ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 37,No. 5,Sep. 2022,pp. 935－951
DOI:10. 16337/j. 1004-9037. 2022. 05. 001

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

可解释的深度TSK模糊系统综述

王士同，谢润山，周尔昊

（江南大学人工智能与计算机学院，无锡 214000）

摘　要：深度神经网络在多个领域取得了突破性的成功，然而这些深度模型大多高度不透明。而在很多高风险领域，如医疗、金融和交通等，对模型的安全性、无偏性和透明度有着非常高的要求。因此，在实际中如何创建可解释的人工智能（Explainable artificial intelligence，XAI）已经成为了当前的研究热点。作为探索XAI的一个有力途径，模糊人工智能因其语义可解释性受到了越来越多的关注。其中将高可解释的Takagi-Sugeno-Kang（TSK）模糊系统和深度模型相结合，不仅可以避免单个TSK模糊系统遭受规则爆炸的影响，也可以在保持可解释性的前提下取得令人满意的测试泛化性能。本文以基于栈式泛化原理的可解释的深度TSK模糊系统为研究对象，分析其代表模型，总结其实际应用场景，最后剖析其所面临的挑战与机遇。

关键词：可解释的人工智能；模糊人工智能；TSK模糊系统；可解释性；深度结构；栈式泛化原理

中图分类号：TP18　　**文献标志码：**A

Survey of Interpretable Deep TSK Fuzzy Systems

WANG Shitong，XIE Runshan，ZHOU Erhao

（School of Artificial Intelligence and Computer Science，Jiangnan University，Wuxi 214000，China）

Abstract：While the existing deep neural networks have earned great successes in various application scenarios，they are still facing black-box challenges that they are not very suitable for some application fields such as healthcare，finance and transportation. Therefore，explainable artificial intelligence（XAI）has been becoming a hot research topic in recent years. Among the existing XAI means，since fuzzy AI systems have the impressive ability to achieve an excellent trade-off between performance and interpretability，interpretable deep Takagi-Sugeno-Kang（TSK）fuzzy systems have been drawing more and more attentions. We first state the concept of the classical TSK fuzzy systems，then give a comprehensive overview of interpretable deep TSK fuzzy systems which are based on stacked generalization principle，including their structures，representative models and application scenarios，and finally discuss their future development direction according to their existing problems.

Key words: explainable artificial intelligence (XAI); fuzzy AI; TSK fuzzy systems; interpretability; deep structures; stacked generalization principle

---

基金项目：国家自然科学基金（61972181）；江苏省自然科学基金（BK20191331）；江苏省研究生科研与实践创新计划项目（KY-CX22-2315）。

收稿日期：2021-08-12；**修订日期：**2022-09-01

引 言

近年来,深度神经网络在多个领域取得了实质性的突破并得到了重要应用,特别是在计算机视觉[1-2]、自然语言处理[3]和医学图像识别[4]方面都取得了巨大的成功。在深度神经网络中,最为主流的结构是卷积神经网络[5],生成式对抗神经网络[6]和残差神经网络[7]。由于传统模型追求的主要目标是使模型与输入数据相匹配,也就是模型要能有效地学习输入数据的潜在分布规律,并能对未知数据做出精准的预测。因此,在选择解决方案时,精确度通常是最重要的性能指标,而深度神经网络拥有从输入数据中准确学习的出色能力,深度神经网络的研究热潮也正好满足了这种高精度要求。然而,这些高精度的深度神经网络大多是高度不透明的,也就是说,人们无法弄清楚是输入数据中的什么信息使它们得出了最终的预测[8-9],因此这些模型也被称为黑箱模型。具体来说,深度神经网络有着若干层完全连接的神经元,第1层致力于从输入数据中提取较简单的、低级的特征,然后在后续层中组合成更复杂的、高级的、因而更有代表性的特征。尽管深度神经网络精准的建模能力令人印象深刻,但这种优点往往与较高的系统复杂性有关。因为神经网络的高复杂度,人们往往无法理解深度神经网络是如何工作的,也无法解读它们为什么会产生某种最终决策。事实上,现有的深度神经网络的有效性可能会因为模型无法向人类用户解释它的决策而在某些应用领域受到限制,从而可能导致不安全和不正确的决策。例如,在美国司法系统中用于累犯风险预测的COMPAS模型就是典型的黑箱模型,它的作用是预测某人在出狱/入狱后的一定时间内是否会被逮捕。由于人们无法弄清COMPAS模型是如何得出最后的决策,曾经发生过人们被错误地拒绝假释的案例[10]。不仅如此,COMPAS模型还被犯罪学家指责有种族偏见[11]。

在当今社会,人们对模型的安全性、无偏性和透明性有着越来越高的需求,模型需要在必要时对其最终决策进行解释[12]。例如,在由政府机构监管的领域,如医疗、金融和交通等,往往要求最终决策有着高度的透明度。也就是说,在这些领域,必须证明模型没有使用或产生任何偏见,即无偏性。在其他领域,例如高风险投资(如投资组合再平衡),或关键任务的应用(如电厂设定点选择),最终决策必须由董事会所批准或由电厂的经营者所接受,然后他们对这些最终决策负责。因此,模型不仅需要追求高的精度,还要针对其他辅助标准进行性能优化,如安全性、无偏性、透明性、隐私性和稳定性等。然而这些辅助标准中的大多数往往不能完全量化,但是如果这个模型是可解释的,即是可解释的人工智能(Explainable AI, XAI),它就能解释自己的推理过程和最终结果,人们就可以验证该推理和最终结果在这些辅助标准方面是否合理和可采纳。

在机器学习背景下,XAI模型的可解释性被定义为模型向人类解释或以可理解的术语呈现其最终决策的能力[13]。也就是说,通过模型的可解释性,人类可以知道模型是如何得出其最终决策的[14]。相反,通过阐明模型内部程序或内部表示往往不能有效地提高模型的可解释性[15]。近年来,一些方法被提出用以尝试解释深度神经网络,例如显著性图[16],它能够可视化输入图像上每个元素对于最终决策的重要程度,从而确定输入图像的哪一部分对深度神经网络的最终决策有最大的影响。然而,显著性图只能找出影响最大的这部分输入图像,却无法告诉用户神经网络利用这部分输入图像做了什么。特别是,多个类别的显著性图可能基本相同,此时显著性图将无法解释输入图像的同一部分为何对于不同类别却有着相同的最终决策[10]。就像文献[10]中所阐述的:“创建第2个(事后)模型来解释第1个黑盒子模型,这本身是有问题的,解释往往是不可靠的,而且可能是误导性的。……因为解释不可能完美地忠实于原始模型。……这就导致了一个危险,即任何对黑箱模型的解释方法都可能是原始模型在部分特征空间中的不准确表示。”因此,相比于构建新的模型去尝试解释黑盒模型,直接构建本质上可解释的模型是更好的选择。这样一来,模型会提供自己的忠实于模型实际最终决策的解释。

众所周知,基于模糊理论的模糊系统[17-18]能够模仿人类的知识推理能力,将复杂的模糊问题清晰化,所生成的模糊规则是可以被人类读懂和理解的IF-THEN语句,因此具有天然的可解释性。实际上,模糊系统存在多种不同的可解释性,而本文重点研究的是语义可解释性。具体来说,语义可解释性使用特定的程度语义值来描述输入特征,如{很低、低、中等、高、非常高}就是一组常用的语义值,这样所得到的模糊规则的前件将具有清晰的语义可解释性。若无具体指明,下文的可解释性都指代语义可解释性。此外,模糊系统已经被证明是一个通用逼近器[19],因此能以任意精度逼近非线性函数,其性能有着坚实的理论保证。模糊系统为处理不确定的数据、代表潜在的知识和展示推理过程提供了一个有效的范式。因此,近年来基于模糊系统的模糊人工智能(Fuzzy AI)[20]得到了广泛的发展。为了改善最初提出的模糊系统的性能,多个主流的模糊系统的变体被相继提出。例如,Mamdani型模糊系统[21]通过给模糊系统中添加模糊化器和解模糊器,使得模糊系统得到了精确值的输出;Wang-Mendel模糊系统[22]给出了一种从数值数据中生成模糊规则的新方法。Takagi-Sugeno-Kang(TSK)模糊系统[17,23-24]通过在模糊系统中使用参数估计的方法来确定系统参数,以使得模糊规则的输出为精确值。作为最常用的模糊系统,TSK模糊系统有着较好的非线性逼近能力,较简洁的规则形式和高可解释性。下面以TSK模糊系统为研究对象,介绍以TSK模糊系统为基础发展而来的TSK人工智能(TSK AI)。

