研究论文

DOI：10.11949/0438-1157.20260133

基于蝙蝠算法与机器学习的渗透气化膜材料高通量筛选与预测

钱锡亮[1,2]，甘嘉添[1,2]，宋宁[1,2]，邹昀[1,2]，童张法[1,2]

（[1]广西大学化学化工学院，广西石化资源加工及过程强化技术重点实验室，广西 南宁 530004；[2]绿色化工新材料广西高校工程研究中心，广西 南宁 530004）

摘要： 渗透气化（PV）是一种低能耗的膜分离技术，在有机溶剂深度脱水中优势显著，然而高性能PV膜材料的开发长期依赖耗时长成本高的实验验证。为实现膜材料的合理设计与高效筛选，本研究设计了一种耦合蝙蝠算法（BAT）与轻量级梯度提升机回归器（LGBMR）的机器学习模型。基于规模化PV数据集，系统对比了BAT与遗传算法（GA）分别耦合随机森林（RF）、极端梯度提升（XGB）和LGBMR三种集成学习模型的预测性能。结果表明，经BAT优化后的LGBMR-BAT模型在测试集上对分离因子和渗透通量的预测性能最优，决定系数（R^2）分别达到0.8891和0.8207，显著优于其他对比模型。特征分析表明，质量比、选择层厚度、操作温度等关键工艺与操作参数，以及分子指纹位点等结构特征，共同对膜性能产生决定性影响。将该模型应用于乙酸乙酯脱水体系的高通量材料筛选，筛选出的高性能聚合物兼具刚性多芳环骨架与适度极性官能团，其性能接近或突破通量-选择性上限。本工作基于LGBMR-BAT框架实现了有机溶剂脱水膜的快速、有效分子设计，为数据驱动的先进膜材料开发提供了新范式。

关键词： 渗透气化；机器学习；蝙蝠算法；轻量级梯度提升机回归器；乙酸乙酯脱水；材料筛选

中图分类号： TQ 028.8　　**文献标志码：** A

文章编号： 0438-1157（2026）10-5975-14

High-throughput screening and prediction of pervaporation membrane materials using bat algorithm and machine learning

QIAN Xiliang[1,2], GAN Jiatian[1,2], SONG Ning[1,2], ZOU Yun[1,2], TONG Zhangfa[1,2]

([1]*Guangxi Key Laboratory of Petrochemical Resource Processing and Process Intensification Technology, School of Chemistry and Chemical Engineering, Guangxi University, Nanning* 530004, *Guangxi, China*; [2]*Guangxi University Engineering Research Center of Green Chemical New Materials, Nanning* 530004, *Guangxi, China*)

Abstract: Pervaporation (PV) is an energy-efficient membrane separation technology that offers distinct advantages in deeply dehydrating organic solvents, while the development of high-performance PV membranes continues to rely on time-consuming and costly experimental verification. To achieve rational design and efficient screening of membrane materials, this study designed a machine learning model coupling the bat algorithm (BAT) and the lightweight gradient boosting regressor (LGBMR). Based on a large-scale literature PV dataset, the predictive performance of BAT and genetic algorithm (GA) respectively coupled with three ensemble learning models—

收稿日期： 2026-01-27　　**修回日期：** 2026-03-01

通信作者： 邹昀(1978—)，男，博士，副教授，zouyun@gxu.edu.cn；童张法(1963—)，男，博士，教授，zftong@gxu.edu.cn

第一作者： 钱锡亮(1999—)，男，硕士研究生，2269565557@qq.com

基金项目： 广西自然科学基金面上项目(2025GXNSFAA0691000)；广西科技重大专项(桂科AA23062019-4)；广西石化资源加工及过程强化技术重点实验室主任基金项目(2024Z006)

引用本文： 钱锡亮, 甘嘉添, 宋宁, 邹昀, 童张法. 基于蝙蝠算法与机器学习的渗透气化膜材料高通量筛选与预测[J]. 化工学报, 2026, 77(10): 5975-5988

Citation: QIAN Xiliang, GAN Jiatian, SONG Ning, ZOU Yun, TONG Zhangfa. High-throughput screening and prediction of pervaporation membrane materials using bat algorithm and machine learning[J]. CIESC Journal, 2026, 77(10): 5975-5988

random forest (RF), eXtreme gradient boosting (XGB), and LGBMR—were systematically compared. The results show that the BAT-optimized LGBMR model performed best among these models on the test set for predicting separation factor and permeation flux, with coefficients of determination (R^2) reaching 0.8891 and 0.8207 respectively, significantly outperforming other comparative models. Feature analysis identified that key process and operation parameters such as mass ratio, selective layer thickness, operating temperature, as well as structural features such as molecular fingerprint sites, collectively exert a decisive influence on membrane performance. When applied to high-throughput screening of materials for the ethyl acetate dehydration, the model identified the high-performing polymers that combine rigid polyaromatic backbone with moderate polar functional groups, with performance approaching or surpassing the flux-selectivity upper bound. This work demonstrates rapid and effective molecular design of organic solvent dehydration membranes based on the LGBMR-BAT framework, providing a new paradigm for data-driven development of advanced membrane materials.

Keywords: pervaporation; machine learning; bat algorithm; light gradient boosting machine regressor; ethyl acetate dehydration; material screening

引 言

渗透气化(PV)技术因其在有机溶剂脱水突破共沸方面的独特优势,已成为一种极具前景的绿色分离方案[1-2]。该技术的核心在于膜材料[3],需兼具高选择性和高渗透通量以实现工业应用的经济性。然而,传统膜材料开发高度依赖“试错法”,通过反复实验调整聚合物结构、填料和工艺参数等,过程耗时耗力,在面对种类繁多的化学结构时难以有效探索[4-5]。

人工智能,特别是机器学习,为打破这一瓶颈提供了新范式。机器学习能够构建材料结构或制备参数与最终性能的复杂映射关系,实现性能的精准预测与虚拟筛选,从而大幅加速研发进程[6-7]。已有研究证明了机器学习在膜材料设计中的潜力,其中随机森林(RF)、极端梯度提升(XGB)与轻量级梯度提升机回归器(LGBMR)三类模型尤为典型[8-9]。RF通过构建多棵决策树,并以平均预测值作为输出,对噪声和异常值不敏感,具备良好的泛化能力与内置特征重要性评估机制[10],可作为稳健的基础模型。XGB通过梯度提升框架集成多棵正则化决策树,利用一阶与二阶梯度信息优化目标函数,对噪声和异常值具有鲁棒性,兼具高精度与内置特征重要性评估能力[11],可作为可靠的回归基模型。LGBMR基于直方图与叶分裂生长策略,在保持高预测性能的同时显著提升计算效率,对大规模数据和高维特征具有强适应性,并提供有效的特征重要性度量。Gao等[9]利用贝叶斯优化指导XGB模型逆向设计聚合物膜,成功获得兼具高渗透性与高选择性的新型PV膜材料,突破了经典权衡限制;Yang等[8]通过比较RF与LGBMR两种回归模型,实现了对百万种聚合物的高效筛选,在保证预测精度的前提下快速识别出适用于乙酸/水体系分离的PV聚合物膜材料,验证了数据驱动策略在复杂材料空间探索中的可行性。

在PV这一多变量耦合的复杂体系中,现有机器学习方法仍面临许多挑战。PV膜的性能不仅受聚合物自身的化学结构影响,还同时受到选择层厚度、交联度、操作温度等多重物化性质与工艺参数的协同影响。尽管分子指纹能够有效将聚合物的结构与拓扑信息数字化[12-13],但其高维特性往往使得模型对超参数高度敏感。若依赖网格搜索易陷入次优解[14],会限制泛化能力与工程适用性[15]。为应对传统膜材料建模方法在复杂多变量体系下存在的模型性能不稳定、泛化能力受限及调参效率低下等问题,需要一种能够自动适配数据特性的智能优化策略。蝙蝠算法(BAT)作为一种受蝙蝠回声定位行为启发的元启发式优化方法,因其参数少、收敛速度快、全局搜索能力强[16]等优势,能够高效探索复杂超参数空间,自动寻找最优或近优配置,但在膜分离领域很少被使用。Alizamir等[16]将BAT应用于气体分离膜性能预测,证实其通过超参数优化提升了模型性能的有效性。然而,该模型所针对的气体分离体系,其传质机理与关键输入特征与PV过程存在本质差异。PV遵循溶解-扩散机制[17],且受亲疏水性[18]、溶胀效应及温度敏感性[19]等多变量耦合影响,直接将BAT优化策略用于PV膜筛选的工作尚未见报道。

本文将BAT算法引入PV膜性能预测模型的超参数自动优化过程。在此基础上，系统比较BAT与遗传算法（GA）分别耦合RF、XGB和LGBMR三种集成学习模型的PV膜性能预测表现；进而将性能最优的模型应用于乙酸乙酯脱水PV膜的高通量筛选，旨在建立一种数据驱动与机理认知协同的新型膜材料设计范式。

