

一种面向中小规模数据集的模糊分类方法^{*}

周塔^{1,2}, 邓赵红¹, 蒋亦樟¹, 王士同¹

¹(江南大学 数字媒体学院, 江苏 无锡 214122)

²(江苏科技大学 电子信息学院, 江苏 张家港 215600)

通讯作者: 周塔, E-mail: jkdzhout@just.edu.cn



摘要: 虽然 Takagi-Sugeno-Kang(TSK)模糊分类器在一些重要场合已经取得了广泛应用,但如何提高其分类性能和增强其可解释性,仍然是目前的研究热点.提出一种随机划分与组合特征且规则具有高可解释性的深度 TSK 模糊分类器(RCC-DTSK-C),但和其他分类器构造不同的是:(1) RCC-DTSK-C 由很多基训练单元构成,这些基训练单元可以被独立训练;(2) 每一个基训练单元的隐含层通过模糊规则的可解释性来表达,而这些模糊规则又是通过随机划分、随机组合来进行特征选择的;(3) 基于栈式结构理论,源数据集作为相同的输入空间被映射到每一个独立的基训练单元中,这样就有效地保证了源数据的所有特征在每一个独立的训练单元中都得以保留.实验结果表明, RCC-DTSK-C 具有良好的分类性能和可解释性.

关键词: Takagi-Sugeno-Kang(TSK);随机模糊划分;特征组合;可解释性;深度学习;栈式结构

中图法分类号: TP18

中文引用格式: 周塔,邓赵红,蒋亦樟,王士同.一种面向中小规模数据集的模糊分类方法.软件学报,2019,30(12):3637-3650.
<http://www.jos.org.cn/1000-9825/5590.htm>

英文引用格式: Zhou T, Deng ZH, Jiang YZ, Wang ST. Fuzzy classification method for small- and medium-scale datasets. Ruan Jian Xue Bao/Journal of Software, 2019, 30(12):3637-3650 (in Chinese). <http://www.jos.org.cn/1000-9825/5590.htm>

Fuzzy Classification Method for Small- and Medium-scale Datasets

ZHOU Ta^{1,2}, DENG Zhao-Hong¹, JIANG Yi-Zhang¹, WANG Shi-Tong¹

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

²(School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang 215600, China)

Abstract: Although Takagi-Sugeno-Kang (TSK) is widely used in practically every profession, how to enhance its classification accuracy and interpretability is still a research focus. In this study, a deep TSK fuzzy classifier is proposed. This classifier (i.e., RCC-DTSK-C) can randomly select features and combine features and own triply concise interpretability for fuzzy rules. There are several other varieties of RCC-DTSK-C such as reasonable structure for rule representation, namely, (1) the proposed RCC-DTSK-C consists of many base-training units and each base-training unit can be trained independently. According to the principle of stacked generalization, the input of the next base-training unit consists of the training set and random result obtained from random projections about prediction results of current base-training unit. (2) In RCC-DTSK-C, the hidden layer of each base-training unit is represented by triply concise interpretable fuzzy rules which are in the sense of randomly selected features. These features are selected by dividing into the not-fixed several fuzzy partitions and randomly combining rules and keeping the same input space in every base-training unit. (3) The source data set is mapped into each of the independent base-training units as the same input space, which effectively ensures that all the features of the source data are preserved in each separate training unit. The extensive experimental results show RCC-DTSK-C can achieve the enhanced classification performance and triply concise interpretability for fuzzy rules.

* 基金项目: 国家自然科学基金(61772239, 61702225, 61572236, 61711540041)

Foundation item: National Natural Science Foundation of China (61772239, 61702225, 61572236, 61711540041)

收稿时间: 2017-09-17; 修改时间: 2018-04-16; 采用时间: 2018-04-23; jos 在线出版时间: 2019-01-21

CNKI 网络优先出版: 2019-01-22 13:49:03, <http://kns.cnki.net/kcms/detail/11.2560.TP.20190122.1348.011.html>

Key words: Takagi-Sugeno-Kang (TSK); random partition; feature combination; interpretability; deep learning; stacked structure

模糊系统作为智能计算领域的一个及其重要的研究分支,由于其自身较强的可解释性和学习能力而被广泛地应用到多个领域^[1,2]. Takagi-Sugeuo-Kang(TSK)模糊系统由于其输出的简洁性,其训练过程通常可以转化为二次规划问题或者线性回归问题进行求解,这就使得 TSK 模型比其他模型训练过程更迅捷、更高效.同时,因其较好的逼近性能而被应用到多个领域,比如模式识别、图像处理和数据挖掘等^[3,4].模糊系统是根据模糊集和模糊推理理论而形成的智能系统,它主要是把自然界的模糊语言向模糊规则进行转换.正是凭借这一特性,模糊系统在日常生活中仍被广泛使用.目前,已经建立或者识别的 TSK 模糊分类器大致可分为以下几种:基于遗传算法的 TSK 模糊分类器、基于神经-模糊混合的 TSK 模糊分类器和类似层次状的 TSK 模糊分类器.对于遗传算法的 TSK 模糊分类器而言,它主要是模仿人类的进化过程去构造结构和识别参数.典型工作包括基于 TSK 模糊分类器的多目标遗传算法,该算法将规则选择问题转化为多目标的组合优化问题.对于神经-模糊混合的 TSK 模糊分类器,它主要是把人工神经网络和模糊系统进行组合.典型的工作包括基于 TSK 模糊分类器的 BP 神经网络,这类工作主要是通过 BP 神经网络和 BP 算法混合而成,用来训练分类器和 SVM 的相关参数.这类的工作可以参考文献[1,5-8].而对于类似层次状的 TSK 模糊分类器,主要包括具有层次模糊系统和具有全局逼近性能的模糊系统.关于这类算法及其改进算法可以参考文献[9-15].这 3 类 TSK 模糊分类器都不可避免地遇到这样的挑战:(1) 当输入的样本维数过高时,会出现维数灾难;(2) 在 TSK 模糊系统中,不恰当的模糊划分在某种程度上也会影响模糊规则意义的表达;(3) 训练完模糊系统后的修整技术在某种程度上会去除一些不合理的模糊划分,但是这种后续处理方法势必也影响了分类的精度.

我们所提出的模糊分类器 RCC-DTSK-C 类似于层次状的 TSK 模糊分类器,但是有本质的区别,它能有效地避免 TSK 模糊分类器所面临的巨大挑战.这一点在文后有详细的报道.

深度学习理论已经成为当前研究的热点,它在很多领域都取得了成功^[16-27].我们知道,深度结构能够快速而有效地捕获源数据中的细节,而恰好这些细节有时候却能更好地表达可解释性.

本文利用栈式结构^[28]构造深度 TSK 模糊分类器来学习模糊规则,以提高分类性能.深度结构能够估算 TSK 模糊分类器的预测误差,栈式结构可以帮助我们解决困难的非凸优化问题,而这些问题也正是深度学习要解决的.

本文基于以下几种考虑构造了可解释性很强的 TSK 模糊分类器.

- (1) 模糊划分数不确定,完全随机生成,比如随机生成 3 个模糊划分,对应高斯隶属函数中心点为[0,0.5,1],其语义表示为{差,中等,好};再如随机生成 5 个模糊划分,对应高斯隶属函数中心点为[0,0.25,0.5,0.75,1],其语义表示为{很差,差,中等,好,非常好};
- (2) 随机选取源数据集中的部分或者大部分特征数据;
- (3) 每个基训练单元中的模糊分类器拥有相同的输入空间;
- (4) 由于 0 阶 TSK 模糊分类器的输出结果是常数,对于系统易于分析和表达,本文将以 0 阶 TSK 模糊系统为基础训练模型,探讨随机模糊划分和规则组合的深度 0 阶 TSK 模糊系统的建模方法.