由于TSK模糊系统有着高可解释性,其最终决策可以被由训练过程得到的若干条模糊规则很好地解释。此外,TSK模糊系统能很好地解决不确定问题,即可以对其他类型得模型难以表达的场景进行有效地建模。特别是,当用户必须处理数据的缺乏问题或输入数据定义的不确定性时,TSK模糊系统将成为一个非常有效的工具。因此,TSK AI被广泛应用于数据挖掘、工业控制和模式识别等领域[25-27]。依据TSK AI中对TSK模糊系统改进方式的不同,它们可以被大致分为3类。

第1类TSK AI是对单个TSK模糊系统使用不同优化方法提高其性能,以使得其能适应不同的应用场景。例如,著名的自适应神经网络的模糊推理系统(Adaptive network-based fuzzy inference system, ANFIS)[28-29]使得模糊系统有了自学习能力;进化模糊系统(Evolutionary fuzzy system,EFS)[30]使用遗传算法对模糊系统的参数进行优化,极大改善了TSK模糊系统的精确度;区间2型TSK模糊系统[31]将一类模糊集的不确定性问题建模为一类区间模糊数,增强了TSK模糊系统应对于高不确定问题的能力。其他的代表模型还包括:基于支持向量机的TSK模糊系统[32]、多任务TSK模糊系统[33]、基于极限学习机的TSK模糊系统[34]、以及可扩展的TSK模糊系统[35]等,它们都在不同的学习任务中很大地增强了TSK模糊系统的性能。然而,由于维度诅咒问题[36-37]的存在,单个TSK模糊系统的精确度和可解释性容易受到规则爆炸问题的影响,尤其是面对近年来出现的大规模高维数据。

第2类TSK AI是将模糊系统和深度神经网络相结合,以增强深度神经网络的可解释性或处理不确定性问题的能力,构成基于神经-模糊混合的TSK模糊神经网络。这类TSK AI的主要创新之处是:利用模糊数的概念来表示神经网络的权重,或用模糊逻辑单元取代神经网络中的感知器,或用模糊系统训练神经网络的参数。例如,文献[38]通过将区间2型TSK模糊系统和神经网络相结合,提出一种简化的区间2型模糊神经网络,在不确定性问题中取得了更好的测试性能和更低的计算复杂度。文献[39]提出了TSK型卷积递归模糊网络(TSK-type convolutional recurrent fuzzy network, TCRFN),它将TSK模糊系统和卷积递归神经网络相结合,提高了网络处理脑电图EEG信号里噪音的能力;文献[40]将1组带有小波函数的TSK模糊系统和模糊小脑模型神经网络相结合,所提出的模型在不确定的非线性系统中取得了比其他神经网络模型更优越的性能;文献[41]提出一种模糊神经网络技术,可以从给定的输入和输出数据集中提取TSK型模糊规则,以用于后续的系统建模。相比于TSK模糊系统,尽管TSK模糊神经网络的性能得到了很大提高,但是它们的整体结构仍然属于深度神经网络这个黑箱模型的范畴,所以模型的可解释性在一定程度上被削弱了。

第3类TSK AI是使用集成的方式来组织多个TSK模糊系统,以获得更好的性能。依据组合方式的不同,可以概括如下:

(1)宽度结构。宽度TSK模糊系统将若干个TSK模糊子系统在宽度层面上进行集成,以保持快速的并行/增量学习过程以及高可解释性。具体来说,一方面一般的集成策略[37](例如,Bagging和Boosting)经常被用于有效地结合若干个TSK模糊子系统,然后在所有子系统上的输出使用常用的聚合策略,如平均法、加权法和多数投票法,来得到整个结构的最终输出。为了增强在TSK模糊子系统之间的多样性,Bagging[42-43]对原始训练数据集进行随机采样,作为新的训练数据集,以减少不同TSK模糊子系统之间的关联性。而Boosting[44-45]依次训练每个新的TSK模糊子系统,并且更加关注那些在前一个TSK模糊子系统中表现不理想的训练实例。另一方面,除了Bagging和Boosting,研究人员还提出了许多新的方法来构建有效的TSK模糊系统的宽度集成。例如,文献[46]提出了利用1个一阶TSK模糊模型来聚合多个TSK子系统的输出,而不是使用传统的线性聚合方法,如平均法。文献[47]通过计算相应输出的权重,使用TSK模糊系统来动态选择有能力的子分类器,然后通过多数投票或可调融合算法将这些被选中的子分类器的输出聚合为最终输出。在笔者最近的工作[48-49]中,一方面基于模糊知识退出的概念[48]提出了由多个TSK模糊子系统组成的宽度集合结构(Wide learning based TSK fuzzy classifier, WL-TSK),它很好地模拟了人类的知识丢弃过程。WL-TSK最后对所有的TSK子系统使用平均法、加权法或多数票法来得到最终输出,WL-TSK可以获得令人满意的分类性能,并具有高可解释性。另一方面,文献[49]通过使用动态正则化模仿人类思维过程中对知识的鲁棒使用,设计了一种称为KAT的新型知识对抗训练方法,以实现零阶TSK模糊分类器增强的泛化性能、可解释性和快速训练,最后将多个知识对抗零阶TSK模糊子分类器进行宽度集成来获得最终输出。

(2)深度结构。依据文献[50]中的猜想,深度神经网络在处理复杂问题上的成功在于:①深度神经网络的深层结构可以捕捉到输入数据的高级的抽象的特征,并以逐层的方式很好地描述输入数据集的特征;②神经网络有着足够高的模型复杂度,即大量的网络可训练参数。这意味着深度学习模型的构件不一定要局限于神经网络。类似的建模思想以前在层级模糊系统中也有过研究[51],层级模糊系统最初是为了克服模糊系统在处理高维问题时的缺点而提出的,它的通用逼近性已经在文献[52]中得到了证明,并在文献[53-54]中得到了进一步的发展,这意味着层级模糊系统的性能已经有了十分坚固的理论保证。一般来说,层级模糊系统由许多低维模糊子系统组成,这些子系统以逐层连接的方式进行连接,这些连接方式主要分为递增式、聚集式和级联式,如图1所示。层级模糊系统近年来的进展可以总结如下:文献[55]提出了一个可扩展的模糊系统框架,该框架通过使用层级表示法来考虑模糊规则的优先级;文献[56]提出一个自适应层级模糊系统,它有利于调整一些控制器的一些参数,同时减少每个处理器中输入变量和模糊规则的数量。其他经典的层级模糊系统还包括文献[51,57-58]。尽管层级TSK模糊系统有效地解决了模糊系统遇到高维数据时产生的规则爆炸问题,然而层级TSK模糊系统的中间变量(即,每个TSK模糊子系统的输出)却变得难以理解,因此中间层和输出层的模糊规则的可解释性也被降低了。特别是,因为前1层

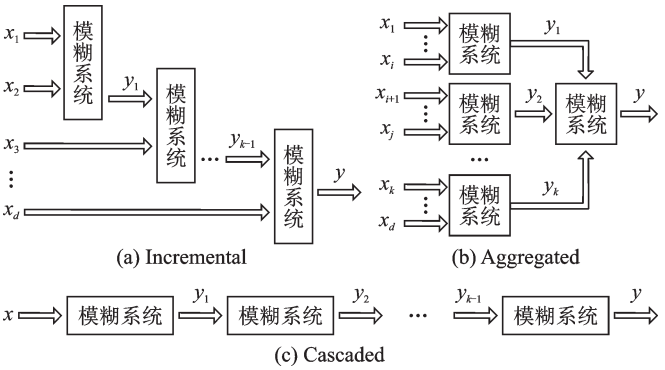

图1 层级TSK模糊系统的结构

Fig.1 Structures of hierarchical TSK fuzzy systems

的输出被当成当前层的输入,这种可解释性上的困难随着层级TSK模糊分类器中层数的增加而变得严重起来。图1(c)中的级联式层级模糊系统虽是深度结构,但其并没有使用深度学习的方法进行优化,因此很难保持良好的泛化性能。为了解决层级结构的可解释性因中间变量变差的问题,同时提高层级结构的泛化性能,最近一系列深度模糊系统在层级模糊系统的基础上以深度学习的方式被开发出来。这类深度模型使用栈式泛化原理[59]来提升模型性能,并将深度学习里的方法引入模糊系统中,从而摆脱了对深度神经网络的依赖,并且保持了模糊系统的高可解释性。因此,这类可解释的深度TSK模糊系统[60-62]正在成为探索新的深度学习模型的一个有希望的潜在途径。

