

文章编号: 1006-3080(2025)02-0158-08

DOI: 10.14135/j.cnki.1006-3080.20240417001

基于多元线性回归和反向传播人工神经网络 预测离子液体的声速

季常征, 万 仁, 时兆翀, 彭昌军, 刘洪来
(华东理工大学化学与分子工程学院, 上海 200237)

摘要: 离子液体的声速可采用实验测定、半经验模型和理论研究方法获得, 其中, 定量结构-性质关系 (QSPR) 模型已受到广泛关注, 但构造一个有效的 QSPR 模型取决于选择合适的分子描述符。本文采用片段活度系数类导体屏蔽模型 (COSMO-SAC) 获得离子液体电荷密度分布片段面积 (S_o) 和空穴体积 (V_{COSMO}) 两个描述符, 并分别采用多元线性回归 (MLR) 和反向传播人工神经网络 (BP-ANN) 构建了用于描述离子液体声速的线性 QSPR 模型 u -MLR 和非线性 QSPR 模型 u -ANN, 模型中包含了温度和离子液体相对分子量, 所涉及的数据集包括 171 种离子液体的 5 114 个数据点。在总的离子液体声速数据集中, u -MLR 和 u -ANN 的决定系数 (R^2) 分别为 0.970 6 和 0.999 5, 平均绝对相对偏差 (AARD) 分别为 1.59% 和 0.10%, 均方根误差 (RMSE) 分别为 30.68 m/s 和 4.12 m/s。结果表明, 基于人工神经网络建立的 u -ANN 模型的预测效果明显优于基于线性回归方法建立的 u -MLR 模型的预测效果。

关键词: 声速; 离子液体; 人工神经网络; 多元线性回归; 定量结构-性质关系

中图分类号: TQ021.4

文献标志码: A

声速是指声波在流体中的传播速度, 它可通过实验直接测定, 也可采用半经验模型和理论模型获得^[1-2]。近些年来, 人们开始关注离子液体 (ILs) 的声速, 其重要意义在于基于声速可确定离子液体的其他热物理性质, 如等压热膨胀系数、等熵和等温压缩系数、热容以及焦耳-汤姆逊系数等^[3-6]。考虑到实验不可能穷尽所有离子液体, 这就需要构建半经验模型和理论模型, 理论上可结合流体的状态方程 (EoS)^[7-11], 也可采用分子动力学 (MD) 模拟进行^[12]。但前者需要大量实验数据拟合模型参数, 后者需要可靠的分子力场且耗时较长。在此背景下, 构建声速的半经验模型受到极大关注。半经验模型可依据声速与其他热物理的关系来建立^[4], 但普遍采用的还是基团贡献法 (Group Contribution Method, GCM) 和定量结构-性质关系 (Quantitative Structure-Property Relationship, QSPR) 模型。

GCM 的核心是将离子液体划分为不同基团, 离子液体的声速则由相应基团的贡献值获得。Sattari 等^[13]采用最小二乘支持向量机 (LS-SVM) 算法, 开发了离子液体声速的非线性基团贡献模型, 模型的决定系数 (R^2) 为 0.997, 平均绝对相对偏差 (AARD) 为 0.36%, 但模型只是在 41 种离子液体的 446 个数据点基础上建立的。Wu 等^[14]针对 96 种离子液体的 4478 个声速数据, 建立了一个二阶对应状态基团贡献模型 (CSGC) 用来预测离子液体的声速, 其 AARD 为 2.34%。GCM 简单明了, 但涉及到新的结构基团时, 还需要有可靠且丰富的实验数据以获得 GCM 中的基团贡献值。此外, Haghbakhsh 等^[15]利用 140 种离子液体的 3 790 个声速数据, 提出了一种离子液体声速的原子贡献模型 (ACM)。ACM 不用统计基团的种类和数量, 而是直接依据离子液体化学式中原子的贡献来推测声速, 其 AARD 为 3.33%。

收稿日期: 2024-04-17

基金项目: 国家自然科学基金 (22078086, 22378111)

作者简介: 季常征 (2000—), 男, 山东青岛人, 硕士生, 主要研究方向为离子液体性质模拟预测。E-mail: Y30220327@mail.ecust.edu.cn

引用本文: 季常征, 万 仁, 时兆翀, 等. 基于多元线性回归和反向传播人工神经网络预测离子液体的声速 [J]. 华东理工大学学报(自然科学版), 2025, 51(2): 158-165.

Citation: JI Changzheng, WAN Ren, SHI Zhaochong, et al. Prediction of Sound Speed of Ionic Liquids Based on Multiple Linear Regression and Back-Propagation Artificial Neural Networks[J]. Journal of East China University of Science and Technology, 2025, 51(2): 158-165.

