第 55 卷　第 4 期
2019 年 7 月

南京大学学报(自然科学版)
JOURNAL OF NANJING UNIVERSITY
(NATURAL SCIENCE)

Vol. 55, No. 4
July, 2019

DOI:10.13232/j.cnki.jnju.2019.04.012

基于时间粒的铝电解过热度预测模型

郭英杰[1*],胡　峰[1],于　洪[1],张红亮[2]

(1.计算智能重庆市重点实验室,重庆邮电大学,重庆,400065;2.中南大学冶金与环境学院,长沙,410083)

摘　要:过热度是铝电解生产过程中的一项重要参数,将过热度保持在适当的范围内可以提高电流效率,减小电解槽损耗,但是过热度测量难度较大且测量过程复杂.因此,基于粒计算理论,提出一种基于时间粒的过热度预测模型.通过在时间序列上构建时间粒,结合时间粒构建新的特征集与样本集,在此基础上,利用分类器对新的样本集进行训练,得到模型.采用山东魏桥铝电有限公司的铝电解生产数据进行实验,结果表明,该方法在预测过热度上较已有模型的预测能力有较大提升.

关键词:过热度,粒计算,时间序列,铝电解

中图分类号:TP391　　**文献标识码:**A

Prediction model of superheat in aluminum electrolysis based on time granularity

Guo Yingjie[1*], Hu Feng[1], Yu Hong[1], Zhang Hongliang[2]

(1. Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China;
2. School of Metallurgy and Environment, Central South University, Changsha, 410083, China)

Abstract: Superheat is an important parameter in the process of aluminum electrolysis. Keeping the superheat within an appropriate range can improve the current efficiency and reduce cell loss. However, the measurement of superheat is difficult and the measurement process is complex. According to granular computing theory, this paper proposes a prediction model of superheat based on time granule. By constructing time granules on time series, new feature sets and sample sets are constructed combinating with time granules. On this basis, new sample sets are trained by the classifier to obtain the model. In this paper, we use the data of aluminum electrolysis production from Shandong Weiqiao Aluminum and Electricity Ltd to test the experiment. The result shows that the supreheat prediction of this method is better than the existing models.

Key words: superheat, granular computing, time series, aluminum electrolysis

2001年中国成为世界最大的铝生产国.截至2017年,中国电解铝、氧化铝产量占全球产量的51.1%和54.8%.2018年中国电解铝产量达2984万吨,同比增长1.83%.尽管发展迅猛,

基金项目:国家自然科学基金(61533020,61876027,61751312)
收稿日期:2019－05－21
*通讯联系人,E-mail:281787765@qq.com

可也有一系列问题,即产能过剩与行业内微利/亏损的矛盾,在短时间内难以调和,这是我国铝电解工业面临的最大挑战,同时,其可持续发展须应对资源、能源和环保等方面的重大难题.

铝电解槽电解质的过热度是指电解质温度与初晶温度的差值.过热度影响电解槽的炉膛形状及稳定性,进而影响电解槽的寿命.其机理是,电解质过热度过大,会造成槽帮减薄、伸腿缩短,严重时使电解槽处于无炉帮运行状态,威胁电解槽寿命,造成原铝品质的降低[1].

很多学者提出一些有效的过热度预测方法.曹丹阳等[2]在Restreken公式的基础上提出初晶温度预测方法,以电解质各种化学成分作为特征来预测初晶温度,再根据电解温度计算出过热度.杨吉森[3]提出一种铝电解过热度软测量预测模型,先对原始数据进行静态规则提取形成规则树,然后基于规则树构造规则增量更新方法.刘运胜[4]提出基于相对密度噪声过滤随机森林的铝电解过热度预测,对历史生产数据进行相对密度噪声过滤后用随机森林方法预测铝电解过热度.但上述模型只能得出过热度与测量属性之间的映射关系,无法反映未来时间过热度与当前已有测量属性之间的关系.因此,开展过热度预测方法的研究很有必要.

