第40卷 第5期 太 阳 能 学 报 Vol. 40, No. 5
2019年5月 ACTA ENERGIAE SOLARIS SINICA May, 2019

文章编号:0254-0096(2019)05-1330-08

基于机器学习的生物质三组分含量预测

邢勇强[1]，邢献军[2,3]，张 静[1]，李永玲[1]，张学飞[1]，张贤文[2,3]

（1. 合肥工业大学机械工程学院，合肥 230009；2. 合肥工业大学先进能源技术与装备研究院，合肥 230009；
3. 合肥工业大学汽车与交通工程学院，合肥 230009）

摘 要：为解决生物质中纤维素、半纤维素和木质素测定时耗时耗力等问题，提出基于支持向量回归机（support vector regression machine，SVR）和热重分析法的生物质三组分含量预测模型。通过对4种不同类型核函数的SVR进行比较，利用K折交叉验证法结合网格搜索法，对SVR的参数进行寻优，以获得最优参数进而训练三组分含量的预测模型，并对该模型进行测试和验证。结果表明：该模型具有较好的预测效果，三组分含量预测模型的相关系数R^2均在0.9532以上；经验证该模型对毛竹、玉米杆和稻草的三组分含量预测绝对误差控制在2.72%以内。

关键词：机器学习；热重分析法；生物质；三组分；预测

中图分类号：TK6 **文献标识码**：A

0 引 言

随着当今社会经济高速发展，资源、环境和能源危机日益突出，生物质能因资源丰富、绿色环保、分布广泛等特点逐渐成为一种可降低温室气体排放、替代化石能源的可再生能源[1-3]。生物质主要由纤维素、半纤维素和木质素3种组分组成[4]，其中纤维素和木质素用于合成可生物降解的聚合物[5]，纤维素还可用于提取纤维素乙醇[6]。因此生物质三组分含量的快速准确测定对加速生物质转化利用有重要意义。

两步硫酸水解方法目前已被广泛应用于三组分含量的测定，其中美国国家可再生能源实验室（National Renewable Energy Laboratory，NREL）法用于测定木质纤维素中三组分含量；美国材料与试验协会E1758-01标准法用于测定生物质中半纤维素含量，但这些方法耗时耗力且成本高[7]。为克服这些缺点，杨海平等[8]针对生物质热解的热重数据，建立预测生物质3种主要组分的含量预测模型。蔡均猛等[9]利用热重分析法（thermogravimetric analysis，TGA）对生物质的纤维素和半纤维素含量进行预测。因生物质中3种组分在热解过程中有不同热解特性，可由生物质热解特性对3种主要组分的含量进行估计并建立相应的模型[8]。故采用TGA快速准确测定三组分含量对加快生物质的利用具有重要意义。

近年来，机器学习随着计算算法的发展逐渐引起人们的关注，在人工智能、计算机科学、化学和生物医药等领域发挥巨大作用[10, 11]。支持向量机（support vector machines，SVM）是以统计学习理论为基础、解决小样本学习问题的机器学习方法，实现结构风险最小化思想，促进机器学习理论的发展。其基本思想是通过一个非线性映射将线性不可分的低维空间数据映射到线性可分的高维空间，并在此空间进行回归和分类，其中用于处理回归问题的SVM称为支持向量回归机（support vector regression machine，SVR）[12]。Raccuglia等[13]利用失败实验数据构建SVM模型来指导无机杂化材料合成，材料合成成功率为89%以上；Fernández等[14]利用SVM和热重分析法实现对树种的分类；Benkedjouh等[15]使用SVR对刀具磨损和剩余寿命进行评估和预测，评估和预测效果较好。

采用SVR和TGA联用的方法对生物质三组分含量进行预测的报道较少。基于此，本文选用不同比例三组分的混合物进行热重实验[16]，通过对4种不同类型核函数的SVR比较，利用K折交叉验证

收稿日期：2016-12-16
基金项目：国家科技支撑计划（2012BAD30B01）
通信作者：邢献军（1964—），男，博士、教授，主要从事生物质能源利用方面的研究。xxianjun@hfut.edu.cn

法（*K*-fold cross validation，*K*-CV）结合网格搜索法，对模型参数进行寻优，以获得最优参数，进而得到不同温度下三组分的失重速率和三组分含量训练模型。随后对该模型的预测效果进行测试，并对棉杆、毛竹和玉米杆三组分含量的预测结果进行验证。

