

面向分类的 TSK 模糊遗忘学习方法^{*}

王 攀, 王士同

(江南大学 人工智能与计算机学院 (软件学院), 江苏 无锡 214122)

通信作者: 王士同, E-mail: wxwangst@aliyun.com



摘 要: 遗忘学习在隐私保护、减少污染数据影响和冗余数据处理等方面具有重要应用价值,但现有的遗忘学习方法多用于神经网络等黑箱模型中,在可解释的 TSK 模糊分类系统中实现高效的单类和多类遗忘仍面临挑战。为此,提出了一种面向分类的 TSK 模糊遗忘学习方法 (TSK-FUC)。首先,通过各规则的前件参数在 (单类或多类) 遗忘数据上的归一化激活强度,将规则库划分为与遗忘数据高相关的删减规则集、与遗忘数据低相关的保留规则集以及和遗忘数据和保留数据关系较为重叠的更新规则集。继而采取差异化处理策略:直接剔除删减规则集,以消除主要信息残留,并降低分类系统参数量;完整保存保留规则集,以缩小遗忘学习过程的参数调整范围;对于更新规则集,通过为每个遗忘类添加噪声,用以进一步消除规则中关于遗忘数据的信息,从而实现单类和多类遗忘。实验结果表明,在 16 个真实数据集的已建好的 0 阶和 1 阶 TSK 分类系统上,TSK-FUC 能够较为准确地划分规则空间,并结合差异化的处理展现出良好的单类和多类遗忘效果。该方法在保持规则库可解释性的同时,使得遗忘学习后的 TSK 模糊分类系统在结构上更加轻量化。

关键词: 遗忘学习; 模糊遗忘; Takagi-Sugeno-Kang (TSK) 模糊系统; 规则分类; 轻量化

中图法分类号: TP18

中文引用格式: 王攀, 王士同. 面向分类的 TSK 模糊遗忘学习方法. 软件学报, 2026, 37(9): 3648–3666. <http://www.jos.org.cn/1000-9825/7503.htm>

英文引用格式: Wang P, Wang ST. TSK Fuzzy Unlearning Method for Classification. Ruan Jian Xue Bao/Journal of Software, 2026, 37(9): 3648–3666 (in Chinese). <http://www.jos.org.cn/1000-9825/7503.htm>

TSK Fuzzy Unlearning Method for Classification

WANG Pan, WANG Shi-Tong

(School of Artificial Intelligence and Computer Science (School of Software), Jiangnan University, Wuxi 214122, China)

Abstract: Unlearning has significant application value in safeguarding privacy, mitigating the impact of contaminated samples, and processing redundant data. However, existing unlearning methods are mostly applied to black-box models such as neural networks, while achieving efficient single-class and multi-class unlearning in interpretable TSK fuzzy classification systems remains challenging. To address this, this study proposes a TSK fuzzy unlearning method for classification (TSK-FUC). First, the rule base is divided into three subsets using the normalized activation strengths of rule antecedent parameters on the (single/multi-class) forgotten data: 1) a deleted rule set that is highly relevant to the forgotten data, 2) a retained rule set with low relevance to the forgotten data, and 3) an updated rule set showing overlapping relevance to both the retained and forgotten data. Subsequently, differential processing strategies are applied: the deleted rule set is directly removed to eliminate major information residues and reduce the number of system parameters; the retained rule set is fully preserved to reduce parameter adjustment scope during unlearning; and for the updated rules, class-specific noise is added to the consequent parameters to further eliminate information related to the forgotten data, thus achieving single-class and multi-class unlearning. Experimental results on 16 benchmark datasets demonstrate that TSK-FUC accurately partitions the rule space and exhibits effective single-class and multi-class unlearning performance through differentiated processing in both 0-order and 1-order established TSK fuzzy

^{*} 基金项目: 国家重点研发计划 (2022YFE0112400)

收稿时间: 2025-04-14; 修改时间: 2025-06-02; 采用时间: 2025-06-27; jos 在线出版时间: 2025-11-26

CNKI 网络首发时间: 2025-11-27

classification systems. This method maintains the interpretability of the rule base while rendering the TSK fuzzy classification system more lightweight in terms of structure after unlearning.

Key words: unlearning; fuzzy unlearning; Takagi-Sugeno-Kang (TSK) fuzzy system; rule classification; lightweight

面向广泛存在的不确定性应用场景,基于模糊集合,模糊系统能够被用来模拟人类的知识推理能力.它通过IF-THEN规则形式(如IF温度高,THEN打开散热)来构成模糊规则库,因而能与人类自然语言所描述的推理规则高度相似,从而在可解释性和透明性方面展现出显著优势^[1],并且已被证明具有全局逼近能力^[2].特别是,作为一种模糊系统类型,TSK模糊分类系统^[3]由于其规则的后件通常采用常数参数或一阶线性函数,因而便于参数学习,并得到广泛的应用^[4,5].随着计算机网络和大数据的发展,各种模糊系统尤其是TSK模糊分类系统在实际场景中所需处理的数据量也在急剧增加.例如,在社会网络分析领域,社交媒体每日产生的信息量极为庞大,信息更新速度也非常快,许多过时的数据已不再适用.同时,这也容易引发在诸如社会网络分析、金融风险评估及医学诊断等对安全性和隐私性要求极高的应用场景中,个人信息泄露和隐私侵犯等社会问题.如果任由数据大量堆积,模糊系统将面临因维度诅咒而导致的规则爆炸问题.因此,如何让现有已建立的模糊系统能够高效且安全地遗忘过时和非法数据,并降低其系统的复杂度,成为一项涉及遗忘学习的很有前景的研究课题.

如Li等人^[6]提到,随着个人信息在各种人工智能模型中的广泛应用,虽然为人们的生活带来了便利,但也导致了个人信息泄露事件的频繁发生.随着相关法规的不断完善以及公众隐私保护意识的提升,从已建立的人工智能模型中删除部分信息,即遗忘学习或机器遗忘(machine unlearning)^[6-8],已成为可信人工智能领域的前沿研究方向^[9-13].遗忘学习的意义在于避免在重新训练模型上耗费大量时间和计算成本,同时使得遗忘模型对保留数据识别精度相对不变的情况下,能够“忘记”从遗忘数据中学习到的信息.进一步分析表明,现实场景中普遍存在针对某一类或多类的遗忘学习需求^[14-18].例如,在智能安防领域,现代人脸识别系统通过动态增量学习持续整合用户的多模态生物特征数据(如虹膜、步态等),但当用户退出系统时需确保模型完全遗忘其所有生物特征.每个用户的生物特征数据可以视为一个独立的数据类,当某个或多个用户共同提出删除个人信息的请求时,已建立的系统需要处理单类和多类的遗忘学习请求.然而,仅从数据库层面删除数据并不能满足隐私保护的要求^[8],因为模型参数中仍然保存着从这些数据中提取的信息.因此,如何安全高效地从已训练好的模型中遗忘一部分信息正是遗忘学习研究的核心内容.尽管现有研究已在深度神经网络领域取得显著进展^[9-18],但在具有显式规则表达能力的TSK模糊分类系统框架下,针对类别遗忘的研究仍然较少.因此,本文将重点关注如何在已建立的可解释的TSK模糊分类系统中实现单类和多类遗忘学习.

在处理TSK模糊分类系统的遗忘学习时,需要关注其与深度神经网络等黑箱模型之间的显著差异.首先,与黑箱模型在构建后结构固化难以调整的特性不同,TSK模糊分类系统通过规则库的灵活增删机制,能够便捷地实现结构优化.其次,黑箱模型中的参数在决策过程中的作用往往难以观察和理解,而TSK模糊系统则以其高可解释性而著称.如文献^[1,19]所述,如果通过一些基于聚类^[20,21]的方法构建模糊规则的前件参数,如模糊C均值聚类(fuzzy C means clustering, FCM)^[22],则能够保证这些规则前件的可解释性.与黑箱模型中全局参数更新的遗忘学习策略不同,这些可解释的前件参数在模糊分类系统的遗忘学习过程中将保持不变.结合上述特性,本文主要研究了,在处理具有可解释性优势的TSK模糊分类系统的分类遗忘问题时,所面临的3个挑战:①如何仅调整规则后件参数,从而在保持规则库可解释性的情况下,实现简单高效地单类和多类遗忘学习.②如何实现特定类别信息的定向擦除,并删减相关参数,从而降低分类系统的空间复杂度.③如何降低遗忘学习过程对现有规则库的影响.为此,本文提出了一种面向数据分类的基于规则空间划分的TSK模糊遗忘学习方法TSK-FUC,本文工作的主要贡献包括如下.

1) 提出了一种规则空间划分方法,该方法依据模糊规则的前件参数在遗忘数据上的归一化激活强度,能够较为准确地将规则分为与遗忘数据高相关的删减规则集、与遗忘数据低相关的保留规则集以及与遗忘数据和保留数据相关性较为重叠的更新规则集.

2) 提出了一种面向分类的TSK模糊遗忘学习方法,称为TSK-FUC.该方法采用先对规则空间进行划分,再分

别进行差异化处理的遗忘学习策略,该策略能够缩小遗忘学习过程中的参数调整范围,并主要通过删减与遗忘数据高度相关的规则,以及修正部分相关性较为重叠的规则的后件参数,实现单类和多类遗忘。

3) 实验结果表明,在保持现有规则库可解释性的前提下(即仅调整规则的后件参数),TSK-FUC 方法能够有效降低 0 阶和 1 阶 TSK 模糊分类系统中关于(单类和多类)遗忘数据的信息,并通过删减规则的方式降低分类系统的参数量,使得遗忘分类系统在结构的意义上更加轻量化。

1 相关工作

在涉及隐私数据、冗余数据及受污染数据等场景下,原始数据(记作 D)被划分为遗忘数据(记作 D_f)和保留数据(记作 D_r),且 $D_r \cup D_f = D$, $D_r \cap D_f = \emptyset$ 。遗忘学习的定义是,给定由全部数据 D 训练出的原始模型 M_{trained} ,通过一定的方法消除其中关于 D_f 的信息,同时保留模型中关于 D_r 的信息,从而获得遗忘模型 M_{unlearn} ^[6]。理想的遗忘学习方法得到的 M_{unlearn} 应该与从 $D_r (=D-D_f)$ 中重新训练所得的模型 M_{retrain} 相似,并且其对资源的消耗应远低于重新训练方法。若从准确率的角度衡量这种相似性, M_{unlearn} 在 D_r 上的分类准确率(记作 Acc^r)和在 D_f 上的准确率(记作 Acc^f)应接近 M_{retrain} 。通常情况下,也可以用 M_{trained} 衡量准确性^[8],即 M_{unlearn} 在 Acc^r 上的表现应接近 M_{trained} ,而 Acc^f 的表现应接近于 0 或低于 M_{trained} 随机猜测水平(即类别数的倒数)。目前的遗忘学习方法主要应用于神经网络中,通常通过从数据端或模型端调整模型参数,以减少模型参数中与 D_f 相关的信息。

最简单直接的遗忘学习方法是从样本中删除遗忘数据并重新训练模型(记为 retrain)。然而,重新建立模型所需的资源消耗非常庞大。Bourtoule 等人^[11]提出了 SISA 方法,该方法首先将训练数据分成多个子集并训练相应的子模型。当发出遗忘学习请求时,仅需找到包含遗忘数据的子集并重新训练对应的子模型。该方法降低了需要重新训练样本的数量。Bad-Teaching^[13]是一种基于知识传递^[23]技术的师生框架遗忘学习方法,该方法通过正面教师和负面教师两个模型,分别选择性地向学生模型传递与保留数据和遗忘数据相关的知识。Zero-shot^[12]方法分别根据代表 D_r 和 D_f 的误差,最小化和最大化噪声,利用知识传递模型在不使用任何训练样本的情况下实现遗忘学习。此外,还有许多通过梯度更新实现遗忘学习的方法。Du 等人^[24]提出一种基于梯度上升(反向训练)的方法,并通过 ReLU 函数和正则化项避免梯度上升过程中的损失爆炸问题。Ma 等人^[25]提出 Forsaken 方法,通过设计损失函数,最小化模型对遗忘样本与随机样本的输出分布。并以此损失计算梯度,用于更新 M_{trained} 以得到 M_{unlearn} 。Ameen 等人^[26]针对联邦学习^[27]框架下因用户端数据污染导致的模型污染问题,提出了一种基于遗忘学习机制的解决方案。该方法同样通过梯度上升策略逆向修正模型参数,以减轻污染数据对已训练模型的影响。

根据不同的场景,遗忘学习可以被划分为样本遗忘、类别遗忘和属性遗忘^[28]等常见类型。限于篇幅限制以及本文关注的单类和多类的类别遗忘场景,更多信息请参考综述文献^[6-8]。关于如何消除模型中一类或多类数据的信息,Golatkhar 等人^[10]提出了 Fisher 方法,该方法通过 KL 散度计算 M_{unlearn} 和 M_{retrain} 输出分布之间的差异,并通过设计特定的损失函数来缩小这一差异,从而实现遗忘学习。Baumhauer 等人^[15]提出了一种基于线性过滤的遗忘学习方法,将需要遗忘的类别的预测转移至其他类别。Tarun 等人^[14]使用误差最大化噪声(error-maximizing noise)来代表遗忘类数据对于模型参数的相反影响,并为每个遗忘类添加该噪声以实现单类和多类遗忘。Shibata 等人^[16]提出了一种在持续学习场景下的分类遗忘方法,该方法通过引入作用于遗忘样本的特定合成信号,称为“助记码”,在不访问遗忘类数据的情况下实现类别遗忘。Wang 等人^[17]提出了针对图像分类任务的联邦遗忘学习算法 FUCP,该算法依据卷积神经网络结构中通道对类别的评分进行剪枝,选择性地遗忘特定类别的贡献。Di 等人^[18]从对抗的角度评估遗忘学习算法的有效性,并设计了一个博弈论框架,利用先进的模型侵入攻击(MIA)来增强遗忘过程。

对于已经建好的可解释的模糊系统,如何处理其遗忘学习问题,近年来引起了学者们的关注。Lughofer 等人^[29]提出了在增量学习场景下,TSK 模糊分类系统的遗忘学习。然而,严格来说,该方法更倾向于经典的增量学习(或称为持续学习)。关于遗忘学习与增量学习之间的区别,详见文献^[6]中的相关讨论。Tang 等人^[30]提出一种基于模糊粗糙集理论的遗忘学习算法 ARU,该研究通过严谨的数学推理验证了 ARU 是一种精确的遗忘学习算法。虽然该方法首次将模糊集理论应用于遗忘学习中,但其主要关注于属性遗忘。Zagardo^[31]设计了一种实用的模糊遗忘方

法, 同时作者仅从实际实现遗忘学习的角度介绍了各种遗忘学习方法.