构建 QSPR 模型的关键是选用一套合适的分子描述符来体现离子液体的结构特征, 并建立声速与描述符之间的线性或非线性关系。Sattari 等^[16]选择离子液体的电子衍射描述符、原子极化率、二阶平均拓扑电荷指数等 9 个描述符, 使用遗传函数逼近 (GFA) 算法建立了常压下离子液体声速的线性 QSPR 模型。模型的 R^2 约为 0.99, AARD 为 0.92%, 但模型的数据集仅囊括了 41 种离子液体的 446 个数据点。Pilarz 等^[17]针对 31 种离子液体的 2257 个实验数据, 以温度、压力、阴/阳离子摩尔质量和异构参数为输入变量, 采用前馈神经网络 (FNN) 算法构建了离子液体声速 QSPR 模型, 模型的 R^2 为 0.989, AARD 为 0.81%。最近, Mohan 等^[18]利用真实溶剂类导体屏蔽模型 (COSMO-RS)^[19]获得描述符, 采用多元线性回归 (MLR) 以及梯度增强树 (GBT) 算法分别建立了离子液体声速的线性与非线性模型, MLR 的 R^2 和 AARD 分别为 0.48 和 13.71%, GBT 的 R^2 和 AARD 分别为 0.99 和 2.07%。可见基于神经网络机器学习算法构建的 QSPR 模型具有更好的发展前景。

我们曾采用片段活度系数类导体屏蔽模型 (COSMO-SAC)(2007)^[20]获得电荷密度分布以及空穴体积, 以特定间隔的电荷密度分布面积 (S_{oi}) 和空穴体积 (V_{COSMO}) 为描述符, 采用 MLR 方法构建了能预测离子液体热导率^[21]、自扩散系数^[22]、电导率^[23]以及无限稀释摩尔电导率^[24]的线性 QSPR 模型。进一步地, 结合反向传播神经网络 (BP-ANN) 算法构建了能预测离子液体自扩散系数^[25]以及电导率^[23]的非线性 QSPR 模型, 发现引入的描述符能令人满意地区分离子液体的种类, 特别是引入 V_{COSMO} 这个描述符后能体现离子液体尺寸对性质的影响。本文沿用这个研究思路, 利用 COSMO-SAC(2007) 模型^[20]获得的片段面积描述符和空穴体积描述符, 分别采用 MLR 和 BP-ANN 算法建立了离子液体声速的线性 QSPR 模型 u -MLR 和非线性 QSPR 模型 u -ANN, 并对模型的预测效果进行了评估。

1 数据集和方法

1.1 数据集

离子液体声速数据均从公开发表的文献中获得。考虑到绝大部分数据都是基于常压下测定的, 数据集中剔除了高压数据, 最后整理的数据集数量 (N) 为 5114 个, 它包含了 171 种离子液体, 涉及 85 种阳离子和 36 种阴离子。数据集的温度范围为 278.15~413.15 K, 声速范围为 1009.68~1935.58 m/s。

1.2 描述符

构建 QSPR 模型的关键是选择描述符。基于先前的工作^[21-25], 离子液体的结构描述符主要基于 COSMO-SAC(2007)^[20]产生, 具体步骤如下: 通过 Materials Studio(V7.0) 中的 DMol3 模块并采用 GGA/VWN-BP 密度泛函和 DNP* 基组^[26-27]计算生成 COSMO 文件。COSMO 文件中包含了电荷密度分布 (σ -profile) 和 V_{COSMO} 等信息, 它们也是 COSMO 基模型预测离子液体性质的关键^[28-29]。一般离子的电荷密度分布主要集中在 $-3.0 \sim 3.5 \text{ e/nm}^2$ 之间。为简便, 本文对阳离子的 σ -profile 在 $-3.0 \sim 2.0 \text{ e/nm}^2$ 之间以 0.5 e/nm^2 为间隔划分成 10 个区域; 对于阴离子, 则在 $-2.0 \sim 3.5 \text{ e/nm}^2$ 之间以 0.5 e/nm^2 为间隔划分成 11 个区域, 然后对每个区域积分, 并将积分面积 S_{oi} 作为描述符。其中, 阳离子的 S_{oi} 分别标记为 $S_{o1}, S_{o2}, \dots, S_{o10}$, 阴离子的 S_{oi} 分别标记为 $S_{o11}, S_{o12}, \dots, S_{o21}$ 。如此, 片段面积描述符共有 21 个。应该指出的是, 对 σ -profile 可以等间隔划分^[21-25], 也可以非等间隔划分^[18]。

考虑到声速与离子液体的尺寸^[3]、相对分子量 (M_w)^[18,30]以及温度 (T)^[4,18] 有关, 声速的 QSPR 模型中也将温度、离子液体的相对分子量和 V_{COSMO} (阳离子与阴离子空穴体积的加和) 作为描述符参与到模型的构建中。

