

# 基于粒计算的极限学习机模型设计与应用

陈丽芳<sup>1</sup> 代琪<sup>1</sup> 付其峰<sup>2</sup>

(华北理工大学理学院 河北唐山 063210)<sup>1</sup> (华北理工大学信息工程学院 河北唐山 063210)<sup>2</sup>

**摘要** 在数据智能处理中属性重要度差异很大且具有高度非线性的特征,在这种情况下直接应用机器学习进行建模处理往往很难获得问题的有效解。针对此问题,文中探索了基于粒计算的属性重要度的排序方法且结合排序结果应用二元关系实现粒层划分算法;应用极限学习机对不同划分获得的粒层空间进行学习,进而对不同粒层空间的学习结果进行对比分析,从而获得最优划分与粒层;此外,将提出的粒度极限学习机模型应用于空气质量的预报问题,不仅加快了预报速度,而且获得的结果与实际预测高度吻合,实证了粒度极限学习机模型的有效性和可靠性。

**关键词** 粒计算,极限学习机,二元关系,粒层空间

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.10.012

## Design and Application of Extreme Learning Machine Model Based on Granular Computing

CHEN Li-fang<sup>1</sup> DAI Qi<sup>1</sup> FU Qi-feng<sup>2</sup>

(College of Science, North China University of Science and Technology, Tangshan, Hebei 063210, China)<sup>1</sup>

(College of Information Engineering, North China University of Science and Technology, Tangshan, Hebei 063210, China)<sup>2</sup>

**Abstract** The importance of attributes in data intelligence processing is not only different from each other, but also highly nonlinear. In such case, it is difficult to obtain effective solutions to the problem by applying machine learning directly. In order to solve this problem, the granularity-based ranking method of attribute importance and the application of the ranking results in binary relationship were explored to perform the granular partitioning algorithm. Then, this paper applied extreme learning machine to granular layer space. The learning results in the layer space were compared and analyzed to obtain the optimal partition and granular layer. In addition, the particle size extreme learning machine model proposed in this paper was applied to the air quality forecasting problem, not only accelerating the forecasting speed, but also being consistent with the actual forecasting, thus empirically proving the validity and reliability of extreme learning machine model.

**Keywords** Granular computing, Extreme learning machine, Binary relationship, Granular space

## 1 引言

数据智能处理作为一个新的研究方向,由于所处理领域问题的差异性,尤其是针对数据高维和非线性特征,采用的处理手段和建模方法大不相同,因此研究适合问题处理的建模方法,达到降维和提升效率的目标是机器学习领域的一个研究热点。

粒计算通过选择合适的粒度来发现问题中较好的、近似的解决方案,从而降低问题求解的复杂度、实现数据的有效降维<sup>[1-2]</sup>,因此成为应对大数据智能处理的有效工具<sup>[3]</sup>。张清华等<sup>[4]</sup>结合粒计算思想提出与人类认知相似的认知模型。梁吉业等<sup>[5]</sup>认为大数据的特征与粒计算的思想高度契合,粒计算的发展有望为大数据的智能处理与挖掘提供一条新的途径。周丹晨<sup>[6]</sup>利用模糊熵空间构建了具有分层递阶结构的熵空间族,并应用粗集计算了各熵空间的属性重要度,最终获得不同

属性的客观权重。刘清等<sup>[7]</sup>在粒的运算机制上,采用非标准分析方法对信息粒的运算规则进行研究,提出了伴随二元关系,并对二元关系信息粒的加粗、加细、交和并运算进行深入分析。该算法可以对粒层空间实现合并和分解,能够有效降低数据的计算强度,简化数据分析过程。考虑到伴随二元关系的计算复杂性,本文借鉴该研究思路,采用一般二元关系实现粒的运算,从而达到粒层空间划分的目的。

本文在分析现有成果的基础上,研究基于粒计算的极限学习机模型,并将其应用于空气质量分类预测<sup>[8]</sup>,对模型进行仿真验证,并对仿真结果进行对比分析,从而筛选出误差最小的粒层,获得决策指导的最优结果。具体实施过程如下:对数据进行离散化,用粒计算方法对条件属性的重要度进行计算并排序;运用二元关系理论进行各粒层属性之间的交计算,得到相应的粒层空间;最后在不同的粒层,分别应用极限学习机构建预测模型,并对结果进行对比分析,得出相应的最优粒层