对于本文所关注的可解释的 TSK 模糊分类系统的遗忘学习问题, 我们认为应结合其所特有的性质. 首先, TSK 模糊分类系统的规则库具有高可解释性与决策透明性. 如前面所述, 在其遗忘学习过程中应当保持规则前件参数不变, 以维持规则的可解释性. 其次, 已建好的 TSK 模糊分类系统也可能包含从专家那里所获取的珍贵的专业知识及通过长期应用积累而得到的经验总结. 这些以模糊规则作为表达的知识在遗忘学习过程中需谨慎修改, 从而尽可能地降低对已有规则库的影响. 可是, 如果直接使用目前的面向黑箱模型的遗忘学习方法, 已建好的 TSK 模糊分类系统将不得不面对随机性、全局性的参数调整过程, 则不仅会损害模糊规则的可解释性, 也会破坏既有的规则库构成. 因此, 在已建好的 TSK 模糊分类系统中如何实现单类、多类遗忘, 以及在遗忘过程中如何保持规则的可解释性、降低对现有规则库的影响, 是一个非常值得研究的新课题.

2 基础知识

2.1 TSK 模糊分类系统

常见的模糊系统包括 Mamdani 模糊系统^[32]和 TSK 模糊系统^[3-5]等. 其中, Mamdani 模糊系统的规则后件为模糊集, 适用于人工经验驱动的控制场景; 而 TSK 模糊系统的规则后件通常为常数参数或更高阶函数, 因而既能表达专家经验知识, 又便于通过数据驱动来优化规则的参数. 特别是, 0 阶与 1 阶 TSK 模糊系统具有计算效率高、可解释性强的优势, 故具有更为广泛的应用. 本文聚焦于面向数据分类的 0 阶和 1 阶 TSK 模糊系统, 并简称之为 0 阶和 1 阶 TSK 模糊分类系统^[4,5,33,34], 然后研究其单类和多类遗忘学习问题. 在 0 阶和 1 阶 TSK 模糊分类系统中, 第 k 条规则可以表示为:

$$\begin{aligned} &\text{If } x_1 \text{ is } A_1^k \text{ and } x_2 \text{ is } A_2^k \text{ and } \dots \text{ and } x_d \text{ is } A_d^k \\ &\text{Then } \omega^{k,c}(x) = \rho_0^{k,c} + \rho_1^{k,c}x_1 + \dots + \rho_d^{k,c}x_d, \quad k = 1, 2, \dots, K; \quad c = 1, 2, \dots, C \end{aligned} \quad (1)$$

其中, x 为输入向量 $x = (x_1, x_2, \dots, x_d)^T$, d 表示输入向量的维度, K 表示规则数, A_d^k 是第 k 条规则在第 d 个特征上的前件模糊集, $\rho_0^{k,c}, \dots, \rho_d^{k,c}$ 是第 k 条规则的后件参数. 当 $\rho_1^{k,c}, \dots, \rho_d^{k,c}$ 全为 0, 即后件为常数参数形式: $\omega^{k,c}(x) = \rho_0^{k,c}$ 时, 此分类系统被称为 0 阶 TSK 模糊分类系统^[33]. 0 阶 TSK 模糊分类系统的模糊规则相对简单, 具有较高的可解释性. 当 $\rho_1^{k,c}, \dots, \rho_d^{k,c}$ 不全为 0, 即后件为线性函数形式时, 此分类系统被称为 1 阶 TSK 模糊分类系统^[34]. 1 阶 TSK 模糊分类系统的分类性能更强, 所需的规则数量更少, 但其规则的后件参数也相对较多. 经过相应的处理和去模糊化操作后, TSK 模糊分类系统在第 c 类上的预测分数可以表示为:

$$y^c(x) = \sum_{k=1}^K \bar{f}^k(x) \cdot \omega^{k,c}(x) \quad (2)$$

其中, $f^k(x)$ 表示第 k 条规则前件部分的激活强度, 可以用连乘公式 (3) 计算得到. $\bar{f}^k(x)$ 表示第 k 条规则前件部分的归一化激活强度.

$$f^k(x) = \prod_{i=1}^d u_{A_i^k}(x_i), \quad \bar{f}^k(x) = \frac{f^k(x)}{\sum_{k=1}^K f^k(x)} \quad (3)$$

其中, $u_{A_i^k}(x_i)$ 表示 x_i 在 A_i^k 上的隶属度, 通常采用高斯函数作为模糊隶属度函数, 可以表示为:

$$u_{A_i^k}(x_i) = \exp\left(-\frac{(x_i - \alpha_i^k)^2}{2(\delta_i^k)^2}\right) \quad (4)$$

其中, α_i^k 和 δ_i^k 分别表示第 k 条规则在第 i 个特征上的中心和标准差, 即模糊规则的前件部分. 如引言中论述, 为了保证这些前件参数的可解释性^[1,19], 可以采用聚类技术^[20,21]等方法进行处理. 本实验采用模糊 C 均值聚类算法 (FCM)^[22]来获得规则的前件参数, 用以模拟已有模型 M_{trained} 规则库中的前件部分. 该计算过程可以用公式 (5) 表示. 其中, $u_{n,k}$ 表示第 k 个簇中第 n 个样本 $x_n = (x_{n,1}, x_{n,2}, \dots, x_{n,d})^T$ 的模糊隶属度, h 为人为设定的尺度参数. 对于模糊规则的后件部分 (即公式 (1) 中的 $\rho_0^{k,c}, \dots, \rho_d^{k,c}$) 有多种求解方法, 本实验采用流行的梯度下降算法^[35]进行求解. 同时, 使用由所用数据集中的全部数据训练得到的 0 阶和 1 阶 TSK 模糊分类系统作为现有模型 M_{trained} .

$$\alpha_i^k = \frac{\sum_{n=1}^N u_{n,k} \cdot x_{n,i}}{\sum_{n=1}^N u_{n,k}}, \delta_i^k = \frac{h \sum_{n=1}^N u_{n,k} (x_{n,i} - \alpha_i^k)^2}{\sum_{n=1}^N u_{n,k}} \quad (5)$$

下面以 1 阶 TSK 模糊分类系统为例, 演示本实验中 M_{trained} 模型的训练过程. 假设分类系统 $(\mathbf{R}, \alpha, \delta, \rho)$ 是在包含 n 个训练样本、 d 个属性和 C 个类别的数据集 D 上训练得到 1 阶 TSK 模糊分类系统. 该分类系统的规则库 $\mathbf{R}$ 中包含 K 条规则. $\mathbf{R}$ 中的前件参数 $\alpha \in \mathbb{R}^{K \times d}$ 和 $\delta \in \mathbb{R}^{K \times d}$ (表示 d 个属性在 K 个聚类簇上的中心和标准差, 分别由公式 (5) 中的 α_i^k 和 δ_i^k 组成) 已通过 FCM 算法从 D 中聚类获得, 而其后件参数 $\rho \in \mathbb{R}^{K \times c \times (d+1)}$ (表示 K 条规则在 C 个类别上的 d 个参数和 1 个偏差, 由公式 (1) 中的 $\rho_0^{k,c}, \rho_1^{k,c}, \dots, \rho_d^{k,c}$ 组成) 则通过梯度下降法求解. 该优化过程采用交叉熵损失计算分类损失, 并使用主流的 Adam 优化器^[34] 更新后件参数. 该求解过程中的优化目标函数可以表示为:

$$\arg \min_{\rho} L = \text{Loss}(\mathbf{R}, \alpha, \delta, \rho, D) \quad (6)$$

其中, L 表示分类系统的预测分类损失, Loss 表示交叉熵损失函数. 假设在第 t 个优化轮次中, TSK 模糊分类系统对训练数据 D 中第 i 个样本 x_i 在全部 C 类上的预测分数为: $y_{t,i} = [y_{t,i}^1, y_{t,i}^2, \dots, y_{t,i}^C]$. 那么该优化轮次中的损失 L^t , 可以表示为公式 (7).

$$L^t = -\frac{1}{n} \sum_{i=1}^n \sum_{c=1}^C I_{t,i}^c \ln(q_{t,i}^c) \quad (7)$$

其中, n 表示训练样本数. $I_{t,i}^c$ 是指示函数, 当 x_i 的真实类别为 c 时其值为 1; 否则为 0. $q_{t,i}^c$ 表示分类系统在第 t 个优化轮次对 x_i 在第 c 类上的预测概率, 该概率由预测分数经过 *Softmax* 函数计算得到, 计算过程如公式 (8).

$$q_{t,i}^c = \text{Softmax}(y_{t,i}^c) = \frac{e^{y_{t,i}^c}}{\sum_{c=1}^C e^{y_{t,i}^c}} \quad (8)$$

根据 Adam 优化器^[36] 的参数更新原理, 假设规则库 $\mathbf{R}$ 的后件参数在第 $t-1$ 次优化轮次时的值为 ρ^{t-1} , 那么其在第 t 次优化轮次的更新过程可以表示为公式 (9):

$$\begin{cases} \mathbf{g}_t = \frac{\partial L^t}{\partial \rho^{t-1}}, \mathbf{w}_t = \beta_1' \mathbf{w}_{t-1} + (1 - \beta_1') \mathbf{g}_t, \hat{\mathbf{w}}_t = \frac{\mathbf{w}_t}{1 - \beta_1'} \\ \mathbf{v}_t = \beta_2' \mathbf{v}_{t-1} + (1 - \beta_2') \mathbf{g}_t^2, \hat{\mathbf{v}}_t = \frac{\mathbf{v}_t}{1 - \beta_2'} \\ \rho^t = \rho^{t-1} - \eta \hat{\mathbf{w}}_t \otimes (\sqrt{\hat{\mathbf{v}}_t} + \epsilon) \end{cases} \quad (9)$$

其中, 根据文献 [34] 中的论述, $\mathbf{g}_t$ 表示交叉熵损失 L^t 对被更新参数 ρ^{t-1} 的梯度, $\mathbf{w}_t$ 是梯度的一阶矩, β_1 是控制一阶动量的指数衰减率, 通常设置为 0.9^[36]. $\mathbf{v}_t$ 是梯度的二阶矩, β_2 是控制二阶动量的指数衰减率, 通常设置为 0.999^[36]. $\hat{\mathbf{w}}_t$ 和 $\hat{\mathbf{v}}_t$ 用作偏差校正. ϵ 是一个非常小的数 (例如 1E-8), 用于防止除以 0 的情况. t 是当前的迭代次数. $\otimes$ 表示两个向量中的每一个元素按第 i 个元素, 如 $\hat{\mathbf{w}}_{t,i} / (\sqrt{\hat{\mathbf{v}}_{t,i}} + \epsilon)$ 的方式, 计算结果向量.

3 面向分类的 TSK 模糊遗忘学习方法 TSK-FUC

本节将详细说明所提方法 (TSK-FUC) 中的规则空间划分方法及其差异化处理策略的理论依据和实施细节. 根据基础知识中对现有模型 M_{trained} 训练过程的描述, 本实验把从全部数据 $D (D_r \cup D_f)$ 中训练得到的 TSK 模糊分类系统作为现有分类系统 M_{trained} , 记作 $(\mathbf{R}, \alpha, \delta, \rho)$. 该规则库 $\mathbf{R}$ 的前件参数 α 和 δ 是通过 FCM^[22] 从 D 中聚类获得的, 用以模拟现有规则库 (如专家提供) 中规则的前件部分, 这些前件参数反映了样本属性数据中的某些规律. 后件参数 ρ 则通过梯度更新求解, 用于模拟现有规则库中规则的后件部分. 可以总结为, 现有 TSK 模糊分类系统的规则前件参数, 用于识别那些属性值符合特定规律 (如公式 (4) 中的高斯分布) 的样本, 而后件参数则用于逼近这些样本的真实标签. 因此, TSK 模糊分类系统的规则空间展现出明显的分类倾向性, 即每条规则倾向于识别符合其前件部分中特定规律的样本. 进一步分析可知, 这些样本可能属于单类, 也可能属于多类, 这取决于前件参数的准确性以及样本空间的重叠程度. 这种倾向性在 0 阶 TSK 模糊分类系统中表现更为明显, 因为其规则的后件参数为常数形式, 结构较为简洁. 而在 1 阶 TSK 模糊系统中, 为了增强单条规则的拟合能力, 引入了参数与样本属性之

间的线性耦合过程. 这一变化增加了单条规则的参数量, 从而提升其学习能力, 减少模型所需的规则数量. 然而, 这也使得规则与数据类之间的关系变得更加复杂, 增加了规则分类的难度.

