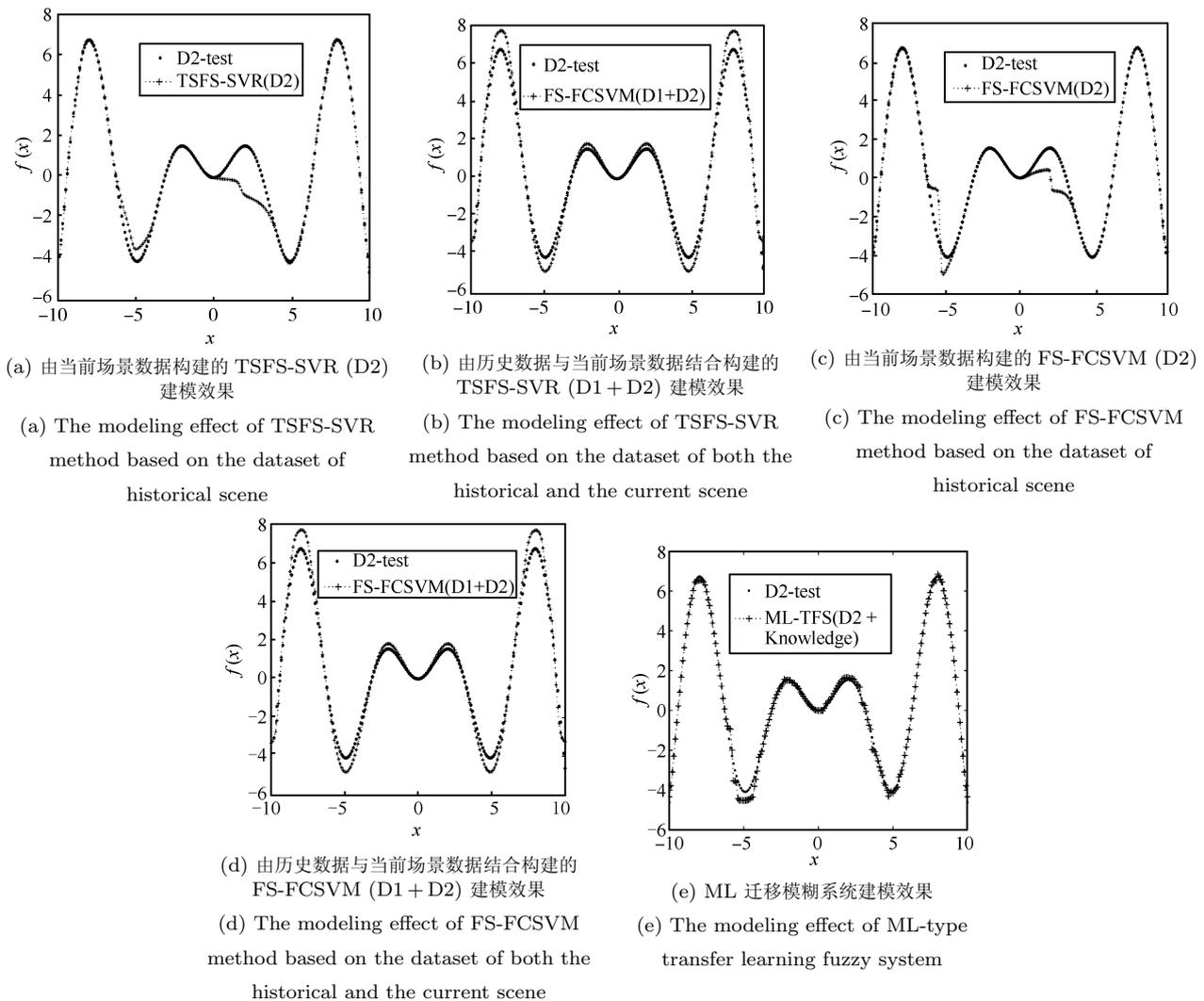


图 7 对当前场景测试集 (D2-test) 测试时各模糊系统建模效果比较

Fig. 7 Comparison of modeling effect of several algorithms on the synthetic test dataset (D2-test)

3.4.2 真实数据集

本节给出了本文方法、TSFS-SVR 方法和 FS-FCSVM 方法在发酵建模数据集上的性能比较. 实验结果如表 5 所示.

表 5 表明, 本文方法较之于 TSFS-SVR 方法和 FS-FCSVM 方法, 在发酵建模数据集亦显示出了更好的泛化性能. 这也进一步表明当引入迁移学习能力后, 本文方法有了从历史知识中学习的能力, 从而提高了受训系统的泛化能力. 虽然 TSFS-SVR 方法和 FS-FCSVM 方法通过支撑学习策略有了一定的鲁棒建模能力, 但由于在当前场景存在重要信息缺失时未能有针对性地对缺失信息进行信息补偿, 造成在该情形下建模效果依然不够理想.