1 Takagi-Sugeno-Kang 模糊系统

1.1 经典 TSK 模糊系统

根据文献[1],经典模糊系统主要有 Takagi-Sugeno-Kang 模糊系统(TSK-FS)、Mamdani-Larsen 模糊系统(ML-FS)^[1]和广义模糊系统(GFM)^[1]这 3 种.对于经典 TSK-FS 而言,其模糊规则表示为

$$\left. \begin{array}{l} \text{If } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \dots \wedge x_d \text{ is } A_d^k \\ \text{Then } y^k = f^k(\mathbf{x}) = p_0^k + p_1^k x_1 + \dots + p_d^k x_d, k = 1, 2, \dots, K \end{array} \right\} \quad (1)$$

这里, $\mathbf{x}=[x_1, x_2, \dots, x_d]^T$ 为输入向量, A_i^k 表示第 i 个输入变量 x_i 所对应的第 k 条规则描述的模糊子集, K 代表模糊规

则数.每条规则都与 \mathbf{x} 相对应,同时把输入空间的模糊子集 $A^k \in R^d$ 映射到输出空间的模糊子集 $f^k(\mathbf{x})$,其中, A^k 表示为 $A^k = (A_1^k, A_2^k, \dots, A_d^k)^T$, $u^k(\mathbf{x})$ 则为模糊子集 $f^k(\mathbf{x})$ 相对应的隶属函数.那么第 k 条规则隶属函数 $u^k(\mathbf{x})$ 表示为

$$u^k(\mathbf{x}) = u_1^k(x_1) \wedge u_2^k(x_2) \wedge \dots \wedge u_d^k(x_d) \quad (2)$$

在反模糊化之前,若采用重心反模糊化操作,则最终的输出 $f(\mathbf{x})$ 即可描述为

$$f(\mathbf{x}) = \frac{\sum_{k=1}^K u^k(\mathbf{x}) f^k(\mathbf{x})}{\sum_{k'=1}^K u^{k'}(\mathbf{x})} = \frac{\sum_{k=1}^K \prod_{i=1}^d u_i^k(\mathbf{x}) f^k(\mathbf{x})}{\sum_{k'=1}^K \prod_{i=1}^d u_i^{k'}(\mathbf{x})} \quad (3)$$

经典模糊系统模型可以分为 TSK 模糊系统、ML 模糊系统和 GFM 模糊系统.对于 ML 模糊系统,其模糊规则表示为

$$\left. \begin{array}{l} \text{If } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \dots \wedge x_d \text{ is } A_d^k \\ \text{Then } y^k = B_k(b_k, v_k), k = 1, 2, \dots, K \end{array} \right\} \quad (4)$$

对于 GFM 模糊系统,其模糊规则表示为

$$\left. \begin{array}{l} \text{If } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \dots \wedge x_d \text{ is } A_d^k \\ \text{Then } y^k = B_k(f^k(x), v_k), k = 1, 2, \dots, K \end{array} \right\} \quad (5)$$

其中, $B^k(\cdot)$ 表示 ML 模型中第 k 条模糊规则对应的模糊集, b_k 和 v_k 分别称为质心和模糊系数.

1.2 0阶和1阶TSK模糊分类器

根据文献[2,29],0阶 TSK 型模糊系统描述的模糊推理规则表达为

$$\left. \begin{array}{l} \text{If } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \dots \wedge x_d \text{ is } A_d^k \\ \text{Then } y^k = f^k(\mathbf{x}) = p_0^k, k = 1, 2, \dots, K \end{array} \right\} \quad (6)$$

然后,经过相应的操作和去模糊化处理,即可得到 0 阶 TSK 模糊系统输出^[2]表达式为

$$y^0 = \frac{\sum_{k=1}^K u^k(\mathbf{x}) f^k(\mathbf{x})}{\sum_{k'=1}^K u^{k'}(\mathbf{x})} = \frac{\sum_{k=1}^K \tilde{u}^k(\mathbf{x}) f^k(\mathbf{x})}{\sum_{k=1}^K \tilde{u}^k(\mathbf{x})} = \sum_{k=1}^K \tilde{u}^k(\mathbf{x}) p_0^k \quad (7)$$

其中,

$$u^k(\mathbf{x}) = \prod_{i=1}^d u_{A_i^k}^k(x_i) \quad (8)$$

$$\tilde{u}^k(\mathbf{x}) = u^k(\mathbf{x}) / \sum_{k=1}^K u^k(\mathbf{x}) \quad (9)$$

如果隶属函数采用高斯函数,公式(8)中的 $u_{A_i^k}^k(x_i)$ 亦可表示成^[2]:

$$u_{A_i^k}^k(x_i) = \exp\left(\frac{-(x_i - c_i^k)^2}{2\delta_i^k}\right) \quad (10)$$

这里,参数 c 和 δ 可以通过一些聚类算法取得.这里以采用 FCM 算法^[1]为例, c 和 δ 可以估计为

$$c_i^k = \frac{\sum_{j=1}^N u_{jk} x_{ji}}{\sum_{j=1}^N u_{jk}} \quad (11)$$

$$\delta_i^k = h \frac{\sum_{j=1}^N u_{jk} (x_{ji} - c_i^k)^2}{\sum_{j=1}^N u_{jk}} \quad (12)$$

在公式(11)和公式(12)中, u_{jk} 表示为 $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jd})^T$ 隶属于第 k 类的隶属程度.这里, h 为尺度参数,该尺度参数可以人为地调节.

根据文献[30],有:

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

即 0 阶 TSK 的输出 y^0 则可表示为

$$y^0 = \mathbf{p}_g^T \mathbf{x}_g \tag{13}$$

以上描述可参见文献[2].由此可知,模糊规则假设的参数确定后,模糊规则的参数学习问题可转化为线性回归问题^[1,31]来求解.这就使得在系统建模时,用线性回归模型替代 TSK 来进行处理.

通常情况下,0 阶 TSK 模糊分类器的分类性能比 1 阶 TSK 差.但是 1 阶 TSK 模糊分类器很难对每条模糊规则下的 $(d+1)$ 参数给出清晰的解释.当每条规则下仅有 1 个参数 p_0^k ,那么 $p_0^k / \max_k(p_0^k)$ 的正负值就很容易解释为属于或者不属于第 k 类的度值^[30].本文提出深度结构的目的是提高分类性能,且保证 RCC-DTSK-C 具有高可解释性.因此,本文使用 0 阶 TSK 模糊分类器作为基训练单元构造深度 TSK 模糊分类器 RCC-DTSK-C.

2 深度 0 阶 TSK 模糊分类器

这里,我们提出一个栈式结构的深度模糊学习模型 RCC-DTSK-C.该模型利用 0-阶 TSK 模糊分类器作为一个基训练模块.

2.1 基训练单元

为了方便解释基训练单元的实现机制,此处以单输出 0 阶 TSK 模糊分类器为例(如图 1 所示).

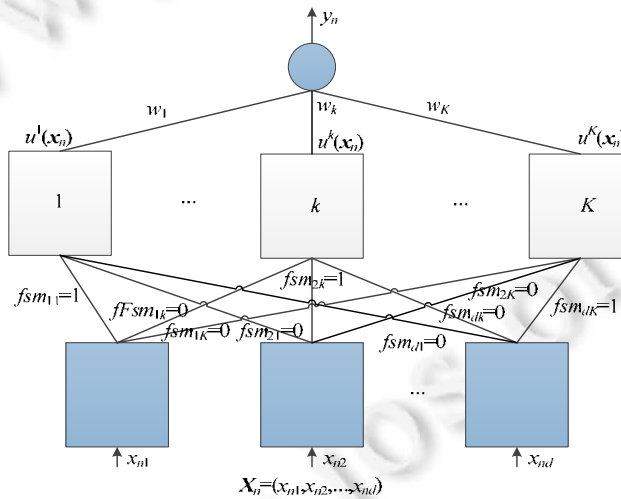


Fig.1 Base training unit corresponding to single output 0-order TSK fuzzy classifier

图 1 单输出 0 阶 TSK 模糊分类器对应的基训练单元

我们的工作过程分为以下几步.