根据规则库 R 中的分类倾向性, 可以将其中的规则概括为 3 类: ① 与 D_f 相关; ② 与 D_r 相关; ③ 与 D_r 和 D_f 都相关. 在遗忘学习过程中, 这 3 类规则应分别处理. 第 1 类规则应予以删减, 以降低模型中从 D_f 获取的主要信息, 同时, 能够降低分类系统的参数量; 第 2 类规则作为与 D_f 无关的知识应予以保留, 以缩小参数调整的范围; 第 3 类规则应予以修正, 以消除模型中关于 D_f 的信息, 并修正删减步骤中可能导致的关于 D_r 的信息损失. 因此, 可以先对 R 中的规则进行分类, 然后针对不同类型的规则进行针对性的处理. 理论上, 这种根据规则空间的分类倾向性对规则进行分类的方法与遗忘数据属于单类还是多类无关, 即这种方法能够适应单类和多类的遗忘学习场景.

3.1 TSK-FUC

基于上述理念, 本文提出了一种面向分类的 TSK 模糊遗忘学习方法, 称为 TSK-FUC. 该方法流程如图 1 所示, 其包含规则分类和遗忘学习两个步骤. $(R, \alpha, \delta, \rho)$ 表示从全部样本数据 $D (D_r \cup D_f)$ 中训练得到的 M_{trained} . 首先, 根据 R 中各规则在 D_f 中所有数据上的总归一化激活强度 (记作 $\bar{f}^k(D_f)$, 将在第 3.1.1 节中详细介绍), 将现有规则分为 3 类: 删减规则集 (记为 R_{del})、保留规则集 (记为 R_{fix}) 和更新规则集 (记为 R_{upd}), 对应的子分类系统分别为: $(R_{\text{del}}, \alpha_{\text{del}}, \delta_{\text{del}}, \rho_{\text{del}})$ 、 $(R_{\text{fix}}, \alpha_{\text{fix}}, \delta_{\text{fix}}, \rho_{\text{fix}})$ 和 $(R_{\text{upd}}, \alpha_{\text{upd}}, \delta_{\text{upd}}, \rho_{\text{upd}})$. R_{upd} 中的规则与 D_r 和 D_f 的相关性较为重叠, 这种情况在多类遗忘学习场景下更为突出. 接下来, 针对这 3 类规则集分别进行删减、保留和更新处理. 删减与 D_f 高度相关的规则集 R_{del} , 旨在删除模型中大部分与 D_f 高度相关的信息, 从而显著降低遗忘学习的难度. 同时, 能够降低模型参数量, 使得遗忘模型更加轻量化; 保留与 D_f 无关的规则集 R_{fix} , 旨在缩小参数调整范围; 对于更新规则集 R_{upd} , 本方法采用基于误差最大化噪声^[14]的遗忘学习方法更新其规则的后件参数, 误差最大化噪声代表了遗忘类数据对模型的反向影响. 通过为每个遗忘类添加噪声, 并利用这些噪声与 D_r 的子集 $D_{r-\text{sub}}$ 结合, 减少 R_{upd} 中关于 D_f 的信息, 同时强化关于 D_r 的信息, 从而实现单类和多类遗忘学习. $D_{r-\text{sub}}$ 的选择应根据具体遗忘场景确定合适的采样比例. 过小的采样比例可能因导致子集无法代表真实数据分布, 从而影响遗忘模型的精度; 而过大的采样比例则会降低方法的执行效率. 经过更新步骤得到的更新规则集用 R'_{upd} 表示, 对应的分类系统为 $(R'_{\text{upd}}, \alpha_{\text{upd}}, \delta_{\text{upd}}, \rho'_{\text{upd}})$. 合并 R_{fix} 与 R'_{upd} 中规则即可得到遗忘学习后的 TSK 模糊分类系统. 规则分类和遗忘学习是 TSK-FUC 的核心步骤, 接下来将详细介绍.

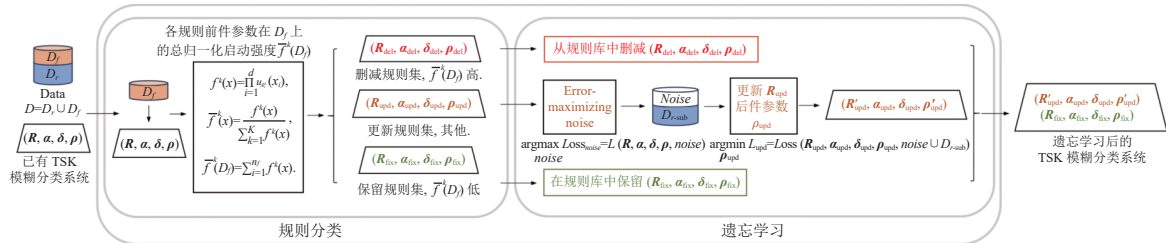


图 1 TSK-FUC 流程图

3.1.1 规则分类

根据 TSK 模糊分类系统中模糊规则的计算过程 (如公式 (1)–公式 (4) 所示), 模糊规则计算对样本数据的预测输出包括以下几个步骤: 首先, 通过规则前件参数 α 和 δ 计算所有规则 (记为 K 条), 在每个样本上的归一化激活强度, 如公式 (3), 公式 (4) 所示; 其次, 利用后件参数 ρ 、归一化激活强度和样本属性值聚合计算出规则的预测输出, 如公式 (1) 所示. 可以看出, 规则的归一化激活强度与其分类倾向性之间存在着显著关系, 因此可以根据各规则在 D_f 上的归一化激活强度对已有分类系统的规则空间进行划分. 对每条规则在所有遗忘样本上的归一化激活强度求和, 得到各规则的总归一化激活强度. 其中, 第 k 条规则的总归一化激活强度记作 $\bar{f}^k(D_f)$. 计算过程如公式 (10) 所示, 并以此作为规则与遗忘数据相关程度的依据来划分规则空间. 其中, n_f 表示遗忘数据中的样本数量, $\bar{f}^k(x_i)$ 是第 k 条规则在第 i 个样本上的归一化激活强度, 计算过程如公式 (3) 所示.

$$\bar{f}^k(D_f) = \sum_{i=1}^{n_f} \bar{f}^k(x_i) \quad (10)$$

显然, 在 D_f 上总归一化激活强度较高的规则表明, D_f 中大部分数据的属性值符合其前件参数中的规律 (如公式 (4) 中的高斯分布), 即这些规则与遗忘数据相关性也较高. 相反, 总归一化激活强度低的规则与遗忘数据的相关性也较低. 将规则库中的 K 条规则用 $R_1, R_2, \dots, R_K$ 表示, 计算这 K 条规则的总归一化激活强度, 完成排序后绘制箱线图, 同时设置两个排序后的分位数 Q_1 和 Q_2 作为分类阈值, 本文所提出的规则空间划分的条件如公式 (11) 所示. 以 $Q_1=0.9$ 、 $Q_2=0.5$ 为例, 总归一化强度前 10% ($0.9-1$) 的规则归为 R_{del} , 后 50% 的规则归为 R_{fix} , 剩余规则归为 R_{upd} . 显然, 考虑到规则空间的重叠性, 划分结果需满足以下约束条件: R_{del} 中的规则占比应小于规则库中与 D_f 相关的规则占比, R_{fix} 中规则占比应小于与 D_f 不相关的规则占比. 当缺乏数据集分布及各类别数据分类难易程度等先验知识时, 我们无法确切得到已有规则库中与 D_f 相关的规则占比. 按照经验, 我们可以取分类系统的类均规则占比 ($1/C$, C 为类别数) 作为调节 Q_1 和 Q_2 时的初始依据, 即 $Q_1 > 1-1/C$, $Q_2 < 1-1/C$. 当然, 它们的具体取值也可以进一步地综合考虑数据集特性、类别分布及 TSK 分类系统类型等因素来调整, 以便在维持良好遗忘准确率的前提下, 尽可能地缩小参数更新范围, 并通过删减更多的规则使得最终得到的 TSK 分类系统在结构上更加轻量化. 也就是说, 调节 Q_1 和 Q_2 在达到良好遗忘学习效果的同时, 使得 R_{upd} 中规则尽量少, R_{del} 和 R_{fix} 中规则尽量多.

$$\begin{cases} R_{\text{del}} = \{R_i \mid \bar{f}^i(D_f) > Q_1\} \\ R_{\text{fix}} = \{R_i \mid \bar{f}^i(D_f) \leq Q_2\} \\ R_{\text{upd}} = \{R_i \mid Q_2 < \bar{f}^i(D_f) \leq Q_1\} \end{cases}, i = 1, 2, \dots, K \quad (11)$$

下面以表 1 中 IS 数据集训练出的 1 阶 TSK 分类系统为例, 演示该分类系统在多类遗忘学习情况下, 遗忘第 5 和第 6 类样本时的规则分类的过程. 根据表 1, 该数据集共包含 7 类样本, 对应的 1 阶分类系统有 6 条规则. 该分类系统的整体准确率为 0.86. 在遗忘第 5 和第 6 类时的准确率分别为 Acc^r : 0.83, Acc^f : 0.95. 首先, 计算这 6 条规则在 D_f 上的总归一化激活强度和准确率, 结果如图 2 所示, 图中从左到右依次对应第 1 条 (R_1)–第 6 条 (R_6) 规则. 图 2 中 Act (activation strength) 表示总归一化激活强度, Acc^r 和 Acc^f 分别表示保留数据和遗忘数据准确率. 对图 2 中 Act 统计箱线图可得到图 3. 假设分别以图 3 中的 90% 分位数 Q_1 和 10% 分位数 Q_2 为阈值, 按照公式 (11) 进行规则空间划分. 则规则分类结果为: 删减规则集 $R_{\text{del}}=[R_1]$, 即图 2(a) 对应的规则. 更新规则集 $R_{\text{upd}}=[R_2, R_3, R_4, R_6]$, 即图 2(b)–(d) 和 (f) 对应的规则. 保留规则集 $R_{\text{fix}}=[R_5]$, 即图 2(e) 对应的规则. 接下来, 分别测试这 3 个子分类系统的准确率表现, 结果如下: ($R_{\text{del}}, \alpha_{\text{del}}, \delta_{\text{del}}, \rho_{\text{del}}$) 的准确率为 Acc^r : 0.04, Acc^f : 0.5; ($R_{\text{upd}}, \alpha_{\text{upd}}, \delta_{\text{upd}}, \rho_{\text{upd}}$) 的准确率为 Acc^r : 0.67, Acc^f : 0.45; ($R_{\text{fix}}, \alpha_{\text{fix}}, \delta_{\text{fix}}, \rho_{\text{fix}}$) 的准确率为 Acc^r : 0.55, Acc^f : 0. 由此可见, R_{del} 为主要与 D_f 相关的规则集, R_{fix} 为主要与 D_r 相关的规则集, R_{upd} 为与 D_r 、 D_f 都相关的规则集. 图 2 中数据也可以印证这一结论.

表 1 实验数据集及原始分类系统详细信息

Datasets	Number of samples	Number of features	Number of classes	Dr-sub	Zero-order TSK		First-order TSK	
					K	Acc	K	Acc
IS	210	19	7	1	26	0.8802	6	0.9286
glass	214	9	6	1	16	0.7209	8	0.7368
led	500	7	10	1	11	0.7534	5	0.7686
flare	1066	11	6	1	26	0.7150	6	0.7852
anemia	1281	14	9	1	60	0.7588	8	0.8876
yeast	1484	8	10	1	56	0.6062	6	0.6110
steel	1941	27	7	1	42	0.6922	6	0.7403
obesity	2111	16	7	1	80	0.7660	6	0.8014
segment	2310	19	7	1	39	0.9016	8	0.9588
handwritten	5620	64	10	0.5	29	0.8810	4	0.9199
satimage	6435	36	6	0.5	18	0.8327	6	0.8859
penbased	10992	16	10	0.2	70	0.9310	6	0.9408
nursery	12960	8	5	0.2	118	0.7112	7	0.8148

表1 实验数据集及原始分类系统详细信息(续)

Datasets	Number of samples	Number of features	Number of classes	Dr-sub	Zero-order TSK		First-order TSK	
					<i>K</i>	<i>Acc</i>	<i>K</i>	<i>Acc</i>
weather	13 200	10	4	0.2	19	0.883 7	7	0.901 9
drybean	13 611	16	7	0.2	35	0.900 8	4	0.914 4
letter	20 000	16	26	0.5	350	0.804 3	13	0.924 7