1　SVR原理

通过引入 ε 不敏感损失函数将 SVM 分类应用到 SVM 回归分析中，用于回归分析的 SVM 称为 SVR，其结构如图 1。

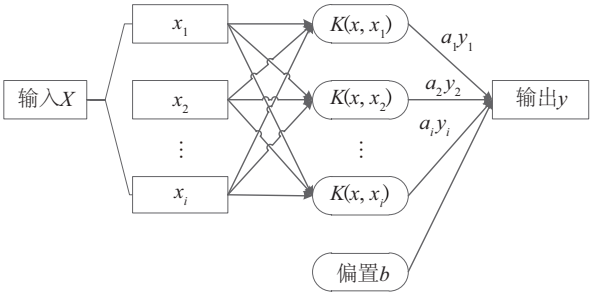


图1　SVR的结构

Fig. 1　Structure of SVR

SVR 通过建立训练集数据中目标向量和支持向量的非线性关系，进而对测试集数据的目标向量进行预测。给定训练集：

$$T=\left\{\left(x_1,y_1\right),\left(x_2,y_2\right),\cdots,\left(x_i,y_i\right)\right\}\in\left(\mathbf{R}_n\times y\right)^l \tag{1}$$

式中，$x_i\in\mathbf{R}^n$；$y_i\in y=\mathbf{R}$；$i=1,\cdots,l$，选取适当的核函数以及适当的精度和惩罚参数 $C>0$[17]。

构造并求解凸二次规划问题：

$$\begin{aligned}&\min_{\alpha^{(*)}\in\mathbf{R}^{2l}}\frac{1}{2}\sum_{i,j=1}^{l}\left(\alpha_i^*-\alpha_i\right)\left(\alpha_j^*-\alpha_j\right)\cdot K\left(x_i,x_j\right)+\\&\quad\varepsilon\sum_{i=1}^{l}\left(a_i^*+a_i\right)\sum_{i=1}^{l}\left(a_i^*-a_i\right)\end{aligned} \tag{2}$$

约束条件：

$$\sum_{i=1}^{l}\left(a_i-a_i^*\right)=0 \tag{3}$$

$$0\leqslant a_i,a_i^*\leqslant C,i=1,\cdots,l \tag{4}$$

构造决策函数：

$$y=\sum_{i=1}^{l}\left(\bar{\alpha}_i^*-\bar{\alpha}_i\right)K\left(x_i,x\right)+b \tag{5}$$

常用的核函数有 4 种：

1）线性核函数：

$$K\left(x,x_i\right)=x^{\mathrm{T}}x_i \tag{6}$$

2）多项式核函数：

$$K\left(x,x_i\right)=\left(g^*x^{\mathrm{T}}x_i+r\right)^p \tag{7}$$

3）RBF 核函数：

$$K\left(x,x_i\right)=\exp\left(-g^*\left\|x-x_i\right\|^2\right) \tag{8}$$

4）Sigmoid 核函数：

$$K\left(x,x_i\right)=\tanh\left(g^*x^{\mathrm{T}}x_i+r\right) \tag{9}$$

式中，α_i、α_i^*——Lagrange 因子；y——决策函数输出值；b——偏置；x——输入样本向量；x_i——核函数中心；g——核函数宽度；r——偏置系数。

2　实　验

2.1　实验原料

实验采用的纤维素为微晶纤维素（CAS 号：9004-34-6），半纤维素由于结构复杂和化学分离较难而普遍采用木聚糖（CAS 号：9014-63-5），木质素为碱性木质素（CAS 号：8068-05-1）。将纤维素、木聚糖和木质素按照一定的比例均匀混合并准确称取 15±0.1 mg 得到 27 组样品；用于模型验证棉杆、毛竹和玉米杆均取自安徽，经粉碎筛分获得 30~80 目粉末制成样品并贮存在 105 ℃的恒温干燥箱中。

2.2　实验装置及方案

热重分析采用法国塞塔拉姆（SETARAM）公司生产的 SETSYS Evo 热重分析仪，通入气体流量为 50 mL/min 的高纯 N_2，采用非等温法加热方式由室温升至 800 ℃，升温速率为 20 ℃/min。实验前做同等条件下空白实验以消除系统误差对实验结果的影响，得到样品的失重曲线，进而求出样品的微商热重（differential thermal gravity，DTG）曲线。

3　三组分含量预测模型构建

图 2 为模型构建流程：包括数据准备及预处理、核函数类型选择、模型参数寻优、模型训练、模型测试和模型验证等[18]。利用 SVR 中核函数强大的非线性处理能力构建基于 SVR 和 TGA 的三组分含量预测模型，采用训练集中输入数据和输出数据间的非线性关系进行模型训练，进一步采用测试集数据对模型进行测试和验证。