结合粒计算的思想,本文提出一种基于时间粒的过热度预测模型,通过在滑动窗口内构建时间粒,通过时间粒来构建新的样本集并采用分类器对样本集进行训练,得到模型.本文首先阐述相关概念,其次详细讨论基于时间粒的过热度预测模型,最后给出实验结果分析.

1　相关概念

1.1　时间序列数据挖掘　自20世纪90年代以来,人们开始进行时间序列数据挖掘(Time-Series Data Mining,TSDM)研究,目标是从这些"形状"序列数据中挖掘出和时间相关的有价值的知识和信息,或发现有规律的变化模式和异常点等,用于指导和改善日常的生产生活[5].

时间序列预测是传统时间序列分析的基本目标,也是TSDM的主要挖掘任务之一.它通过已知的时间序列的历史观测值去预测未来某一时刻或者某一时段的值,通常指下一时刻.传统时间序列预测模型主要是基于概率统计学的建模方法,最典型的代表是Box and Jenkins[6]提出的差分整合移动平均自回归模型(Autoregressive Integrated Moving Average Model,ARIMA)系列(主要包括自回归模型(Autoregressive Model,AR),滑动平均模型(Moving Average Model,MA),自回归滑动平均模型(Auto-Regressive Moving Average Model,ARMA)和ARIMA).随着人工智能技术的飞速发展,一些经典机器学习方法如神经网络(Artificial Neural Network,ANN)[7−9]和支持向量回归机(Support Vactor Regerssion,SVR)[10−12]已广泛应用于时间序列预测.例如,Moore et al[13]利用径向神经网络对水质时间序列进行预测,Zadeh[14]在递归神经网络和贝叶斯Levenberg-Marquardt学习算法的基础上提出了一种时间序列预测模型.这类模型对精确时间序列数据具有较好的预测精度.

1.2　粒计算　粒计算是借助现有的一些形式体系和技术如区间(集合)[15]、模糊集[16]、粗糙集[17]、阴影集[18]和概率集[19]等来解决问题的方法论.在粒计算框架下,根据实际问题选择合适的形式体系,将问题抽象为信息粒,对这些信息粒进行计算、处理并返回处理的结果.

粒计算在本质上与人类处理复杂问题的基本思路相同,即从实际问题出发,将复杂问题分解转化为若干较简单的问题(信息粒),并在分析和解决问题的过程中,通过不断地调整信息粒的粒度,将问题进行简化,提高求解问题的效率,从而帮助人们更好地解决问题.

2　基于时间粒的铝电解过热度预测模型

为解决时间序列上的过热度预测问题,本文提出一种基于时间粒的铝电解过热度预测模

型. 第一步,对数据进行频率统一并选择用于构建特征集的属性;第二步,确定滑动窗口和时间粒的大小,将时间序列内的子列进行区间信息粒化,得到若干个相同大小的时间粒;第三步,在每个时间粒内和相邻时间粒之间构建特征集,使用特征集构建样本并对样本进行打标,将相同电解槽的数据进行聚合,得到小样本集,将所有小样本集取并集得到样本集;第四步,对样本集中数据不平衡的小样本集进行过采样处理,将所有样本集取并集得到处理后的样本集;第五步,将样本集划分为训练集、验证集和测试集,使用XGBoost[20],RandomForest[21],IBK(即KNN算法)[22]和J48(即Decision Tree算法)[23]进行训练得到模型. 其模型构建流程图如图1所示.