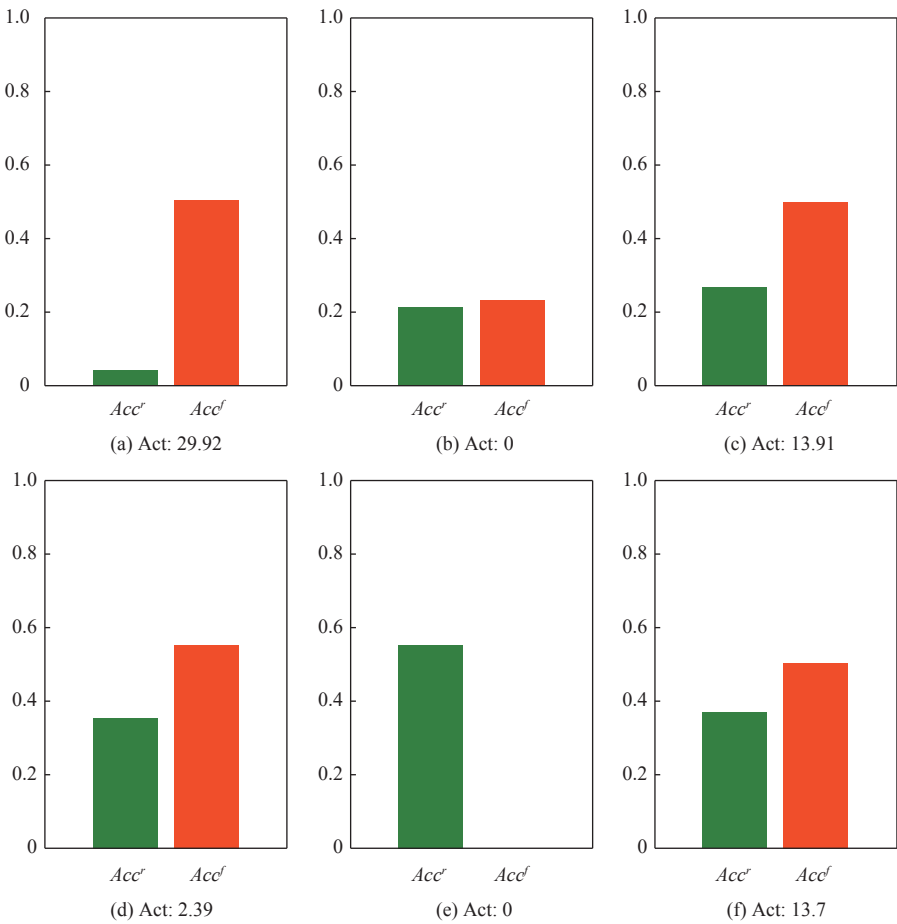


图2 各规则准确率和前件总归一化激活强度

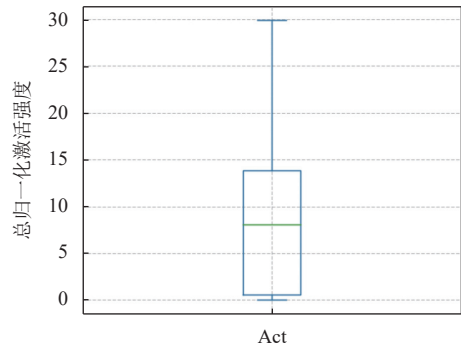


图3 总归一化激活强度箱线图

综上所述, 理论分析和实验结果均表明, 在单类和多类遗忘学习场景下, 本文提出的基于规则前件参数在 D_f 上的总归一化激活强度划分规则空间的方法, 在 0 阶和 1 阶 TSK 模糊分类系统上都能够较为准确地对规则进行分类. 并且, 该方法的分类过程原理简明、执行效率较高. 规则分类步骤是 TSK-FUC 的核心, 准确的规则空间划分能够为后续遗忘学习步骤奠定良好的基础. 规则空间划分的准确性与分类方式的有效性、规则空间本身的可分性 (重叠性) 以及样本空间的可分性 (重叠性) 密切相关. 一些分类系统由于本身规则库中规则数或每个数据类的平均 (规则数/类别数) 规则数较少, 如表 1 中的 handwritten、drybean 和 letter 数据集的 1 阶分类系统. 这导致规则与数据类别之间的关系更为复杂. 因此, 明确区分规则的难度较大, 且删减和保留规则集的选择会显著影响遗忘分类系统的准确率. 在这种情况下, 规则分类步骤中的 R_{del} 和 R_{fix} 都可能为空集 ($Q_1=1, Q_2=0$). 这时, TSK-FUC 将退化为与 UNSIR^[14]相似的方法, 即通过调整规则库中所有规则的后件参数实现遗忘学习.

3.1.2 遗忘学习

得益于规则分类步骤良好的分类准确性, 遗忘学习步骤主要对 3 类规则集进行差异化处理. 如图 1 所示, 其中 R_{del} 和 R_{fix} 中的规则处理相对简单, 仅需从 $M_{trained}$ 中删除 R_{del} 中规则的前后件参数, 并在后续优化过程中保持 R_{fix} 中规则的前后件参数不变. 接下来, 主要针对 R_{upd} 中规则继续进行遗忘学习, 并进一步增强 R_{upd} 中规则对 D_r 的学习表现. TSK-FUC 通过添加误差最大化噪声的方式来进一步消除 R_{upd} 中关于 D_f 的信息. 根据文献 [14], 误差最大化噪声的生成过程如下: 首先, 对于遗忘类集合中的每个遗忘类, 随机初始化一个符合标准正态分布的噪声 $noise$, 其大小为 $[b, d]$, b 为 $batch_size$ 的大小, d 为样本属性数量. 然后, 以 $M_{trained}((R, \alpha, \delta, \rho))$ 作为预测分类系统, 向分类损失最大化方向调整 $noise$ 的参数, 该优化目标函数可以表示为:

$$\arg \max_{noise} L_{noise} = Loss(R, \alpha, \delta, \rho, noise) + \lambda \|\omega\| \quad (12)$$

其中, L_{noise} 表示噪声生成过程中的预测分类损失, $Loss$ 表示交叉熵损失函数, ω 是噪声 $noise$ 的参数, $\lambda \|\omega\|$ 是用于平衡损失的正则化项. 假设在第 t 次优化轮次中, $M_{trained}$ 对 $noise$ 中第 j 个样本 x_j 在所有类上的预测输出为 $y_{t,j} = [y_{t,j}^1, y_{t,j}^2, \dots, y_{t,j}^c]$. 则在误差最大化训练过程中, 第 t 次优化轮次的损失 L_{noise}^t 可以表示为:

$$L_{noise}^t = \frac{1}{b} \sum_{i=1}^b \sum_{c=1}^C l_{t,j}^c \ln(q_{t,j}^c) + \lambda \|\omega\| \quad (13)$$

该优化过程沿着误差最大化的方向调整参数, 公式 (13) 与公式 (7) 的主要区别在于公式 (13) 添加了负号用于取反, 并且引入了正则化项. 公式 (13) 中 $l_{t,j}^c$ 为符号函数, 如果 x_j 的真实类别标签等于 c , 则其值为 1; 否则为 0. $q_{t,j}^c$ 表示在第 t 次轮次中, $M_{trained}$ 对 x_j 在 c 类别的预测概率, 由 $Softmax$ 函数计算得出, 计算过程与公式 (8) 相同. $noise$ 中的参数 ω 由 Adam 优化器^[36]进行更新, ω 在每个优化轮次的更新过程遵循公式 (9).

在得到所有遗忘类所对应的误差最大化噪声以后, 将其合并为 $Noise$, 再与 D_{r-sub} 合并为 $Noise \cup D_{r-sub}$. 以 $Noise \cup D_{r-sub}$ 为数据更新 R_{upd} 中的后件参数 ρ_{upd} . 如图 1 所示, 该参数更新过程与噪声生成过程中的优化目标方向相反, 即噪声生成过程旨在最大化预测分类损失 (L_{noise}), 而参数更新过程则力求最小化预测分类损失 (L_{upd}). 更新 ρ_{upd} 的过程类似公式 (6)–公式 (9), 主要区别为将训练数据替换为 $noise \cup D_{r-sub}$; 将被训练模型更换为 $(R_{upd}, \alpha_{upd}, \delta_{upd}, \rho_{upd})$, 并设置合适的学习率. 学习率过高和过低都影响遗忘分类系统的准确率, 在本文实验中, 遗忘学习步骤的学习率需针对不同数据集、遗忘数据类或其组合以及规则分类结果 (R_{upd} 中规则数) 进行调节, 取值范围为 0.01–0.9.

3.2 TSK-FUC 方法描述与计算复杂度分析

算法 1 给出 TSK-FUC 方法的遗忘学习过程.

算法 1. TSK-FUC 遗忘学习方法.

输入: 原始 TSK 分类系统 $M_{trained}((R, \alpha, \delta, \rho))$; 遗忘类别 $F=[F_0, F_1, \dots, F_f]$; 保留数据子集 D_{r-sub} ; 遗忘数据集 D_f ; 规则分类阈值 Q_1 和 Q_2 ; $batch_size$ b ; 数据维度 d ; 遗忘过程学习率 lr , 初始化噪音集合 $Noise$;

输出: 遗忘学习后的分类系统 $M_{unlearn}$.

1. 根据公式 (3)、公式 (4) 和公式 (10) 计算 R 中规则的前件参数在 D_f 上的总归一化激活强度, 并统计箱线图.

2. 根据规则分类阈值 Q_1 和 Q_2 , 按照公式 (11) 进行规则空间划分, 得到 R_{del} 、 R_{fix} 和 R_{upd} , 以及 $(R_{del}, \alpha_{del}, \delta_{del}, \rho_{del})$ 、 $(R_{fix}, \alpha_{fix}, \delta_{fix}, \rho_{fix})$ 和 $(R_{upd}, \alpha_{upd}, \delta_{upd}, \rho_{upd})$. //规则分类
3. For i in F : //误差最大化噪声生成
4. 随机生成符合标准正态分布, 大小为 $[b, d]$ 的噪声样本 $noise_{i,0}$.
5. 以 $(R, \alpha, \delta, \rho)$ 为预测分类系统, 根据公式 (13) 优化公式 (12), 得到噪音 $noise_i$, 并加入 $Noise$.
6. End For
7. 将 $Noise$ 与 D_{r-sub} 合并为 $Noise \cup D_{r-sub}$.
8. 以 $Noise \cup D_{r-sub}$ 为数据, 根据公式 (6) 和公式 (7), 优化 $(R_{upd}, \alpha_{upd}, \delta_{upd}, \rho_{upd})$ 后件参数 ρ_{upd} 两个回合得到 $(R'_{upd}, \alpha_{upd}, \delta_{upd}, \rho'_{upd})$. 优化过程学习率为 lr . //参数更新
9. 合并 R_{fix} 与 R'_{upd} 中的规则参数得到遗忘学习后的分类系统 $M_{unlearn}$.

分析 TSK-FUC 方法的计算复杂度, 根据算法 1 的描述, TSK-FUC 方法执行过程中的计算主要集中在步骤 1、5 和 8 中. 由于 0 阶 1 阶模糊规则的主要区别在后件参数部分, 因此, TSK-FUC 在这两种场景下的计算复杂度之区别主要在步骤 5 和 8 中涉及后件参数运算的部分. 步骤 1 涉及公式 (3)、公式 (4) 和公式 (10) 中的计算, 该过程的计算复杂度为 $O(n_f K d + n_f K) = O(n_f K d)$. 步骤 5 涉及计算 K 条规则在 b 个样本上的预测输出和梯度更新噪声参数, f 个遗忘类共需要进行 f 遍计算. 其中, 前件部分的计算过程与步骤 1 相同, 后件部分的聚合运算过程如公式 (2). 对于 0 阶 1 阶 TSK 分类系统, 该过程的计算复杂度分别为 $O(bKC)$ 和 $O(bKCd)$. 因此, 这两种分类系统对 K 条规则计算 b 个样本的预测输出 (前件部分+后件部分) 的计算复杂度分别为 $O(bK(C+d))$ 和 $O(bKCd + bKd) = O(bKCd)$. 梯度更新过程与需要更新的参数量和更新轮次相关, 根据步骤 4, 该过程需要更新的参数量为 $b \times d$. 假设每个遗忘类对应的噪声的平均优化轮次为 T_{noise} , 则这两种分类系统在步骤 5 的计算复杂度为 $O(fT_{noise}(bK(C+d) + bd)) = O(fT_{noise}bK(C+d))$ 和 $O(fT_{noise}(bKCd + bd)) = O(fT_{noise}bKCd)$; 步骤 8 涉及计算 K_{upd} 条规则在 $Noise \cup D_{r-sub}$ 上的预测输出以及梯度更新 K_{upd} 条规则的后件参数. 根据公式 (1) 可知, 0 阶 1 阶 TSK 分类系统需要更新的参数量分别为 $K_{upd} \times C$ 和 $K_{upd} \times C \times (d+1)$. 由算法 1 可知, 该过程优化轮次 $T_{upd}=2$, 假设 $Noise \cup D_{r-sub}$ 中样本数量为 n_{upd} , 则对于 0 阶 1 阶 TSK 分类系统, 步骤 8 的计算复杂度分别为 $O(2n_{upd}K_{upd}Cd + 2K_{upd}C) = O(2n_{upd}K_{upd}Cd)$ 和 $O(2n_{upd}K_{upd}Cd + 2K_{upd}C(d+1)) = O(2n_{upd}K_{upd}Cd)$. 综上所述, 对于 0 阶和 1 阶 TSK 模糊分类系统, TSK-FUC 的计算复杂度分别为 $O(n_f K d + fT_{noise}bK(C+d) + 2n_{upd}K_{upd}Cd)$ 和 $O(n_f K d + fT_{noise}bKCd + 2n_{upd}K_{upd}Cd)$. 此外, 需要说明的是, 尽管这里仅聚焦于 0 阶 1 阶 TSK 模糊分类系统. 对于高阶 TSK 模糊分类系统. 我们也可以类似地推导出相应的 TSK-FUC 的计算复杂度. 鉴于篇幅限制, 此处不再展开讨论.

综上所述, 该方法的时间复杂度计算主要与方法执行过程中所用的 D_f 样本数量 n_f 、 $Noise \cup D_{r-sub}$ 样本数量 n_{upd} 、 R_{upd} 规则数 K_{upd} 及优化过程中的执行轮次相关. n_{upd} 又主要与 D_{r-sub} 采样比例相关, 而 K_{upd} 又主要与规则分类时的分类阈值相关. 这些超参数共同影响遗忘学习后分类系统的准确率表现以及方法的执行效率, 因此在遗忘过程中应针对具体数据集和遗忘学习场景进行适当调节.

