

基于迭代模糊聚类算法与K近邻和数据字典的集成TSK模糊分类器

张雄涛*^{①②} 蒋云良^② 潘兴广^{①③} 胡文军^② 王士同^①

^①(江南大学数字媒体学院 无锡 214122)

^②(湖州师范学院信息工程学院 湖州 313000)

^③(贵州民族大学工程实训中心 贵阳 550025)

摘要: 该文提出一种新型的集成TSK模糊分类器(IK-D-TSK), 首先通过并行学习的方式组织所有0阶TSK模糊子分类器, 然后每个子分类器的输出被扩充到原始(验证)输入空间, 最后通过提出的迭代模糊聚类算法(IFCM)作用在增强验证集上生成数据字典, 从而利用KNN对测试数据进行快速预测。IK-D-TSK具有以下优点: 在IK-D-TSK中, 每个0阶TSK子分类器的输出被扩充到原始输入空间, 以并行方式打开原始(验证)输入空间中存在的流形结构, 根据堆栈泛化原理, 可以保证提高分类精度; 和传统TSK模糊分类器相比, IK-D-TSK以并行方式训练所有的子分类器, 因此运行速度可以得到有效保证; 由于IK-D-TSK是在以IFCM & KNN所获得的数据字典的基础上进行分类的, 因此具有强鲁棒性。理论和实验验证了模糊分类器IK-D-TSK具有较高的分类性能、强鲁棒性和高可解释性。

关键词: TSK模糊分类器; 迭代模糊聚类算法; 数据字典; 可解释性

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2020)03-0746-09

DOI: [10.11999/JEIT190214](https://doi.org/10.11999/JEIT190214)

Iterative Fuzzy C-means Clustering Algorithm & K-Nearest Neighbor and Dictionary Data Based Ensemble TSK Fuzzy Classifiers

ZHANG Xiongtao^{①②} JIANG Yunliang^② PAN Xingguang^{①③}

HU Wenjun^② WANG Shitong^①

^①(School of Digital Media, Jiangnan University, Wuxi 214122, China)

^②(School of Information Engineering, Huzhou University, Huzhou 313000, China)

^③(Engineer Training Center, Guizhou Minzu University, Guiyang 550025, China)

Abstract: A new ensemble TSK fuzzy classifier (i.e. IK-D-TSK) is proposed. First, all zero-order TSK fuzzy sub-classifiers are organized in a parallel way, then the output of each sub-classifier is augmented to the original (validation) input space, finally, the proposed Iterative Fuzzy C-Means (IFCM) clustering algorithm generates dictionary data on augmented validation dataset, and then KNN is used to predict the result for test data. IK-D-TSK has the following advantages: the output of each zero-order TSK subclassifier is augmented to the original input space to open the manifold structure in parallel, according to the principle of stack generalization, the classification accuracy can be improved; Compared with traditional TSK fuzzy classifiers which trains sequentially, IK-D-TSK trains all the sub-classifiers in parallel, so the running speed can be effectively guaranteed; Because IK-D-TSK works based on dictionary data obtained by IFCM & KNN, it has strong robustness. The theoretical and experimental results show that IK-D-TSK has high classification performance, strong robustness and high interpretability.

Key words: TSK fuzzy classifier; Iterative Fuzzy C-Means (IFCM) clustering algorithm; Dictionary data; Interpretability

收稿日期: 2019-04-03; 改回日期: 2019-11-08; 网络出版: 2019-11-18

*通信作者: 张雄涛 1047897965@qq.com

基金项目: 国家自然科学基金(61572236, 61300151, 61772198, 61771193), 中央高校基本科研业务费专项资金(JUDCF13030)

Foundation Items: The National Natural Science Foundation of China (61572236, 61300151, 61772198, 61771193), The Fundamental Research Funds of the Central Universities (JUDCF13030)

1 引言

Takagi-Sugeno-Kang(TSK)模糊分类器^[1,2]具有强大的逼近能力和高可解释性,在很多领域,如模式识别^[1]、智能数据处理^[2]和自适应模糊控制^[3]等已经被得到成功的应用。但是,在一些场景中,比如众包问题^[4],网络图像标识^[5]等,通常拥有宝贵但少量的真实数据及其标签(在本文中称为验证数据),并且很容易地收集大量原始甚至多源训练数据,通常这些训练数据不可避免地含有噪音数据。但如果在这些训练数据上构建分类模型,其性能将会严重降低。换句话说,用原始训练数据直接训练TSK模糊分类器或集成TSK模糊分类器(即,多个原始训练数据子集)将明显不适合以上考虑的场景,由于以下原因:(1)含有噪音的原始训练数据,所训练得到的分类器可能不可靠;(2)现有的TSK模糊分类器的训练算法大多采用对验证数据进行交叉验证策略来确定相关参数,宝贵的验证数据没有得到直接和充分的利用;(3)当采用传统的集成TSK模糊分类器^[6]对原始数据进行分类时,由于中间变量的存在,中间子分类器并没有可解释性。因此,对于这样的应用场景,自然希望在充分利用宝贵验证数据的基础上进而构建一个高性能的集成模糊分类器。

基于以上原因,本文提出了一种新型的基于IFCM & KNN和数据字典的集成TSK模糊分类器,即IK-D-TSK。采用以下策略:(1)在每个TSK模糊子分类器中,将每个输入特征等间隔划分成5个固定模糊集,由此产生5种高斯隶属函数。它们的固定中心是 $[0, 0.25, 0.50, 0.75, 1.00]$ 。以这种方式,即使这5个高斯隶属函数的所有核宽不同,它们也会具有相应的语义解释^[6,7]:非常小、小、中、大、非常大;(2)为每个TSK模糊子分类器指定不同的模糊规则数。这是为了增加TSK模糊子分类器之间的多样性^[8],因为不同的专家由于具有不同的经验和知识对模糊规则的数量可能会有不同的意见或建议;(3)在IK-D-TSK集成框架中,第1阶段每个TSK模糊子分类器的输出以并行学习的方式被扩充到原始(验证)输入空间,然后该扩充的输入空间被反馈至第2阶段,即IFCM & KNN阶段,通过使用K最近邻(KNN)方法^[9]可以有效地加速测试速度;(4)利用最小学习机(LLM)^[6,10],能够快速训练每个模糊子分类器。

本文贡献总结如下:

(1)提出了一种集成TSK模糊分类器IK-D-TSK。从原理上来看,本文方法和文献^[11]类似,然而,文献^[11]的决策机制是建立在多源或含噪的训练集上,而IK-D-TSK建立在无噪的验证集上;并且,

IK-D-TSK也不同于现有的堆栈TSK模糊分类器^[6],要么是将上一层子分类器的输出作为下一层子分类器的输入,要么将上一层子分类器的输出加上原始输入空间作为下一层子分类器的输入。IK-D-TSK中所有子分类器以并行的方式快速构建,每个子分类器的输出被反馈到验证数据集上形成增强验证数据集,根据堆栈泛化原理^[12],如此可以不断地打开原始(验证)数据空间中的流行结构,使得分类变得相对容易。

(2)提出了一种新的基于IFCM & KNN和数据字典的快速预测机制。也就是在通过所提出的IFCM聚类算法生成数据字典之后,对于任何一个预测样本,使用KNN找到数据字典中距离其最近的 k 个点,最后基于投票策略,得到预测样本的标签。

