DOI:10.13718/j.cnki.xdzk.2025.04.020
江山,杨金瑞,武士裕,等. 基于模糊隶属度的最优间隔分布矩阵分类器[J]. 西南大学学报(自然科学版),2025,47(4):230-238.

# 基于模糊隶属度的最优间隔分布矩阵分类器

江山[1], 杨金瑞[2], 武士裕[2], 张里博[2], 杨帆[3]

1. 西南大学 计算机与信息科学学院,重庆 400715;2. 西南大学 人工智能学院,重庆 400715;
3. 重庆市教育信息技术与装备中心,重庆 400020

**摘要**:提出了一种模糊最优间隔分布矩阵分类器(Fuzzy Optimal-margin Distribution Matrix Classifier,FODMC)。该模型通过整合模糊隶属度理论与间隔分布优化机制,实现了矩阵结构信息的有效提取与异常值的鲁棒处理。具体而言,FODMC 采用基于间隔分布的损失函数来优化分类边界,结合核范数正则化策略保持矩阵的低秩特性,并利用交替方向乘子法(Alternating Direction Method of Multipliers,ADMM)实现模型的高效训练。在多个基准数据集上的实验结果表明:与现有方法相比,FODMC 在分类准确率、鲁棒性和泛化能力等方面均展现出显著优势,为矩阵数据分类问题提供了一种有效的解决方案。

**关 键 词**:机器学习;支持矩阵机;支持向量机;间隔分布;模糊隶属度
**中图分类号**:**TP18** **文献标志码**:**A** 开放科学(资源服务)标识码(OSID):
**文 章 编 号**:1673-9868(2025)04-0230-09

# Optimal-margin Distribution Matrix Classifier Based on Fuzzy Membership

JIANG Shan[1], YANG Jinrui[2], WU Shiyu[2], ZHANG Libo[2], YANG Fan[3]

*1. College of Computer and Information Science, Southwest University, Chongqing 400715, China;*
*2. College of Artificial Intelligence, Southwest University, Chongqing 400715, China;*
*3. Chongqing Education Information Technology and Equipment Center, Chongqing 400020, China*

**Abstract**: This study proposes an innovative Fuzzy Optimal-margin Distribution Matrix Classifier (FODMC). The model integrates fuzzy membership theory with margin distribution optimization mechanisms to

收稿日期:2024-11-28
基金项目:国家重点研发计划项目(2021YFB 3101500);国家自然科学基金项目(62106205);大学生创新创业训练计划资助项目(S202410635028)。
作者简介:江山,实验师,主要从事网络工程与信息安全研究。
通信作者:杨帆,中学一级教师。

achieve effective extraction of matrix structural information and robust handling of outliers. Specifically, FODMC employs a margin distribution-based loss function to optimize classification boundaries, incorporates a nuclear norm regularization strategy to maintain the low-rank characteristics of matrices, and utilizes the Alternating Direction Method of Multipliers (ADMM) for efficient model training. Experimental results on multiple benchmark datasets demonstrated that compared to existing methods, FODMC showed significant advantages in classification accuracy, robustness, and generalization capability, providing an effective solution for matrix data classification problems.

**Key words**: machine learning; support matrix machine; support vector machine; margin distribution; fuzzy membership

支持向量机(Support Vector Machine，SVM)作为一种经典的分类算法，自提出以来便在机器学习领域引起了广泛关注。其核心思想是通过最小化正则化项来实现分类误差的最小化，同时最大化两类样本之间的几何间隔[1-2]。然而，基于间隔分布(Margin Distribution，MD)理论的研究表明，相较于单纯优化最大间隔，优化整体间隔分布对于提升模型性能具有更重要的意义[3-4]。在此理论基础上，文献[5]对传统支持向量机进行了重要改进，通过引入间隔分布的均值和方差这两个二阶统计量，提出了大间隔分布机(Large-margin Distribution Machine，LDM)。这一方法不仅确保了所有样本点都能对分类超平面的构建产生影响，而且显著提升了模型的泛化性能。

在现实世界的数据表征中，矩阵形式因其多维信息承载能力已成为灰度图像、多通道脑电信号等复杂数据的标准表示形式[6]。大量实证研究表明，此类矩阵信号普遍呈现出低秩或近似低秩的固有特性[7]。这种结构先验为回归矩阵的低秩建模提供了理论依据。基于此，学界相继提出多种支持矩阵机(Support Matrix Machine，SMM)的改进框架。文献[8]提出了双线性支持向量机，但确定回归矩阵秩的上界在实际操作中较为困难。文献[9]首次将回归矩阵的秩直接纳入目标函数，然而这个做法也使得优化问题变得更加复杂。文献[10]将矩阵的核范数并入正则化的支持矩阵机，在矩阵分类领域表现卓越，并衍生出如迁移最小二乘矩阵机[11]和辛几何矩阵机[12]等多种改进算法。文献[13]延续了核范数正则化的思想，利用高阶张量核范数将支持矩阵机推广到张量数据集上，提出了多分类低秩支持张量机。此外，针对矩数据阵的半监督机器学习，文献[14]基于流形正则化设计了矩阵图嵌入分类器。文献[15]还提出了非平行有界支持矩阵机，利用了一种约束范数组将损失函数有界化。文献[16]最近提出的最优间隔分布矩阵机(Optimal-margin Distribution Matrix Machine，ODMM)在保留矩阵结构信息的同时优化了间隔分布，尽管其分类效果良好，但在抗噪性能方面仍有提升空间。