到稿日期:2018-04-17 返修日期:2018-05-22 本文受河北省自然科学基金面上项目(F2014209086)资助。

陈丽芳(1973—),女,博士,教授,主要研究方向为机器学习、智能计算、数据挖掘,E-mail: hblg\_cif@163.com(通信作者);代琪(1991—),男,硕士生,主要研究方向为粒计算、神经网络;付其峰(1996—),男,主要研究方向为机器学习。

空间。该模型算法将大大简化数据的计算量,实现数据的有效降维,提高学习机的运行效率,降低数据的处理成本。

## 2 理论基础

### 2.1 粒计算

粒计算是当前人工智能研究领域中模拟人类思维和解决复杂问题的新方法<sup>[9]</sup>,越来越受到学者们的关注,其能够有效解决复杂问题求解、大数据挖掘、不确定性信息处理等问题<sup>[10]</sup>。

粒计算中最基本的问题是粒化和粒计算<sup>[11]</sup>。粒化的目的是将问题空间分解为多个子空间,将原始数据划分成适合领域问题的粒;粒计算则是根据粒的划分,在不同粒层空间中进行问题求解,并实现解的最优化计算<sup>[12]</sup>。近年来,对粒化方法的研究一直是学者们关注的焦点。

### 2.2 二元关系

**定义 1** 令  $U = \{x_1, x_2, \dots, x_n\}$  是一个论域,  $R$  是  $U$  上的一个二元关系。若  $R$  是自反的、对称的及传递的,则称  $R$  为等价关系<sup>[13]</sup>。用等价关系实现的划分,称为等价划分。

**定义 2** 设  $R$  和  $S$  是  $U$  上的两个二元关系,定义:

$$R \cap S = \{(x, y) | xRy \wedge xSy\} \quad (1)$$

$$R \cup S = \{(x, y) | xRy \vee xSy\} \quad (2)$$

式(1)为二元关系的交运算,式(2)为二元关系的并运算<sup>[14]</sup>。

本文首先分别用关系  $R$  和  $S$  获得划分的粒,然后应用二元关系的交运算对合并后的权重粒集进行粒层细化计算,得到细化后的粒层,新的粒被分别包含在两个参与运算的旧粒中。

## 3 粒划分与粒层空间

在数据智能处理中,属性重要度不仅差异很大且具有高非线性特征<sup>[15]</sup>,在这种情况下直接应用机器学习进行建模处理往往很难获得问题的有效解。因此,研究基于粒计算的属性重要度排序算法,并结合排序结果应用二元关系进行粒层划分,获得不同的粒层空间,进而应用极限学习机建模,将会大大加快数据的处理速度,精简网络结构<sup>[16]</sup>。

考虑到数据集中的多属性特征,以属性重要度权重为粒子,实施权重粒化。

粒划分与粒层空间构造算法的步骤如下。

**步骤 1** 在数据处理过程中,样本集数据各属性的量纲和取值范围不同,若直接对数据进行计算将产生很大误差,导致结果不收敛。针对这一问题,研究数据规范化处理算法,将样本中各属性转换到固定统一的区间上,以便后续计算。

**步骤 2** 利用规范化处理后的数值计算样本与样本之间的相似关系:

$$R(x_i, x_j) = \begin{cases} 1, & i=j \\ 0 \leq r_{ij} < 1, & i \neq j \end{cases} \quad (3)$$

构建所有样本的模糊相似矩阵。根据得到的样本集模糊相似矩阵创建具有分层递阶结构的熵空间族。传统方法中,一般采用平方法求出  $R$  的传递闭包,从而得到聚类结果<sup>[17]</sup>以构建熵空间族。传递闭包运算的复杂度较高,运算困难,但文献<sup>[18]</sup>证明了传递闭包运算与模糊相似关系运算可以互换,并给出了一种基于模糊相似矩阵的聚类算法,得到一个具有熵