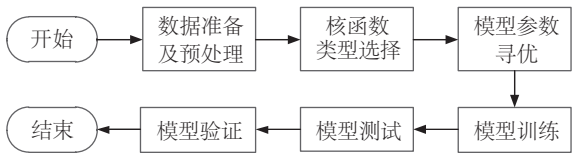


图2　模型构建流程图

Fig. 2　Flow diagram of model building

3.1 数据准备及预处理

文中27组数据样本的输入数据为温度范围在150~800 ℃间样品的失重速率，输出数据为相应混合物的三组分含量。数据样本随机分成22个训练集和5个测试集，用于三组分含量预测模型的构建。为消除样品噪声和漂移的影响，提高SVR模型的泛化能力，在进行三组分含量预测模型训练之前，需对27个样本数据进行[0,1]归一化处理。

3.2 核函数类型选择

在SVM分类基础上引进ε不敏感损失函数，增强了SVR的鲁棒性和稀疏性[19]，因此文中选用ε不敏感损失函数，选取线性核函数，多项式核函数，径向基函数(radial basis function, RBF)核函数和Sigmoid核函数作为SVR的核函数，并根据训练集和测试集数据对核函数类型进行选取。

3.3 模型参数寻优

交叉验证(cross validation，CV)是一种用来验证SVR模型回归性能的统计方法，其基本思想是将原始实验数据进行分组，一部分作为训练集，另一部分作为验证集，先用训练集对模型进行训练，再利用验证集来测试训练得到的模型，把最低均方误差作为评价指标。常用CV法有留出法、留一交叉验证法和*K*-CV法。在SVR核函数类型确定的基础上，模型参数的选取对于三组分含量预测模型的预测精度具有较大影响。本文采用*K*-CV法结合网格搜索法来确定模型参数中惩罚参数C和核函数宽度g的最优值。*K*-CV法是指在进行K折交叉验证时将训练样本分成大致相等的K组，再利用其中$K-1$组样本训练模型，测试其余的样本，依次轮流进行K次。*K*-CV法可有效避免模型过学习和欠学习状态，得到的结果较为可靠，故选用*K*-CV法进行寻优。

3.4 模型训练、测试和验证

利用最优核函数和模型参数将被随机选取的22个训练集用于训练三组分含量预测模型，并用5个测试集数据对模型进行测试。棉杆、毛竹和玉米杆被用于验证模型的有效性。通过SVR和TGA模型预测三组分含量。

本文采用均方差(mean square error，*MSE*)、相关系数R^2和绝对误差e评价该模型。定义公式如式(10)~式(12)所示：

$$MSE=\sum_{i=1}^{n}\frac{(\hat{y}_i-y_i)^2}{n} \tag{10}$$

$$R^2=1-\sum_{i=1}^{n}\frac{(\hat{y}_i-y_i)^2}{(y_i-\bar{y}_i)^2} \tag{11}$$

$$e=\hat{y}_i-y_i \tag{12}$$

式中，n——样本数；y_i——实验值；$\hat{y}_i$——计算值；$\bar{y}_i$—实验值的平均值。

4 结果与分析

4.1 纤维素含量预测模型构建与测试

表1给出了基于不同类型核函数的SVR对纤维素含量训练和测试的结果对比。由表1可知，纤维素含量预测模型采用不同核函数所对应的回归效果存在较大的差异。训练集中线性核函数、多项式核函数和RBF核函数的*MSE*均小于0.0002，R^2均大于0.9988，表明回归效果较好；测试集中RBF核函数的*MSE*和R^2分别是0.0005和0.9714，预测效果明显好于线性核函数和多项式核函数。由此可知，选用RBF核函数的SVR对纤维素含量预测能够取得更准确的结果。

表1 不同核函数的SVR纤维素含量预测模型比较

Table 1 Comparison of SVR prediction model with different kernel functions for cellulose content

项目	参数	核函数			
		线性	多项式	RBF	Sigmoid
训练集	*MSE*	0.0001	0.0001	0.0002	0.0951
	R^2	0.9988	0.9989	0.9988	4.569×10^{-7}
测试集	*MSE*	0.0195	0.0043	0.0005	0.0118
	R^2	0.7596	0.7698	0.9714	0.7664

取$K=3$，采用3-CV法结合网格搜索法对RBF核函数SVR的参数C和g进行寻优，首先确定惩罚参数C的搜索范围为$[2^{-8},2^{8}]$，RBF核函数的核函数宽度g的搜索范围为$[2^{-8},2^{8}]$，得到预测模型的最优参数C和g。参数选择如图3所示，当$\log_2 c=-0.8$，$\log_2 g=-8$，即当$c=0.5743$，$g=0.0039$时交叉验证的均方差有最小值0.0331，此时的C和g值即为预测模型最优参数。纤维素含量预测模型训练结果如图4a所示，可看出构建的纤维素含量预测模型对训练集的均方差*MSE*=0.0002，输出误差较低；相关系数R^2=0.9988，拟合效果较好。

