

基于机器学习的 $\text{BiFeO}_3\text{-PbTiO}_3\text{-BaTiO}_3$ 固溶体居里温度预测

焦志翔¹, 贾帆豪^{1,2}, 王永晨¹, 陈建国¹, 任伟², 程晋荣¹

(上海大学 1. 材料科学与工程学院; 2. 物理系, 量子与分子结构国际中心, 上海 200444)

摘要: 钙钛矿(ABO_3)型压电陶瓷的发展已有几十年历史, 现存有大量数据, 从这些数据中寻找出材料结构与性能之间的关系很有意义。本工作收集了 $\text{BiFeO}_3\text{-PbTiO}_3\text{-BaTiO}_3$ 钙钛矿型压电陶瓷居里温度(T_c)实验数据, 通过机器学习, 构建钙钛矿型压电陶瓷 T_c 的预测模型。热力学角度, T_c 与约合质量符合二次多项式关系, 但偏差较大。选择元素信息、物理量、空间群编号等基础描述符, 利用基于压缩感知原理的 SISO(Sure Independence Screening and Sparsifying Operator)方法进行机器学习, 找出了 T_c 与成分之间的相关性。比较不同描述符在不同维度上的均方根误差 RMSE (Root Mean Square Error), 发现描述符越多、越基础, 维数越大、RMSE 越小。同时比较相同个数描述符在同一维度下的 RMSE, 用约合质量、A 位和 B 位的离子半径比、A 位和 B 位的未填充电子数比和 Ba、Pb、Bi 的元素含量等六个描述符构建出最优的四维模型, 其 RMSE 为 $0.59\text{ }^\circ\text{C}$, 最大绝对误差(MaxAE)为 $1.38\text{ }^\circ\text{C}$, 外部测试的平均相对误差 MRE (Mean Relative Error)为 1.00%。结果表明, 利用 SISO 可以进行有限样本钙钛矿型压电陶瓷 T_c 的机器学习预测。

关键词: 钙钛矿型压电陶瓷; 机器学习; 居里温度; SISO

中图分类号: TQ174 文献标志码: A

Curie Temperature Prediction of $\text{BiFeO}_3\text{-PbTiO}_3\text{-BaTiO}_3$ Solid Solution Based on Machine Learning

JIAO Zhixiang¹, JIA Fanhao^{1,2}, WANG Yongchen¹, CHEN Jianguo¹, REN Wei², CHENG Jinrong¹

(1. School of Materials Science and Engineering, Shanghai University, Shanghai 200444, China; 2. Department of Physics, International Center for Quantum and Molecular Structures, Shanghai University, Shanghai 200444, China)

Abstract: Perovskite (ABO_3) piezoceramics have been developed for several decades, and there are a lot of data available. It is of great significance to find relationships between structure and properties of materials from these data. In this work, experimental data of Curie temperature (T_c) of $\text{BiFeO}_3\text{-PbTiO}_3\text{-BaTiO}_3$ solid solution of perovskite piezoelectric ceramics was collected to build the model to predict the T_c . From the perspective of thermodynamics, the quadratic polynomial relationship between T_c and reduced mass was introduced but the deviation was relatively large. More descriptors (including element information, physical quantities, space groups number) and SISO (Sure Independen-

收稿日期: 2022-02-17; 收到修改稿日期: 2022-03-29; 网络出版日期: 2022-08-04

基金项目: 水声对抗技术重点实验室开放基金(JCKY2020207CH02); 国家自然科学基金(51872180, 51672169)

Open Fund project of Key Laboratory of Underwater Acoustic Countermeasures Technology (JCKY2020207CH02); National Natural Science Foundation of China (51872180, 51672169)

作者简介: 焦志翔(1996-), 男, 硕士研究生. E-mail: jzxxxzj@163.com

JIAO Zhixiang (1996-), male, Master candidate. E-mail: jzxxxzj@163.com

通信作者: 程晋荣, 研究员. E-mail: jrcheng@shu.edu.cn; 贾帆豪, 博士. E-mail: fanhaojia@shu.edu.cn