1 方 法

本文提出了一种耦合机器学习与群智能优化算法的方法，用于研究聚合物PV膜在脱水过程中的性能。基于机器学习和优化算法的聚合物筛选流程如图1所示，流程包括数据集预处理、模型训练与超参数优化、模型评估与物理解释机制和聚合物筛选4个步骤。

1.1 数据集预处理

采用Yang等[8]整理的PV实验数据作为输入数据集，该数据集包含2341组聚合物膜在有机物/水体系中的分离性能数据。数据涵盖聚乙烯醇（PVA）、聚二甲基硅氧烷（PDMS）和海藻酸钠（SA）等52种常见聚合物。所有样本包含膜结构参数、有机物参数、操作条件及分离因子与渗透通量的性能指标，以确保建模所需信息的完整性。在预处理阶段，将数据集按8∶2的比例随机划分为训练集与测试集。为了确保输入特征的真实性和模型的泛化能力，本研究根据变量的物理意义与实际工况采取相应的处理方式。对于填料尺寸、填料浓度及交联剂浓度等可选性工艺参数，其缺失值为0表示该膜体系未引入填料或未进行交联处理；而对于选择层厚度、溶解度参数等本征属性或操作温度、质量比等操作条件，为了体现典型实验条件下的基准水平，避免偏差过大，缺失值则设定为训练集对应变量的均值。PV相关输入特征变量共12个，分为膜材料参数、有机物组分物理性质和操作条件3类，其中膜材料参数包括膜组件的构型、选择层厚度（μm）、填料尺寸（nm）、填料质量分数（%）、交联剂质量分数（%）和有效面积（cm^2），有机物组分物理性质包括溶解度参数（$MPa^{1/2}$）、摩尔体积（$cm^3 \cdot mol^{-1}$）和分离类型，操作条件包括温度（℃）、压力[mbar（1 bar=0.1 MPa）]和进料中目标组分与水的质量比。在数据处理中，膜组件构型作为分类变量进行编码，平板膜编码为1，管式膜编码为0。在分离类型的编码中，膜用于水优先渗透分离体系的编码为1，用于有机物优先渗透分离体系的编码为0。模型输出变量为分离因子与渗透通量，为提升模型的收敛速度与预测精度，需对两个输出变量进行对数变换。数据集的参数统计汇总于表1中。本文使用的数据集在关键实验参数上呈现出较大的离散性，其标准差偏高。这一宽泛分布并非数据噪声所致，而是为了反映PV膜在真实研发与应用场景中的高度多样性，且已通过对特征进行归一化处理，从而增强模型的适应性与普适性。

为表征聚合物膜材料的分子结构特征，采用Morgan指纹[20]对重复单元的SMILES字符串进行编码。SMILES是一种基于字符串的分子表示方法，能够以简洁的ASCII文本形式描述分子的原子类型、连接顺序及拓扑结构。以SMILES序列作为输入，利用RDKit化学信息学工具包生成固定长度的二进制指纹向量。RDKit是一个开源的化学信息学软件工具包，广泛应用于分子结构处理、特征提取与定量构效关系建模等任务。设置原子环境扩展半径为2及指纹维度为1024位，即分子化学结构的特征变量数目为1024个。生成每个分子的Morgan指纹，将所得位向量转换为长度为1024的二进制字符串，作为该分子的结构描述符。该指纹方案能够有效

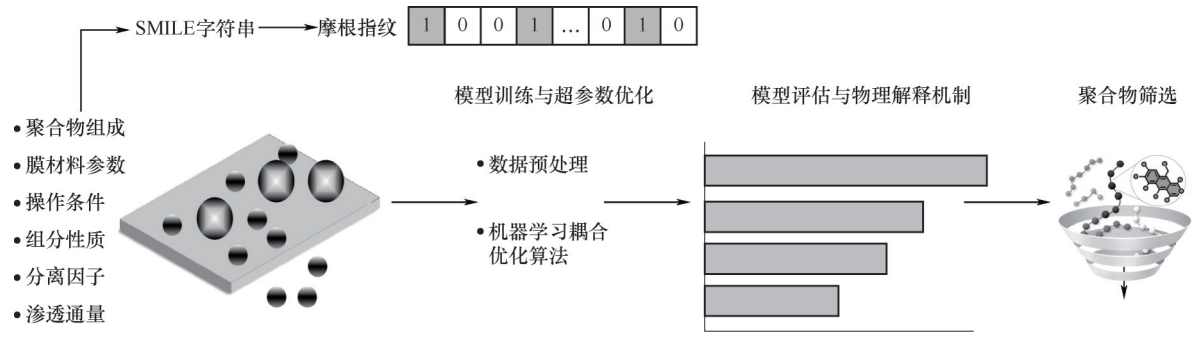


图1 基于机器学习和优化算法的聚合物PV膜筛选流程图

Fig.1 Polymer PV membrane screening flowchart based on machine learning and optimization algorithms

表1 数据集的参数统计

Table 1 Parameter statistics of the dataset

数据集	单位	训练集			测试集		
		范围	平均值	标准差	范围	平均值	标准差
选择层厚度	μm	0.05~1400.00	66.06	152.35	0.10~1400.00	57.33	111.99
填料尺寸	nm	0.00~8000.00	240.13	731.85	0.00~8000.00	282.12	894.09
填料质量分数	%	0.00~67.00	4.35	10.65	0.00~67.00	4.35	10.58
交联剂质量分数	%	0.00~50.00	1.86	4.80	0.00~25.00	1.76	4.23
有效面积	cm^2	1.43~1600.00	31.97	65.06	0.24~1600.00	42.36	146.01
溶解度参数	$MPa^{1/2}$	10.00~32.95	24.09	3.32	12.14~32.95	23.05	3.10
摩尔体积	$cm^3·mol^{-1}$	37.80~128.60	70.72	26.09	40.49~128.60	71.38	24.73
质量比(A/B)	—	0.01~13182.30	70.03	533.11	0.01~8787.89	92.97	622.28
操作温度	℃	15.00~120.00	44.10	16.36	15.00~100.00	43.07	15.95
实验压力	mbar	0.00~333.30	9.59	31.37	0.00~266.60	8.99	28.40
分离因子	—	0.20~16481.60	45.60	8.41	0.20~16481.60	41.67	7.92
渗透通量	$kg·m^{-2}·h^{-1}$	0.0001~12.9300	0.1823	5.5861	0.0001~11.9400	0.1936	4.7108

捕获分子局部子结构及其拓扑环境,提供适用于机器学习模型的数值化输入。

1.2 模型构建与超参数优化

通过调用Python中Sklearn机器学习开源库[8]构建了RF、XGB、LGBMR三种主流核心回归模型,并使用GA、BAT两种群智能优化算法代码,构建了基于智能优化算法的机器学习模型算法框架。其中包括独立的RF、XGB、LGBMR三种基础模型和RF-GA、RF-BAT、XGB-GA、XGB-BAT、LGBMR-GA、LGBMR-BAT六种耦合模型,所有模型均在统一的数据集下运行训练。图2是机器学习耦合智能优化算法框架流程图。首先对原始数据集进行划分,并初始化算法参数。在迭代过程中,由GA或BAT生成候选超参数组合S_i,并以此训练RF、XGB或LGBMR等模型。采用5折交叉验证策略,在训练集上计算模型预测值与真实值之间的均方根误差(RMSE),以其平均值作为BAT算法的适应度函数。算法通过不断更新个体位置或基因结构,逐步逼近最优超参数组合。种群规模设为30,最大迭代次数为50。当满足收敛条件(适应度值连续10代未改善或达到最大迭代次数)时,提取全局最优超参数组合。全局最优超参数组合优化完成后,在训练集上

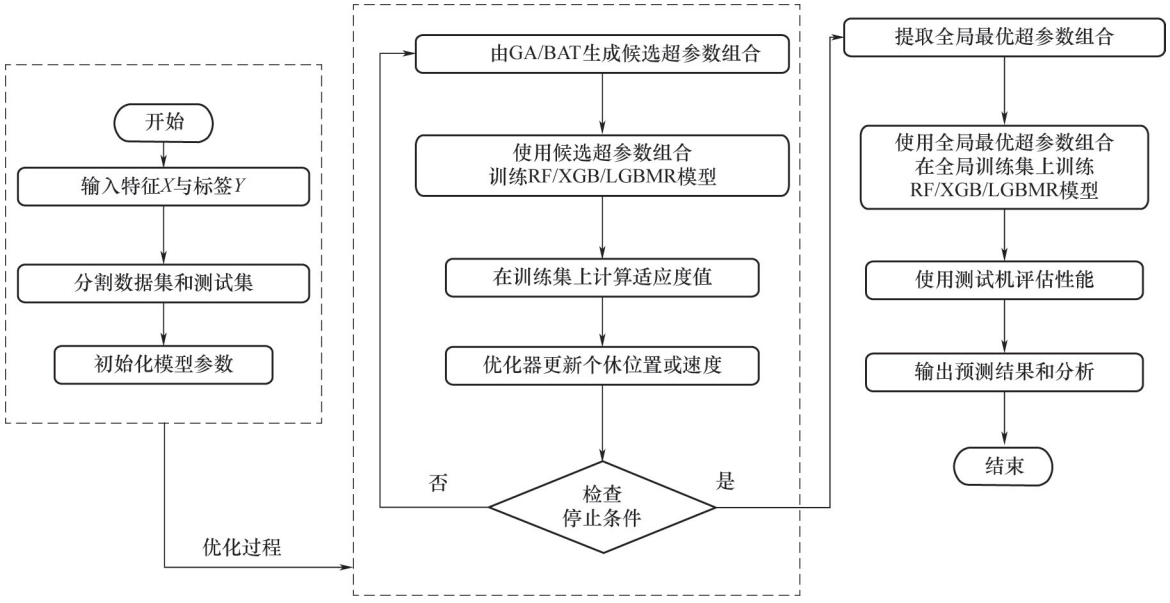


图2 机器学习耦合智能优化算法框架流程图

Fig.2 Flowchart of the machine learning coupled with intelligent optimization algorithm framework