- (1) 直接采用 P 个高斯隶属函数,分别为 F_1, F_2, \dots, F_P ,中心为 $[0/(P-1), 1/(P-1), \dots, (P-1)/(P-1)]$.例如,如果 $P=3$,那么隶属函数表示为 F_1, F_2, F_3 ,中心为 $[0, 0.5, 1]$,它们的语义解释为:差、中等、好.
- (2) 随机生成特征选择矩阵 FSM ,其每一元素值通过随机分配 0,1 进行赋值.即 $FSM[fsm_{ik}]_{d \times K}$:当 $fsm_{ik}=0$ 时,表示第 i 维属性未被选中;否则,表示已被选中.其中, $i=1, 2, \dots, d, k=1, 2, \dots, K$.
- (3) 随机生成规则组合矩阵 RRC (random rule-combination matrix),其元素值由随机生成的 0,1 二值构成. $RRC[3,1,4]=1$ 表示第 4 条规则的第 3 个输入特征采用“非常差”的高斯函数,即第 3 维属性的 F_1 被选中,它将被用在第 4 规则中.

通过以上分析,在这种 0 阶 TSK 模糊分类器中所有规则可改写为(以 5 维输入空间的第 k 条规则为例)

$$\left. \begin{array}{l} \text{If } x_1 \text{ is bad with } fsm_{1k} \text{ and } x_2 \text{ is high with } fsm_{2k} \text{ and } x_3 \text{ is "Can be ignored" with } fsm_{3k} \\ \text{and } x_4 \text{ is medium with } fsm_{4k} \text{ and } x_5 \text{ is very high with } fsm_{5k} \\ \text{Then } y = p_0^k \text{ for Class 2} \end{array} \right\} \quad (14)$$

其中,“Can be ignored”表示当前这一维特征丢弃(未被选中).

2.2 栈式结构

关于本文为什么要使用栈式结构,后文有详细的解释,此处只介绍栈式结构的构成. \mathbf{X} 作为第 j 个基训练单元的输出,即 $\mathbf{X}_j = \mathbf{X}$.当运行第 j 个基训练单元后,它就生成预测结果.根据栈式原理,RCC-DTSK-C 将预测结果随机投影到源训练集 \mathbf{X}_j 中.最终,第 $j+1$ 个 0 阶 TSK 模糊分类器的输出由源训练集 \mathbf{X}_j 与第 j 个基训练单元后的随机投影(如图 2 所示),即

$$\mathbf{X}_{j+1} = \mathbf{X} + \gamma \times \mathbf{Y}_j \times \mathbf{R}_j \quad (15)$$

其中, \mathbf{Y}_j 是第 j 个基训练单元的输出; \mathbf{R}_j 是一个随机投影矩阵,其元素值由 0 或者 1 随机赋值; γ 是一个很小的常数.

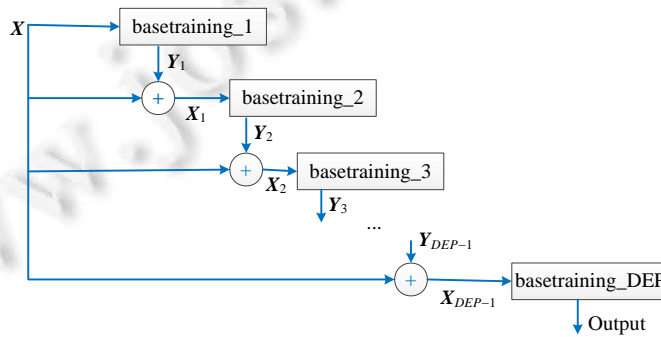


Fig.2 Structure of RCC-DTSK-C with single output

图 2 单输出的 RCC-DTSK-C 结构

2.3 深度 TSK 模糊分类器结构

近年来报道了很多关于 TSK 模糊模型及其应用工作^[32-34],这些工作大都构造了一些不同的模糊分类器去验证分类性能和获取高的可解释性.基于栈式结构的思想,图 1 提出了一个基训练单位模型.该模型由单一输出的 0 阶 TSK 模糊分类器构成.图 2 构造了一个新的深度 TSK 模糊分类器.为了保证可解释性和满足 Kuncheva 在文献[35]中的陈述:如果隶属函数的整个选择是不一致的或者隶属函数的形状是不规则的,那么它们也不太可能与语言准确地关联.与其他报道算法不同的是,我们所提的方法通过确定隶属函数的中心很好地解释了隶属函数不规则和语言标签歧义的问题.我们随机引用了 P 个高斯型的隶属函数 F_1, F_2, \dots, F_P , 比如 $P=3$, 它们的标签标记为 F_1 :低, F_2 :中等, F_3 :高.它们分别中心化为 $[0, 0.5, 1]$.我们认为,规则的可解释性是设置分类器时值得考虑的一个非常重要的部分.

通常情况下,我们考虑有 N 对数据构成的训练集 (x_n, t_n) , 其中, x_n 是特征向量, $x_n \in \mathbf{R}^d$; t_n 是相应输入特征向量的标签.为了方便,我们定义一个矩阵 \mathbf{X} 作为训练集, T 训练样本的类标签.设置 \mathbf{X}_1 作为原始输入的第 1 层, 即 $\mathbf{X}_1 = \mathbf{X}$.

我们这里提出的深度 0 阶 TSK 模糊分类器 RCC-DTSK-C 是深度学习的又一次尝试.该分类器构建机制类似于多层的极限学习机 ELM, 但是有本质的不同.在 RCC-DTSK-C 中, 把模糊规则映射到多层 TSK 的每一个隐含层中.

与随机构建 TSK 的策略类似, RCC-DTSK-C 随机分配了高斯函数的标准差、随机规则组合矩阵 RRC 以及特征选择矩阵 FSM . RCC-DTSK-C 利用由 0 或 1 二进制构成的 RRC 决定在某一个规则中某一个输入属性的哪个隶属函数被使用. RCC-DTSK-C 提出的最大的优点是允许隶属函数不连续.比如, 基于 RRC , 对于某个规则的

输入属性,隶属函数 1 或者隶属函数 4 被选中,这就意味着 RCC-DTSK-C 仅仅考虑 F_1 和 F_4 ,而隶属函数则不予考虑.然而这与文献[36]中提出的方法有区别,文献[36]方法仅限于使用连续的隶属函数.那么对于上面例子而言,文献[36]中提到的方法显然不能使用,它应该修改为:对于该条规则的输入属性而言,所有隶属函数都应该被考虑.换句话说,在 F_1, F_2, \dots, F_P 之间的每一个隶属函数都应该被考虑.当然,这就不可避免地计算了很多无关紧要的隶属函数值,最终肯定会影响分类的精度.此外,根据后续迭代层次的数目划分了输入空间,我们使用上述设置标签的方法(F_1 :特别特别差, F_2 :特别差, \dots, F_P :特别特别好)对输入属性进行划分,于是对每一输入属性就得到了 P 个划分.这里我们还设置了“Can be ignored”,利用“Can be ignored”除去部分属性,这样,模糊 if-then 规则就可以被重新定义.正如上面提到的每一个计算出的模糊函数值代表了一个有效的模糊输入属性,这个模糊输入属性被翻译成非常有用的规则.举例如下:

If x_1 is bad OR good And x_2 is Medium And x_3 is “Can be ignored” x_4 is very good then Class 2.

该条规则就可以重新写成:

If x_1 is bad OR good And x_2 is Medium And x_4 is very good, then Class 2.

对于 RCC-DTSK-C,随机输入规则数(隐含层节点数),输出 y 的值通过一个方程快速的计算得到.

图 1 展示了典型的 3 层 TSK 结构.与图 1 类似,图 2 展示了 RCC-DTSK-C 的基本思想.这里值得注意的是,我们所提的方法可以应用于多输入和多输出系统,具体细节将在后面章节详细讨论.

