

# 基于粒球计算的多粒度支持向量回归算法

华有霖<sup>1</sup>, 邵亚斌<sup>1,2\*</sup>, 朱学勤<sup>1</sup>

(1.重庆邮电大学理学院, 重庆 400065; 2.网络空间大数据智能安全教育部重点实验室, 重庆 400065)

**摘要:**为了实现支持向量回归算法的高效性和鲁棒性,本文将多粒度粒球计算融合到支持向量回归算法中,提出了一种基于粒球计算的多粒度粒球支持向量回归算法。该算法将粒球中的半径信息加入到约束条件中,将原本基于样本点的支持向量算法替换为基于粒球的支持向量回归算法。同时,本文研究了多粒度粒球支持向量回归机的对偶模型。实验结果表明,采用人工数据集和加州大学欧文分校(University of California-Irvine, UCI)公开数据集时,多粒度粒球支持向量回归机的计算效率和鲁棒性均得到提升。

**关键词:**数据挖掘;粒计算;多粒度表示;多粒度粒球计算;支持向量回归机

**中图分类号:**TP391 **文献标志码:**A

**引用格式:**华有霖,邵亚斌,朱学勤. 基于粒球计算的多粒度支持向量回归算法[J]. 山东大学学报(理学版),2025,60(7):104-115.

## Multi-granularity support vector regression algorithm based on granular ball computing

HUA Youlin<sup>1</sup>, SHAO Yabin<sup>1,2\*</sup>, ZHU Xueqin<sup>1</sup>

(1. School of Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. Key Laboratory of Cyberspace Big Data Intelligent Security, Ministry of Education, Chongqing 400065, China)

**Abstract:** To achieve both efficiency and robustness in the support vector regression algorithm, multi-granularity granular ball computing is integrated into the support vector regression algorithm. A multi-granularity granular ball support vector regression algorithm is proposed based on granular ball computing. The radius information from granular balls is incorporated into the model's constraint conditions, replacing the traditional sample point based support vector regression algorithm with a granular ball based support vector regression algorithm. Additionally, the dual model of the multi-granularity granular ball support vector regression is investigated, and a particle swarm optimization algorithm is utilized to solve it. Experimental results show that on both artificial datasets and University of California-Irvine(UCI) publicly available datasets, computational efficiency and robustness are improved by the multi-granularity granular ball support vector regression.

**Key words:** data mining; granular computing; multi-granularity representation; multi-granularity granular ball computing; support vector regression

## 0 引言

支持向量回归机<sup>[1]</sup>(support vector regression, SVR)是一种基于统计学理论和 VC 维(vapnik chervonenkis dimension, VC)理论的回归模型。作为一种强大的模式识别方法,SVR 已广泛应用在时间序列预测等各个领域<sup>[2-3]</sup>。SVR 的求解本质上是解决一个凸二次优化问题(quadratic programming problem, QPP)。尽管 SVR 可以取得较好的拟合效果,但是计算复杂度较高,尤其是处理大规模数据集时更为明显。于是针对如何提升 SVR 的计算效率问题提出了很多解决方法<sup>[4]</sup>,例如 Peng 等<sup>[5]</sup>提出的孪生支持向量回归机(twin support vector regression, TSVR)对 2 个独立的 QPPs 确定决策函数求解上下界。由于 TSVR 的每个 QPP 都小于传统的 QPP,

收稿日期:2024-10-29;网络出版时间:2025-05-08 14:59:00

基金项目:国家自然科学基金资助项目(12061067, 62176033);重庆市自然科学基金面上资助项目(CSTB2023NSCQ-MSX0707)

第一作者:华有霖(2000—),男,硕士研究生,研究方向为机器学习、数据分析的不确定方法等。E-mail:s220603003@stu.cqupt.edu.cn

\*通信作者:邵亚斌(1974—),男,教授,博士,研究方向为模糊分析学、粗糙集与粒计算、数据分析的不确定方法等。E-mail:shaoyb@cqupt.edu.cn

因此 TSVR 的计算效率明显高于传统 SVR。迄今为止 TSVR 也衍生了许多变体模型<sup>[6]</sup>。

人类认知具有大规模优先处理的特征<sup>[7]</sup>,人脑会根据情况从大到小或从粗到细处理事物。例如,当人类看到一幅海滩画时,即使其中的细节描绘得不够精细,人类仍然能够通过画面的整体特征(粗粒度)认出这是一个海滩。如果要更细致地辨别图中的细节特征,如海浪的形状或沙滩上的贝壳,则应关注更小的特征(细粒度)。然而,大多数现有的支持向量回归模型的输入是点或像素,并且它们的训练过程始终以最小粒度执行,没有对数据进行聚类整合,缺乏对数据结构的充分理解使得它们效率低下。为了提升支持向量算法效率,Tang 等<sup>[8]</sup>将粒计算与支持向量算法结合,提出粒支持向量机(granular support vector machine, GSVM)模型。GSVM 模型首先构建信息粒,然后根据需要在一些粒子中构造 SVM。这种方法实现了线性可分问题的推广,同时也增强了线性不可分问题的线性可分性。但是经典的 GSVM 模型用粒子信息代替样本点,导致泛化性能下降。为了解决这个问题郭虎升等<sup>[9]</sup>提出一种动态粒度支持向量回归机(dynamical granular support vector regression, DGSVR)模型,DGSVR 模型在不同层次的粒进行回归训练,在有效压缩训练集的同时,使含有重要信息的样本在最终训练集中保留下来并保证了模型的泛化性能。但是这种粒支持向量回归算法只是在输入数据中找到重要的粒样本信息然后进行训练,没有在 DGSVR 模型中考虑粒子半径信息。

基于以上问题,本文提出一种基于多粒度粒球计算的支持向量回归(multi-granularity granular ball support vector regression, MGBSVR)模型,该模型将粒球的半径信息纳入约束条件中,替代了传统的基于样本点的支持向量回归模型,直接将粗粒度的粒球作为输入,提高了模型训练的效率。此外,研究了 MGBSVR 模型的对偶模型,使用粒子群算法求解,且与传统支持向量回归的对偶模型形式一致,进一步验证了模型的正确性。最后,基于人工数据集和加州大学欧文分校(University of California-Irvine, UCI)数据集,提出的 MGBSVR 模型有效地提升了传统支持向量回归模型的性能。

## 1 相关工作

### 1.1 支持向量回归算法

SVR 在统计学习理论框架中是一种有效的回归模型<sup>[10-11]</sup>,对于数据集  $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ ,  $n$  为样本数, SVR 基本模型为

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C(\mathbf{e}^T \boldsymbol{\xi} + \mathbf{e}^T \boldsymbol{\xi}^*) \\ \text{s.t.} \quad & y_i - f(\mathbf{x}_i) \leq \varepsilon + \xi_i, \quad \xi_i \geq 0, \\ & f(\mathbf{x}_i) - y_i \leq \varepsilon + \xi_i^*, \quad \xi_i^* \geq 0, \end{aligned} \quad (1)$$

式中:  $i = 1, 2, \dots, n$ ,  $f(\mathbf{x}_i)$  为预测值且  $f(\mathbf{x}) = \mathbf{w}\mathbf{x} + b$ ,  $\mathbf{w}$  是权重向量,  $b$  是偏置项,  $\mathbf{e}$  为单位向量,  $y_i$  是实际值,  $\varepsilon$  是设定的一个阈值,用于决定模型对误差的容忍度称之为不敏感参数,  $C$  是惩罚系数,用于平衡模型复杂度和训练误差,  $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ 、 $\boldsymbol{\xi}^* = (\xi_1^*, \xi_2^*, \dots, \xi_n^*)$  为松弛变量,用于处理不满足  $\varepsilon$  不敏感带的样本。