有效表征样本的可信度是提升矩阵分类器抗噪性能的关键。在实际应用场景中，由于噪声干扰和数据不确定性等因素的影响，不同样本往往具有显著的置信度差异[17]。在这一背景下，模糊集理论因其在处理模糊性和不确定性方面的独特优势而得到广泛应用，其中模糊隶属度函数能够有效量化样本属于特定类别的程度，从而为样本置信水平的评估提供了数学基础[18]。距离类中心较近的样本通常具有较高的可信度，而异常值则表现出较低的隶属度值。尽管模糊 C-均值算法[19]通过基于距离的模糊隶属函数在多种任务中取得了显著成效，但作为无监督的软聚类算法，其难以充分利用监督学习中的标签信息。针对这一局限，文献[20-21]创新性地将模糊隶属度引入监督分类框架，构建了性能优越的监督分类器。在矩阵数据分类领域，文献[22]提出的多分类模糊支持矩阵机通过融合模糊理论，显著提升了模型的分类性能。基于上述研究进展，对 ODMM 进行相应的改进已成为必然趋势。

针对上述问题，本文提出了一种新型矩阵分类器——模糊最优间隔分布矩阵分类器(Fuzzy Optimal-margin Distribution Matrix Classifier，FODMC)。该模型通过引入模糊隶属度机制动态调节各样本对分类器训练的贡献度，同时整合核范数正则化项与基于间隔分布的损失函数，显著提升了模型的泛化性能。

具体而言，FODMC具有以下优势：首先，通过低秩优化有效捕获矩阵数据的结构信息；其次，对训练集中的异常值表现出优异的稳健性。考虑到核范数导致目标函数非光滑的特性，本研究采用交替方向乘子法(Alternating Direction Method of Multipliers，ADMM)对优化问题进行高效求解。通过在多个真实数据集上的系统性实验验证，FODMC在分类准确率等关键指标上均展现出显著优势，充分证明了所提方法的有效性。

## 1 分类算法研究

### 1.1 模糊支持向量机

假设数据集$\{(\boldsymbol{x}_i, y_i) \mid i=1, 2, \cdots, m\}$包含正、负两类样本，其中$\boldsymbol{x}_i \in \mathbb{R}^n$并且$y_i \in \{1, -1\}$，$i=1, 2, \cdots, m$。SVM通过优化如下的损失函数来得到一个分类平面$\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b = 0$：

$$\min_{\omega, b} \frac{1}{2} \| \boldsymbol{\omega} \|^2 + C_0 \sum_{i=1}^{m} \xi_i$$
$$\text{s.t. } \xi_i \geqslant 1 - y_i(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b) \qquad \xi_i \geqslant 0 \qquad i = 1, \cdots, m \tag{1}$$

式中：$\boldsymbol{\omega} \in \mathbb{R}_n$为权重，$b$为偏置，$\xi_i$为第$i$个样本的松弛变量，$C_0$为误分类的惩罚系数。尽管支持向量机是分类问题的有用工具，但它依然忽略了一些问题。有些训练点实际上比其他训练点更重要，因此要求有意义的训练点被正确地分类。假设每个训练点都有一个模糊隶属度$s_i$，表示它属于一个类的可能性。例如，一个训练点有90%可能属于一个类，有10%可能属于另一个类。同样，另一个训练点有20%可能属于一个类，有80%属于另一个类。考虑到支持向量机对噪声非常敏感[23]，文献[24]提出了一种模糊支持向量机模型，该模型将模糊隶属度设置为点与类中心之间距离的函数。假设有一系列训练点：$(y_i, \boldsymbol{x}_i, s_i)$，…，$(y_m, \boldsymbol{x}_m, s_m)$，模糊支持向量机通过模糊隶属度来削弱不确定性高的样本对训练模型的影响，它的模型表示为

$$\min_{\omega, b} \frac{1}{2} \| \boldsymbol{\omega} \|^2 + C_0 \sum_{i=1}^{m} s_i \xi_i$$
$$\text{s.t. } \xi_i \geqslant 1 - y_i(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b) \qquad \xi_i \geqslant 0 \qquad i = 1, \cdots, m \tag{2}$$

### 1.2 支持矩阵机

由于SVM在处理矩阵形式的数据时，会将其扁平化为向量而丢失行列之间的关系。文献[25]提出将矩阵的核范数并入到SVM的正则化项中来利用矩阵数据的结构信息，建立了支持矩阵机(SMM)。考虑矩阵形式的数据集$\{(\boldsymbol{X}_i, y_i) \mid i=1, 2, \cdots, m\}$，其中$\boldsymbol{X}_i$是一个$d_1 \times d_2$的矩阵，$y_i$是它对应的标签。分类函数为$f(\boldsymbol{X}) = \mathrm{sgn}(\langle \boldsymbol{W}, \boldsymbol{X} \rangle + b)$，其中$\boldsymbol{W}$是与$\boldsymbol{X}$大小相同的权重矩阵，$b$是偏置。对于样本$(\boldsymbol{X}_i, y_i)$，它的间隔定义为$\gamma_i = y_i(\langle \boldsymbol{W}, \boldsymbol{X}_i \rangle + b)$。支持矩阵机的模型表示为

$$\min C \sum_{i=1}^{m} L_h(\gamma_i) + \frac{1}{2} \| \boldsymbol{W} \|_F^2 + \tau \| \boldsymbol{W} \|_* \tag{3}$$

式中：$C > 0$，铰链损失函数$L_h(\gamma) = \max\{1-\gamma, 0\}$，$\| \cdot \|_F$是矩阵的Frobenius范数，$\| \cdot \|_*$是矩阵的核范数(Nuclear Norm)，参数$\tau$是用于权衡核范数和其他项之间的惩罚系数。公式(3)中目标函数的后两项之和称为谱弹性网络正则化。