1.3 u -MLR 模型

常压下, 离子液体的声速 (u) 一般可用以下线性方程表示:

$$u = u_0 + \frac{A}{M_w} + B \times T \quad (1)$$

其中, u_0 为线性方程的截距; A 和 B 为常数, 其值与离子液体的结构和分子间相互作用有关, 可由 S_{oi} 和 V_{COSMO} 共 22 个描述符线性加和得到, 具体为:

$$A = a_0 + \sum_{i=1}^{21} a_i S_{oi} + a_{22} V_{\text{COSMO}} \quad (2)$$

$$B = b_0 + \sum_{i=1}^{21} b_i S_{oi} + b_{22} V_{\text{COSMO}} \quad (3)$$

式(1)~(3)即为基于多元线性回归构建的 u -MLR。其中, a 和 b 为模型参数, 可采用实验数据回归得到。

1.4 u -ANN 模型

人工神经网络 (ANN) 是一种灵感来源于生物神经系统的机器学习算法^[31]。本文采用的 BP-ANN 是一种常见的神经网络, 它由输入量、隐藏层、输出层和输出量组成。在 BP-ANN 拓扑结构中, 每一层神经网络都由被称为节点的简单处理单元组成。节点

处理数据的过程如下^[32]:

$$y = \varphi_2 [w_2 \varphi_1 (w_1 x + b_1) + b_2] \quad (4)$$

式中, x 代表输入样本, y 是输出层的结果, φ_1 和 φ_2 分别代表隐藏层和输出层的激活函数, w_1 和 w_2 分别代表隐藏层和输出层的权值, b_1 和 b_2 分别代表隐藏层和输出层的阈值。

通常来说, 前一层的输入数据使用特定的权重和偏差以及激活函数进行处理后传递到下一层, 并作为最终输出或者后续层的输入。在整个传递过程中, 信号传递通过“输入-隐藏层-输出层”前向传播, 误差通过“输出层-隐藏层-输入”反向传播, 以优化权重和偏差, 如此循环往复使计算值和目标值间的误差朝着减小的方向进行。本文使用 Matlab R2020b 中的深度学习工具完成。隐藏层使用双曲正切激活函数 (tansig), 输出层使用线性激活函数 (purelin)。以计算值和目标值的均方误差 (MSE) 最小为目标确定网络的权重和偏差。在优化和训练神经网络模型时, Matlab 默认将 MSE 用作损失函数。

BP-ANN 模型中的输入量为 T 、 M_w 、 V_{COSMO} 以及 S_{oi} , 输出量为 u 。基于声速的数据集, 则可建立网络模型, 即:

$$u = f(S_{\sigma 1}, S_{\sigma 2}, \dots, S_{\sigma 21}, V_{\text{COSMO}}, T, M_w) \quad (5)$$

为寻找最优模型, 将整理后的数据集随机划分为训练集、验证集和测试集。10% 的数据点 (511 个数据点) 为测试集, 它们并不参与模型的建立, 仅用于评估模型的预测能力 (外部验证); 80% 的数据点 (4091 个) 作为训练集用于拟合模型参数; 10% 的数据点 (512 个) 作为验证集用于评估模型的稳定性 (内部验证), 内部验证采用 9 折交叉验证。实际训练前, 利用“mapminmax”函数将输入样本和目标 y 值归一化到 [0,1]。最高失败次数设置为 6 次, 即当验证集的 MSE 计算 6 次仍不下降时, 神经网络不会再将误差反传, 训练停止模型输出。

运行神经网络首先要确定隐藏层节点数, 本文中最优隐藏层节点数的确定参照 Paduszyński 等^[33] 的方法。训练过程中以验证集最低的 MSE 平均值所对应的节点作为最佳节点 (S_{opt}), 最优模型为最佳节点下验证集最低的 MSE 对应折数的模型。

1.5 模型评价

R^2 、AARD、相对偏差 (RD)、均方根误差 (RMSE) 和 MSE 被用来评估本文所建模型, 计算公式如下:

$$R^2 = 1 - \sum_{i=1}^N (u_{i,\text{cal}} - u_{i,\text{exp}})^2 / \sum_{i=1}^N (\bar{u}_{\text{cal}} - u_{i,\text{exp}})^2 \quad (6)$$

$$\text{AARD} = \left[(1/N) \sum_{i=1}^N |(u_{i,\text{cal}} - u_{i,\text{exp}}) / u_{i,\text{exp}}| \right] \times 100\% \quad (7)$$

$$\text{RD} = (u_{i,\text{cal}} - u_{i,\text{exp}}) / u_{i,\text{exp}} \times 100\% \quad (8)$$

$$\text{RMSE} = \left[(1/N) \sum_{i=1}^N (u_{i,\text{cal}} - u_{i,\text{exp}})^2 \right]^{1/2} \quad (9)$$

$$\text{MSE} = (1/N) \sum_{i=1}^N (u_{i,\text{cal}} - u_{i,\text{exp}})^2 \quad (10)$$

式中, 下标“exp”和“cal”分别代表声速的实验值和模型计算值, $\bar{u}_{i,\text{cal}}$ 表示 $u_{i,\text{cal}}$ 平均值。