SVR 采用  $L_2$  范数正则化项估计最优回归函数,有效克服了过拟合问题。SVR 及其变体不断发展以适应在不同特定应用<sup>[3]</sup>。受支持向量分类器的启发,在原始 SVR 基础上提出了不敏感的拉普拉斯损失,忽略  $\varepsilon$  内两个软边缘之间误差,避免 SVR 训练中的过度拟合。 $\varepsilon$  决定了 SVR 的泛化能力。Schölkopf 等<sup>[12]</sup>没有直接选择  $\varepsilon$ ,而是引入参数  $\nu$  控制模型的泛化能力,从而提出了  $\nu$ -SVR 模型

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \left( \nu \varepsilon + \frac{1}{n} \sum_{i=1}^n (\xi_i + \xi_i^*) \right) \\ \text{s.t.} \quad & (\mathbf{w}\mathbf{x}_i + b) - y_i \leq \varepsilon + \xi_i, \\ & y_i - (\mathbf{w}\mathbf{x}_i + b) \leq \varepsilon + \xi_i^*, \quad \xi_i^* \geq 0, \quad \varepsilon \geq 0, \end{aligned} \quad (2)$$

式中  $\mathbf{w}\mathbf{x}_i$  表示内积运算。

考虑数据中的噪声分布 SVR 并结合统计机制提高未知集合中的泛化能力,Suykens 等<sup>[13]</sup>将最小二乘的方法引入到支持向量回归机中,提出最小二乘支持向量回归机(least squares support vector machines, LS-SVR)模型,即

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} C \boldsymbol{\xi}^T \boldsymbol{\xi} \\ \text{s.t.} \quad & y_i = \varphi(\mathbf{x}_i) \mathbf{w} + b \mathbf{e} + \boldsymbol{\xi}, \end{aligned} \quad (3)$$

式中,  $\varphi(\mathbf{x}_i)$  为高维空间中的样本点。假设噪声服从正态分布, 对式(3)的拉格朗日函数求导, 利用矩阵形式描述 LS-SVR 模型, LS-SVR 的解则可以用一组线性方程表示。

为了进一步提升传统支持向量回归算法的计算效率, 将粒计算与支持向量算法结合提出粒支持向量机。GSVM 模型创建一系列信息粒, 必要时在粒中学习。它具有以下优点: GSVM 模型将整个空间划分为子空间, 从较小的样本集中获得更多的局部分类信息, 并且超平面的间隔更大。利用分类超平面的间隔能够衡量泛化性能, 因此 GSVM 模型的泛化性能优于传统 SVM。一般来说, GSVM 模型在高维空间中训练, 在低维空间中进行粒划分并用粒替换数据。用生成的粒替换数据后, 在高维特征空间中无法反映原数据固有分布特征, 导致预测函数可能不合适。为了解决这个问题, Guo 等<sup>[14]</sup> 提出核粒支持向量机(model-kernel granular support vector machine, KGSVM)模型, 该模型首先将原始数据映射到高维空间, 通过聚类、神经网络、决策树或粗糙集等策略对粒进行划分, 从而获得高维粒, 最后由高维粒进行 SVM 的训练。KGSVM 模型可以解决低维空间和高维空间中数据点分布不同的问题, 因此, KGSVM 模型在很大程度上提高了 SVM 的泛化性能, 但是这种用部分粒代替样本点信息降低了模型泛化性能。为了解决这一问题, 郭虎升等<sup>[9]</sup> 提出了 DGSVR 模型。该模型首先将原始数据映射到高维空间, 揭示原始样本空间中隐含的特征, 然后将数据划分为一些粒, 并提取近似超平面附近的粒, 定义粒密度和距离函数。根据粒密度程度和半径程度在精细层面上进行重新粒化。最后, 通过不同层次上的所有粒信息获得决策超平面。DGSVR 模型在很大程度上提高了泛化性能和学习效率。

## 1.2 粒球计算

粒计算<sup>[15-16]</sup> 与人脑“大范围首先”的思维方式相似, 并且具有鲁棒性、高效性和可扩展性。通过粒化, 人类获得对客观世界和主观世界多层次、多角度的描述和理解。粒计算不是求解精确解, 而是找寻简单、低成本的近似解<sup>[17]</sup>。Zadeh 提出另一种粒计算模型, 即词计算<sup>[18]</sup>, 该模型利用自然语言变量的模糊约束规则作为推理模式, 解决基于感知计算问题, 该问题无法利用传统的“符号计算”和“数值计算”范式得出结果。张钹院士提出了商空间理论<sup>[19]</sup>, 基本思想是把信息的不同属性放在不同的粒度空间, 寻找合适的商空间, 再合成得出信息处理的一般方法。Yao 等<sup>[20]</sup> 将三支决策理论与粒度计算方法结合, 推进了粒计算的发展。随着粒计算研究的深入, 近年来粒计算在神经网络的应用中取得了丰富的成果。为了提高分类器的鲁棒性和效率, Xia 等<sup>[21]</sup> 提出了粒球计算分类器, 在该分类器的数学模型中, 粒球代替样本点作为输入, 由于超球体几何形状完全对称性并且数学表达简单, 在任何维度下都只需用中心和半径表示, 因此粒球计算与其他分类器的数学模型结合相对容易<sup>[22]</sup>。

目前, 粒球的生成方法主要采用粒球分裂的方式。在该算法中, 为了实现大范围优先的认知规律, 开始生成是最粗粒度, 在满足约束条件时停止。这一过程不仅能够实现从粗粒度到细粒度的逐步精细化, 还能确保粒球数尽可能少。初始时, 所有数据被视作一个单一的粒球, 质量较差, 无法有效反映数据的分布特征, 因此不满足约束条件, 无法满足优化目标。算法通过统计粒球中不同类别的标签数量进行分裂。但此时粒球的质量依旧较低, 因此继续分裂。随着分裂的进行, 粒球质量逐步提升, 决策边界愈加明确。最终, 当所有粒球的质量达到预定要求时, 算法收敛, 粒球边界与数据集边界高度一致。现有的粒球分裂生成方法最早采用的是高效的  $k$ -means, 以保证粒球分类器整体过程的高效性。每个粒球分裂时的初始中心为该粒球内不同标签的点。Xia 等<sup>[22]</sup> 提出基于划分的方法取代  $k$ -means, 将粒球生成效率提高到数倍到数十倍, 同时降低了粒球的时间复杂度。由于之前的粒球生成方法中的初始粒球中心选择具有随机性, 因此粒球生成方法结果不稳定。针对该问题, Xie 等<sup>[23]</sup> 将粒球中不同类别的中心作为子粒球的初始点, 提出有序的、稳定的粒球生成方法, 并且提高了粒球分裂生成的收敛速度。这些策略对于其他粒球计算方法中的粒球生成同样具备一定的启发意义<sup>[24]</sup>。另外, 结合粒球计算概念, Xie 等<sup>[25]</sup> 提出一种无标签的粒球生成方法。首先, 将整个数据集视为一个粒球  $B$ , 随后识别粒球  $B$  中距离最大的 2 个点, 将粒球  $B$  分割成 2 个子球  $B_1$  和  $B_2$ 。粒球的中心对应于所有点的重心, 半径由样本点到中心的最远距离确定。利用局部密度的度量进行粒球分裂, 如果粒球分割后的局部密度超过分割前的密度, 则要进行额外的分割, 直到粒球数量稳定。此外, 半径过大的粒球可能会受到边界点或噪声的影响。会继续分割粒球列表中半径超过最大均值和中位数值值的粒球。无标签粒球生成过程如图 1 所示。