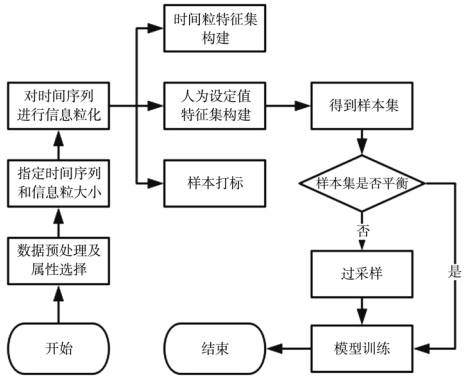

图1 模型构建流程图

Fig. 1 The flow chart of model construction

2.1 数据介绍 数据由山东魏桥铝电有限公司提供,包含288个电解槽从2015年10月到2016年7月的生产数据,分为生产数据1和生产数据2两部分. 生产数据1主要是能够实时监测的属性的值,数量为47个,包括设定电压、槽电压、槽电流、槽电阻、基准下料间隔、实际下料间隔、针振、摆动等,测量频率约为每槽每天720条. 生产数据2主要是通过对产出原铝进行化学分析得到的值,数量为31个,包括铁含量、硅含量、温度、铝水平、电解质水平、槽温度、析出铝量、原铝质量、分子比等,频率为每槽每天一条. 表1列出了两部分数据的基本信息.

表1 铝电解数据集基本信息

Table 1 The information of aluminum electrolysis datasets

属性	样本数	测量频率(每槽每天)	取值范围	属性数量
生产数据1	3500万	720条	2015.10—2016.7	47
生产数据2	65536	1条	2015.10—2016.7	31

为了构建样本集将数据进行预处理. 首先将数据进行频率统一. 为了解决铝电解的原始数据存在采样频率不同的问题,本文根据生产数据1的统计数据对生产数据2进行填充,统计生产数据1中每个属性每天的平均值,根据电解槽编号和采样时间填充到生产数据2中.

其次,选择合适的属性用于特征集构建. 由于历史数据中属性过多,且其中一些属性与过热度毫无关系,本文结合冶金专家经验,选出一部分属性用于构建特征集,包括:设定电压、槽电压、槽电流、槽电阻、过滤电阻、平滑电阻、设定最高电压、设定最低电压、效应等待间隔、针振、摆动等20个属性,使用这些属性构造属性集F,如式(1)所示:

$$F=\{X_1,X_2,\ldots X_n\},n\subseteq\{1,2,\ldots 20\} \quad (1)$$

其中,X_n表示选择的属性. 表2列出F中每个属性的取值范围.

2.2 算法思路

2.2.1 确定滑动窗口和时间粒大小 进行特征集和样本集构建的一个重要前提就是确定滑动窗口和滑动窗口内时间粒的大小. 如果滑动窗口和时间粒太小,会导致样本集数量过大,算法运行时间过长;如果滑动窗口和时间粒太大,会导致构建样本集所需原始数据数量太大.

在基于时间粒的铝电解过热度预测模型中,采用的是经验值划分,令时间窗口天数大小$T_b=9$,时间粒天数大小$T_s=3$,即每个时间窗口

表2　选出的属性的取值范围

Table 2　The range of values for selected attributes

属性	取值范围
设定电压	[4200,4290]
槽电流	[4010,4200]
过滤电阻	[4200,4400]
设定最高电压	[4080,4480]
效应等待间隔	[7200,57600]
摆　动	[0,164]
总变化斜率	[1820,2206]
实际下料间隔	[463,1533]
硅含量	[0.04,0.18]
电解质水平	[24,33]
槽电压	[4160,4320]
槽电阻	[4250,4400]
平滑电阻	[4300,4500]
设定最低电压	[4000,4430]
针　振	[0,347]
电阻变化斜率	[1740,2169]
基准下料间隔	[820,1140]
铁含量	[0.02,0.16]
铝水平	[19,23.5]
析出铝	[2950,5600]

中包含三个长度为3的时间粒.公式如下所示:

$$TW=\{W_1,W_2,W_3\} \tag{2}$$

$$W_n=\{T_{3n-2},T_{3n-1},T_{3n}\},n\subseteq\{1,2,3\} \tag{3}$$

其中,TW表示一个滑动时间窗口,W表示一个时间粒,T_n表示一个时间窗口内第n天的测量数据.

2.2.2　构建特征集和样本集　确定滑动窗口和时间粒大小后,可将原始数据处理成用于模型训练的数据集,具体包括两部分:特征集的构建和样本集的构建.