4 实验分析

4.1 实验数据

如表 1 所示, 本实验采用了 16 个不同数据量、属性数量和分类数量的多分类数据集, 以展示 TSK-FUC 的单类和多类遗忘性能. 这些数据集均来自 UCI^[37]、KEEL^[38]和 Kaggle^[39]公开数据集平台. 表 1 详细列出了各数据集的相关信息, 以及在这些数据集上由全部样本训练出的 0 阶和 1 阶 TSK 原始分类系统的详细信息, 这些模型将作为已建好的 $M_{trained}$ 用于后续的遗忘学习实验. 表 1 中的 D_{r-sub} 表示从 D_r 随机采样得到子集 D_{r-sub} 的采样比例, K 表示模型规则数, Acc 表示模型的准确率. Zero-order TSK 和 First-order TSK 分别表示 0 阶、1 阶 TSK 模糊分类系统.

4.2 对比方法及评价指标

本实验选取了重新训练方法作为参考,并选取两种新近提出的类别遗忘方法 Bad-Teaching^[13]和 UNSIR^[14]作为对比方法,以综合对比 TSK-FUC 的遗忘学习性能.接下来将进行详细介绍.

1) Retrain. 保持原有模型结构不变,从样本中删除遗忘数据后重新训练模型.该方法的遗忘效果最为理想,但其资源消耗也很大.

2) Bad-Teaching^[13]. 基于知识传递的师生框架模型,通过正面教师和负面教师分别向学生模型传递与保留数据和遗忘数据相关的知识,该方法具有较好的泛化能力.

3) UNSIR^[14]. 基于反向误差噪声,通过修正现有模型参数实现单类和多类遗忘学习.该方法比较高效,且无需访问遗忘数据.

根据文献[6-8],本实验主要从遗忘学习可用性^[6]、遗忘学习速度^[6]和遗忘学习完成度^[6]这3个方面评估所提方法的遗忘学习性能.因此,在本实验中,我们采用常用的评价指标^[6-8],包括准确率、执行时间和重新学习时间(记作 RT, relearn time),这些指标广泛用于评估先前的遗忘学习方法^[10-14].接下来将详细介绍.

1) 准确率,用于评估遗忘学习可用性.包括保留样本集 D_r 上的识别准确率,记为 Acc^r ,以及遗忘样本集 D_f 上的识别准确率,记为 Acc^f . Acc^r 和 Acc^f 分别表示模型对 D_r 和 D_f 中样本做出正确分类预测的比例.一个好的遗忘模型 M_{unlearn} 的 Acc^r 表现应接近原始模型 M_{trained} ,而 Acc^f 应接近 0 或低于模型随机猜测的准确率(类别数的倒数).

2) 执行时间,用于评估遗忘学习速度.在相近的遗忘性能下,所需时间越少,遗忘效率越高.

3) RT, 用于评估遗忘学习完成度.该指标的定义为: M_{unlearn} 重新学习 D_f 至 Acc^f 达到 M_{trained} 对应的精度所需的学习轮次.该指标用于衡量 M_{unlearn} 中关于遗忘数据信息的残留量, M_{unlearn} 中残留的关于 D_f 的信息越多,RT 值越小;相反,RT 值越高,说明 M_{unlearn} 的遗忘学习完成度越高.

4.3 实验环境,参数设置和实验方法

本实验中的对比实验程序在配备了 Windows 11 操作系统, Intel i9-12900H 2.50 GHz 处理器, 32 GB 内存和 NVIDIA 3060 laptop 显卡的笔记本电脑上执行.实验中使用的程序编辑器为 PyCharm 2023.2.1.对比实验中的超参数设置如下: batch-size 为 256;对于 M_{trained} 的训练过程,最大优化轮次为 400,提前停止条件为 20,学习率范围为 0.001-0.2,FCM 聚类的模糊因子范围为 1.1-2.5;对于 Retrain 方法,学习率范围为 0.001-0.2,FCM 聚类的模糊因子范围为 1.1-2.5;对于 RT 评价指标中的再学习过程,所有对比方法使用相同的参数设置:训练数据为全部遗忘数据 D_f ,学习率为 0.01,最大学习轮次为 100;对于 TSK-FUC 方法, $D_{r-\text{sub}}$ 的采样比例同样如表 1 所示,规则分类阈值 Q_1 和 Q_2 的取值范围为 [0, 1],遗忘学习步骤的学习率范围为 0.01-0.9;对于 UNSIR 方法,损伤步骤的学习率范围为 0.001-0.9,修复步骤的学习率范围为 0.01-0.9, $D_{r-\text{sub}}$ 采样比例同 TSK-FUC 方法,如表 1 所示;对于 Bad-Teaching 方法,学习率范围为 0.001-0.9.

在实验方案设置上,我们通过比较各数据集中不同类别的平均遗忘学习表现,来评估不同方法的遗忘学习性能.具体而言,在单类遗忘学习实验中,对于类别数低于 10 的数据集,对其每一类数据依次执行 10 次遗忘实验,并对结果取平均值,以作为每一类别的遗忘学习表现;而对于类别数超过 10 的数据集,均匀选取 10 类进行上述实验.随后,计算所有类别的平均表现,以此作为该数据集上的平均遗忘学习表现.在多类遗忘学习场景中,我们每次遗忘两个类别.根据条件 $\{<c_i, c_j> \mid i, j=1, 2, \dots, C \text{ and } i < j\}$ 在数据类中均匀的选取遗忘类组合.对于组合数少于 10 的数据集,选择其最大组合数;而对于组合数量超过 10 的,则均匀选取 10 种组合,以覆盖数据集中大多数类别.每种组合的遗忘实验进行 10 次,取平均值作为每组遗忘学习实验的结果,再对所有组合的结果取平均,作为各数据集的遗忘学习平均表现.在每组实验中,调节所有方法中的超参数,以实现最佳的遗忘学习效果.需要注意的是,因数据分布不均衡或其他原因导致的 M_{trained} 在某些类别上精度较差的情况将不予考虑.也就是说,如果模型本身未能从某类中学习得到有效信息,则无需遗忘学习.

4.4 实验结果与分析

4.4.1 单类遗忘学习

根据实验方案,我们首先在 16 个数据集的 0 阶和 1 阶 TSK 模糊分类系统上进行了单类遗忘学习的对比实

验,并统计平均实验结果.准确率的对比结果见表2,TSK-FUC方法对于单类遗忘学习后的规则分类结果见表3,执行时间对比结果见表4,RT对比结果见表5.各表中的粗体字表示对比方法中的最佳结果,各表中Zero-order TSK和First-order TSK分别表示0阶、1阶TSK模糊分类系统.

表2 单类遗忘学习后的准确率对比 ($Acc \pm Std$ 和 $Acc' \pm Std$)

Type	Datasets	Trained	Retrain	Bad-Teaching	UNSIR	TSK-FUC
Zero-order TSK	IS	0.8238±0	0.8727±0.0075	0.6773±0.0070	0.8444±0	0.8421±0
		0.8238±0	0±0	0.0800±0.0125	0±0	0±0
	glass	0.6885±0	0.7521±0.0035	0.5427±0.0087	0.7345±0.0023	0.7388±0.0004
		0.7136±0	0±0	0.2166±0.0199	0.0158±0	0.0026±0
	led	0.7400±0	0.7351±0.0012	0.7025±0.0334	0.7505±0.0004	0.7480±0
		0.7422±0	0±0	0.0891±0.0847	0.0019±0.0006	0±0
	flare	0.6754±0	0.7391±0.0008	0.6002±0.0263	0.7127±0.0076	0.7137±0
		0.7715±0	0±0	0.2342±0.1212	0±0	0±0
	anemia	0.7562±0	0.8004±0.0020	0.7310±0.0016	0.7770±0.0035	0.7600±0.0007
		0.7334±0	0±0	0.3052±0.0141	0.0625±0.0066	0.0288±0.0016
	yeast	0.5939±0	0.6539±0.0003	0.5935±0.0143	0.6385±0.0025	0.6366±0.0001
		0.6042±0	0±0	0.1762±0.0691	0.0009±0	0.0047±0
	steel	0.6814±0	0.7180±0	0.6900±0.0020	0.7211±0.0023	0.7182±0
		0.7046±0	0±0	0.3116±0.0111	0.0042±0.0057	0±0
	obesity	0.7398±0	0.7849±0.0007	0.7016±0.0061	0.7754±0.0027	0.7593±0
		0.7340±0	0±0	0.1502±0.0238	0.1383±0.0098	0.0144±0
	segment	0.9000±0	0.9086±0.0018	0.8958±0.0016	0.9089±0.0056	0.9068±0
		0.9000±0	0±0	0.0242±0.0086	0.3889±0.0242	0.0177±0
	handwritten	0.8811±0	0.8896±0.0005	0.8866±0	0.8897±0.0006	0.8883±0
		0.8814±0	0±0	0.0305±0	0.2967±0.0004	0.0259±0
	satimage	0.8327±0	0.8519±0.0001	0.8440±0.0011	0.8531±0.0046	0.8576±0
		0.7878±0	0±0	0.1591±0.0006	0.0806±0.0003	0.0012±0
First-order TSK	penbased	0.9327±0	0.9622±0.0008	0.8962±0.0063	0.9461±0.0008	0.9349±0
		0.9334±0	0±0	0.1170±0.0474	0.0702±0.0060	0.0030±0
	nursery	0.6874±0	0.8586±0.0004	0.6368±0.0263	0.8093±0.0011	0.7876±0.0001
		0.7102±0	0±0	0.2545±0.0790	0.0289±0.0002	0±0
	weather	0.8771±0	0.8963±0.0001	0.8808±0.0015	0.9017±0.0009	0.8848±0
		0.8771±0	0±0	0.2650±0.1412	0±0	0.0028±0
	drybean	0.8970±0	0.9159±0	0.9009±0.0007	0.9157±0.0003	0.9110±0
		0.9052±0	0±0	0.4297±0.0556	0.0172±0.0005	0±0
	letter	0.8033±0	0.8225±0	0.7890±0.0011	0.7908±0.0019	0.8018±0
		0.8133±0	0.0010±0.0001	0.3600±0.0009	0.4148±0.0003	0.0597±0.0001
	IS	0.8619±0	0.9768±0.0064	0.7894±0.0128	0.9286±0	0.9254±0
		0.8619±0	0±0	0.1043±0.0163	0.0238±0	0.0190±0
	glass	0.7293±0	0.8076±0.0129	0.6179±0.0163	0.7864±0.0129	0.7662±0.0087
		0.8297±0	0.0159±0.0062	0.2169±0.0463	0.0281±0.0093	0.0636±0.0247
	led	0.7680±0	0.8017±0.0045	0.7405±0.0078	0.7847±0.0047	0.7790±0.0033
		0.7692±0	0±0	0.0410±0.0290	0.1020±0.0540	0.0039±0
	flare	0.7494±0	0.8071±0.0007	0.7233±0.0069	0.7828±0.0080	0.7486±0.0052
		0.8463±0	0±0	0.2421±0.0260	0.1071±0.0025	0.0150±0.0118
	anemia	0.8691±0	0.9289±0.0037	0.8443±0.0049	0.8738±0.0061	0.8690±0.0034
		0.8076±0	0±0	0.2168±0.0299	0.0316±0.0207	0.0515±0.0148
	yeast	0.6045±0	0.6634±0.0009	0.6063±0.0076	0.6356±0.0052	0.6343±0.0018
		0.6083±0	0.0002±0.0003	0.2505±0.0856	0.0018±0.0012	0.0036±0.0012
	steel	0.7362±0	0.8249±0.0026	0.6886±0.0040	0.7702±0.0050	0.7609±0.0014
		0.7238±0	0.0003±0.0002	0.1673±0.0126	0.0004±0.0003	0.0035±0.0001
	obesity	0.7863±0	0.9703±0.0020	0.7358±0.0096	0.8266±0.0073	0.8162±0.0026
		0.7833±0	0±0	0.1614±0.0280	0.0128±0.0057	0.0070±0.0023

表 2 单类遗忘学习后的准确率对比 ($Acc^r \pm Std$ 和 $Acc^d \pm Std$)(续)

Type	Datasets	Trained	Retrain	Bad-Teaching	UNSIR	TSK-FUC
First-order TSK	segment	0.9584±0	0.9806±0.0021	0.9444±0.0023	0.9594±0.0043	0.9548±0
		0.9584±0	0±0	0.0283±0.0054	0.0333±0.0224	0.0134±0
	handwritten	0.9237±0	0.9523±0.0016	0.9165±0	0.9356±0.0018	0.9321±0
		0.9239±0	0.0050±0.0053	0.0927±0	0.0139±0.0027	0.0066±0
	satimage	0.8847±0	0.9027±0.0020	0.8523±0.0080	0.8910±0.0077	0.8864±0.0019
		0.8638±0	0.0383±0.0095	0.3240±0.0874	0.0896±0.0574	0.0083±0.0005
	penbased	0.9460±0	0.9929±0.0001	0.9412±0.0016	0.9700±0.0029	0.9691±0.0011
		0.9462±0	0±0	0.0137±0.0052	0.0027±0.0034	0.0001±0.0002
	nursery	0.8002±0	0.8671±0.0003	0.7419±0.0165	0.8799±0.0026	0.8829±0.0006
		0.8296±0	0±0	0.2908±0.0670	0.0277±0.0043	0±0
	weather	0.8877±0	0.9238±0.0001	0.8788±0.0065	0.9105±0.0027	0.9056±0.0012
		0.8877±0	0.0014±0.0003	0.1246±0.1088	0.0178±0.0132	0.0026±0.0007
	drybean	0.9152±0	0.9367±0.0005	0.9158±0.0016	0.9257±0.0028	0.9272±0.0009
		0.9241±0	0±0	0.0841±0.0066	0.1438±0.0004	0.0080±0.0007
	letter	0.9247±0	0.8716±0	0.8820±0.0020	0.9173±0.0008	0.9046±0.0002
		0.9280±0	0.0103±0	0.1375±0.0122	0.0141±0.0061	0.0738±0.0320