### 1.3 模糊最优间隔分布矩阵分类器模型

本文所提出的FODMC将损失函数定义为间隔分布损失(MDB loss)与谱弹性网络正则化之和，即

$$\min \frac{\lambda}{m} \sum_{i=1}^{m} s_i L_{\mathrm{MDB}}(\gamma_i) + \frac{1}{2} \| \boldsymbol{W} \|_F^2 + \tau \| \boldsymbol{W} \|_* \tag{4}$$

式中：$s_i$为第$i$个样本的模糊隶属度。第一项中的MDB损失$L_{\mathrm{MDB}}$定义如下，它能够优化间隔分布，使得训练数据集的间隔方差尽可能的小，并让间隔均值尽可能大：

$$L_{\mathrm{MDB}}(\gamma)=\begin{cases}\dfrac{(1-\theta-\gamma)^2}{(1-\theta)^2} & \gamma<1-\theta \\ \mu\dfrac{(\gamma-\theta-1)^2}{(1-\theta)^2} & \gamma>1+\theta \\ 0 & \text{其他}\end{cases} \tag{5}$$

式中 $\gamma$ 表示间隔 $\gamma=y(\langle \boldsymbol{W},\boldsymbol{X}\rangle+b)$。而公式(4)中的第二项和第三项之和被称为谱弹性网络正则化，它能够捕获矩阵的结构信息来使得矩阵的秩尽可能低。$\lambda$，$\mu$ 和 $\tau$ 均为正的超参数。当模糊隶属度均为 1 时，FODMC 将退化为 ODMM；当模糊隶属度均为 1 并且 $\tau=0$ 时，FODMC 将退化为 ODM。

为了提高模型的分类性能，引入模糊隶属度 $s_i$。本文的策略是，样本距离类别中心越远，其模糊隶属度就越小。正类和负类的类别中心均定义为所有该类的样本的均值，即

$$\overline{\boldsymbol{X}}_+=\frac{1}{|\{i: y_i>0\}|}\sum_{i: y_i>0}\boldsymbol{X}_i \qquad \overline{\boldsymbol{X}}_-=\frac{1}{|\{i: y_i<0\}|}\sum_{i: y_i<0}\boldsymbol{X}_i \tag{6}$$

同时，定义正、负类别半径为 $R_+=\max\limits_{i: y_i>0}\|\boldsymbol{X}_i-\overline{\boldsymbol{X}}_+\|_F^2$ 和 $R_-=\max\limits_{i: y_i<0}\|\boldsymbol{X}_i-\overline{\boldsymbol{X}}_-\|_F^2$。于是，模糊隶属度被定义为

$$s_i=\begin{cases}1-\dfrac{\|\boldsymbol{X}_i-\overline{\boldsymbol{X}}_+\|_F^2}{R_++\hat{\delta}} & y_i>0 \\ 1-\dfrac{\|\boldsymbol{X}_i-\overline{\boldsymbol{X}}_-\|_F^2}{R_-+\hat{\delta}} & y_i<0\end{cases} \tag{7}$$

式中 $\hat{\delta}$ 是一个很小的正数，用来防止作除法时可能带来的数值问题。

### 1.4　模型求解算法

FODMC 的目标函数是一个非光滑的凸函数，其非光滑性来源于谱弹性正则化中的核范数 $\|\boldsymbol{W}\|_*$。处理这种非光滑函数的一个常用方法是使用交替拉格朗日乘子法(ADMM)。引入辅助变量 $\boldsymbol{Z}$ 将目标函数(4)表示为

$$\begin{gathered}\min_{\boldsymbol{W},b,\boldsymbol{Z}} G_1(\boldsymbol{W},b)+G_2(\boldsymbol{Z}) \\ \text{s. t. } \boldsymbol{Z}-\boldsymbol{W}=\boldsymbol{O}\end{gathered} \tag{8}$$

式中

$$G_1(\boldsymbol{W},b)=\frac{\lambda}{m}\sum_{i=1}^{m}s_iL_{\mathrm{MDB}}(\gamma_i)+\frac{1}{2}\|\boldsymbol{W}\|_F^2 \qquad G_2(\boldsymbol{Z})=\tau\|\boldsymbol{Z}\|_* \tag{9}$$

于是，可以给出其增广拉格朗日函数

$$L(\boldsymbol{W},b,\boldsymbol{Z};\boldsymbol{\Lambda})=G_1(\boldsymbol{W},b)+G_2(\boldsymbol{Z})+\langle\boldsymbol{\Lambda},\boldsymbol{Z}-\boldsymbol{W}\rangle+\frac{\rho}{2}\|\boldsymbol{Z}-\boldsymbol{W}\|_F^2 \tag{10}$$

式中：$\boldsymbol{\Lambda}$ 是拉格朗日乘子，$\rho>0$ 是违反约束的乘法项的系数。在分离出非光滑项并消除硬约束后，将交替地关于 $\boldsymbol{W}$，$b$ 和 $\boldsymbol{Z}$ 最小化增广拉格朗日函数，并在每一步的最后更新拉格朗日乘子。

**定理 1**　关于 $\boldsymbol{Z}$ 对增广拉格朗日函数最小化，即

$$\arg\min_{\boldsymbol{Z}} G_2(\boldsymbol{Z})+\langle\boldsymbol{\Lambda},\boldsymbol{Z}\rangle+\frac{\rho}{2}\|\boldsymbol{W}-\boldsymbol{Z}\|_F^2 \tag{11}$$