4 结论

本文通过采用知识迁移的思想, 针对多场景相

关联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 ML 迁移模糊系统 (ML-TFS) 建模方法. 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, 且允许当前场景下采集到的数据存在一定程度的信息缺失. 由于这些特性的存在, 使得本方法不但能够根据历史知识进行当前场景下的信息补偿, 又能够起到对历史数据的隐私保护. 这些特性是传统的模糊建模方法所不具备的. 通过合成数据实验以及发酵过程的仿真实验, 仿真结果亦表明了本文方法较之于传统方法的更好适应性. 但目前本文方法仍存在一些不足之处. 例如, 由于本文方法在参数优化时得不到参数 $\sigma_{q,j}$ 的解析解, 在优化的过程中采用了梯度下降法学习此参数, 从而造成了本系统在训练数据尺度以及模糊规则数增加的情况下, 所运行的时间效率较长. 为了克服以上的缺陷, 我们在今后的工作中将致力于寻找一种快速算法来替代梯度下降法. 进一步

地研究其他类型具有迁移学习能力的模糊系统也是我们拟将探讨的工作。

附录 A 公式的证明

对于目标函数 (9):

$$\arg \min_{q_j, \mathbf{u}_{q,j}, \sigma_{q,j}} J = \arg \min_{q_j, \mathbf{u}_{q,j}, \sigma_{q,j}} \left(-2 \sum_{i=1}^C \sum_{j=1}^C p_i q_j \mathbf{u}_{p,i} \times G(\mathbf{u}_{p,i}, \mathbf{u}_{q,j}, \sigma_{p,i}^2 + \sigma_{q,j}^2) + (1 + \lambda) \sum_{i=1}^C \sum_{j=1}^C q_i q_j G(\mathbf{u}_{q,i}, \mathbf{u}_{q,j}, \sigma_{q,i}^2 + \sigma_{q,j}^2) - 2\lambda \sum_{i=1}^N \sum_{j=1}^C \frac{1}{N} q_j G(\mathbf{x}_i, \mathbf{u}_{q,j}, \sigma_{q,j}^2) \right)$$

为了便于描述和分析, 将其分为三个部分:

$$J_1 = -2 \sum_{i=1}^C \sum_{j=1}^C p_i q_j G(\mathbf{u}_{p,i}, \mathbf{u}_{q,j}, \sigma_{p,i}^2 + \sigma_{q,j}^2) \quad (A1)$$

$$J_2 = (1 + \lambda) \sum_{i=1}^C \sum_{j=1}^C q_i q_j G(\mathbf{u}_{q,i}, \mathbf{u}_{q,j}, \sigma_{q,i}^2 + \sigma_{q,j}^2) \quad (A2)$$

$$J_3 = -2\lambda \sum_{i=1}^N \sum_{j=1}^C \frac{1}{N} q_j G(\mathbf{x}_i, \mathbf{u}_{q,j}, \sigma_{q,j}^2) \quad (A3)$$

则 $J = J_1 + J_2 + J_3$, 进而得到:

$$\frac{\partial J}{\partial \mathbf{u}_{q,j}} = \frac{\partial J_1}{\partial \mathbf{u}_{q,j}} + \frac{\partial J_2}{\partial \mathbf{u}_{q,j}} + \frac{\partial J_3}{\partial \mathbf{u}_{q,j}} \quad (A4)$$

$$\frac{\partial J}{\partial \sigma_{q,j}} = \frac{\partial J_1}{\partial \sigma_{q,j}} + \frac{\partial J_2}{\partial \sigma_{q,j}} + \frac{\partial J_3}{\partial \sigma_{q,j}} \quad (A5)$$

1) 式 (11a), 式 (11b) 和式 (11c) 的推导过程:

$$G_{1ij} = G(\mathbf{u}_{p,i}, \mathbf{u}_{q,j}, \sigma_{p,i}^2 + \sigma_{q,j}^2) = (2\pi)^{-\frac{d}{2}} \times (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-\frac{d}{2}} \times \exp\left(-\frac{1}{2} (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{p,i} - \mathbf{u}_{q,j})^T \times (\mathbf{u}_{p,i} - \mathbf{u}_{q,j})\right) \quad (A6)$$

则

$$\frac{\partial J_1}{\partial \mathbf{u}_{q,j}} = -2 \sum_{i=1}^C p_i q_j \frac{\partial G_{1ij}}{\partial \mathbf{u}_{q,j}} \quad (A7)$$

$$\begin{aligned} \frac{\partial G_{1ij}}{\partial \mathbf{u}_{q,j}} &= G_{1ij} \times \left(-\frac{1}{2}\right) \left[(\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \times \right. \\ &\quad \left. (-1) + (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \times (-1) \right] = \\ &= G_{1ij} \times \frac{1}{2} \times \left[(\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) + \right. \\ &\quad \left. (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \right] = \\ &= \frac{1}{2} G_{1ij} \times \left[(\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} + (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} \right] \\ &= G_{1ij} \times (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} \times (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \end{aligned} \quad (A8)$$