2.2.2.1　特征集的构建　每个样本的特征集由测量值特征集和经验值特征集组成.其中,测量值特征集表示使用测量得到的数据构建的特征集,经验值特征集表示使用经验的数据构建的特征集.如式(4)所示:

$$A=N_c\cup N_m \tag{4}$$

其中,N_c表示测量值特征集,N_m表示经验值特征集.下面给出两部分特征集的构建过程.

(1)构建测量值特征集:对于TW中的所有Ω,将F中的每个属性在Ω中求出其均值、方差、最大值,令得到的属性集分别为N_1,N_2,N_3,如式(5)、式(6)和式(7)所示:

$$N_1=\left\{\frac{\sum_{i=1}^{3}F(T_i)}{3},\frac{\sum_{i=4}^{6}F(T_i)}{3},\frac{\sum_{i=7}^{9}F(T_i)}{3}\right\} \tag{5}$$

$$N_2=\left\{\frac{\sum_{i=1}^{3}F(T_i)-N_{11}}{3},\frac{\sum_{i=4}^{6}F(T_i)-N_{12}}{3},\frac{\sum_{i=7}^{9}F(T_i)-N_{13}}{3}\right\} \tag{6}$$

$$N_3=\{\mathrm{MAX}(F(W_1)),\mathrm{MAX}(F(W_2)),\mathrm{MAX}(F(W_3))\} \tag{7}$$

其中N_{11},N_{12},N_{13}为在N_1中对应属性集合中的第1,2,3个值,$F(T_i)$为F中的每个属性在时间窗口内第i天的测量数据.

在时间粒迁移的过程中,即从前一个时间粒迁移到下一个时间粒的过程中,将式(5)至式(7)中求出的属性集中的每一项与其前一项的差的绝对值作为新的特征集.将上述所有特征集取并集作为测量值特征集N_c,如式(8)、式(9)和式(10)所示:

$$N_4=\{|F(N_{12}-N_{11})|,|F(N_{13}-N_{12})|\} \tag{8}$$

$$N_5=\{|F(N_{22}-N_{21})|,|F(N_{23}-N_{22})|\} \tag{9}$$

$$N_6=\{|F(N_{32}-N_{31})|,|F(N_{33}-N_{32})|\} \tag{10}$$

则:

$$N_c=N_1\cup N_2\cup N_3\cup N_4\cup N_5\cup N_6 \tag{11}$$

(2)构建经验值特征集:对于经验值属性M,包括设定电压、设定最高电压、设定最低电压、基准下料间隔,计算出当天的设定值与滑动窗口内每一天的设定值之差的绝对值,作为特征集N_7.将当天所有设定值作为特征集N_8.将N_7和N_8取并集得到经验值特征集N_m,如式(12)、式(13)和式(14)所示:

$$N_7=\{|F(M-M_i)|\},i\subseteq\{1,2,\ldots,9\} \tag{12}$$

$$N_8=M \tag{13}$$

$$N_m = N_7 \cup N_8 \tag{14}$$

2.2.2.2 样本集的构建 对于数据集中每个电解槽的数据,分别采用2.2.2.1得到的特征集A作为样本的特征向量,采用过热度类别*label*作为样本的标签,将相同的槽的样本进行聚合得到一个小样本集Y_n. 由于过热度这个属性测量难度较大,因此并无测量数据,所以通过冶金专家提供的过热度计算公式计算出对应的过热度. 具体来说,过热度指的是电解质初晶温度与电解质温度的差值,计算公式如式(15)所示:

$$SHD = T_p - T_e \tag{15}$$

其中,SHD表示过热度,T_p表示电解槽温度,T_e表示初晶温度. 由于初晶温度在线测量难度较大,因此使用分子比的值对初晶温度进行估算. 本文采用的初晶温度计算公式如下:

$$T_e = 35MR + 846 \tag{16}$$

其中,MR表示分子比.

