

## 面向特征演化数据流的增量学习方法研究

陈燕菲, 刘三民

安徽工程大学 计算机与信息学院, 安徽 芜湖 241000

**摘要:**目的 特征演化数据流的特征空间随时间推移而动态变化,传统增量学习方法囿于固定特征空间的假设,无法直接应用于特征演化数据流的学习场景,因此针对挖掘特征演化数据流时面对的分类模型与当前数据特征不匹配而失效、模型预测性能受噪声干扰等问题,提出了一种面向特征演化数据流的增量学习方法。方法 首先,通过引入模糊隶属度函数并结合增量孪生支持向量机模型,鲁棒地训练与更新分类器;当出现新特征时,重新训练新分类器,同时结合局部线性加权回归算法拟合新旧特征之间的映射关系,从而在旧特征消失时,利用所学到的映射关系,将已训练好的旧分类器投影至新特征空间继续更新;最后,结合两种不同的集成策略以合并新旧两分类器实现共同预测。结果 通过大量仿真实验,所提方法分类准确率相较于对比方法提升了0.3%~21.7%;在含不同信噪比数据集上,分类模型性能整体优于对比模型,并随着人工增加噪声比例,模型分类效果受负面影响较小。结论 所提方法得以构建性能高效稳定的分类模型,在提升模型预测精度的同时能减少噪声对分类性能的干扰,增强了模型对特征演化数据流自适应学习能力。

**关键词:**数据流挖掘;特征演化;增量学习;动态数据流;集成学习

**中图分类号:**TP181 **文献标识码:**A **doi:**10.16055/j.issn.1672-058X.2025.0001.013

### Incremental Learning Method for Feature-evolving Data Streams

CHEN Yanfei, LIU Sanmin

School of Computer and Information, Anhui Polytechnic University, Anhui Wuhu 241000, China

**Abstract: Objective** The feature space of feature-evolving data streams dynamically changes over time. Traditional incremental learning methods are constrained by the assumption of a fixed feature space and cannot be directly applied to the learning scenario of feature-evolving data streams. Therefore, to address the problems of classification models mismatching with current data features and model prediction performance being affected by noise interference when mining feature-evolving data streams, an incremental learning method tailored to feature-evolving data streams was proposed.

**Methods** Firstly, by introducing fuzzy membership functions and combining them with an incremental twin support vector machine model, classifiers were robustly trained and updated. When new features appeared, new classifiers were retrained, and the mapping relationship between new and old features was fitted using a local linear weighted regression algorithm. Thus, when old features disappeared, the trained old classifiers were projected into the new feature space for continued updating using the learned mapping relationship. Finally, two different ensemble strategies were combined to

**收稿日期:**2023-04-16 **修回日期:**2023-07-21 **文章编号:**1672-058X(2025)01-0094-11

**基金项目:**安徽省高校自然科学研究重点项目(2022AH050972, KJ2021A0516, KJ2019ZD15);安徽省自然科学基金项目(2108085MF213)。

**作者简介:**陈燕菲(1998—),女,河南巩义人,硕士研究生,从事数据流分类与增量学习研究。

**通讯作者:**刘三民(1978—),男,安徽岳西人,教授,博士,从事数据挖掘与机器学习研究。Email: aqlsm@163.com。

**引用格式:**陈燕菲,刘三民.面向特征演化数据流的增量学习方法研究[J].重庆工商大学学报(自然科学版),2024,42(1):94-104.

CHEN Yanfei, LIU Sanmin. Incremental learning method for feature-evolving data streams[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 42(1): 94-104.

merge the new and old classifiers for joint prediction. **Results** Through extensive simulation experiments, the proposed method improved classification accuracy by 0.3% to 21.7% compared with baseline methods. On datasets with different signal-to-noise ratios, the overall performance of the classification model was superior to that of the baseline model, and the model's classification effectiveness was less affected by artificially increasing the noise ratio. **Conclusion** The proposed method is verified to construct an efficient and stable classification model, which not only enhances model prediction accuracy but also reduces the interference of noise on classification performance, thus strengthening the adaptive learning capability of the model for feature-evolving data streams.

**Keywords:** data stream mining; feature evolution; incremental learning; dynamic data streams; ensemble learning

## 1 引言

大数据时代,海量数据往往以流的形式出现并随时间积累,如何挖掘数据流中潜在重要信息成为机器学习研究热点之一<sup>[1]</sup>。由于数据流学习环境的动态可变性,导致特征空间随时间推移在特征维度、特征属性上动态改变,通常将数据流的特征空间动态可变的现象称为特征演化<sup>[2]</sup>。这种现象经常出现在环境监控<sup>[3]</sup>、智慧医疗<sup>[4-5]</sup>和热点话题等<sup>[6]</sup>领域中,如何对特征演化数据流进行分析将成为动态数据流挖掘场景中新涌现的问题之一。

特征演化数据流挖掘任务面临诸多挑战:如高速海量数据的实时处理、分类模型与当前数据特征不匹配而无法预测等。近年来,针对数据流中特征演化(Feature evolution)问题的研究工作相继展开,并取得一系列研究成果。Hou 等<sup>[7]</sup>提出了一种新的特征演化流学习范式(Feature Evolvable Streaming Learning, FESL),通过特征映射恢复消失特征空间数据并利用在线梯度下降算法对模型进行更新,结合集成策略实现特征演化流的学习。基于 FESL 算法的思想,Liu 等<sup>[8]</sup>提出一种基于被动攻击更新策略的特征演化学习方法,该方法采用一阶在线学习算法加快模型收敛速度和提升分类性能。随后,在文献[9]中提出特征加权的置信-加权学习算法,进一步挖掘数据特征相关性和二阶信息提升了准确性,但同时带来较高的时空复杂度。文献[10]提出具有增量和递减特征的一遍学习方法,通过压缩与拓展两个阶段实现特征信息的保留与传递,以辅助对当前数据的分类预测精度。面对同时具有重叠实例和重叠特征的特征演化场景,文献[11]提出面向特征继承性增减的在线分类算法,通过结合在线被动-主动算法和结构风险最小化原则进行模型训练,并运用在线矩阵补全方法恢复原特征数据实现模型重用,但缺失特征补全的策略可能使特征冗余甚至引入噪声。上述方法大多聚焦于对特征演化流的适应性学习问题研究上,随着对特征演化数据流研究的深入,Hou

等<sup>[12]</sup>意识到更实际的情况是原特征会不可预测地消失,从而导致一个不完整重叠周期,因此提出具有不可预测特征演化的预测,运用矩阵补全的方法来解决重叠时期的特征不完全问题。由于特征演化的同时也可能导致数据分布变化,这将使特征映射或补全方法不再可靠,因此文献[13]提出特征空间和分布可进化流学习,为充分利用旧特征空间中的数据,关键在于弥合新旧特征空间与新特征空间之间的差距,针对这一问题提出一种新的差异度量方法——进化差异,利用横跨两个不同特征空间的差异度量来提高模型泛化能力。考虑大多时间步后不提供标签,虽然流形正则化技术可以很好地解决标签缺失问题,但是需要一个缓冲区来存储部分以往数据来辅助学习,于是 Hou 等<sup>[14]</sup>提出存储适合特征演化流学习,利用流形正则化技术将 FESL 算法中的损失函数转换为“风险函数”,从而重新使用在线梯度下降进行计算。当进入大数据的在线深度学习环境下,对存在特征演化的高维数据流的分类则需要权衡收敛速度和模型深度,Lian 等<sup>[15]</sup>提出新的在线深度学习范式,其核心思想是用变分逼近发现新旧特征之间的共享潜在子空间,由此实现在线分类遗憾快速最小化和特征级关系的实施逼近。

解决特征演化问题的关键是在特征动态变化的情况下,如何尽可能保留已学习到的知识以构建动态自适应分类模型。上述相关工作结合特征映射、矩阵补全与特征划分等方法,在学习适应特征演化的分类模型、提高模型预测性能等方面的研究做出贡献。但考虑学习环境的复杂性,数据采集过程中不可避免地容易将噪声数据引入特征演化数据流中,这将对数据挖掘模型的泛化能力产生负面影响<sup>[16]</sup>。因此,为构建高效鲁棒的特征演化数据流自适应分类模型,本文结合现有研究成果,提出一种面向特征演化流的增量学习方法(Incremental Learning with Feature Evolvable data streams, ILFE),引入增量式模糊孪生支持向量机,在提升模型预测准确率的同时降低噪声干扰;然后,通过投