(3)在IK-D-TSK中,由于所有的子分类器都是并行构建的,增加或删除一个子分类器变得相对容易。当增加一个子分类器时,只需要独立训练此分类器并将其输出反馈到增强验证空间中;当删除一个子分类器时,只需要去除增强验证空间中的对应输出即可,并且,不需要任何代码的改变。

(4)此外,由于所提出的集成TSK模糊分类器中所有模糊规则的前件部分,通过采用对每一个输入特征随机选择一个固定的高斯隶属度函数,使其具有简洁的物理意义,从而保证IK-D-TSK具有高可解释性。

2 IK-D-TSK

2.1 经典0阶TSK模糊分类器

在0阶TSK模糊分类器中,第 k 条模糊规则定义如下: IF x_1 is $A_1^k \wedge x_2$ is $A_2^k \wedge \dots \wedge x_n$ is A_n^k , THEN $f^k(\mathbf{x}) = a_0^k, k = 1, 2, \dots, K$ 。其中, A_i^k 是第 k 条模糊规则对应的第 i 个输入变量的模糊子集^[13], \wedge 是模糊结合算子, K 是模糊规则数, $f^k(\mathbf{x})$ 是第 k 条模糊规则的输出,并且 $\mathbf{x} = [x_1, x_2, \dots, x_d]$ 是输入向量。在经过一系列操作之后,TSK模糊分类器的输出可以被表述为^[6,14]

$$\begin{aligned} y^0 &= \sum_{k=1}^K \left(\frac{\omega^k(\mathbf{x})}{\sum_{k'=1}^K \omega^{k'}(\mathbf{x})} \right) f^k(\mathbf{x}) \\ &= \sum_{k=1}^K \tilde{\omega}^k f^k(\mathbf{x}) \end{aligned} \quad (1)$$

其中, ω^k 和 $\tilde{\omega}^k$ 分别表示与模糊子集 A_i^k 相关联的模糊隶属度和归一化模糊隶属度,当使用高斯函数作为模糊隶属函数,它们可以被计算为

$$\left. \begin{aligned} \omega_{A_i^k}(x_i) &= \exp(-(x_i - v_i^k)^2 / 2\delta_i^k) \\ \omega^k(\mathbf{x}) &= \prod_{i=1}^d \omega_{A_i^k}(x_i) \end{aligned} \right\} \quad (2)$$

其中, 参数 v_i^k 和 δ_i^k 分别是它的中心和核宽。

对于输入向量 \mathbf{x}

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

$$\mathbf{a}_g = (a_0^1, a_0^2, \dots, a_0^K)^T \quad (2b)$$

然后, 式(1)中的0阶TSK模糊分类器的输出可被表示为式(3)线性回归模型^[9]

$$\begin{aligned} \min J_{\text{IFCM}}^q(\mathbf{U}, \mathbf{V}) &= \sum_{i=1}^{N_q} \sum_{j=1}^{Q \cdot c} (\mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q))^2 \|\mathbf{x}_i^q - \mathbf{v}_j^q\|^2 + \sum_{i=N_q+1}^{N_q+(Q-1) \cdot c} \sum_{j=1}^{Q \cdot c} (\mu^q(\mathbf{v}_i^{\bar{q}}, \mathbf{v}_j^q))^2 \|\mathbf{v}_i^{\bar{q}} - \mathbf{v}_j^q\|^2 \quad (4) \\ \text{s.t. } \sum_{j=1}^c \mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q) + (1+\lambda) \sum_{j=c+1}^{Q \cdot c} \mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q) &= 1, \\ i &\in \{1, 2, \dots, N_q\} \end{aligned} \quad (5)$$

$$\begin{aligned} (1+\lambda) \sum_{j=1}^c \mu^q(\mathbf{v}_i^{\bar{q}}, \mathbf{v}_j^q) + \sum_{j=c+1}^{Q \cdot c} \mu^q(\mathbf{v}_i^{\bar{q}}, \mathbf{v}_j^q) &= 1, \\ i &\in \{N_q+1, N_q+2, \dots, N_q+(Q-1) \cdot c\} \end{aligned} \quad (6)$$

其中 \mathbf{x}_i^q 是第 q 类的第 i 个样本点, \mathbf{v}_j^q 是第 q 类的第 j 个中心点, N_q 是第 q 类的样本数, c 是每一类的中心点数, $\mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q)$ 表示第 q 类中第 i 个样本点 \mathbf{x}_i^q 属于第 j 个中心点 \mathbf{v}_j^q 的隶属度, $\mu^q(\mathbf{v}_i^{\bar{q}}, \mathbf{v}_j^q)$ 表示 q 类外第 i 个中心点属于 q 类的第 j 个中心点的隶属度, λ 是平衡因子, 用于控制 q 类内的中心点和 q 类外中心点的疏

$$\begin{aligned} \bar{J}_{\text{IFCM}}^q &= J_{\text{IFCM}}^q + \sum_{i=1}^{N_q} \beta_i \left(1 - \sum_{j=1}^c \mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q) - (1+\lambda) \sum_{j=c+1}^{Q \cdot c} \mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q) \right) \\ &\quad + \sum_{i=N_q+1}^{N_q+(Q-1) \cdot c} \beta_i \left(1 - (1+\lambda) \sum_{j=1}^c \mu^q(\mathbf{v}_i^{\bar{q}}, \mathbf{v}_j^q) - \sum_{j=c+1}^{Q \cdot c} \mu^q(\mathbf{v}_i^{\bar{q}}, \mathbf{v}_j^q) \right) \end{aligned} \quad (7)$$

其中, β_i 是拉格朗日乘子。可以导出关于 q 类模糊隶属度和中心点的更新方程为

$$\mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q) = \frac{\beta_i}{2\|\mathbf{x}_i^q - \mathbf{v}_j^q\|^2}, j \in \{1, 2, \dots, c\} \quad (8)$$

$$\mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q) = \frac{(1+\lambda)\beta_i}{2\|\mathbf{x}_i^q - \mathbf{v}_j^q\|^2}, j \in \{c+1, c+2, \dots, Q \cdot c\} \quad (9)$$

$$\mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q) = \frac{1}{\|\mathbf{x}_i^q - \mathbf{v}_j^q\|^2 \left\{ \sum_{k=1}^c (1/\|\mathbf{x}_i^q - \mathbf{v}_k^q\|^2) + \sum_{k=c+1}^{Q \cdot c} [(1+\lambda)^2/\|\mathbf{x}_i^q - \mathbf{v}_k^q\|^2] \right\}}, i \in \{1, 2, \dots, N_q\}, j \in \{1, 2, \dots, c\} \quad (12)$$

$$\mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q) = \frac{1+\lambda}{\|\mathbf{x}_i^q - \mathbf{v}_j^q\|^2 \left\{ \sum_{k=1}^c (1/\|\mathbf{x}_i^q - \mathbf{v}_k^q\|^2) + \sum_{k=c+1}^{Q \cdot c} [(1+\lambda)^2/\|\mathbf{x}_i^q - \mathbf{v}_k^q\|^2] \right\}}, i \in \{1, 2, \dots, N_q\}, j \in \{c+1, c+2, \dots, Q \cdot c\} \quad (13)$$