CHENG Jinrong, professor. E-mail: jrcheng@shu.edu.cn; JIA Fanhao, PhD. E-mail: fanhaojia@shu.edu.cn

dence Screening and Sparsifying Operator) were used for machine learning to find the correlation between T_c and components. Comparing the root mean square error (RMSE) of different descriptors and dimensions, it's found that more descriptors, more fundamental the descriptors are, and larger dimension will result in smaller RMSE to be used. Meanwhile, RMSE of the same number of descriptors in the same dimension are compared. The optimal four-dimensional model is build using six descriptors: reduced mass, the ratio of A- and B-site ion radii, the ratio of A- and B-site unfilled electrons and element contents of Ba, Pb and Bi. RMSE and maximum absolute error (MaxAE) of our model are $0.59\text{ }^\circ\text{C}$ and $1.38\text{ }^\circ\text{C}$, respectively. The average relative error (MRE) of external test is 1.00%. Our results indicate that SISSO machine learning based on limited samples is suitable for the predication of T_c of perovskite piezoelectric ceramics.

Key words: perovskite piezoelectric ceramics; machine learning; Curie temperature; SISSO

钙钛矿结构 $\text{BiFeO}_3\text{-PbTiO}_3$ (BF-PT)基压电陶瓷具有准同型相界(MPB)与高居里温度(T_c), 在智能机电系统传感器、滤波器和驱动器等方面有广泛的应用前景^[1-4]。图 1 是典型 ABO_3 钙钛矿结构示意图^[5], 通常, A 位是低价、半径较大的离子; B 位为高价、半径较小的离子, 处于氧八面体中心。居里温度 T_c 是 ABO_3 钙钛矿铁电/压电陶瓷的重要参数, 它表示铁电相与顺电相之间转变温度。压电陶瓷的制备通常采用固相反应法, 用材多、耗时长。因此, 从实验上获得的 T_c 数据集较小较稀疏。传统研究在寻找合适 T_c 压电陶瓷的过程中, 需要反复进行实验验证。这种实验室试错的方法不适用研究组分较多的压电陶瓷体系, 因为每增加一种新的元素, 样本空间将呈现数量级式地增加。

近年来, 数据驱动的机器学习(Machine Learning, ML)材料研究方法引起了学界的广泛关注^[6-9]。ML 越来越多地被应用于分子科学、材料科学和生物科学等领域^[10-12]。数据驱动+机器学习的研究范式显著加速了探索复杂陶瓷样本空间的材料研究。与传统实验或演绎方法不同, ML 通过正交展开、模式识别、人工神经网络和遗传算法等方式对材料进行优化设计并取得了良好的效果^[13-16]。然而, 关于

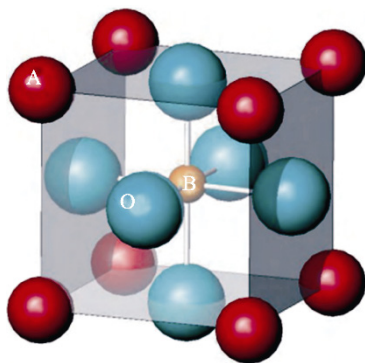


图 1 ABO_3 钙钛矿结构示意图

Fig. 1 Schematic diagram of ABO_3 perovskite structure

钙钛矿结构压电材料的宏观极化与组分关系方面的仿真设计结果仍差强人意^[17]; 利用高通量密度泛函理论筛选高性能压电材料在准确性等方面也值得商榷^[18]。

Lookman 等^[19]小数据集主动学习模型加速了目标性能陶瓷的实验合成, 具体步骤包括: (i)训练已有数据集并建立预测模型; (ii)预测最高性能未知体系的组分; (iii)实验验证该组分的性能; (iv)把新的数据加入原始数据集, 并重新训练得到新的模型; (v)重复以上过程若干次, 直到获得足够满意的高性能陶瓷。这套方法适用于目标性能陶瓷的快速设计。然而, 一个非常重要的技术问题是构建小数据集机器学习模型时, 如何选择好的特征描述符? Askanazi 等^[20]对比了一系列“黑箱”回归算法(如 K-近邻算法、支持向量机、随机森林等), 并指出输入回归模型的特征变量的选择对小数据集模型非常重要。