2 结果与讨论

2.1 u -MLR 最优模型确定与性能评估

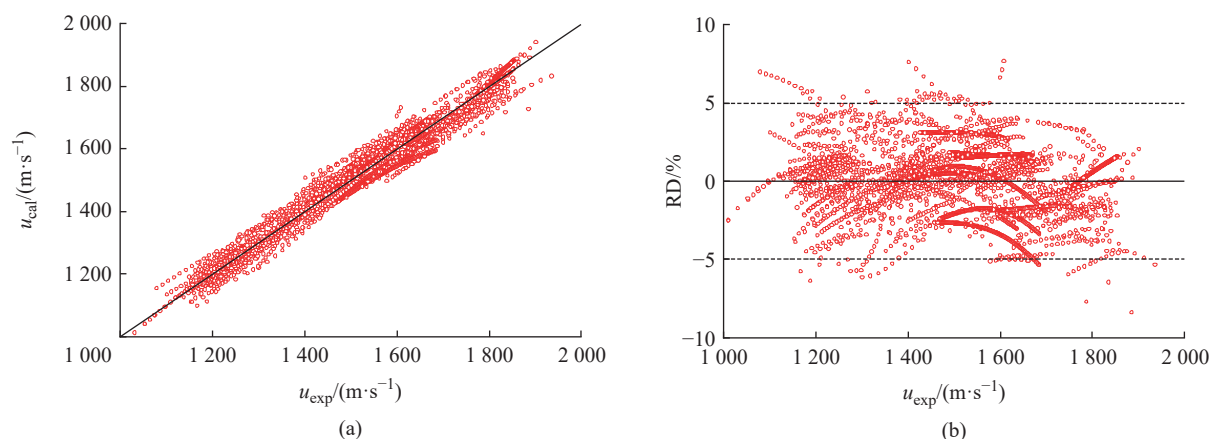
为了验证 u -MLR 模型的可行性, 首先将声速数据集 (5114 个数据点) 随机划分为训练集 (4114 个数据点) 和测试集 (1000 个数据点)。训练集用来检验模型的相关性, 通过多元线性回归得到模型参数 a_i 、 b_i 和 u_0 。测试集则用来检验模型的预测性能。拟合发现, 训练集的 R^2 、AARD 和 RMSE 分别为 0.9708、1.57% 和 30.54 m/s, 说明 u -MLR 模型的计算值与实验值的整体偏差控制在较低水平。进一步使用训练集获得的 u -MLR 定量模型对测试集数据进行预测, 发现测试集的 R^2 为 0.9696, AARD 为 1.63%, RMSE 为 31.42 m/s, 证明建立的模型对测试集的声速数据预测非常精准。

基于上述分析, 为提高 u -MLR 模型参数的可靠性, 增大样本容量, 对全部数据点 (合集) 进行了整体建模。结果显示, R^2 为 0.9706, AARD 为 1.59%, RMSE 为 30.68 m/s, 说明模型具有极佳的关联和预测效果。

合集的计算值与实验值对比如图 1 所示, 图中离散点紧密分布在等值线附近, 说明高达 5114 个声速数据点的计算值与实验值非常接近。其计算值与实验值的 RD 分布情况如图 1(b) 所示, 可以看出几乎所有离散点的相对偏差均在 $\pm 5\%$ 之间, 证明模型对声速的计算偏差可控制在较低程度。训练集、测试集和合集的统计参数汇总于表 1 中, 3 个集合的统计参数几乎一致, 证实模型能准确预测离子液体的声速, 且具有稳定性。

2.2 u -ANN 最优模型确定与性能评估

在神经网络模型中, 隐藏层节点数 S 直接关系到模型的最终表现形式, 依据的 Klimasauskas^[32] 方法如式 (11) 所示:

图1 u -MLR 模型的计算值与实验值的比较Fig. 1 Comparisons between the calculated and experimental values of the u -MLR model表1 u -MLR 模型性能参数Table 1 Performance parameters of the u -MLR model

Data set	N	R^2	AARD/%	RMSE/($\text{m}\cdot\text{s}^{-1}$)	RD _{max} /%
Training set	4 114	0.970 8	1.57	30.54	8.19
Testing set	1 000	0.969 6	1.63	31.42	7.70
Total	5 114	0.970 6	1.59	30.68	8.73

$$r = \frac{N_{\text{train}}}{S(D+2)+1} \quad (11)$$

式中, r 为训练集数据量与神经网络权值个数的比值, $r \geq 5$; N_{train} 为训练集数据点个数; S 为隐藏层节点数; D 为输入变量数。将 $r_{\text{min}}=5$ 、 $N_{\text{train}}=4091$ 和 $D=24$ 代入式(11), 计算得到的最大隐藏层节点数 $S_{\text{max}}=32$ 。