它的最优解是

$$\boldsymbol{Z}^*=\frac{1}{\rho}\boldsymbol{D}_\tau(\rho\boldsymbol{W}-\boldsymbol{\Lambda}) \tag{12}$$

**证**　奇异值阈值算子 $\boldsymbol{D}_\tau$ 的定义与详细证明过程参见文献[16]。

关于对 $\boldsymbol{W}$，$b$ 的优化，首先将损失函数转化为以下的等价形式：

$$\min_{\boldsymbol{W},b,\xi,\varepsilon} \frac{\lambda}{m}\sum_{i=1}^{m} s_i \frac{\xi_i^{\ 2}+\mu\varepsilon_i^{\ 2}}{(1-\theta)^2}+\frac{1}{2}\|\boldsymbol{W}\|_F^2-\langle\boldsymbol{\Lambda},\boldsymbol{W}\rangle+\frac{\rho}{2}\|\boldsymbol{W}-\boldsymbol{Z}\|_F^2$$

$$\text{s. t. } 1-\theta-\xi_i \leqslant y_i(\langle\boldsymbol{W},\boldsymbol{X}_i\rangle+b)\leqslant 1+\theta+\varepsilon_i \qquad i=1,2,\cdots,m \tag{13}$$

式中 $\xi_i$ 和 $\varepsilon_i$ 是由 $L_{\mathrm{MDB}}$ 转变来的松弛变量。这是一个凸优化问题。引入新的拉格朗日乘子 $\zeta_i>0$ 和 $\beta_i>0$，并给出(13)式的拉格朗日函数，即

$$L_1(\boldsymbol{W},b,\boldsymbol{\xi},\boldsymbol{\varepsilon};\boldsymbol{\zeta},\boldsymbol{\beta})=\frac{\lambda(\boldsymbol{\xi}^{\mathrm{T}}\mathrm{diag}(\boldsymbol{s})\boldsymbol{\xi}+\mu\boldsymbol{\varepsilon}^{\mathrm{T}}\mathrm{diag}(\boldsymbol{s})\boldsymbol{\varepsilon})}{m(1-\theta)^2}+\frac{1}{2}\|\boldsymbol{W}\|_2^2-\langle\boldsymbol{\Lambda},\boldsymbol{W}\rangle+\frac{\rho}{2}\|\boldsymbol{W}-\boldsymbol{Z}\|_2^2+$$

$$\sum_{i=1}^{m}\zeta_i(1-\theta-\xi_i-y_i(\langle\boldsymbol{W},\boldsymbol{X}_i\rangle+b))+\sum_{i=1}^{m}\beta_i(y_i(\langle\boldsymbol{W},\boldsymbol{X}_i\rangle+b)-1-\theta-\varepsilon_i) \tag{14}$$

式中：$\boldsymbol{\xi}=(\zeta_1,\cdots,\zeta_m)^{\mathrm{T}}$，$\boldsymbol{\omega}=(\omega_1,\cdots,\omega_m)^{\mathrm{T}}$，$\boldsymbol{s}=(s_1,\cdots,s_m)^{\mathrm{T}}$。将拉格朗日函数关于原始变量的偏导数置零后，可以得到

$$\boldsymbol{W}=\frac{1}{1+\rho}(\boldsymbol{\Lambda}+\rho\boldsymbol{Z}+\sum_{i=1}^{m}(\zeta_i-\beta_i)y_i\boldsymbol{X}_i)$$

$$\boldsymbol{\xi}=\frac{m(1-\theta)^2\boldsymbol{\zeta}}{2\lambda\boldsymbol{s}} \qquad \boldsymbol{\varepsilon}=\frac{m(1-\theta)^2\boldsymbol{\beta}}{2\mu\lambda\boldsymbol{s}} \qquad \sum_{i=1}^{m}y_i(\zeta_i-\beta_i)=0。 \tag{15}$$

这里向量之间的除法是逐元素进行的。于是，忽略常数项，该对偶问题将化为如下的形式：

$$L_1=\frac{(\hat{\rho}_-\boldsymbol{e}+\boldsymbol{u})^{\mathrm{T}}}{(1+\rho)}\boldsymbol{\zeta}-\frac{m(1-\theta)^2}{4\lambda\mu}\left(\left(\frac{\boldsymbol{\zeta}}{\boldsymbol{s}}\right)^{\mathrm{T}}\left(\frac{\boldsymbol{\zeta}}{\boldsymbol{s}}\right)+\left(\frac{\boldsymbol{\beta}}{\boldsymbol{s}}\right)^{\mathrm{T}}\left(\frac{\boldsymbol{\beta}}{\boldsymbol{s}}\right)\right)+\frac{(\hat{\rho}_+\boldsymbol{s}-\boldsymbol{u})^{\mathrm{T}}\boldsymbol{\beta}}{(1+\rho)\boldsymbol{s}}$$

$$+\frac{1}{2(1+\rho)}\left(\frac{\boldsymbol{\zeta}-\boldsymbol{\beta}}{\boldsymbol{s}}\right)^{\mathrm{T}}\boldsymbol{K}\left(\frac{\boldsymbol{\zeta}-\boldsymbol{\beta}}{\boldsymbol{s}}\right) \tag{16}$$

式中：矩阵 $\boldsymbol{K}$ 的元素为 $K_{ij}=s_iy_is_jy_j\langle\boldsymbol{X}_i,\boldsymbol{X}_j\rangle$，向量 $\boldsymbol{u}$ 的各分量为 $u_i=s_iy_i\langle\boldsymbol{\Lambda}+\rho\boldsymbol{Z},\boldsymbol{X}_i\rangle$，常数 $\hat{\rho}_\pm=(1+\rho)(\theta\pm 1)$。