计算值小于等于25的为低过热度,*label*为0;大于25的为高过热度,*label*为1. 如式(17)所示:

$$label = \begin{cases} 0 & SHD \leqslant 25 \\ 1 & SHD > 25 \end{cases} \tag{17}$$

将所有Y_n取并集可以得到总样本集YS,如式(18)所示:

$$YS = \{Y_1, Y_2, \cdots, Y_n\}, n \subseteq \{1, 2, \ldots, 288\} \tag{18}$$

2.2.3 对样本集进行过采样操作 对于2.2.2.2中得到的样本集合YS,其中一些Y_n存在样本不平衡现象,即Y_n中两类样本比例不等于1:1. 样本不平衡往往会导致模型对样本数较多的分类造成过拟合. 为了解决这个问题,本文采用weka平台下的SMOTE[24]算法对Y_n进行过采样操作. 算法流程如下:

(1)对于少数类中每一个样本,以欧氏距离[25]为标准计算它到少数类样本集中其余样本的距离,取距离前k近的样本作为其k近邻. 欧氏距离的计算如式(19)所示:

$$\rho = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{19}$$

(2)根据样本不平衡比例设置采样倍率N使得采样后样本比大致为1:1,对于每一个少数类样本,设置最近邻数量k=5,从其k近邻中随机选择若干个样本,令选择的近邻为X_n.

(3)对于每一个随机选出的近邻X_n,在X与X_n之间生成新的少数类样本X_{new},如式(20)所示[24]:

$$X_{new} = X + rand(0,1) \times \left| X - X_n \right| \tag{20}$$

2.3 算法描述 基于时间粒的铝电解过热度预测模型算法思想描述如下:

算法 基于时间粒的铝电解过热度预测模型(Prediction Model Based on Time Granularity,PMBTG)

输入:n个电解槽的数据集$D=\{D_1, D_2, \cdots, D_n\}$

输出:分类模型PMBTG-XGB,PMBTG-RF,PMBTG-IBK和PMBTG-J48.

Step1:对D进行频率统一和属性选择

Step2:初始化:

滑动窗口大小T_b=9.

时间粒大小T_s=3.

Step3:利用式(5)至式(11)构建测量值特征集N_c.

利用式(12)、式(13)和式(14)构建经验值特征集N_m.

利用式(4)得到特征集A

for each D_i in D

用A作为样本的特征向量,利用式(15)、式(16)和式(17)得到过热度类别*label*作为样本标签. 将所有样本进行聚合得到样本集Y_n.

end for

将所有Y_n取并集得到总样本集YS.

Step4:for each Y_n in YS

If Y_n类别比例≠1:1 then

利用式(19)和式(20)对Y_n进行过采样操作.

end if

end for

将处理后的所有小样本集Y_n合成样本集YN.

Step5:采用文中提出的PMBTG算法,并采用XGBoost[20],RandomForest[21],IBK[22]和J48[23]分类器对YN进行分类训练,RandomForest[21]算法参数与RDNF-RF[4]保持一致,XGBoost[20],IBK[22]和J48[23]算法的参数采用默认值. 得到分类模型PMBTG-XGB,PMBTG-RF,PMBTG-IBK和PMBTG-J48.

算法的时间复杂度分析:令数据集中有n

个电解槽,每个电解槽有m条原始数据.在算法中,Step1的时间复杂度为$O(m \times n)$;Step2的时间复杂度为$O(1)$;Step3的时间复杂度为$O(n \times m^2)$;Step4的时间复杂度为$O(n)$;Step5的时间复杂度为$O(n\log n)$.故,算法的时间复杂度为$O(n \times m^2)$.