当 $i \in \{N_q+1, N_q+2, \dots, N_q+(Q-1) \cdot c\}$ 时, 将式(10)和式(11)代入式(6)后, 可以得到

$$y^0 = \mathbf{a}_g^T \mathbf{x}_g \quad (3)$$

2.2 迭代模糊聚类算法IFCM

给定数据集 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in R^{N \times D}$, 其中, N 是总的样本数, D 是样本维度, 假定共有 Q 类, 用FCM初始化每一类样本的中心点 $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_Q\}$, N_q 为第 q 类的样本数, 其中 $1 \leq q \leq Q$ 。当固定第 q 类时, IFCM的目标函数形式为

离程度。当固定 q 类中心点时, 将使其与其他类的中心点尽量远离, 因此, λ 的值应该大于0。在此, 本文将 λ 的值设为0.3。 $\|\mathbf{x}_i^q - \mathbf{v}_j^q\|$ 表示在 q 类中, 第 i 个样本和第 j 个中心点之间的欧氏距离; $\|\mathbf{v}_i^{\bar{q}} - \mathbf{v}_j^q\|$ 表示 q 类外的第 i 个中心点和 q 类内的第 j 个中心点之间的距离, 其中 $\mathbf{v}_i^{\bar{q}}$ 表示不属于 q 类的中心点。如式(4)—式(6)所示, 当 $i \in \{1, 2, \dots, N_q\}$, 式(5)通过加强该约束中的第2项来迫使所有的样本更多地落在 q 类内部, 当 $i \in \{N_q+1, N_q+2, \dots, N_q+(Q-1) \cdot c\}$ 时, 式(6)通过加强该约束中的第1项来迫使所有样本更多地落入第 q 类之外。

根据拉格朗日优化方法^[9], 对于式(4)—式(6), 可以得到式(7)的拉格朗日函数

$$\mu^q(\mathbf{v}_i^{\bar{q}}, \mathbf{v}_j^q) = \frac{(1+\lambda)\beta_i}{2\|\mathbf{v}_i^{\bar{q}} - \mathbf{v}_j^q\|^2}, j \in \{1, 2, \dots, c\} \quad (10)$$

$$\mu^q(\mathbf{v}_i^{\bar{q}}, \mathbf{v}_j^q) = \frac{\beta_i}{2\|\mathbf{v}_i^{\bar{q}} - \mathbf{v}_j^q\|^2}, j \in \{c+1, c+2, \dots, Q \cdot c\} \quad (11)$$

当 $i \in \{1, 2, \dots, N_q\}$ 时, 将式(8)和式(9)代入式(5)后, 可以得到

$$\mu^q(\mathbf{v}_i^q, \mathbf{v}_j^q) = \frac{1+\lambda}{\|\mathbf{x}_i - \mathbf{v}_j^q\|^2 \left\{ \sum_{k=1}^c \left((1+\lambda)^2 / \|\mathbf{v}_i^q - \mathbf{v}_k^q\|^2 \right) + \sum_{k=c+1}^{Q \cdot c} \left[1 / \|\mathbf{v}_i^q - \mathbf{v}_k^q\|^2 \right] \right\}},$$

$$i \in \{N_q + 1, N_q + 2, \dots, N_q + (Q - 1) \cdot c\}, j \in \{1, 2, \dots, c\} \quad (14)$$

$$\mu^q(\mathbf{v}_i^q, \mathbf{v}_j^q) = \frac{1}{\|\mathbf{x}_i - \mathbf{v}_j^q\|^2 \left\{ \sum_{k=1}^c \left((1+\lambda)^2 / \|\mathbf{v}_i^q - \mathbf{v}_k^q\|^2 \right) + \sum_{k=c+1}^{Q \cdot c} \left[1 / \|\mathbf{v}_i^q - \mathbf{v}_k^q\|^2 \right] \right\}},$$

$$i \in \{N_q + 1, N_q + 2, \dots, N_q + (Q - 1) \cdot c\}, j \in \{c + 1, c + 2, \dots, Q \cdot c\} \quad (15)$$

$$\frac{\partial \bar{J}_{\text{IFCM}}^q}{\partial \mathbf{v}_j^q} = -2(\mu(\mathbf{x}_i^q, \mathbf{v}_j^q))^2 \|\mathbf{x}_i^q - \mathbf{v}_j^q\| - 2(\mu^q(\mathbf{v}_i^q, \mathbf{v}_j^q))^2 \|\mathbf{v}_i^q - \mathbf{v}_j^q\| = 0 \quad (16)$$

$$\mathbf{v}_j^q = \frac{\sum_{i=1}^{N_q} (\mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q))^2 \mathbf{x}_i^q + \sum_{i=N_q+1}^{N_q+(Q-1) \cdot c} (\mu^q(\mathbf{v}_i^q, \mathbf{v}_j^q))^2 \mathbf{v}_i^q}{\sum_{i=1}^{N_q} (\mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q))^2 + \sum_{i=N_q+1}^{N_q+(Q-1) \cdot c} (\mu^q(\mathbf{v}_i^q, \mathbf{v}_j^q))^2} \quad (17)$$

上述更新方程为迭代模糊聚类算法IFCM提供了核心参数优化公式。通过在数据集中的每个类上运行FCM, IFCM从每个类的初始中心点开始。然后, 从 $q=1$ 到 Q , 依次执行迭代式(12)、式(13)、式(14)、式(15)和式(17), 直到所有中心点保持几乎不变或达到最大迭代次数为止。将IFCM算法总结如下, 详见表1。

表1 IFCM算法

<p>输入: 数据集$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in R^{N \times D}$, 及其标签$\mathbf{Y} = \{y_1, y_2, \dots, y_N\}$, 真实类别数$Q$, 每一类的聚类中心点$c$, 每一类的样本数$N_1, N_2, \dots, N_Q$, 最大误差阈值$\tau$。</p> <p>输出: 中心点矩阵$\mathbf{V}$及其标签。</p> <p>(1)通过FCM初始化每类中的中心点, 然后形成中心点矩阵\mathbf{V}。初始化$q=1$, 其中$1 \leq q \leq Q$。</p> <p>(2)重复:</p> <p>(a)重复:</p> <p>①当$i \in \{1, 2, \dots, N_q\}$时, 通过式(12)和式(13)计算隶属度$\mu^q(\mathbf{x}_i^q, \mathbf{v}_j^q)$; 当$i \in \{N_q+1, N_q+2, \dots, N_q+(Q-1) \cdot c\}$时, 通过式(14)和式(15)计算隶属度$\mu^q(\mathbf{v}_i^q, \mathbf{v}_j^q)$;</p> <p>②通过式(17)计算中心点$\mathbf{v}_j^q$;</p> <p>(b)直到中心点矩阵保持几乎不变或达到内部迭代的最大次数为止;</p> <p>(c)利用\mathbf{v}_j^q更新中心点矩阵\mathbf{V}, 并且$q = (q+1) \bmod Q$;</p> <p>(3)直到$\max_{j \in \{1, 2, \dots, Q \cdot c\}} \ \mathbf{v}_j^q - \mathbf{v}_j^{q-1}\ < \tau$或达到外部最大迭代次数为止;</p> <p>(4)根据中心点矩阵\mathbf{V}输出所有的中心点及其标签。</p>
--