将式 (A8) 代入式 (A7), 得:

$$\begin{aligned} \frac{\partial J_1}{\partial \mathbf{u}_{q,j}} &= -2 \sum_{i=1}^C p_i q_j \frac{\partial G_{1ij}}{\partial \mathbf{u}_{q,j}} = \\ &= -2 \sum_{i=1}^C p_i q_j (G_{1ij} \times (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} \times (\mathbf{u}_{p,i} - \mathbf{u}_{q,j})) = \\ &= -2 \sum_{i=1}^C p_i q_j G_{1ij} (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \end{aligned} \quad (A9)$$

令

$$G_{2ij} = G(\mathbf{u}_{q,i}, \mathbf{u}_{q,j}, \sigma_{q,i}^2 + \sigma_{q,j}^2) = (2\pi)^{-\frac{d}{2}} \times (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-\frac{d}{2}} \times \exp\left(-\frac{1}{2} (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{q,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{q,i} - \mathbf{u}_{q,j})\right) \quad (A10)$$

则

$$\begin{aligned} \frac{\partial J_2}{\partial \mathbf{u}_{q,j}} &= (1 + \lambda) \left[\sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j \left(\frac{\partial G_{2ij}}{\partial \mathbf{u}_{q,j}} \right) + \right. \\ &\quad \left. \sum_{\substack{i=1, \\ i \neq j}}^C q_j q_i \left(\frac{\partial G_{2ji}}{\partial \mathbf{u}_{q,j}} \right) + 2 q_j^2 \frac{\partial G_{2jj}}{\partial \mathbf{u}_{q,j}} \right] \end{aligned} \quad (A11)$$

当 $i \neq j$ 时,

$$\begin{aligned} \frac{\partial G_{2ij}}{\partial \mathbf{u}_{q,j}} &= G_{2ij} \times \left(-\frac{1}{2}\right) \left[(\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \times \right. \\ &\quad \left. (-1) + (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \times (-1) \right] = \\ &= G_{2ij} (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \end{aligned} \quad (A12)$$

$$\begin{aligned} \frac{\partial G_{2ji}}{\partial \mathbf{u}_{q,j}} &= G_{2ji} \times \left(-\frac{1}{2}\right) \times \\ &\quad \left[2 (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{q,j} - \mathbf{u}_{q,i}) \times (-1) \right] = \\ &= G_{2ji} \times \left(-\frac{1}{2}\right) \left[2 (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \times (-1) \right] = \\ &= G_{2ji} (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \end{aligned} \quad (A13)$$

当 $i = j$ 时,

$$\frac{\partial G_{2ij}}{\partial \mathbf{u}_{q,j}} = \mathbf{0} \quad (A14)$$

将式 (A12)、式 (A13) 和式 (A14) 分别代入式 (A11), 得:

$$\begin{aligned} \frac{\partial J_2}{\partial \mathbf{u}_{q,j}} &= (1 + \lambda) \left[\sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j \left(\frac{\partial G_{2ij}}{\partial \mathbf{u}_{q,j}} \right) + \right. \\ &\quad \left. \sum_{\substack{i=1, \\ i \neq j}}^C q_j q_i \left(\frac{\partial G_{2ji}}{\partial \mathbf{u}_{q,j}} \right) + q_j^2 \frac{\partial G_{2jj}}{\partial \mathbf{u}_{q,j}} \right] = \\ &= (1 + \lambda) \sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j \left(2 \times G_{2ij} (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \right) \end{aligned} \quad (A15)$$

令

$$G_{3ij} = G(\mathbf{x}_i, \mathbf{u}_{q,j}, \sigma_{q,j}^2) = (2\pi)^{-\frac{d}{2}} \times (\sigma_{q,j}^2)^{-\frac{d}{2}} \times \exp\left(-\frac{1}{2} \sigma_{q,j}^{-2} (\mathbf{x}_i - \mathbf{u}_{q,j})^T (\mathbf{x}_i - \mathbf{u}_{q,j})\right) \quad (A16)$$

则

$$\frac{\partial J_3}{\partial \mathbf{u}_{q,j}} = -2\lambda \sum_{i=1}^N \frac{1}{N} q_j \frac{\partial G_{3ij}}{\partial \mathbf{u}_{q,j}} \quad (\text{A17})$$

$$\begin{aligned} \frac{\partial G_{3ij}}{\partial \mathbf{u}_{q,j}} &= G_{3ij} \times \left(-\frac{1}{2} \right) \left[\sigma_{q,j}^2{}^{-1} (\mathbf{x}_i - \mathbf{u}_{q,j}) (-1) + \right. \\ &\quad \left. \sigma_{q,j}^2{}^{-1} (\mathbf{x}_i - \mathbf{u}_{q,j}) (-1) \right] = \\ &\quad G_{3ij} \times \sigma_{q,j}^2{}^{-1} (\mathbf{x}_i - \mathbf{u}_{q,j}) \end{aligned} \quad (\text{A18})$$