依据表 1 和图 3 确定的核函数类型、惩罚参数和核函数宽度可建立基于 SVR 和 TGA 的纤维素含量预测模型的最佳决策函数。在 150~800 ℃下的不同失重速率作为测试集的输入数据，即可得到 SVR 模型下的纤维素含量预测结果（图 4b）。测试集的

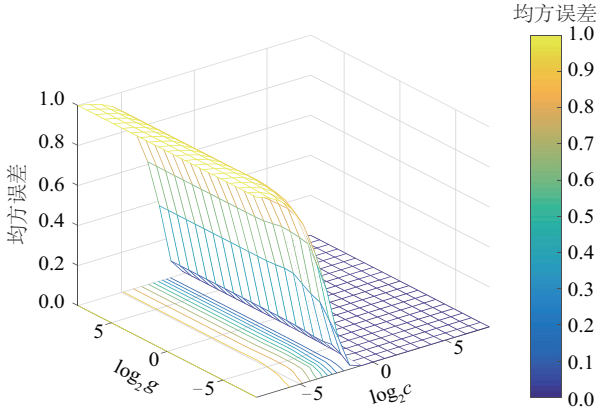
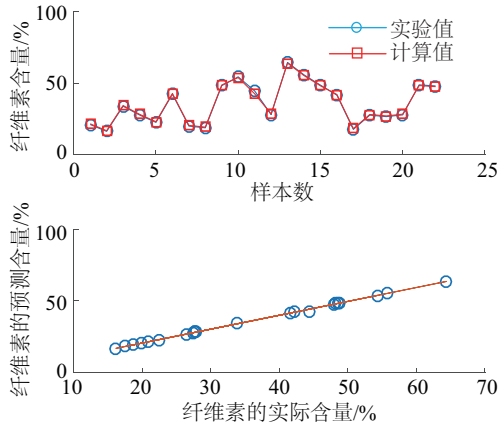
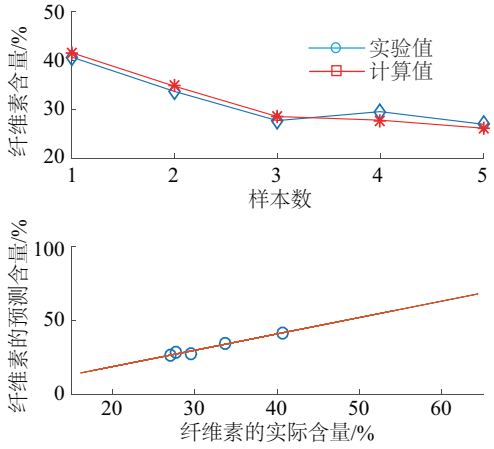


图3　SVR纤维素含量预测模型参数选择

Fig. 3　Parameter selection of SVR prediction model for cellulose content



a. 训练集



b. 测试集

图4　训练集和测试集实测值与计算值对比

Fig. 4　Comparison of the measured value and calculated value from training set and testing set

相关系数 R^2=0.9714，均方差 *MSE*=0.0005。通过对测试集的预测结果分析，SVR 纤维素含量预测模型可以较好地预测纤维素的含量。

4.2　木聚糖含量预测模型构建与测试

由表 2 给出的不同类型核函数的 SVR 对木聚糖含量预测模型训练和预测结果对比，可看出木聚糖含量预测模型采用不同类型核函数所对应的回归效果存在明显差异。训练集中线性核函数、多项式核函数和 RBF 核函数的 *MSE* 和 R^2 明显好于 Sigmoid 核函数的 *MSE*=0.0884 和 R^2=0.0131；测试集中 RBF 核函数的 *MSE* 和 R^2 分别是 0.0030 和 0.9532，预测效果明显好于线性核函数和多项式核函数。可见选用 RBF 核函数的 SVR 木聚糖含量预测模型具有较好的泛化性能，能够较好地描述输入参数和输出参数之间的复杂非线性关系。

表 2　不同核函数的SVR木聚糖含量预测模型比较

Table 2　Comparison of SVR prediction model with different kernel functions for lignin content