再次训练对应的回归模型，并在独立测试集上评估其预测性能，评估指标包括均方根误差（RMSE）、相关系数（R^2）、平均绝对误差（MAE）和残差预测偏差（RPD）及最终保留的特征数量。

其中GA算法遵循“适者生存”机制，通过选择、交叉、变异[21]对二进制编码的个体不断演化，利用适应度差异驱动种群向高适应度区域收敛；BAT算法模拟蝙蝠回声定位原理，在特征空间中高效搜索全局最优解[22]，每只“蝙蝠个体”通过调整频率发出脉冲，并依据回声调整位置与响度，响度与脉冲发射率共同调控算法在局部精细搜索与全局跳跃之间的行为，从而平衡广域探索与邻域挖掘能力[16]。与粒子群优化算法（PSO）不同，BAT算法不依赖人工设定或线性调整的惯性权重来平衡探索与开发，而是通过响度衰减与脉冲发射率自增长的内在反馈机制，实现搜索行为的自主动态调节，这显著提升了BAT算法对复杂优化问题的适应能力。此外，BAT算法引入了基于当前最优解的局部随机扰动策略，而PSO算法虽然收敛速度快但缺乏此类局部开发机制[23]，因此在多函数优化中更易陷入局部最优。相比之下，BAT算法能更有效地跳出局部极值，增强全局搜索的鲁棒性。

在算法中，每只“蝙蝠个体”i由位置x_i、速度v_i、频率f_i、响度A_i、脉冲发射率r_i等5个状态变量描述，其迭代过程遵循频率动态调整、基于频率的速度与位置更新、局部随机扰动、响度与脉冲率的自适应演化4个规则，分别由式（1）~式（4）表示：

$$f_i = f_{\min} + (f_{\max} - f_{\min})\beta, \beta \in (0,1) \tag{1}$$

$$v_i^t = v_i^{t-1} + (x_i^{t-1} - x^*)f_i, x_i^t = x_i^{t-1} + v_i^t \tag{2}$$

$$x_{\text{new}} = x^* + \epsilon A_t, \epsilon \sim [-1,1] \tag{3}$$

$$A_i^{t+1} = \alpha A_i^t, \alpha \in (0,1)$$

$$r_i^{t+1} = r_i^0\left[1 - \exp(-\gamma t)\right], \gamma > 0 \tag{4}$$

其中，式（1）是频率调整机制，个体根据式（1）动态调整步长，f_i越高，搜索步长越小，越利于局部精细探索；反之，则跳跃范围越大，增强全局搜索能力。式（2）是速度与位置更新机制，其中x^*当前最优解，此形式虽与PSO相似，但引入了f_i，使个体运动更具随机性与多样性，有效避免陷入局部最优。式（3）通过对x^*施加局部扰动，基于平均响度A_i进行邻域精细开发寻找新解，若新解优于当前解且满足随机条件，则接受新解，并根据式（4）更新A_i和r_i，在初期高响度、低脉冲率时促进广域探索；后期低响度、高脉冲率则强化邻域开发，从而无须人工干预即可自动平衡全局探索与局部开发能力。

为确保研究的可重复性，本文对所采用的机器学习模型与超参数优化算法设定了关键超参数的搜索空间或固定值，表2汇总了各方法的关键参数设置，其他参数为默认值。

表2 机器学习模型及超参数优化算法的关键参数配置

Table 2 Key parameter settings of machine learning models and algorithm for optimizing hyperparameters

类别	方法	关键参数	参数定值或范围
机器学习模型	RF	决策树数量（n_estimators）	[50,500]
		最大深度（max_depth）	[2,50]
		最小分割样本数（min_samples_split）	[2,10]
		叶节点最小样本数（min_samples_leaf）	[1,10]
	XGB	决策树数量（n_estimators）	[50,500]
		最大深度（max_depth）	[2,50]
		学习率（learning_rate）	[0.01,0.30]
		子采样比例（subsample）	[0.6,1.0]
	LGBMR	决策树数量（n_estimators）	[50,500]
		学习率（learning_rate）	[0.01,0.30]
		最大深度（max_depth）	[2,50]
		叶子节点数（num_leaves）	[3,150]
		子采样比例（subsample）	[0.6,1.0]
优化算法	BAT	种群规模	30
		最大迭代次数	50
		频率范围	[0,2]
	GA	种群规模	30
		最大迭代次数	50
		交叉概率	0.1
		变异概率	0.2

1.3 模型性能评估与物理解释机制

为全面评估模型性能，采用RMSE、R^2、MAE和RPD四种评估指标[8]。其中，RMSE通过计算预测值与真实值之间偏差的标准差，衡量模型预测的整体误差水平，其对较大误差较为敏感，能够有效反映模型在极端样本上的预测稳定性。MAE通过计算预测误差的平均绝对值，反映模型的平均预测偏差，其数值单位与目标变量一致，结果直观且对异常值不敏感，适用于评估整体稳健性。R^2表征模型对目标变量总变异的解释比例，其取值范围通常为[0,1]，用于衡量模型的整体拟合优度。RPD定义为测试集因变量的标准差与RMSE的比值，是一种无量纲的综合性能评价指标。在化学计量学与材料

信息学中，RPD常用于评估模型的实用预测能力。通常认为，RPD大于2.0且小于3.0时，表明模型具有良好的预测性能；RPD大于3.0时，代表预测性能优异，可用于实际应用。其数学表达式分别见式(5)~式(8)：

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i-\overline{Y}_i\right)^2} \tag{5}$$

$$R^2=1-\frac{\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i-\overline{Y}_i\right)^2} \tag{6}$$

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|Y_i-\hat{Y}_i\right| \tag{7}$$

$$\mathrm{RPD}=\frac{\mathrm{SD_Y}}{\mathrm{RMSE}} \tag{8}$$

式中，n表示样本的数量；Y_i表示第i个样本的真值；$\hat{Y}_i$表示预测值；$\overline{Y}_i$表示真值的平均值；$\mathrm{SD_Y}$表示预测集中样本的标准差。

为增强模型的可解释性，本文采用Python的SHAP库计算Shapley值，以量化各特征对模型预测的贡献，并通过可视化手段揭示其决策机制。SHAP方法基于合作博弈模型中的Shapley值，通过评估每个特征在所有可能特征子集组合下对模型输出的边际贡献，提供理论上一致且局部准确的特征重要性度量[24]。为进一步阐释SHAP分析所识别的关键聚合物描述符背后的化学本质，本文对具有高贡献度的分子指纹位点进行溯源，还原其对应的原始ECFP化学子结构，从而建立机器学习预测结果与具体分子结构特征之间的可解释性关联。

1.4 乙酸乙酯脱水的高性能聚合物膜筛选

乙酸乙酯是化工和制药工业中常用的典型溶剂，但其与水形成共沸物，导致传统精馏提纯过程能耗高、操作复杂[25-26]。这一挑战促使研究者致力于开发高性能聚合物脱水膜，以实现高效节能的分离。为系统评估聚合物在乙酸乙酯/水体系PV脱水过程中的分离性能，本研究采用经过优化算法筛选得到的最优超参数组合，构建了最终的机器学习预测模型，并将其应用于大规模聚合物数据库PI1M[27]，进行候选材料筛选。针对数据库中的每种聚合物，首先基于其分子结构生成1024位Morgan指纹作为化学特征表示，进而计算该聚合物与已知参照物的Tanimoto相似性得分[28]，仅保留得分不低于0.6的聚合物结构用于后续筛选。Tanimoto相似性的数学表达式如式(9)所示。

$$T_{\mathrm{AB}}=\frac{N_{\mathrm{AB}}}{N_{\mathrm{A}}+N_{\mathrm{B}}-N_{\mathrm{AB}}} \tag{9}$$