2.3 IK-D-TSK

图1描述了IK-D-TSK的框架, 其中 \mathbf{D}_{tr} 表示总训练数据集, $\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_L$ 表示 \mathbf{D}_{tr} 的子集, TSK_1, TSK_2, ..., TSK_L表示 L 个0阶TSK子分类

器, 以及 $F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_L(\mathbf{x})$ 表示它们的输出函数。IK-D-TSK包括两个阶段, 即增强验证数据生成和数据字典生成及预测。将数据集划分为训练集、验证集和测试集。在第1阶段, IK-D-TSK从训练子集 $\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_L$ 开始, 其中 \mathbf{D}_i 是从训练子集 \mathbf{D}_{tr} 中随机抽样生成, 并且 $\mathbf{D}_1 \cup \mathbf{D}_2 \cup \dots \cup \mathbf{D}_L = \mathbf{D}_{\text{tr}}$ 。这些子集被作为 L 个0阶TSK模糊子分类器的输入, 以并行学习的方式训练。每个0阶TSK模糊分类器具有不同数量的模糊规则, 以保证 L 个0阶TSK模糊子分类器之间的多样性。对于每个子分类器, 本文将每个输入特征等间隔划分成5个固定的模糊集, 构造5种高斯隶属度函数。它们的固定中心是0, 0.25, 0.50, 0.75, 1.00, 它们具有各自的语义解释: 非常小、小、中、大、非常大。每个高斯隶属函数的宽度可以不同, 但不会显著影响其语义解释。本文将随机正值作为每个隶属函数的宽度。基于以上考虑, 相应地使用LLM^[6,10]独立地快速学习每个子分类器的后件参数。所有子分类器进行并行训练后, 将 L 个输出函数作用在验证数据集上, 对于每个验证数据, 生成 L 个增强特征。将所有验证数据的原始特征和增强特征合并, 形成具有 $(d+L)$ 维的增强验证数据集。在这样一个 $(d+L)$ 维的增强验证空间中, 可以将原始验证空间中的流形打开, 从而可以以堆栈方式实现更好的线性可分离性。换句话说, 根据堆栈泛化原理^[12], 与原始验证数据集相比, 增强验证数据集更有助于提高分类性能。

在生成增强验证数据集之后, IK-D-TSK进入到第2阶段。为了实现利用KNN^[9]对训练或隐含样本的快速预测, 希望有少量且有代表性的数据样本作为数据字典供KNN使用。然而, 增强验证数据

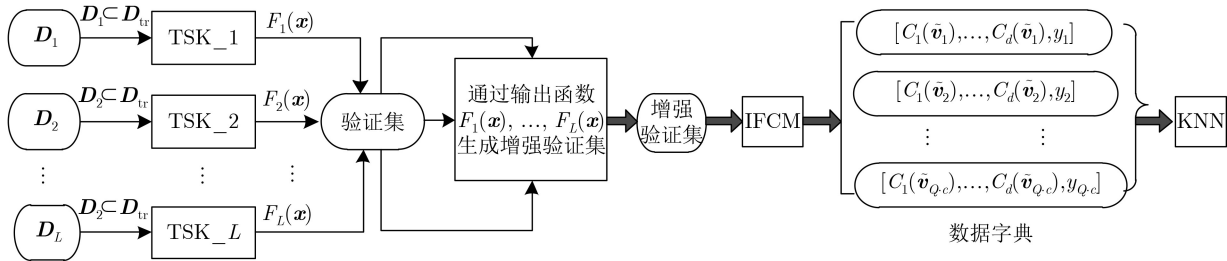


图1 IK-D-TSK的框架

集有时数据量相对太大。在这个阶段，使用本文提出的IFCM聚类算法来找出增强验证数据集中每个类的代表性中心点。在运行IFCM之后，所有去掉增强特征的代表性中心点及其标签被称为数据字典，即 $[C_1(\tilde{\mathbf{v}}_i), C_2(\tilde{\mathbf{v}}_i), \dots, C_d(\tilde{\mathbf{v}}_i), y_i], 1 < i < (Q \cdot c)$ ，如图1所示，用于KNN对训练或隐含样本的预测。也就是说，对于一个训练样本或隐含样本，KNN从所有具有代表性的中心点中选择 k 个最近邻，然后基于投票策略，根据所有 k 个所选中心点的标签来决定其标签。显然，这种机制确实比直接在原始验证数据集上执行KNN实现更快的预测。

基于上述思想，本文提出IK-D-TSK学习算法详见表2。它包含两个过程，训练过程从增强验证数据集中生成具有代表性的中心点及其标签，即数据字典，预测过程使用KNN方法利用数据字典计算出所有训练样本或任何隐含样本的标签。

2.4 理论证明

引理1^[15] 不一致概率 $P(h^1 \neq h^2)$ 是两个子分类器分类误差的严格上界。

当IFCM & KNN作用于增强验证数据集 $\tilde{\mathbf{X}}_v = [\mathbf{X}_v, \tilde{\mathbf{X}}_v]$ 上的任何训练或隐含样本时，它的分类行为如图2所示

显然，IFCM & KNN可以等价地看作是具有两个分类器 h^1, h^2 的多任务分类方法，其中他们都采用相同的KNN方法。根据引理1，通过降低不一致概率 $P(h^1 \neq h^2)$ ，可以提高两个任务场景中每个子分类器的分类性能。为达到此目的，可以采取两种措施。一种是在目标函数中设计一个关于不一致概率的显式惩罚项，如文献[15]中式(3)；另一种是考虑隐式的方法，使分类器 h^1 和 h^2 尽可能一致。IFCM & KNN属于第2种方法。本质上，在IK-D-TSK中，通过以堆栈方式将每个子分类器的输出对原始验证数据集进行扩充来生成增强验证数据集。因此，原始验证数据集中的流形结构通过增强特征相继被打开。根据堆栈泛化原理^[12]，对于任何作用在该增强验证数据集上的分类器，其分类性能肯定会高于在原始验证数据集上的结果。这意味着当要对任何训练或隐含样本进行预测时，与原始验证数

据相比，增强验证数据有更强大的分类潜能。也就是说，作用在增强验证数据集而不是原始验证数据集，具有同样KNN的两个分类器 h^1 和 h^2 ，能够有强大的能力，以达到一致的分类结果。此外，作为

表2 IK-D-TSK学习算法

输入：训练数据集 $\mathbf{D}_{tr} = [\mathbf{X}_{tr} \ \mathbf{Y}_{tr}]$ ，验证数据集 $\mathbf{D}_v = [\mathbf{X}_v \ \mathbf{Y}_v]$ ，其中 \mathbf{X}_{tr} 和 \mathbf{X}_v 分别表示训练数据和验证数据，对应的标签集为 \mathbf{Y}_{tr} 和 \mathbf{Y}_v ，子分类器数 L, K_1, K_2, \dots, K_L 表示每个子分类器的模糊规则数

输出：IK-D-TSK的结构，数据字典

训练过程

(1)初始化：为每个子分类器从 \mathbf{D}_{tr} 中随机抽样训练数据集

$$\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_L, \text{ 并且 } \mathbf{D}_1 \cup \mathbf{D}_2 \cup \dots \cup \mathbf{D}_L = \mathbf{D}_{tr}$$