近年来,以深度模糊系统为研究对象的综述类文章受到了研究人员越来越多的关注。例如,文献[63]通过分析模糊系统的发展历程和近年来的研究进展,总结出在大数据时代下,结合模糊系统的可解释性和深度神经网络的学习能力将是未来解决高维数据问题的有力途径。文献[64]总结了模糊系统和神经网络的发展历程,并分析了两者在脑科学上的联系,最后得出两者的结合将会是脑综合研究领域中一个十分有潜力的尝试。文献[65]讨论了神经网络和模糊系统两者结合的优点,然后介绍了由两者结合得到的模糊神经网络在工业上的应用情况。可以看出,当前关于深度模糊系统的综述文章基本以深度模糊神经网络为研究对象,而本文的研究重点则是基于栈式泛化原理的深度模糊系统,其目的是在保持可解释性的前提下取得令人满意的测试泛化性能。而在实际应用中,相比于训练性能,测试泛化性能往往更为重要,这也增强了基于栈式泛化原理的深度模糊系统的实际应用能力。以此为出发点,本文总结了深度模糊系统的代表模型、应用场景和未来发展趋势。

1 TSK模糊系统

作为最常用的模糊系统,TSK模糊系统[17,23-24]的模糊规则(以第k条为例)的形式可以表达为

$$\begin{aligned}&\text{If } x_1 \text{ is } A_{k1} \text{ and } x_2 \text{ is } A_{k2} \text{ and } \cdots \text{ and } x_d \text{ is } A_{kd}\\&\text{Then } y_k = f_k(\boldsymbol{x}) \qquad k = 1, 2, \cdots, K\end{aligned} \tag{1}$$

式中:x_1代表输入向量$\boldsymbol{x}$的第1个特征;A_{ki}代表第k条规则在第i个输入特征x_i上的前件模糊集;$f_k(\boldsymbol{x})$代表第k条规则的后件;K代表模糊规则的数量;and代表模糊操作符;d是输入向量$\boldsymbol{x}=[x_1, x_2, \cdots, x_d]^{\mathrm{T}}$的特征数量。所有这些模糊规则构成了一个模糊知识库[48-49],其中每个模糊规则都可以被看作是一块模糊知识。可以看出,式(1)中的If-Then模糊规则和人类的语言非常接近,也就是说模糊规则可以被人类直接读懂和理解,因此TSK模糊系统有着高可解释性[66-68]。除了式(1)中的模糊规则外,还有如式(2)和式(3)中的其他模糊规则形式,它们的特点是带有用于评估模糊规则重要程度的指标。

$$\begin{aligned}&\text{If } x_1 \text{ is } A_{k1} \text{ and } x_2 \text{ is } A_{k2} \text{ and } \cdots \text{ and } x_d \text{ is } A_{kd}\\&\text{Then } y_k = f_k(\boldsymbol{x})\\&\text{confidence} = r_k, \text{support} = s_k \qquad k = 1, 2, \cdots, K\end{aligned} \tag{2}$$

式中r_k和s_k分别代表模糊规则的置信度和支持度,是两个最常用的评估模糊规则重要程度的统计指标[69]。

$$\begin{aligned}&\text{If } x_1 \text{ is } A_{k1} \text{ and } x_2 \text{ is } A_{k2} \text{ and } \cdots \text{ and } x_d \text{ is } A_{kd}\\&\text{Then } y_k = f_k(\boldsymbol{x}) \text{ with } \mathrm{CF}_k \qquad k = 1, 2, \cdots, K\end{aligned} \tag{3}$$

式中CF_k代表模糊规则的规则权重(即不确定度),用于综合地评估模糊规则的重要程度[70]。借助于式(2)和式(3)中的评估指标,人们可以从模糊系统得到的所有规则中找出较为重要的模糊规则,依据这些较为重要的模糊规则来对模糊系统的最终决策进行解释。由于参与解释的模糊规则数量的减少,模糊系统的可解释性得到了提高。

通常来说,TSK模糊规则的前件模糊集A_{ki}使用高斯函数作为其模糊隶属度函数,即

$$\phi_{ki}(x_i)=\exp\left(\frac{-(x_i-c_{ki})^2}{\delta_{ki}}\right) \tag{4}$$

式中:c_{ki}和δ_{ki}分别代表高斯隶属度函数的中心和带宽;x_i代表输入向量$\boldsymbol{x}$的第i个特征。依据文献[71-72],TSK模糊系统的输出可以表示为以下两种方式。

(1)经过解模糊处理

$$Y=\sum_{k=1}^{K}\frac{\phi_k(\boldsymbol{x})}{\sum_{l=1}^{K}\phi_l(\boldsymbol{x})}f_k(\boldsymbol{x})=\sum_{k=1}^{K}\tilde{\phi}_k(\boldsymbol{x})f_k(\boldsymbol{x}) \tag{5}$$

(2)没有经过任何解模糊处理

$$Y=\sum_{k=1}^{K}\phi_k(\boldsymbol{x})f_k(\boldsymbol{x}) \tag{6}$$

式中$\phi_k(\boldsymbol{x})=\prod_{i=1}^{d}\phi_{ki}(x_i)$,此时模糊操作符and使用相乘操作来实现。大多数时候,TSK模糊系统的后件$f_k(\boldsymbol{x})$有不同的两种表达形式:一种是待定常数$f_k(\boldsymbol{x})=p_{k0}$;另一种是线性函数,即$f_k(\boldsymbol{x})=p_{k0}+p_{k1}x_1+\cdots+p_{kd}x_d$。前者的TSK模糊系统被称为零阶TSK模糊系统[73],后者的TSK模糊系统被称为一阶TSK模糊系统[60]。很明显,零阶TSK模糊系统有更加简单的后件形式。TSK模糊系统已经被证明是一个通用逼近器[23],并在实际生活和工业生产中得到了广泛的应用[25-27]。

TSK模糊系统的基本特征之一就是可解释性。正如Kuncheva[66]所指出的:“一旦可解释性被作为对系统的要求而被否定,模糊分类器就会落入众多以其性能来判断的其他设计中。这些设计包括统计分类器和神经网络,而模糊分类器很难成为它们的最佳对手。”可以看出,与深度神经网络相比,可解释性是TSK模糊系统的一大优势。影响模糊系统可解释性的相关关因素[74]可以总结如下:(1)模糊分区的可解释性;(2)基于模糊规则的系统的规模;(3)IF-THEN规则的复杂性;(4)推理过程和解模糊处理的简易性。

目前,对于模糊系统来说,常用的可解释性评价指标[66,75]可以总结如下:(1)模糊规则的数量,更少的模糊规则通常意味着更高的可解释性;(2)模糊规则包含的特征数量,即模糊规则的长度,更短的模糊规则通常意味着更高可解释性,这也是短规则受到研究人员青睐的原因[71,76];(3)模糊规则的后件复杂度,更简单的模糊规则后件形式通常意味着更高的可解释性,例如后件形式更为简单的零阶TSK模糊规则通常比一阶TSK模糊规则有着更高的可解释性[77]。

具体来说,零阶TSK模糊系统的分类性能比一阶TSK模糊系统要差[62]。然而,与一阶TSK模糊系统相比,零阶TSK模糊系统具有更简洁的可解释性。因为每条模糊规则的后件部分只涉及一个参数p_0^k,因此$p_0^k/\max_k(|p_0^k|)$的正负值可以明确地解释为支持或反对被归入第k类的确定度。相反,一阶TSK模糊系统很难对每条模糊规则的后件部分所涉及的$(d+1)$个后件参数作出明确的解释。

关于如何确定TSK模糊规则的前件部分(即模糊隶属度函数的参数c_i^k和δ_i^k),有多种方法已经被提出。例如Wang-Mendel方法[78]和基于聚类的方法[79-80],如模糊C均值聚类(Fuzzy c-means clustering, FCM)[81],这些方法可以保证TSK模糊规则前件的可解释性。此外,在最近的工作中[24,48,62,71,76,82],一种快速确定模糊规则的语义前件的方法被提出,即使用固定的语义分区来确定模糊前件。该方法首先将输入数据的每个特征分为5个相等的部分,然后生成5个中心分别固定在{0, 0.25, 0.5, 0.75, 1.0}的模糊隶属函数,这5个中心分别与5个明确的语义值一一对应,例如:{很低、低、中、高、非常高}。这些语义值之间虽然有时很难划清他们的界限,但它们的含义一般都是可以被正确理解的,不会引起误会。

因此每个模糊规则的前件部分可以通过随机选择这5个模糊隶属度函数中的一个来生成,以使得每个模糊规则的前件具有高可解释性。使用固定的语义分区的另一个原因是,这样的做法更加符合人们表达问题的方式,例如人们常说某件事发生的可能性“比较小”,而不习惯于使用一个具体的数来指出程度的大小,并且在多数情况下人们很难给出一个表示程度大小的数。