集包含关系的样本熵空间族,避免了传递闭包繁琐的求解过程。本文采用基于模糊相似矩阵的聚类算法构建熵空间族。

**步骤 3** 对规范化数据进行离散化处理,并对离散化样本属性表进行分析,得到删除各属性后的等价类。以熵空间族所形成的熵空间样本聚类作为单一决策属性进行分类。

**步骤 4** 根据式(4)<sup>[6]</sup>计算各粒度空间下各属性的重要度,根据式(5)求出各属性的最终重要度。

$$sig(a_k, C, D) = \gamma(C, D) - \gamma(C - \{a_k\}, D) \quad (4)$$

$$sig(a_k, C, D) = \frac{1}{q} \sum_{\rho=1}^q sig^{X^{(\rho)}}(a_k, C, D) \quad (5)$$

基于以上 4 个步骤,得到数据集的属性重要度排序结果  $\{a_1, a_2, a_3, \dots, a_n\}$ 。

**步骤 5** 以属性权重  $\omega$  为依据,实施等价划分以获得不同的粒子,并构建粒层空间。

1) 令  $\omega_1 = \{a_1\}$ , 对数据集进行粒划分,构成第一粒层空间  $\Omega_1 = \{[x]_1\}$ ;

2) 令  $\omega_2 = \omega_1 \cup \{a_2\}$ , 应用二元关系的交运算式(1)对数据集进行粒划分,构成第二粒层空间  $\Omega_2 = \{[x]_2\}$ ;

3) 以此类推,令  $\omega_n = \omega_{n-1} \cup \{a_n\}$ , 对数据集实施粒划分,得到第  $n$  粒层空间  $\Omega_n = \{[x]_n\}$ 。

## 4 模型构建

经过以上粒划分算法,可获得不同的粒层空间,针对不同粒层空间应用极限学习机分别构建模型并实施粒计算;另外,对不同粒层的学习结果进行性能对比分析,最终获得最佳粒划分与最优粒层。

### 4.1 极限学习机

极限学习机(Extreme Learning Machine, ELM)作为一种单隐层前馈网,因具有实现过程简单、执行速度快、泛化性能好等优点<sup>[19]</sup>,成为机器学习领域研究的热点,其学习速度和效率明显优于传统的 BP 神经网络。其算法流程<sup>[20]</sup>如下: 1) 根据任意连续概率分布函数随机生成输入权值  $\omega_i$  和偏置  $b_i$ ; 2) 计算隐含层输出矩阵  $H$ ; 3) 计算输出权值矩阵  $\hat{\beta}$ 。

与传统的机器学习算法相比,ELM 具有结构简单、学习速度快和良好的全局寻优能力,因而受到越来越广泛的关注<sup>[21]</sup>,已被成功应用于生物科技、光学、医学治疗、电力、图像与模式识别等领域,是目前机器学习领域中最受欢迎的学习机之一<sup>[22]</sup>。特别地,该模型的初始权值和偏置值可以随机给定,只要满足激活函数是连续可微的,输出权值就被唯一确定;且权值是通过求隐含层输出矩阵的广义逆得到的,在样本规模适中的情况下,其具有不需要反复调整参数、学习速度快<sup>[23]</sup>的优势。

### 4.2 模型构建

针对不同粒层空间应用极限学习机分别建模,并对学习结果进行对比分析,获得最佳划分与粒层。

模型构建的步骤如下。

**步骤 1** 研究属性重要度排序的粒化方法,实现科学、约简的粒化过程,构成二元关系序列  $R_1, R_2, \dots, R_n$ , 其中  $R_i = R_{i-1} \cup \{a_i\}$ , 二元关系的个数与属性个数相同。

**步骤 2** 等价类划分,利用排序的属性结果实施粒划分,获得不同的粒层空间  $\Omega_1, \Omega_2, \dots, \Omega_n$ 。

步骤 3 用等价类划分结果重新设计构建 ELM,以粒子(等价类)代替原来的单个数据样本,从学习机的内涵上对 ELM 进行重新构造,构造过程如下:

1)以  $N$  个不同的等价类粒子  $(x_i, t_i)$  为学习样本,若具有  $\tilde{N}$  个隐含层神经元,激活函数为  $g(x)$ ,则 ELM 算法的输出函数为:

$$\sum_{i=1}^{\tilde{N}} \beta_j g(\omega_i X_j + b_i) = o_j, j=1, \dots, N \quad (6)$$