式中，T_{AB}表示A、B两分子间的结构相似性，取值范围为[0,1]；N_{A}与N_{B}分别为分子A与分子B的Morgan指纹中取值为1的特征位总数；N_{AB}为两分子指纹中共同取值为1的特征位总数。

将乙酸乙酯的溶解度参数、摩尔体积及分离类型等物性参数输入模型中，对于膜材料参数与工艺条件参数等非结构类变量，统一采用训练集中对应变量的平均值进行赋值。加载经优化算法后的模型，读取包含分子指纹与工艺参数的新数据集，获得分离因子与渗透通量的预测值，并计算PV分离指数(PSI)作为聚合物筛选的关键指标。基于预测结果，绘制分离因子与渗透通量的双对数坐标图，用于直观识别兼具高选择性与高通量的优选膜材料。

2 结果与讨论

2.1 机器学习模型对分离因子的预测性能分析

表3列出了前文所述的9种模型在测试集上对分离因子的预测性能。所有经超参数优化的模型均优于其对应的基础原始模型，验证了群智能算法在提升PV膜性能预测精度方面的有效性。其中，LGBMR-BAT表现最优，其RMSE为0.6891，R^2为0.8891，RPD为3.0061，这组数据表明该模型具备优异的预测能力与实际筛选价值。RF整体性能相对较弱，XGB模型性能介于二者之间。此外，BAT算法在优化LGBMR和XGB模型时均优于GA算法，说明BAT凭借其高效的全局-局部平衡搜索机制，更

表3 3种原始模型与6种优化超参数耦合模型在分离因子预测中的4种性能对比

Table 3 Comparison of four performance metrics for separation factor prediction among three original models and six optimized hyperparameter coupled models

方法	RMSE	MAE	R^2	RPD
RF	0.8278	0.5278	0.8400	2.5026
RF-GA	0.8193	0.5245	0.8432	2.5284
RF-BAT	0.7814	0.5022	0.8574	2.6510
XGB	0.8026	0.5001	0.8496	2.5811
XGB-GA	0.7166	0.4379	0.8801	2.8909
XGB-BAT	0.7108	0.4289	0.8820	2.9143
LGBMR	0.8170	0.5480	0.8441	2.5355
LGBMR-GA	0.7002	0.4408	0.8855	2.9587
LGBMR-BAT	0.6891	0.4321	0.8891	3.0061

适用于PV膜性能预测任务中的超参数空间精细调优。

图3是9种耦合模型对分离因子的真实值与预测值散点分布图。其中横坐标$\lg\alpha_{real}$为真实值，纵坐标$\lg\alpha_{pre}$为模型预测值，虚线为$y=x$对称轴，实线为回归拟合线，用于评估模型的系统偏差与拟合度。对比可见，LGBMR-BAT模型的散点最接近对称轴，其线性回归系数(r)最高，达到0.9434；LGBMR-GA模型的r值为0.9415，二者显著优于其余模型。

图4展示了LGBMR-BAT与LGBMR-GA在超参数优化过程中的适应度值随迭代次数的变化情况。二者收敛趋势存在明显差异。BAT算法在初始阶段即表现出较快的收敛速度，前10次迭代内RMSE从0.6878降至0.6860以下，随后持续优化并最终稳定在约0.6857；相比之下，GA的收敛速度较慢，且在整个优化过程中波动较为显著。该结果表明，BAT算法在本研究任务中具有更强的全局搜索能力与优化稳定性，能够更有效地为LGBMR模型

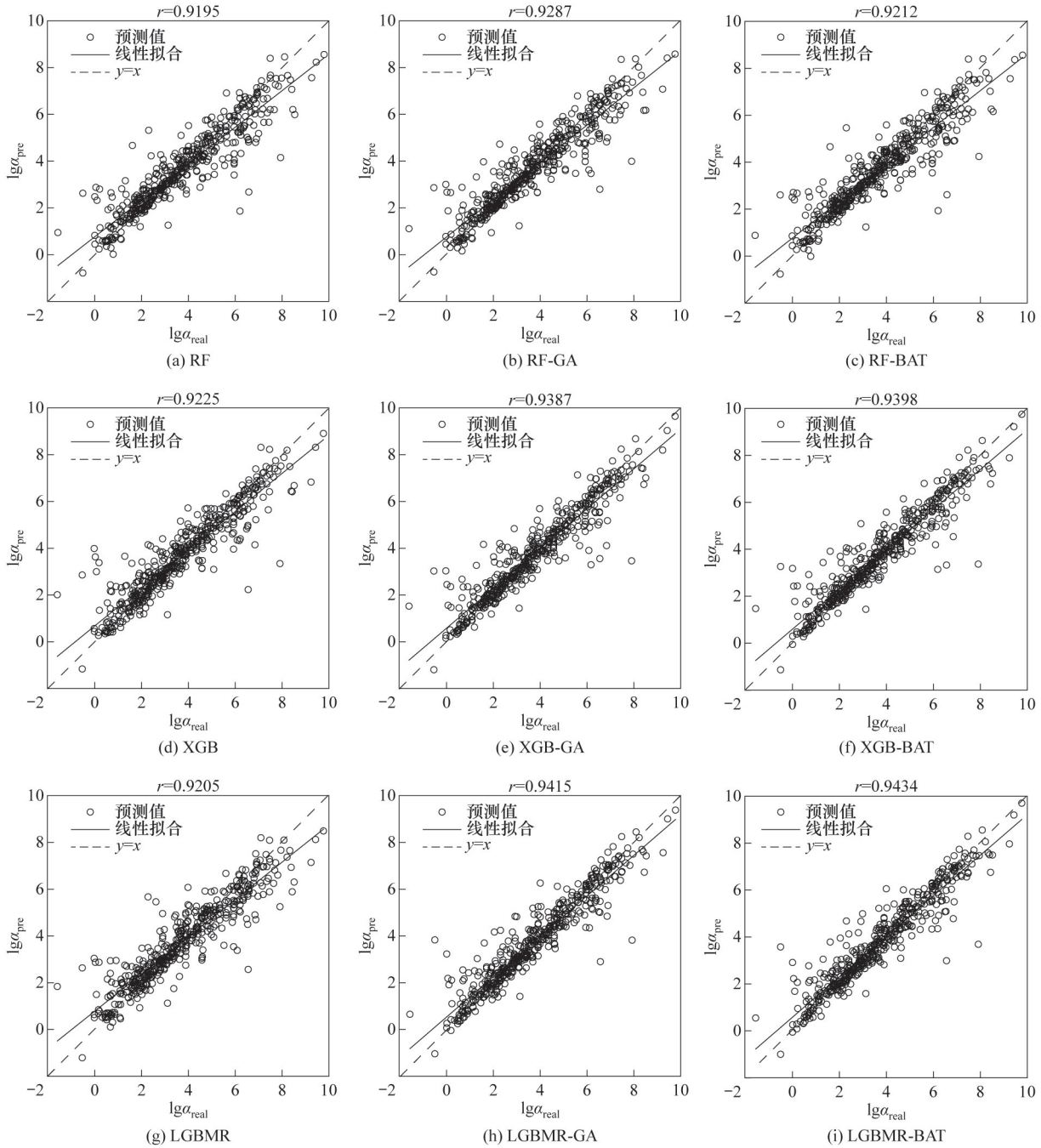


图3 9种模型分离因子的真实值和预测值散点图

Fig.3 Scatter plots of true *versus* predicted separation factor across nine models

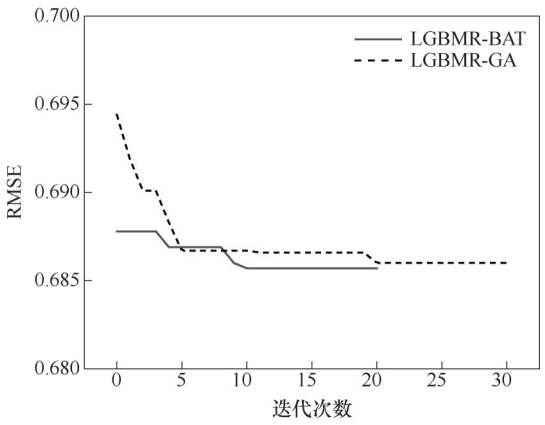


图4 LGBMR模型在BAT与GA超参数优化过程中的适应度收敛曲线

Fig.4 Fitness convergence curves of LGBMR model during BAT and GA hyperparameter optimization process

寻得高性能的超参数配置。因此,综合考虑后,本研究选用LGBMR-BAT进行聚合物分离因子的筛选。

2.2 机器学习模型对渗透通量的预测性能分析

表4列出了9种建模策略在测试集上对渗透通量的预测性能。结果显示,3种集成学习模型在经过超参数优化后,其性能均得到了不同程度的提升。其中,基础LGBMR模型表现最差,R^2只有0.7573;但经BAT优化后性能显著提升,LGBMR-BAT取得了最优的综合指标:RMSE为0.6584,R^2为0.8207,RPD为2.3644。这表明LGBMR模型对超参数较为敏感,且BAT算法能够有效挖掘其性能潜力。相比之下,XGB系列的整体表现较为稳健;而RF模型无论是否经过优化,其性能均相对较低,R^2在0.7900左右,且GA与BAT算法对其提升幅度有限。