OUYANG 等^[21]基于压缩感知理论提出了一套 SISSO (Sure Independence Screening and Sparsifying Operator)特征工程研究框架^[22-23]。该方法从基础物理量出发, 构造大量稀疏的高维描述符, 然后根据关联性降维筛选最有效的描述符, 并给出描述符与目标之间的函数关系。他们应用 SISSO 得到了不同氧化态和任意配位时的离子半径, 扩展了 Shannon 离子半径数据库。

本研究以课题组 $\text{BiFeO}_3\text{-PbTiO}_3\text{-BaTiO}_3$ (BF-PT-BT)三元系固溶体压电陶瓷的 T_c 实验数据为对象, 利用 SISSO 进行数据挖掘, 建立材料组分和 T_c 之间的量化关系。首先, 从热力学角度, 采用约合质量作为反映组分变化的描述符; 然后, 根据化学式构造更多基础描述符, 通过机器学习筛选描述符并得出合适的预测模型; 最后, 通过外部验证预测模型的可靠性。最终建立的一种多维度、多描述符、反映组分和性能关系的量化模型能够快速、简单、有效地预测钙钛矿结构压电陶瓷的居里温度。

1 数据集和描述符

首先, 选取 22 个 BF-PT-BT 三元系压电陶瓷的组分和居里温度实验数据^[24], 建立内部训练集; 再选取 3 个相同体系实验数据作为外部测试集^[25], 这些数据均在同一实验环境下采用传统的固相反应法制得。通过测量材料的介电常数随温度升高的变化, 将 100 kHz 频率下的介电常数变化的峰值对应的温度作为居里温度。居里温度为材料的本征参数与组分的关联性较大, 因此机器学习预测的准确率会较高且实用性较强。

Goldschmidt 容忍因子作为特征参数被应用在描述钙钛矿晶体和陶瓷材料上已约 100 年, 现被大量回归算法机器学习模型所使用。然而该容忍因子主要用于描述氧转模式和铁电模之间的关联性, 或者说结构相关性, 其本身就是基础特征 r_A 和 r_B (A 位和 B 位离子半径) 的高维复合的一种数学表示。但该容忍因子在描述居里温度上存在一定的局限性。基于 SISO 方法可以自动地构造出比它更优异的“新容忍因子”, 在判断是否为钙钛矿结构的精度可以达到 92%, 而 Goldschmidt 容忍因子只有近 75%。针对 BF-PT-BT 固溶体居里温度预测, 本工作只选择了 r_A 和 r_B 作为基础特征, 而构造类似容忍因子的过程则交给 SISO 自动实现。最终选出了 4 个最重要的高维变量 (即 x_i 的函数) 并用其构造出了回归模型函数 y , 后续章节将验证该高维变量模型的精度。

表 1 给出选择的特征量及其物理意义, 包括元素信息、物理量、空间群编号在内共 12 个基础描述符^[27]。其中, μ 为原胞约合质量, A 、 B 分别为 ABO₃ 钙钛矿中占据 A、B 位元素性质的组分加权平均值。除 μ 和元素含量外, 其余特征量均采用 A 和 B 比值的形式, 以体现出 A 位和 B 位元素的相对性质。

2 数据处理和机器学习

从热力学角度, 钙钛矿结构氧化物原胞的涨落可以用约合质量 μ 来描述^[28], 用下列公式计算:

$$\mu^{-1} = \sum_{i=A,B,3O} \frac{1}{m_i} \quad (1)$$

$$m_{i=(A \text{ or } B)} = (1 - \sum_j x_j) m_{M_0} + \sum_j x_j m_{M_j} \quad (2)$$