其中,  $\omega_i = [\omega_{i1}, \omega_{i2}, \dots, \omega_{im}]^T$  为第  $i$  个输入节点与隐含层节点的权重,  $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{in}]^T$  和  $b_i$  分别为第  $i$  个隐含层节点与输出节点的权重和偏置。

2)通过寻求最优的网络参数使目标函数  $t$  的取值最小:

$$\sum_{j=1}^{\tilde{N}} \|o_j - t_j\| = 0 \quad (7)$$

即存在  $\beta_i, \omega_i$  和  $b_i$  满足:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(\omega_i \cdot X_j + b_i) = t_j, j=1, \dots, N \quad (8)$$

式(8)可以简化为:

$$H\beta = T \quad (9)$$

其中:

$$H(\omega_1, \dots, \omega_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}, X_1, \dots, X_N) = \begin{bmatrix} g(\omega_1 \cdot X_1 + b_1) & \dots & g(\omega_{\tilde{N}} \cdot X_1 + b_{\tilde{N}}) \\ \vdots & \dots & \vdots \\ g(\omega_1 \cdot X_N + b_1) & \dots & g(\omega_{\tilde{N}} \cdot X_N + b_{\tilde{N}}) \end{bmatrix} \quad (10)$$

其中,  $H$  为神经网络隐含层输出矩阵;  $H$  的第  $i$  列是输入向量  $(x_1, x_2, \dots, x_N)$  的隐含层节点输出。

当激活函数  $g$  无限可微时,输入权值  $\omega$  和隐含层节点阈值  $b$  在网络训练前随机给定,输出权值  $\beta$  是唯一需要训练确定的参数,即:

$$\min_{\beta} E = \min \|H\beta - t\| \quad (11)$$

目标求解过程可转化为求取线性函数的最小二乘解,可采用式(12)计算:

$$\beta^* = H^+ t \quad (12)$$

其中,  $H^+$  是矩阵  $H$  的 Moore-penrose 广义逆矩阵<sup>[23]</sup>。

3)将以上学习机构建过程应用于不同的粒层空间  $\Omega_i$ , 对比不同粒层空间的学习效率和精度,从而获得最优粒层。

整个模型构建过程中,最关键的步骤是对粒层空间  $\Omega_i$  中的参数  $i$  进行优化,找到最佳的  $i$  值,通过模型输出结果对比来确定最优的粒层空间。

## 5 仿真与应用

### 5.1 粒化计算

采集空气质量站点监测原始数据,考虑了  $\text{SO}_2, \text{NO}_2, \text{PM}_{10}, \text{PM}_{2.5}, \text{CO}, \text{O}_3$  6 个条件属性进行粒化建模和极限学习寻优。

空气质量的各项数据属性和量化标准不同,且数值差异较大<sup>[18]</sup>。为了避免训练过程中产生误差,对所有数据进行离散化处理,使处理后的数据范围为  $[0, 100]$ , 公式如下:

$$x'_{ik} = \frac{x_{ik} - \min_{i=1}^n(x_{ik})}{\max_{i=1}^n(x_{ik}) - \min_{i=1}^n(x_{ik})} \times 100 \quad (13)$$

其中,  $x'_k$  是预测模型的输入数据,  $x_{ik}$  表示实际观测数据,  $\min_{i=1}^n(x_{ik})$  和  $\max_{i=1}^n(x_{ik})$  表示条件属性中的最小值与最大值。

空气质量的原始数据如表 1 所列,其中  $a_1$  表示  $\text{SO}_2, a_2$  表示  $\text{NO}_2, a_3$  表示  $\text{PM}_{10}, a_4$  表示  $\text{PM}_{2.5}, a_5$  表示  $\text{CO}, a_6$  表示  $\text{O}_3$ 。使用式(13)对数据进行规范化处理(限于篇幅,此处省略规范化后的表格),并将前 5 个属性数据按照如下连续值区间进行离散化处理:  $[0, 30) = 1; [30, 60) = 2; (60, 100] = 3; a_6$  属性按平均划分,划分区间为  $[0, 8) = 1; [8, 16] = 2; (16, 24] = 3$ 。离散化结果如表 2 所列。