3　实验评价

3.1　实验方法　为考察文中提出的基于时间粒的过热度预测模型的性能,本文将其与杨吉森提出的过热度软测量模型(Soft Measuring Model of Superheat Degree,SMM)[3]和刘运胜提出的基于相对密度噪声过滤随机森林的铝电解过热度模型(Prediction of Aluminum Electrolysis Superheat Based on Relative Density Noise Filtering Random Forest,RDNF-RF)[4]进行对比.采用文中提出的PMBTG算法,使用sklearn中的XGBoost[20]分类器和weka平台下的RandomForest[21],IBK[22]和J48[23]分类器分别对样本集进行训练(模型记作PMBTG-XGB,PMBTG-RF,PMBTG-IBK和PMBTG-J48),将训练得到的模型与SMM和RDNF-RF模型进行比较.

3.2　评价方法　本实验采用经典的查准率P(Precision)、查全率R(Recall)和F值(F-Score)作为评测标准,在二分类问题中,可将样例根据其真实类别与学习器预测类别的组合划分为真正例(true positive,TP)、假正例(false positive,FP)、真反例(true negative,TN)和假反例(false negative,FN)四种情形,令TP,FP,TN和FN分别表示其对应的样例数.计算如下所示:

$$P=\frac{TP}{TP+FP} \tag{21}$$

$$R=\frac{TP}{TP+FN} \tag{22}$$

$$F=\frac{2PR}{P+R} \tag{23}$$

在本文中,TP表示预测过热度为低同时实际过热度也是低的样本数,FP表示预测过热度为低但实际过热度是高的样本数,TN表示预测过热度为高同时实际过热度也是高的样本数,FN表示预测过热度为高但实际过热度是低的样本数.

3.3　样本集　根据2.2.2所述,每个时间窗口内的数据可以构建一条样本,每个电解槽的数据可以构建一个样本集.表3展示了一个时间窗口内的样本特征及标签.其中,N_c表示测量值特征集,N_m表示经验值特征集.Label为过热度类别,由式(15)和式(16)计算获得.

表3　样本特征及标签

Table 3　Sample characteristics and labels

N_c	N_m	Label
N_1,N_2,N_3 N_4,N_5,N_6	N_7,N_8	*label*

3.4　实验结果　从原始数据集aeData中随机抽取若干个电解槽的数据作为实验样本集,其基本信息如表4所示.基本信息包括样本数、特征数、测量日期与类分布.其中类分布为高过热度样本数与低过热度样本数之比.数据集名称前的5,10,30,50表示随机抽取5,10,30,50个电解槽的数据形成的数据集.

表4　实验数据集基本信息

Table 4　Basic information of experimental datasets

Dataset	样本数	特征数	测量日期	类分布
5-aeData	1306	78	2015.11.6—2016.7.4	868:438
10-aeData	2477	78	2015.11.6—2016.7.4	1605:872
30-aeData	7606	78	2015.11.6—2016.7.4	3948:3658
50-aeData	12362	78	2015.11.6—2016.7.4	7307:5055

将表4中数据集以6∶3∶1的比例划分为训练集、测试集、验证集,使用PMBTG-RF,PMBTG-XGB,SMM和RDNF-RF训练,得到各项评测标准的对比,实验结果如表5、表6和表7所示,表中黑体字表示评测标准的最优值.

为提高实验的客观性和公平性,PMBTG-RF的各项参数与RDNF-RF中的各项参数保持一致.其余分类器的各项参数采用默认值.

表5 各个模型的Precision值对比

Table 5 Precision values for each model

Dataset	PMBTG-RF	PMBTG-XGB	PMBTG-J48	PMBTG-IBK	RDNF-RF	SMM
5-aeData	0.7904	**0.8002**	0.7055	0.7735	0.5507	0.2353
10-aeData	0.7806	**0.7975**	0.7171	0.7702	0.6030	0.3095
30-aeData	0.7943	**0.8204**	0.7303	0.7801	0.6604	0.4063
50-aeData	0.8086	**0.8101**	0.7235	0.7909	0.6803	0.6050