注: 每组上下两个数值分别对应 $Acc^r \pm Std$ 和 $Acc^d \pm Std$

表 3 TSK-FUC 方法对于单类遗忘学习后的规则分类结果

Datasets	D_f/D	Zero-order TSK								D_f/D	First-order TSK							
	(%)	K	K_{del}	K_{del}/K (%)	K_{fix}	K_{fix}/K (%)	K_{upd}	K_{upd}/K (%)	(%)	K	K_{del}	K_{del}/K (%)	K_{fix}	K_{fix}/K (%)	K_{upd}	K_{upd}/K (%)		
IS	14.3	26	2.6	9.9	20.9	80.2	2.6	9.9	14.3	6	0.9	14.3	0.9	14.3	4.3	71.4		
glass	18.4	16	2.4	15.0	8.4	52.5	5.2	32.5	18.4	8	1.0	12.5	2.0	25.0	5.0	62.5		
led	10.0	11	1.0	9.1	9.9	90.0	0.1	0.9	10.0	5	0.4	8.0	2.1	42.0	2.5	50.0		
flare	21.8	26	4.5	17.3	8.0	30.8	13.5	51.9	21.8	6	1.2	20.8	1.0	16.7	3.8	62.5		
anemia	12.3	60	6.1	10.2	21.6	36.0	32.2	53.8	11.1	8	0.8	9.7	1.0	12.5	6.2	77.8		
yeast	13.8	56	5.6	9.9	42.0	75.0	8.4	15.1	12.2	6	0.6	10.4	2.2	37.5	3.1	52.1		
steel	15.3	42	4.3	10.3	28.3	67.5	9.3	22.2	14.3	6	0.6	9.5	1.1	19.0	4.3	71.4		
obesity	14.3	80	8.4	10.5	55.6	69.5	16.0	20.0	14.3	6	0.7	11.9	1.1	19.0	4.1	69.0		
segment	14.3	39	3.6	9.2	25.9	66.3	9.6	24.5	14.3	8	1.0	12.5	1.9	23.2	5.1	64.3		
handwritten	10.0	29	2.6	9.0	25.7	88.6	0.7	2.4	10.0	4	0.3	7.5	1.0	25.0	2.7	67.5		
satimage	16.7	18	2.3	13.0	13.2	73.1	2.5	13.9	16.7	6	0.8	13.9	1.8	30.6	3.3	55.6		
penbased	10.0	70	5.6	8.0	62.7	89.6	1.7	2.4	10.0	6	0.5	8.3	0.6	10.0	4.9	81.7		
nursery	32.5	118	28.0	23.7	24.0	20.3	66.0	55.9	32.5	7	1.7	23.8	0.3	4.8	5.0	71.4		
weather	25.0	19	3.5	18.4	12.5	65.8	3.0	15.8	25.0	7	1.5	21.4	1.0	14.3	4.5	64.3		
drybean	14.3	35	4.9	13.9	28.7	82.0	1.4	4.1	14.3	4	0.4	10.7	2.1	53.6	1.4	35.7		
letter	3.9	350	12.2	3.5	323.5	92.4	14.3	4.1	3.9	13	0.4	3.1	3.2	24.6	9.4	72.3		

表 4 单类遗忘学习的执行时间对比 (s)

Datasets	Zero-order TSK				First-order TSK			
	Retrain	Bad-Teaching	UNSIR	TSK-FUC	Retrain	Bad-Teaching	UNSIR	TSK-FUC
IS	0.3848	0.0080	0.0817	0.0707	0.6340	0.0089	0.1015	0.1120
glass	0.3571	0.0073	0.0703	0.0670	0.7426	0.0092	0.0992	0.1087
led	0.6617	0.0176	0.0847	0.0251	1.9485	0.0204	0.1170	0.1289
flare	1.3406	0.0394	0.1095	0.1083	1.5804	0.0472	0.1364	0.1587
anemia	11.7507	0.0522	0.1492	0.1366	6.7411	0.0551	0.1349	0.1564
yeast	1.7248	0.0546	0.1274	0.1215	1.5770	0.0587	0.1324	0.1620
steel	2.3743	0.0602	0.1367	0.1191	9.4387	0.0792	0.1755	0.1925
obesity	3.7942	0.0685	0.1346	0.1483	18.0499	0.0862	0.1721	0.1783
segment	5.1611	0.0746	0.1691	0.1700	8.0638	0.0955	0.1831	0.1900

表 4 单类遗忘学习的执行时间对比 (s)(续)

Datasets	Zero-order TSK				First-order TSK			
	Retrain	Bad-Teaching	UNSIR	TSK-FUC	Retrain	Bad-Teaching	UNSIR	TSK-FUC
handwritten	17.998 7	0.099 9	0.165 5	0.135 1	21.244 9	0.119 4	0.187 8	0.286 7
satimage	6.876 0	0.124 4	0.186 1	0.230 0	15.625 3	0.149 2	0.195 3	0.310 4
penbased	76.810 5	0.101 1	0.153 1	0.269 0	17.748 1	0.129 4	0.158 7	0.365 0
nursery	79.794 8	0.189 7	0.130 8	0.359 0	17.569 2	0.240 7	0.161 1	0.420 1
weather	21.947 3	0.461 9	0.357 9	0.473 7	17.727 7	0.216 2	0.177 0	0.419 6
drybean	24.161 4	0.486 4	0.408 3	0.300 3	52.260 6	0.177 6	0.185 8	0.408 2
letter	289.494 7	0.329 7	0.369 6	0.679 5	228.833 0	0.759 7	0.638 4	1.066 0

表 5 单类遗忘学习的基于 Relearn Time (学习轮次) 度量的对比

Datasets	Zero-order TSK				First-order TSK			
	Retrain	Bad-Teaching	UNSIR	TSK-FUC	Retrain	Bad-Teaching	UNSIR	TSK-FUC
IS	>100	27.71	59.86	29.14	>100	7.34	13.14	29.71
glass	>100	19.08	44.1	48.26	40	9.60	28.66	17.24
led	>100	6.07	18.79	51.69	42.31	11.93	24.82	31.65
flare	>100	4.95	69.48	45.77	>100	16.82	45.48	82.33
anemia	>100	48.48	> 100	> 100	>100	29.50	68.93	45.37
yeast	>100	7.60	86.76	34.41	>100	8.44	38.18	44.96
steel	>100	34.58	> 100	> 100	81.49	4.96	27.77	18.61
obesity	>100	15.89	> 100	31.01	>100	8.86	46.49	56.10
segment	>100	87.99	> 100	> 100	>100	35.03	79.61	98.71
handwritten	>100	10.80	> 100	> 100	>100	3	47.26	79.10
satimage	>100	32.10	> 100	> 100	>100	4.87	18.80	30.40
penbased	>100	8.10	84.98	27.58	69.52	3.80	26.38	19.12
nursery	68.13	1	21.93	18.33	24.93	1.40	10.27	10.27
weather	44.90	1.50	35.5	20.8	18.95	1.50	5.75	11.60
drybean	>100	3.17	> 100	49.57	48.57	2.66	13.89	38.69
letter	>100	84.35	> 100	> 100	93.35	9.35	32.7	32.25

由表 2 可知,在大多数数据集上,TSK-FUC 方法在 0 阶和 1 阶 TSK 模糊分类系统上的准确率表现与 UNSIR^[14] 相当,达到了相似的单类遗忘学习效果,即 Acc^r 接近 M_{trained} , Acc^f 接近 0. UNSIR 方法在多数数据集上的准确率表现略有优势,但对于一些数据集(如 handwritten 和 letter 等)的 0 阶 TSK 分类系统,其中许多类别的准确率较高(接近于 1 或高于平均准确率),即这些类较易于被分类系统准确识别.这种情况下基于反向噪声的 UNSIR 方法在仅更新规则后件的情况下,难以完全抹去规则库中关于这些类的信息.如表 2 中数据所示,UNSIR 方法在这些数据集上的 Acc^f 仍有 0.2 以上.而我们提出的基于规则分类的方法在这种情况下则能够更准确地进行分类,往往仅需删除与 D_f 相关的规则即可实现遗忘学习,即更新规则的比例 (K_{upd}/K) 很低,这在表 3 中(上述提到的两数据集 K_{upd}/K 分别为 2.4% 和 4.1%)也能得到印证.与 Bad-Teaching^[13] 方法相比,TSK-FUC 方法在大多数数据集上的准确率表现略有优势.

为了展现 TSK-FUC 中规则空间划分方法及其差异化处理策略的有效性,本实验统计了 TSK-FUC 方法对于单类遗忘学习后的规则分类结果.结果见表 3,其中, D_f/D 表示遗忘数据比例, K 、 K_{del} 、 K_{fix} 和 K_{upd} 分别表示原始分类系统规则数、删减规则集规则数、保留规则集规则数和更新规则集规则数; K_{del}/K 、 K_{fix}/K 和 K_{upd}/K 分别表示删减规则、保留规则和更新规则的占比.根据表 3 中的数据,对于大部分数据集的 0 阶和 1 阶 TSK 分类系统,TSK-FUC 方法在单类遗忘过程中 K_{del}/K 的值也都小于或接近数据集的类均规则占比 ($1/C$),说明 TSK-FUC 在实现良好的遗忘性能的同时,较大程度地降低了分类系统的参数量,使得遗忘分类系统在结构上更加轻量化.对于表 3 中大多数数据集的 0 阶 TSK 模糊分类系统,当 D_f/D 低于 20% 时,TSK-FUC 方法保留了超过 60% 的规则不变.这些数据集对应的 K_{upd}/K 结果也都较低,这一结果也说明 TSK-FUC 的遗忘学习过程较大程度地缩小了参数

更新的范围. 而对于 1 阶 TSK 分类系统, 由于 M_{trained} 本身规则数较少, 规则与数据类别之间的关系更加复杂, 图 2 中的结果也能印证这一观点. 因此, 更加难以对这些分类系统的规则库进行准确分类. 如表 3, TSK-FUC 方法在大多数数据集的 1 阶 TSK 分类系统也保留了 20% 以上的规则不变. 相应地, 与对应的 0 阶 TSK 分类系统相比, 1 阶 TSK 分类系统的 K_{upd}/K 较高, 但也一定程度上缩小了参数更新的范围. 此外, 表 3 中 0 阶和 1 阶的 D_f/D 可能存在差异, 例如 anemia、yeast 和 steel 数据集. 这是由于类别数据分布不均匀, 导致两种分类系统在部分类上的学习效果不同. 因此, 根据实验方案, 对于同一数据集, 两种分类系统在遗忘实验时所选取的类别可能会出现少许差异.

由表 4 中的数据可知, TSK-FUC 方法在执行效率上与以快速高效著称的 UNSIR 和 Bad-Teaching 遗忘学习方法相比并无优势. 虽然 Bad-Teaching 方法执行效率高, 但其遗忘准确率表现并不突出 (如前文对表 2 的分析). TSK-FUC 与 UNSIR 的总体差距较小, 在多数数据集 (10/16) 的 0 阶 1 阶 TSK 分类系统上执行效率相近. 然而, 对于 nursery、letter 等较大规模数据集, 受限 TSK 模糊系统在高维大数据场景下的性能, 且 UNSIR 在该类场景表现优异, TSK-FUC 执行效率低于 UNSIR. 此外, TSK-FUC 在执行效率上显著优于 Retrain 方法, 这也进一步验证 TSK-FUC 中规则空间划分方法与基于反向误差噪声的遗忘学习策略的高效性. 特别需要指出的是, 当 UNSIR、Bad-Teaching 和 Retrain 方法大范围改动原有规则库中的规则参数时, TSK-FUC 能够较大程度地降低对原有规则库的影响, 并能够始终保证模糊规则具有可解释性. 同时, 实现结构轻量化.

根据表 5, 在 0 阶 TSK 模糊分类系统上, TSK-FUC 在一半数据集上达到或超过 UNSIR 方法的 RT 表现. 同时, 与 Bad-Teaching 方法相比, TSK-FUC 在 RT 表现上优势明显. 对于 1 阶 TSK 模糊分类系统, TSK-FUC 在大多数数据集上, RT 表现优于对比方法. 这也说明在遗忘学习完成度上, TSK-FUC 在 0 阶和 1 阶上与对比方法互有优势. 这表明 TSK-FUC 能够消除分类系统中大部分关于遗忘数据的信息, 整体上达到良好水平.

4.4.2 多类遗忘学习

与单类遗忘学习实验相似, 我们按照实验方案在 16 个数据集的 0 阶、1 阶 TSK 分类系统上进行了多类遗忘学习对比实验, 并统计平均实验结果. 由于 TSK-FUC 的遗忘学习策略在单类和多类遗忘学习场景下并无本质区别, 因此多类遗忘学习场景的执行时间和 RT 对比结果与单类遗忘学习场景类似. 由于篇幅限制, 接下来仅展示准确率对比结果 (见表 6), 以及 TSK-FUC 方法对于多类遗忘学习后的规则分类结果 (见表 7).