传统上,TSK模糊规则的后件部分可以由一些流行的梯度下降算法[83]求解,即根据输入数据的标签和TSK模糊系统的输出之间的差异来反复迭代确定,然而当输入数据有很大的规模时,训练过程通常是非常耗时的。因此,为了加快TSK模糊系统的训练过程并提高其性能,许多有效的学习算法已经被提出[84-88]。例如,最小二乘法[84]、伪逆法[85]或极限学习机(Extreme learning machine, ELM)[89]。此外,依据最近关于最小学习机(Least learning machine,LLM)[86-88]的工作,证明了ELM和岭回归的等价性,从而LLM成为比ELM的最初伪逆版本更灵活的解法。因此,LLM可以一次性快速求解模糊规则的后件参数,从而有效地避免了耗时的训练过程。LLM的有效性已经得到了广泛的验证[60-62]。

2 可解释的深度TSK模糊系统

相比于TSK模糊系统的其他变体结构,可解释的深度TSK模糊系统有着以下4点明显的优势:

(1)与单个TSK模糊系统相比。面对复杂或者高维输入数据,为了取得令人满意的分类性能,单个TSK模糊系统通常需要大量的模糊规则,而过多的模糊规则不可避免地降低了模糊系统的可解释性。因此,单个TSK模糊系统的性能和可解释性很容易受由维度诅咒[36-37]引发的规则爆炸问题所影响。相反,因为可解释的深度TSK模糊系统可以通过不断加深其结构来增强其分类性能,所以其包含的TSK模糊子系统可以使用相对少的模糊规则。具体来说,在相同学习任务下,借助栈式泛化原理[59]、深度结构和深度学习的方法,可解释的深度TSK模糊系统可以学习到原始输入样本里高级的抽象的特征,因此所需的模糊规则数量往往要比单个TSK模糊系统要少得多,从而有着更高的可解释性。

(2)与宽度TSK模糊系统相比。虽然宽度TSK模糊系统通常有着更高的可解释性,但是可解释的深度TSK模糊系统可以从原始输入特征中学习到更加高级的和抽象的特征,因此可解释的深度TSK模糊系统可以处理更为复杂的学习任务。

(3)与深度模糊神经网络相比。可解释的深度TSK模糊系统的每个子系统都可以借助LLM[86-88]实现只需要1次的快速训练,因此其整个结构不需要像深度模糊神经网络那样使用基于反向传播的梯度下降算法[83]来反复迭代其网络参数,从而保持了快速训练的优势,也避免了如梯度消失[90]等问题的产生。此外,它的分类性能可以随着深度的增加(即TSK模糊子系统数量的增加)不断地增强。层数的增加也可以被视为是一种增量学习,使得深度TSK模糊系统可以动态地更新其网络结构,以适应不同难度的学习任务,而不需要像模糊神经网络那样重新训练整个网络结构,这大大增强了其实用性。

(4)与层级模糊系统相比。可解释的深度TSK模糊系统通常是将若干TSK模糊子系统在栈式泛化原理[59]下进行栈式堆叠得到的,每个TSK模糊子系统依旧保持着单独训练的方式和高可解释性。因此,整个结构训练完成之后,可以根据不同需求选取不同层的TSK模糊子系统得到的模糊规则来对整个结构的最终决策进行解释,这使得整个结构都具有良好的可解释性。此外,借助于深度学习里的方法,可解释的深度TSK模糊系统通常可以取得更好的泛化性能。

栈式泛化原理最早由文献[59]提出,是一种提高模型泛化能力的深度集成方法,其核心思想是利用前一层模型的输出来提高当前层模型的泛化性能。实现栈式泛化原理的方式有很多,其中一种最为经典的方式如图2所示。栈式泛化原理首先在第1层分别训练了若干个弱分类器,然后将第1层分类器的

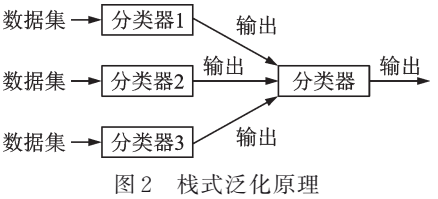

图2 栈式泛化原理

Fig.2 Stacked generalization principle

输出作为第2层分类器的输入,以提高第2层分类器的泛化性能。此外,栈式泛化原理可以在每一层训练不同类型的分类器,这大大增强了其实用性。尽管栈式泛化原理没有Bagging和Boosting[37]那么流行,但它的有效性已经在提高无监督学习和有监督学习的性能方面得到了证明。此外,栈式泛化原理可以通过不断地打开原始输入空间的流形结构来保证其增强的泛化能力。更为重要的是,栈式泛化原理可以有效地避免去解决一个困难的非凸优化问题,而目前大多数的深度学习方法却都困扰于这个问题。因此,从这个角度来看,将栈式泛化原理引入建立深度TSK模糊系统中是十分合适。近年来,可解释的深度TSK模糊系统因其优越的分类性能和高可解释性得到了研究人员越来越多的关注,其中的代表模型可以总结如下。

为了改善传统的层级TSK模糊系统的中间变量难以解释的问题,可解释的深度TSK模糊系统首先在文献[62]中被提出。具体来说,基于栈式泛化原理,作者提出了一种新颖的深度TSK模糊分类器D-TSK-FC。如图3所示,D-TSK-FC栈式堆叠了若干个零阶TSK模糊子分类器,除了第1个子分类器建立在原始输入数据外,剩下的子分类器的输入都设置为原始输入数据加上前一个子系统输出的随机偏移。此外,D-TSK-FC借助三重简洁的模糊规则,即随机选择的原始特征加上固定的语义分区、随机的规则组合和相同的子分类器输入空间,实现了增强的分类精度和高可解释性。通过引入深度学习的技术,D-TSK-FC改善了传统层级TSK模糊系统的分类性能和可解释性。

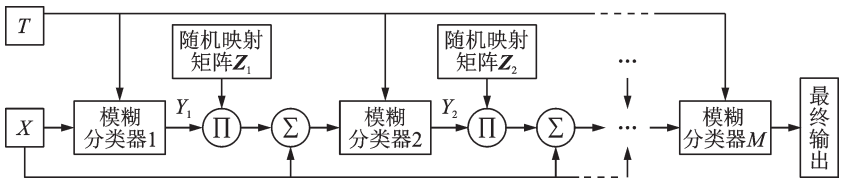

图3 D-TSK-FC的结构

Fig.3 Structure of D-TSK-FC

在D-TSK-FC提出之后,可解释的深度TSK模糊系统吸引了越来越多研究人员的关注。例如,文献[61]通过将零阶TSK模糊分类器以一种特殊的堆叠方式组装起来,提出了一种可解释的高阶深度TSK模糊分类器DHO-TSK。相比于高阶TSK模糊分类器,DHO-TSK不仅有着更好的可解释性,而且它理论上等价于1个高级TSK模糊分类器。DHO-TSK的结构如图4所示。从图4可以看出,它的每层都包含1个零阶TSK子分类器,除了第1层的输出是第1个子分类器的输出外,其余层的输出都是当前层子分类器的输出乘上对应的随机选中的原始特征,再加上前一层子系统的输出。这样一来,DHO-TSK实现了令人满意的分类性能和高可解释性。

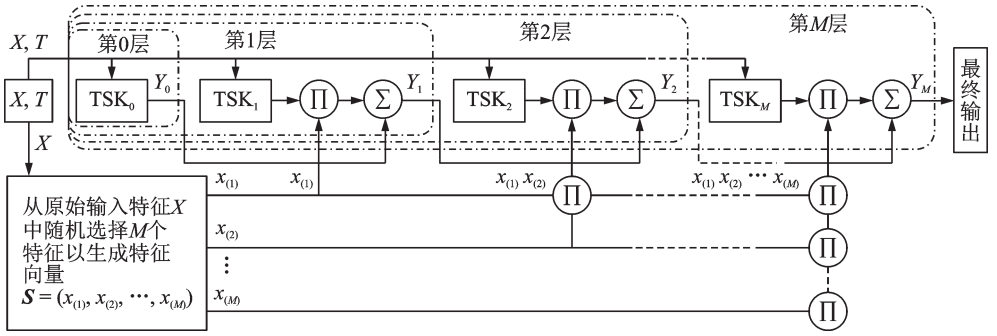

图4 DHO-TSK的结构

Fig.4 Structure of DHO-TSK

文献[60]首先提出了1个基于输出扰动的对抗TSK模糊分类器TSKa,在理论上TSKa有着增强的泛化性能。然后,基于栈式泛化原理,作者将若干个TSKa进行栈式堆叠,得到1个深度对抗TSK模糊系统DSA-FC,其结构如图5所示。除了第1个子分类器外,其余的子分类器不仅利用了原始输入数据,还同时利用到了前一个子分类器的平滑梯度信息G,用以避免在各层输入数据空间的生成中出现不均匀现象。DSA-FC在分类精度、抗噪性能和可解释性方面都取得了令人满意的结果。