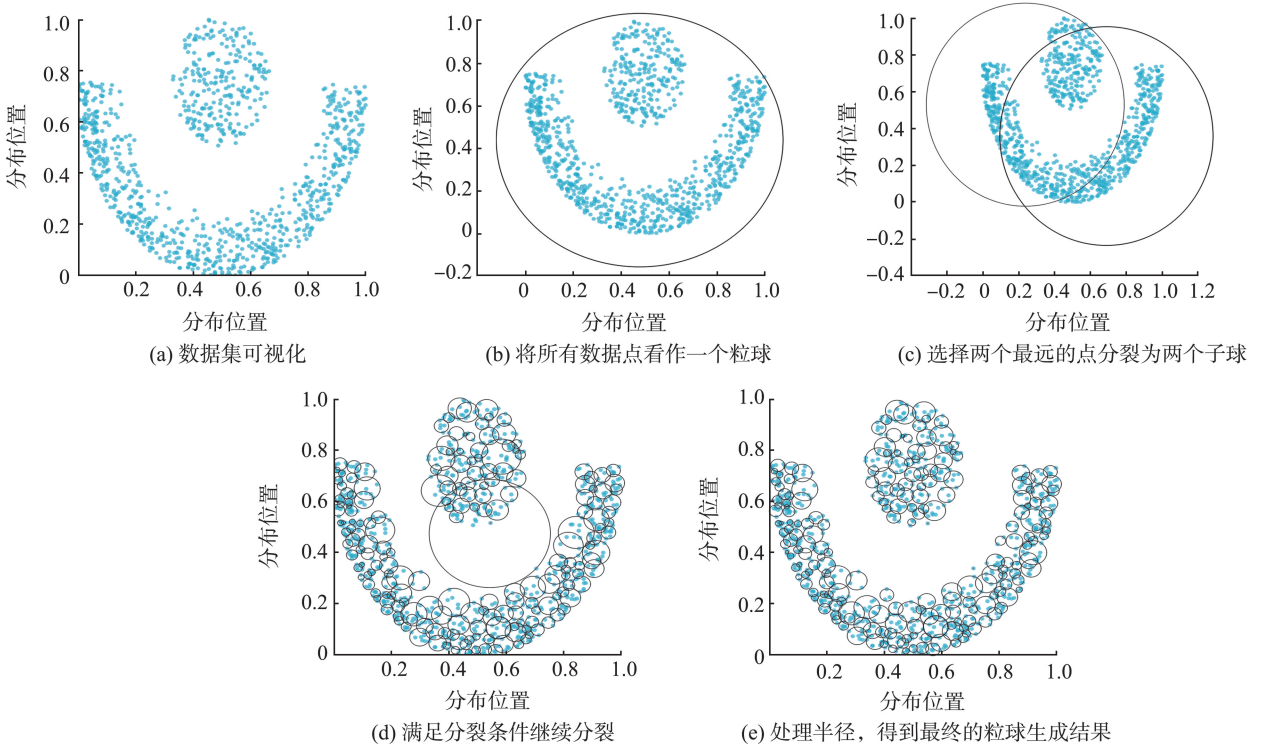


图 1 无标签粒球生成过程

Fig.1 The process of generating label-free granular balls

## 2 多粒度粒球支持向量回归机模型

支持向量回归模型中采用最细的信息粒度,输入样本或像素,然而,这种方法在处理大型数据集时效率低。为了解决这一问题,利用各种方法增强回归器的计算效率。粒球计算对大规模数据处理非常有效且不依赖额外的技术,通过生成全面覆盖原始数据集的粒球实现。给定训练样本  $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ ,  $y_i \in \mathbf{R}$ , 得到一个回归模型  $g(\mathbf{x}_i) = \boldsymbol{\omega}'\mathbf{x}_i + b'$ , 使得  $g(\mathbf{x}_i)$  尽可能接近  $y_i$ ,  $\boldsymbol{\omega}'$  和  $b'$  是待确定的模型参数。对于样本  $(\mathbf{x}_i, y_i)$ , 传统的回归模型通常直接根据模型输出  $g(\mathbf{x}_i)$  与真实输出  $y_i$  之间的差异计算损失。假设  $g(\mathbf{x}_i)$  和  $y_i$  之间最大误差为  $\varepsilon'$ , 当  $g(\mathbf{x}_i)$  和  $y_i$  之间误差大于  $\varepsilon'$  时才计算损失。如图 2(a) 所示, 以  $g(\mathbf{x})$  为中心, 宽度为  $2\varepsilon'$  的区间带上建立一个间隔带。如果训练样本落入这个间隔带内, 则认为预测是正确的。为了提高传统支持向量回归机的效率, 原始数据集通过粒化得到粒球数据  $B_j (j = 1, 2, \dots, N)$ , 其中  $N$  是训练样本的粒球数。希望得到一个回归模型  $h(\mathbf{c}_j, r_j) = \boldsymbol{\omega}\mathbf{c}_j + b_1$  预测数据。粒球取代了原始数据, 如果训练粒球样本落入以  $h(\mathbf{c}_i, r_i)$  为中心、宽度为  $2\varepsilon'$  的间隔带内, 则认为预测是正确的, 如图 2(b) 中的粒球  $B_1$  落入支持超平面  $l_1$  和  $l_2$  内。此时的回归模型不是在单个点  $\mathbf{x}_j$  上进行的, 而是在粒球中心  $\mathbf{c}_j$  和粒球半径  $r_j$  上进行的。

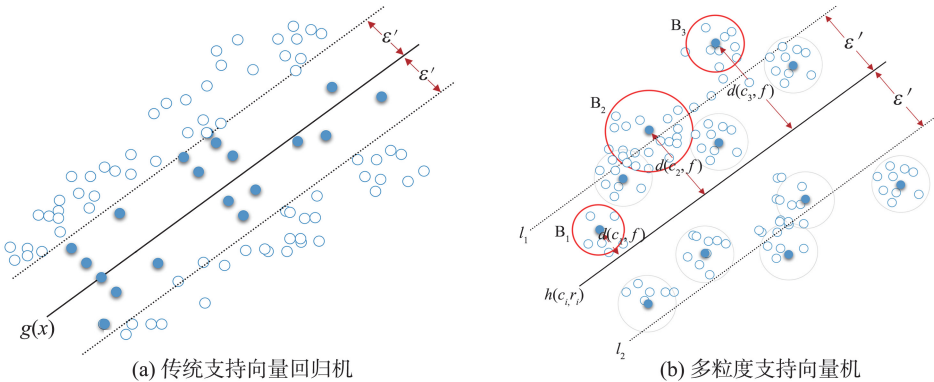


图 2 模型对比示意图

Fig.2 Model of comparison schematic diagram

## 2.1 多粒度粒球支持向量回归机

为了阐明多粒度粒球支持向量回归机模型,下面介绍 MGBSVR 模型的推导过程。如图 2(b) 所示的支持平面  $l_1$  和  $l_2$  分别为

$$\omega \mathbf{x}_1 + b_1 = \varepsilon', \quad (4)$$

$$\omega \mathbf{x}_2 + b_1 = -\varepsilon', \quad (5)$$

式中  $\mathbf{x}_1, \mathbf{x}_2$  为支持向量。由式(4)、(5)得

$$\omega \mathbf{x}_1 = \varepsilon' - b_1, \quad (6)$$

$$\omega \mathbf{x}_2 = -\varepsilon' - b_1. \quad (7)$$

从  $\mathbf{c}_j$  到支持平面的垂直距离应该等于相应的  $r_j$ , 则有

$$\frac{\omega(\mathbf{c}_j - \mathbf{x}_j)}{\|\omega\|} = r_j. \quad (8)$$

将式(6)、(7)代入式(8),得

$$\omega \mathbf{c}_{j_1} - \|\omega\| r_{j_1} + b_1 = \varepsilon', \quad (9)$$

$$\omega \mathbf{c}_{j_2} - \|\omega\| r_{j_2} + b_1 = -\varepsilon', \quad (10)$$