将式 (A18) 代入式 (A17), 得:

$$\begin{aligned} \frac{\partial J_3}{\partial \mathbf{u}_{q,j}} &= -2\lambda \sum_{i=1}^N \frac{1}{N} q_j \frac{\partial G_{3ij}}{\partial \mathbf{u}_{q,j}} = \\ &\quad -2\lambda \sum_{i=1}^N \frac{1}{N} q_j G_{3ij} \times \sigma_{q,j}^2{}^{-1} (\mathbf{x}_i - \mathbf{u}_{q,j}) \end{aligned} \quad (\text{A19})$$

将式 (A9)、式 (A15) 和式 (A19) 分别代入式 (A4), 并由 $\frac{\partial J}{\partial \mathbf{u}_{q,j}} = \frac{\partial J_1}{\partial \mathbf{u}_{q,j}} + \frac{\partial J_2}{\partial \mathbf{u}_{q,j}} + \frac{\partial J_3}{\partial \mathbf{u}_{q,j}} = 0$, 可得固定 q_j , $\sigma_{q,j}$ 不变时, 式 (9) 取极值的必要条件, 即

$$\begin{aligned} &-2 \sum_{i=1}^C p_i q_j G_{1ij} (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) + (1 + \lambda) \times \\ &\quad \sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j \left(2G_{2ij} (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \right) - \\ &\quad 2\lambda \sum_{i=1}^N \frac{1}{N} q_j G_{3ij} \sigma_{q,j}^2{}^{-1} (\mathbf{x}_i - \mathbf{u}_{q,j}) = 0 \end{aligned} \quad (\text{A20})$$

由式 (A20) 进一步化简可得:

$$\mathbf{u}_{q,j} = \frac{\mathbf{b}}{a} \quad (\text{A21})$$

其中

$$\begin{aligned} a &= - \sum_{i=1}^C p_i q_j G_{1ij} (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} + (1 + \lambda) \times \\ &\quad \sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j \left(G_{2ij} (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} \right) - \\ &\quad \lambda \sum_{i=1}^N \frac{1}{N} q_j G_{3ij} \sigma_{q,j}^2{}^{-1} \end{aligned} \quad (\text{A22})$$

$$\begin{aligned} \mathbf{b} &= - \sum_{i=1}^C p_i q_j G_{1ij} (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} \mathbf{u}_{p,i} + (1 + \lambda) \times \\ &\quad \sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j \left(G_{2ij} (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} \right) \mathbf{u}_{q,i} - \\ &\quad \lambda \sum_{i=1}^N \frac{1}{N} q_j G_{3ij} \times \sigma_{q,j}^2{}^{-1} \mathbf{x}_i \end{aligned} \quad (\text{A23})$$

式 (A21) ~ (A23) 即为正文中式 (11a) ~ (11c).

2) 式 (12b) 的推导过程:

根据式 (A1) 和式 (A6), 得:

$$\frac{\partial J_1}{\partial \sigma_{q,j}} = -2 \sum_{i=1}^C p_i q_j \frac{\partial G_{1ij}}{\partial \sigma_{q,j}} \quad (\text{A24})$$

$$\begin{aligned} \frac{\partial G_{1ij}}{\partial \sigma_{q,j}} &= (2\pi)^{-\frac{d}{2}} \times \\ &\quad \exp \left(-\frac{1}{2} (\mathbf{u}_{p,i} - \mathbf{u}_{q,j})^T (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \right) \times \\ &\quad \left(-\frac{d}{2} \right) \times (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-\frac{d+2}{2}} \times 2\sigma_{q,j} + \\ &\quad G_{1ij} \left(-\frac{1}{2} \times (-1) \times (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-2} \times 2\sigma_{q,j} \times \right. \\ &\quad \left. (\mathbf{u}_{p,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \right) = \\ &\quad \left(-\frac{d}{2} \right) G_{1ij} \times (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} \times 2\sigma_{q,j} + \\ &\quad \left(\frac{1}{2} \right) G_{1ij} \left((\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-2} \times 2\sigma_{q,j} \times \right. \\ &\quad \left. (\mathbf{u}_{p,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \right) = \\ &\quad \left(-\frac{1}{2} \right) G_{1ij} \left(d (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} \times 2\sigma_{q,j} - \right. \\ &\quad \left. (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-2} \times 2\sigma_{q,j} \times (\mathbf{u}_{p,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \right) = \\ &\quad (-1) G_{1ij} \left(d (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} - (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-2} \times \right. \\ &\quad \left. (\mathbf{u}_{p,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \right) \times \sigma_{q,j} \end{aligned} \quad (\text{A25})$$

将式 (A25) 代入式 (A24), 得:

$$\begin{aligned} \frac{\partial J_1}{\partial \sigma_{q,j}} &= -2 \sum_{i=1}^C p_i q_j \frac{\partial G_{1ij}}{\partial \sigma_{q,j}} = \\ &\quad 2 \sum_{i=1}^C p_i q_j G_{1ij} \left(d (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} - \right. \\ &\quad \left. (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-2} \times (\mathbf{u}_{p,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \right) \times \sigma_{q,j} \end{aligned} \quad (\text{A26})$$