(2)并行训练 L 个零阶TSK模糊子分类器：

(a)为每个子分类器分配模糊规则数；

(b)构造5个高斯型隶属度函数，在每一维上从中心点集合 $\{0, 0.25, 0.50, 0.75, 1.00\}$ 中随机指定一个值并构造规则组合矩阵 $\Theta_l = [\theta_{ik}^l]_{K_l \times d}$ 。通过给每个元素分配一个随机正数来构造核宽度矩阵 $\Phi_l = [\delta_{ik}^l]_{K_l \times d}$ ，利用式(2)计算模糊隶属度，正则化并构造矩阵

$$\mathbf{X}_g = \begin{bmatrix} \tilde{\omega}_1^1 & \tilde{\omega}_1^2 & \dots & \tilde{\omega}_1^{K_1} \\ \tilde{\omega}_2^1 & \tilde{\omega}_2^2 & \dots & \tilde{\omega}_2^{K_1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\omega}_{N_l}^1 & \tilde{\omega}_{N_l}^2 & \dots & \tilde{\omega}_{N_l}^{K_1} \end{bmatrix}_{N_l \times K_l} \quad (18)$$

通过LLM计算后件参数 \mathbf{a}_g ，即

$$\mathbf{a}_g = ((1/C) \mathbf{I} + \mathbf{X}_g^T \mathbf{X}_g)^{-1} \mathbf{X}_g^T \mathbf{y} \quad (19)$$

其中 \mathbf{I} 是 $K \times K$ 单位矩阵， C 是给定的正则化参数；

(c)通过式(3)生成 L 个子分类器的输出函数 $F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_L(\mathbf{x})$ ；

(3)生成增强验证数据集；

对于验证数据集的每个样本，计算对应每个输出函数 $F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_L(\mathbf{x})$ 的值并将其作为增强特征，将原始特征和增强特征合并，从而形成增强验证数据集 $\mathbf{D}_v^{\text{new}} = [\mathbf{X}_v, \tilde{\mathbf{X}}_v, \mathbf{Y}_v]$ ，其中 $\tilde{\mathbf{X}}_v$ 表示验证数据的增强特征集；

(4)生成数据字典；

在 $\mathbf{D}_v^{\text{new}}$ 上调用IFCM算法后，生成代表性的中心点及其对应的标签，去掉增强特征，即得到数据字典。

预测过程

(1)对于任何测试样本，利用KNN方法在数据字典上找到最近的 k 个点，基于投票策略，确定其类标；

(2)输出测试样本的标签。

$$\begin{aligned}
 J_{\text{IFCM}}^q(\mathbf{U}, \mathbf{V}) &= \sum_{i=1}^{N_q} \sum_{j=1}^{Q_c} (\mu^q(\mathbf{x}_i^q \mathbf{v}_j^q))^2 \|\mathbf{x}_i^q - \mathbf{v}_j^q\|^2 + \\
 &+ \sum_{i=N_{q+1}}^{N_q+(Q-1)c} \sum_{j=1}^{Q_c} (\mu^q(\mathbf{v}_i^q \mathbf{v}_j^q))^2 \|\mathbf{v}_i^q - \mathbf{v}_j^q\|^2 \quad \left. \vphantom{J_{\text{IFCM}}^q} \right\} \text{IFCM\&KNN} \\
 &\quad \downarrow \text{KNN} \\
 J_{\text{IFCM}}^q(\mathbf{U}, \mathbf{V}) &= \underbrace{\sum_{i=1}^{N_q} \sum_{j=1}^{Q_c} (\mu^q(\mathbf{x}_i^q \mathbf{v}_j^q))^2 \|\mathbf{x}_i^q - \mathbf{v}_j^q\|^2 + \sum_{i=1}^{N_q} \sum_{j=1}^{Q_c} (\mu^q(\mathbf{x}_i^q \mathbf{v}_j^q))^2 \|\mathbf{x}_i^q - \bar{\mathbf{v}}_j^q\|^2}_{\text{分类器 } h^1} + \underbrace{\sum_{i=N_{q+1}}^{N_q+(Q-1)c} \sum_{j=1}^{Q_c} (\mu^q(\mathbf{v}_i^q \mathbf{v}_j^q))^2 \|\mathbf{v}_i^q - \mathbf{v}_j^q\|^2 + \sum_{i=N_{q+1}}^{N_q+(Q-1)c} \sum_{j=1}^{Q_c} (\mu^q(\mathbf{v}_i^q \mathbf{v}_j^q))^2 \|\mathbf{v}_i^q - \bar{\mathbf{v}}_j^q\|^2}_{\text{分类器 } h^2}
 \end{aligned}$$

图2 IFCM & KNN

一种非常流行的分类技术，KNN在各种实际应用中已被证明是非常高效的。基于以上分析，同时根据引理1，对IK-D-TSK中采用的IFCM & KNN方法理论，其高效性得以证明。

3 实验与分析

3.1 实验安排

本实验运行的硬件环境为：i7-8700K 3.70GHz CPU and 64 GB RAM, Microsoft Windows 10, Matlab 2016a。在本文的实验中，所有6个数据集都来自于UCI^[16]或KEEL^[17]，详细信息如表3所示。

当IK-D-TSK学习算法在表3的数据集上运行时，实验设置应该考虑其中的所有子分类器之间的多样性，及与数据集的大小、维度和类别数之间的关系。为了做到这一点，如表4所示，本文为每个子分类器设置不同的模糊规则数。根据数据集的大小，为3个数据集安排3个子分类器，其余数据集安排5个子分类器。为了从多样性的角度观察IK-D-TSK

在每个数据集上运行20次所得平均实验结果，对于所有数据集，给每个子分类器分配一个不同的模糊规则数区间，然后从预先指定的规则区间内随机抽取规则数训练每个子分类器。为了方便展示IK-D-TSK的工作原理及可解释性，对于小样本数据集WDB，其上的每个分类器，模糊规则区间都设置为2~15，搜索最优的规则数组组合。对于采用的IFCM每类聚类中心点数，表4给出了搜索范围，即数据字典的大小范围。对于KNN中的最近邻k搜索区间范围设置为{3, 5, 7}。

在本实验中，当IK-D-TSK算法被执行时，每个数据集被随机抽取70%作为训练数据集，20%作为验证数据集，10%作为测试数据集。为了模拟本文中研究的真实场景，在训练数据集中分别加入5%和10%的高斯噪声。在实验中，组织了5个比较分类器进行性能比较。包括经典的0阶(zero-order-TSK-FC)^[1]和1阶TSK模糊分类器(first-order-TSK-FC)^[14]。为了突出展示IK-D-TSK在增强验证集上的分类性能，本文设计了分类器IFCM-KNN-C，即直接在验证集上使用IFCM & KNN和IK-D-TSK进行性能比较。除此之外，本文也将IK-D-TSK和最近流行的深度信念网络(DBN)^[18]和宽度学习系统(BLS)^[9]进行比较。DBN的层数为3，每层的隐节点数根据数据的规模进行设置。BLS的参数用网格搜索寻最优参数， $N_{11} \in [4, 10]$ 间隔为2， $N_2 \in [2, 10]$ 间隔为2， $N_3 \in [60, 200]$ 间隔为20。