对于一个含有 N 个数据的训练集(含有标签)可用向量 (x_n, t_n) 表示, $(x_n, t_n) \in R^n \times R^c$, 其中, x_n 表示特征数据, t_n 表示标签.

2.4 RCC-DTSK-C 训练算法描述

根据第 2.3 节的描述,本节给出 RCC-DTSK-C 训练算法,具体训练步骤如下.

输入和输出:

输入:训练集 $X=[x_1, x_2, \dots, x_N]^T$, 标签 $T=[t_1, t_2, \dots, t_N]^T$, $x_i \in R^n, t_i \in R^c$;

输出:预测函数以及每个基训练单元的模糊规则.

初始化:

随机选择模糊规则数 L ; 随机选择深度 DEP ; 模糊划分数 P ;

随机生成 P 个高斯函数的核宽 $\phi_p (\phi_p \in R^+, p=1, 2, \dots, P)$

$X_1=X$

训练过程:

for $dep=1$ to DEP **do**

1: 初始化第 dep 层模糊规则数 K_{dep}

2: 模糊化成 P 个隶属函数 F_1, F_2, \dots, F_P , 其中心确定为 $cn=[0/(P-1), 1/(P-1), \dots, (P-1)/(P-1)]$

3: 生成 RRC_{dep} 和 FSM_{dep}

4: 对所有训练数据 $(x_n, t_n) \in X_{dep} \times T, n=1, 2, \dots, N$, 执行如下步骤:

5: 对所有的输入属性 x_{ni} , 计算高斯隶属函数值 $u(p, x_{ni}) = \exp(-(x_{ni} - cn_p)^2 / \phi^2 / 2)$, 其中, $n=1, 2, \dots, N, i=1, 2, \dots, d$, $cn_p \in \{0/(P-1), 1/(P-1), \dots, (P-1)/(P-1)\}, p=1, 2, \dots, P$

6: 在一个规则中, 计算输入属性的值 val :

➤ 当 $fsm_{ie}=0$ 时, $val_{ie}(x_i^n) = 1 - \prod_{p=1}^P (1 - RRC_{dep}(i, p, e)u(p, x_{ni}))$;

➤ 当 $fsm_{ie}=1$ 时, $val_{ie}(x_i^n) = 1$

其中, $n=1, 2, \dots, N, i=1, 2, \dots, d, p=1, 2, \dots, P, e=1, 2, \dots, K_{dep}$

7: 根据步骤 6 中的 val 计算 wei_{ne} : $wei_{ne} = \prod_{i=1}^d val_{ie}(x_i^n)$, 其中, $wei_{ne} \in R^{d \times K_{dep}}$

$$8: \text{构造规则层输出矩阵 } \mathbf{O}_{dep} = \begin{bmatrix} wei_{11} & \cdots & wei_{1K_{dep}} \\ \vdots & \ddots & \vdots \\ wei_{N1} & \cdots & wei_{NK_{dep}} \end{bmatrix}_{N \times K_{dep}}$$

9: 计算输出权重 $\mathbf{W}_{dep} = \mathbf{O}_{dep}^T \left(\frac{1}{\varepsilon} \mathbf{I} + \mathbf{O}_{dep} \mathbf{O}_{dep}^T \right)^{-1} \mathbf{T}$ [37], 其中, ε 为参数, \mathbf{T} 为训练数据集中的标签, $\mathbf{I}_{K_{dep} \times K_{dep}}$ 是一个单位矩阵

10: 计算整个输出矩阵 $\mathbf{Y}_{dep} = \mathbf{O}_{dep} \mathbf{W}_{dep}$

11: 生成随机投影矩阵 $\mathbf{R}_{dep} \in R^{m \times d}$

12: 根据第 2.2 节栈式结构的描述计算 \mathbf{X}_{dep+1}

13: $dep = dep + 1$

End for

预测输出:

输出预测函数 $\mathbf{Y}_{DEP} = \mathbf{O}_{DEP} \mathbf{W}_{DEP}$

for $dep = 1$ to DEP **do**

根据 RRC_{dep} 和 FSM_{dep} , 按照式(14)输出 K_{dep} 模糊规则

End for

• 算法分析 1

算法 1 中提出的规则组合矩阵 RRC 、特征选择矩阵 FSM 以及生成 P 个高斯函数核宽 ϕ , 它们元素的值都是随机生成, 很显然, 所得到的预测结果一般来讲不是很精确, 但是该结果要比随机猜想的值要好一些. 我们观察算法 1 为什么可以从定性的角度获得增强的分类性能. 在步骤 12 中, 我们可以清楚地看到, 算法的每一个基训练单元都在相同的数据空间上运行, 而原来的训练集落在这个数据空间中. 因此, 参与各基训练单元的模糊规则的每个特征都具有相同的物理解释. 更重要的是, 每个基训练单元运行在原来的训练集加上随机投影, 即 $\gamma \mathbf{Y}_{dep} \mathbf{R}_{dep}$. 因此, 即使这些附近的数据具有相同的标签, 由于 \mathbf{R}_{dep} 的随机性, 相同的输出 \mathbf{Y}_{dep} 也可能引发不同的转移到原始训练集中. 我们说, 对于 $RCC-DTSK$, 更好的线性可分性最终可以预测, 算法 1 的确可以提高分类性能.

• 算法分析 2

对于算法中表达式 $\mathbf{W}_{dep} = \mathbf{O}_{dep}^T \left(\frac{1}{\varepsilon} \mathbf{I} + \mathbf{O}_{dep} \mathbf{O}_{dep}^T \right)^{-1} \mathbf{T}$ 和 $\mathbf{X}_{j+1} = \mathbf{X} + \gamma \times \mathbf{Y}_j \times \mathbf{R}_j$, ε 和 γ 是及其重要的两个参数. 根据文献[37], 参数 ε 的取值应该稍大些. 在本实验中, 我们取 $\varepsilon = 500$. 而对于参数 γ , 其取值应该相对较小, 本实验中, 我们取 $\gamma = 0.05$, $DEP = 3$ 或者 4. 关于其他一些参数在下一节讨论.

• 算法分析 3

下面分析算法的时间复杂度. 我们首先分析第 dep 个基训练单元的时间复杂度. 根据算法, 第 dep 个基训练单元的时间复杂度主要包括随机生成规则组合矩阵 RRC 的时间复杂度、生成特征选择矩阵 FSM 的时间复杂度、计算规则层输出矩阵 \mathbf{O} 的时间复杂度、计算输出权重 \mathbf{W} 的时间复杂度. 对于 RRC_{dep} , 其时间复杂度是 $O(PdK_{dep})$. 对于 FSM_{dep} , 其时间复杂度是 $O(dK_{dep})$. \mathbf{O}_{dep} 的时间复杂度可由步骤 6~步骤 8 得到, 其时间复杂度是 $O(PNd^2K_{dep})$. 很明显, 步骤 9 中 \mathbf{W}_{dep} 的时间复杂度是 $O(N^3 + NK_{dep} + Nm)$, 步骤 10 的时间复杂度是 $O(NmK_{dep})$, 步骤 11 和步骤 12 的时间复杂度是 $O(Nd + Nmd)$. 所以, 由于 m 非常小, 那么第 dep 个基训练单元的时间复杂度为

$$O(PdK_{dep} + dK_{dep} + PNd^2K_{dep} + N^3 + NK_{dep} + Nm + NmK_{dep} + Nd + Nmd) \approx O(PNd^2K_{dep} + N^3) \quad (16)$$

其中, d 是特征数, K_{dep} 是模糊规则数. 训练模型的深度是 DEP , 所以整个 $RCC-DTSK-C$ 时间复杂度粗略地表示为

$O\left(\sum_{dep=1}^{DEP} (PNd^2K_{dep} + N^3)\right)$. 在每个基训练单元里, 模糊规则数 K 相对来说是比较小的, 如果样本数 N 不是太大, 那么该时间复杂度还是可以接受的.