表6 各个模型的Recall值对比

Table 6 Recall values for each model

Dataset	PMBTG-RF	PMBTG-XGB	PMBTG-J48	PMBTG-IBK	RDNF-RF	SMM
5-aeData	0.7735	**0.8035**	0.6846	0.7467	0.5706	0.4392
10-aeData	0.7697	**0.7904**	0.6905	0.7403	0.6035	0.4227
30-aeData	0.7940	**0.8202**	0.7107	0.7677	0.6307	0.4743
50-aeData	0.8002	**0.8238**	0.7198	0.7803	0.6715	0.4856

表7 各个模型的*F*-Score值对比

Table 7 *F*-Score values for each model

Dataset	PMBTG-RF	PMBTG-XGB	PMBTG-J48	PMBTG-IBK	RDNF-RF	SMM
5-aeData	0.7819	**0.8018**	0.7052	0.7502	0.5605	0.3064
10-aeData	0.7751	**0.7939**	0.6987	0.7433	0.6032	0.3573
30-aeData	0.7941	**0.8103**	0.7099	0.7707	0.6452	0.4377
50-aeData	0.8044	**0.8269**	0.7236	0.7799	0.6759	0.5388

从表5、表6和表7可以看出,在所有数据集上,无论是Precision值,Recall值还是*F*-Score值,本文提出的PMBTG算法得到的模型相比较其他两种算法在各个指标上都存在一定的优势.

4 结束语

本文提出一种基于时间粒的铝电解过热度预测方法,用于预测未来某一时刻的过热度.首先,对数据进行频率统一和属性选择.其次,根据经验值确定滑动窗口和时间粒的天数大小,进行特征集的构建,包括测量值特征集和经验值特征集.然后,用得到的特征集作为样本的特征向量并对样本进行打标,得到小样本集,将若干小样本取并集得到初始样本集.再次,对初始样本集中不平衡的小样本集进行过采样处理,将所有平衡的小样本集取并集得到训练样本集.最后,采用XGBoost,RandomForest,IBK,J48算法对训练样本集进行训练,得到模型.实验结果表明,相比已有的预测模型,本文提出的方法在各个指标上都有明显提高.但本文算法中的时间窗口和时间粒大小均由经验值决定,如何划分出更优的滑动窗口和时间粒大小从而提高模型效果将是今后研究的重点.

参考文献

[1] 裴海灵,周乃君,姜昌伟.电解质过热度对铝电解槽物理场的影响.湖南冶金,2005,33(1):12－16. (Pei H L, Zhou N J, Jiang C W. The influence of super heat on three field of aluminum reduction cells. Hunan Metallurgy, 2005, 33(1): 12－16.)

[2] 曹丹阳,曾水平,李晋宏.铝电解质过热度预测模型研究.轻金属,2010(10):35－38. (Cao D Y, Zeng S P, Li J H. Research on prediction model of aluminium electrolyte superheat degree. Light Metals, 2010(10): 35－38.)

[3] 杨吉森.一种铝电解过热度软测量预测模型.中国自动化学会,济南市人民政府.2017中国自动化大会(CAC2017)暨国际智能制造创新大会(CIMIC2017)论文集.中国自动化学会,济南市人民政府:中国自动化学会,2017:161－167. (Yang J S. Soft measuring model of superheat degree in the aluminum electrolysis production. Chinese association of automation, Ji′nan Municipal People′s Government//2017 Chinese Automation Congress(CAC2017) and Intelligent Manufacturing International Conference (CIMIC2017) Proceeding. Chinese Association of Automation, Ji′nan Municipal People′s Government: Chinese Association of Automation, 2017: 161－167.)

[4] 刘运胜.基于相对密度噪声过滤随机森林的铝电解过热度预测.中国自动化学会,济南市人民政府.2017中国自动化大会(CAC2017)暨国际智能制造创新大会(CIMIC2017)论文集.中国自动化学会、济南市人民政府:中国自动化学会,2017,324－329. (Liu Y S. Prediction of aluminum electrolysis superheat based on relative density noise filtering random forest. Chinese Association of Automation, Ji′nan Municipal People′s Government//2017 Chinese Automation Congress (CAC2017) &. Intelligent Manufacturing International Conference (CIMIC2017) Proceeding. Chinese Association of Automation, Ji′nan Municipal People′s Government: Chinese Association of Automation, 2017, 324－329).