表 1 空气质量的原始数据

Table 1 Original data of air quality

样本	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$
$x_1$	68	65	170	40	1.228	112
$x_2$	60	60	158	41	1.231	117
$x_3$	27	52	140	50	1.002	111
$x_4$	69	89	141	50	3.729	61
$x_5$	68	90	156	59	2.884	54
$x_6$	45	65	113	56	1.756	63
$x_7$	50	79	111	49	2.932	41
$x_8$	62	89	100	44	4.621	29
$x_9$	83	102	151	78	3.755	18
$x_{10}$	194	120	248	124	9.217	6
$x_{11}$	166	108	203	125	7.598	11
$x_{12}$	161	112	206	139	7.183	10
$x_{13}$	197	123	232	129	8.080	12
$x_{14}$	211	63	297	123	7.387	31
$x_{15}$	137	78	298	136	4.562	61

表 2 离散化结果属性

Table 2 Discretization result attribute

样本	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$
$x_1$	1	2	1	1	1	3
$x_2$	1	1	1	1	1	3
$x_3$	1	1	1	1	1	3
$x_4$	1	2	1	1	2	2
$x_5$	1	2	1	1	1	2
$x_6$	1	2	1	1	1	2
$x_7$	1	2	1	1	1	1
$x_8$	1	2	1	1	2	1
$x_9$	1	2	1	1	2	1
$x_{10}$	2	3	2	2	3	1
$x_{11}$	2	2	1	2	3	1
$x_{12}$	2	2	1	2	3	1
$x_{13}$	2	3	1	2	3	1
$x_{14}$	2	2	2	2	3	1
$x_{15}$	2	2	2	2	2	2

计算出属性的权重粒排序结果为:

$\text{SO}_2 = 0.536, \text{NO}_2 = 0.214, \text{PM}_{10} = 0.178$

$\text{PM}_{2.5} = 0.036, \text{CO} = 0.018, \text{O}_3 = 0.018$

根据式(1),使用粒划分与粒层空间构造算法中的步骤 5 进行计算。

1)选择  $\omega_1 = \{a_1\}$  作为  $R, U/R = \{\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}, \{x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}\}\}$ , 将论域  $U$  划分成 2 类, 构成第一粒层空间  $\Omega_1 = \{[x]_1\}$ 。

2)选择  $\{a_2\}$  作为  $S, U/S = \{\{x_2, x_3\}, \{x_1, x_4, x_5, x_6, x_7, x_8, x_9, x_{11}, x_{12}, x_{14}, x_{15}\}, \{x_{10}, x_{13}\}\}$ , 将论域  $U$  划分成 3 类。以权重粒  $\omega_2 = \omega_1 \cup \{a_2\}$ , 利用二元关系的交对  $\omega_2$  实现粒层细

化计算: $R \cap S = UR_1 \cap US_1 = \{a_1 = 1 \cap a_2 = 1\} = \{x_2, x_3\}$ ,  $R \cap S = UR_1 \cap US_2 = \{a_1 = 1 \cap a_2 = 2\} = \{x_1, x_4, x_5, x_6, x_7, x_8, x_9\}$ ;  $R \cap S = UR_2 \cap US_2 = \{a_1 = 2 \cap a_2 = 2\} = \{x_{11}, x_{12}, x_{14}, x_{15}\}$ ,  $R \cap S = UR_2 \cap US_3 = \{a_1 = 2 \cap a_2 = 3\} = \{x_{10}, x_{13}\}$ 。获得4类分类结果,分别为: $\{x_2, x_3\}$ ,  $\{x_1, x_4, x_5, x_6, x_7, x_8, x_9\}$ ,  $\{x_{11}, x_{12}, x_{14}, x_{15}\}$ ,  $\{x_{10}, x_{13}\}$ 。其他交运算的结果均为 $\emptyset$ 。通过 $\omega_2$ ,将论域划分成4类,构成第二粒层空间 $\Omega_2 = \{[x]_2\}$ 。