图5展示了9种耦合模型中渗透通量的真实值与预测值散点分布。其中横坐标$\lg J_{real}$为真实值,纵坐标$\lg J_{pre}$为模型预测值,LGBMR-BAT与LGBMR-GA表现较好,r分别为0.9060和0.9023。鉴于二者在最终预测性能上较为接近,为深入评估两种优化策略的效率,本研究进一步对比了它们在超参数搜索过程中的收敛行为。

图6展示了LGBMR-BAT与LGBMR-GA在渗透通量预测任务中超参数优化过程中适应度值变化情况。二者的优化过程呈现显著差异:BAT算法在前5次迭代内迅速将RMSE从0.6039降至0.5945,并在后续迭代中保持稳定,表现出优异的收敛速度与搜索稳定性;而GA优化后的RMSE始终高于

表4 3种原始模型与6种优化超参数耦合模型在渗透通量预测中的4种性能对比

Table 4 Comparison of four performance metrics for permeation flux prediction among three original models and six optimized hyperparameter coupled models

方法	RMSE	MAE	R^2	RPD
RF	0.7225	0.4413	0.7841	2.1544
RF-GA	0.7074	0.4435	0.7930	2.2005
RF-BAT	0.7080	0.4445	0.7927	2.1988
XGB	0.6790	0.4150	0.8094	2.2928
XGB-GA	0.6764	0.3750	0.8108	2.3014
XGB-BAT	0.6709	0.3922	0.8139	2.3204
LGBMR	0.7661	0.4947	0.7573	2.0320
LGBMR-GA	0.6715	0.4068	0.8135	2.3182
LGBMR-BAT	0.6584	0.3750	0.8207	2.3644

BAT。结果表明,BAT显著提升了优化效率,并实现了更优的全局搜索性能,说明其在处理高维超参数空间时具备更强的适应性与鲁棒性。因此,本研究最终选用LGBMR-BAT模型进行聚合物渗透通量的筛选。

2.3 机器学习预测结果的化学可解释性分析

为了深入理解模型的决策机制并揭示影响膜性能的关键物理化学因素,本研究采用SHAP方法对最优模型LGBMR-BAT进行了可解释性分析,结果如图7所示。其中,图7(a)、(b)展示了各输入特征的全局SHAP值分析图,图7(c)、(d)进一步揭示了特征值高低对预测结果的系统性影响。在分离因子预测中,质量比是最重要的特征,其高值对应显著的正向SHAP值,表明提高目标组分比例可有效提升膜的选择性。分离类型的重要性排名位列第二,SHAP分析表明,脱水任务普遍贡献正向影响,而脱有机物任务则多呈现负SHAP值,这可能是由于在脱有机物体系中,极性有机物与亲水膜之间的强相互作用易引发非选择性溶胀,从而削弱膜的选择性;有机物分子体积位列第三,其高值带来正向的SHAP值,体现了分子尺寸筛分效应的积极作用。有效面积与渗透通量呈正相关,面积增大可提供更多传质路径[29]。

在渗透通量预测中,主导因素转变为膜结构参数与操作条件。选择层厚度最为关键,其低值与正向SHAP值强烈关联,直接影响传质阻力。根据Fick扩散定律,渗透通量与选择层厚度成反比[30],因此减小选择层厚度是提高通量的有效方法之一。操作温度和质量比(A/B)也具有显著影响:温度升

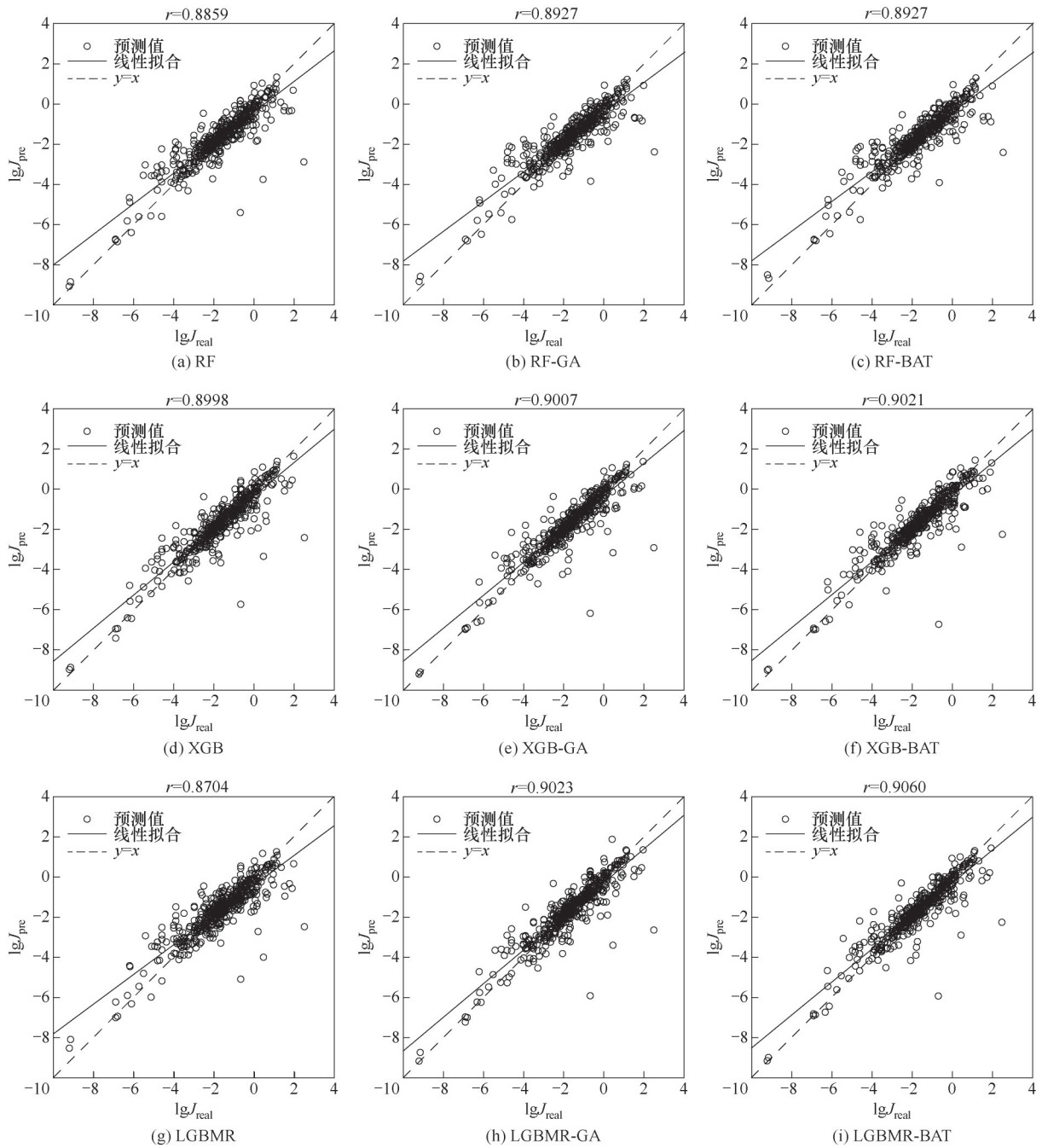


图5 9种模型的真实值和预测值渗透通量散点图

Fig.5 Scatter plots of the true and predicted permeation fluxes of the nine models

高对应正向SHAP值，表明温度通过增强高分子链段运动，扩大膜内自由体积[31]，可指数级提升组分扩散系数，并增强下游气化推动力，从而显著促进渗透通量。调节实验温度是强化膜性能的有效手段，但在实际应用中，必须在膜的热稳定性和过程经济性之间取得平衡，避免因温度过高导致膜结构破坏或能耗显著增加。渗透侧压力作为传质驱动力的直观体现，与渗透通量成反比，降低渗透侧压力可显著提高通量[32]。高填料浓度则集中于负SHAP区域，表明过量填料可能造成孔道堵塞或界面缺陷，反而抑制渗透性能的提升。

本研究基于SHAP特征分析，识别出4个对膜性能具有显著影响的分子结构位点，其结构如图8所示。其中，位点_807和位点_137是决定乙酸乙酯脱水分离因子的关键因素，二者的SHAP值均大于0，表现出显著的正向贡献。这表明在相应位点引入亲水性羟基[33]或离子化羧酸钠基团，可有效增强膜对水分子的选择性吸附与传递，从而提升分离性