3 实验研究

3.1 实验数据集

为了进一步表现 RCC-DTSK-C 的分类性能,我们采用了如表 1 所罗列的 6 个数据集^[38],这些数据集可以从 <https://archive.ics.uci.edu/ml/datasets.html> 下载,其中包括小/中等/大样本数据集,大样本数据集 *Airline* 可以从 <http://stat-computing.org/dataexpo/2009/> 下载.我们还采用了 0 阶 TSK 模糊分类器、1 阶 TSK 模糊分类器以及 KEEL 软件工具箱中的两个进化模糊分类器(FURIA & C4.5)在这些数据集中进行了对比.KEEL(基于进化学习的知识提取)是一种免费软件(GPLv3)Java 套件,它允许用户评估不同类型的数据挖掘问题的进化学习和基于软件计算的的行为:回归、分类、聚类、模式挖掘等等.KEEL 软件工具箱可从 <http://www.keel.es/download.php> 下载.受论文版面的限制,关于数据集更多的细节可以参考各自的网页.对于数据集 *Airline*,1987 年 10 月~2008 年 4 月,航空公司的数据集包括与美国的所有商业航班的航班到达和离港详细信息.这是一个大型数据集,共有近 1.2 亿条记录,占用了 1.6 千兆字节的压缩空间和 12 千兆字节^[39].我们挑选了 1990 年~1993 年这 4 年的数据.在我们的实验中,所有数据集都被归一化.我们将每个数据集样本的 75% 数据用于训练,剩余部分用于测试.我们使用分类精度和训练/测试时间作为性能指标来评估所有比较分类器的性能,其中,分类精度定义为正确分类的样本数与总样本数的比率.所有实验都在具有 64GB 内存的 E5-2609 v2 2.5GHZ CPU(2 个处理器)的计算机上进行.

Table 1 Datasets

表 1 数据集

数据集	训练样本数	测试样本数	特征数	类别数
Balloons (BAL)	57	19	5	2
Climate-Model-Simulation-Crashes (CLI)	405	135	21	2
Airline (AIR)	300 000	100 000	29	2
Balance-Scale (BAS)	469	156	5	3
Abalone (ABA)	3 133	1 044	9	3
Yeast (YEA)	1 113	371	9	10

3.2 本文方法与其他对比方法性能比较

我们知道,虽然有很多不同的分类器被开发出来,比如 BP 神经网络和支持向量机,但是我们这里采用常见的 0 阶和 1 阶 TSK 模糊分类器^[1,2]作为比较的方法,因为分类的准确性和可解释性可以同时从它们观察得到.而其他分类器,比如支持向量机 SVM 和 BP 网络就像黑盒子.RCC-DTSK-C 与其他非模糊深度分类器相比有如下优势.

- 训练计算量小:大多数非模糊深度分类器训练通常需要很多次迭代,无疑会增加训练的计算量;而 RCC-DTSK-C 在训练过程中无需迭代,极大地提高了训练效率.
- 无需大量训练样本:大多数非模糊深度分类器在很大程度上要求大量的训练样本,而 RCC-DTSK-C 在训练过程中只需要随机挑选部分样本数据即可.
- 训练结果具有强的可解释性:大多数非模糊深度分类器通常输出结果难以解释,而 RCC-DTSK-C 的输出具有强的可解释性.

3.2.1 分类器参数设置

下面我们列出这几种分类器各自的参数设置.因为 0 阶 TSK 和 1 阶 TSK 模糊分类器都用到模糊聚类方法(fuzzy c-means,简称 FCM)和 SVM,所以先介绍 FCM 和 SVM 的参数设置.SVM 的正则化参数设置通过网络搜索从 0.01 到 100,步长是 0.1,FCM 中的聚类数和模糊规则数相等,尺度参数 r 值的个数可以设置网络搜索从 0.01 到 100,步长是 0.1.对于分类器 FURIA 和 C4.5,它们的参数则采用 KEEL 软件工具箱中的默认值.对于 RCC-DTSK-C,层数 DEP 的取值为 2 或 3,数据集 *Balloons* 每一层的模糊规则数搜索范围设置为 2~4,步长为 1;数据集 *Climate-Model-Simulation-Crashes* 每一层的模糊规则数搜索范围设置为 3~5,步长为 1;数据集 *Airline* 每一层的

模糊规则数搜索范围设置为 150~400,步长为 50;数据集 Balance-Scale 每一层的模糊规则数搜索范围设置为 5 到 30,步长为 1;数据集 Abalone 每一层的模糊规则数搜索范围设置为 10~25,步长为 1;数据集 Yeast 每一层的模糊规则数搜索范围设置为 5~15,步长为 1.

3.2.2 分类性能比较

由于对输入特征和模糊隶属函数都是随机选择的,那么对于一个数据集而言,RCC-DTSK-C 的结构就有多种组合.对每个数据集,我们稍微改变每层规则数,并同时运行 10 次,取平均值,得到了平均模糊规则数、平均训练精度/平均测试精度、平均训练时间/平均测试时间.最后,我们也列出了所有数据集的平均模糊规则数、平均训练精度/平均测试精度.结果见表 2、表 3.

Table 2 Average number of fuzzy rules and average classification accuracies (%)

表 2 平均模糊规则数和平均分类精度 (%)

数据集		BAL	CLI	AIR	BAS	ABA	YEA	Mean
0 阶 TSK 模糊分类器	规则数	2.50	5.00	—	15.25	20.50	16.00	11.85
	训练	43.86(1.24)	91.60(0.69)	—	56.08(6.63)	55.54(0.75)	52.85(2.50)	59.99(1.89)
	测试	52.63(3.72)	91.11(2.09)	—	50.64(4.08)	52.96(0.25)	48.64(2.50)	59.20(2.53)
1 阶 TSK 模糊分类器	规则数	2.50	4.50	—	12.00	15.50	12.25	9.35
	训练	49.12(7.44)	98.27(0)	—	91.47(0.90)	56.43(0.23)	64.43(1.21)	71.94(1.96)
	测试	68.42(0)	92.59(1.57)	—	88.46(3.17)	54.11(2.72)	56.52(3.66)	72.02(2.22)
FURIA	规则数	3.00	7.50	—	16.50	17.50	16.50	12.2
	训练	74.36(0.04)	99.07(0)	—	88.95(0)	56.21(0)	64.01(0.03)	76.52(0.01)
	测试	68.42(0.05)	91.85(0.02)	—	83.36(0.01)	54.59(0.02)	58.89(0.04)	71.42(0.03)
C4.5	规则数	—	—	—	—	—	—	—
	训练	77.63(0.01)	99.07(0)	—	89.91(0.01)	56.77(0.01)	79.31(0.01)	80.54(0.01)
	测试	61.84(0.01)	91.66(0.02)	—	78.40(0.01)	56.01(0.01)	55.86(0.02)	68.75(0.01)
RCC-DTSK-C	规则数	2.25	3.75	300.00	8.75	21.50	11.25	9.50
	训练	80.63(2.15)	99.20(0.41)	65.61(1.13)	91.98(1.09)	57.81(0.69)	74.58(2.61)	80.84(1.39)
	测试	77.50(1.26)	92.83(1.28)	61.78(2.11)	90.73(2.00)	56.32(2.36)	61.52(1.91)	75.78(1.76)