式中  $\mathbf{c}_{j_1}, \mathbf{c}_{j_2}, r_{j_1}, r_{j_2}$  为分别位于支持超平面上粒球的球心和半径。如果粒球位于 2 个支持平面间隔内, 则 MGBSVR 模型的约束表达式为

$$\omega \mathbf{c}_{j_1} - \|\omega\| r_{j_1} + b_1 \geq \varepsilon',$$

$$\omega \mathbf{c}_{j_2} - \|\omega\| r_{j_2} + b_1 \leq -\varepsilon'. \quad (11)$$

该模型的目标是将粒球放入以  $h(\mathbf{c}_j, r_j)$  为中心、宽度为  $2\varepsilon'$  的间隔带中, 因此定义粒球到超平面的距离和损失函数。

**定义 1** (粒球到超平面的距离) 对于一个粒球  $B_j, \mathbf{c}_j, r_j$  分别为

$$\mathbf{c}_j = \frac{1}{m} \sum_{k=1}^m \mathbf{x}_k, \quad (12)$$

$$r_j = \frac{1}{m} \sum_{k=1}^m \|\mathbf{x}_k - \mathbf{c}_j\|, \quad (13)$$

式中  $m$  是粒球当中包含的样本  $\mathbf{x}_k$  的数量。粒球  $B_j$  到超平面  $h(x) = \omega \mathbf{x} + b_1$  的距离是

$$d(B_j, h) = \frac{(\omega, -1) \mathbf{c}_j^T + b_1}{\sqrt{\omega^2 + 1}} - r_j. \quad (14)$$

如果使用映射  $\phi$  将样本点映射到高维空间, 则粒球的中心、半径和距离变为:

$$\mathbf{c}'_j = \frac{1}{m} \sqrt{\sum_{k=1}^m \sum_{k'=1}^m K(x_k, x_{k'})}, \quad (15)$$

$$r'_j = \frac{1}{m} \sum_{i=1}^m \left( \sqrt{K(x_k, x_k) - \frac{2}{m} \sum_{k'=1}^m K(x_k, x_{k'}) + \frac{1}{m^2} \sum_{k', k_1=1}^m K(x_{k'}, x_{k_1})} \right), \quad (16)$$

$$d(B_j, h) = \frac{\frac{1}{m} \sum_{k=1}^m \sum_{j=1}^{|\text{SV}_j|} \alpha_j y_j K(x_j, x_k) + b_1}{\sqrt{\sum_{k=1}^m \sum_{j=1}^{|\text{SV}_j|} \alpha_j \cdot \alpha_k \cdot y_j \cdot y_k \cdot K(x_j, x_k)}} - r_j, \quad (17)$$

式中:  $|\text{SV}_j|$  是支持粒球的个数,  $K(x_j, x_k)$  是样本进行核函数内积运算。

**定义 2** (粒球  $\varepsilon'$ -不敏感损失函数) 给定一个粒球  $B_j, B_j$  的  $\varepsilon'$ -不敏感损失函数为

$$l_{\varepsilon'}(z) = |y_j - h(\mathbf{c}_j, r_j)| = \max\{0, d(B_j, h) + r_j - \varepsilon'\}. \quad (18)$$

引入粒球的  $\varepsilon'$ -不敏感损失函数, 改善了原始损失函数的泛化性能, 如图 3 所示。

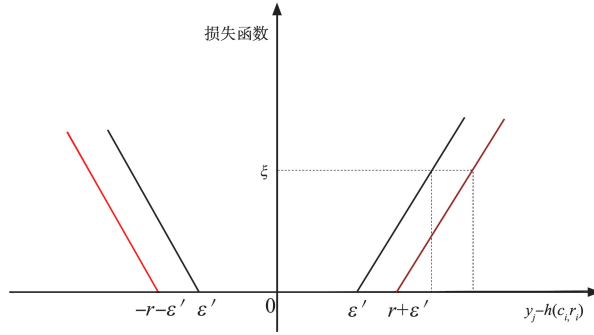


图 3 粒球  $\varepsilon$ -不敏感损失函数  
Fig.3  $\varepsilon$ -Insensitive loss function of granular ball

粒球支持向量回归机问题形式转化为

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N l_{\varepsilon}(z) \quad (19)$$

引入松弛变量  $\xi_j, \xi_j^*$ , 方程(19)可以重写为

$$\begin{aligned} \min_{\omega, \xi_j, \xi_j^*, b} \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{j=1}^N (\xi_j + \xi_j^*) \\ \text{s.t.} \quad & (\omega c_j + b_1) - r_j - y_j \leq \varepsilon + \xi_j, \\ & y_j - (\omega c_j + b_1) - r_j \leq \varepsilon + \xi_j^*, \end{aligned} \quad (20)$$

式中  $\xi_j, \xi_j^* \geq 0$ 。

值得注意的是,当  $r_j=0$  时, MGBSVR 模型等效于传统的 SVR,在这种情况下,包含样本的粒球尺寸最小,2 个模型的等价性证实了 MGBSVR 模型的正确性。

### 2.2 多粒度粒球支持向量回归机对偶模型

引入拉格朗日乘子  $\alpha_j, \alpha_j^*, \mu_j, \mu_j^* \geq 0$ , 利用拉格朗日乘法得到式(20)的拉格朗日函数

$$\begin{aligned} & L(\omega, b_1, \alpha, \alpha^*, \xi, \xi^*, \mu, \mu^*) \\ &= \frac{1}{2} \|\omega\|^2 + C \sum_{j=1}^N (\xi_j + \xi_j^*) - \sum_{j=1}^N (\mu_j \xi_j + \mu_j^* \xi_j^*) \\ &+ \sum_{j=1}^N \alpha_j ((\omega c_j + b_1) - r_j - y_j - \varepsilon - \xi_j) \\ &+ \sum_{j=1}^N \alpha_j^* (y_j - (\omega c_j + b_1) - r_j - \varepsilon - \xi_j^*). \end{aligned} \quad (21)$$

令  $L(\omega, b_1, \alpha, \alpha^*, \xi, \xi^*, \mu, \mu^*)$  关于  $\omega, b_1, \xi, \xi^*$  的偏导数为零, 由式(21)得

$$\omega = \sum_{j=1}^N (\alpha_j - \alpha_j^*) c_j, \quad (22)$$

$$\sum_{j=1}^N (\alpha_j - \alpha_j^*) = 0, \quad (23)$$

$$C = \sum_{j=1}^N (\mu_j + \alpha_j) = \sum_{j=1}^N (\mu_j^* + \alpha_j^*). \quad (24)$$

将式(22)、(23)、(24)带入式(21), 得到对偶模型

$$\begin{aligned} \max_{\alpha_j, \alpha_j^*} \quad & \sum_{j=1}^N [\alpha_j^* (y_j - \varepsilon) - \alpha_j (y_j + \varepsilon)] - \sum_{i=1}^N (\alpha_i + \alpha_i^*) r_i - \frac{1}{2} \sum_{j=1}^N \sum_{j'=1}^N (\alpha_j - \alpha_j^*) (\alpha_{j'} - \alpha_{j'}^*) (c_j c_{j'}), \\ \text{s.t.} \quad & \sum_{j=1}^N (\alpha_j - \alpha_j^*) = 0, \end{aligned} \quad (25)$$