影策略实现模型重用以适应特征演化;最后,采用新的集成策略保证模型的稳定性。

## 2 问题背景

本文所研究的特征演化数据流问题如图 1 所示。数据流被切分成一系列的微批次 (mini-batch) 进行收集并进行增量训练。沿时间轴在  $1 \sim T_1 - B$  阶段,数据样本均来自特征空间  $S_1$ ;当进入  $T_1 - B + 1 \sim T_1$  阶段时,新的特征(简称  $S_2$ )出现,与  $S_1$  共同描述数据。这样每个数据样本的特征空间都可被分为两个部分,因此本文将这些数据称之为“共现数据”;随后,进入  $T_1 + 1 \sim T_2$  阶段,新到达数据仅来自特征空间  $S_2$  下。该过程构成一个变化周期,可以循环往复,每个周期包含两个特征空间  $S_1 = \{\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_{d_1}\} \in \mathbf{R}^{d_1}$  和特征空间  $S_2 = \{\tilde{f}_{d_1+1}, \tilde{f}_{d_1+2}, \dots, \tilde{f}_{d_2}\} \in \mathbf{R}^{d_2}$ 。值得注意的是,各阶段的数据特征保持相对稳定,但当前阶段无法预知下一阶段的特征空间及其变化,同时整个数据流上均含有噪声干扰。各阶段的特征变化规律总结如下:

(1)  $i = 1 \sim T_1 - B$  时,分类器观测到数据样本  $\mathbf{x}_i^{(1)} \in \mathbf{R}^{d_1}$ ,所有样本均来自特征空间为  $S_1$ ,其中  $d_1$  为特征空间  $S_1$  的维度。

(2)  $i = T_1 - B + 1 \sim T_1$  时,分类器获取样本  $\mathbf{x}_i^{(c)} = \{([\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)}], y_i)\}_{T_1 - B + 1}^{T_1}$ ,其中  $\mathbf{x}_i^{(1)}$  和  $\mathbf{x}_i^{(2)}$  分别来自特征空间  $S_1$  和  $S_2$ ,其中  $d_2$  为特征空间  $S_2$  的维度,则  $d_c = d_1 + d_2$ 。

(3)  $i = T_1 + 1 \sim T_2$ ,分类器观测到数据样本  $\mathbf{x}_i^{(2)} \in \mathbf{R}^{d_2}$ ,所有样本均来自特征空间为  $S_2$ 。

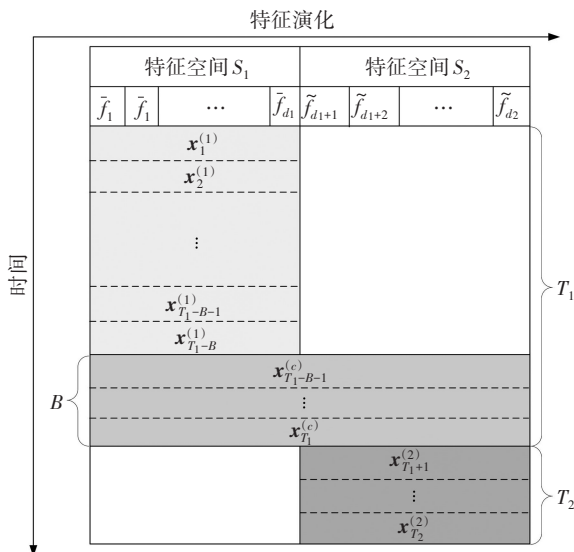


图 1 特征演化数据流

Fig. 1 Feature-evolving data stream

为便于描述,表 1 中列举了本文所用符号及对应含义。

表 1 符号及相关概念

Table 1 Notations and concepts

符 号	含 义
$\mathbf{x}_i$	第 $i$ 轮输入的数据样本
$y_i$	样本 $\mathbf{x}_i$ 的真实标签
$\hat{y}_i$	样本 $\mathbf{x}_i$ 的预测标签
$d_i$	样本 $\mathbf{x}_i$ 的特征维度
$X_{\pm}^{(1)}$	来自特征空间 $S_1$ 的实例集合
$X_{\pm}^{(c)}$	共现数据实例集合
$X_{\pm}^{(2)}$	来自特征空间 $S_2$ 的实例集合
$\mathbf{M}$	$S_1$ 和 $S_2$ 之间的映射关系矩阵
$l_{\pm}$	正类或负类实例数量
$(\bar{\mathbf{w}}_{\pm}, \bar{b}_{\pm})$	特征空间 $S_1$ 中分类器
$(\tilde{\mathbf{w}}_{\pm}, \tilde{b}_{\pm})$	特征空间 $S_2$ 中分类器, $\tilde{\mathbf{w}}_{\pm} = \mathbf{M}\bar{\mathbf{w}}_{\pm}$
$f_{\pm}$	正类和负类样本的超平面函数, $f_{\pm} = \mathbf{w}_{\pm}\mathbf{x} + b_{\pm}$
$\ell_i$	分类模型在 $(\mathbf{x}_i, y_i)$ 上逻辑损失
batch	一个 mini-batch 的样本个数

## 3 增量学习框架模型

为解决本文所设置的研究场景,通过增量学习方法建立预测模型,需要解决如下 3 个关键问题:

(1) 如何消除噪声干扰,增量训练鲁棒的分类模型;

(2) 如何在异构特征场景下实现模型拓展与重用,充分利用历史知识提升分类模型的准确性;

(3) 如何利用集成策略,结合更新重用的旧模型和重新训练的新模型进行共同预测。

本节结合上述归纳的 3 个关键问题,将从模型训练与更新、模型拓展与重用和集成学习策略 3 个方面详细阐述本文所设计的增量学习框架。

### 3.1 模型训练与更新

由于增量孪生支持向量机可以快速高效地训练分类模型,同时考虑到模糊隶属度在鲁棒分类学习中起着关键作用,本文采取文献[17]中模糊有界孪生支持向量机算法 (Fuzzy Bounded Twin Support Vector Machine, FBTWSVM) 对数据流进行增量训练与更新模型。

#### 3.1.1 模糊隶度函数

在解决含噪声或异常点的数据流分类问题时,可通过设计合适的模糊隶度函数,为每个样本计算模糊

隶度值,以观察不同的样本对决策函数的影响程度。本文采用的模糊隶度函数根据每个数据样本到类中心距离来计算样本的模糊隶度值。首先选取正类实例和负类实例的均值点  $\bar{x}_+$  和  $\bar{x}_-$  作为类中心,如式(1)所示:

$$\bar{x}_+ = \frac{1}{l_+} \sum_{i=1, y_i > 0}^{l_+} x_i, \bar{x}_- = \frac{1}{l_-} \sum_{i=1, y_i < 0}^{l_-} x_i \quad (1)$$

通过式(2)计算实例与其类中心之间的距离,得出正类和负类的超球半径:

$$\begin{aligned} r_+ &= \max \{ \|x_i - \bar{x}_+\| \mid y_i > 0 \} \\ r_- &= \max \{ \|x_i - \bar{x}_-\| \mid y_i < 0 \} \end{aligned} \quad (2)$$

那么,数据样本的模糊隶度函数如式(3)所示<sup>[18]</sup>:

$$s_i = \begin{cases} \mu(1 - \|x_i - \bar{x}_+\| / (r_+ + \varepsilon)), \|x_i - \bar{x}_+\| \geq \|x_i - \bar{x}_-\| \ \& y_i = +1 \\ (1 - \mu)(1 - \|x_i - \bar{x}_+\| / (r_+ + \varepsilon)), \|x_i - \bar{x}_+\| < \|x_i - \bar{x}_-\| \ \& y_i = +1 \\ \mu(1 - \|x_i - \bar{x}_-\| / (r_- + \varepsilon)), \|x_i - \bar{x}_-\| \geq \|x_i - \bar{x}_+\| \ \& y_i = -1 \\ (1 - \mu)(1 - \|x_i - \bar{x}_-\| / (r_- + \varepsilon)), \|x_i - \bar{x}_-\| < \|x_i - \bar{x}_+\| \ \& y_i = -1 \end{cases} \quad (3)$$