项目	参数	核函数			
		线性	多项式	RBF	Sigmoid
训练集	*MSE*	8.894×10^{-5}	0.0001	9.599×10^{-5}	0.0884
	R^2	0.9991	0.9989	0.9995	0.0131
测试集	*MSE*	0.0126	0.0056	0.0030	0.0276
	R^2	0.8617	0.9248	0.9532	0.6901

采用 3-CV 法结合网格搜索法对 RBF 核函数 SVR 模型的参数 C 和 g 进行寻优，确定惩罚参数 C 的搜索范围为$[2^{-8},2^{8}]$，RBF 核函数的核函数宽度 g 的搜索范围为$[2^{-8},2^{8}]$，确定预测模型最优参数 C 和 g。当 $\log_2 c=0$，$\log_2 g=-8$，即 $c=1$，$g=0.0039$ 时交叉验证的均方差有最小值 0.0238，即为模型最优参

数，参数选择如图 5 所示。木聚糖含量预测模型训练结果如图 6a 所示，可看出实验值和计算值大致分布在对角线附近，其中均方差 MSE=9.599×10^{-5}，相关系数 R^2=0.9995，表明该模型具有较好的拟合效果。

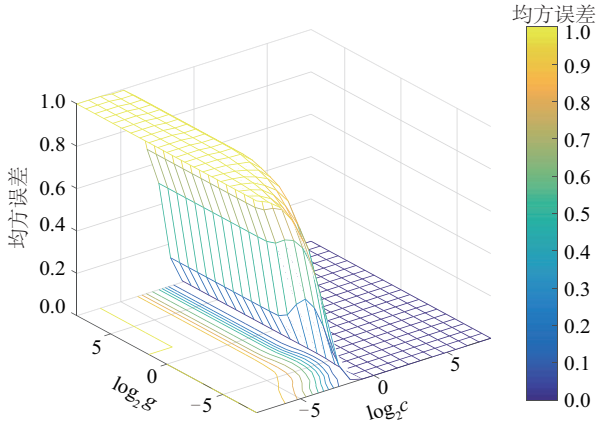
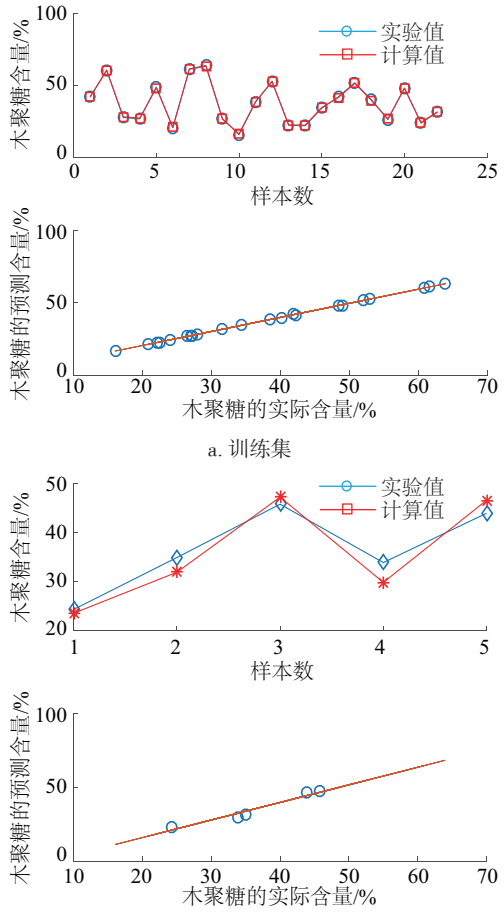


图5 SVR木聚糖含量预测模型参数选择

Fig. 5 Parameter selection of SVR prediction model for lignin content



a. 训练集

b. 测试集

图6 训练集和测试集实测值与计算值对比

Fig. 6 Comparison of measured value and calculated value from training set and testing set

依据表 2 和图 5 确定的核函数类型、惩罚参数和核函数宽度构建基于 SVR 和 TGA 的木聚糖含量预测模型的最佳决策函数。选用测试集中样品 150~800 ℃下不同失重速率为输入数据，得到预测模型下的木聚糖含量预测结果如图 6b。因测试集的相关系数 R^2=0.9532，均方差 MSE=0.0030，得出木聚糖含量预测模型可较好地预测木聚糖的含量。

4.3 木质素含量预测模型构建与测试

由表 3 可知采用不同类型核函数的木质素含量预测模型的回归精度存在较大的差异。训练集中 Sigmoid 核函数训练结果较差（MSE=0.0816，R^2=0.0679）；测试集中 RBF 核函数的 MSE 和 R^2 分别是 0.0017 和 0.9599，预测效果明显好于线性核函数和多项式核函数。因此选用 RBF 核函数的 SVR 木质素含量预测模型能较好地木质素含量。