3)选择 $\{a_3\}$ 作为 $Q$ ,将论域 $U$ 分成2类。以权重粒 $\omega_3 = \omega_2 \cup \{a_3\}$ ,利用二元关系运算对 $\omega_3$ 实现粒层细化计算,获得6类分类结果,分别为: $\{x_2, x_3\}$ ,  $\{x_1, x_4, x_5, x_6, x_7, x_8, x_9\}$ ,  $\{x_{11}, x_{12}\}$ ,  $\{x_{14}, x_{15}\}$ ,  $\{x_{10}\}$ ,  $\{x_{13}\}$ 。其余二元关系运算的结果均为 $\emptyset$ 。通过 $\omega_3$ ,将论域划分为6类,构成第三粒层空间 $\Omega_3 = \{[x]_3\}$ 。目前空气质量等级分为6类,因此这个粒层的划分结果与实际情况相符,但是否为最佳粒层还需要结合机器学习算法进行进一步验证。

4)选择 $\{a_4\}$ 作为 $P$ ,对权重粒 $\omega_4 = \omega_3 \cup \{a_4\}$ 进行粒层计算,形成第四粒层空间 $\Omega_4 = \{[x]_4\}$ ,分类并没有增加,仍为6类,说明添加属性 $P$ 对粒层的变化并没有太大的影响,但其是否对预测结果产生影响,还需要进行学习机建模后才可验证。

5)以 $\omega_5 = \omega_4 \cup \{a_5\}$ 作粒层空间划分,构成第五粒层空间 $\Omega_5 = \{[x]_5\}$ ,分为8类;以 $\omega_6 = \omega_5 \cup \{a_6\}$ 作粒层空间划分,构成第六粒层空间 $\Omega_6 = \{[x]_6\}$ ,分为11类,限于篇幅,省略具体计算过程。

因此,通过二元关系运算,获得了6个粒层空间,第一粒层 $\Omega_1$ 的分类结果为2类;第二粒层 $\Omega_2$ 的分类结果为4类;第三粒层 $\Omega_3$ 、第四粒层 $\Omega_4$ 的分类结果均为6类;第五粒层 $\Omega_5$ 的分类结果为8类;第六粒层 $\Omega_6$ 的分类结果为11类。该计算结果中,第三、第四2个粒层空间与空气质量实际分为6类相符合,说明这两种粒层划分是正确且有效的,因此,针对第三、第四2个粒层空间分别构建极限学习机,对空气质量的AQI进行计算分类,对比分类结果,最终确定最优粒层。

## 5.2 ELM 仿真实证

基于Python编程,实现极限学习机模型的设计与仿真。将第三粒层、第四粒层的数据作为输入,利用ELM算法模型对AQI进行预测,并将预测结果与实际观测值进行对比,第三粒层的输入为 $\omega_3$ ,第四粒层的输入为 $\omega_4$ 。

使用第三粒层数据进行学习建模,得到的预测结果如图1所示,可以看出预测值与实际观测值有一定的误差。

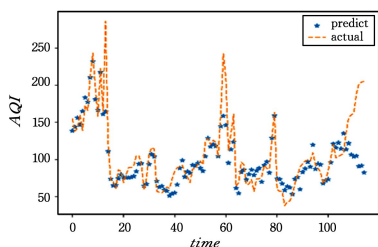


图1 第三粒层数据预测结果与实际观测值的对比

Fig. 1 Comparison between third set of data prediction and actual observation

第三粒层拟合度大大提高,如图2所示。

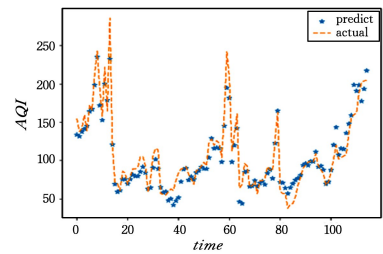


图2 第四粒层数据预测结果与实际观测值的对比

Fig. 2 Comparison between fourth set of data prediction and actual observation

按照第3部分的粒层划分算法,增加权重粒 $\omega_4$ ,粒层分类结果不变。极限学习机建模预测精度比第三粒层更高。进一步用12组测试数据对极限学习进行验证,结果如图3所示。