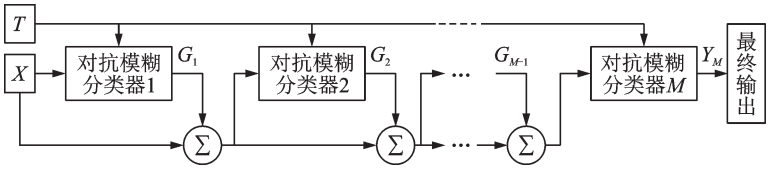

图5 DSA-FC的结构

Fig.5 Structure of DSA-FC

文献[91]为了解决DSA-FC[60]存在的面对大规模数据时训练速度慢的问题,开发了一种针对DSA-FC的快速训练算法FTA。FTA的训练工作流程如图6所示。FTA首先在每个子对抗模糊分类器的所有模糊规则中选出前k条模糊规则(见图6中虚线圆里的模糊规则)。然后从这些选择的模糊规则中生成一阶平滑的梯度引导信息。最后根据这些信息快速更新当前的输入,也就是说,这些信息将加入到下一层子分类器的输入中。FTA在理论上能提高DSA-FC的泛化能力,同时实验上表明了其对DSA-FC加速能力的有效性。

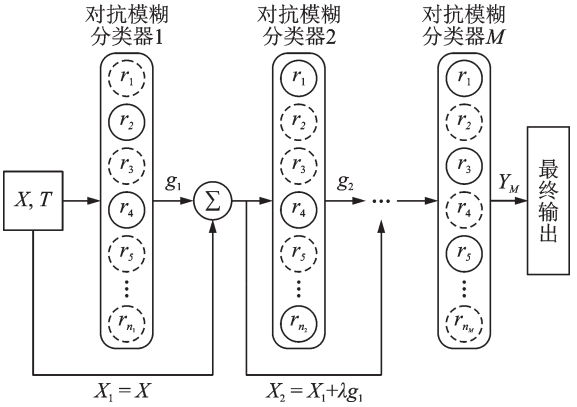

图6 FTA-FC的训练工作流程

Fig.6 Training workflow of FTA-FC

文献[71]提出一种基于共享语义模糊规则的深度TSK模糊分类器HID-TSK-FC,其结构如图7所示。HID-TSK-FC使用栈式结构堆叠了若干个TSK模糊子分类器,除了第1个子分类使用零阶TSK模糊系统,其余的子分类器均使用特殊的TSK模糊系统,即将前面所有子系统的输出对原始的输入数据进行扩维处理,以打开输入空间的流形结构,并体现到模糊规则的后件形式中。为了取得更好的分类的性能,HID-TSK-FC使用梯度下降法来更新后件里的所有参数。此外,HID-TSK-FC在数学上等价于1个具有共享可解释语言模糊规则的新型TSK模糊分类器,因此其每一条模糊规则都是可解释的。

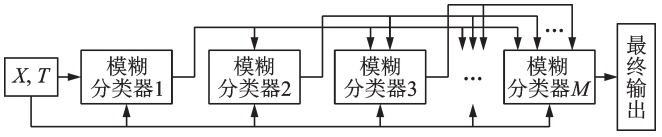

图7 HID-TSK-FC的结构

Fig.7 Structure of HID-TSK-FC

文献[92]提出一种基于栈式堆叠结构的深度TSK模糊分类器SHFA-TSK-FC,以解决现有层级式模糊分类器在解释中间层变量和模糊规则方面的不足。图8给出了SHFA-TSK-FC的结构。SHFA-TSK-FC每一层的模糊子分类器的输入都设置为原始输入样本的所有输入特征加上前一层的模糊子分

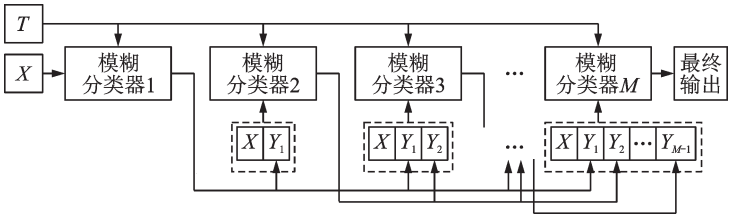

图8 SHFA-TSK-FC的结构

Fig.8 Structure of SHFA-TSK-FC

类器的输出。这样做的好处是,这些扩维后的输入特征可以从本质上打开原始输入空间的流形结构,从而增强模糊子分类的分类性能。因此,SHFA-TSK-FC实现了增强的分类性能和高可解释性。

文献[93]借助于栈式泛化原理,通过在深度集成中对少数类及其问题区域栈式堆叠若干个零阶TSK模糊子分类器,提出了一种深度TSK模糊分类器IDE-TSK-FC,用以处理不平衡数据分类任务。IDE-TSK-FC的结构如图9所示。从图9可以看出,除了第1个零阶TSK模糊子分类器是建立在原始训练数据集外,后续的所有零阶TSK模糊子分类器都被逐层堆叠在原始训练数据集中由K近邻(K-nearest neighbor, KNN)识别的问题区域和之前所有子分类器的平均输出上。借助于栈式泛化原理[59],IDE-TSK-FC在类不平衡问题上实现了良好分类性能和高可解释性。

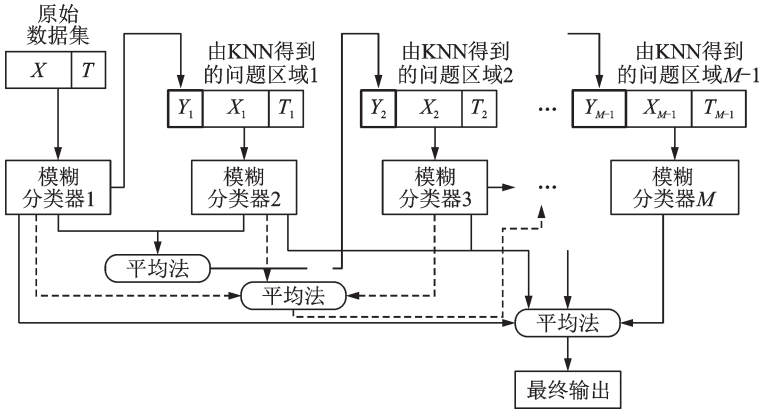

图9 IDE-TSK-FC的结构

Fig.9 Structure of IDE-TSK-FC

文献[94]提出一种多视角深度TSK模糊系统DVR-TSK-FS用于检测癫痫性脑电信号,其结构如图10所示。依据图10,DVR-TSK-FS使用栈式泛化原理堆叠了若干个1阶TSK模糊子分类器,每个子分类器都构建在p个不同视角下的数据集上。此外,除了第1个子分类器构建在p个原始数据集上外,其余的子分类器使用前面所有子分类器的输出对p个原始数据集进行了扩维处理。相比于单视角的模糊系统,DVR-TSK-FS在检测癫痫性脑电信号上取得了更好的效果。

除了可解释的深度TSK模糊系统外,以其他模糊系统为基本构件的可解释的深度模糊系统也得到了广泛的发展。其代表模型包括文献[78]依据栈式泛化原理,提出了一种可解释的深度Wang-Mendel模糊系统DFRBCS,DFRBCS采用逐层堆叠的构建方式,每层子系统的输入是通过对前一层所有模糊子系统的输出进行洗牌和滑动窗口操作产生的。滑动窗口的使用既实现了降维的效果,也最大限度地保持了原始的输入特征,因此取得了良好的分类性能和可解释性。文献[95]使用卷积操作进行特征提

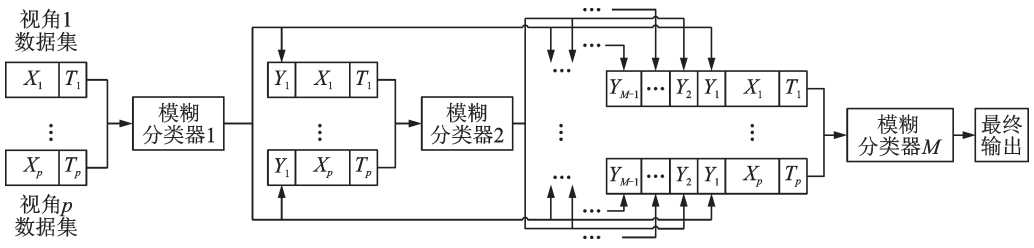

图10 DVR-TSK-FS的结构

Fig.10 Structure of DVR-TSK-FS

取,逐层构建了一种可解释的深度Wang-Mendel模糊系统DCFS。DCFS的第1层包含若干弱的子分类器,它们使用卷积操作(1个移动窗口)在原始输入数据中提取特征。然后,第2层以相同的方式构建在第1层子分类器的输出上面,逐层重复这个过程,直到达到满意性能或最大层数。DCFS在拟合真实香港股票市场数据上取得了令人满意的性能。

在实际应用方面,因为可解释的深度模糊系统可以在保持可解释性的前提下取得令人满意的测试泛化能力。在实际应用中,测试泛化能力往往比训练性能更为重要。因此,以TSK模糊系统及其他模糊系统为基本构件的可解释的深度模糊系统在以下方面已经得到了重要的应用:

(1)中小规模数据集、数据缺乏和不平衡问题。在关注与可解释的深度模糊系统相关的鲁棒性属性时,首先要考虑的是它们在提取领域内密度较低的中小规模数据集中信息时的有效性,即所谓的"缺乏数据"问题[96]。原因很简单:通过沿着输入数据原始特征的总体领域定义模糊前件的语义空间,模糊前件可以对输入空间进行整体覆盖。此外,当前件模糊集之间存在相互重叠时,建模的信息颗粒之间可以获得一个更为平滑的过渡。除了缺乏数据问题之外,可解释的深度模糊系统在处理不精确和不确定数据方面也非常有效[62,92]。此时,模糊分区的定义以及隶属函数的灵活性是十分重要的,例如,可以使用传统的Ⅰ型模糊集的不同扩展来增加表示中的额外自由度[97]。

(2)社会网络分析问题。在过去的几年里,由于社会媒体互动的增加,社会网络分析已经成为一个热门话题。企业和学术界对这些关系的概念化、模型化、分析、解释和预测非常感兴趣。社会网络分析所基于的图论与模糊集理论之间存在着自然的联系,这允许模糊系统提供一种更容易和更强大的方式来表达这些网络中节点之间的关系。此外,一些研究已经讨论了基于模糊集的社会数据的理论、概念模型和实际应用[98]。

(3)金融。金融数据的固有不确定性造成了人们难以对其规律进行准确的预测,同时金融领域也急切需要可解释的模型,以便于用户可以放心地使用。在金融环境中,理解输入和输出是如何相互关联对于能够做出操作性和战略性的决策至关重要。因此,可解释的深度模糊系统已经成功应用于许多金融领域,例如股票走势预测[95]。

(4)医学。医学应用中的任何决定对于医生和病人来说都至关重要,因此医生采取的任何行动都必须有十足信心。这意味着在这种情况下使用的任何决策支持系统必须是可信的和透明的。换句话说,它必须向医生和病人解释某项诊断背后的原因,即模型必须是可解释的。在这个意义上,可解释的深度模糊系统将是合适的选择[27,99]。

(5)入侵检测系统。信息系统的广泛使用和建立安全策略和规则的需要,使不想要的系统访问被区分开来。其中,可解释的深度模糊系统有着广阔的应用前景,原因可归结如下。首先,入侵检测问题有一个共同的结构,事实上它们是由数字数据描述的,因此清晰的阈值会导致低的检测精度。此外,合法行为和异常行为之间的界限本身是模糊的。换句话说,入侵行为中的微小变化可能不会被识别,而

正常情况下的微小偏差可能会产生一个错误的警报[60,91]。

综上所述,近年来,可解释的深度TSK模糊系统得在理论、模型和实际应用方面都得到了广泛的发展。但其还面临着以下的挑战和机遇:

(1)更复杂数据的处理能力。虽然可解释的TSK模糊系统在面对中小规模数据集上有竞争性优势,但面对大规模数据集的性能还有进一步的探索空间。深度神经网络的成功经验已经证明了越深的结构和越多的参数量是处理复杂问题的关键。借助残差神经网络[100],深度神经网络解决了梯度消失[90]问题,因此其层数可以轻松实现几十层甚至是上百层,以帮助其在大规模的和更具挑战性的数据集上取得令人满意的精确度。然而,现阶段的可解释的深度TSK模糊系统往往采用逐层训练和栈式堆叠的方式来加深其层数,因此其结构大多在10层以下。虽然更深的结构是提高深度TSK模糊系统处理更复杂数据的一个有效方式,但是深度的增加也意味着模型参数和复杂度的增加,这不可避免会降低所得到的模糊规则的可解释性。如果在深度TSK模糊系统的设计中,更多关注的是精确度而不是可解释性,那么得到的模糊系统就很难与其他更可取的、更复杂的解决方案相比较,比如深度神经网络。因此,面对更复杂数据,如何在发挥深度TSK模糊系统的优势,仍是困扰深度TSK模糊系统的一个难题。

(2)多种可解释性。目前,深度TSK模糊系统的可解释性仅仅考虑的是语义可解释性,然而可解释性的含义是广泛的,仅仅考虑语义可解释性显然是不够的也是不全面的[101]。具体来说,模糊系统的可解释性还包含后件的可解释性,整条模糊规则的可解释和可视化的可解释性等。在深度神经网络里,可视化的可解释性也是一种去理解深度神经网络行为的常用方法,例如可视化的饼图[102]。因此,未来的评价可解释的指标应该是多个指标的综合,而不单单是语义上的可解释性。如何让深度TSK模糊系统在更加综合的可解释性指标下依旧取得令人满意的测试泛化性能将会是一个有趣的方向。

(3)更多深度学习领域的技术。深度TSK模糊系统正是因为使用了深度学习领域的技术,所以取得了比传统的层级式TSK模糊系统更好的性能和可解释性。近年来,深度学习领域出现一些颇具潜力的新兴技术,如著名的拥有注意力机制的Transformer[103]。Transformer不仅仅在自然语言处理方面取得了巨大的成功[104],最近能有效地处理图像数据的Transformer被相继提出,例如国际顶会ICCV的2021最佳论文里提出的Swin transformer[3]就在多个图像数据集上达到了最先进的性能。然而,深度TSK模糊系统和Transformer的结合还鲜有研究,未来两者的结合或许能给深度TSK模糊系统更加有效地处理文本和图像数据另一种有趣的思路。

3 结束语

数据科学的世界已经改变了对模型性能的要求。以往模型只需要一味地追求高精确度,因此模型的复杂度被不断提高。但目前,模型的核心不仅要达到尽可能高的精度,而且要使其对研究人员和从业人员具有可解释性。在这个意义上,可解释的深度TSK模糊系统保留了模糊系统可解释的原始本质,也通过深度结构提升了其建模能力,因此提供了比其他范式更多的优势。本文从可解释的深度TSK模糊系统出发,分析了深度TSK模糊系统相对于其他TSK模糊系统变体存在的优势;总结了深度TSK模糊系统当前主流的模型和实际的应用场景;并据此分析了深度TSK模糊系统未来可能面临的挑战和机遇。

参考文献:

[1] GONG M, ZHAO J, LIU J, et al. Change detection in synthetic aperture radar images based on deep neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(1): 125-138.

[2] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[3] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2021: 10012-10022.

[4] GU Z, CHENG J, FU H, et al. CE-Net: Context encoder network for 2D medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2019, 38(10): 2281-2292.

[5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2021-07-20]. http://arxiv.org/abs/1409.1556.

[6] CRESWELL A, WHITE T, DUMOULIN V, et al. Generative adversarial networks: An overview[J]. IEEE Signal Processing Magazine, 2018, 35(1): 53-65.

[7] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2016: 770-778.

[8] SAMEK W, WIEGAND T, MÜLLER K R. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models[EB/OL]. (2017-08-28)[2021-07-20]. http://arxiv.org/abs/1708.08296.

[9] CASTELVECCHI D. Can we open the black box of AI?[J]. Nature News, 2016, 538(7623): 20.

[10] RUDIN C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead[J]. Nature Machine Intelligence, 2019, 1(5): 206-215.

[11] LARSON J, MATTU S, KIRCHNER L, et al. How we analyzed the COMPAS recidivism algorithm[J]. ProPublica, 2016, 9(1): 3.

[12] GOODMAN B, FLAXMAN S. European Union regulations on algorithmic decision-making and a "right to explanation"[J]. AI Magazine, 2017, 38(3): 50-57.

[13] DOSHI-VELEZ F, KIM B. Towards a rigorous science of interpretable machine learning[EB/OL]. (2017-01-01)[2021-07-20]. http://arxiv.org/abs/1702.08608.

[14] LIPTON Z C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery[J]. Queue, 2018, 16(3): 31-57.

[15] MONTAVON G, SAMEK W, MÜLLER K R. Methods for interpreting and understanding deep neural networks[J]. Digital Signal Processing, 2018, 73: 1-15.

[16] SIMONYAN K, VEDALDI A, ZISSERMAN A. Deep inside convolutional networks: Visualising image classification models and saliency maps[EB/OL]. (2013-03-28)[2021-07-20]. http://arxiv.org/arxiv.org/abs/1312.6034.

[17] ZADEH L A. Fuzzy sets[J]. Information & Control, 1965, 8(3): 338-353.

[18] ZADEH L A. Fuzzy sets as a basis for a theory of possibility[J]. Fuzzy Sets and Systems, 1978, 1(1): 3-28.

[19] WANG L X. Fuzzy systems are universal approximators[C]//Proceedings of IEEE International Conference on Fuzzy Systems. [S.l.]: IEEE, 1992: 1163-1170.