根据式 (A2) 和式 (A10), 得:

$$\begin{aligned} \frac{\partial J_2}{\partial \sigma_{q,j}} &= (1 + \lambda) \left[\sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j \left(\frac{\partial G_{2ij}}{\partial \sigma_{q,j}} \right) + \right. \\ &\quad \left. \sum_{\substack{i=1, \\ i \neq j}}^C q_j q_i \left(\frac{\partial G_{2ji}}{\partial \sigma_{q,j}} \right) + 2q_j^2 \frac{\partial G_{2jj}}{\partial \sigma_{q,j}} \right] \end{aligned} \quad (\text{A27})$$

当 $i \neq j$ 时,

$$\begin{aligned} \frac{\partial G_{2ij}}{\partial \sigma_{q,j}} &= (2\pi)^{-\frac{d}{2}} \times \\ &\quad \exp \left(-\frac{1}{2} (\mathbf{u}_{q,i} - \mathbf{u}_{q,j})^T (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \right) \times \\ &\quad \left(-\frac{d}{2} \right) \times (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-\frac{d+2}{2}} \times 2\sigma_{q,j} + \\ &\quad G_{2ij} \left(-\frac{1}{2} \times (-1) \times (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-2} \times 2\sigma_{q,j} \times \right. \\ &\quad \left. (\mathbf{u}_{q,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \right) = \end{aligned}$$

$$\begin{aligned}
 & \left(-\frac{d}{2}\right) G_{2ij} \times (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} \times 2\sigma_{q,j} + \\
 & \left(\frac{1}{2}\right) G_{2ij} \left((\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-2} \times 2\sigma_{q,j} \times \right. \\
 & \left. (\mathbf{u}_{q,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \right) = \\
 & \left(-\frac{1}{2}\right) G_{2ij} \left(d (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} \times 2\sigma_{q,j} - \right. \\
 & \left. (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-2} \times 2\sigma_{q,j} \times (\mathbf{u}_{q,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \right) = \\
 & (-1) G_{2ij} \left(d (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} - (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-2} \times \right. \\
 & \left. (\mathbf{u}_{q,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \right) \times \sigma_{q,j} \tag{A28}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial G_{2ji}}{\partial \sigma_{q,j}} &= (2\pi)^{-\frac{d}{2}} \times \\
 & \exp\left(-\frac{1}{2} (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})^T (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})\right) \times \\
 & \left(-\frac{d}{2}\right) \times (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-\frac{d+2}{2}} \times 2\sigma_{q,j} + \\
 & G_{2ij} \left(-\frac{1}{2} \times (-1) \times (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-2} \times 2\sigma_{q,j} \times \right. \\
 & \left. (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})^T (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})\right) = \\
 & \left(-\frac{d}{2}\right) G_{2ij} \times (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} \times 2\sigma_{q,j} + \\
 & \left(\frac{1}{2}\right) G_{2ij} \left((\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-2} \times 2\sigma_{q,j} \times \right. \\
 & \left. (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})^T (\mathbf{u}_{q,j} - \mathbf{u}_{q,i}) \right) = \\
 & \left(-\frac{1}{2}\right) G_{2ij} \left(d (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} \times 2\sigma_{q,j} - \right. \\
 & \left. (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-2} \times 2\sigma_{q,j} \times (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})^T (\mathbf{u}_{q,j} - \mathbf{u}_{q,i}) \right) = \\
 & (-1) G_{2ij} \left(d (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} - (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-2} \times \right. \\
 & \left. (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})^T (\mathbf{u}_{q,j} - \mathbf{u}_{q,i}) \right) \times \sigma_{q,j} = \frac{\partial G_{2ij}}{\partial \sigma_{q,j}} \tag{A29}
 \end{aligned}$$

当 $i = j$ 时,

$$\begin{aligned}
 \frac{\partial G_{2ij}}{\partial \sigma_{q,j}} &= (2\pi)^{-\frac{d}{2}} \times \\
 & \exp\left(-\frac{1}{2} (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})^T (2\sigma_{q,j}^2)^{-1} (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})\right) \times \\
 & \left(-\frac{d}{2}\right) \times (2\sigma_{q,j}^2)^{-\frac{d+2}{2}} \times 4\sigma_{q,j} + \\
 & G_{2ij} \left(-\frac{1}{2} \times (-1) \times (2\sigma_{q,j}^2)^{-2} \times 4\sigma_{q,j} \times \right. \\
 & \left. (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})^T (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})\right) =
 \end{aligned}$$