Table 3 Average training time and test time

表 3 平均训练时间和测试时间

数据集		BAL	CLI	AIR	BAS	ABA	YEA	Mean
0 阶 TSK 模糊分类器	训练	0.05(0.01)	1.09(0.09)	—	0.16(0)	5.27(0.58)	2.41(0.50)	1.80(0.24)
	测试	0	0.02	—	0.03	0.57	0.07	0.14
1 阶 TSK 模糊分类器	训练	0.06(0.01)	6.56(0.57)	—	0.14(0)	36.13(0.60)	8.70(0.35)	10.32(0.31)
	测试	0	0.08	—	0	4.56	0.45	1.02
FURIA	训练	—	—	—	—	—	—	—
	测试	—	—	—	—	—	—	—
C4.5	训练	—	—	—	—	—	—	—
	测试	—	—	—	—	—	—	—
RCC-DTSK-C	训练	0.05(0.01)	2.52(0.17)	2.09e+04(66.59)	0.43(0.01)	17.81(1.29)	6.95(0.52)	5.55(0.4)
	测试	0	0.04	4.78e+03	0.01	2.92	0.44	0.77

根据表 2,我们发现,RCC-DTSK-C 几乎优于所有其他几个分类器,且取得了最好的平均分类训练精度,分别是 80.63%,99.20%,91.98%和 57.81%.对于 AIR 大样本数据集,0 阶 TSK、1 阶 TSK、FURIA 和 C4.5 均运行相当缓慢,表 2 中用“—”标记;而 RCC-DTSK-FC 也能正常运行,这也直接说明了 RCC-DTSK-C 适用于大样本数据.表 2 中,RCC-DTSK-C 的测试精度也明显高于其他分类器,这就说明 RCC-DTSK-C 具有良好的泛化性能.根据表 3,由于 FURIA 和 C4.5 是基于 JAVA 平台的软件系统,时间对比没有意义,在表 3 中用“—”标记,这里用 RCC-DTSK-C 与 0 阶 TSK 和 1 阶 TSK 模糊分类器进行了对比.我们发现,RCC-DTSK-C 运行慢于 0 阶 TSK 模糊分类器,但是明显快于 1 阶 TSK 模糊分类器.

接下来,我们研究 RCC-DTSK-C 随着层数变化而引起性能的变化.表 4 列出了每个数据集运行 10 次得到的平均训练精度和平均测试精度.通过实验发现,RCC-DTSK-FC 在大多数情况下可以通过多层结构达到令人满

意的分类性能,层数的搜索范围为 2~3.因此本文认为,RCC-DTSK-C 的深度为 2 或 3.因为在这个范围内,RCC-DTSK-C 的分类精度非常接近或者高于对比分类器的分类精度.根据分析,层数的选择对分类精度有影响.然而如何在每个数据集上确定 RCC-DTSK-C 的合适层数,是将来一个有趣的研究课题.

Table 4 Training accuracies and test accuracies of RCC-DTSK-C for different layers (%)

表 4 RCC-DTSK-C 对于不同层的训练精度和测试精度 (%)

数据集	第1层		第2层		第3层		第4层	
	训练	测试	训练	测试	训练	测试	训练	测试
BAL	74.23(1.23)	69.30(1.17)	75.27(2.03)	72.15(1.91)	80.92(1.82)	77.95(1.28)	-	-
CLI	99.17(0.02)	91.63(1.02)	99.24(0.26)	92.98(1.25)	-	-	-	-
AIR	63.55(1.27)	60.85(0.42)	65.91(0.92)	62.42(1.02)	65.98(1.01)	62.35(2.52)	-	-
BAS	87.95(0.98)	79.62(0.62)	89.85(0.47)	77.95(1.52)	-	-	-	-
ABA	56.61(0.98)	50.93(0.53)	58.91(0.29)	58.18(1.08)	-	-	-	-
YEA	76.44(0.33)	75.27(0.59)	76.85(0.47)	71.95(1.52)	76.99(1.08)	74.86(2.52)	-	-

3.2.3 可解释性

为了更好地描述 RCC-DTSK-C 的可解释性,我们记录了当 RCC-DTSK-C 在每个数据集取得最好的精度时对应的结构.RCC-DTSK-C 规则结构的表示形式为“第 1 层模糊规则数-第 2 层模糊规则数-...-第 DEP 层模糊规则数”.表 4 描述了各个数据集对应最好的精度.从表 4 可以看出,RCC-DTSK-C 获取最好的精度时,其层数为 2 或 3;RCC-DTSK-C 获得最好的精度时,对应的结构分别为 4-3-2;5-3-2;350-100-50;20-10-2;25-5-2;15-4-2.比如,4-3-2 意味着 RCC-DTSK-FC 有 3 层结构:第 1 层、第 2 层和第 3 层的模糊规则数分别是 4,3,2.

限于文章篇幅,我们这里以数据集 BAL 为例进一步展示 RCC-DTSK-C 的可解释性.由于 RCC-DTSK-C 的可解释性与 RCC-DTSK-C 的相应结构和模糊规则有关,在前面的实验中,RCC-DTSK-C 在数据集 BAL 运行的最好精度是 80.92%,其对应的结构是 4-3-2.为了方便观察模糊规则的可解释性,我们取 5 个模糊划分数,且在 RCC-DTSK-C 获得的所有模糊规则中提取了前 4 个规则,然后在表 5 中对这些规则进行了总结.

Table 5 Rule presentation

表 5 规则展示

规则 <i>i</i>	IF					THEN output <i>y</i> (class label)
	特征1	特征2	特征3	特征4	特征5	is
规则1	low	medium	very high	high	Can be ignored	0.098(+1)
规则2	very low	low	Medium	Can be ignored	Can be ignored	-0.442(-1)
规则3	medium	very high	High	very low	Can be ignored	0.706(+1)
规则4	Can be ignored	very high	very low	high	medium	-0.529(-1)

表 5 中,“Can be ignored”表示在相应的模糊规则中没有选择相应的特征.比如,我们可以方便地把表 5 中的规则 1 表示为:

```
IF      feature_1 is low
AND    feature_2 is medium
AND    feature_3 is very high
AND    feature_4 is high
AND    feature_5 is Can be ignored
```

THEN $y = 0.098(+1)$

其中,表 5 中的+1,-1 分别表示数据集 BAL 中对应的“inflated T”和“inflated F”.

很明显,这种模糊规则具有很高的可解释性.

为了对 RCC-DTSK-C 的可解释性进行更深入的研究,表 6 列出了模糊划分、特征选择矩阵、规则组合矩阵、前 4 个模糊规则的学习参数以及每一层模糊规则“THEN-part”的输出.对于模糊划分(模糊隶属函数),不同的专家有自己的建议和理解.换句话说,对于模糊规则,其解释可能因不同的专家而异,从而只能提供模糊规则的底层解释.例如,第 1 个模糊规则中的模糊划分可以解释为语言意义由气球决策专家决定.因此,对于 DEP=3,

我们获得了相应的语言规则,见表 6.

Table 6 Four rules presentation for BAL dataset
表 6 对于数据集 BAL 的 4 条规则展示

Fuzzy rule k:
 If x_1 is $A_1^k \wedge x_2$ is $A_2^k \wedge \dots \wedge x_d$ is A_d^k then $y^k = f^k(\mathbf{X}) = p_0^k + p_1^k x_1 + \dots + p_d^k x_d, k = 1, 2, \dots, K$