式中:  $0 \leq \alpha_j, \alpha_j^* \leq C$ 。当  $r_j$  变为零时, 方程(25)转化为与原始的 SVR 对偶模型相同的形式, 进一步验证了

MGBSVR 模型的对偶模型的正确性。

### 2.3 多粒度粒球支持向量回归机算法

利用粒子群算法 (particle swarm optimization algorithm, PSO) 求解 MGBSVR 模型的对偶模型 (25), 如算法 1 所示。首先, 将  $D$  中的样本划分为多个粒球得到粒球列表  $T = \{(B_j, y_j)\}_{j=1}^N$ , 计算每个粒球的半径  $r_j$  和中心  $c_j$ 。再利用 MGBSVR 模型问题的对偶形式, 定义一个整合了数据集拟合和模型复杂度惩罚项的适应度函数。初始化粒子的位置和速度, 使用 PSO 算法更新。更新位置后, 遵循 MGBSVR 模型的约束条件, 根据适应度函数更新个体最佳解和全局最佳解。算法在这些步骤之间迭代, 达到最大迭代次数。如果调整后的粒子位置导致适应度函数值增加, 算法返回到步骤 (3); 否则, 优化过程终止。最后, 利用提供的方程式, 确定模型的权重和偏置项。

**算法 1** 多粒度粒球支持向量回归机算法。

**输入** 数据集  $D = \{(x_i, y_i)\}_{i=1}^n$ , 惯性权重  $\eta$ , 最大迭代次数  $I$ , 随机数  $r_1, r_2$ , 更新系数及搜索上下界  $d_1, d_2, L, U$ ;

**输出** 权重和偏置项。

(1) 创建粒球列表  $T = \{(B_j, y_j)\}_{j=1}^N$ , 其中  $N$  表示数据集  $D$  生成粒球的数量。对于每个粒球,  $c_j = \frac{1}{m} \sum_{k=1}^m x_k$ ,  $r_j = \frac{1}{m} \sum_{k=1}^m \|x_k - c_j\|$ ;

(2) 根据式 (25), 定义适应度函数

$$f = -\frac{1}{2} \sum_{j=1}^N \sum_{j'=1}^N (\alpha_j - \alpha_j^*) (\alpha_{j'} - \alpha_{j'}^*) (c_j \cdot c_{j'}) + \sum_{j=1}^N [\alpha_j^* (y_j - \varepsilon) - \alpha_j (y_j + \varepsilon)] - \sum_{j=1}^N (\alpha_j + \alpha_j^*) r_j;$$

(3) 在  $L$  和  $U$  范围内初始化粒子位置  $z_s$  和速度  $v_s$ 。

For  $t = 1$  to  $I$  do

For 每个粒子  $s$

更新粒子速度  $v_s^{t+1} = \eta v_s^t + d_1 r_1 (P_s - x_s^t) + d_2 r_2 (G_s - x_s^t)$

更新粒子位置  $z_s^{t+1} = \text{clip}(z_s^t + v_s^{t+1}, L, U)$

调整粒子位置满足式 (25) 的约束

End For

根据适应度函数更新粒子  $P_s$  和  $G_s$

End For

(4) If 适应度函数随着粒子的调整而增加

重复步骤 (3) else

优化过程结束

End If

(5) 根据式 (22) 计算权重;

根据式 (6) 计算偏置项。

## 3 实验与分析

为了证明 MGBSVR 模型的有效性, 利用人工合成数据集和 UCI 基准数据集验证, 将 MGBSVR 模型与  $\varepsilon$ -SVR、 $\nu$ -SVR 进行比较。所有实验都是在 intel i7 CPU (3.0 GHz)、内存 32 GB 和 Python3.10 的环境中进行。

为了评价所有算法的性能, 评价指标包括总平方和  $T$ 、误差平方和  $E$ 、回归平方和  $R$ 、未解释变异占比  $R_{\text{unexplained}}$ 、决定系数  $R^2$ , 相应的计算式分别为:

$$T = \sum_{i=1}^n (y_{a_i} - \bar{y}_{a_i})^2, \quad (26)$$

$$E = \sum_{i=1}^n (y_{p_i} - y_{a_i})^2, \quad (27)$$

$$R = \sum_{i=1}^n (y_{p_i} - \bar{y}_{p_i})^2, \tag{28}$$

$$R_{\text{unexplained}} = \frac{E}{T} = \frac{\sum_{i=1}^n (y_{p_i} - y_{a_i})^2}{\sum_{i=1}^n (y_{a_i} - \bar{y}_{a_i})^2}, \tag{29}$$

$$R^2 = \frac{R}{T} = \frac{\sum_{i=1}^n (y_{p_i} - \bar{y}_{p_i})^2}{\sum_{i=1}^n (y_{a_i} - \bar{y}_{a_i})^2}, \tag{30}$$

式中,  $y_{p_i}$ 、 $y_{a_i}$  分别代表预测值和观测值,  $\bar{y}_{p_i}$ 、 $\bar{y}_{a_i}$  分别为预测平均值和观测平均值。

误差平方和越小, 预测平均值越精确。当噪声用于测试样本时, 误差平方越小模型越容易出现过拟合。 $R_{\text{unexplained}}$  越小, 说明模型解释的变异性越多,  $R^2$  表示模型对数据变异性的解释程度,  $R^2 \in [0, 1]$ ,  $R^2$  越接近 1, 说明模型解释的变异性越大, 拟合效果越好。

### 3.1 人工数据集

为了验证 MGBSVR 模型的有效性和鲁棒性, 实验中使用 6 个由传统数学函数生成的人工数据集, 如表 1 所示。每个函数生成 500 样本, 其中  $f_1(x)$  为一元函数, 其余均为多元函数。多元函数  $f_4$ 、 $f_5$ 、 $f_6$  中  $\alpha = (x_1, x_2, x_3, x_4, x_5)$ , 其中  $x_1, x_2, \dots, x_5$  表示为向量的分量, 取值范围列在表 1 中。构建模型之前, 数据已经被归一化到 0~1 的范围, 这种情况下, 数据之间没有显著的相关性, 避免了计算的困难。

表 1 生成人工数据集的函数  
Table 1 Functions used to generate artificial datasets

函数类型	表达式	变量范围
Type1	$f_1(x) = \sin\left(\frac{9\pi}{0.35x+1}\right)$	$x \in [0, 10]$
Type2	$f_2(x_1, x_2) = \frac{\sin\sqrt{x_1^2+x_2^2}}{\sqrt{x_1^2+x_2^2}}$	$x_1, x_2 \in [-4\pi, 4\pi]$
Type3	$f_3(x_1, x_2, x_3, x_4, x_5) = 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5$	$x_1, x_2, x_3, x_4, x_5 \in [0, 1]$
Type4	$f_4(\alpha) = \ \alpha - 1\  (1 + 10 \sin\ \alpha + 1\  ) + 1$	$x_1, x_2, x_3, x_4, x_5 \in [-10, 10]$
Type5	$f_5(\alpha) = \left\  \frac{\alpha - 1}{4} \right\  + \left  \sin\left(\pi \left\  1 + \frac{\alpha - 1}{4} \right\  \right) \right  + 1$	$x_1, x_2, x_3, x_4, x_5 \in [-10, 10]$
Type6	$f_6(\alpha) = \sin(\ \alpha\ ) \cos(\alpha^T \alpha)$	$x_1, x_2, x_3, x_4, x_5 \in [0, 6]$