图2示出了训练集、验证集和测试集中误差(MSE平均值)随节点数 S 的变化。很明显, 当 S 大于30时, 验证集和测试集的误差均有增加的趋势, 因此, 根据上述 Paduszyński 等^[33] 的方法, 确定最优节点数(S_{opt})为30, 其值小于 S_{max} , 说明节点数选择是合理的。在 S_{opt} 为30的状态下, 进行9次交叉验证, 以验证9个神经网络模型的表现能力。结果发现, 每个折数下所有测试集的MSE都非常接近。其中, 第8次交叉验证时验证集的MSE最小, 对应的模型为最优模型。后续研究都基于此模型展开。

图3(a)示出了声速的计算值与实验值之间的对比。值得注意的是, 图中的训练集、验证集和测试集的结果几乎分布在对角线上, 表明 u -ANN 模型的计算值与实验值高度吻合。为了更进一步了解计算值与实验值之间的偏差, 运用式(8)计算了训练集、验证集和测试集中各个数据点的相对偏差, 具体结果如图3(b)所示。由图可见, 绝大部分离散点的相对偏差在 $\pm 1\%$ 之间, 占总数据点的98.8%, 最大偏差也

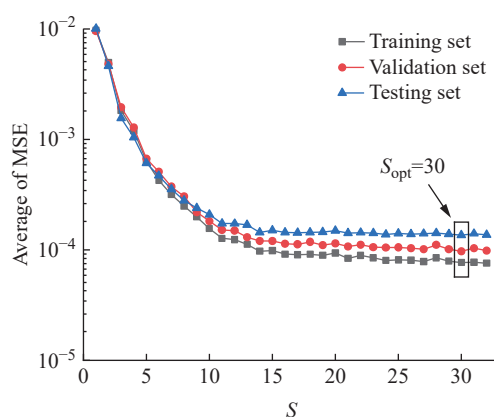


图2 隐藏层最佳节点数的选取

Fig. 2 Selection of the optimum number of nodes in hidden layer

未超出 $\pm 4\%$, 这也说明 u -ANN 模型能以极高的精度预测离子液体声速。

表2示出了 u -ANN 模型的相关统计参数, 数据集的 R^2 为 0.9995, RMSE 为 4.12 m/s, AARD 为 0.10%, u -ANN 模型表现优异。进一步研究发现, 训练集、验证集与测试集的 R^2 均接近于1, 而且验证集和测试集的 R^2 、AARD 和 RMSE 与训练集的对应值较为接近, 说明 u -ANN 模型不仅内部稳定, 而且在预测其他离子液体声速方面同样能表现出良好性能。此外, 训练集、验证集与测试集的 AARD 保持在 0.1% 左右, 最大相对偏差(RD_{max})也未超过 4%, 说明 u -ANN 模型具有极高的预测精度。

实际上, u -ANN 与 u -MLR 使用了相同的数据集和同样的描述符, 只不过前者采用了 BP-ANN, 后者采用了 MLR, 但二者的效果相差较大。对于合集而言, u -ANN 模型的 R^2 为 0.9995, AARD 为 0.10%, RMSE 为 4.12 m/s, RD_{max} 为 3.82%。可见, u -MLR 模型比 u -ANN 模型预测效果差。这或许是因为离子液体的声速与描述符间并不存在简单的线性关系, 需要采用

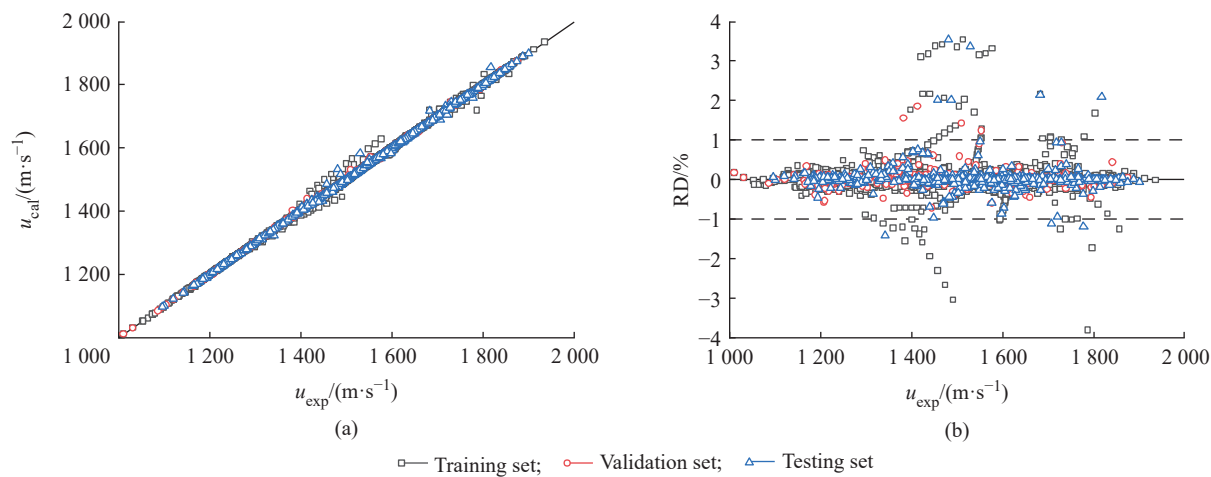


图 3 u -ANN 模型的计算值与实验值比较
Fig. 3 Comparisons between the calculated and experimental values of the u -ANN model