为了简洁起见，将 $\boldsymbol{\zeta}$ 和 $\boldsymbol{\beta}$ 合并为一个变量，使得 $\boldsymbol{\alpha}^{\mathrm{T}}=\left[\left(\frac{\boldsymbol{\zeta}}{\boldsymbol{s}}\right)^{\mathrm{T}},\left(\frac{\boldsymbol{\beta}}{\boldsymbol{s}}\right)^{\mathrm{T}}\right]$。于是 $\boldsymbol{\zeta}=[\mathrm{diag}(\boldsymbol{s}),\boldsymbol{0}]\boldsymbol{\alpha}$，$\boldsymbol{\beta}=[\boldsymbol{0},\mathrm{diag}(\boldsymbol{s})]\boldsymbol{\alpha}$，并且 $\boldsymbol{\zeta}-\boldsymbol{\beta}=[\mathrm{diag}(\boldsymbol{s}),-\mathrm{diag}(\boldsymbol{s})]\boldsymbol{\alpha}$。最终，(13)式的对偶问题变为

$$\min_{\boldsymbol{\alpha}}\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}K\boldsymbol{\alpha}+\boldsymbol{p}^{\mathrm{T}}\boldsymbol{\alpha} \qquad \text{s. t. } \boldsymbol{\alpha}\geqslant 0 \tag{17}$$

式中 $\boldsymbol{H}\in\mathbb{R}^{2m\times 2m}$ 和 $\boldsymbol{p}\in\mathbb{R}^{2m}$ 定义为

$$\boldsymbol{H}=\begin{bmatrix}\boldsymbol{K}+\dfrac{m\hat{\rho}_-^{\ 2}}{2\lambda(1+\rho)}\boldsymbol{I} & -\boldsymbol{K}\\ -\boldsymbol{K} & \boldsymbol{K}+\dfrac{m\hat{\rho}_-^{\ 2}}{2\lambda\mu(1+\rho)}\boldsymbol{I}\end{bmatrix} \qquad \boldsymbol{p}=\begin{bmatrix}\hat{\rho}_-\boldsymbol{s}+\boldsymbol{u}\\ \hat{\rho}_+\boldsymbol{s}-\boldsymbol{u}\end{bmatrix} \tag{18}$$

这里 $\boldsymbol{I}$ 代表 $m\times m$ 的单位矩阵，$\boldsymbol{e}$ 代表所有元素是 1 的 $n$ 维向量。因此，可以得到如下定理。

**定理 2**　假设 $\boldsymbol{\alpha}^*\in\mathbb{R}^{2m}$ 是(17)式的解。令 $m^*=|\{i:\xi_i>0 \text{ 或 } \varepsilon_i>0\}|$，$\boldsymbol{v}=\boldsymbol{Y}[\mathrm{diag}(\boldsymbol{s}),-\mathrm{diag}(\boldsymbol{s})]\boldsymbol{\alpha}^*$，其中 $\boldsymbol{Y}=\mathrm{diag}(y_1,\cdots,y_m)$。于是(13)式的一个最优解是

$$\boldsymbol{W}^*=\frac{\sum_{i=1}^{m}v_i\boldsymbol{X}_i+\boldsymbol{\Lambda}+\rho\boldsymbol{Z}}{(1+\rho)} \tag{19}$$

$$b^*=\frac{1}{m^*}\sum_{i:\xi_i>0}(y_i(1-\theta-\xi_i)-\langle\boldsymbol{W},\boldsymbol{X}_i\rangle)+\frac{1}{m^*}\sum_{i:\varepsilon_i>0}(y_i(1+\theta+\varepsilon_i)-\langle\boldsymbol{W},\boldsymbol{X}_i\rangle) \tag{20}$$

在求解二次规划(17)后，就可以通过(19)，(20)式来更新 $\boldsymbol{W}$ 和 $b$。随后，使用以下公式来更新乘子 $\boldsymbol{\Lambda}$：

$$\boldsymbol{\Lambda}^{(k+1)}=\boldsymbol{\Lambda}^{(k)}+\rho(\boldsymbol{Z}^{(k)}-\boldsymbol{W}^{(k)}) \tag{21}$$

完整的迭代优化过程见算法 1。

---

算法 1　FODMC 的训练算法

---

输入：训练集$\{(\boldsymbol{X}_i, y_i) \mid i = 1, 2, \cdots, m\}$，超参数 $\lambda$，$\tau > 0$ 和 $0 < \mu$，$\theta < 1$

输出：$\boldsymbol{W}$，$b$

1. 初始化 $\boldsymbol{Z}^{(0)} \in \mathbb{R}^{d_1 \times d_2}$，$\boldsymbol{\Lambda}^{(0)} = \in \mathbb{R}^{d_1 \times d_2}$，$\rho > 0$，$t^{(1)} = 1$
2. 通过公式(7)计算每个样本的模糊隶属度 $s_i$，$i = 1, 2, \cdots, m$：

$$s_i = \begin{cases} 1 - \dfrac{\| \boldsymbol{X}_i - \overline{\boldsymbol{X}}_+ \|_F^2}{R_+ + \hat{\delta}} & y_i > 0 \\ 1 - \dfrac{\| \boldsymbol{X}_i - \overline{\boldsymbol{X}}_- \|_F^2}{R_- + \hat{\delta}} & y_i < 0 \end{cases}$$

3. for $k \leftarrow 1$ to 迭代次数
4. 使用公式(19)更新 $\boldsymbol{W}^{(k)}$：

$$\boldsymbol{W}^{(k)} = \frac{\sum_{i=1}^{m} \{ v_i \boldsymbol{X}_i \} + \boldsymbol{\Lambda}^{(k-1)} + \rho \boldsymbol{Z}^{(k-1)}}{(1+\rho)}$$