表 2 列出了 MGBSVR 模型与  $\varepsilon$ -SVR、 $\nu$ -SVR 模型在不同噪声水平时的  $R^2$ , MGBSVR 模型的  $R^2$  较高。在实验结果中, 当噪声水平为 0%~30% 时, MGBSVR 模型的  $R^2$  均大于其他模型的  $R^2$ , 如图 4 所示, 而且在高噪声条件下 MGBSVR 模型保持稳定的预测能力。在无噪声条件下, MGBSVR 模型在 Type1 和 Type6 数据类型中  $R^2$  分别达到了 0.672 7 和 0.786 9, 明显大于  $\varepsilon$ -SVR 和  $\nu$ -SVR 模型。中等噪声水平为 10%~20% 时, MGBSVR 模型在 Type6 数据类型中,  $R^2$  值大于  $\varepsilon$ -SVR 和  $\nu$ -SVR 模型。当高噪声水平为 22.5%~30% 时, MGBSVR 模型的  $R^2$  还是大于  $\varepsilon$ -SVR 和  $\nu$ -SVR 模型, 尤其在 Type6 数据类型中的  $R^2$  分别达到 0.911 4、0.917 6、0.927 6 和 0.901 0, 显著大于  $\varepsilon$ -SVR 和  $\nu$ -SVR 模型。 $\varepsilon$ -SVR 和  $\nu$ -SVR 模型在多个噪声水平和数据类型均为  $R^2$  负值, 回归性能较差。总体而言, MGBSVR 模型在不同噪声条件下展示了出色的鲁棒性和有效性, 尤其是在噪声环境中, 其抗噪性能和回归效果显著优于  $\varepsilon$ -SVR 和  $\nu$ -SVR 模型。MGBSVR 模型之所以能够在不同噪声条件下展示出色的鲁棒性和有效性, 因为它将传统的 SVR 模型的点输入改为了粒球输入。多粒度粒球通过聚合样本点形成粗粒度的表示, 减少了对噪声的敏感性, 从而提升了模型的鲁棒性。尤其在高噪声环境中, 粒球输入有效地平滑了噪声干扰, 保证了模型的稳定性和预测精度。此外, MGBSVR 模型还将多粒度粒球的半径信息纳入模型, 进一步提升了计算效率和性能。

表2 不同噪声水平、不同人工数据集的3个模型的决定系数

Table 2 Coefficients of determination of three models on different synthetic datasets under various noise levels

噪声水平	模型	Type1	Type2	Type3	Type4	Type5	Type6
0%	MGBSVR	<b>0.672 7</b>	<b>0.633 0</b>	0.604 2	0.508 3	<b>0.901 8</b>	<b>0.786 9</b>
	$\varepsilon$ -SVR	-0.048 4	0.101 3	0.996 9	0.617 4	0.784 4	-0.040 8
	$\nu$ -SVR	-0.017 3	-0.025 8	0.997 6	0.605 4	0.793 5	-0.001 9
2.5%	MGBSVR	<b>0.573 3</b>	<b>0.281 7</b>	0.853 9	<b>0.754 7</b>	0.612 2	<b>0.911 2</b>
	$\varepsilon$ -SVR	-0.052 6	0.084 4	0.990 2	0.618 0	0.786 4	-0.113 8
	$\nu$ -SVR	-0.010 8	-0.017 6	0.993 0	0.608 3	0.785 2	-0.009 2
5%	MGBSVR	<b>0.572 4</b>	<b>0.371 0</b>	0.850 9	<b>0.804 1</b>	<b>0.889 9</b>	<b>0.735 9</b>
	$\varepsilon$ -SVR	-0.052 4	0.089 5	0.986 2	0.615 9	0.785 0	-0.043 6
	$\nu$ -SVR	-0.000 6	-0.043 9	0.985 2	0.602 2	0.786 9	0.000 1
7.5%	MGBSVR	<b>0.600 2</b>	<b>0.237 9</b>	0.886 3	<b>0.828 1</b>	<b>0.888 0</b>	<b>0.889 9</b>
	$\varepsilon$ -SVR	-0.049 2	-0.051 3	0.979 1	0.621 9	0.782 4	-0.044 1
	$\nu$ -SVR	-0.139 8	0.015 5	0.981 0	0.613 6	0.772 3	-0.135 6
10%	MGBSVR	<b>0.663 2</b>	<b>0.337 1</b>	0.864 1	0.524 4	0.727 7	<b>0.937 9</b>
	$\varepsilon$ -SVR	-0.044 3	0.055 0	0.976 3	0.616 1	0.783 6	-0.041 7
	$\nu$ -SVR	-0.019 3	-0.025 4	0.978 5	0.585 2	0.787 9	-0.042 5
12.5%	MGBSVR	<b>0.621 2</b>	<b>0.248 0</b>	0.847 4	<b>0.727 5</b>	<b>0.844 8</b>	<b>0.937 9</b>
	$\varepsilon$ -SVR	-0.047 6	-0.051 4	0.968 4	0.613 5	0.774 1	-0.035 2
	$\nu$ -SVR	-0.076 5	0.070 1	0.967 0	0.578 7	0.743 0	-0.164 5
15%	MGBSVR	<b>0.656 9</b>	<b>0.348 4</b>	0.880 2	<b>0.730 9</b>	<b>0.929 7</b>	<b>0.909 8</b>
	$\varepsilon$ -SVR	-0.034 9	-0.051 8	0.963 6	0.606 7	0.770 3	-0.040 6
	$\nu$ -SVR	-0.049 8	-0.023 9	0.966 9	0.577 4	0.781 9	-0.124 0
17.5%	MGBSVR	<b>0.582 7</b>	<b>0.417 2</b>	0.860 1	<b>0.891 9</b>	<b>0.758 2</b>	<b>0.9059</b>
	$\varepsilon$ -SVR	-0.077 9	-0.047 4	0.946 2	0.599 8	0.760 2	-0.042 0
	$\nu$ -SVR	0.003 7	-0.025 3	0.948 6	0.583 5	0.766 0	-0.038 6
20%	MGBSVR	<b>0.528 1</b>	<b>0.284 2</b>	0.859 8	0.590 4	<b>0.831 9</b>	<b>0.915 5</b>
	$\varepsilon$ -SVR	-0.039 2	-0.061 0	0.960 1	0.596 3	0.767 8	-0.096 8
	$\nu$ -SVR	-0.003 7	-0.025 0	0.952 6	0.570 0	0.767 1	-0.475 2
22.5%	MGBSVR	<b>0.617 9</b>	<b>0.268 5</b>	0.876 6	<b>0.902 7</b>	<b>0.780 7</b>	<b>0.9114</b>
	$\varepsilon$ -SVR	-0.055 7	-0.049 1	0.942 2	0.613 7	0.764 1	-0.027 4
	$\nu$ -SVR	-0.035 2	-0.025 2	0.950 0	0.599 7	0.754 3	0.001 0
25%	MGBSVR	<b>0.543 1</b>	<b>0.365 9</b>	0.831 6	<b>0.764 8</b>	<b>0.816 2</b>	<b>0.917 6</b>
	$\varepsilon$ -SVR	-0.045 9	-0.046 5	0.947 2	0.614 9	0.739 0	-0.033 5
	$\nu$ -SVR	-0.099 2	-0.033 3	0.935 6	0.597 5	0.749 4	-0.078 8
27.5%	MGBSVR	<b>0.574 5</b>	<b>0.387 0</b>	0.779 1	<b>0.801 8</b>	<b>0.780 1</b>	<b>0.927 6</b>
	$\varepsilon$ -SVR	-0.032 5	-0.038 5	0.931 8	0.564 6	0.734 5	-0.033 8
	$\nu$ -SVR	-0.001 8	-0.021 0	0.939 0	0.582 5	0.727 2	-0.003 0
30%	MGBSVR	<b>0.667 3</b>	<b>0.366 3</b>	0.825 4	<b>0.909 1</b>	<b>0.759 4</b>	<b>0.901 0</b>
	$\varepsilon$ -SVR	-0.053 1	-0.052 9	0.905 8	0.578 4	0.728 6	-0.023 2
	$\nu$ -SVR	-0.038 6	-0.029 4	0.889 3	0.571 1	0.728 6	0.001 4