其中,  $0 \leq \mu \leq 1$  是用于平衡普通实例和噪声的参数,  $\varepsilon > 0$  是为了避免  $s_i = 0$ 。式(3)表明:如果实例距离类中心较远,说明该实例重要程度越低,则通常会按比例赋予实例更小的  $s_i$  值。

### 3.1.2 线性分类模型

为对数据流进行实时处理,可通过增量学习模糊孪生支持向量机,同时引入式(3)中有效的模糊隶度函数,既降低异常值所带来的噪声,又考虑支持向量的重要性。首先,本文所提 ILFE 算法采用孪生支持向量机作为基分类器,目标是求解式(4)以学习两个非平行的超平面:

$$w_+^T x + b_+ = 0 \text{ and } w_-^T x + b_- = 0 \quad (4)$$

通过引入的模糊隶度函数和边际参数,具有线性核的 ILFE 的原始问题定义如式(5)和式(6)所示<sup>[17]</sup>:

$$\begin{aligned} \min_{w_+, b_+, \xi_-} & c_1 \eta (\|w_+\|^2 + b_+^2) + \\ & \eta \|X_+ w_+ + e_+ b_+\|^2 + c_2 s_+^T \xi_- \\ \text{s. t. } & (-1) * (X_- w_+ + e_+ b_+) + \xi_- \geq e_-, \xi_- \geq 0 \end{aligned} \quad (5)$$

$$\begin{aligned} \min_{w_-, b_-, \xi_+} & c_3 \eta (\|w_-\|^2 + b_-^2) + \\ & \eta \|X_- w_- + e_- b_-\|^2 + c_4 s_-^T \xi_+ \\ \text{s. t. } & (+1) * (X_+ w_- + e_- b_-) + \xi_+ \geq e_+, \xi_+ \geq 0 \end{aligned} \quad (6)$$

其中,  $c_1$  和  $c_3$  是间隔参数,  $c_2$  和  $c_4$  是惩罚参数,并加入  $b_+, b_-$  使结构风险最小化,  $\xi_+, \xi_-$  为松弛变量,  $e_+$  和  $e_-$  是单位列向量,平衡系数  $\eta = 0.5$ ,同时  $s_+ \in R^{l_+}$  和  $s_- \in R^{l_-}$  分别是正类、负类实例相关的模糊隶度值,它们为加权正则化模型的鲁棒性提供了保障。

**定理 1** 式(5)和式(6)对应原始问题的对偶形式为<sup>[18]</sup>

$$\begin{aligned} \max_{\alpha} & e_-^T \alpha - \frac{1}{2} \alpha^T H_- (H_-^T H_- + c_1 E_+)^{-1} H_-^T \alpha \\ \text{s. t. } & 0 \leq \alpha \leq c_2 s_- \end{aligned} \quad (7)$$

$$\begin{aligned} \max_{\beta} & e_+^T \beta - \frac{1}{2} \beta^T H_+ (H_+^T H_+ + c_3 E_-)^{-1} H_+^T \beta \\ \text{s. t. } & 0 \leq \beta \leq c_4 s_+ \end{aligned} \quad (8)$$

这里  $H_+ = [X_+, e_+]$ ,  $H_- = [X_-, e_-]$ , 并且  $E_{\pm} = \begin{pmatrix} I \\ 0 \end{pmatrix}$ 。原始问题式(5)和式(6)及其对偶问题式(7)和式(8)的最优解为

$$\begin{aligned} u_+^* &= (-1) * (H_+^T H_+ + c_1 E_+)^{-1} H_+^T \alpha^* \\ u_-^* &= (+1) * (H_-^T H_- + c_3 E_-)^{-1} H_-^T \beta^* \end{aligned} \quad (9)$$

**证明** 定理 1 的详细证明可参考文献[18]。

通过求解式(9)中  $u_{\pm}^* = [w_{\pm}^{*T}, b_{\pm}^*]^T$ ,  $\alpha^*$  和  $\beta^*$  分别为  $\alpha$  和  $\beta$  的最优解,  $w_{\pm}$  和  $b_{\pm}$  分别表示为正类和负类样本构造的超平面系数。用  $\|\cdot\|$  是绝对值函数,那么可得出预测函数,如式(10)所示:

$$f(x) = \arg \min_{\pm} \frac{|w_{\pm}^T x + b_{\pm}|}{\|w_{\pm}\|} \quad (10)$$

本文算法将新的数据点连续地集成到现有模型中,而无需完全重构模型,这使它适合于增量学习数据流。所采用的增量式 FBTWSVM<sup>[17]</sup> 基于收缩启发式 (Shrinking heuristic) 方法,考虑新值的模糊信息,通过只选择新的样本点来更新模型,从上一个训练步骤中推断出投影梯度的最小值  $\min_{i-1}$  和最大值  $\max_{i-1}$ ,并由式(11)计算新到样本点的投影梯度:

$$\nabla f(\alpha) = -H_i u_{\pm} - 1 \quad (11)$$

判断其数值大小是否在  $(\min_{i-1}, \max_{i-1})$  区间内来区分支持向量和影响分类边界的异常点。同时为避免增量过程导致模型维数的持续增长,不断去除对模型精度干扰较小或没有干扰的数据来控制模型维数。本文所提算法 ILFE 的伪代码具体如算法 1 所示。

## 算法 1 ILFE

输入: 训练样本  $X_{\pm}^{(1)} = \{(\mathbf{x}_{\pm i}^{(1)}, y_i)\}_1^{T_1}$  以及  
 共现数据样本  $X_{\pm}^{(c)} = \{([\mathbf{x}_{\pm i}^{(1)}, \mathbf{x}_{\pm i}^{(2)}], y_i)\}_1^{T_1}$   
 和新特征样本  $X_{\pm}^{(2)} = \{(\mathbf{x}_{\pm i}^{(2)}, y_i)\}_1^{T_2}$ .

输出: 对  $X_{\pm}^{(2)}$  的标签预测结果.

- (1) **for**  $i=1, 2, \dots, T_1$  **do**
- (2) 接收样本  $\mathbf{x}_{\pm i}^{(1)}$ ;
- (3) 选择合适的参数  $c_j (j=1, 2, 3, 4)$ ;
- (4) 利用式(3)计算各样本点模糊隶度  $s_i$ ;
- (5) 通过求解式(8)和式(9)的对偶问题,得到最优解  $\alpha^*$  和  $\beta^*$ ;
- (6) 计算得到原始问题式(6)和式(7)的最优解  $\mathbf{w}_{\pm}^*$  和  $b_{\pm}^*$ ;
- (7) 通过式(10)更新超平面  $\bar{\mathbf{u}}_{\pm}^* = ((\bar{\mathbf{w}}_{\pm}^*)^T, \bar{b}_{\pm}^*)$ ;
- (8) **if**  $i > T_1 - B$  **then**
- (9) 接收共现数据样本  $\mathbf{x}_{\pm i}^{(c)}$ ;
- (10) 调用算法 2, 学习一个映射矩阵  $\mathbf{M}$ ;
- (11) 利用  $\tilde{\mathbf{w}}_{\pm}^* = \mathbf{M}\bar{\mathbf{w}}_{\pm}^*$  将模型由  $S_1$  投影至  $S_2$ ;
- (12) 获取超平面  $\tilde{\mathbf{u}}_{\pm}^* = ((\tilde{\mathbf{w}}_{\pm}^*)^T, \tilde{b}_{\pm}^*)$ ;
- (13) 同步骤(2)-(7), 利用共现数据中新特征样本训练新的超平面  $\tilde{\mathbf{v}}_{\pm}^* = ((\tilde{\mathbf{w}}_{\pm}^*)^T, \tilde{b}_{\pm}^*)$ ;
- (14) **end if**
- (15) **end for**
- (16) **for**  $i=T_1+1, T_1+2, \dots, T_1+batch$  **do**
- (17) 接收样本  $\mathbf{x}_{\pm i}^{(2)}$ ;
- (18) 同步骤(3)-(7), 继续更新两个超平面  $\tilde{\mathbf{u}}_{\pm}^* = ((\tilde{\mathbf{w}}_{\pm}^*)^T, \tilde{b}_{\pm}^*)$ ,  $\tilde{\mathbf{v}}_{\pm}^* = ((\tilde{\mathbf{w}}_{\pm}^*)^T, \tilde{b}_{\pm}^*)$ ;
- (19) **end for**
- (20) 调用算法 3, 对样本  $X_{\pm}^{(2)} = \{(\mathbf{x}_{\pm i}^{(2)}, y_i)\}_1^{T_2}$  进行标签预测;