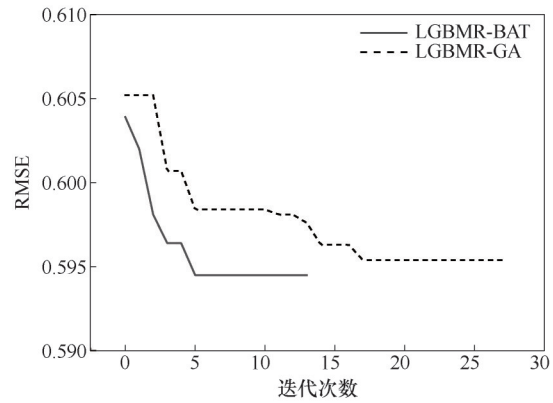


图6 LGBMR模型在BAT与GA超参数优化下的收敛曲线对比

Fig.6 Comparison of convergence curves for LGBMR model during BAT and GA hyperparameters optimization process

能。位点_332和位点_879主要影响渗透通量。位点_332的碳碳双键结构虽因增强刚性而略微降低选择性，但可通过增大链间自由体积显著促进整体渗透速率；而位点_879的氯代烷基虽具弱疏水性，却因适度增加链段间距而对通量产生轻微正向作用。

综上所述，通过引入强极性与多氢键官能团[34]并调控链段柔性，可在分子层面实现对水-有机物分离行为的有效调控。该结论不仅为理解膜材料构效关系提供了化学层面的可解释性依据，也为面向乙酸乙酯脱水体系的新型聚合物的设计提供了数据驱动的指导原则。

2.4 面向乙酸乙酯脱水的高性能聚合物膜筛选

为聚焦于结构新颖且具有潜在应用价值的候选材料，本研究首先基于Tanimoto相似性对原始PI1M虚拟聚合物库进行去冗余初筛，筛选出与已知高性能聚合物 T_{AB}≥0.6的结构，最终保留21830个多样性较高的虚拟聚合物，作为后续LGBMR-BAT模型高通量性能预测的输入集合。如图9所示，基于LGBMR-BAT模型对PI1M虚拟聚合物库进行的分离性能预测结果呈现出典型的性能权衡关系：绝大多数聚合物集中在中等分离因子与中等通量区间，而少数材料表现出卓越性能。右上方区域红色框内聚集了一批高性能候选聚合物，其性能显著优于

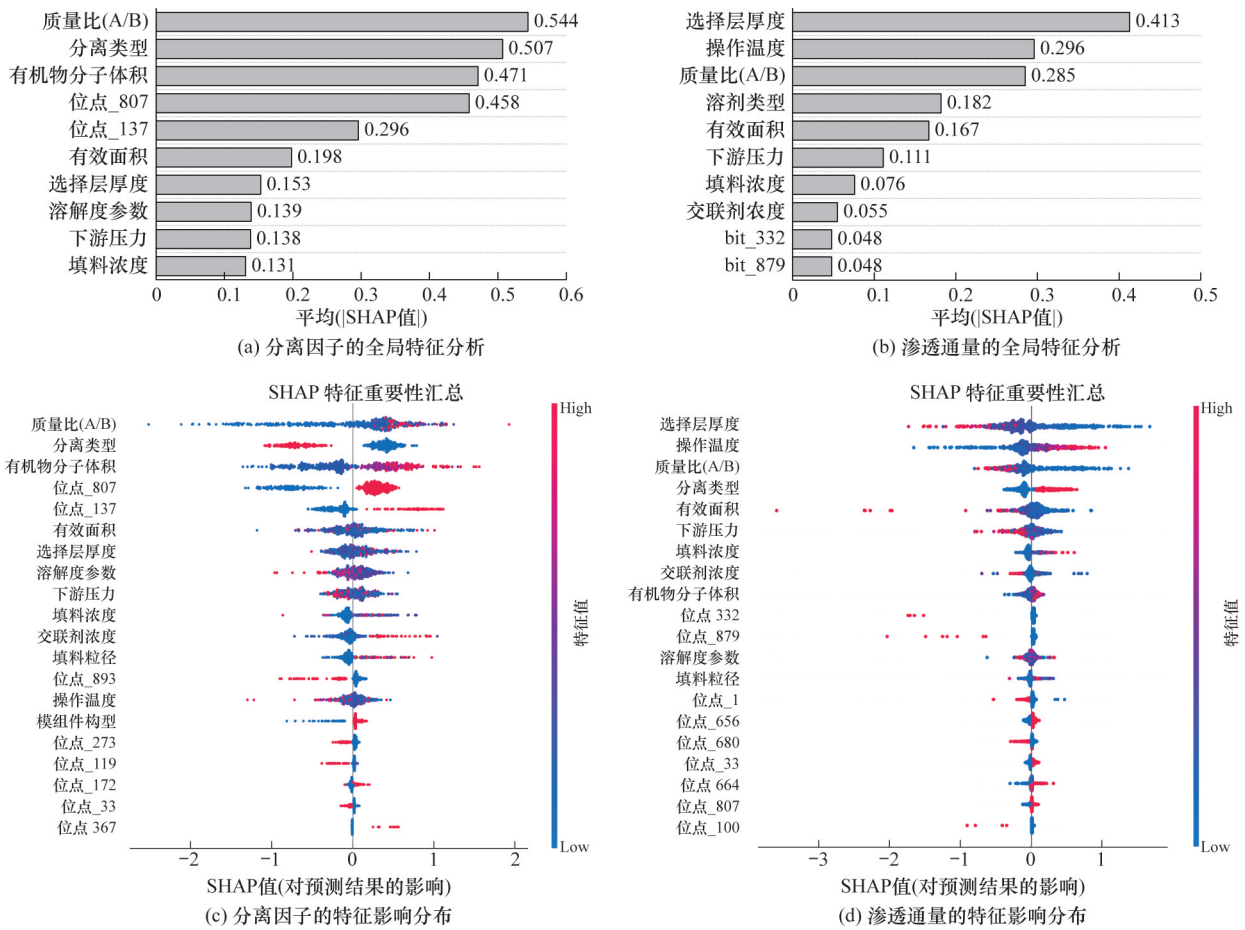


图7 基于SHAP方法的分离因子与渗透通量预测模型特征分析图

Fig.7 Analysis diagram of feature importance of separation factor and permeation flux prediction models based on SHAP method

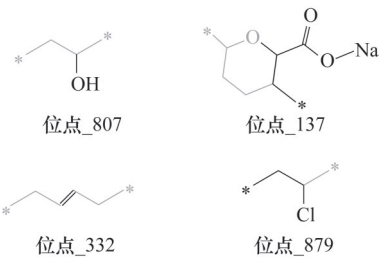


图8 基于LGBMR-BAT特征分析识别出的对分离性能有显著影响的聚合物子结构

Fig.8 Key polymer substructures on separation performance identified by LGBMR-BAT feature analysis

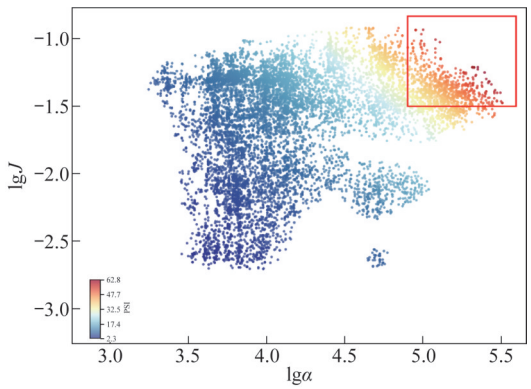


图9 基于LGBMR-BAT下不同聚合物的分离性能预测结果

Fig.9 Prediction results of separation performance for different polymers based on LGBMR-BAT

其他材料，甚至逼近或突破传统通量-选择性上限。该区域颜色偏红的聚合物普遍具有较高的PSI值，这表明高溶解性可能是实现高分离性能的关键因素之一。

图10是使用LGBMR-BAT模型在乙酸乙酯脱水中识别出的PSI最高的前10种聚合物分子结构图，按性能从高到低排列。从图中可以看出，高性能聚合物如1~4号，普遍具有两个关键结构特征：含有多苯环刚性骨架，并引入了醚键或磺酸基等适度的极性官能团。这类结构赋予主链高刚性，抑制链段过度堆积，同时能在保持一定自由体积的基础上调控膜的亲水性。其中，磺酸基的强极性可与水分子形成偶极-偶极相互作用，有利于水的选择性渗透，而苯环的疏水性则有助于排斥乙酸乙酯，从而在脱水过程中实现高选择性与适中通量的协同提升。

相比之下，5~7号聚合物虽保留了多苯环结构，但缺乏强极性基团，主链主要依赖碳氢键（非极性键），导致其对水的亲和力不足，在乙酸乙酯/水混合物中难以有效富集水分子，因而分离因子下降。而8~10号聚合物则表现出明显的结构劣势：其主链中含有长柔性烷基链或脂环结构。这些柔性单元增加了链段的运动性，易导致自由体积分布不均，且膜材料更易被乙酸乙酯溶胀。