5. 使用公式(20)更新 $b^{(k)}$：

$$b^{(k)} = \frac{1}{m^*} \sum_{i:\xi_i > 0} (y_i(1-\theta-\xi_i) - \langle \boldsymbol{W}^{(k)}, \boldsymbol{X}_i \rangle) + \frac{1}{m^*} \sum_{i:\varepsilon_i > 0} (y_i(1+\theta+\varepsilon_i) - \langle \boldsymbol{W}^{(k)}, \boldsymbol{X}_i \rangle)$$

6. 使用公式(12)更新 $\boldsymbol{Z}^{(k)}$：$\boldsymbol{Z}^{(k)} = \dfrac{1}{\rho} \boldsymbol{D}_\tau (\rho \boldsymbol{W}^{(k)} - \boldsymbol{\Lambda}^{(k-1)})$
7. 使用公式(21)更新 $\boldsymbol{\Lambda}^{(k)}$：$\boldsymbol{\Lambda}^{(k)} = \boldsymbol{\Lambda}^{(k-1)} + \rho(\boldsymbol{Z}^{(k-1)} - \boldsymbol{W}^{(k-1)})$
8. end for
9. return $\boldsymbol{W}^{(k)}$，$b^{(k)}$

---

## 2　实验结果与分析

### 2.1　实验设置

为了通过实验验证模型的性能，选取 SVM、ODM、SMM 和 ODMM 作为对照模型。为确保实验的公正性，采用了五折交叉验证结合网格搜索策略，以筛选出最优的超参数组合。具体操作如下：首先，为模型的各个参数设定了一系列候选值；接着，通过网格搜索法探寻这些参数的所有可能组合。在此过程中，利用五折交叉验证来评估每组参数组合的表现，从而在多个数据子集中准确衡量其准确性。依据五折交叉验证的结果，筛选出达到最高准确率的超参数组合，作为针对该数据集的模型最佳配置。这种严谨的调参方法确保了模型能够针对特定数据集实现最优性能。超参数的设置沿用文献[16]的设置。对于 SVM 和 SMM 的惩罚系数 $C$，从$\{2^i \mid i = 0, 1, \cdots, 10\}$中搜索。而关于 FODMC、ODMM 以及 ODM，$\lambda$ 选择自$\{2^i \mid i = 0, 2, 4, 6, 8\}$，并且 $\mu$ 和 $\theta$ 两者都选择自$\{0.2, 0.4, 0.6, 0.8\}$。而核范数的乘法系数 $\tau$ 则从$\{2^i \mid i = 0, 2, 4, 6, 8\}$中搜索。在所有使用 ADMM 的模型中，都把算法的惩罚系数 $\rho$ 设置为 1。实验平台为 MATLAB R2021b 以及具有 3.3 GHz 的 CPU 和 32 GB 的内存的个人电脑。

### 2.2　在真实数据集上的实验结果

为了评估 FODMC 的性能，从 Kaggle 网站(https://www.kaggle.com/)上收集了 8 个来自真实世界的数据集。这些数据集中的每个样本都是矩阵形式的，代表了图片或者是多通道的时间序列，并且每个样本都具有相应的类别标签。由于本文仅探究二分类任务，因此具有多个类别标签的数据集只有两类样本被挑选来进行实验。在包含不同大小图像的数据集中，所有图像都被统一缩放到统一的大小。数据集中的每个样本都提取了图像的灰度像素，从而得到一个与相应图像大小相同的样本矩阵。表 1 总结了每

个数据集的正类和负类样本数量以及样本矩阵的维度。每个数据集被分为一个训练集和一个测试集，训练集包含80%的数据，测试集包含剩余20%的数据。

**表1 数据集信息**

| 数据集 | 尺寸 | 正类/负类 | 描述 |
|---|---|---|---|
| Animals | 256×256 | 500/500 | 猫、狗和野生动物的面部照片[25]，狗和猫分别为正类和负类 |
| Gender | 110×90 | 500/500 | 该数据集由男性和女性的面部照片组成 |
| Alcohol | 256×64 | 475/461 | 包含酗酒者和对照组两类受试者的脑电图信号 |
| Concrete | 227×227 | 250/250 | 此数据集包含有或没有裂缝的混凝土表面图像 |
| Flowers | 512×512 | 108/114 | 两种花的图像，包含大丽花和天竺葵 |
| Fashion | 224×224 | 500/500 | 两类女性面部图像：时装模特和普通女性 |
| Instruments | 224×224 | 500/500 | 乐器的图像，包含架子鼓和班卓琴 |
| Balls | 200×200 | 400/400 | 不同体育运动球的图像排球和羽毛球分别被选为正负类样本 |

本文选择了准确率 $I$ 和F1分数 $F$ 作为评估指标。令 $T_P$，$T_N$，$F_P$，$F_N$ 分别表示真正类、真负类、假正类、假负类的样本数量，定义精准度 $p=\dfrac{T_P}{T_P+F_P}$ 和召回率 $r=\dfrac{T_P}{T_P+F_N}$，则准确率和F1分数由如下公式计算：

$$I=\frac{T_P+T_N}{T_P+T_N+F_P+F_N}\times 100\% \qquad F=\frac{2pr}{p+r}\times 100\% \tag{22}$$