表 2 u -ANN 模型的性能参数

Table 2 Performance parameters of the u -ANN model					
Data set	N	R^2	RMSE/($\text{m}\cdot\text{s}^{-1}$)	AARD/%	RD _{max} /%
Training set	4 091	0.999 5	4.04	0.10	3.82
Validation set	512	0.999 7	3.01	0.09	1.86
Testing set	511	0.999 1	5.47	0.13	3.54
Total	5 114	0.999 5	4.12	0.10	3.82

神经网络算法来体现函数间复杂的非线性关系。MLR 模型的优点在于模型形式简单明了,缺点是模型的拟合精度很难有质的提升,除非重新设计更为合理的映射关系。ANN 模型的优点在于其具有极高的预测精度,缺点是需要大量底层数据支持才能保证模型的可靠性。若追求计算快捷和方便可选用

MLR 模型,若追求预测的准确性可选用 ANN 模型。

2.3 不同模型的比较

表 3 给出了采用不同方法计算离子液体声速的结果比较。实际上,很难直接分辨出哪个方法更具优势。一方面虽然预测精度是可比拟的,但建立模型所使用的数据量不同;另一方面,当数据量相近时,数据集中包含的离子液体种类也可能不同。可以发现,与 Mohan 等^[18]工作类似,本文包含的离子液体种类以及数据集也非常丰富。表 3 也给出了不同声速模型所采用的方法,如 SAFT 方程 (ePC-SAFT 和 SAFT)^[10-11]、基团贡献法 (GCM)^[13-14]和原子贡献法 (ACM)^[15],以及采用 QSPR 构建思路^[16-18]。从表 3 的结果可知,基于 SAFT 方程的 AARD 均较大^[10-11],采用 GCM 方法的 AARD 虽然较小,但涉及的离子液体种类较少^[13-14]。总之,无论是 Mohan 等^[18]的工作,还

表 3 本文模型与相关模型的比较

Table 3 Comparison of this work with related models						
Item	Method	ILs	N	T/K	R^2	AARD/%
Literature [10]	ePC-SAFT	52	1 570	273.15—413.10	—	>10
Literature [11]	PC-SAFT	11	—	273.00—413.00	—	35.4
Literature [13]	GCM	41	446	278.15—343.15	0.9970	0.36
Literature [14]	GCM	96	4 478	278.14—343.15	—	2.34
Literature [15]	ACM	140	3 790	278.15—343.15	—	3.33
Literature [16]	GFA-QSPR	41	446	278.15—343.15	0.9862	0.92
Literature [17]	FFAN-QSPR	31	2 257	278.15—413.10	0.9890	0.81
Literature [18]	MLR-QSPR	218	5 702	273.13—413.10	0.8700	3.63
Literature [18]	GBT-QSPR	218	5 702	273.13—413.10	0.9900	0.58
u -MLR	MLR-QSPR	171	5 114	278.15—413.15	0.9706	1.59
u -ANN	BPANN-QSPR	171	5 114	278.15—413.15	0.9995	0.10

是本文工作,都证实基于 ANN 建立的 QSPR 模型的预测效果要明显优于基于 MLR 的效果。

即使都采用了 MLR 方法,但由于描述符的差异以及线性方程形式的不同也会导致预测结果的不同。在 Mohan 等^[18]构建的线性方程中,他们对 σ -profile 采用了非等间隔的方式以获得片段面积描述符 S_{ai} ,并假定离子液体声速与温度、离子液体相对分子量、片段面积描述符以及离子液体空穴面积等线性相关。如此建立的方程形式虽然简单,但 R^2 仅为 0.87, AARD 可达 3.63%。导致这个结果的原因在于线性方程中无法体现温度对不同离子液体声速影响的差异。本文建立的线性方程 (1) 中,温度对不同离子液体声速的影响程度可通过不同离子液体的片段面积以及空穴体积来体现。线性方程 (1) 的形式稍复杂,但模型计算的 R^2 高达 0.9706, AARD 仅为 1.59%。这也再次说明,要提升 MLR 模型的拟合精度需要重新选择合适的描述并设计更为合理的映射关系。

总之,以 R^2 和 AARD 两个指标来评估模型,本文以及 Mohan 等^[18]建立的模型的预测效果均要优于 Sattari 等^[16]和 Pilarz 等^[17]利用较少离子液体种类和较少数据集所建立模型的预测效果。特别是本文建立的 u -ANN 模型,其 R^2 几乎接近于 1, AARD 也仅为 0.10%,证明 u -ANN 模型要稍优于已有的其他模型。

3 结 论

(1) 采用 COSMO-SAC 获得的电荷密度分布面积和空穴体积为描述符,同时考虑相对分子量以及温度的影响,并采用多元线性回归和 BP-ANN 构建了可用于预测离子液体声速的定量结构-性质关系的 u -MLR 和 u -ANN。

(2) 在总的数据集中, u -ANN 和 u -MLR 的 R^2 分别为 0.9995 和 0.9706, AARD 分别为 0.10% 和 1.59%, RMSE 分别为 4.12 m/s 和 30.68 m/s,说明基于神经网络建立的 u -ANN 模型的预测能力明显优于基于线性回归方法建立的 u -MLR 模型。

参考文献:

[1] 张锐, 胡宇军, 黑恩成, 等. 乙醇和环己烷混合物的某些热力学性质 [J]. 华东理工大学学报 (自然科学版), 1998, 24(2): 96-100.