## 3.2 模型拓展与重用

为提升异构特征空间的分类效率, 本文选择局部加权线性回归方法, 以学习异构空间的线性映射关系对应的系数矩阵  $\mathbf{M}$ , 实现模型重用, 利用共现数据可学习  $\mathbf{M}$ , 如式(12)所示:

$$\min_{\mathbf{M} \in R^{d_2 \times d_1}} \sum_{i=T_1-B+1}^{T_1} K^i (\mathbf{M}^T \mathbf{x}_i^{(2)} - \mathbf{x}_i^{(1)})^2 \quad (12)$$

其中, 高斯核权重  $K$  形式如式(13):

$$K(i, i) = e^{-\frac{(\mathbf{x}_i^{(2)} - \mathbf{x}_i^{(1)})^2}{2\kappa^2}} \quad (13)$$

式(14)中, 问题的最优解  $\mathbf{M}^*$  如下:

$$\mathbf{M}^* \left( \sum_{i=T_1-B+1}^{T_1} \mathbf{x}_i^{(2)} \mathbf{K} \mathbf{x}_i^{(2)T} \right)^{-1} \left( \sum_{i=T_1-B+1}^{T_1} \mathbf{x}_i^{(2)} \mathbf{K} \mathbf{x}_i^{(1)T} \right) \quad (14)$$

计算得到该线性映射矩阵后, 可以通过  $\mathbf{M}^T \mathbf{w}_{\pm}^*$  将已训练好的模型参数投影至新特征空间下, 从而对新到来样本继续预测。该线性映射矩阵最优解  $\mathbf{M}^*$  的具体学习流程如算法 2 所示。

算法 2 学习线性映射最优解  $\mathbf{M}^*$ 

输入: 共现数据样本  $X_{\pm}^{(c)} = \{([\mathbf{x}_{\pm i}^{(1)}, \mathbf{x}_{\pm i}^{(2)}], y_i)\}_1^{T_1}$ .

输出: 线性映射最优解  $\mathbf{M}^*$ .

- (1) 初始化  $\mathbf{M}_1 = \mathbf{O}^{d_2 \times d_1}$  和  $\mathbf{M}_2 = \mathbf{O}^{d_2 \times d_2}$ ;
- (2) **for**  $i=T_1-B+1, T_1-B+2, \dots, T_1$  **do**
- (3) 接收样本  $\mathbf{x}_{\pm i}^{(c)} = [\mathbf{x}_{\pm i}^{(1)}, \mathbf{x}_{\pm i}^{(2)}]$ ;
- (4) 计算核矩阵  $K(i, i)$ ;
- (5) 交替计算系数矩阵  $\mathbf{M}_1 = \mathbf{M}_1 + \mathbf{x}_{\pm i}^{(2)} K(\mathbf{x}_{\pm i}^{(2)})^T$  和系数矩阵  $\mathbf{M}_2 = \mathbf{M}_2 + \mathbf{x}_{\pm i}^{(2)} K(\mathbf{x}_{\pm i}^{(1)})^T$ ;
- (6) **end for**
- (7) 计算得到最优解  $\mathbf{M}^* = \mathbf{M}_1^{-1} \mathbf{M}_2$ ;

## 3.3 模型集成

处理特征演化带来的新特征空间数据时, ILFE 算法中通过拓展重用与重新训练得到两个基分类器, 并对应产生两个预测结果。本节提出两种新的集成方法: 加权组合和选择最优, 利用两个基分类器进行共同预测, 保证模型整体分类性能高效稳定。

## 3.3.1 加权组合预测

该方法基于当前逻辑损失计算分类器权重, 通过加权组合的方法进行预测标签。两个基分类器预测值  $f_{1,i}$  和  $f_{2,i}$  如式(15)所示:

$$f_{1,i} = \arg \min_{\pm} \frac{|(\tilde{\mathbf{w}}_{1\pm}^*)^T \mathbf{x} + \tilde{b}_{1\pm}^*|}{\|\tilde{\mathbf{w}}_{1\pm}^*\|}$$

$$f_{2,i} = \arg \min_{\pm} \frac{|(\tilde{\mathbf{w}}_{2\pm}^*)^T \mathbf{x} + \tilde{b}_{2\pm}^*|}{\|\tilde{\mathbf{w}}_{2\pm}^*\|} \quad (15)$$

第  $i$  轮的预测结果是对两个基分类器预测值  $f_{1,i}$  和  $f_{2,i}$  的加权平均值, 如式(16)所示:

$$\hat{p} = \theta_{1,i} f_{1,i} + \theta_{2,i} f_{2,i} \quad (16)$$

其中,  $\theta_{1,i}$  和  $\theta_{2,i}$  是两个基分类器预测值的权重。根据各个基分类器上一轮的损失值, 可通过式(17)来计算当前分类器权重:

$$\theta_{a,i+1} = \frac{\theta_{a,i} \left( 1 - \frac{1}{(1 - \ell_i) + \zeta} \right)}{\sum_{j=1}^2 \theta_{j,i} \left( 1 - \frac{1}{(1 - \ell_i) + \zeta} \right)}, a = 1, 2 \quad (17)$$

本文设置一个很小的常数项  $\zeta = 0.001$ , 用来避免

分母为零。本文将采用加权组合策略的算法称为 ILFE-c,如算法 3 所示。

算法 3 ILFE-c

---

输入: 测试样本  $X_{\pm}^{(2)} = \{(\mathbf{x}_{\pm i}^{(2)}, y_i)\}_{T_1+batch+1}^{T_2}$ ;

分类器  $\tilde{\mathbf{u}}_{\pm}^* = ((\tilde{\mathbf{w}}_{1\pm}^*)^T, (\tilde{b}_{1\pm}^*)^T)$  和  
分类器  $\tilde{\mathbf{v}}_{\pm}^* = ((\tilde{\mathbf{w}}_{2\pm}^*)^T, (\tilde{b}_{2\pm}^*)^T)$ .

输出: 测试样本标签的预测结果.

(1) **for**  $t=T_1+batch+1, T_1+batch+2, \dots, T_1+T_2$  **do**

(2) 通过分类器  $\tilde{\mathbf{u}}_{\pm}^* = ((\tilde{\mathbf{w}}_{1\pm}^*)^T, (\tilde{b}_{1\pm}^*)^T)$  对  $X_{\pm}^{(2)}$  进行标签预测得到  $f_{1,i}$ ;

(3) 通过分类器  $\tilde{\mathbf{v}}_{\pm}^* = ((\tilde{\mathbf{w}}_{2\pm}^*)^T, (\tilde{b}_{2\pm}^*)^T)$  对  $X_{\pm}^{(2)}$  进行标签预测得到  $f_{2,i}$ ;

(4) 利用式(16)计算预测结果  $\hat{p}$ ;

(5) 揭示真实标签并计算当前损失  $\ell_i$ ;

(6) 利用式(17)更新分类器权重;

(7) **end for**

---

### 3.3.2 选择最优预测

上述加权组合策略可通过集成多个基分类器预测结果以提升整体分类性能。该方法的前提假设是各个基分类器性能都不能低于某个阈值,这一假设在本文研究场景中并不总是成立,因为在特征演化初期分类器  $\tilde{\mathbf{v}}_{\pm}^*$  性能较差,随着样本积累  $\tilde{\mathbf{u}}_{\pm}^*$  可能逐渐不适合对当前样本分类。因此,提出另一种基于选择最优预测的策略,在本轮中挑选最优基分类器预测作为最终结果,如式(18)所示:

$$\hat{p} = f_{a,i}, a = 1 \text{ 或 } 2 \quad (18)$$

通过式(19)中基权重分布来决定挑选哪个基分类器预测值:

$$\rho_{a,i+1} = \frac{\theta_{a,i}}{\sum_{j=1}^2 \theta_{j,i}}, a = 1, 2 \quad (19)$$