3.2 实验结果与分析

图3给出了IK-D-TSK和5个比较分类器在6个数据集上的平均测试精度，图4给出了IK-D-TSK和2个模糊分类器在6个数据集上的平均模糊规则数，表5给出了各分类器运行时间比较结果。

(1)就分类精度而言，与其他分类器相比，IK-D-TSK在平均测试精度方面表现出色。具体而言，IK-D-TSK在6个数据集中分别有5个(5%高斯噪声，图3(a))和5个(10%高斯噪声，图3(b))数据集上具有最高平均测试精度。此外，在所有的数据集上，IK-D-TSK的分类性能比zero-order-TSK-FC和

表3 数据集

数据集	类别数	特征数	样本数
SATimage(SAT)	6	36	6435
MUShroom(MUS)	2	22	8124
WAVeform3(WAV)	3	21	5000
PENBased(PENB)	10	16	10992
WDBc(WDB)	2	14	569
ADUlt(ADU)	2	14	48841

表4 IK-D-TSK参数设置

数据集	分类器	规则数	数据字典
(WDB)	3	2~15	3~4
(WAV)		1.10~120	17~20
(PENB)		2.15~140	10~13
(SAT)	5	3.18~160	10~13
(ADU)		1.5~90	40~45
(MUS)		2.8~120	20~23
		3.10~150	
		4.13~170	
		5.15~190	

IFCM-KNN-C具有绝对优势,不管是5%的高斯噪声还是10%的高斯噪声数据。只有在WDB数据集上,IK-D-TSK稍逊于first-order-TSK-FC和DBN,IK-D-TSK之所以有如此出色地表现,主要原因在于,根据堆栈泛化原理^[19],IK-D-TSK中IFCM所获得的增强特征能够打开增强验证数据集中的流形结构,使得分类变得相对容易。同时,仔细观

察图3(a)和图3(b),可以发现:当高斯噪声从5%增加到10%时,WDB,SAT,MUS和WAV在分类器zero-order-TSK-FC和first-order-TSK-FC上的平均测试精度均迅速下降。然而,作为建立在增强验证数据集上的可解释TSK模糊分类器,IK-D-TSK仍然保持高可靠性。综上,IK-D-TSK在具有高分类精度的同时,还具有很强的鲁棒性。

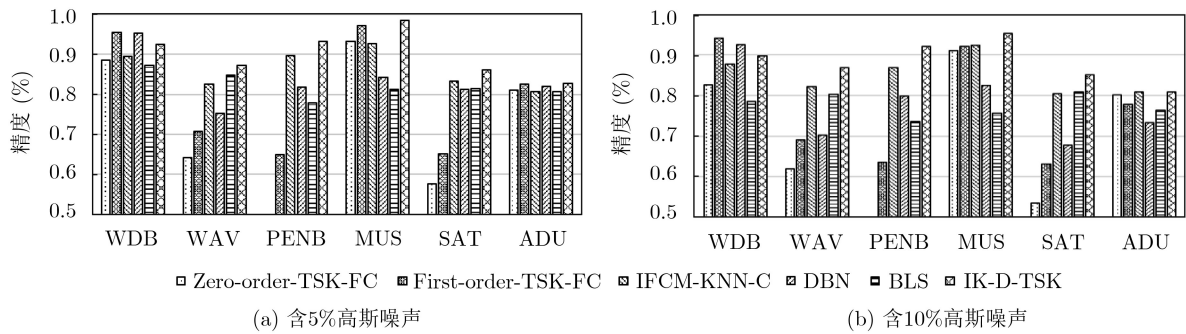


图3 分类器在各数据集上的测试精度

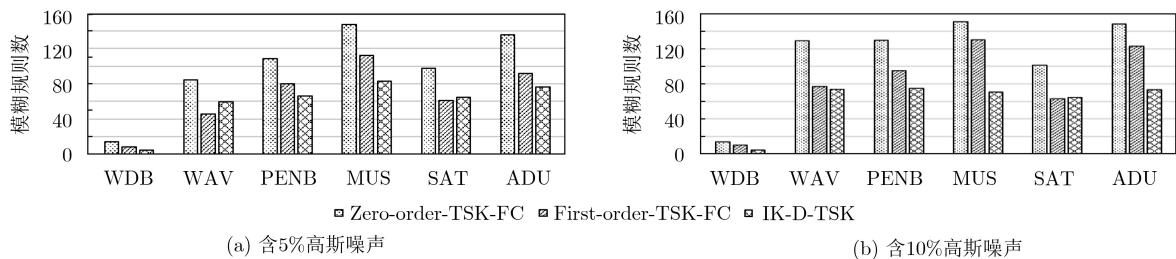


图4 模糊分类器在各数据集上的模糊规则数

表5 各分类器运行时间比较结果(s)

数据集	Zero-order-TSK-FC ^[1]		First-order-TSK-FC ^[14]		IFCM-KNN-C		DBN ^[18]		BLS ^[19]		IK-D-TSK	
	5%噪音	10%噪音	5%噪音	10%噪音	5%噪音	10%噪音	5%噪音	10%噪音	5%噪音	10%噪音	5%噪音	10%噪音
	训练时间	测试时间	训练时间	测试时间	训练时间	测试时间	训练时间	测试时间	训练时间	测试时间	训练时间	测试时间
WDB	0.0216	0.0224	0.0237	0.0243	0.0162	0.0141	4.1844	4.1555	0.0122	0.0122	0.0209	0.0205
	(0.0039)	(0.0057)	(0.0034)	(0.0023)	(0.0019)	(0.0018)	(0.1861)	(0.1592)	(0.0013)	(0.0011)	(0.0032)	(0.0023)
	0.0004	0.0005	0.0004	0.0004	0.0016	0.0016	0.0086	0.0079	0.0102	0.0104	0.0021	0.0020
WAV	0.7982	0.7984	3.8207	4.1065	0.2863	0.2808	35.4047	35.2445	0.0256	0.0261	0.3333	0.3130
	(0.0256)	(0.0346)	(0.0719)	(0.2303)	(0.0222)	(0.0181)	(0.2407)	(0.1511)	(0.0028)	(0.0016)	(0.0366)	(0.0409)
	0.0050	0.0071	0.0059	0.0112	0.0128	0.0129	0.0430	0.0391	0.0155	0.0170	0.0143	0.0142
PENB	0.9656	0.9794	3.7465	3.9261	0.5067	0.4809	15.1945	15.2313	0.0189	0.0191	0.6105	0.5659
	(0.0181)	(0.0320)	(0.1615)	(0.1764)	(0.0225)	(0.0151)	(0.1656)	(0.1790)	(0.0013)	(0.0012)	(0.0372)	(0.0323)
	0.0098	0.0097	0.0196	0.0224	0.0353	0.0311	0.0086	0.0086	0.0124	0.0125	0.0352	0.0340
MUS	0.9496	0.9965	7.6208	8.1693	0.8053	0.8124	47.1515	47.3102	0.0323	0.0321	0.9432	0.9513
	(0.0230)	(0.0377)	(0.2844)	(0.2367)	(0.0629)	(0.0223)	(0.2267)	(0.3248)	(0.0038)	(0.0032)	(0.0415)	(0.0323)
	0.0123	0.0125	0.0283	0.0361	0.0253	0.0231	0.0469	0.0602	0.0189	0.0187	0.0241	0.0244
SAT	1.2282	1.2301	13.3579	14.2199	0.3393	0.3221	338.383	346.9789	0.1491	0.1578	0.4881	0.4528
	(0.0720)	(0.0738)	(0.4825)	(0.6745)	(0.0262)	(0.0134)	(1.2035)	(4.4332)	(0.0052)	(0.0099)	(0.0441)	(0.0383)
	0.0073	0.0062	0.0167	0.0254	0.0183	0.0184	0.2492	0.2039	0.0644	0.0658	0.0209	0.0209
ADU	5.9016	6.0366	15.9947	17.3695	3.1255	3.0311	56.4922	64.3266	0.0337	0.0389	5.9502	5.5299
	(0.1901)	(0.1239)	(0.8757)	(0.8218)	(0.0415)	(0.0215)	(0.3625)	(0.6555)	(0.0028)	(0.0051)	(0.7296)	(0.5056)
	0.0322	0.0370	0.0768	0.1047	0.1126	0.1127	0.0305	0.0656	0.0200	0.0230	0.1549	0.1536