其中, m_{M_j} ($j=0, 1, \dots, n$) 代表占据 A 或 B 位第 j 个元素的相对原子质量, x_j 为对应元素含量。将钙钛矿氧化物的组成变化归纳为一个参量 μ , 可以简化组分与性能之间关系。本研究首先对 BF-PT-BT 三元固溶体居里温度实验数据 T_c 与约合质量 μ 以及化学压 μ^*A/B 关系进行拟合, 获得了组分和 T_c 的量化关系; 然后引入更多特征量, 采用 SISO 筛选出最合适的基础描述符并确定描述符和模型的维度。最后, 通过外部数据集验证模型的可靠性。

3 结果和讨论

3.1 二次多项式拟合

图 2 给出了 T_c 与 μ 和 μ^*A/B 的关系。彩色点表示实验数据, 红色实线是根据文献报道的二次多项式的拟合结果^[28]。可以看出, 实验结果和拟合结果总体是一致的, 即随 μ 和 μ^*A/B 增大, T_c 分别上升和下降, 具体到组分, 即 BT 和 PT 含量增加导致 T_c 下降。以 μ 为自变量, 二次多项式拟合结果的决定系数 R^2 更大, 拟合效果更好。 μ^*A/B 虽然提升了描述符的复杂程度, 但是拟合结果误差反而增大。总的说来, 仅用 μ 或 μ^*A/B 作为描述符的拟合结果误差是比较大的。

表 1 基础描述符及其物理意义

Table 1 Basic descriptor and related physical meaning

Feature	Name	Physical attributes
μ	Reduced mass	Reduced mass of atoms
A/B	Ionic radius	Shannon ion radius
A/B_C	Covalent radius	The covalent bond radius of an atom
A/B_E	Electronegativity	Electronegativity of atoms
A/B_NV	NValence	The number of electrons of unfilled orbitals
A/B_NU	NUfilled	The number of electrons of unfilled orbitals
A/B_S	Space group numbering	The serial number of the element's space group in the space group table
Ti,Fe,Ba,Pb,Bi	Element content	Element content ratio of metal ions

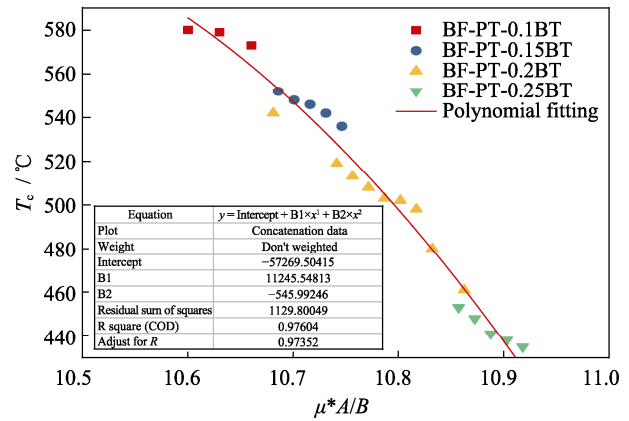
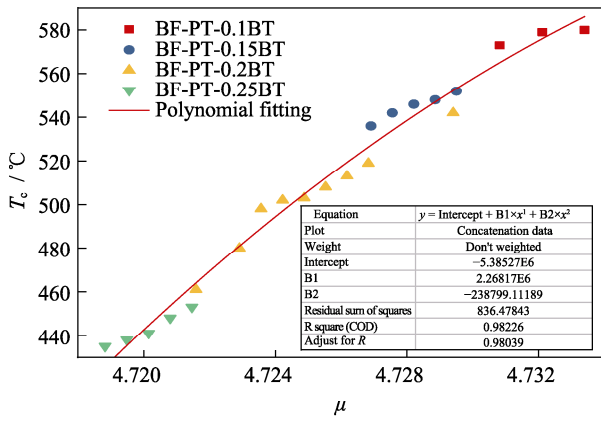