其中,权重  $v_{1,i}$  和  $v_{2,i}$  的权重更新策略如式(20)所示:

$$v_{a,i+1} = \theta_{a,i} \left( 1 - \frac{1}{(1-\ell) + \zeta} \right)$$

$$\theta_{1,i+1} = \mu \frac{V_i}{2} + (1-\mu)\theta_{1,i}, a = 1, 2 \quad (20)$$

其中,对于  $V_i$  和  $\mu$  定义如式(21)所示:

$$V_i = v_{1,i} + v_{2,i} \quad (21)$$

$$\mu = \frac{1}{T_2 - batch - 1}$$

本文将采用挑选最优预测策略的算法称为 ILFE-s,如算法 4 所示。

算法 4 ILFE-s

---

输入: 测试样本  $X_{\pm}^{(2)} = \{(\mathbf{x}_{\pm i}^{(2)}, y_i)\}_{T_1+batch+1}^{T_2}$ ;

分类器  $\tilde{\mathbf{u}}_{\pm}^* = ((\tilde{\mathbf{w}}_{1\pm}^*)^T, (\tilde{b}_{1\pm}^*)^T)$  和  
分类器  $\tilde{\mathbf{v}}_{\pm}^* = ((\tilde{\mathbf{w}}_{2\pm}^*)^T, (\tilde{b}_{2\pm}^*)^T)$ .

输出: 测试样本标签的预测结果.

(1) **for**  $t=T_1+batch+1, T_1+batch+2, \dots, T_1+T_2$  **do**

(2) 通过分类器  $\tilde{\mathbf{u}}_{\pm}^* = ((\tilde{\mathbf{w}}_{1\pm}^*)^T, (\tilde{b}_{1\pm}^*)^T)$  对  $X_{\pm}^{(2)}$  进行标签预测得到  $f_{1,i}$ ;

(3) 通过分类器  $\tilde{\mathbf{v}}_{\pm}^* = ((\tilde{\mathbf{w}}_{2\pm}^*)^T, (\tilde{b}_{2\pm}^*)^T)$  对  $X_{\pm}^{(2)}$  进行标签预测得到  $f_{2,i}$ ;

(4) 利用式(18)计算预测结果  $\hat{p}$ ;

(5) 揭示真实标签并计算当前损失  $\ell_i$ ;

(6) 利用式(20)更新分类器权重;

(7) **end for**

---

## 4 实验和结果分析

### 4.1 数据集

本文选用了来自不同领域的 10 个 UCI 数据集(所有数据集均可在 <http://archive.ics.uci.edu> 进行下载),通过采取和文献[7]相同处理方法,生成合成数据以模拟特征演化数据流。具体来说,原始数据集最初仅有一个特征空间,借助随机高斯矩阵将原始数据集人工映射到新的特征空间下,并在学习过程中逐个读取使数据依次到来,那么将获得来自两个不同的特征空间  $S_1$  和  $S_2$  的数据流,即特征演化数据流。合成的人工数据集的详细信息如表 2 所示,其中  $d_1$  表示原始数据集的特征空间  $S_1$  的维度,  $d_2$  表示人工生成的新特征空间  $S_2$  的维度。

表 2 数据集描述

Table 2 Detailed description of datasets

数据集	样本个数	$d_1$	$d_2$
Australian	690	42	29
Credit-a	653	15	10
Credit-g	1 000	20	14
Diabetes	768	8	5
DNA12	940	180	125
German	1 000	59	41
Kr-vs-kp	3 196	36	25
Svmguide3	1 284	22	15
Magic04	19 020	10	7
HTRU_2	17 898	8	5

### 4.2 对比算法及实验设置

将本文算法 ILFE-c、ILFE-s 与文献[7]中所提算法 FESL-c、FESL-s 以及其过程中涉及的算法 NOGD、ROGD-f 和 ROGD-u,通过实验评估所提算法在测试数据上的分类性能。同时对对比算法还包括算法 NX、RX-u 进行对比,其中 NX 是指从头调用 FBTWSVM 算

法对共现数据进行训练,并对新特征空间下数据的预测;RX- $u$  是指通过线性映射,将新特征空间下的数据恢复至旧特征空间,利用 FBTWSVM 继续更新已学到的模型进行预测。

为保证实验对比的公平性,对比算法 NOGD、ROGD- $u$ 、FESL- $s$  以及 FESL- $c$  的参数设置中,对于 HTRU\_2 数据集和 Magic04 数据集,经过调优后参数  $c$  均设置为 200,其他数据集的参数  $c$  的选取保留了文献 [7] 中的设置。所有算法的分类精度是通过 10 次独立重复实验,每次重复随机打乱数据顺序,最后预测结果取 10 次结果平均值得到。参数敏感性实验中,研究参

数  $c_i (i=1,2,3,4)$  对分类准确率的影响,其中  $c_1=c_3$  取值范围为  $[2^1, 2^2, 2^3, 2^4]$ ,  $c_2=c_4$  其取值范围为  $[2^0, 2^1, 2^2, 2^3]$ 。

### 4.3 实验结果分析

#### 4.3.1 无噪声数据集上对比实验

为验证本文算法的可行性,本文采用上述 8 种算法在 10 个不同不含噪声的人工数据集上进行实验,实验结果如表 3 所示。其中,通过分组与整体对比不同算法性能,并用粗体来标注每个网格中较好的实验结果,用粗体加下划线的组合标记采取集成策略的算法中整体最优预测结果。

表 3 无噪声人工数据集上各对比算法准确率

Table 3 Accuracy of the compared algorithms on synthetic datasets without noise

Dataset	Australian	Credit- $a$	Credit- $g$	Diabetes	DNA12
NOGD	0.592±0.034	0.670±0.040	0.597±0.025	0.623±0.038	0.509±0.035
NX	<b>0.868±0.021</b>	<b>0.845±0.028</b>	<b>0.720±0.028</b>	<b>0.710±0.029</b>	<b>0.890±0.014</b>
ROGD- $u$	0.860±0.028	0.837±0.022	<b>0.732±0.025</b>	0.625±0.030	<b>0.890±0.014</b>
RX- $u$	<b>0.869±0.023</b>	<b>0.847±0.026</b>	0.718±0.030	<b>0.700±0.038</b>	0.890±0.026
FESL- $c$	0.850±0.030	0.824±0.025	0.732±0.025	0.633±0.032	0.885±0.017
FESL- $s$	0.842±0.033	0.811±0.022	<b>0.733±0.023</b>	0.623±0.035	0.888±0.013
ILFE- $c$	0.865±0.023	0.845±0.027	0.719±0.028	0.707±0.033	0.889±0.023
ILFE- $s$	<b>0.869±0.022</b>	<b>0.847±0.026</b>	0.721±0.029	<b>0.712±0.033</b>	<b>0.894±0.020</b>
Dataset	German	Kr-vs-kp	Svmguide3	HTRU_2	Magic04
NOGD	0.599±0.023	0.531±0.018	0.615±0.024	0.952±0.002	0.515±0.271
NX	<b>0.717±0.023</b>	<b>0.867±0.016</b>	<b>0.779±0.015</b>	<b>0.969±0.004</b>	<b>0.758±0.021</b>
ROGD- $u$	0.706±0.017	0.781±0.016	0.770±0.014	0.968±0.002	0.683±0.004
RX- $u$	<b>0.715±0.015</b>	<b>0.870±0.014</b>	<b>0.776±0.017</b>	<b>0.970±0.005</b>	<b>0.757±0.018</b>
FESL- $c$	0.706±0.017	0.779±0.015	0.768±0.013	0.968±0.002	0.547±0.289
FESL- $s$	0.701±0.020	0.769±0.022	0.770±0.014	0.966±0.005	0.547±0.288
ILFE- $c$	0.719±0.013	0.869±0.014	0.777±0.017	<b>0.971±0.004</b>	<b>0.764±0.015</b>
ILFE- $s$	<b>0.723±0.012</b>	<b>0.870±0.014</b>	<b>0.779±0.016</b>	<b>0.971±0.004</b>	0.763±0.015