实验结果如表2所示。显然，我们提出的FODMC在几乎所有数据集上都取得了最高的准确率，并且在大部分数据集上的F1分数也领先于其他模型。从这个表中还可以看出，SMM和SVM没有优化间隔分布，导致了它们的准确率低于其他基于间隔分布的模型，这凸显了优化间隔分布对于提高模型泛化能力的重要性。与FODMC相比，虽然ODMM也用了基于间隔分布的损失函数，并且同样通过核范数来捕获矩阵数据的结构信息，但是由于它未考虑每个样本对模型训练的重要程度，因此ODMM的平均准确率和平均F1分数均不及本文所提出的FODMC，表明使用模糊隶属度来控制每个样本对训练模型的影响是有意义的。

**表2 在真实数据集上的实验结果** /%

| 数据集 | FODMC | | ODMM | | SMM | | SVM | | ODM | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 准确率 | F1分数 | 准确率 | F1分数 | 准确率 | F1分数 | 准确率 | F1分数 | 准确率 | F1分数 |
| Animals | 85.00 | 84.88 | 81.00 | 81.45 | 81.00 | 80.98 | 79.50 | 78.79 | 80.00 | 80.00 |
| Gender | 82.50 | 82.50 | 82.00 | 83.17 | 75.50 | 72.53 | 75.50 | 72.53 | 76.00 | 76.00 |
| Alcohol | 90.48 | 92.68 | 88.89 | 90.90 | 89.42 | 91.65 | 89.42 | 92.12 | 89.95 | 89.95 |
| Concrete | 91.00 | 91.79 | 92.00 | 91.37 | 92.00 | 88.46 | 88.00 | 82.57 | 91.00 | 87.00 |
| Flowers | 84.44 | 85.54 | 75.56 | 72.67 | 77.78 | 83.44 | 82.22 | 84.39 | 84.44 | 87.13 |
| Fashion | 73.00 | 70.46 | 72.50 | 69.07 | 69.00 | 64.55 | 69.00 | 65.15 | 68.00 | 64.11 |
| Instruments | 91.30 | 92.07 | 84.06 | 84.06 | 81.16 | 88.14 | 85.51 | 90.64 | 84.06 | 88.24 |
| Balls | 76.88 | 76.46 | 76.25 | 73.22 | 74.38 | 77.86 | 74.38 | 74.38 | 73.12 | 74.56 |
| 平均值 | 84.33 | 84.55 | 81.53 | 80.74 | 80.03 | 80.95 | 80.44 | 80.07 | 80.82 | 80.87 |

### 2.3 模型稳健性分析

为了进一步证实所提出的方法能够提高分类器的稳健性，在具有异常值的数据集上评估模型的性能。先在数据集中添加不同比例的标签噪声，即将一部分的样本标签设置为错误的标签，然后再训练分类器并评估其分类效果。依旧使用前文提到的8个真实数据集，同时为训练集设置10%，20%，30%，40%和

50%的异常值。图 1 展示了每个模型在所有数据集上的平均准确率随异常值比例的变化。显然，每个模型的准确率都逐步下降的趋势。然而，本文所提出的 FODMC 比其他模型的性能更好。表 3 统计了每个模型在各个数据集上不同噪声水平的平均准确率。容易发现，FODMC 在大多数数据集上都具有优势，并且平均值也最高。

**表 3　在具有标签噪声的数据集上的平均准确率**　　/%

| 数据集 | FODMC | ODMM | SMM | SVM | ODM |
|---|---|---|---|---|---|
| Animals | 68.70 | 67.87 | 63.80 | 63.90 | 64.17 |
| Gender | 68.40 | 69.40 | 62.03 | 62.10 | 62.63 |
| Alcohol | 62.62 | 59.92 | 61.38 | 61.45 | 62.62 |
| Concrete | 84.07 | 78.93 | 67.93 | 68.87 | 69.47 |
| Flowers | 62.52 | 62.81 | 62.52 | 62.52 | 64.00 |
| Fashion | 62.13 | 61.33 | 58.20 | 58.07 | 58.97 |
| Instruments | 64.66 | 64.17 | 60.20 | 61.36 | 60.49 |
| Balls | 64.13 | 60.25 | 53.54 | 53.54 | 54.71 |
| 平均值 | 67.15 | 65.59 | 61.20 | 61.48 | 62.13 |

## 3　结论

本文提出的模糊最优间隔分布矩阵分类器(FODMC)通过引入模糊隶属度机制，有效减轻了噪声对模型训练的干扰，显著提升了模型的稳健性。FODMC 利用核范数来挖掘矩阵数据的内在结构信息，并通过优化间隔分布策略进一步增强了模型的泛化能力，确保了对矩阵形式数据集进行分类时的有效性。FODMC 目标函数的非光滑性使得一般的优化算法不再适用，为了解决这一问题，设计了基于交替乘子法的训练算法。通过在真实数据集上的综合实验，成功地验证了所提方法在分类性能上的优秀表现。此外，本文还在含有标签噪声的数据集上对模型的抗噪能力进行了深入评估。实验结果一致表明，方法在面对噪声干扰时展现出了最佳的稳健性，这一特性使得 FODMC 在实际工程应用中更具竞争力和实用价值。总体而言，FODMC 的分类性能和抗噪能力均得到了实验的充分验证，为矩阵数据分类提供了新的研究视角和解决方案。未来的研究方向包括将 FODMC 拓展至多分类问题，以及确定泛化误差上界等理论性能分析。