### 3.2 UCI 数据集

为了进一步验证 MGBSVR 模型的有效性,将 MGBSVR 模型与  $\varepsilon$ -SVR、 $\nu$ -SVR 模型进行比较,从 UCI 数据库中选择 6 个数据集进行对比实验,数据集信息见表 3 所示,其中 R\_wpbc 数据集为修订版威斯康星乳腺癌预后(revised wisconsin prognostic breast cancer, R\_wpbc)数据集。

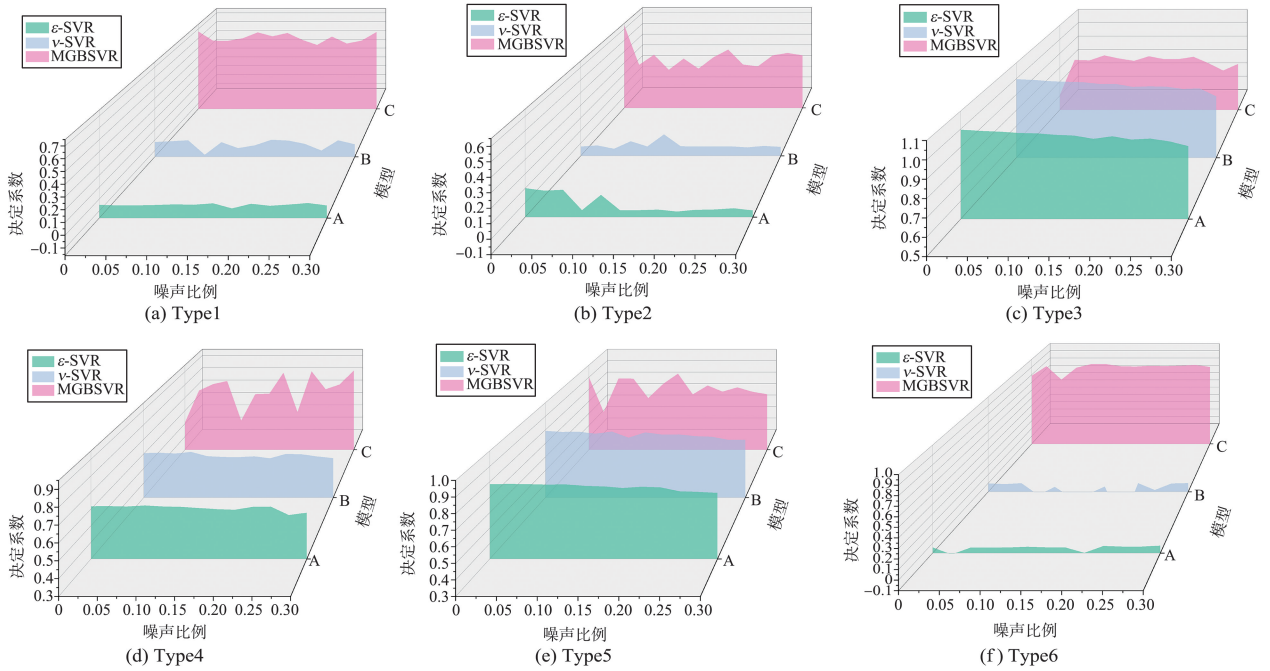


图 4 不同噪声水平下性能对比

Fig.4 Performance comparison at different noise levels

实验之前,对所有数据集进行归一化处理,样本所有特征在 $[0, 1]$ 进行缩放。对于每一个数据集,随机选择 80%的数据集作为训练集,其余的 20%作为测试集。

MGBSVR 模型相比于  $\epsilon$ -SVR 和  $\nu$ -SVR 模型表现更优,实验结果见表 4—7 所示。传统的 SVR 训练过程对所有样本点进行优化,计算复杂度通常为  $O(M^3)$  或  $O(M^2, d)$ ,其中  $M$  为训练的样本数量,  $d$  为特征维度。而在 MGBSVR 模型中,在输入数据

预处理阶段,对样本点进行多粒度粒球表示时得到的粒球数量远小于样本点数量,因此计算复杂度为  $O(N^3)$  或  $O(N^2, d_r)$ ,其中  $N$  为粒球数量,  $d_r$  为粒球表示的维度(包括粒球半径)。表 8 给出了粒度大小对 MGBSVR 模型计算效率的影响,  $\epsilon$ -SVR 和  $\nu$ -SVR 模型均为输入样本点训练,粒度对模型没有影响。由于粒球数量远小于样本数量,模型训练优化的变量数量大幅减少,降低了计算复杂度。MGBSVR 模型的预测精度更高且更稳定,能够解释的数据变异比例大。虽然  $\epsilon$ -SVR 和  $\nu$ -SVR 模型在某些数据集上的预测误差平方较高,预测误差占总平方和比例小说明数据过拟合的风险。MGBSVR 模型引入了多粒度粒球的半径信息,能在训练过程中使模型更好地捕捉数据的结构信息,增强了模型的泛化能力,减少了过拟合的风险。总之, MGBSVR 模型在回归任务中具有全面和可靠的性能,引入粒球和半径信息,提升了计算效率,还增强了对数据结构的理解和适应能力。

表 4 MGBSVR 模型的误差平方和与其它模型的平均误差平方和对比

Table 4 Comparison of the sum of squared errors of the MGBSVR model with the mean squared errors of other models

数据集	MGBSVR	$\epsilon$ -SVR	$\nu$ -SVR
Price	<b>1.143 5</b>	2.110 4	2.251 3
Housing	1.299 4	9.664 9	1.037 3
PoleTelecomm	<b>0.058 5</b>	3.143 2	0.168 7
R_wpbc	<b>4.080 1</b>	31.046 9	4.480 6
Day	5.333 5	0.002 5	0.000 2
Hour	<b>0.005 6</b>	0.081 7	0.006 8

表 3 UCI 数据集

Table 3 UCI Datasets

编号	数据集	样本数	特征数
1	Price	159	14
2	Housing	506	13
3	PoleTelecomm	15 000	48
4	R_wpbc	1 000	33
5	Day	731	14
6	Hour	17 379	15

表 5 MGBSVR 模型的  $R_{\text{unexplained}}$  与其它模型的  $R_{\text{unexplained}}$  对比

Table 5 Comparison of MGBSVR model  $R_{\text{unexplained}}$  with that of other models

数据集	MGBSVR	$\epsilon$ -SVR	$\nu$ -SVR
Price	<b>0.424 2</b>	0.594 5	0.567 0
Housing	<b>0.282 7</b>	0.985 0	1.008 8
PoleTelecomm	0.253 6	0.913 6	0.909 4
R_wpbc	0.427 4	0.102 0	0.198 9
Day	<b>0.322 1</b>	0.999 9	0.999 9
Hour	<b>0.248 8</b>	0.999 9	0.999 9

表6 MGBSVR模型的 $R^2$ 与其它模型的 $R^2$ 对比  
Table 6 Comparison of MGBSVR model  $R^2$  with that of other models

数据集	MGBSVR	$\varepsilon$ -SVR	$\nu$ -SVR
Price	0.777 4	0.837 8	0.827 0
Housing	<b>0.936 1</b>	0.828 2	0.815 6
PoleTelecomm	0.636 2	0.887 8	0.906 8
R_wpbc	<b>0.821 2</b>	0.108 8	0.082 2
Day	0.832 5	0.002 5	0.000 2
Hour	0.999 8	0.999 9	0.999 9