规则 序号	dep=1		dep=2		dep=3	
	$P_{dep=3, cn}=[0,0.5,1]$ $\phi=[0.2156,0.7759]$		$P_{dep=5, cn}=[0,0.25,0.5,0.75,1]$ $\phi=[0.7781,0.5621,0.5610,0.9378]$		$P_{dep=3, cn}=[0,0.5,1]$ $\phi=[0.0981,0.2983,1]$	
	if-part parameters	then-part parameters	if-part parameters	then-part parameters	if-part parameters	then-part parameters
1	$FSM_{dep(:,1)}=[1,0,1,0,0]$ $RRC_{dep(:,1)}$ $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	$[-0.1029]$	$FSM_{dep(:,1)}=[0,0,1,0,1]$ $RRC_{dep(:,1)}$ $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$	$[-0.7812, 0.2638]$ $wei(\mathbf{X}) = e^{-\frac{1}{2} \left(\frac{y_1(x)-0.25}{0.3172} \right)^2}$	$FSM_{dep(:,1)}=[1,0,1,0,0]$ $RRC_{dep(:,1)}$ $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$[0.3539, -0.2019, 4852]$ $wei(\mathbf{X}) = e^{-\frac{1}{2} \left(\frac{y_1(x)-1}{0.2983} \right)^2 + \left(\frac{y_2(x)-0}{1} \right)^2}$
2	$FSM_{dep(:,2)}=[1,1,0,1,0]$ $RRC_{dep(:,2)}$ $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$	$[-0.3351]$	$FSM_{dep(:,2)}=[1,0,1,1,1]$ $RRC_{dep(:,2)}$ $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$	$[0.6668, -0.2497]$ $wei(\mathbf{X}) = e^{-\frac{1}{2} \left(\frac{y_1(x)-0.5}{0.4439} \right)^2}$	$FSM_{dep(:,2)}=[0,1,1,1,0]$ $RRC_{dep(:,2)}$ $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$[-0.7953, 0.0058, 0.7013]$ $wei(\mathbf{X}) = e^{-\frac{1}{2} \left(\frac{y_1(x)-0.75}{1} \right)^2 + \left(\frac{y_2(x)-0.25}{0.0981} \right)^2}$
3	$FSM_{dep(:,3)}=[1,1,1,0,0]$ $RRC_{dep(:,3)}$ $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$	$[-0.7068]$	$FSM_{dep(:,3)}=[0,0,1,1,1]$ $RRC_{dep(:,3)}$ $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$	$[0.6533, -0.0201]$ $wei(\mathbf{X}) = e^{-\frac{1}{2} \left(\frac{y_1(x)-0}{0.7823} \right)^2}$	$FSM_{dep(:,3)}=[0,1,1,1,1]$ $RRC_{dep(:,3)}$ $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	$[-0.6652, 0.0478, 0.0680]$ $wei(\mathbf{X}) = e^{-\frac{1}{2} \left(\frac{y_1(x)-0.75}{0.2983} \right)^2 + \left(\frac{y_2(x)-0}{0.0981} \right)^2}$
4	$FSM_{dep(:,4)}=[1,1,1,1,0]$ $RRC_{dep(:,4)}$ $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$	$[-0.9523]$	$FSM_{dep(:,4)}=[1,0,1,0,1]$ $RRC_{dep(:,4)}$ $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$	$[-0.4841, -0.6692]$ $wei(\mathbf{X}) = e^{-\frac{1}{2} \left(\frac{y_1(x)-0.75}{0.8864} \right)^2}$	$FSM_{dep(:,4)}=[1,1,0,1,1]$ $RRC_{dep(:,4)}$ $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$[0.8792, 0.6530, 0.9008]$ $wei(\mathbf{X}) = e^{-\frac{1}{2} \left(\frac{y_1(x)-0}{0.0981} \right)^2 + \left(\frac{y_2(x)-0.5}{1} \right)^2}$
output	$y_1(x)=-0.4450$		$y_2(x)=-0.0891$		$y_3(x)=-0.7455$	

3.2.4 非参数统计分析

Milton Friedman^[40]开发了非参数统计测试,即 Friedman 等级测试,用于检测多个测试中的差异.这里,我们对表 1 中列出的所有数据集进行 Friedman 排名测试.Friedman 排名测试用于评估在这些数据集的多个比较方法中是否存在差异.图 3 显示了 Friedman 排名测试中这 5 个分类器对所有数据集的排名结果.从图 3 可以看出,RCC-DTSK-C 在这些分类器中保持最好的排名.Friedman 测试得到的相应 p 值为 0,这表明在所有的分类器中确实存在显著的差异.总之,这些结果也表明 RCC-DTSK-C 明显优于其他几种分类器.

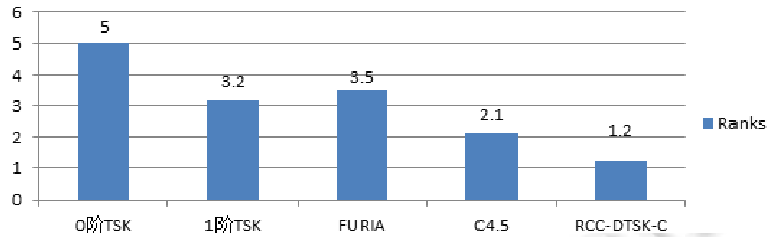


Fig.3 Nonparametric statistical analysis

图3 非参数统计分析

4 结 论

本文通过栈式结构原理,以提高分类性能和较强的可解释性为目的,提出一种深度 TSK 模糊分类器 RCC-DTSK-C. RCC-DTSK-C 以栈式方式构建,提出随机选取特征,不固定模糊划分和随机规则组合,生成每一个 basetraining 中的模糊规则.在 RCC-DTSK-C 的第 1 层和其他隐含层中始终保持相同的数据空间,使得每个隐含层的每个特征仍然保持与输入层相同的物理意义.我们对所有数据集的实证结果表明,RCC-DTSK-C 在分类性能上明显优于其他几种分类器.更重要的是,通过对数据集 BAL 的进一步研究发现,RCC-DTSK-C 还具有较强的可解释性.

References:

- [1] Deng ZH, Choi KS, Chung FL, *et al.* Scalable TSK fuzzy modeling for very large datasets using minimal-enclosing-ball approximation. *IEEE Trans. on Fuzzy Systems*, 2011,19(2):210–226.
- [2] Jiang YZ, Deng ZH, Wang ST. 0 order L_2 norm Takagi Sugeno Kang type transfer learning fuzzy system. *Acta Electronica Sinica*, 2013,41(5):897–904 (in Chinese with English abstract).
- [3] Jiang YZ, Chung FL, Ishibuchi H, *et al.* Multitask TSK fuzzy system modeling by mining intertask common hidden structure. *IEEE Trans. on Cybernetics*, 2015,45(3):548–561.
- [4] Fadali S, Jafarzadeh S. TSK observers for discrete type-1 and type-2 fuzzy systems. *IEEE Trans. on Fuzzy Systems*, 2014,22(2):451–458.
- [5] Jiang YZ, Chung FL, Ishibuchi H, Deng ZH, Wang ST. Multitask TSK fuzzy system modeling by mining intertask common hidden structure. *IEEE Trans. on Cybernetics*, 2015,45(3):534–547.
- [6] Mitrakis NE, Theocharis JB. A diversity-driven structure learning algorithm for building hierarchical neuro-fuzzy classifiers. *Information Sciences*, 2012,186(1):40–58.
- [7] Chen Y, Wang DZ, Tong SC. Forecasting studies by designing Mamdani interval type-2 fuzzy logic systems: With the combination of BP algorithms and KM algorithms. *Neurocomputing*, 2016,174(22):1133–1146.
- [8] Zhou SM, Gan JQ. Constructing L_2 -SVM-based fuzzy classifiers in high-dimensional space with automatic model selection and fuzzy rule ranking. *IEEE Trans. on Fuzzy Systems*, 2007,15(3):398–409.
- [9] Chung FL, Duan JC. On multistage fuzzy neural network modeling. *IEEE Trans. on Fuzzy Systems*, 2000,8(2):25–142.
- [10] Mantas CJ, Puche JM. Artificial neural networks are zero-order TSK fuzzy systems. *IEEE Trans. on Fuzzy Systems*, 2008,16(3):630–643.
- [11] Alcalá-Fdez J, Alcalá R, González S, Nojima Y, García S. Evolutionary fuzzy rule-based methods for monotonic classification. *IEEE Trans. on Fuzzy Systems*, 2017,25(6):1376–1390.
- [12] Li T, Wang ST. Zero-order TSK fuzzy classifier based on LLM for large-scale data sets. *Control and Decision*, 2017,32(1):21–30 (in Chinese with English abstract).
- [13] Ye SZ, Zhang B, Wu MR, Zheng WB. A fuzzy classifier based on the constructive covering approach in neural networks. *Ruan Jian Xue Bao/Journal of Software*, 2003,14(3):429–434 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/429.htm>