[2] 栾文耕, 张伟伟, 董红, 等. 聚(二甲基-二乙基)硅氧烷共聚物与长链烷烃组成的二元体系的混合性质研究 [J]. 高

校化学工程学报, 2021, 35(3): 407-414.

- [3] DZIDA M, ZOREBSKI E, ZOREBSKI M, *et al.* Speed of sound and ultrasound absorption in ionic liquids[J]. *Chemical Reviews*, 2017, 117(5): 3883-3929.
- [4] GARDAS R L, COUTINHO J A P. Estimation of speed of sound of ionic liquids using surface tensions and densities: A volume based approach[J]. *Fluid Phase Equilibria*, 2008, 267(2): 188-192.
- [5] ESTRADA-ALEXANDERS A F, JUSTO D. New method for deriving accurate thermodynamic properties from speed-of-sound[J]. *The Journal of Chemical Thermodynamics*, 2004, 36(5): 419-429.
- [6] LUSTIG R. Direct molecular NVT simulation of the isobaric heat capacity, speed of sound and Joule-Thomson coefficient[J]. *Molecular Simulation*, 2011, 37(6): 457-465.
- [7] MAGHARI A, ZIAMAJIDI F, PASHAEI E. Thermophysical properties of alkyl-imidazolium based ionic liquids through the heterosegmented SAFT-BACK equation of state[J]. *Journal of Molecular Liquids*, 2014, 191: 59-67.
- [8] MAGHARI A, ZIAMAJIDI F. Prediction of thermodynamic properties of pure ionic liquids through extended SAFT-BACK equation of state[J]. *Fluid Phase Equilibria*, 2013, 356: 109-116.
- [9] POLISHUK I. Implementation of perturbed-chain statistical associating fluid theory (PC-SAFT), generalized (g) SAFT+cubic, and cubic-plus-association (CPA) for modeling thermophysical properties of selected 1-alkyl-3-methyl-imidazolium ionic liquids in a wide pressure range[J]. *The Journal of Physical Chemistry A*, 2013, 117(10): 2223-2232.
- [10] SUN Y H, SCHEMANN A, HELD C, *et al.* Modeling thermodynamic derivative properties and gas solubility of ionic liquids with ePC-SAFT[J]. *Industrial & Engineering Chemistry Research*, 2019, 58(19): 8401-8417.
- [11] MAJDI H S, RAHEEM A B F, ABDULLAH S J, *et al.* Prediction of speed of sound and specific heat capacity of ionic liquids using predictive SAFT-based equation of state[J]. *Chemical Engineering Science*, 2023, 265: 118246.
- [12] RIBEIRO M C C. High-frequency acoustic modes in an ionic liquid[J]. *The Journal of Chemical Physics*, 2013, 139(11): 114505.
- [13] SATTARI M, GHARAGHEIZI F, ILANI-KASHKOULI P, *et al.* Determination of the speed of sound in ionic liquids using a least squares support vector machine group contribution method[J]. *Fluid Phase Equilibria*, 2014, 367: 188-193.
- [14] WU K J, CHEN Q L, HE C H. Speed of sound of ionic liquids: Database, estimation, and its application for thermal conductivity prediction[J]. *AIChE Journal*, 2014, 60(3): 1120-1131.