(2)从可解释性方面来观察，虽然在WDB数据集上，IK-D-TSK的分类精度稍逊于first-order-TSK-FC，但是IK-D-TSK所需要的规则数比zero-order-TSK-FC和first-order-TSK-FC都要少。通常，规则数越少，往往意味着较高的可解释性。并且，IFCM-KNN-C不具有可解释性。在5%高斯噪声时，IK-D-TSK在6个数据集中有4个数据集的平均规则数达到最少；当高斯噪声达到10%时，IK-D-TSK具有最少平均规则数的数据集达到了5个。IK-D-TSK需要较少的规则数，原因在于每个子分类器的输出仅用作鉴别信息以形成一个增强特征，而不用作最终分类结果，从而可以不断地打开原始验证数据的流形结构，提高数据可分离性。

(3)在平均训练时间方面，IK-D-TSK比IFCM-KNN-C和BLS要慢，比zero-order-TSK-FC，first-order-TSK-FC和DBN都要快。这是因为zero-order-TSK-FC和first-order-TSK-FC的训练需要网格搜索寻找最优的模糊规则数，而IK-D-TSK中的每

个子分类器规则数都是随机的，每次随机选择相应范围内的模糊规则数，以保证所有子分类器之间具有不同的模糊规则数即可。此外，所有的子分类器都是并行运行的，因此，IFCM-KNN-C的训练时间很大程度上由具有最大模糊规则数的子分类器决定。

接下来，本文将以WDB数据集为例，阐述IK-D-TSK的工作原理。WDB在IK-D-TSK中具有3个分类器，经过第1阶段和第2阶段，所形成的数据字典如表6所示共有两类，每一类3个代表性中心点，分别带有标签。以数据集WDB中一个随机的样本点 $\mathbf{x}^1 = [0.4972, 0.7808, 0.3479, 0.2901, 0.6077, 0.5781, 0.4806, 0.4234, 0.4806, 0.4234, 0.4610, 0.3860, 0.3928, 0.2244, 0.6969, 0.2889]$ 为例，它的真实标签是-1。通过KNN找到离这个样本点最近的3个代表性中心点 $\mathbf{v}_{1,1}, \mathbf{v}_{1,2}, \mathbf{v}_{2,1}$ ，其标签分别为-1, -1, +1，根据投票原则，样本点的预测标签是-1，和这个样本点的真实标签一致。

表 6 WDB数据集在IK-D-TSK上生成的数据字典

$\mathbf{v}_{1,1} = [0.3221, 0.6299, 0.3633, 0.3023, 0.5487, 0.5950, 0.5260, 0.3796, 0.4162, 0.4037, 0.5162, 0.2613, 0.7203, 0.4236, -1]$
$\mathbf{v}_{1,2} = [0.3589, 0.5702, 0.3630, 0.2741, 0.5715, 0.5258, 0.5245, 0.4388, 0.4216, 0.3926, 0.4954, 0.2346, 0.5913, 0.3333, -1]$
$\mathbf{v}_{1,3} = [0.2962, 0.5501, 0.4035, 0.2355, 0.5358, 0.5635, 0.5233, 0.4925, 0.3430, 0.3778, 0.5045, 0.4081, 0.7043, 0.5754, -1]$
$\mathbf{v}_{2,1} = [0.3555, 0.5604, 0.3788, 0.2586, 0.5516, 0.5644, 0.5155, 0.4579, 0.4592, 0.3885, 0.5256, 0.3284, 0.5952, 0.1384, 1]$
$\mathbf{v}_{2,2} = [0.3646, 0.3985, 0.2364, 0.2755, 0.4574, 0.5489, 0.4467, 0.4598, 0.3965, 0.4276, 0.4772, 0.4100, 0.4240, 0.2729, 1]$
$\mathbf{v}_{2,3} = [0.3582, 0.6097, 0.2785, 0.3392, 0.3736, 0.6051, 0.5651, 0.4549, 0.4203, 0.3447, 0.4312, 0.4583, 0.5412, 0.1683, 1]$

为了更好地阐述IK-D-TSK的可解释性，仍然以数据集WDB为例，记录了当IK-D-TSK在一次运行中取得最高精度时所对应的结构。即当数据集WDB在IK-D-TSK上取得最优分类性能时，它的最优规则组合形式是：3-4-6，即第1个分类器，第2个分类器，第3个分类器模糊规则数分别是3,4,6。以第1个分类器为例。

展示其中的一条模糊规则。IK-D-TSK中5个高斯隶属度函数的中心都取自集合 $\{0, 0.25, 0.50, 0.75, 1.00\}$ ，虽然核宽不同，但他们仍然具有各自的语义解释，例如：非常小、小、中、大、非常大，虽然不同的专家根据自己的经验和知识不同可能会给出不同的解释。很明显，IK-D-TSK的规则形式简洁且具有高可解释性。

IK-D-TSK中的模糊规则形式如下

$$\text{IF } x_1 \text{ is exp} \left(-\left(\frac{x_1 - 0.25}{0.6288}\right)^2 \right),$$

$$x_2 \text{ is exp} \left(-\left(\frac{x_2 - 1}{0.6543}\right)^2 \right),$$

$$x_3 \text{ is exp} \left(-\left(\frac{x_3 - 0}{0.6560}\right)^2 \right),$$

$$x_4 \text{ is exp} \left(-\left(\frac{x_4 - 1}{0.0615}\right)^2 \right),$$

$$x_5 \text{ is exp} \left(-\left(\frac{x_5 - 0.5}{0.2584}\right)^2 \right),$$

$$x_6 \text{ is exp} \left(-\left(\frac{x_6 - 0}{0.9008}\right)^2 \right),$$

$$x_7 \text{ is exp} \left(-\left(\frac{x_7 - 0.25}{0.8650}\right)^2 \right),$$

$$x_8 \text{ is exp} \left(-\left(\frac{x_8 - 0.25}{0.8231}\right)^2 \right),$$

$$x_9 \text{ is exp} \left(-\left(\frac{x_9 - 1}{0.3662}\right)^2 \right),$$

$$x_{10} \text{ is exp} \left(-\left(\frac{x_{10} - 1}{0.6899}\right)^2 \right),$$

$$x_{11} \text{ is } \exp\left(-\left(\frac{x_{11}-0.75}{0.1231}\right)^2\right),$$

$$x_{12} \text{ is } \exp\left(-\left(\frac{x_{12}-0.5}{0.0634}\right)^2\right),$$

$$x_{13} \text{ is } \exp\left(-\left(\frac{x_{13}-1}{0.3488}\right)^2\right),$$

$$x_{14} \text{ is } \exp\left(-\left(\frac{x_{14}-0.25}{0.3248}\right)^2\right)$$

THEN $f(\mathbf{x}) = -0.6877$.