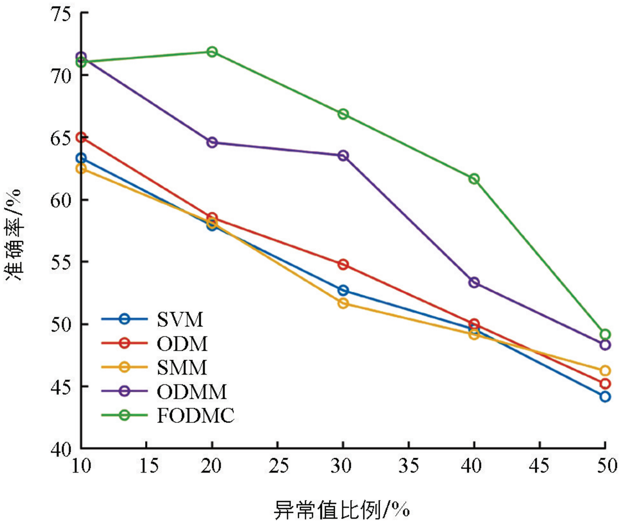


**图 1　每个模型在不同程度噪声下的准确率**

**参考文献：**

[1]　汪海燕，黎建辉，杨风雷. 支持向量机理论及算法研究综述 [J]. 计算机应用研究，2014，31(5)：1281-1286.

[2]　李春祥，殷潇. 基于小波支持向量机的非高斯空间风压内外插预测 [J]. 上海交通大学学报，2018，52(11)：1516-1523.

[3]　GAO W，ZHOU Z H. On the Doubt about Margin Explanation of Boosting [J]. Artificial Intelligence，2013，203：1-18.

[4]　WANG L W，SUGIYAMA M，JING Z X，et al. A Refined Margin Analysis for Boosting Algorithms via Equilibrium

Margin [J]. Journal of Machine Learning Research, 2011, 12: 1835-1863.
[5] ZHANG T, ZHOU Z H. Large Margin Distribution Machine [C] //Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2014: 313-322.
[6] 武柏芸. 支持矩阵机 $I_{0/1}$-ADMM 算法研究 [D]. 北京: 北京交通大学, 2021.
[7] 潘海洋, 徐海锋, 郑近德, 等. 基于双加权不平衡矩阵分类器的机械故障诊断方法 [J]. 机械工程学报, 2024, 60(3): 170-180.
[8] PIRSIAVASH H, RAMANAN D, FOWLKES C. Bilinear Classifiers for Visual Recognition [C] //23rd International Conference on Neural Information Processing Systems. New York: ACM, 2009.
[9] WOLF L, JHUANG H, HAZAN T. Modeling Appearances with Low-Rank SVM [C] //2007 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2007.
[10] LUO L, XIE Y, ZHANG Z, et al. Support Matrix Machines [J]Journal of Machine Learning Research, 2015, 37: 938-947.
[11] 伍毅, 盛丽, 潘海洋, 等. 基于迁移最小二乘支持矩阵机的滚动轴承故障诊断方法 [J]. 振动与冲击, 2022, 41(21): 53-59.
[12] 潘海洋. 基于辛几何模态分解和支持矩阵机的机械故障诊断方法 [D]. 长沙: 湖南大学, 2019.
[13] YANG J R, FAN S Y, ZHANG L B, et al. A Low-Rank Support Tensor Machine for Multi-Classification [J]. Information Sciences, 2025, 688: 121398.
[14] PAN H, XU H, ZHENG J, et al. A Semi-Supervised Matrixized Graph Embedding Machine for Roller Bearing Fault Diagnosis Under Few-Labeled Samples [J]. IEEE Transactions on Industrial Informatics, 2024, 20(1): 854-863.
[15] PAN H, XU H, ZHENG J, et al. Non-Parallel Bounded Support Matrix Machine and Its Application in Roller Bearing Fault Diagnosis [J]. Information Sciences, 2023, 624: 395-415.
[16] YANG J, FAN S, LIU L, et al. Optimal Margin Distribution Matrix Machine [J]. Expert Systems with Applications, 2024, 240: 122497.
[17] 张法滢, 吕莉, 韩龙哲, 等. 直觉模糊的结构化最小二乘孪生支持向量机 [J]. 应用科学学报, 2024, 42(2): 350-363.
[18] 张翔, 肖小玲, 徐光祐. 模糊支持向量机中隶属度的确定与分析 [J]. 中国图象图形学报, 2006, 11(8): 1188-1192.
[19] 孙晓霞, 刘晓霞, 谢倩茹. 模糊 C-均值(FCM)聚类算法的实现 [J]. 计算机应用与软件, 2008, 25(3): 48-50.
[20] DONG D H, FENG M Y, KURTHS J, et al. Fuzzy Large Margin Distribution Machine for Classification [J]. International Journal of Machine Learning and Cybernetics, 2024, 15(5): 1891-1905.
[21] JIN Q, FAN S, DONG D, et al. Fuzzy Twin Bounded Large Margin Distribution Machines [C] // 5th Chinese Conference on Pattern Recognition and Computer Vision(PRCV). Berlin: Springer, 2022.
[22] PAN H Y, XU H F, ZHENG J D, et al. Multi-Class Fuzzy Support Matrix Machine for Classification in Roller Bearing Fault Diagnosis [J]. Advanced Engineering Informatics, 2022, 51: 101445.
[23] ATLA A, TADA R, SHENG V, et al. Sensitivity of Different Machine Learning Algorithms to Noise [J]. Journal of Computing Sciences in Colleges, 2011, 26(5): 96-103.
[24] LIN C F, WANG S D. Fuzzy Support Vector Machines [J]. IEEE Trans Neural Netw, 2002, 13(2): 464-471.
[25] CHOI Y, UH Y, YOO J, et al. Stargan v2: Diverse Image Synthesis for Multiple Domains [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2020. New York: IEEE Press, 2020.

**责任编辑** 张枸