图 2 T_c 与 μ 和 μ^*A/B 的关系
Fig. 2 Relationship between T_c and μ or μ^*A/B

3.2 特征变量选择

为降低预测误差, 引入表 1 中更多的基础描述符。筛选掉冗余的描述符, 帮助降低输入空间维数, 同时尽可能地保证了信息的完整性^[29]。图 3 显示了 T_c 和候选描述符之间的相关性(Correlation, 用皮尔森相关系数 R 表示)^[30]。可以看出, μ 、 A/B_E 、 A/B_NV 、 A/B_NU 、 Fe 、 Pb 、 Bi 和 T_c 呈正相关, A/B 、 A/B_C 、 A/B_S 、 Ti 、 Ba 则和 T_c 呈负相关; μ 与 A/B 是与 T_c 相关性最大的两个描述符。结合候选描述符与 T_c 之间的相关性系数, 进一步利用 SISSO 得到的 RMSE^[31]来选择最终的基础描述符。RMSE 越小, 表示拟合程度越好。输出公式中描述符出现的次数可以用来判断特征量的重要性。 R 和 RMSE 定义如下:

$$R = \frac{n \sum_{i=1}^n x_i e_i - \sum_{i=1}^n x_i \sum_{i=1}^n e_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \sqrt{n \sum_{i=1}^n e_i^2 - \left(\sum_{i=1}^n e_i\right)^2}} \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - e_i)^2} \quad (4)$$

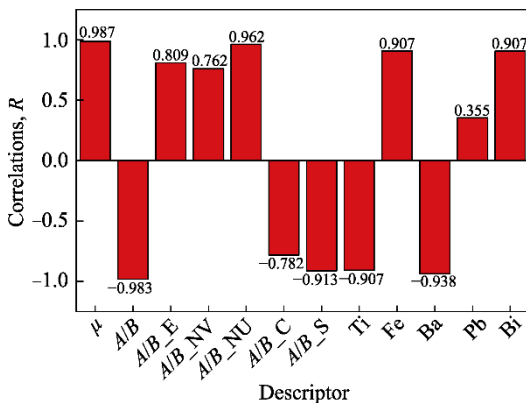


图 3 T_c 与候选描述符的相关性
Fig. 3 Correlations between the T_c and the candidate descriptors

其中, x_i 是第 i 个样本的描述符参数, p_i 和 e_i 分别为 T_c 的预测值和实验值, n 是样本数量。

图 4 给出了不同运算方式下, 以 μ 或 μ^*A/B 为描述符训练模型的 RMSE 和 MaxAE 随维度变化的趋势。运算方式 1 代表加、减、乘、除、倒数、平方和立方; 运算方式 2 在运算方式 1 基础上增加了绝对值、自然指数、对数、平方根和立方根。SISSO 预测模型一般不超过四维, 即多项式项数加上常数项不超过五项。维数越多, 一方面会过拟合, 另一方面会占用更多的运算时间和资源。比较图 4(a~d), 发现维度越大, 运算方式越复杂, RMSE 和 MaxAE 越低, 这说明高维度和复杂运算方式能产生更好的组合描述符, 充分挖掘和利用描述符的信息。从图 4 还可以看出, 仅用一个描述符时, 描述符越复杂, RMSE 越大, 说明蕴含过多信息的描述符会对运算产生较大干扰。总之, 选择更基础描述符和运用更复杂的运算方式得到的拟合结果误差更小。

采用运算方式 2, 图 5 给出了两个描述符(图 5(a)) 和三个描述符(图 5(b))的拟合误差随维度的变化趋势。对比图 5 和图 4(a, c), 发现使用多个基础描述符比组合描述符的输出误差更低。