4 结束语

针对传统集成模糊分类器可解释性低,鲁棒性差等缺点,本文提出了一种新型基于IFCM & KNN和数据字典的集成TSK模糊分类器IK-D-TSK,首先通过并行学习的方式组织所有0阶TSK模糊子分类器,然后每个子分类器的输出被扩充到原始输入空间,以便以并行方式打开原始输入空间中存在的流形结构,最后通过本文提出的模糊聚类算法IFCM生成数据字典,从而可以利用KNN对测试数据进行快速预测。理论和实验验证了本方法的高效性和高可解释性。

参 考 文 献

- [1] TEH C Y, KERK Y W, TAY K M, *et al.* On modeling of data-driven monotone zero-order TSK fuzzy inference systems using a system identification framework[J]. *IEEE Transactions on Fuzzy Systems*, 2018, 26(6): 3860–3874. doi: [10.1109/TFUZZ.2018.2851258](https://doi.org/10.1109/TFUZZ.2018.2851258).
- [2] PEDRYCZ W and GOMIDE F. *Fuzzy Systems Engineering: Toward Human-Centric Computing*[M]. Hoboken, NJ: Wiley, 2007: 85–101.
- [3] TAKAGI T and SUGENO M. Fuzzy identification of systems and its applications to modeling and control[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1985, SMC-15(1): 116–132. doi: [10.1109/TSMC.1985.6313399](https://doi.org/10.1109/TSMC.1985.6313399).
- [4] TAO Dapeng, CHENG Jun, YU Zhengtao, *et al.* Domain-weighted majority voting for crowdsourcing[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(1): 163–174. doi: [10.1109/TNNLS.2018.2836969](https://doi.org/10.1109/TNNLS.2018.2836969).
- [5] HU Mengqiu, YANG Yang, SHEN Fumin, *et al.* Robust Web image annotation via exploring multi-facet and structural knowledge[J]. *IEEE Transactions on Image Processing*, 2017, 26(10): 4871–4884. doi: [10.1109/TIP.2017.2717185](https://doi.org/10.1109/TIP.2017.2717185).
- [6] ZHANG Yuanpeng, ISHIBUCHI H, and WANG Shitong. Deep Takagi-Sugeno-Kang fuzzy classifier with shared linguistic fuzzy rules[J]. *IEEE Transactions on Fuzzy Systems*, 2018, 26(3): 1535–1549. doi: [10.1109/TFUZZ.2017.2729507](https://doi.org/10.1109/TFUZZ.2017.2729507).
- [7] CORDON O, HERRERA F, and ZWIR I. Linguistic modeling by hierarchical systems of linguistic rules[J]. *IEEE Transactions on Fuzzy Systems*, 2002, 10(1): 2–20. doi: [10.1109/91.983275](https://doi.org/10.1109/91.983275).
- [8] NASCIMENTO D S C, BANDEIRA D R C, CANUTO A M P, *et al.* Investigating the impact of diversity in ensembles of multi-label classifiers[C]. 2018 International Joint Conference on Neural Networks, Rio de Janeiro, Brazil, 2018: 1–8. doi: [10.1109/IJCNN.2018.8489660](https://doi.org/10.1109/IJCNN.2018.8489660).
- [9] BISHOP C M. *Pattern Recognition and Machine Learning*[M]. New York: Springer, 2006: 51–75.
- [10] 王士同, 钟富礼. 最小学习机[J]. 江南大学学报: 自然科学版, 2010, 9(5): 505–510. doi: [10.3969/j.issn.1671-7147.2010.05.001](https://doi.org/10.3969/j.issn.1671-7147.2010.05.001). WANG Shitong and CHUNG K F L. On least learning machine[J]. *Journal of Jiangnan University. Natural Science Edition*, 2010, 9(5): 505–510. doi: [10.3969/j.issn.1671-7147.2010.05.001](https://doi.org/10.3969/j.issn.1671-7147.2010.05.001).
- [11] TUR G, DENG Li and HAKKANI-TÜR D, *et al.* Towards deeper understanding: Deep convex networks for semantic utterance classification[C]. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing, Kyoto, Japan, 2012: 5045–5048. doi: [10.1109/ICASSP.2012.6289054](https://doi.org/10.1109/ICASSP.2012.6289054).
- [12] WOLPERT D H. Stacked generalization[J]. *Neural Networks*, 1992, 5(2): 241–259. doi: [10.1016/s0893-6080\(05\)80023-1](https://doi.org/10.1016/s0893-6080(05)80023-1).
- [13] ZADEH L A. Fuzzy sets[J]. *Information and Control*, 1965, 8(3): 338–353. doi: [10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).
- [14] DENG Zhaohong, JIANG Yizhang, CHUNG F L, *et al.* Knowledge-leverage-based fuzzy system and its modeling[J]. *IEEE Transactions on Fuzzy Systems*, 2013, 21(4): 597–609. doi: [10.1109/TFUZZ.2012.2212444](https://doi.org/10.1109/TFUZZ.2012.2212444).
- [15] GU Xin, CHUNG F L, ISHIBUCHI H, *et al.* Multitask coupled logistic regression and its fast implementation for large multitask datasets[J]. *IEEE Transactions on Cybernetics*, 2015, 45(9): 1953–1966. doi: [10.1109/TCYB.2014.2362771](https://doi.org/10.1109/TCYB.2014.2362771).
- [16] BACHE K and LICHMAN M. UCI machine learning repository[EB/OL]. <http://archive.ics.uci.edu/ml>, 2015.
- [17] ALCALÁ-FDEZ J, FERNÁNDEZ A, LUENGO J, *et al.* KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework[J]. *Journal of Multiple-Valued Logic & Soft Computing*, 2011, 17(2/3): 255–287.
- [18] HINTON G E, OSINDERO S, and TEH Y W. A fast learning algorithm for deep belief nets[J]. *Neural Computation*, 2006, 18(7): 1527–1554. doi: [10.1162/neco.2006.18.7.1527](https://doi.org/10.1162/neco.2006.18.7.1527).
- [19] CHEN C L P and LIU Zhulin. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(1): 10–24. doi: [10.1109/TNNLS.2017.2716952](https://doi.org/10.1109/TNNLS.2017.2716952).

张雄涛: 男, 1984年生, 博士, 研究方向为模式识别、模糊系统。
 蒋云良: 男, 1967年生, 教授, 主要研究方向为智能计算、数据挖掘。
 潘兴广: 男, 1979年生, 博士, 研究方向为神经网络、数据挖掘。
 胡文军: 男, 1977年生, 教授, 主要研究方向为机器学习、模式识别。
 王士同: 男, 1964年生, 教授, 博士生导师, 主要研究方向为模式识别、人工智能。