- [15] HAGHBAKHS R, KESHTKARI S, RAEISSI S. Simple estimations of the speed of sound in ionic liquids, with and without any physical property data available[J]. *Fluid Phase Equilibria*, 2020, 503: 112291.
- [16] SATTARI M, GHARAGHEIZI F, ILANI-KASHKOULI P, *et al.* A chemical structure based model for the determination of speed of sound in ionic liquids[J]. *Journal of Molecular Liquids*, 2014, 196: 7-13.
- [17] PILARZ J, POLISHUK I, CHORAŻEWSKI M. Prediction of sound velocity for selected ionic liquids using a multilayer feed-forward neural network[J]. *Journal of Molecular Liquids*, 2022, 347: 118376.
- [18] MOHAN M, SMITH M D, DEMERDASH O N, *et al.* Quantum chemistry-driven machine learning approach for the prediction of the surface tension and speed of sound in ionic liquids[J]. *ACS Sustainable Chemistry & Engineering*, 2023, 11(20): 7809-7821.
- [19] KLAMT A, ECKERT F. COSMO-RS: A novel and efficient method for the a priori prediction of thermophysical data of liquids[J]. *Fluid Phase Equilibria*, 2000, 172(1): 43-72.
- [20] WANG S, SANDLER S I, CHEN C C. Refinement of COSMO-SAC and the applications[J]. *Industrial & Engineering Chemistry Research*, 2007, 46(22): 7275-7288.
- [21] WAN R, LI M Y, SONG F, *et al.* Predicting the thermal conductivity of ionic liquids using quantitative structure-property relationship[J]. *Industrial & Engineering Chemistry Research*, 2022, 61(32): 12032-12039.
- [22] XIAO Y J, SONG F, AN S H, *et al.* Quantitative structure-property relationship for predicting diffusion coefficient of ionic liquids[J]. *Journal of Molecular Liquids*, 2022, 349: 118476.
- [23] SHI Z C, SONG F, JI C Z, *et al.* QSPR Models based on multiple linear regression and back-propagation artificial neural network for predicting electrical conductivity of ionic liquids[J]. *Industrial & Engineering Chemistry Research*, 2024, 63(10): 4571-4584.
- [24] SONG F, AN S H, XIAO Y J, *et al.* Prediction of infinite dilution molar conductivity for unconventional ions: A quantitative structure-property relationship study[J]. *Industrial & Engineering Chemistry Research*, 2021, 60(40): 14625-14634.
- [25] 肖拥君, 时兆翀, 万仁, 等. 反向传播神经网络用于预测离子液体的自扩散系数 [J]. *化工学报*, 2024, 75(2): 429-438.
- [26] DELLEY B. From molecules to solids with the DMol³ approach[J]. *The Journal of Chemical Physics*, 2000, 113(18): 7756-7764.
- [27] GRIMME S, ANTONY J, EHRLICH S, *et al.* A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu[J]. *The Journal of Chemical Physics*, 2010, 132(15): 154104.
- [28] 邓芙蓉, 季常征, 牛亚鹏, 等. 含烟碱多元系统气-液相平衡的理论预测 [J]. *华东理工大学学报(自然科学版)*, 2024, 50(4): 517-523.
- [29] 李明宴, 李进龙, 彭昌军, 等. 基于 COSMO-SAC 模型研究离子液体对氨水溶液汽液平衡的影响 [J]. *化工学报*, 2022, 73(3): 1044-1053.
- [30] LINDER C J. University physics students' conceptualizations of factors affecting the speed of sound propagation[J]. *International Journal of Science Education*, 1993, 15(6): 655-662.
- [31] AGATONOVIC-KUSTRIN S, BERESFORD R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research[J]. *Journal of Pharmaceutical and Biomedical Analysis*, 2000, 22(5): 717-727.
- [32] KLIMASAUSKAS C C. Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real World Performance[M]. Chicago, Illinois, USA: Probus Publishing Company, 1993.
- [33] PADUSZYŃSKI K. In silico calculation of infinite dilution activity coefficients of molecular solutes in ionic liquids: Critical review of current methods and new models based on three machine learning algorithms[J]. *Journal of Chemical Information and Modeling*, 2016, 56(8): 1420-1437.

Prediction of Sound Speed of Ionic Liquids Based on Multiple Linear Regression and Back-Propagation Artificial Neural Networks

Ji Changzheng, WAN Ren, SHI Zhaochong, PENG Changjun, LIU Honglai

(School of Chemistry and Molecular Engineering, East China University of Science and Technology, Shanghai 200237, China)

Abstract: Sound speed of ionic liquids (ILs) can be obtained using various methods, including experimental measurements, semi-empirical models, and theoretical research. Among these methods, the quantitative structure-property relationship (QSPR) model has gained wide attention. However, constructing an effective QSPR model depends on the selection of appropriate molecular descriptors. In this study, there are two descriptors for model construction, namely, the charge density distribution area of ILs at a specific interval obtained using the conductor-like screening model for the segment activity coefficient (COSMO-SAC) and the cavity volume of ILs. Using the multiple linear regression (MLR) approach and the back-propagation artificial neural network (BP-ANN) method, two quantitative structure-property relationship (QSPR) models, namely u -MLR and u -ANN, were proposed respectively to describe the sound speed (u) of ILs. These models incorporated temperature and molecular weight of ILs. The evaluated data set comprised 171 ILs with 5 114 data points. For the entire data set, u -MLR and u -ANN exhibited determination coefficients (R^2) of 0.9706 and 0.9995, respectively. Furthermore, the calculated average absolute relative deviations (AARD) were 1.59% and 0.10% with root mean square errors (RMSE) of 30.68 m/s and 4.12 m/s, respectively. Therefore, the prediction performance of the u -ANN model based on back-propagation artificial neural networks is significantly better than that of the u -MLR based on multiple linear regression.

Key words: sound speed; ionic liquids; artificial neural networks; multiple linear regression; quantitative structure-property relationship

(责任编辑: 李娟)