[5] 邓伟辉.时间序列的多粒度智能分析方法研究.博士学位论文.重庆:中国科学院大学(中国科学院重庆绿色智能技术研究院),2017. (Deng W H. Multi－granularity intelligent analyzing for time series. Ph.D. Dissertation. Chongqing: Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, 2017.)

[6] Box G E P, Jenkins G M. Time series analysis: forecasting and control. Journal of Time, 2010, 31(4):303－305.

[7] Yan W Z. Toward automatic time－series forecasting using neural networks. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23(7): 1028－1039.

[8] Chandra R. Competition and collaboration in cooperative coevolution of elman recurrent neural networks for time－series prediction. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(12): 3123－3136.

[9] Miranian A, Abdollahzade M. Developing a local least-squares support vector machines-based neuro-fuzzy model for nonlinear and chaotic time series 0prediction. IEEE Transactions on Neural Networks and Learning Systems, 2013, 24(2): 207－218.

[10] Chen T T, Lee S J. A weighted LS-SVM based learning system for time series forecasting. Information Sciences, 2015, 299: 99－116.

[11] Ristanoski G, Liu W, Bailey J. A time-dependent enhanced support vector machine for time series regression//Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2013, DOI: 10.1145/2487575.2487655.

[12] Laboissiere L A, Fernandes R A S, Lage G G. Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. Applied Soft Computing, 2015, 35: 66－74.

[13] Moore R E, Kearfott R B, Cloud M J. Introduction to interval analysis. Philadelphia: Society for Industrial and Applied Mathematics, 2009, 235.

[14] Zadeh L A. Fuzzy sets. Information and Control, 1965, 8(3): 338－353.

[15] Zadeh L A. Fuzzy algorithms. Information and Control, 1968, 12(2): 94－102.

[16] Zadeh L A. Quantitative fuzzy semantics. Information Sciences, 1971, 3(2): 159－176.

[17] Pawlak Z. Rough sets. International Journal of Computer & Information Sciences, 1982, 11(5): 341－356.

[18] Pawlak Z, Skowron A. Rough sets: some extensions. Information Sciences, 2007, 177(1): 28－40.

[19] Pedrycz W. From fuzzy sets to shadowed sets: interpretation and computing. International Journal of Intelligent Systems, 2009, 24(1): 48－61.

[20] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Machine Learning, 2016, ArXiv: 1603.02754.

[21] Yen S J, Lee Y S. Cluster-based under-sampling approaches for imbalanced data distributions. Expert Systems with Applications, 2009, 36(3): 5718－5727.

[22] Atallah D M, Badawy M, El-Sayed A, et al. Predicting kidney transplantation outcome based on hybrid feature selection and KNN classifier. Multimedia Tools and Applications, 2019(10): 1－25.

[23] Sahu S, Mehtre B M. Network intrusion detection system using J48 Decision Tree//International Conference on Advances in Computing, Communications and Informatics. Kochi, India: IEEE, 2015, DOI: 10.1109/ICACCI. 2015. 7275914.

[24] 胡峰, 王蕾, 周耀. 基于三支决策的不平衡数据过采样方法. 电子学报, 2018, 46(1): 135－144. (Hu F, Wang L, Zhou Y. An oversampling method for imbalance data based on three-way decision model. Acta Electronica Sinica, 2018, 46(1): 135－144.)

[25] 吴青华. 基于相对同步欧氏距离筛选的在线GPS轨迹数据压缩算法. 计算机应用与软件, 2018, 35(3): 282－288. (Wu Q H. Online GPS Online gps trajectory data compression algorithm based on relative synchronous Euclidean distance filtering. Computer Applications and Software, 2018, 35(3): 282－288.)

(责任编辑 杨可盛)