表3 不同核函数的SVR木质素含量预测模型比较

Table 3 Comparison of SVR prediction model with different kernel functions for xylan content

项目	参数	核函数			
		线性	多项式	RBF	Sigmoid
训练集	MSE	0.0001	0.0001	0.0060	0.0816
	R^2	0.9987	0.9987	0.9566	0.0679
测试集	MSE	0.0932	0.0339	0.0017	0.0600
	R^2	0.6417	0.8979	0.9599	0.9123

采用 3-CV 法结合网格搜索法对 RBF 核函数 SVR 模型参数 C 和 g 进行寻优，确定 C 的搜索范围为[2^{-8}，2^8]，RBF 核函数的 g 的搜索范围为[2^{-8}，2^8]，参数选择如图 7 所示，当 $\log_2 c$=−1.6，$\log_2 g$=−8，即 c=0.3299，g=0.0039 时交叉验证的均方差有最小值 0.0535，C=0.3299 和 g=0.0039 即为预测模型的最优参数。木质素含量预测模型训练结果如图 8a 所示，可看出建立的木质素含量预测模型对训练集的输出误差较低，MSE=0.0060，拟合效果较好，R^2=0.9566，由此可知，该训练模型具有较好的拟合性能。

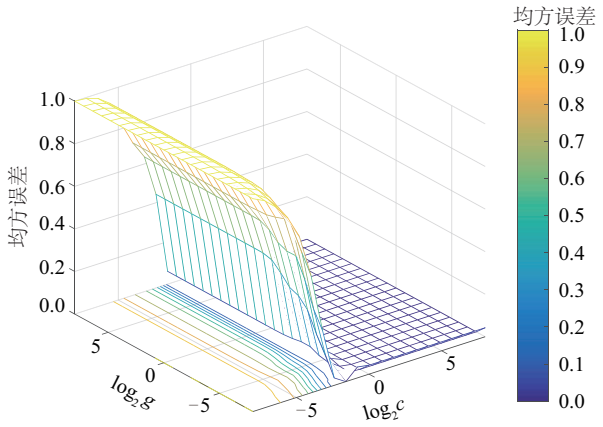


图7　SVR木质素含量预测模型参数选择

Fig.7　Parameter selection of SVR prediction model for xylan content

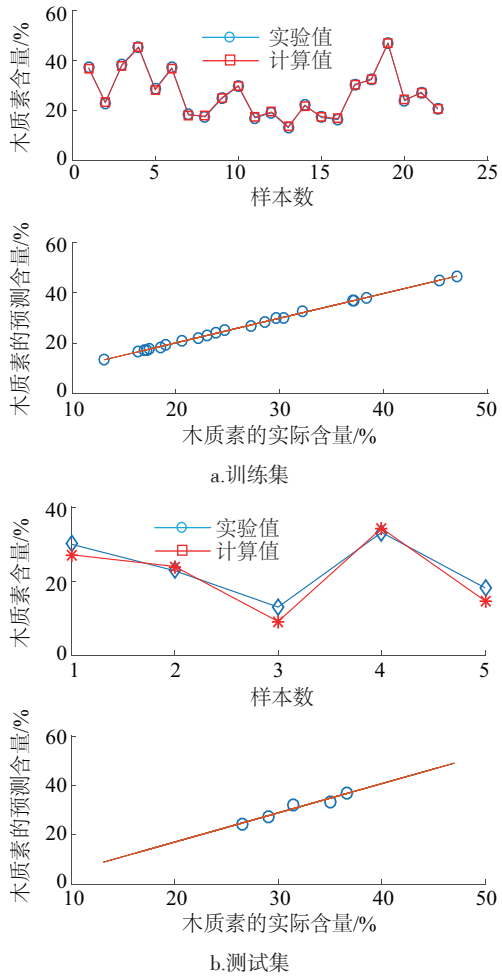


图8　训练集和测试集实测值与计算值对比

Fig. 8　Comparison of measured value and calculated value from training set and testing set

以测试集中样品150~800 ℃下的不同失重速率为输入数据，根据表3和图7已确定的核函数类型、惩罚参数和核函数宽度可构建基于SVR和TGA的木质素含量预测模型的最佳决策函数并得到SVR模型下木质素含量预测结果如图8b。测试集的相关系数 R^2=0.9599，均方差 *MSE*=0.0017。测试集的测试结果分析，SVR木质素含量预测模型可较好地预测木质素的含量，具有较好的拟合和预测能力。