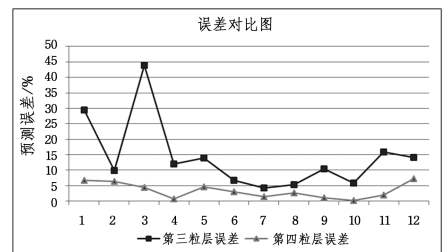


图3 12组数据预测误差的对比

Fig. 3 Prediction error between twelve sets

由图3可以看出,第四粒层的预测结果误差明显小于第三粒层,基本控制在7%以内,准确率达93%以上,因此第四粒层能够正确获得空气质量分类结果,为人们的出行、生产生活提供指导。

利用第五粒层数据对AQI进行预测,相比第四粒层数据,预测值精度反而降低,这是因为粒层分类逐渐加细,不仅增加了学习机的计算量,还会导致误差增大。因此,该应用中第四粒层是最优粒层,能够实现较为准确的空气质量预报,指导人们的行为决策。

**结束语** 本文针对目前数据建模中属性众多且存在高度非线性关系,直接应用学习机建模的学习效果不理想、误差大等问题,提出将粒计算思想引入算法中,对数据进行粒化处理得到不同粒层,并在不同粒层寻找最优粒层;对最佳粒层数据应用极限学习机建模,获得最优的预测结果。该方法能够大大缩短机器学习时间,提升结果精度。仿真实证结果表明,随着粒层的增加,数据的预测精度不断提升,但是当粒层过于细化时,预测精度反而有所下降,这说明了本文粒化算法的有效性和合理性。通过实证仿真,证明了利用粒计算选择最优粒层对预测精度的提升有很大的帮助,同时能够大大降低计算成本和学习时间,且获得的结果与实际预测结果高度吻合,验证了模型的有效性和可靠性。

## 参考文献

- [1] LIANG J Y, FENG C J, SONG P. A Survey on Correlation Analysis of Big Data[J]. Chinese Journal of Computers, 2016, 39(1): 1-15. (in Chinese)