表7 MGBSVR模型的训练时间和与其它模型的训练时间对比  
Table 7 Comparison of the training time of the MGBSVR model with that of other models

数据集	MGBSVR	$\varepsilon$ -SVR	$\nu$ -SVR
Price	<b>0.002</b>	0.010	0.010
Housing	<b>0.002</b>	0.180	0.030
PoleTelecomm	<b>0.150</b>	0.170	1.540
R_wpbc	0.010	0.010	0.010
Day	<b>0.020</b>	0.200	0.040
Hour	0.814	18.400	0.730

表8 粒度对MGBSVR模型的计算效率的影响  
Table 8 The impact of granularity on the computational efficiency of the MGBSVR model

数据集	算法	支持向量个数	粒球半径均值	粒球半径极差
Price	MGBSVR	<b>36</b>	0.307 7	0.931 4
	$\varepsilon$ -SVR	126		
	$\nu$ -SVR	64		
Housing	MGBSVR	<b>110</b>	0.271 5	0.802 4
	$\varepsilon$ -SVR	393		
	$\nu$ -SVR	214		
PoleTelecomm	MGBSVR	<b>1 116</b>	0.202 5	0.845 2
	$\varepsilon$ -SVR	3 893		
	$\nu$ -SVR	2011		
R_wpbc	MGBSVR	<b>42</b>	0.637 8	0.862 2
	$\varepsilon$ -SVR	154		
	$\nu$ -SVR	79		
Day	MGBSVR	<b>159</b>	0.384 5	1.026 1
	$\varepsilon$ -SVR	548		
	$\nu$ -SVR	292		
Hour	MGBSVR	<b>5 123</b>	0.312 5	0.792 4
	$\varepsilon$ -SVR	13 822		
	$\nu$ -SVR	6 983		

## 4 结语

本文提出了一种基于多粒度粒球计算的支持向量回归算法。MGBSVR模型不仅可以直接将粒球作为输入而且还将粒球的半径信息加入到模型的约束条件中,提升了计算效率,增强了模型的鲁棒性。MGBSVR模型比现有的 $\varepsilon$ -SVR和 $\nu$ -SVR模型更高效、更鲁棒。大量的数值实验表明,MGBSVR模型在性能上明显优于传统的 $\varepsilon$ -SVR和 $\nu$ -SVR模型。

尽管MGBSVR模型在效率和鲁棒性方面表现优异,但在处理粗粒度数据时,精度和鲁棒性之间仍存在一定的权衡。本文提出了改进支持向量算法的思路,基于粒球计算的理念可扩展至更多现有的高级模型,这些模型在模式识别领域具有广泛应用前景,如人脸识别、手写字符识别、语音识别、时间序列预测等。未来,粒球计算在这些领域的发展将具有重要意义。

### 参考文献:

- [1] VAPNIK V N. The nature of statistical learning theory[M]. New York: Springer, 2000: 117-125.
- [2] AHMED H U, MOSTAFA R R, MOHAMMED A, et al. Support vector regression (SVR) and grey wolf optimization (GWO) to predict the compressive strength of GGBFS-based geopolymer concrete[J]. Neural Computing and Applications, 2023, 35(3): 2909-2926.

- [3] INAPAKURTHI R K, MITRA K. Optimal surrogate building using SVR for an industrial grinding process[J]. *Materials and Manufacturing Processes*, 2022, 37(15):1701-1707.
- [4] WANG Yougan, WU Jinran, HU Zhihua, et al. A new algorithm for support vector regression with automatic selection of hyperparameters[J]. *Pattern Recognition*, 2023, 133:108989.
- [5] PENG Xinjun. TSVR: an efficient twin support vector machine for regression[J]. *Neural Networks*, 2010, 23(3):365-372.
- [6] DING Shifei, ZHANG Zichen, GUO Lili. Survey on twin support vector regression[J]. *Acta Electronica Sinica*, 2023, 51(4):1117-1134.
- [7] CHEN Lin. Topological structure in visual perception[J]. *Science*, 1982, 218(4573):699-700.
- [8] TANG Yuchun, BO Jin, ZHANG Yanqing. Granular support vector machines with association rules mining for protein homology prediction[J]. *Artificial Intelligence in Medicine*, 2005, 35(1/2):121-134.
- [9] 郭虎升,王文剑. 动态粒度支持向量回归机[J]. *软件学报*, 2013, 24(11):2535-2547.  
GUO Husheng, WANG Wenjian. Dynamical granular support vector regression machine[J]. *Journal of Software*, 2013, 24(11):2535-2547.
- [10] BURGESS C J C. A tutorial on support vector machines for pattern recognition[J]. *Data Mining and Knowledge Discovery*, 1998, 2(2):121-167.
- [11] DRUCKER H, BURGESS C J, KAUFMAN L, et al. Support vector regression machines[J]. *Advances in Neural Information Processing Systems*, 1996, 9:155-161.
- [12] SCHÖLKOPF B, SMOLA A J, WILLIAMSON R C, et al. New support vector algorithms[J]. *Neural Computation*, 2000, 12(5):1207-1245.
- [13] SUYKENS J A K, VANDEWALLE J, DE MOOR B. Optimal control by least squares support vector machines[J]. *Neural Networks*, 2001, 14(1):23-35.
- [14] GUO Husheng, WANG Wenjian, MEN Changqian. A novel learning model-kernel granular support vector machine[C]// 2009 International Conference on Machine Learning and Cybernetics. Baoding: IEEE, 2009, 2:930-935.
- [15] ZADEH L. Fuzzy sets and information granularity[M]// GUPTA N, RAGADE R, YAGER R. *Advances in Fuzzy Set Theory and Applications*. Amsterdam: World Scientific Publishing; 3-18.
- [16] PAWLAK Z. Rough sets[J]. *International Journal of Computer & Information Sciences*, 1982, 11(5):341-356.
- [17] WANG Guoyin, YAO Yiyu, YU Hong. A survey on rough set theory and applications[J]. *Journal of Computer Research and Development*, 2009, 32(7):1229-1246.
- [18] ZADEH L A. Fuzzy logic computing with words[J]. *IEEE Transactions on Fuzzy Sets and Systems*, 1999, 4(2):103-111.
- [19] ZHANG Lin, ZHANG Bo. Theory of fuzzy quotient space (methods of fuzzy granular computing)[J]. *Journal of Software*, 2003, 14(4):770-776.
- [20] YAO Yiyu, YANG Jilin. Granular rough sets and granular shadowed sets: three-way approximations in pawlak approximation spaces[J]. *International Journal of Approximate Reasoning*, 2022, 142:231-247.
- [21] XIA Shuyin, ZHENG Shaoyuan, WANG Guoyin, et al. Granular ball sampling for noisy label classification or imbalanced classification[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 34(4):2144-2155.
- [22] XIA Shuyin, DAI Xiaochuan, WANG Guoyin, et al. An efficient and adaptive granular-ball generation method in classification problem[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(4):5319-5331.
- [23] XIE Qin, ZHANG Qinghua, XIA Shuyin, et al. GBG++: a fast and stable granular ball generation method for classification[J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024, 8(2):2022-2036.
- [24] XIE Jiang, LIAN Jiang, XIA Shuyin, et al. An adaptive density clustering approach with multi-granularity fusion[J]. *Information Fusion*, 2024, 106:102273.
- [25] XIE Jiang, KONG Weiyu, XIA Shuyin, et al. An efficient spectral clustering algorithm based on granular-ball[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(9):9743-9753.

(编辑:陈丽萍)