$$\begin{aligned}
 & \left(-\frac{d}{2}\right) G_{2ij} \times (2\sigma_{q,j}^2)^{-1} \times 4\sigma_{q,j} + \\
 & \left(\frac{1}{2}\right) G_{2ij} \left((2\sigma_{q,j}^2)^{-2} \times 4\sigma_{q,j} \times \right. \\
 & \left. (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})^T (\mathbf{u}_{q,j} - \mathbf{u}_{q,i}) \right) = \\
 & \left(-\frac{1}{2}\right) G_{2ij} \left(d (2\sigma_{q,j}^2)^{-1} \times 4\sigma_{q,j} - (2\sigma_{q,j}^2)^{-2} \times \right. \\
 & \left. 4\sigma_{q,j} \times (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})^T (\mathbf{u}_{q,j} - \mathbf{u}_{q,i}) \right) = \\
 & (-2) G_{2ij} \left(d (2\sigma_{q,j}^2)^{-1} - (2\sigma_{q,j}^2)^{-2} \times \right. \\
 & \left. (\mathbf{u}_{q,j} - \mathbf{u}_{q,i})^T (\mathbf{u}_{q,j} - \mathbf{u}_{q,i}) \right) \times \sigma_{q,j} = \\
 & -2 G_{2ij} \times d \times (2\sigma_{q,j}^2)^{-1} \times \sigma_{q,j} \tag{A30}
 \end{aligned}$$

将式 (A28)、式 (A29) 和式 (A30) 分别代入式 (A27), 得:

$$\begin{aligned}
 \frac{\partial J_2}{\partial \sigma_{q,j}} &= (1 + \lambda) \left[\sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j \frac{\partial G_{2ij}}{\partial \sigma_{q,j}} + \right. \\
 & \left. \sum_{\substack{i=1, \\ i \neq j}}^C q_j q_i \frac{\partial G_{2ji}}{\partial \sigma_{q,j}} + q_j^2 \frac{\partial G_{2jj}}{\partial \sigma_{q,j}} \right] = \\
 & (1 + \lambda) \left[-2 \sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j G_{2ij} \left(d (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} - \right. \right. \\
 & \left. \left. (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-2} \times (\mathbf{u}_{q,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \right) \times \right. \\
 & \left. \sigma_{q,j} - 2q_j^2 G_{2ij} \times d \times (2\sigma_{q,j}^2)^{-1} \times \sigma_{q,j} \right] \tag{A31}
 \end{aligned}$$

根据式 (A3) 和式 (A16), 得:

$$\frac{\partial J_3}{\partial \sigma_{q,j}} = -2\lambda \sum_{i=1}^N \frac{1}{N} q_j \frac{\partial G_{3ij}}{\partial \sigma_{q,j}} \tag{A32}$$

$$\begin{aligned}
 \frac{\partial G_{3ij}}{\partial \sigma_{q,j}} &= (2\pi)^{-\frac{d}{2}} \times \\
 & \exp\left(-\frac{1}{2} (\mathbf{x}_i - \mathbf{u}_{q,j})^T \sigma_{q,j}^2^{-1} (\mathbf{x}_i - \mathbf{u}_{q,j})\right) \times \\
 & \left(-\frac{d}{2}\right) \times (\sigma_{q,j}^2)^{-\frac{d+2}{2}} \times 2\sigma_{q,j} + \\
 & G_{3ij} \left(-\frac{1}{2} \times (-1) \times \sigma_{q,j}^2^{-2} \times \right. \\
 & \left. (\mathbf{x}_i - \mathbf{u}_{q,j})^T (\mathbf{x}_i - \mathbf{u}_{q,j})\right) \times 2\sigma_{q,j} = \\
 & -1 \times G_{3ij} \times \left\{ d \sigma_{q,j}^2^{-1} - \right. \\
 & \left. \sigma_{q,j}^2^{-2} (\mathbf{x}_i - \mathbf{u}_{q,j})^T (\mathbf{x}_i - \mathbf{u}_{q,j}) \right\} \times \sigma_{q,j} \tag{A33}
 \end{aligned}$$

将式 (A33) 代入式 (A32), 得:

$$\begin{aligned} \frac{\partial J_3}{\partial \sigma_{q,j}} &= -2\lambda \sum_{i=1}^N \frac{1}{N} q_j \frac{\partial G_{3ij}}{\partial \sigma_{q,j}} = \\ &2\lambda \sum_{i=1}^N \frac{1}{N} q_j G_{3ij} \left(d \sigma_{q,j}^2{}^{-1} - \right. \\ &\left. \sigma_{q,j}^2{}^{-2} (\mathbf{x}_i - \mathbf{u}_{q,j})^T (\mathbf{x}_i - \mathbf{u}_{q,j}) \right) \times \sigma_{q,j} \end{aligned} \quad (\text{A34})$$