4.4　生物质三组分含量预测模型验证

为进一步验证模型的有效性，采用NREL法测定棉杆、毛竹和玉米杆原料中三组分的含量作为输出数据，采用棉杆、毛竹和玉米杆的输入和输出数据作为验证集。通过该模型对毛竹、玉米杆和稻草的三组分含量进行预测，表4给出了模型预测结果与实际结果的比较分析。

表4　对不同生物质三组分预测的绝对误差对比

Table 4　Comparison of absolute error of prediction value for different biomass three components

生物质	三组分	计算值/%	实验值/%	绝对误差/%
毛竹	纤维素	34.75	32.44	2.31
	半纤维素	27.37	25.69	1.68
	木质素	24.26	22.90	1.36
玉米杆	纤维素	32.16	35.18	−2.02
	半纤维素	28.09	25.37	2.72
	木质素	22.31	19.74	2.57
稻草	纤维素	33.36	35.76	−2.40
	半纤维素	27.36	26.20	1.16
	木质素	12.89	10.66	2.23

基于三组分含量预测模型，对3种生物质三组分含量的计算值和实验值进行比较。从表4可看出：该模型对毛竹的三组分含量预测效果较好，其中对木质素的预测绝对误差为1.36%，对半纤维素预测的误差为1.68%，对纤维素预测误差为2.31%；对玉米杆的三组分含量预测中木质素的绝对误差为2.57%，半纤维素的绝对误差最大为2.72%，纤维素的预测误差为−2.02%；稻草的三组分含量预测中对木质素的预测绝对误差为2.23%，对半纤维素含量预测的绝对误差为1.16%，对纤维素的预测误差为−2.40%。通过对毛竹、玉米杆和稻草预测结果的

分析表明该模型对于实际生物质具有一定的预测效果。

5 结 论

1)结合SVR和TGA方法,研究基于SVR的4种常见核函数类型对三组分含量预测模型的影响,得出RBF核函数更适用于三组分含量预测模型。在核函数类型确定的基础下,采用*K*-CV结合网格搜索法确定模型的最优参数,进而构建三组分含量预测模型。

2)基于该模型对纤维素、木聚糖和木质素含量进行预测,输出误差分别为*MSE*=0.0005、*MSE*=0.0030和*MSE*=0.0017,拟合度分别为R^2=0.9714、R^2=0.9532和R^2=0.9599。表明该模型对三组分含量预测的输出误差小且模型拟合度高,能够较好地预测三组分的含量。

3)利用该模型实现对毛竹、玉米杆和稻草的三组分含量预测,可看出该模型对于实际生物质三组分含量具有一定预测效果,对毛竹、玉米杆和稻草的木质素含量预测的绝对误差均小于2.57%,对于纤维素含量预测的绝对误差均小于2.40%,对于半纤维素含量预测的绝对误差均小于2.72%。

构建基于SVR和TGA的生物质三组分含量预测模型,为生物质三组分含量预测提供一种新颖、简易的方法,同时拓宽热重分析的应用范围。

[参考文献]

[1] Binder J B, Raines R T. Simple chemical transformation of lignocellulosic biomass into furans for fuels and chemicals[J]. Journal of the American Chemical Society, 2009, 131(5): 1979—1985.

[2] Neville A. New technologies advance biomass for power generation[J]. Power, 2012, 156(7): 62—65.

[3] 赵 军, 王述洋. 我国生物质能资源与利用[J]. 太阳能学报, 2008, 29(1): 90—94.

[3] Zhao Jun, Wang Shuyang. Bio-energy resource and its utilization in China[J]. Acta Energiae Solaris Sinica, 2008, 29(1): 90—94.

[4] 李 睿, 金保昇, 仲兆平, 等. 高斯多峰拟合用于生物质热解三组分模型的研究[J]. 太阳能学报, 2010, 31(7): 806—810.

[4] Li Rui, Jin Baosheng, Zhong Zhaoping, et al. Research on biomass pyrolysis three-pseudo component model by Gaussian multi-peaks fitting[J]. Acta Energiae Solaris Sinica, 2010, 31(7): 806—810.

[5] Siracusa V, Rocculi P, Romani S, et al. Biodegradable polymers for food packaging: A review[J]. Trends in Food Science & Technology, 2008, 19(12): 634—643.

[6] Yang B, Wyman C E. Pretreatment: The key to unlocking low-cost cellulosic ethanol[J]. Biofuels Bioproducts & Biorefining, 2008, 2(2): 26—4026.

[7] Hames B R, Thomas S R, Sluiter A D, et al. Rapid biomass analysis[J]. Applied Biochemistry & Biotechnology, 2003, 105(1-3): 5—16.