综上，在乙酸乙酯脱水体系中，理想的聚合物分离膜应同时具备刚性多芳环骨架与适度极性官能团。这一结论已得到文献支持。不同聚合物膜在乙酸乙酯脱水性能的文献数据比较结果如表5所示，具备此特征的苯并咪唑连接聚合物（BILP）、聚苯并咪唑（PBI）与聚醚酰亚胺（PEI）等聚合物，其PSI均大于2000.0，其优异的性能源于刚性骨架的分子筛分效应与适度极性官能团和水分子的亲和力，且后者未引发严重的膜溶胀。相比之下，PVA、壳聚糖（CS）等传统亲水膜虽依赖强极性基团实现选择性分离，但其柔性主链在溶剂中易溶胀，破坏了筛分结构，导致高通量与高选择性难以兼顾，PSI值因而较低。

3 结 论

本文构建了一个基于LGBMR与BAT相结合的机器学习框架，用于高效预测与筛选适用于乙酸乙酯脱水的高性能聚合物PV膜材料。研究结果表明，所提出的LGBMR-BAT模型在预测精度、模型稳定性和计算效率方面均优于其他对比模型，展现出优异的综合性能。

SHAP可解释性分析表明，乙酸乙酯脱水膜的

表5 不同聚合物膜在乙酸乙酯脱水的PV性能文献数据比较

Table 5 Comparison of literature data on PV performance of different polymer membranes for ethyl acetate dehydration

聚合物名称	结构特征	渗透通量/(kg·m^{-2}·h^{-1})	分离因子	PSI	文献
BILP	刚性芳环骨架 + 含氮极性基团	0.3600	13800.00	4968.0	[35]
PBI/PEI	刚性杂环骨架+极性侧链	0.8200	2478.00	2031.9	[36]
PVA/Tac	柔性聚乙烯醇主链	0.2000	5000.00	1000.0	[37]
PVA/Mac	柔性聚乙烯醇主链	0.0458	4241.00	194.2	[38]
CS/PVA	柔性糖环骨架	2.2200	500.00	1110.0	[39]
CS/PVP	柔性糖环骨架	0.9530	746.00	710.9	[40]

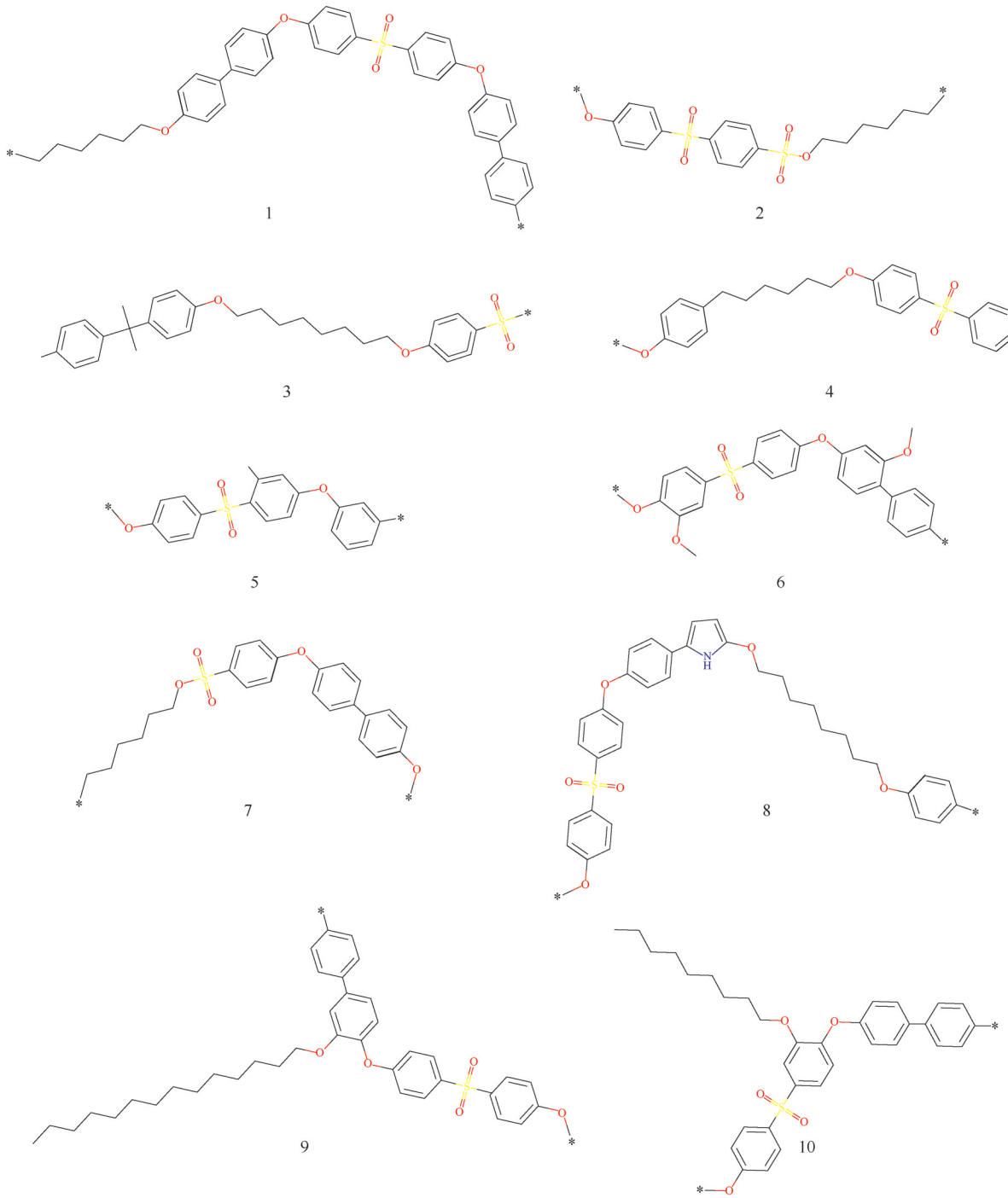


图10 基于PSI指标筛选的乙酸乙酯脱水膜候选材料分子结构

Fig.10 Molecular structure of candidate materials for ethyl acetate dehydration membrane screened based on PSI index

分离因子主要受化学组成调控，其中质量比(A/B)、溶剂极性及分子尺寸筛分效应起关键作用；而亲水位点_807和离子化位点_137能显著提升膜对水的选择性。渗透通量则主要由膜结构及操作条件决定，减薄选择层、提高温度可有效增强渗透通量，位点_332和位点_879通过增大自由体积或链间距来促进水分子的传输。

基于LGBMR-BAT模型，对初筛后的21830个虚拟聚合物进行了高通量筛选，成功识别出一批突破传统性能权衡的高性能乙酸乙酯脱水膜候选材料。其中PSI排名前10的聚合物普遍具备刚性多芳环骨架与适度极性官能团。该方法实现了对海量虚拟聚合物的快速初筛，并定位出若干极具潜力的新型脱水膜材料，为后续实验验证与分子设计提供

了明确方向。尽管本文构建的LGBMR-BAT框架在乙酸乙酯脱水体系中表现优异，但要将其推广至通用膜材料智能设计平台，仍需克服以下挑战：不同体系的高质量实验数据往往规模有限且分布不均，直接迁移训练容易导致偏差。本文所用的数据集中，其他体系样本量相对不足，需要补充不同溶剂-聚合物组合的渗透性、选择性等数据，并建立跨体系的数据标准化方法和迁移学习策略。不同溶剂与聚合物之间的相互作用（如极性、氢键、自由体积分布等）存在显著差异，导致膜分离性能受体系特异性规律主导，仅依赖通用分子描述符难以充分刻画此类复杂的物理化学耦合效应，可能限制模型在新体系中的泛化能力，因此，有必要建立面向特定分离任务的多组分特征工程方法。此外，高性能膜候选材料的合成与测试周期长，可结合主动学习策略，优先对预测性能优良的候选材料进行实验验证，从而以最少的实验轮次实现模型的快速迭代优化，有效降低研发成本。

综上所述，本文为高性能PV膜材料的开发提供了一种数据驱动的理性设计新路径。未来的研究可将该机器学习框架扩展至其他膜材料体系与应用场景，进一步验证并提升其普适性与推广价值。

参考文献

[1] Chapman P D, Oliveira T, Livingston A G, et al. Membranes for the dehydration of solvents by pervaporation[J]. Journal of Membrane Science, 2008, **318**(1/2): 5-37.

[2] Liu S, Zhou G Y, Cheng G B, et al. Emerging membranes for separation of organic solvent mixtures by pervaporation or vapor permeation[J]. Separation and Purification Technology, 2022, **299**: 121729.

[3] 陈子健, 钱锡亮, 宋宁, 等. 非溶剂诱导相分离法制备纤维素复合膜及乙酸乙酯渗透气化脱水性能[J]. 化工进展, 2025, **44**(11): 6514-6523.
Chen Z J, Qian X L, Song N, et al. Fabrication of cellulose composite membranes *via* non-solvent induced phase separation and their pervaporation dehydration performance for ethyl acetate [J]. Chemical Industry and Engineering Progress, 2025, **44**(11): 6514-6523.