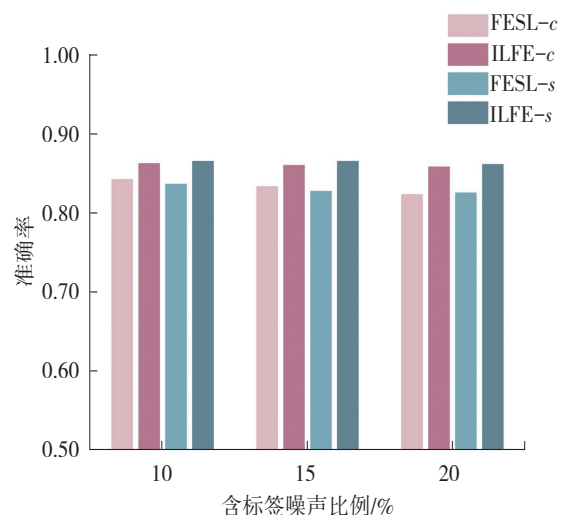
由表 3 可知,本文算法 ILFE- $c$  和 ILFE- $s$  在其中 9 个数据集上的准确率均优于其他对比算法。尤其在 Diabetes、Kr-vs-kp 和 Magic04 数据集上,本文算法 ILFE 比 FESL 分别高出 7.9%、9.1% 和 21.7%;在除 Credit- $g$  外的其他数据集上,本文算法实验结果高出约 0.3%~2.3%。这表明在不同数量级的数据集中 ILFE 算法都能取得不错的分类效果,说明本文算法较好地适应特征演化场景,具有现实应用价值。

实验结果表明:从各自对应网格中可以看出, NX 算法在 10 个数据集上整体表现都优于 NOGD 算法,在 Australian、Kr-vs-kp 数据集上准确率甚至分别高出 42.0% 和 33.6%;而在 Credit- $a$ 、Credit- $g$ 、German 和 Svmguide3 数据集上,均高出 11.0% 以上, RX- $u$  算法在 8 个数据集上相对于 ROGD- $u$  也表现出分类优势。再次体现本文算法对于特征演化数据流的学习效果整体有明显的提升。

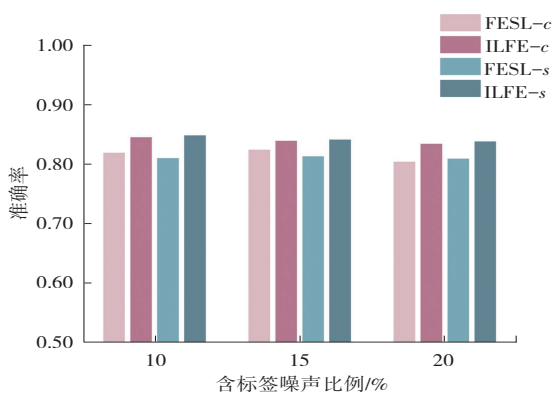
#### 4.3.2 不同比例噪声数据集上对比实验

为了验证本文算法的抗噪声能力,将算法 ILFE- $c$ 、

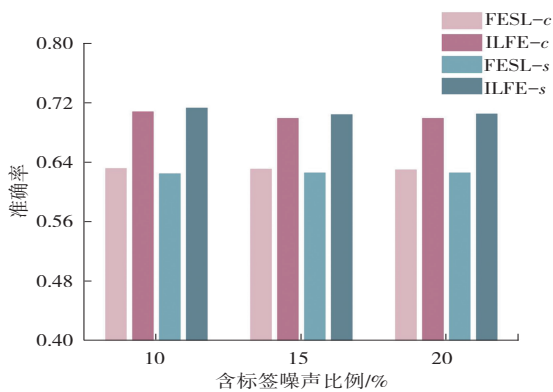
ILFE- $s$ 、FESL- $c$ 、FESL- $s$  部署在不同比例的标签噪声和属性噪声环境中,以分类准确率作为评价标准进行实验,实验结果分别如图 2 和图 3 所示。



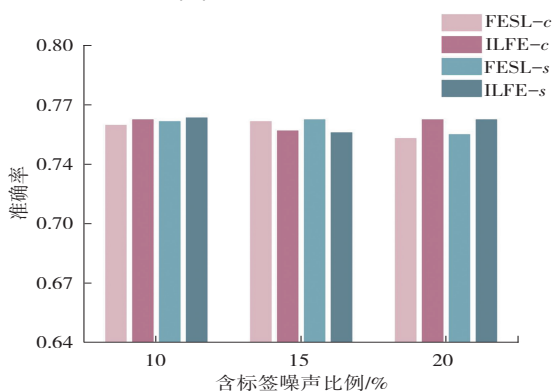
(a) Australian 数据集



(b) Credit-a 数据集



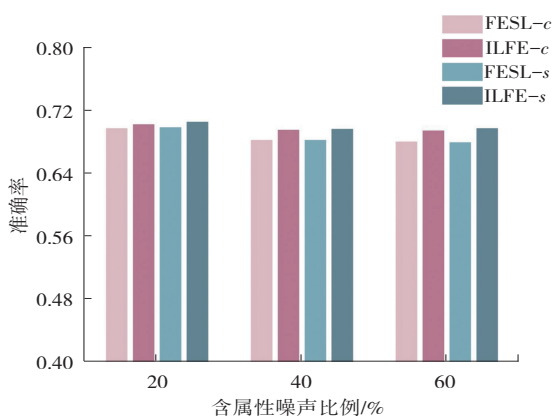
(c) Diabetes 数据集



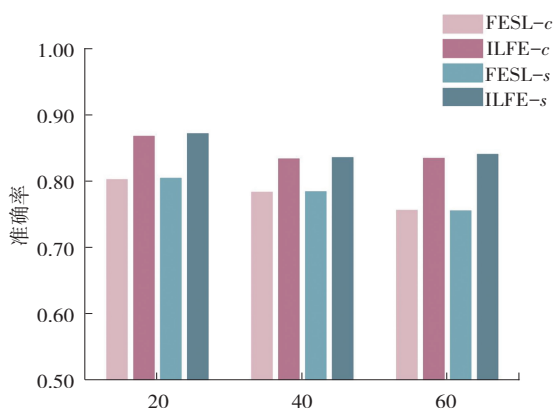
(d) Svmguide3 数据集

图 2 不同比例标签噪声数据集上实验结果

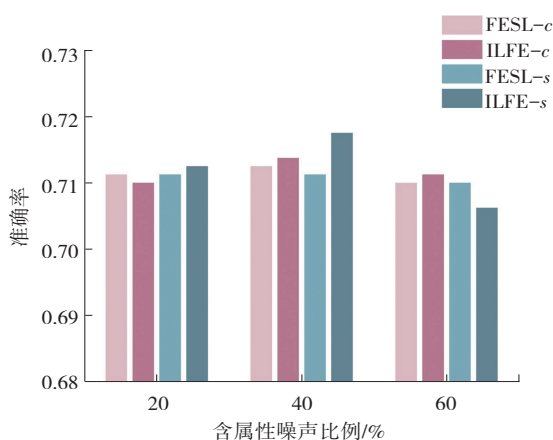
Fig. 2 Experimental results on datasets with different proportions of label noise



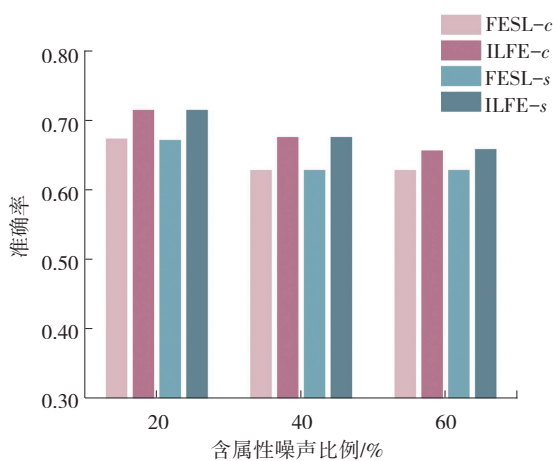
(a) Credit-g 数据集



(b) DNA12 数据集



(c) German 数据集



(d) Kr-vs-kp 数据集

图 3 不同比例属性噪声数据集上实验结果

Fig. 3 Experimental results on datasets with different proportions of attribute noise