- [14] Xu ML, Wang ST. Extracting fuzzy rules from the maximum ball containing the homogeneous data. *Journal of Electronics & Information Technology*, 2017,39(5):1130–1135 (in Chinese with English abstract).
- [15] Zhai JH, Wang XZ, Zhang SF. Ensemble incomplete wavelet packet subspaces for face recognition based on fuzzy integral. *Pattern Recognition and Artificial Intelligence*, 2014,27(9):794–801 (in Chinese with English abstract).
- [16] Wang ST, Jiang YZ, Chung FL, Qian PJ. Feedforward kernel neural networks, generalized least learning machine, and its deep learning with application to image classification. *Applied Soft Computing*, 2015,37:125–141.
- [17] Bell IE, Baranowski GVG. Reducing the dimensionality of plant spectral databases. *IEEE Trans. on Geoscience and Remote Sensing*, 2004,42(3):570–576.
- [18] Papa JP, Rosa GH, Pereira DR, Yang XS. Quaternion-based deep belief networks fine-tuning. *Applied Soft Computing*, 2017,60:328–335.
- [19] Li F, Zhang J, Shang C, Huang DX, Oko E, Wang MH. Modelling of a post-combustion CO₂ capture process using deep belief network. *Applied Thermal Engineering*, 2018,130:997–1003.
- [20] Chen CLP, Zhang CY, Chen L, Gan M. Fuzzy restricted Boltzmann machine for the enhancement of deep learning. *IEEE Trans. on Fuzzy Systems*, 2015,23(6):2163–2173.
- [21] Long MS, Wang JM, Cao Y, Sun JG, Yu PS. Deep learning of transferable representation for scalable domain adaptation. *IEEE Trans. on Knowledge and Data Engineering*, 2016,28(8):2027–2040.
- [22] Wu ZD, Wang YN, Zhang JW. Fouling and damaged fingerprint recognition based on deep learning. *Journal of Electronics & Information Technology*, 2017,39(7):1585–1591 (in Chinese with English abstract). [doi: 10.11999/JEIT161121]
- [23] Liu Q, Zhai JW, Zhong S, Zhang ZC, Zhou Q, Zhang P. A deep recurrent q-network based on visual attention mechanism. *Journal of Computers*, 2017,40(6):1353–1366 (in Chinese with English abstract).
- [24] Gao JY, Yang XS, Zhang TZ, Xu CS. Robust visual tracking method via deep learning. *Chinese Journal of Computers*, 2016,39(7):1419–1434 (in Chinese with English abstract).
- [25] Li HY, Bi DY, Yang Y, Zha YF. Research on visual tracking algorithm based on deep feature expression and learning. *Journal of Electronics & Information Technology*, 2015,37(9):2033–2039 (in Chinese with English abstract).
- [26] He YY, Li BQ. A combination form learning rate scheduling for deep learning model. *Acta Electronica Sinica*, 2016,42(6):953–958 (in Chinese with English abstract).
- [27] Sun R, Zhang GH, Gao J. Pedestrian recognition method based on depth hierarchical feature representation. *Journal of Electronics & Information Technology*, 2016,38(6):1528–1535 (in Chinese with English abstract).
- [28] Anifowose F, Labadin J, Abdulraheem A. Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines. *Applied Soft Computing*, 2015,26:483–496.
- [29] Juang CF, Chen TC, Cheng WY. Speedup of implementing fuzzy neural networks with high-dimensional inputs through parallel processing on graphic processing units. *IEEE Trans. on Fuzzy Systems*, 2011,19(4):717–728.
- [30] Yuan YF, Zhuang HJ. A genetic algorithm for generating fuzzy classification rules. *Fuzzy Sets and Systems*, 1996,84:1–19.
- [31] Leski J. TSK-Fuzzy modeling based on ϵ -insensitive learning. *IEEE Trans. on Fuzzy Systems*, 2005,13(2):181–193.
- [32] Lin YY, Chang JY, Lin CT. A TSK-type-based self-evolving compensatory interval type-2 fuzzy neural network and its applications. *IEEE Trans. on Industrial Electronics*, 2014,6(1):447–459.
- [33] Deng ZH, Cao LB, Jiang YZ, Wang ST. Minimax probability TSK fuzzy system classifier: A more transparent and highly interpretable classification model. *IEEE Trans. on Fuzzy Systems*, 2015,23(4):813–826.
- [34] Zheng YJ, Ling HF, Chen SY, Xue JY. A hybrid neuro-fuzzy network based on differential biogeography-based optimization for online population classification in earthquakes. *IEEE Trans. on Fuzzy Systems*, 2014,23(4):1070–1083.
- [35] Guenounou O, Dahhou B, Chabour F. TSK fuzzy model with minimal parameters. *Applied Soft Computing*, 2015,30(2):748–757.
- [36] Bortolan G, Brohet C, Fusaro S. Possibilities of using neural networks for ECG classification. *Journal of Electrocardiology*, 1995,29:10–16.
- [37] Valdes JJ. Extreme learning machines with heterogeneous data types. *Neurocomputing*, 2018,280:38–52.
- [38] 1987. <https://archive.ics.uci.edu/ml/datasets.html>
- [39] 2009. <http://stat-computing.org/dataexpo/2009/>

- [40] Li T, Li J, Liu ZL, Li P, Jia CF. Differentially private Naïve Bayes learning over multiple data sources. Information Sciences, 2018, 444:89–104.

附中中文参考文献:

- [2] 蒋亦樟,邓赵红,王士同.0阶 L_2 型 TSK 迁移学习模糊系统.电子学报,2013,41(5):897–904.
- [12] 李滔,王士同.适合大规模数据集且基于 LLM 的 0 阶 TSK 模糊分类器.控制与决策,2017,32(1):21–30.
- [13] 叶少珍,张钺,吴鸣锐,郑文波.一种基于神经网络覆盖构造法的模糊分类器.软件学报,2003,14(3):429–434. <http://www.jos.org.cn/1000-9825/14/429.htm>
- [14] 徐明亮,王士同.由最大同类提取模糊分类规则.电子与信息学报,2017,39(5):1130–1135.
- [15] 翟俊海,王熙照,张素芳.基于模糊积分的不完全小波包子空间集成人脸识别.模式识别与人工智能,2014,27(9):794–801.
- [22] 吴震东,王雅妮,章坚武.基于深度学习的污损指纹识别研究.电子与信息学报,2017,39(7):1585–1591. [doi: 10.11999/JEIT161121]
- [23] 刘全,翟建伟,钟珊,章宗长,周倩,章鹏.一种基于视觉注意力机制的深度循环 Q 网络模型.计算机学报,2017,40(6):1353–1366.
- [24] 高君宇,杨小汕,张天柱,徐常胜.基于深度学习的鲁棒性视觉跟踪方法.计算机学报,2016,39(7):1419–1434.
- [25] 李寰宇,毕笃彦,杨源,查宇飞.基于深度特征表达与学习的视觉跟踪算法研究.电子与信息学报,2015,37(9):2033–2039.
- [26] 贺昱曜,李宝奇.一种组合型的深度学习模型学习率策略.自动化学报,2016,42(6):953–958.
- [27] 孙锐,张广海,高隼.基于深度分层特征表示的行人识别方法.电子与信息学报,2016,38(6):1528–1535.



周塔(1980—),男,江苏盐城人,博士,讲师,主要研究领域为模式识别,数据挖掘.



蒋亦樟(1988—),男,博士,副教授,CCF 高级会员,主要研究领域为模式识别,系统建模.



邓赵红(1981—),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为模式识别,数据挖掘.



王士同(1964—),男,教授,博士生导师,CCF 专业会员,主要研究领域为模式识别,人工智能.