[8] Yang Haiping, Yan Rong, Chen Hanping, et al. In-depth investigation of biomass pyrolysis based on three major components: Hemicellulose, cellulose and lignin [J]. Energy Fuels, 2005, 20(1): 388—393.

[9] Cai Junmeng, Wu Weixuan, Liu Ronghon, et al. A distributed activation energy model for the pyrolysis of lignocellulosic biomass[J]. Green Chemistry, 2013, 15(5): 1331—1340.

[10] Zubek J, Tatjewski M, Boniecki A, et al. Multi-level machine learning prediction of protein-protein interactions in *Saccharomyces cerevisiae*[J]. Peer-Reviewed & Open Access, 2015, 3(1): 1—21.

[11] Libbrecht M W, Noble W S. Machine learning applications in genetics and genomics[J]. Nature Reviews Genetics, 2015, 16(6): 321—332.

[12] Vapnik V. Statistical learning theory[M]. New York: John Wiley & Sons Inc., 1998.

[13] Raccuglia P, Elbert K C, Adler P D F, et al. Machine-learning-assisted materials discovery using failed experiments[J]. Nature, 2016, 533(7601): 73—76.

[14] Francisco-Fernández M, Tarrío-Saavedra J, Naya S, et al. Classification of wood using differential thermogravimetric analysis[J]. Journal of Thermal Analysis & Calorimetry, 2014, 120(1): 541—551.

[15] Benkedjouh T, Medjaher K, Zerhouni N, et al. Health assessment and life prediction of cutting tools based on support vector regression[J]. Journal of Intelligent Manufacturing, 2015, 26(2): 213—223.

[16] Qu Tingting, Guo Wanjun, Shen Laihong, et al. Experimental study of biomass pyrolysis based on three major components: Hemicellulose, cellulose, and lignin [J]. Industrial & Engineering Chemistry Research, 2011, 50(18): 10424—10433.

[17] Mohammadi K, Shamshirband S, Anisi M H, et al. Support vector regression based prediction of global solar radiation on a horizontal surface[J]. Energy

Conversion & Management, 2015, 91: 433—441.

[18] 王　霞, 王占岐, 金　贵, 等. 基于核函数支持向量回归机的耕地面积预测[J]. 农业工程学报, 2014, 30(4): 204—211.

[18] Wang Xia, Wang Zhanqi, Jin Gui, et al. Land reserve prediction using different kernel based support vector regression[J]. Transactions of the Chinese Society of Agricultural Engineering, 2014, 30(4): 204—211.

[19] 陈善静, 胡以华, 孙杜娟, 等. 基于非线性核空间映射与人工免疫网络的高光谱遥感图像分类[J]. 红外与毫米波学报, 2014, 33(3): 289—296.

[19] Chen Shanjing, Hu Yihua, Sun Dujuan, et al. Classification of hyperspectral remote sensing image based on nonlinear kernel mapping and artificial immune network[J]. Journal of Infrared & Millimeter Waves, 2014, 33(3): 289—296.

A PREDICTION MODEL OF BIOMASS THREE-COMPONENTS CONTENTS BASED ON MACHINE LEARNING AND THERMOGRAVIMETRIC ANALYSIS

Xing Yongqiang[1], Xing Xianjun[2,3], Zhang Jing[1], Li Yongling[1], Zhang Xuefei[1], Zhang Xianwen[2,3]

(1. *School of Mechanical Engineering*, *Hefei University of Technology*, *Hefei* 230009, *China*;
2. *Institute of Advanced Energy Technology & Equipment*, *Hefei University of Technology*, *Hefei* 230009, *China*;
3. *School of Automobile and Transportation Engineering*, *Hefei University of Technology*, *Hefei* 230009, *China*)

Abstract: A prediction model of biomass three-components contents was proposed on basis of support vector regression machine (SVR) and thermogravimetric analysis (TGA) in order to address the existing weakness of time-consuming and labor intensive process for determination of cellulose, hemi-cellulose and lignin of biomass. SVR model with four types of kernel functions were compared and corresponding parameters were optimized through the *K*-fold cross-validation method combined with the grid search method, based on which the prediction model of three components contents using training dataset was built. Subsequently, the steps of model testing and validation were carried out. The results showed that the model had a satisfied performance with correlated coefficient over 0.9532. The absolute error of the model for the prediction of biomass three-component contents was limited within 2.72%. It was shown that the prediction of biomass three-components contents combined machine learning and TGA was successfully achieved and it also extend the application of thermogravimetric analyzer.

Keywords: machine learning; thermogravimetric analysis; biomass; three components; prediction