梁吉业,冯晨娇,宋鹏. 大数据相关分析综述[J]. 计算机学报,

利用第四粒层数据进行建模预测,得到的测试结果相比

- 2016,39(1):1-15.
- [2] LIANG J Y, QIAN Y H, LI D Y, et al. Research development on granular computing theory and method for big data[J]. *Big Data Research*, 2016, 38(2): 13-22. (in Chinese)  
梁吉业, 钱宇华, 李德玉, 等. 面向大数据的粒计算理论与方法研究进展[J]. *大数据*, 2016, 38(2): 13-22.
- [3] XU J, WANG G Y, YU H. Review of Big Data Processing Based on Granular computing [J]. *Chinese Journal of Computers*, 2015, 38(8): 1497-1507. (in Chinese)  
徐计, 王国胤, 于洪. 基于粒计算的大数据处理[J]. *计算机学报*, 2015, 38(8): 1497-1507.
- [4] ZHANG Q H, ZHOU Y L, TENG H T. Cognition model based on granular computing[J]. *Journal of Chongqing University of Posts and Telecommunications(Natural Science Edition)*, 2009, 21(4): 494-501. (in Chinese)  
张清华, 周玉兰, 滕海涛. 基于粒计算的认知模型[J]. *重庆邮电大学学报(自然科学版)*, 2009, 21(4): 494-501.
- [5] LIANG J Y, QIAN Y H, LI D Y, et al. Granular Data Calculation Theory and Method for Big Data Mining[J]. *China Science (Information Science)*, 2015, 45(11): 1355-1369. (in Chinese)  
梁吉业, 钱宇华, 李德玉, 等. 大数据挖掘的粒计算理论与方法[J]. *中国科学(信息科学)*, 2015, 45(11): 1355-1369.
- [6] ZHOU D C. A method for ascertaining the weight of attributes based on granular computing[J]. *Journal of Intelligent Systems*, 2015, 10(2): 273-279. (in Chinese)  
周丹晨. 采用粒计算的属性权重确定方法[J]. *智能系统学报*, 2015, 10(2): 273-279.
- [7] LIU Q, QIU T R, LIU L. The Research of Granular computing Based on Nonstandard Analysis[J]. *Chinese Journal of Computers*, 2015, 38(8): 1618-1624. (in Chinese)  
刘清, 邱桃荣, 刘澜. 基于非标准分析的粒计算研究[J]. *计算机学报*, 2015, 38(8): 1618-1624.
- [8] FU Y L, HONG Y. Air Quality Forecasting Based on IPSO-ELM [J]. *Environmental Science and Technology*, 2017, 40(S1): 324-328. (in Chinese)  
付亚丽, 洪亚. 基于 IPSO-ELM 算法的空气质量预测[J]. *环境科学与技术*, 2017, 40(S1): 324-328.
- [9] 张铃, 张钊. 问题求解理论与应用[M]. 北京: 清华大学出版社, 2007.
- [10] MIAO D Q, XU F F, YAO Y Y, et al. Set-theoretic formulation of granular computing[J]. *Chinese Journal of Computers*, 2012, 35(2): 351-363. (in Chinese)  
苗夺谦, 徐菲菲, 姚一豫, 等. 粒计算的集合论描述[J]. *计算机学报*, 2012, 35(2): 351-363.
- [11] SANCHEZ M A, CASTILLO O, CASTRO J R. An Overview of Granular Computing Using Fuzzy Logic Systems[M]// *Nature-Inspired Design of Hybrid Intelligent Systems*. Springer International Publishing, 2017, 667: 19-38.
- [12] XIE K M, LU X H, CHEN Z H. Basic Problem and Research of Granular computing [J]. *Computer Engineering and Applications*, 2007, 43(16): 41-44. (in Chinese)  
谢克明, 遂新红, 陈泽华. 粒计算的基本问题和研究[J]. *计算机工程与应用*, 2007, 43(16): 41-44.
- [13] WANG C Z, HE Q, SHAO M W, et al. A unified information measure for general binary relations[J]. *Knowledge-Based Systems*, 2017, 135: 18-28.
- [14] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.
- [15] HSU C H. Intelligent big data processing[J]. *Future Generation Computer Systems*, 2014, 36(3): 16-18.
- [16] TANG C H, SHU L. An Improved Attribute Reduction Algorithm based on Granular Computing[J]. *International Journal of Computers Communications and Control*, 2015, 10(6): 96.
- [17] 杨纶标, 高英仪. 模糊数学原理及应用[M]. 广州: 华南理工大学出版社, 2008: 66-77.
- [18] TANG X Q, ZHU P, CHENG J X. Cluster analysis based on fuzzy quotient space[J]. *Journal of Software*, 2008, 19(4): 861-868.
- [19] VALDÉS J J. Extreme learning machines with heterogeneous data types[J]. *Neurocomputing*, 2018, 277: 38-52.
- [20] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2005, 70(1): 489-501.
- [21] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications [J]. *Neurocomputing*, 2006, 70(1): 489-501.
- [22] ZHANG H X. Research on Extreme Learning Machine Theory and Algorithms[D]. Shenyang: Shenyang Aerospace University, 2017. (in Chinese)  
张海霞. 极限学习机理论与算法研究[D]. 沈阳: 沈阳航空航天大学, 2017.
- [23] GAN L. Research and Application of Extreme Learning Machine [D]. Xi'an: Xi'an University of Electronic Science and Technology, 2014. (in Chinese)  
甘露. 极限学习机的研究与应用[D]. 西安: 西安电子科技大学, 2014.

(上接第 53 页)

- [11] LI J H, WU W Z. Granular computing approach for formal concept analysis and its research outlooks [J]. *Journal of Shandong University(Natural Science)*, 2017, 52(7): 1-12. (in Chinese)  
李金海, 吴伟志. 形式概念分析的粒计算方法及其研究展望[J]. *山东大学学报(理学版)*, 2017, 52(7): 1-12.
- [12] XU W, LI W. Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets [J]. *IEEE Transactions on Cybernetics*, 2016, 46(2): 366-379.
- [13] LI J, MEI C, XU W, et al. Concept learning via granular computing: a cognitive viewpoint [J]. *Information Sciences*, 2015, 298(1): 447-467.
- [14] LOIA V, ORCIUOLIF, PEDRYCZ W. Towards a granular computing approach based on Formal Concept Analysis for discovering periodicities in data[J]. *Knowledge-Based Systems*, 2018, 146: 1-11.
- [15] YAO Y Y. Concept lattices in rough set theory[C]// *Fuzzy Information IEEE Annual Meeting*. IEEE, 2004, 2: 796-801.