将式 (A26)、式 (A31) 和式 (A34) 代入式 (A5), 由 $\frac{\partial J}{\partial \sigma_{q,j}} = 0$ 可得, 可固定 $q_j, \mathbf{u}_{q,j}$ 不变时, 式 (9) 取极值的必要条件为

$$\begin{aligned} \frac{\partial J}{\partial \sigma_{q,j}} &= 2 \sum_{i=1}^C p_i q_j G_{1ij} \left(d (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-1} - \right. \\ &\left. (\sigma_{p,i}^2 + \sigma_{q,j}^2)^{-2} \times (\mathbf{u}_{p,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{p,i} - \mathbf{u}_{q,j}) \right) \times \\ &\sigma_{q,j} + (1 + \lambda) \left[-2 \sum_{\substack{i=1, \\ i \neq j}}^C q_i q_j G_{2ij} \left(d (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-1} - \right. \right. \\ &\left. \left. (\sigma_{q,i}^2 + \sigma_{q,j}^2)^{-2} \times (\mathbf{u}_{q,i} - \mathbf{u}_{q,j})^T (\mathbf{u}_{q,i} - \mathbf{u}_{q,j}) \right) \times \sigma_{q,j} - \right. \\ &\left. 2 q_j^2 G_{2ij} \times d \times (2 \sigma_{q,j}^2)^{-1} \times \sigma_{q,j} \right] + \\ &2\lambda \sum_{i=1}^N \frac{1}{N} q_j G_{3ij} \left(d \sigma_{q,j}^2{}^{-1} - \right. \\ &\left. \sigma_{q,j}^2{}^{-2} (\mathbf{x}_i - \mathbf{u}_{q,j})^T (\mathbf{x}_i - \mathbf{u}_{q,j}) \right) \times \sigma_{q,j} = 0 \end{aligned} \quad (\text{A35})$$

由于上式得不到解析解, 我们可用如下的梯度下降参数学习公式:

$$\sigma_{q,j}(t+1) = \sigma_{q,j}(t) - \eta \frac{\partial J}{\partial \sigma_{q,j}(t)} \quad (\text{A36})$$

式 (A36) 即为正文中式 (12b).

References

- Zadeh L A. Probability theory and fuzzy logic are complementary rather than competitive. *Technometrics*, 1995, **37**(3): 271–276
- Dubois D, Prade H. Fuzzy sets and probability: misunderstandings, bridges, and gaps. In: *Proceedings of the 2nd IEEE Conference on Fuzzy Systems*. San Francisco, CA: IEEE, 1993. 1059–1068
- Mendel J M. *Uncertain Rule-based Fuzzy Logic Systems Introduction and New Directions*. New York: Prentice-Hall, 2000
- Zhu L, Chung F L, Wang S T. Generalized fuzzy C-means clustering algorithm with improved fuzzy partitions. *IEEE Transactions on Systems, Man, and Cybernetics*, 2009, **39**(3): 578–591
- Aström K J, McAvoy T J. Intelligent control. *Journal of Process Control*, 1993, **2**(3): 115–127
- 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. *IEEE Transactions on Fuzzy Systems*, 2011, **19**(2): 210–226
- Chung F L, Deng Z H, Wang S T. An adaptive fuzzy-inference-rule-based flexible model for automatic elastic image registration. *IEEE Transactions on Fuzzy Systems*, 2009, **17**(5): 995–1010
- Dai X L, Khorrarn S. A feature-based image registration algorithm using improved chain-code representation combined with invariant moments. *IEEE Transactions on Geoscience and Remote Sensing*, 1999, **37**(5): 2351–2362
- Le Moigne J, Campbell W J, Cromp R F. An automated parallel, image registration technique based on the correlation of wavelet features. *IEEE Transactions on Geoscience and Remote Sensing*, 2002, **40**(8): 1849–1864
- Juang C F, Chiu S H, Chang S W. A self-organizing TS-type fuzzy network with support vector learning and its application to classification problems. *IEEE Transactions on Fuzzy Systems*, 2007, **15**(5): 998–1008
- Juang C F, Chiu S H, Shiu S J. Fuzzy system learned through fuzzy clustering and support vector machine for human skin color segmentation. *IEEE Transactions on Systems, Man, and Cybernetics*, 2007, **37**(6): 1077–1087
- Juang C F, Hsieh C D. TS-fuzzy system-based support vector regression. *Fuzzy Sets and Systems*, 2009, **160**(17): 2486–2504
- Lin C T, Yeh C M, Liang S F, Chung J F, Kumar N. Support-vector-based fuzzy neural network for pattern classification. *IEEE Transactions on Fuzzy Systems*, 2006, **14**(1): 31–41
- Pan J L, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010, **22**(10): 1345–1359
- Raina R, Battle A, Lee H, Packer B, Ng A Y. Self-taught learning: transfer learning from unlabeled data. In: *Proceedings of the 24th International Conference on Machine Learning*. New York, NY, USA: ACM, 2007. 759–766
- Xue G R, Dai W Y, Yang Q, Yu Y. Topic-bridged PLSA for cross-domain text classification. In: *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM, 2008. 627–634
- Pan S J, Tsang I W, Kwok J T, Yang Q. Domain adaptation via transfer component analysis. In: *Proceedings of the 21st international joint conference on Artificial intelligence*. San Francisco, CA, USA: ACM, 2009. 1187–1192
- Glorot X, Bordes A, Bengio Y. Domain adaptation for large-scale sentiment classification: a deep learning approach. In: *Proceedings of the 28th International Conference on Machine Learning*. Bellevue, Washington, USA: ACM, 2011. 513–520
- Duan L X, Tsang I W, Xu D. Domain transfer multiple kernel learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(3): 465–479
- Dai W Y, Yang Q, Xue G R, Yu Y. Self-taught clustering. In: *Proceedings of the 25th International Conference on Machine Learning*. New York, NY, USA: ACM, 2008. 200–207