图 2 展示了分别含 10%、15% 和 20% 的标签噪声数据集中,对比算法的分类结果。通过分析该柱状图可以看出,在 Australian、Credit-a、Diabetes 和 Svmguide3 这 4 个数据集上,整体 ILFE 系列算法平均准确率高于 FESL 系列算法。其中,在 Diabetes 数据集上这一优势表现更为明显。同时观察发现,随着标签噪声比例的增大,所有算法准确率都会有不同程度的下降,但本文

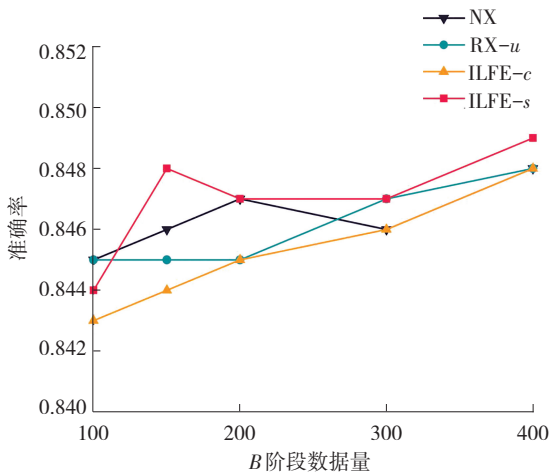


算法相比于 FESL-c 和 FESL-s 下降相对缓慢,甚至在 Australian 数据集上基本呈现平稳趋势。进一步说明,对于给算法分类性能带来更大消极影响的标签噪声,本文方法具有较好的健壮性。

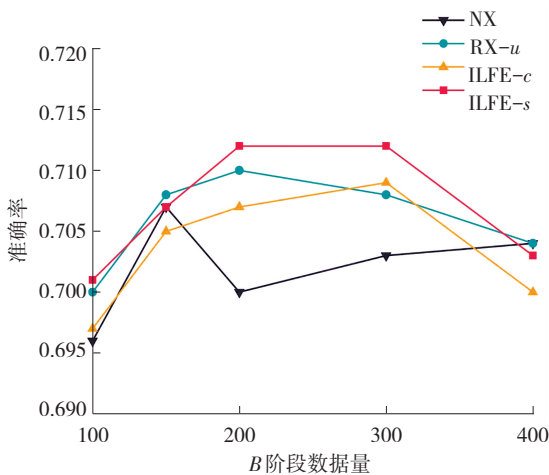
图 3 展示了分别含 20%、40% 和 60% 的属性噪声数据集中,不同对比算法的分类准确率。比较在 Credit-g、DNA12、German 和 Kr-vs-kp 这 4 个数据集上的实验结果,发现随着噪声比例的增大,整体算法性能都会下降,但由于设置的属性噪声干扰较大,有时甚至会出现准确率先减后增、先增后减这样的情况。但是,在有属性噪声干扰的环境中,整体本文算法的学习性能普遍仍高于 FESL-c 和 FESL-s。这表明 ILFE-c 和 ILFE-s 能够同样对属性噪声进行有效的抗噪声处理。

### 4.3.3 不同的 B 阶段数据量的对比实验

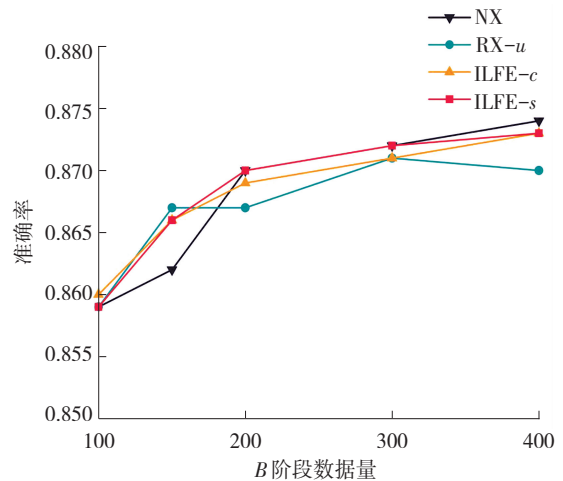
为观察不同的新旧特征共存期,即共现数据出现的阶段 B 长度,对算法分类性能的影响。本文在无噪声干扰的 4 个不同人工数据集中进行实验,分别设置在 B=100,200,300,400,对比本文算法相关的 4 个算法,实验结果如图 4 所示。



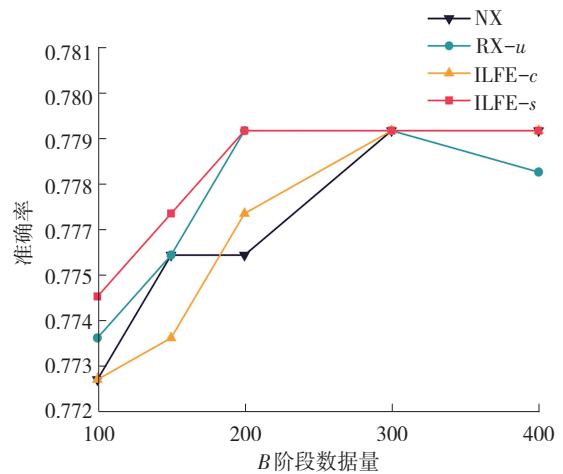
(a) Credit-a 数据集



(b) Diabetes 数据集



(c) Kr-vs-kp 数据集



(d) Svmguide3 数据集

图 4 不同 B 值的对比实验

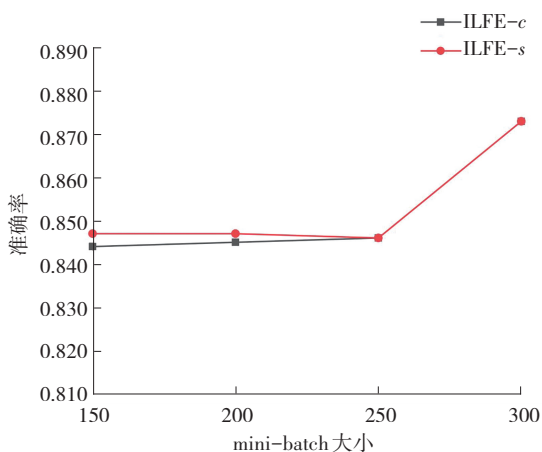
Fig. 4 Experimental results with different B values

通过观察图 4 中折线图,可以发现 ILFE-s 在不同数据量的 B 阶段数据集中平均分类效果较好。而 ILFE-c 在 B 值较小的情况下分类性能较差,这主要是受到该算法受到基分类算法 NX 的影响。对比两个基分类算法 NX 和 RX-u, B 阶段数据量较小时, RX-u 整体表现优于 NX。在现实应用平台中 B 值一般不大,即新旧特征重叠期短,恰恰说明将旧的分类模型投影至新特征下进行重用是有必要的。当然,随着 B 阶段数据量增大,投影的旧模型性能会逐渐下降,而 NX 重新训练的分类模型性能会得到提升,其原因还是用来训练的新特征数据量增多,这也是十分合理的。并且基本在 B 取 300 时,本文算法能够稳中有升地表现出良好的预测效果。

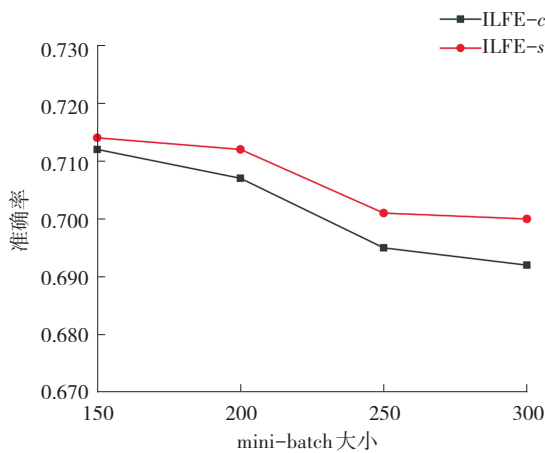
基于上述可知,两个基分类模型性能表现不稳定,因此通过加权组合或者选择最优的集成方法,能够得到更为稳定且高效的分类能力。