[4] Zhong S F, Zhang K, Bagheri M, et al. Machine learning: new ideas and tools in environmental science and engineering[J]. Environmental Science & Technology, 2021, **55**(19): 12741-12754.

[5] Liu G P, Jin W Q. Pervaporation membrane materials: recent trends and perspectives[J]. Journal of Membrane Science, 2021, **636**: 119557.

[6] Audus D J, de Pablo J J. Polymer informatics: opportunities and challenges[J]. ACS Macro Letters, 2017, **6**(10): 1078-1082.

[7] Chen L H, Pilania G, Batra R, et al. Polymer informatics: current status and critical next steps[J]. Materials Science and Engineering: R: Reports, 2021, **144**: 100595.

[8] Yang M Q, Zhu J J, McGaughey A L, et al. Machine learning for polymer design to enhance pervaporation-based organic recovery [J]. Environmental Science & Technology, 2024, **58**(23): 10128-10139.

[9] Gao H P, Zhong S F, Zhang W L, et al. Revolutionizing membrane design using machine learning-Bayesian optimization[J]. Environmental Science & Technology, 2022, **56**(4): 2572-2581.

[10] Alizamir M, Kim S, Zounemat-Kermani M, et al. Modelling daily soil temperature by hydro-meteorological data at different depths using a novel data-intelligence model: deep echo state network model[J]. Artificial Intelligence Review, 2021, **54**(4): 2863-2890.

[11] Li H, Wang Y K, Wang Y Q. Machine learning for predicting the dynamic extraction of multiple substances by emulsion liquid membranes[J]. Separation and Purification Technology, 2023, **313**: 123458.

[12] Axelrod S, Schwalbe-Koda D, Mohapatra S, et al. Learning matter: materials design with machine learning and atomistic simulations[J]. Accounts of Materials Research, 2022, **3**(3): 343-357.

[13] Xu X Y, Zhao W L, Hu Y X, et al. Discovery of thermosetting polymers with low hygroscopicity, low thermal expansivity, and high modulus by machine learning[J]. Journal of Materials Chemistry A, 2023, **11**(24): 12918-12927.

[14] Japa L, Serqueira M, MendonçA I, et al. A population-based hybrid approach for hyperparameter optimization of neural networks[J]. IEEE Access, 2023, **11**: 50752-50768.

[15] Zhang X W, Chen Z G, Jiao F. Application of the combination method based on RF and LE in near infrared spectral modeling[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2023, **289**: 122247.

[16] Alizamir M, Keshavarz A, Abdollahi F, et al. Accurately predicting the performance of MOF-based mixed matrix membranes for CO_2 removal using a novel optimized extreme learning machine by BAT algorithm[J]. Separation and Purification Technology, 2023, **325**: 124689.

[17] Qiu B Y, Wang Y Y, Fan S Q, et al. Ethanol mass transfer during pervaporation with PDMS membrane based on solution-diffusion model considering concentration polarization[J]. Separation and Purification Technology, 2019, **220**: 276-282.

[18] Sang C, Liu C, Ying Y P, et al. Recent progress in hydrophobic pervaporation membranes for phenol recovery[J]. Green Energy & Environment, 2025, **10**(9): 1828-1837.

[19] da Silva D A R O, Zuge L C B, de Paula Scheer A. Pretreatments for seawater desalination by pervaporation using the developed green silica/PVA membrane[J]. Journal of Environmental Chemical Engineering, 2021, **9**(6): 106327.

[20] Maleki R, Shams S M, Chellehbari Y M, et al. Materials discovery of ion-selective membranes using artificial intelligence[J]. Communications Chemistry, 2022, **5**: 132.

[21] Chai Z H, Yang X, Liu Z L, et al. Correlation analysis-based neural network self-organizing genetic evolutionary algorithm[J]. IEEE Access, 2019, **7**: 135099-135117.

[22] Yan C H, Liang J, Zhao M J, et al. A novel hybrid feature selection strategy in quantitative analysis of laser-induced

breakdown spectroscopy[J]. Analytica Chimica Acta, 2019, **1080**: 35-42.

[23] Yang X S, Hossein Gandomi A. Bat algorithm: a novel approach for global engineering optimization[J]. Engineering Computations, 2012, **29**(5): 464-483.

[24] Wang Z H, Zhou T, Sundmacher K. Interpretable machine learning for accelerating the discovery of metal-organic frameworks for ethane/ethylene separation[J]. Chemical Engineering Journal, 2022, **444**: 136651.

[25] 张宇, 钱锡亮, 甘嘉添, 等. 基于仿生矿化的PVA/SA/CPO混合基质膜制备与酯化体系脱水性能[J]. 化工学报, 2026, **77**(3): 1339-1351.
Zhang Y, Qian X L, Gan J T, et al. Fabrication of PVA/SA/CPO mixed matrix membranes *via* biomimetic mineralization for dehydration of esterification systems[J]. CIESC Journal, 2026, **77**(3): 1339-1351.

[26] 肖武, 李明月, 阮雪华, 等. 响应面法优化一水硫酸氢钠流化催化精馏生产乙酸乙酯工艺条件[J]. 化工学报, 2014, **65**(11): 4465-4471.
Xiao W, Li M Y, Ruan X H, et al. Optimization of ethyl acetate process conditions for sodium bisulfate fluidized catalytic distillation using response surface methodology[J]. CIESC Journal, 2014, **65**(11): 4465-4471.

[27] Ma R M, Luo T F. PI1M: a benchmark database for polymer informatics[J]. Journal of Chemical Information and Modeling, 2020, **60**(10): 4684-4690.

[28] Zhang B J, Vogt M, Maggiora G M, et al. Design of chemical space networks using a Tanimoto similarity variant based upon maximum common substructures[J]. Journal of Computer-Aided Molecular Design, 2015, **29**(10): 937-950.

[29] Park H B, Kamcev J, Robeson L M, et al. Maximizing the right stuff: the trade-off between membrane permeability and selectivity [J]. Science, 2017, **356**(6343): eaab0530.

[30] Sanders D F, Smith Z P, Guo R L, et al. Energy-efficient polymeric gas separation membranes for a sustainable future: a review[J]. Polymer, 2013, **54**(18): 4729-4761.

[31] 刘燕青, 胡听听, 鲁落义, 等. PDMS/ZSM-5膜的制备及渗透汽化分离水中乙酸正丁酯和乙酸乙酯[J]. 化工学报, 2020, **71**(2): 843-853.
Liu Y Q, Hu T T, Lu L Y, et al. Preparation of PDMS/ZSM-5 membranes and pervaporation separation of *n*-butyl acetate and ethyl acetate from aqueous media[J]. CIESC Journal, 2020, **71**(2): 843-853.

[32] Lipnizki F, Field R W, Ten P K. Pervaporation-based hybrid process: a review of process design, applications and economics [J]. Journal of Membrane Science, 1999, **153**(2): 183-210.

[33] Kueh T C, Hung Y M. Enhanced film-wise water evaporation through graphene nanostructures: a molecular dynamics insight[J]. International Journal of Heat and Mass Transfer, 2024, **233**: 126010.

[34] Esposito R, Abdulhamid M A, Upadhyaya L, et al. Carboxyl-functionalized polyimide for polar/non-polar organic solvent separation by pervaporation[J]. Journal of Membrane Science, 2025, **713**: 123277.

[35] Wang X Y, Cong S Z, Yan X R, et al. Crosslinked benzimidazole-linked polymer membranes for dehydration of organics[J]. Separation and Purification Technology, 2023, **316**: 123802.

[36] Wang Y. Pervaporation dehydration of ethyl acetate *via* PBI/PEI hollow fiber membranes[J]. Industrial & Engineering Chemistry Research, 2015, **54**(11): 3082-3089.

[37] Salt Y, Hasanoğlu A, Salt İ, et al. Pervaporation separation of ethylacetate-water mixtures through a crosslinked poly (vinylalcohol) membrane[J]. Vacuum, 2005, **79**(3/4): 215-220.

[38] Yuan H K, Ren J, Ma X H, et al. Dehydration of ethyl acetate aqueous solution by pervaporation using PVA/PAN hollow fiber composite membrane[J]. Desalination, 2011, **280**(1/2/3): 252-258.

[39] Zhu Y X, Xia S S, Liu G P, et al. Preparation of ceramic-supported poly(vinyl alcohol)-chitosan composite membranes and their applications in pervaporation dehydration of organic/water mixtures[J]. Journal of Membrane Science, 2010, **349**(1/2): 341-348.

[40] Zhang X H, Liu Q L, Xiong Y, et al. Pervaporation dehydration of ethyl acetate/ethanol/water azeotrope using chitosan/poly(vinyl pyrrolidone) blend membranes[J]. Journal of Membrane Science, 2009, **327**(1/2): 274-280.