[20] GARIBALDI J M. The need for fuzzy AI[J]. IEEE/CAA Journal of Automatica Sinica, 2019, 6(3): 610-622.

[21] LIU P. Mamdani fuzzy system: Universal approximator to a class of random processes[J]. IEEE Transactions on Fuzzy Systems, 2002, 10(6): 756-766.

[22] WANG L X, MENDEL J M. Generating fuzzy rules by learning from examples[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1992, 22(6): 1414-1427.

[23] YING H, CHEN G. Necessary conditions for some typical fuzzy systems as universal approximators[J]. Automatica, 1997, 33(7): 1333-1338.

[24] WONG S Y, YAP K S, YAP H J, et al. On equivalence of FIS and ELM for interpretable rule-based knowledge representation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 26(7): 1417-1430.

[25] HERRERA L J, PÉREZ M M, SANTANA J, et al. A data mining approach based on a local-global fuzzy modelling for prediction of color change after tooth bleaching using vita classical shades[C]//Proceedings of the Ninth International Conference on Intelligent Systems Design and Applications. Washing D C, NW: IEEE, 2009: 1268-1273.

[26] ZHONG L. Fuzzy chaos generators for nonlinear dynamical systems[C]//Proceedings of International Conference Physics and

Control. Tokyo, Japan: IEEE, 2003: 429-433.

[27] DENG Z, XU P, XIE L, et al. Transductive joint-knowledge-transfer TSK FS for recognition of epileptic EEG signals[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2018, 26(8): 1481-1494.

[28] KARABOGA D, KAYA E. An adaptive and hybrid artificial bee colony algorithm (aABC) for ANFIS training[J]. Applied Soft Computing, 2016, 49: 423-436.

[29] PRAMOD C, PILLAI G. K-means clustering based extreme learning ANFIS with improved interpretability for regression problems[J]. Knowledge-Based Systems, 2021, 215: 106750.

[30] ISHIBUCHI H, NOZAKI K, YAMAMOTO N, et al. Selecting fuzzy if-then rules for classification problems using genetic algorithms[J]. IEEE Transactions on Fuzzy Systems, 1995, 3(3): 260-270.

[31] LIN C T, PAL N R, WU S L, et al. An interval type-2 neural fuzzy system for online system identification and feature elimination[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 26(7): 1442-1455.

[32] JUANG C, CHEN G. A TS fuzzy system learned through a support vector machine in principal component space for real-time object detection[J]. IEEE Transactions on Industrial Electronics, 2012, 59(8): 3309-3320.

[33] JIANG Y, CHUNG F L, ISHIBUCHI H, et al. Multitask TSK fuzzy system modeling by mining intertask common hidden structure[J]. IEEE Transactions on Cybernetics, 2015, 45(3): 534-547.

[34] YEOM C, KWAK K. A design of TSK-based ELM for prediction of electrical power in combined cycle power plant[C]// Proceedings of 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS). Bang Kok, Thailand: IEEE, 2018: 226-229.

[35] DENG Z, CHOI K, CHUNG F, et al. Scalable TSK fuzzy modeling for very large datasets using minimal-enclosing-ball approximation[J]. IEEE Transactions on Fuzzy Systems, 2011, 19(2): 210-226.

[36] STAVRAKOUDIS D G, GITAS I Z, THEOCHARIS J B. A hierarchical genetic fuzzy rule-based classifier for high-dimensional classification problems[C]//Proceedings of 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011). [S.l.]: IEEE, 2011: 1279-1285.

[37] HU X, PEDRYCZ W, WANG X. Random ensemble of fuzzy rule-based models[J]. Knowledge-Based Systems, 2019, 181: 104768.

[38] LIN Y, LIAO S, CHANG J, et al. Simplified interval type-2 fuzzy neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(5): 959-969.

[39] DU G, WANG Z, LI C, et al. A TSK-type convolutional recurrent fuzzy network for predicting driving fatigue[J]. IEEE Transactions on Fuzzy Systems, 2021, 29(8): 2100-2111.

[40] ZHAO J, LIN C M. Wavelet-TSK-type fuzzy cerebellar model neural network for uncertain nonlinear systems[J]. IEEE Transactions on Fuzzy Systems, 2019, 27(3): 549-558.

[41] CHEN-SEN O, WAN-JUI L, SHIE-JUE L. A TSK-type neurofuzzy network approach to system modeling problems[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2005, 35(4): 751-767.

[42] LUGHOFER E, PRATAMA M, ŠKRJANC I. Online bagging of evolving fuzzy systems[J]. Information Sciences, 2021, 570: 16-33.

[43] SIAMI M, NADERPOUR M, LU J. A choquet fuzzy integral vertical bagging classifier for mobile telematics data analysis [C]//Proceedings of 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). [S.l.]: IEEE, 2019: 1-6.

[44] HOFFMANN F. Boosting a genetic fuzzy classifier[C]//Proceedings of the 9th IFSA World Congress and 20th NAFIPS International Conference. [S.l.]: IEEE, 2001: 1564-1569.

[45] MIYAJIMA H, SHIGEI N, FUKUMOTO S, et al. A learning algorithm with boosting for fuzzy reasoning model[C]// Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007). [S.l.]: IEEE, 2007: 85-90.

[46] COCOCCIONI M, LAZZERINI B, MARCELLONI F. A TSK fuzzy model for combining outputs of multiple classifiers[J]. IEEE Annual Meeting of the Fuzzy Information, 2004, 2: 871-876.

[47] KRYSMANN M. Takagi-Sugeno-Kanga fuzzy fusion in dynamic multi-classifier system[C]//Proceedings of the 2nd World Congress on Electrical Engineering and Computer Systems and Science (EECSS'16). Budapest, Hungary: Elsevier, 2016: 108.

[48] QIN B, CHUNG F L, WANG S. Biologically Plausible fuzzy-knowledge-out and its induced wide learning of interpretable TSK fuzzy classifiers[J]. IEEE Transactions on Fuzzy Systems, 2019, 28(7): 1276-1290.

[49] QIN B, CHUNG F L, WANG S. KAT: A knowledge adversarial training method for zero-order takagi-Sugeno-Kang fuzzy classifiers[J]. IEEE Transactions on Cybernetics, 2020. DOI:10.1109/TCYB.2020.3034792.

[50] ZHOU Z H, FENG J. Deep forest[J]. National Science Review, 2019, 6(1): 74-86.

[51] RAJU G, ZHOU J, KISNER R A. Hierarchical fuzzy control[J]. International Journal of Control, 1991, 54(5): 1201-1216.

[52] WANG L X. Universal approximation by hierarchical fuzzy systems[J]. Fuzzy Sets and Systems, 1998, 93(2): 223-230.

[53] WANG L X. Analysis and design of hierarchical fuzzy systems[J]. IEEE Transactions on Fuzzy systems, 1999, 7(5): 617-624.

[54] ZENG X J, GOULERMAS J Y, LIATSIS P, et al. Hierarchical fuzzy systems for function approximation on discrete input spaces with application[J]. IEEE Transactions on Fuzzy Systems, 2008, 16(5): 1197-1215.

[55] YAGER R R. On a hierarchical structure for fuzzy modeling and control[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1993, 23(4): 1189-1197.

[56] RAJU G, ZHOU J. Adaptive hierarchical fuzzy controller[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1993, 23(4): 973-980.

[57] ZHAO X, YANG H, XIA W, et al. Adaptive fuzzy hierarchical sliding-mode control for a class of MIMO nonlinear time-delay systems with input saturation[J]. IEEE Transactions on Fuzzy Systems, 2016, 25(5): 1062-1077.

[58] JOO M G, LEE J S. Universal approximation by hierarchical fuzzy system with constraints on the fuzzy rule[J]. Fuzzy Sets and Systems, 2002, 130(2): 175-188.

[59] WOLPERT D H. Stacked generalization[J]. Neural Networks, 1992, 5(2): 241-259.

[60] GU S, CHUNG F L, WANG S. A novel deep fuzzy classifier by stacking adversarial interpretable TSK fuzzy sub-classifiers with smooth gradient information[J]. IEEE Transactions on Fuzzy Systems, 2020, 28(7): 1369-1382.

[61] QIN B, NOJIMA Y, ISHIBUCHI H, et al. Realizing deep high-order TSK fuzzy classifier by ensembling interpretable zero-order TSK fuzzy subclassifiers[J]. IEEE Transactions on Fuzzy Systems, 2021, 29(11): 3441-3455.

[62] ZHOU T, CHUNG F L, WANG S. Deep TSK fuzzy classifier with stacked generalization and triplely concise interpretability guarantee for large data[J]. IEEE Transactions on Fuzzy Systems, 2016, 25(5): 1207-1221.