### 4.3.4 不同的 mini-batch 大小的对比实验

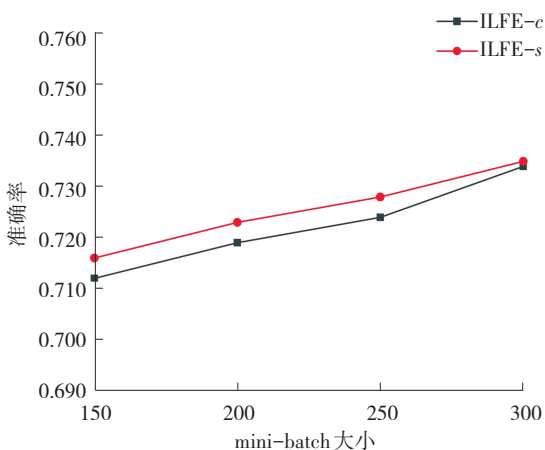
为了进一步分析 mini-batch 的大小对 ILFE-c 和 ILFE-s 算法的影响。本文设置 mini-batch 在 {150, 200, 250, 300} 取值范围中,分别部署两个算法在 Credit-a、Diabetes、German 和 Kr-vs-kp 人工数据集上进行对比实验,结果如图 5 所示。



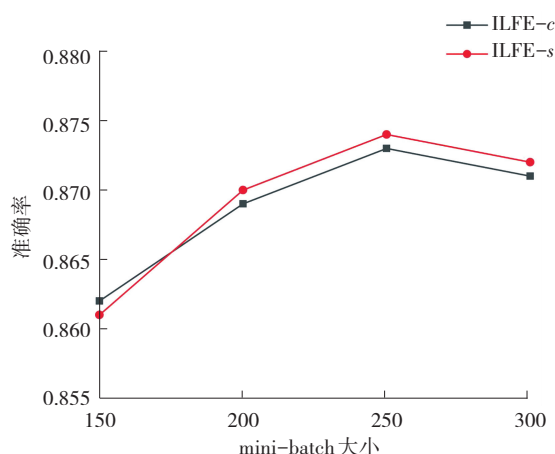
(a) Credit-a 数据集



(b) Diabetes 数据集



(c) German 数据集



(d) Kr-vs-kp 数据集

图 5 不同 mini-batch 大小的对比实验

Fig. 5 Experimental results with different sizes of mini-batch

从图 5 可以看出,在 4 个数据集的实验中 ILFE-s 的平均准确率稍优于 ILFE-c。通过比较不同数据集上折线趋势,除了 Diabetes 数据集的其他数据集上,两个算法的预测准确度一般随着 mini-batch 的增大而增大,但也会出现先增后减甚至一直递减的现象。

因此,说明针对不同数量的数据集仍需要选取合适的 mini-batch 的大小,其取值对预测结果具有一定程度的影响。

## 5 结论

本文提出了一种面向含噪声特征演化流的增量学习算法 ILFE。首先,利用模糊隶属度函数区分噪声与支持向量,增量训练与更新孪生支持向量机分类模型;然后,新特征出现时,利用对应阶段收集的共现数据学习新旧特征之间的映射函数,同时重新训练一个新分类器;继而,在旧特征消失时,通过特征映射的方法实现模型投影至新特征空间,并提出两个新的集成方法:加权组合预测和选择最优预测。最后,大量实验验证了本文算法具有对特征演化数据流自适应学习能力,同时能减少噪声对分类性能的干扰。本文重点关注于新特征出现而旧特征消失这类特征演化场景,而现实应用中的特征演化形式多样而复杂,如何拓展增量学习方法以挖掘特征任意变化的数据流是未来重要研究工作之一。

### 参考文献(References):

[1] 许冠英, 韩萌, 王少峰, 等. 数据流集成分类算法综述[J]. 计算机应用研究, 2020, 37(1): 1-8, 15.

XU Guan-ying, HAN Meng, WANG Shao-feng, et al.

- Summarization of data stream ensemble classification algorithm[J]. *Application Research of Computers*, 2020, 37(1): 1-8, 15.
- [2] 侯博建. 特征变化环境的机器学习方法研究[D]. 南京: 南京大学, 2020.  
HOU Bo-jian. Research on machine learning method in feature changing environment[D]. Nanjing: Nanjing University, 2020.
- [3] WANG C, XIE L, WANG W, et al. Probing into the physical layer: moving tag detection for large-scale RFID systems[J]. *IEEE Transactions on Mobile Computing*, 2020, 19(5): 1200-1215.
- [4] AHMAD QADRI Y, NAUMAN A, ZIKRIA Y B, et al. The future of healthcare Internet of Things: A survey of emerging technologies[J]. *IEEE Communications Surveys & Tutorials*, 2020, 22(2): 1121-1167.
- [5] 张金润, 胡森荣, 洪炎. 基于 LoRa 技术的便携式健康监测系统设计[J]. *重庆工商大学学报(自然科学版)*, 2022, 39(1): 56-61.  
ZHANG Jin-run, HU Sen-rong, HONG Yan. Design of portable health monitoring system based on LoRa technology[J]. *Journal of Chongqing Technology and Business University (Natural Science Edition)*, 2022, 39(1): 56-61.
- [6] 赵旭剑, 杨春明, 李波, 等. 一种基于特征演变的新闻话题演化挖掘方法[J]. *计算机学报*, 2014, 37(4): 819-832.  
ZHAO Xu-jian, YANG Chun-ming, LI Bo, et al. A topic evolution mining algorithm of news text based on feature evolving[J]. *Chinese Journal of Computers*, 2014, 37(4): 819-832.
- [7] HOU B J, ZHANG L, ZHOU Z H. Learning with feature evolvable streams[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 33(6): 2602-2615.
- [8] 刘艳芳, 李文斌, 高阳. 基于被动-主动的特征演化流学习[J]. *计算机研究与发展*, 2021, 58(8): 1575-1585.  
LIU Yan-fang, LI Wen-bin, GAO Yang. Passive-aggressive learning with feature evolvable streams[J]. *Journal of Computer Research and Development*, 2021, 58(8): 1575-1585.
- [9] 刘艳芳, 李文斌, 高阳. 特征演化的置信-加权学习方法[J]. *软件学报*, 2022, 33(4): 1315-1325.  
LIU Yan-fang, LI Wen-bin, GAO Yang. Confidence-weighted learning for feature evolution[J]. *Journal of Software*, 2022, 33(4): 1315-1325.
- [10] HOU C, ZHOU Z H. One-pass learning with incremental and decremental features[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(11): 2776-2792.
- [11] 刘兆清, 古仕林, 侯臣平. 面向特征继承性增减的在线分类算法[J]. *计算机研究与发展*, 2022, 59(8): 1668-1682.  
LIU Zhao-qing, GU Shi-lin, HOU Chen-ping. Online classification algorithm with feature inheritably increasing and decreasing[J]. *Journal of Computer Research and Development*, 2022, 59(8): 1668-1682.
- [12] HOU B J, ZHANG L, ZHOU Z H. Prediction with unpredictable feature evolution[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(10): 5706-5715.
- [13] ZHANG Z, ZHAO P, JIANG Y, et al. Learning with feature and distribution evolvable streams[C]//International Conference on Machine Learning. PMLR, 2020: 11317-11327.
- [14] HOU B J, YAN Y H, ZHAO P, et al. Storage fit learning with feature evolvable streams[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(9): 7729-7736.
- [15] LIAN H, ATWOOD J S, HOU B J, et al. Online deep learning from doubly-streaming data[C]//Proceedings of the 30th ACM International Conference on Multimedia. New York: ACM, 2022: 3185-3194.
- [16] 佟强, 刁恩虎, 李丹, 等. 分类任务中标签噪声的研究综述[J]. *科学技术与工程*, 2022, 22(31): 13626-13635.  
TONG Qiang, DIAO En-hu, LI Dan, et al. Survey of label noise in classification[J]. *Science Technology and Engineering*, 2022, 22(31): 13626-13635.
- [17] MELLO A R, STEMMER M R, KOERICH A L. Incremental and decremental fuzzy bounded twin support vector machine[J]. *Information Sciences*, 2020, 526: 20-38.
- [18] GAO B B, WANG J J, WANG Y, et al. Coordinate descent fuzzy twin support vector machine for classification[C]//Proceedings of the IEEE 14th International Conference on Machine Learning and Applications. Piscataway: IEEE Press, 2015: 7-12.

责任编辑:陈 芳