表 6 多类遗忘学习后的准确率对比 ($Acc^f \pm Std$ 和 $Acc^l \pm Std$)

Type	Datasets	Trained	Retrain	Bad-Teaching	UNSIR	TSK-FUC
Zero-order TSK	IS	0.8267±0	0.9048±0.0011	0.5737±0.0135	0.8753±0	0.8727±0
		0.8167±0	0±0	0.3020±0.0186	0.0050±0	0.0017±0
	glass	0.6740±0	0.7863±0.0044	0.5181±0.0216	0.7675±0.0024	0.7548±0.0006
		0.7437±0	0±0	0.4637±0.0156	0.0316±0.0112	0.0167±0
	led	0.7434±0	0.7227±0.0012	0.7274±0.0141	0.7644±0.0010	0.7577±0
		0.7277±0	0±0	0.0672±0.0417	0.0009±0.0004	0±0
	flare	0.6276±0	0.7706±0.0004	0.5785±0.0240	0.7336±0.0030	0.7057±0
		0.7808±0	0±0	0.2475±0.0536	0±0	0.0099±0
	anemia	0.7626±0	0.7857±0.0007	0.7505±0.0020	0.8177±0.0043	0.7882±0.0011
		0.7393±0	0±0	0.2054±0.0056	0.0689±0.0099	0.0470±0.0040
	yeast	0.5910±0	0.6796±0.0003	0.6057±0.0123	0.6559±0.0026	0.6534±0.0001
		0.6062±0	0.0004±0	0.1573±0.0445	0.0043±0.0011	0.0050±0
	steel	0.6778±0	0.7277±0.0001	0.6873±0.0019	0.7232±0.0040	0.7195±0
		0.7228±0	0±0	0.3081±0.0091	0.1226±0.0271	0.0570±0
	obesity	0.7425±0	0.8220±0.0010	0.7454±0.0101	0.8080±0.0041	0.7840±0.0001
		0.7286±0	0±0	0.1780±0.0342	0.1578±0.0004	0.0337±0.0001

表 6 多类遗忘学习后的准确率对比 ($Acc^r \pm Std$ 和 $Acc^f \pm Std$)(续)

Type	Datasets	Trained	Retrain	Bad-Teaching	UNSIR	TSK-FUC
Zero-order TSK	penbased	0.9327±0	0.9628±0.0005	0.9229±0.0050	0.9498±0.0008	0.9374±0
		0.9334±0	0±0	0.0791±0.0213	0.0367±0.0002	0.0088±0
	nursery	0.6589±0	0.9701±0	0.7203±0.0357	0.9421±0.0003	0.9371±0.0001
		0.7143±0	0±0	0.1568±0.0274	0.0248±0.0033	0.0025±0
	weather	0.8771±0	0.9376±0	0.8832±0.0146	0.9325±0.0005	0.9259±0
		0.8771±0	0±0	0.1302±0.0271	0.0006±0	0.0077±0
	drybean	0.8949±0	0.9318±0	0.9130±0.0015	0.9289±0.0004	0.9273±0
		0.9008±0	0±0	0.0901±0.0154	0.0549±0.0024	0±0
	letter	0.8040±0	0.8314±0	0.7980±0	0.8069±0	0.8019±0
		0.8004±0	0.0006±0	0.3327±0	0.4082±0	0.0767±0
First-order TSK	IS	0.8560±0	0.9832±0.0005	0.7103±0.0192	0.9293±0	0.9333±0
		0.8767±0	0.0022±0.0043	0.1387±0.0200	0.0417±0	0.0250±0
	glass	0.7088±0	0.9366±0.0045	0.6114±0.0200	0.8170±0.0078	0.7881±0.0099
		0.8225±0	0.0022±0.0034	0.3806±0.0254	0.0643±0.0247	0.0592±0.0099
	led	0.7671±0	0.8197±0.0008	0.7522±0.0112	0.8061±0.0059	0.7922±0.0036
		0.7732±0	0±0	0.0822±0.0425	0.0238±0.0185	0.0485±0.0383
	flare	0.7083±0	0.8283±0.0015	0.6692±0.0141	0.7846±0.0087	0.7749±0.0064
		0.8452±0	0.0001±0.0001	0.3100±0.0277	0.0396±0.0224	0.0847±0.0295
	anemia	0.8700±0	0.9752±0.0047	0.8109±0.0090	0.8884±0.0069	0.8775±0.0052
		0.8584±0	0.0001±0.0002	0.1898±0.0212	0.0240±0.0121	0.0558±0.0105
	yeast	0.6022±0	0.6903±0.0012	0.6095±0.0058	0.6569±0.0059	0.6563±0.0017
		0.6451±0	0.0003±0.0005	0.2521±0.0467	0.0072±0.0027	0.0072±0.0009
	steel	0.7243±0	0.8098±0.0027	0.6327±0.0119	0.7689±0.0070	0.7603±0.0020
		0.7868±0	0.0069±0.0090	0.2851±0.0177	0.0015±0.0006	0.0027±0.0010
	obesity	0.7997±0	0.9757±0.0024	0.7251±0.0125	0.8549±0.0058	0.8329±0.0030
		0.7485±0	0.0009±0.0009	0.1946±0.0316	0.0092±0.0040	0.0107±0.0034
	segment	0.9624±0	0.9864±0.0008	0.9432±0.0038	0.9725±0.0030	0.9678±0
		0.9485±0	0.0003±0.0003	0.1087±0.0176	0.0295±0.0036	0.0136±0
	handwritten	0.9237±0	0.9517±0.0015	0.9150±0	0.9421±0.0018	0.9361±0
		0.9237±0	0.0074±0.0074	0.0441±0	0.0360±0.0031	0.0456±0
	satimage	0.8852±0	0.9276±0.0014	0.8739±0.0040	0.9130±0.0065	0.9106±0.0028
		0.8739±0	0.0144±0.0106	0.2013±0.0872	0.0355±0.0196	0.0251±0.0091
	penbased	0.9460±0	0.9946±0.0003	0.8721±0.0121	0.9711±0.0021	0.9601±0.0009
		0.9462±0	0.0002±0.0001	0.1479±0.0345	0.0005±0.0003	0.0001±0
	nursery	0.7695±0	0.9798±0.0002	0.8058±0.0114	0.9529±0.0032	0.9518±0.0013
		0.8316±0	0±0	0.1321±0.0096	0.0531±0.0034	0.0207±0.0013
	weather	0.8877±0	0.9508±0.0002	0.8818±0.0046	0.9394±0.0022	0.9323±0.0005
		0.8877±0	0.0014±0.0007	0.1509±0.0469	0.0154±0.0055	0.0251±0.0034
	drybean	0.9122±0	0.9488±0.0001	0.9272±0.0026	0.9374±0.0016	0.9375±0.0007
		0.9226±0	0±0	0.1301±0.0718	0.0257±0.0003	0.0257±0.0007
	letter	0.9256±0	0.8755±0	0.8959±0	0.9151±0	0.9064±0
		0.9147±0	0.0065±0	0.1951±0	0.0025±0	0.0574±0

注: 每组上下两个数值分别对应 $Acc^r \pm Std$ 和 $Acc^f \pm Std$

根据表 6, UNSIR^[14]方法在大多数数据集的 0 阶和 1 阶 TSK 模糊分类系统上准确率表现略有优势, 但 TSK-FUC 与其之间的差距较小, 整体上达到相近水平. 与 Bad-Teaching^[13]方法相比, TSK-FUC 在准确率上略有优势. 这表明在所用数据集上, TSK-FUC 也拥有良好的多类遗忘学习可用性. 根据表 7 中数据, 对于大部分所用数据集的 0 阶和 1 阶 TSK 分类系统, 部分规则较少或类均规则 (Number of rules/Number of classes) 较少的 1 阶分类系统除外, 如 handwritten、drybean 和 letter, TSK-FUC 方法在多类遗忘学习过程中删减规则比例 (K_{del}/K) 也都小于或接近数据集的遗忘类的类均规则占比 ($2/C$). 这表明 TSK-FUC 在多类遗忘学习中也能有效降低分类系统的参数量,

且与单类遗忘学习相比,随着遗忘比例的增加,使模型轻量化的作用更加明显.根据表 7,对于大部分数据集,当遗忘数据占比 (D_f/D) 小于 30% 时,TSK-FUC 在 0 阶 TSK 分类系统上保留了超过 50% 的规则不变,而对于 1 阶 TSK 分类系统,TSK-FUC 方法也能保留 20% 以上的规则不变.这说明 TSK-FUC 对规则库的差异化处理策略在 multi-class 遗忘学习中也能够一定程度缩小参数更新范围.

表 7 TSK-FUC 方法对于多类遗忘学习后的规则分类结果

Datasets	D_f/D	Zero-order TSK								D_f/D	First-order TSK							
	(%)	K	K_{del}	K_{del}/K (%)	K_{fix}	K_{fix}/K (%)	K_{upd}	K_{upd}/K (%)	(%)	K	K_{del}	K_{del}/K (%)	K_{fix}	K_{fix}/K (%)	K_{upd}	K_{upd}/K (%)		
IS	28.60	26	6.7	25.8	13.3	51.2	6.0	23.1	28.60	6	1.2	20.0	1.3	21.7	3.5	58.3		
glass	33.30	16	3.8	23.6	6.6	41.0	5.7	35.4	33.30	8	2.2	27.8	1.8	22.2	4.0	50.0		
led	19.00	11	2.0	18.2	8.8	80.0	0.2	1.8	19.00	5	0.6	12.0	1.8	36.0	2.6	52.0		
flare	43.50	26	6.8	26.3	5.3	20.5	13.8	53.2	43.50	6	1.8	30.6	0.7	11.1	3.5	58.3		
anemia	28.30	60	7.5	12.5	25.5	42.5	27.0	45.0	28.30	8	2.0	25.0	1.6	20.0	4.4	55.0		
yeast	23.20	56	10.0	17.9	36.9	65.9	9.1	16.3	23.20	6	0.9	15.0	2.6	43.3	2.5	41.7		
steel	24.60	42	7.3	17.4	24.7	58.8	10.0	23.8	24.60	6	1.0	16.7	0.9	15.0	4.1	68.3		
obesity	28.50	80	10.8	13.5	44.8	56.0	24.4	30.5	28.50	6	1.0	16.7	0.5	8.3	4.5	75.0		
segment	28.60	39	8.8	22.6	23.0	59.0	7.2	18.5	28.60	8	1.9	23.8	1.2	15.0	4.9	61.3		
handwritten	20.00	29	5.8	20.0	21.	75.5	1.3	4.5	20.00	4	0.3	7.5	1.0	25.0	2.7	67.5		
satimage	30.90	18	4.7	26.1	11.0	61.1	2.3	12.8	30.90	6	1.5	25.0	0.7	11.7	3.8	63.3		
penbased	20.00	70	14.1	20.1	55.9	79.9	0.0	0.0	20.00	6	1.0	16.7	1.3	21.7	3.7	61.7		
nursery	65.00	118	62.7	53.1	16.0	13.6	39.3	33.3	65.00	7	3.3	47.6	0.3	4.8	3.3	47.6		
weather	50.00	19	8.0	42.1	8.5	44.7	2.5	13.2	50.00	7	3.0	42.9	1.5	21.4	2.5	35.7		
drybean	29.10	35	10.5	30.0	24.5	70.0	0.0	0.0	29.10	4	0.5	12.5	1.4	35.0	2.1	52.5		
letter	7.70	350	19.6	5.6	295.7	84.5	34.7	9.9	7.70	13	0.5	3.8	2.4	18.5	10.1	77.7		

综上所述,在单类和 multi-class 遗忘学习场景中,TSK-FUC 在遗忘学习可用性^[6]和完成度^[6]上与对比方法相当,仅在遗忘学习速度^[6]上与对比方法略有差距.UNSIR 和 Bad-Teaching 方法虽然效率高,但如相关工作中所述,其遗忘学习策略会破坏规则库的可解释性,且导致已有规则库的大幅更改.相比之下,TSK-FUC 能够针对 TSK 模糊分类系统的特性,先对现有规则空间进行划分,再实施差异化处理.所提出的基于规则总归一化激活强度的规则分类方法兼具执行高效性和分类准确性.差异化处理一方面可最大限度缩小参数更新范围,降低 TSK-FUC 基于反向误差噪声的遗忘学习策略中,随机参数更新过程对原有规则库的影响;另一方面,通过删减与遗忘类高度相关的规则,即降低了遗忘学习难度,也减少了遗忘分类系统参数量,实现结构轻量化.这是包括对比方法在内的适用于神经网络模型的现有遗忘学习方法难以实现的.

5 总 结

本文针对 TSK 模糊分类系统,在保持现有规则库可解释性的前提下,提出了一种面向分类的 TSK 模糊遗忘学习方法,称为 TSK-FUC.实验数据表明,该方法在所用数据集上的 0 阶和 1 阶 TSK 模糊分类系统上均展现出与 UNSIR 方法相近的遗忘学习性能.该方法中基于规则前件归一化激活强度的规则空间划分方法及其差异化处理策略在遗忘学习速度、可用性和完成度上表现良好.该遗忘学习策略能够结合模糊系统便于结构调整的特性,较大幅度地缩小参数更新的范围,从而降低遗忘学习过程对原有规则库的影响.同时,直接删减与遗忘数据高相关的规则即降低了遗忘学习的难度,也精简了分类系统的规则库,降低了其参数量,使得遗忘分类系统在结构上更加轻量化.此外,针对现有规则空间的差异化处理机制也赋予了 TSK-FUC 良好的泛化能力,使其能够适应更为多样化的数据分布场景.

然而,目前 TSK-FUC 的遗忘学习能力仅限于分类遗忘,其规则空间划分方法仍需进一步优化.未来的工作中,可进一步提升所提规则空间划分方法的准确性,并探索样本级、属性级等更细粒度的数据遗忘类型.此外,针对 Mamdani 模糊系统和区间 II 型模糊系统等其他类型的模糊系统,研究其遗忘学习方法亦是未来的工作方向.