[63] 陈德旺,蔡际杰,黄允浒. 面向可解释性人工智能与大数据的模糊系统发展展望[J]. 智能科学与技术学报, 2019, 1(4): 327-334.
CHEN Dewang, CAI Jijie, HUANG Yunhu. Development prospect of fuzzy system oriented to interpretable artificial intelligence and big data[J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(4): 327-334.

[64] 王素芬. 模糊系统与神经网络结合的现状[J]. 网络与信息, 2007, 21(5): 69-69.
WANG Sufen. Current status of combining fuzzy systems with neural networks[J]. Networks and Information, 2007, 21(5): 69-69.

[65] 张凯,钱锋,刘漫丹. 模糊神经网络技术综述[J]. 信息与控制, 2003, 32(5): 431-435.
ZHANG Kai, QIAN Feng, LIU Mandan. A survey on fuzzy neural network technology[J]. Information and Control, 2003, 32(5): 431-435.

[66] KUNCHEVA L I. How good are fuzzy if-then classifiers?[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2000, 30(4): 501-509.

[67] SONBOL A H, FADALI M S, JAFARZADEH S. TSK fuzzy function approximators: Design and accuracy analysis[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2012, 42(3): 702-712.

[68] GACTO M J, ALCALÁ R, HERRERA F. Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures[J]. Information Sciences, 2011, 181(20): 4340-4360.

[69] AGRAWAL R, SRIKANT R. Fast algorithms for mining association rules[C]//Proceedings of the 20th Int Conf Very Large Data Bases (VLDB). [S.l.]: [s.n.], 1994, 1215: 487-499.

[70] ISHIBUCHI H, YAMAMOTO T, NAKASHIMA T. Determination of rule weights of fuzzy association rules[C]//Proceedings of the 10th IEEE International Conference on Fuzzy Systems (Cat. No.01CH37297). [S.l.]: IEEE, 2001, 3: 1555-1558.

[71] ZHANG Y, ISHIBUCHI H, WANG S. Deep Takagi-Sugeno-Kang fuzzy classifier with shared linguistic fuzzy rules[J]. IEEE Transactions on Fuzzy Systems, 2017, 26(3): 1535-1549.

[72] WANG S, CHUNG F L, HONGBIN S, et al. Cascaded centralized TSK fuzzy system: Universal approximator and high interpretation[J]. Applied Soft Computing, 2005, 5(2): 131-145.

[73] MANTAS C J, PUCHE J M. Artificial neural networks are zero-order TSK fuzzy systems[J]. IEEE Transactions on Fuzzy Systems, 2008, 16(3): 630-643.

[74] ISHIBUCHI H, YAMAMOTO T. Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining[J]. Fuzzy Sets and Systems, 2004, 141(1): 59-88.

[75] ISHIBUCHI H, NOJIMA Y. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning[J]. International Journal of Approximate Reasoning, 2007, 44(1): 4-31.

[76] FENG S, CHEN C L P, XU L, et al. On the accuracy-complexity tradeoff of fuzzy broad learning system[J]. IEEE Transactions on Fuzzy Systems, 2021, 29(10): 2963-2974.

[77] SONBOL A H, FADALI M S, JAFARZADEH S. TSK fuzzy function approximators: Design and accuracy analysis[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2011, 42(3): 702-712.

[78] WANG Y, LIU H, JIA W, et al. Deep fuzzy rule-based classification system with improved wang-mendel method[J]. IEEE Transactions on Fuzzy Systems, 2022, 30(8): 2957-2970.

[79] GU X, CHUNG F L, ISHIBUCHI H, et al. Imbalanced TSK fuzzy classifier by cross-class Bayesian fuzzy clustering and imbalance learning[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2016, 47(8): 2005-2020.

[80] YAZDANBAKHSH O, DICK S. Forecasting of multivariate time series via complex fuzzy logic[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017, 47(8): 2160-2171.

[81] BEZDEK J C, EHRLICH R, FULL W. FCM: The fuzzy c-means clustering algorithm[J]. Computers & Geosciences, 1984, 10(2/3): 191-203.

[82] WANG Z, PAN X, WEI G, et al. A faster convergence and concise interpretability TSK fuzzy classifier deep-wide-based integrated learning[J]. Applied Soft Computing, 2019, 85: 105825.

[83] LECUN Y, BOSER B, DENKER J, et al. Handwritten digit recognition with a back-propagation network[J]. Advances in Neural Information Processing Systems, 1990, 2(2): 396-404.

[84] RUBIO J D J. SOFMLS: Online self-organizing fuzzy modified least-squares network[J]. IEEE Transactions on Fuzzy Systems, 2009, 17(6): 1296-1309.

[85] CHEUNG N J, DING X M, SHEN H B. OptiFel: A convergent heterogeneous particle swarm optimization algorithm for takagi-sugeno fuzzy modeling[J]. IEEE Transactions on Fuzzy Systems, 2014, 22(4): 919-933.

[86] WANG S, CHUNG F L, WU J, et al. Least learning machine and its experimental studies on regression capability[J]. Applied Soft Computing, 2014, 21: 677-684.

[87] WANG S, CHUNG F L. On least learning machine[J]. Journal of Jiangnan University (Natural Science Edition), 2010, 9: 505-510.

[88] WANG S, JIANG Y, CHUNG F L, et al. Feedforward kernel neural networks, generalized least learning machine, and its deep learning with application to image classification[J]. Applied Soft Computing, 2015, 37: 125-141.

[89] HUANG G B, ZHOU H, DING X, et al. Extreme learning machine for regression and multiclass classification[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2011, 42(2): 513-529.

[90] HOCHREITER S. The vanishing gradient problem during learning recurrent neural nets and problem solutions[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 1998, 6(2): 107-116.

[91] GU S, VONG C M, WONG P K, et al. Fast training of adversarial deep fuzzy classifier by downsizing fuzzy rules with gradient guided learning[J]. IEEE Transactions on Fuzzy Systems, 2022, 30(6): 1967-1980.

[92] ZHOU T, ISHIBUCHI H, WANG S. Stacked-structure-based hierarchical Takagi-Sugeno-Kang fuzzy classification through feature augmentation[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2017, 1(6): 421-436.

[93] WANG G, ZHOU T, CHOI K S, et al. A deep-ensemble-level-based interpretable Takagi-Sugeno-Kang fuzzy classifier for imbalanced data[J]. IEEE Transactions on Cybernetics, 2022, 52(5): 3805-3818.

[94] ZHOU Z, ZHANG Y, JIANG Y. Deep view-reduction TSK fuzzy system: A case study on epileptic EEG signals detection [C]//Proceedings of 2019 IEEE Symposium Series on Computational Intelligence (SSCI). [S.l.]: IEEE, 2019: 387-392.

[95] WANG L X. Fast training algorithms for deep convolutional fuzzy systems with application to stock index prediction[J]. IEEE Transactions on Fuzzy Systems, 2020, 28(7): 1301-1314.

[96] LÓPEZ V, FERNÁNDEZ A, GARCÍA S, et al. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics[J]. Information Sciences, 2013, 250: 113-141.

[97] LEE C H, CHANG F Y, LIN C M. An efficient interval type-2 fuzzy CMAC for chaos time-series prediction and synchronization[J]. IEEE Transactions on Cybernetics, 2013, 44(3): 329-341.

[98] WANG L X. Hierarchical fuzzy opinion networks: Top-down for social organizations and bottom-up for election[J]. IEEE Transactions on Fuzzy Systems, 2020, 28(7): 1265-1275.

[99] TIAN X, DENG Z, YING W, et al. Deep multi-view feature learning for EEG-based epileptic seizure detection[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2019, 27(10): 1962-1972.

[100] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2016: 770-778.

[101] MENDEL J M, BONISSONE P P. Critical thinking about explainable AI (XAI) for rule-based fuzzy systems[J]. IEEE Transactions on Fuzzy Systems, 2021, 29(12): 3579-3593.

[102] KROL M, FUHRMAN A, PAVONE L, et al. Fuzzy presentation of medical data[C]//Proceedings of the 14th IEEE Symposium on Computer-Based Medical Systems. [S.l.]: IEEE, 2001: 241-244.

[103] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017. DOI: 1048550/arxiv.1706.03762.

[104] DAI Z, YANG Z, YANG Y, et al. Transformer-XL: Attentive language models beyond a fixed-length context[EB/OL]. (2019-03-01)[2021-07-20]. http://arxiv.org/abs/1901.02860.

作者简介:

王士同(1964-),男,教授,博士生导师,研究方向:人工智能、模式识别、生物信息,E-mail:wxwangst@aliyun.com。

谢润山(1994-),**通信作者**,男,博士研究生,研究方向:模式识别,E-mail:runshan_xie@foxmail.com。

周尔昊(1995-),男,博士研究生,研究方向:模式识别,E-mail:erhaozhou@163.com。

(编辑:刘彦东)