- 21 Jang J S R, Sun C T, Mizutani E. *Neural-Fuzzy and Soft-Computing*. Upper Saddle River, NJ: Prentice-Hall, 1997
- 22 Wang Shi-Tong. *Neural-Fuzzy Systems and Their Application*. Beijing: Beihang University Press, 1998
(王士同. 神经模糊系统及其应用. 北京: 北京航空航天大学出版社, 1998)
- 23 Mamdani E H. Application of fuzzy logic to approximate reasoning using linguistic synthesis. *IEEE Transactions on Computers*, 1977, **26**(12): 1182–1191
- 24 Larsen P M. Industrial applications of fuzzy logic control. *International Journal of Man-Machine Studies*, 1980, **12**(1): 3–10
- 25 Takagi T, Sugeno M. Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 1985, **15**(1): 116–132
- 26 Azeem M F, Hanmandlu M, Ahmad N. Generalization of adaptive neuro-fuzzy inference systems. *IEEE Transactions on Neural Networks*, 2000, **11**(6): 1332–1346
- 27 Chung F L, Deng Z H, Wang S T. From minimum enclosing ball to fast fuzzy inference system training on large datasets. *IEEE Transactions on Fuzzy Systems*, 2009, **17**(1): 173–184
- 28 Mark G, Chao H. Probability density estimation from optimally condensed data samples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(10): 1253–1264
- 29 Deng Z H, Chung F L, Wang S T. FRSDE: fast reduced set density estimator using minimal enclosing ball approximation. *Pattern Recognition*, 2008, **41**(4): 1363–1372
- 30 Domeniconi C, Gunopulos D, Ma S, Yan B J, Al-Razgan M, Papadopoulos D. Locally adaptive metrics for clustering high dimensional data. *Data Mining and Knowledge Discovery Journal*, 2007, **14**(1): 63–97
- 31 Wu K L, Yu J, Yang M S. A novel fuzzy clustering algorithm based on a fuzzy scatter matrix with optimality tests. *Pattern Recognition Letters*, 2005, **26**(5): 639–652
- 32 Deng Z H, Choi K S, Chung F L, Wang S T. Enhanced soft subspace clustering integrating within-cluster and between-cluster information. *Pattern Recognition*, 2010, **43**(3): 767–781
- 33 Yu J, Cheng Q S, Huang H K. Analysis of the weighting exponent in the FCM. *IEEE Transactions on Systems Man, and Cybernetics-Part B: Cybernetics*, 2004, **34**(1): 643–639
- 34 Yang S, Yan S C, Zhang C, Tang X O. Bilinear analysis for kernel selection and nonlinear feature extraction. *IEEE Transactions on Neural Networks*, 2007, **18**(5): 1442–1452
- 35 Jang J S. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 1993, **23**(3): 665–685
- 36 Fan R E, Chan P H, Lin C J. Working set selection using second order information for training support vector machines. *Journal of Machine Learning Research*, 2005, **6**: 1889–1918
- 37 Frey B J, Dueck D. Clustering by passing messages between data points. *Science*, 2007, **315**(5814): 972–976
- 38 Gan M T, Hanmandlu M, Tan A H. From a Gaussian mixture model to additive fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 2005, **13**(3): 303–316



蒋亦樟 江南大学数字媒体学院博士研究生。主要研究方向为模式识别, 智能计算及应用。本文通信作者。

E-mail: jyz0512@163.com

(**JIANG Yi-Zhang** Ph.D. candidate at the School of Digital Media, Jiangnan University. His research interest covers pattern recognition, intel-

ligent computation and their applications. Corresponding author of this paper.)



邓赵红 博士, 江南大学数字媒体学院副教授。主要研究方向为模糊建模和智能计算。

E-mail: dzh666828@yahoo.com.cn

(**DENG Zhao-Hong** Ph.D., associate professor at the School of Digital Media, Jiangnan University. His research interest covers fuzzy modeling

and intelligent computation.)



王士同 江南大学数字媒体学院教授。主要研究方向为人工智能, 模式识别和生物信息。

E-mail: wxwangst@yahoo.com.cn

(**WANG Shi-Tong** Professor at the School of Digital Media, Jiangnan University. His research interest covers artificial intelligence, pattern recognition,

and bioinformatics.)