References

- [1] Wang ST, Xie RS, Zhou EH. Survey of interpretable deep TSK fuzzy systems. *Journal of Data Acquisition and Processing*, 2022, 37(5): 935–951 (in Chinese with English abstract). [doi: [10.16337/j.1004-9037.2022.05.001](https://doi.org/10.16337/j.1004-9037.2022.05.001)]
- [2] Wang LX. Fuzzy systems are universal approximators. In: *Proc. of the 1992 IEEE Int'l Conf. on Fuzzy Systems*. San Diego: IEEE, 1992. 1163–1170. [doi: [10.1109/FUZZY.1992.258721](https://doi.org/10.1109/FUZZY.1992.258721)]
- [3] Wong SY, Yap KS, Yap HJ, Tan SC, Chang SW. On equivalence of FIS and ELM for interpretable rule-based knowledge representation. *IEEE Trans. on Neural Networks and Learning Systems*, 2015, 26(7): 1417–1430. [doi: [10.1109/TNNLS.2014.2341655](https://doi.org/10.1109/TNNLS.2014.2341655)]
- [4] 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 with English abstract). <http://www.jos.org.cn/1000-9825/5590.htm> [doi: [10.13328/j.cnki.jos.005590](https://doi.org/10.13328/j.cnki.jos.005590)]
- [5] Zhou T, Deng ZH, Jiang YZ, Wang ST. Multi-module TSK fuzzy system based on training space reconstruction. *Ruan Jian Xue Bao/Journal of Software*, 2020, 31(11): 3506–3518 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5846.htm> [doi: [10.13328/j.cnki.jos.005846](https://doi.org/10.13328/j.cnki.jos.005846)]
- [6] Li ZT, Meng XF, Wang LX, Hao XL. Survey on machine unlearning. *Ruan Jian Xue Bao/Journal of Software*, 2025, 36(4): 1637–1664 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/7237.htm> [doi: [10.13328/j.cnki.jos.007237](https://doi.org/10.13328/j.cnki.jos.007237)]
- [7] Xu H, Zhu TQ, Zhang LF, Zhou WL, Yu PS. Machine unlearning: A survey. *ACM Computing Surveys*, 2023, 56(1): 9. [doi: [10.1145/3603620](https://doi.org/10.1145/3603620)]
- [8] Nguyen TT, Huynh TT, Ren Z, Le Nguyen P, Liew AWC, Yin HZ, Nguyen QVH. A survey of machine unlearning. *arXiv:2209.02299*, 2024.
- [9] Graves L, Nagisetty V, Ganesh V. Amnesiac machine learning. In: *Proc. of the 35th AAAI Conf. on Artificial Intelligence*. AAAI, 2021. 11516–11524. [doi: [10.1609/aaai.v35i13.17371](https://doi.org/10.1609/aaai.v35i13.17371)]
- [10] Golatkar A, Achille A, Soatto S. Eternal sunshine of the spotless net: Selective forgetting in deep networks. In: *Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. Seattle: IEEE, 2020. 9301–9309. [doi: [10.1109/CVPR42600.2020.00932](https://doi.org/10.1109/CVPR42600.2020.00932)]
- [11] Bourtole L, Chandrasekaran V, Choquette-Choo CA, Jia HR, Travers A, Zhang BW, Lie D, Papernot N. Machine unlearning. In: *Proc. of the 2021 IEEE Symp. on Security and Privacy (SP)*. San Francisco: IEEE, 2021. 141–159. [doi: [10.1109/SP40001.2021.00019](https://doi.org/10.1109/SP40001.2021.00019)]
- [12] Chundawat VS, Tarun AK, Mandal M, Kankanhalli M. Zero-shot machine unlearning. *IEEE Trans. on Information Forensics and Security*, 2023, 18: 2345–2354. [doi: [10.1109/TIFS.2023.3265506](https://doi.org/10.1109/TIFS.2023.3265506)]
- [13] Chundawat VS, Tarun AK, Mandal M, Kankanhalli M. Can bad teaching induce forgetting? Unlearning in deep networks using an incompetent teacher. In: *Proc. of the 37th AAAI Conf. on Artificial Intelligence*. Washington: IEEE, 2023. 7210–7217. [doi: [10.1609/aaai.v37i6.25879](https://doi.org/10.1609/aaai.v37i6.25879)]
- [14] Tarun AK, Chundawat VS, Mandal M, Kankanhalli M. Fast yet effective machine unlearning. *IEEE Trans. on Neural Networks and Learning Systems*, 2024, 35(9): 13046–13055. [doi: [10.1109/TNNLS.2023.3266233](https://doi.org/10.1109/TNNLS.2023.3266233)]
- [15] Baumhauer T, Schöttle P, Zeppelzauer M. Machine unlearning: Linear filtration for logit-based classifiers. *Machine Learning*, 2022, 111(9): 3203–3226. [doi: [10.1007/s10994-022-06178-9](https://doi.org/10.1007/s10994-022-06178-9)]
- [16] Shibata T, Irie G, Ikami D, Mitsuzumi Y. Learning with selective forgetting. In: *Proc. of the 30th Int'l Joint Conf. on Artificial Intelligence*. *ijcai.org*, 2021. 989–996. [doi: [10.24963/ijcai.2021/137](https://doi.org/10.24963/ijcai.2021/137)]
- [17] Wang JX, Guo S, Xie X, Qi H. Federated unlearning via class-discriminative pruning. In: *Proc. of the 2022 ACM Web Conf.* ACM, 2022. 622–632. [doi: [10.1145/3485447.3512222](https://doi.org/10.1145/3485447.3512222)]
- [18] Di ZL, Yu SX, Vorobeychik Y, Liu Y. Adversarial machine unlearning. *arXiv:2406.07687*, 2024.
- [19] Zhou EH, Chung FL, Wang ST. Multiview fully interpretable TSK fuzzy classifier enhanced by multiview accompanying GMMs. *IEEE Trans. on Fuzzy Systems*, 2024, 32(11): 6288–6302. [doi: [10.1109/TFUZZ.2024.3445389](https://doi.org/10.1109/TFUZZ.2024.3445389)]
- [20] Liu JF, Jiang YZ, Wang J, Deng ZH, Wang ST. Single pass Bayesian fuzzy clustering. *Ruan Jian Xue Bao/Journal of Software*, 2018, 29(9): 2664–2680 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5265.htm> [doi: [10.13328/j.cnki.jos.005265](https://doi.org/10.13328/j.cnki.jos.005265)]
- [21] Zhang YP, Zhou J, Deng ZH, Chung FL, Jiang YZ, Hang WL, Wang ST. Multi-view fuzzy clustering approach based on medoid invariant constraint. *Ruan Jian Xue Bao/Journal of Software*, 2019, 30(2): 282–301 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5625.htm> [doi: [10.13328/j.cnki.jos.005625](https://doi.org/10.13328/j.cnki.jos.005625)]
- [22] Chaira T. A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images. *Applied Soft Computing*, 2011, 11(2): 1711–1717. [doi: [10.1016/j.asoc.2010.05.005](https://doi.org/10.1016/j.asoc.2010.05.005)]
- [23] Micaelli P, Storkey A. Zero-shot knowledge transfer via adversarial belief matching. In: *Proc. of the 33rd Int'l Conf. on Neural*

- Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 9551–9561.
- [24] Du M, Chen Z, Liu C, Oak R, Song D. Lifelong anomaly detection through unlearning. In: Proc. of the 2019 ACM SIGSAC Conf. on Computer and Communications Security. London: ACM, 2019. 1283–1297. [doi: [10.1145/3319535.3363226](https://doi.org/10.1145/3319535.3363226)]
 - [25] Ma Z, Liu Y, Liu XM, Liu J, Ma JF, Ren K. Learn to forget: Machine unlearning via neuron masking. IEEE Trans. on Dependable and Secure Computing, 2023, 20(4): 3194–3207. [doi: [10.1109/TDSC.2022.3194884](https://doi.org/10.1109/TDSC.2022.3194884)]
 - [26] Ameen M, Khan RU, Wang PF, Batool S, Alajmi M. Addressing unreliable local models in federated learning through unlearning. Neural Networks, 2024, 180: 106688. [doi: [10.1016/j.neunet.2024.106688](https://doi.org/10.1016/j.neunet.2024.106688)]
 - [27] Yurdem B, Kuzlu M, Gullu MK, Catak FO, Tabassum M. Federated learning: Overview, strategies, applications, tools and future directions. Heliyon, 2024, 10(19): e38137. [doi: [10.1016/j.heliyon.2024.e38137](https://doi.org/10.1016/j.heliyon.2024.e38137)]
 - [28] Warnecke A, Pirch L, Wressnegger C, Rieck K. Machine unlearning of features and labels. arXiv:2108.11577, 2023.
 - [29] Lughofer E. Process safety enhancements for data-driven evolving fuzzy models. In: Proc. of the 2006 Int'l Symp. on Evolving Fuzzy Systems. Ambelside: IEEE, 2006. 42–48. [doi: [10.1109/ISEFS.2006.251173](https://doi.org/10.1109/ISEFS.2006.251173)]
 - [30] Tang YX, Zhao SY, Chen H, Li CP, Zhai JH, Zhou QJ. Fuzzy rough unlearning model for feature selection. Int'l Journal of Approximate Reasoning, 2024, 165: 109102. [doi: [10.1016/j.ijar.2023.109102](https://doi.org/10.1016/j.ijar.2023.109102)]
 - [31] Zagardo D. A more practical approach to machine unlearning. arXiv:2406.09391, 2024.
 - [32] Liu PY. Mamdani fuzzy system: Universal approximator to a class of random processes. IEEE Trans. on Fuzzy Systems, 2002, 10(6): 756–766. [doi: [10.1109/TFUZZ.2002.805890](https://doi.org/10.1109/TFUZZ.2002.805890)]
 - [33] Gu XQ, Jiang YZ, Wang ST. Zero-order TSK-type fuzzy system for imbalanced data classification. Acta Automatica Sinica, 2017, 43(10): 1773–1788 (in Chinese with English abstract). [doi: [10.16383/j.aas.2017.c160200](https://doi.org/10.16383/j.aas.2017.c160200)]
 - [34] Zhou EH, Vong CM, Nojima Y, Wang ST. A fully interpretable first-order TSK fuzzy system and its training with negative entropic and rule-stability-based regularization. IEEE Trans. on Fuzzy Systems, 2023, 31(7): 2305–2319. [doi: [10.1109/TFUZZ.2022.3223700](https://doi.org/10.1109/TFUZZ.2022.3223700)]
 - [35] Cui YQ, Wu DR, Huang J. Optimize TSK fuzzy systems for classification problems: Minibatch gradient descent with uniform regularization and batch normalization. IEEE Trans. on Fuzzy Systems, 2020, 28(12): 3065–3075. [doi: [10.1109/TFUZZ.2020.2967282](https://doi.org/10.1109/TFUZZ.2020.2967282)]
 - [36] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv:1412.6980, 2017.
 - [37] Dua D, Graff C. UC Irvine machine learning repository. 2023. <http://www.archive.ics.uci.edu/>
 - [38] Alcalá-Fdez J, Sánchez L, García S, del Jesus MJ, Ventura S, Garrell JM, Otero J, Romero C, Bacardit J, Rivas VM, Fernández JC, Herrera F. KEEL: A software tool to assess evolutionary algorithms for data mining problems. Soft Computing, 2009, 13(3): 307–318. [doi: [10.1007/s00500-008-0323-y](https://doi.org/10.1007/s00500-008-0323-y)]
 - [39] Kaggle. Datasets. 2025. <https://www.kaggle.com/datasets>.

附中文参考文献

- [1] 王士同, 谢润山, 周尔昊. 可解释的深度 TSK 模糊系统综述. 数据采集与处理, 2022, 37(5): 935–951. [doi: [10.16337/j.1004-9037.2022.05.001](https://doi.org/10.16337/j.1004-9037.2022.05.001)]
- [4] 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> [doi: [10.13328/j.cnki.jos.005590](https://doi.org/10.13328/j.cnki.jos.005590)]
- [5] 周塔, 邓赵红, 蒋亦樟, 王士同. 基于训练空间重构的多模块 TSK 模糊系统. 软件学报, 2020, 31(11): 3506–3518. <http://www.jos.org.cn/1000-9825/5846.htm> [doi: [10.13328/j.cnki.jos.005846](https://doi.org/10.13328/j.cnki.jos.005846)]
- [6] 李梓童, 孟小峰, 王雷霞, 郝新丽. 机器遗忘综述. 软件学报, 2025, 36(4): 1637–1664. <http://www.jos.org.cn/1000-9825/7237.htm> [doi: [10.13328/j.cnki.jos.007237](https://doi.org/10.13328/j.cnki.jos.007237)]
- [20] 刘解放, 蒋亦樟, 王骏, 邓赵红, 王士同. 单趟贝叶斯模糊聚类算法. 软件学报, 2018, 29(9): 2664–2680. <http://www.jos.org.cn/1000-9825/5265.htm> [doi: [10.13328/j.cnki.jos.005265](https://doi.org/10.13328/j.cnki.jos.005265)]
- [21] 张远鹏, 周洁, 邓赵红, 钟富礼, 蒋亦樟, 杭文龙, 王士同. 代表点一致性约束的多视角模糊聚类算法. 软件学报, 2019, 30(2): 282–301. <http://www.jos.org.cn/1000-9825/5625.htm> [doi: [10.13328/j.cnki.jos.005625](https://doi.org/10.13328/j.cnki.jos.005625)]
- [33] 顾晓清, 蒋亦樟, 王士同. 用于不平衡数据分类的 0 阶 TSK 型模糊系统. 自动化学报, 2017, 43(10): 1773–1788. [doi: [10.16383/j.aas.2017.c160200](https://doi.org/10.16383/j.aas.2017.c160200)]

作者简介

王攀, 博士生, CCF 学生会员, 主要研究领域为模式识别, 人工智能.

王士同, 教授, 博士生导师, 主要研究领域为模式识别, 人工智能.