图 6 反映了描述符的变化和模型维度对拟合结果 RMSE 和 MaxAE 的影响。图 6(a, b) 分别是以 μ 和 A/B 为前两个描述符改变第三个描述符, 和以 μ 、 A/B 、 Ba 、 Pb 和 Bi 为前五个描述符改变第六个描述符, 对 RMSE 和 MaxAE 的影响。添加更多描述符明显降低了预测模型误差。第三和第六个描述符的变化对 RMSE 的影响较小, 而对 MaxAE 的影响较大。这说明不同描述符在运算过程中有交互作用, 在共同减小整体误差的同时, 对局部误差影响可能增大。值得注意的是, A/B_NU 分别作为第三和第六个描述符时, 拟合结果的 RMSE 和 MaxAE 都是最

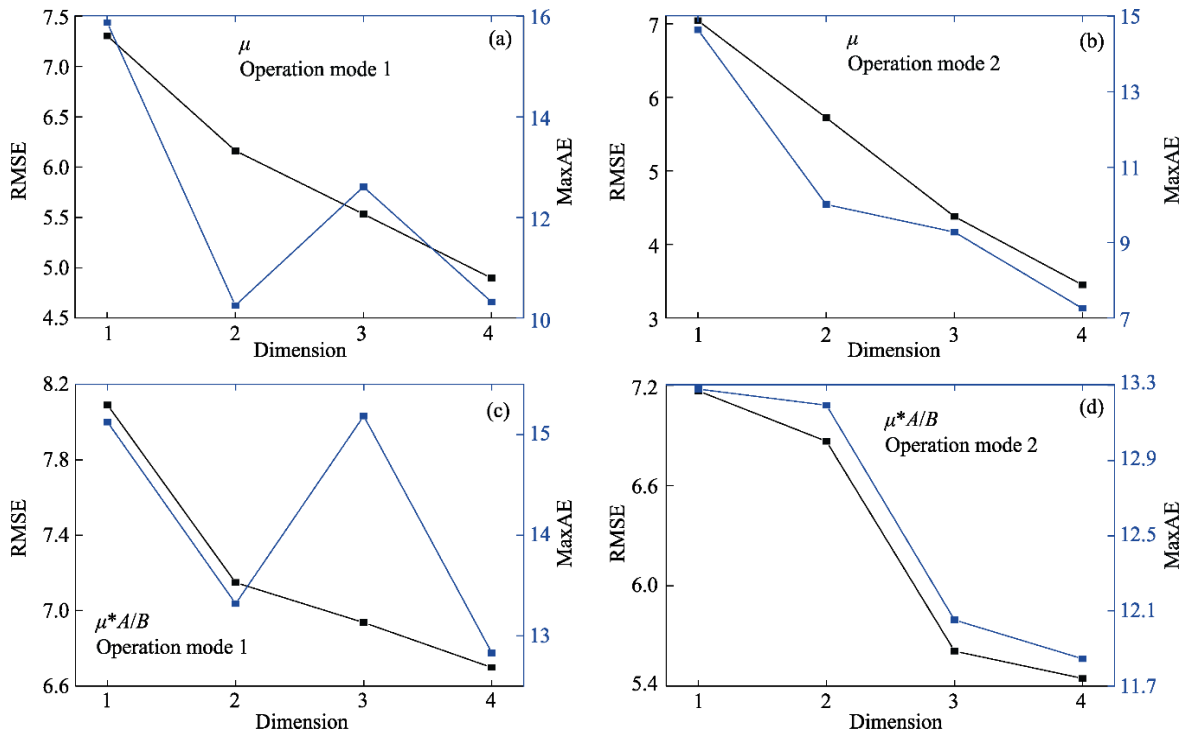


图 4 不同运算方式下 μ 和 $\mu^*/A/B$ 随维度变化的 RMSE 和 MaxAE

Fig. 4 RMSE and MaxAE of μ and $\mu^*/A/B$ varied with dimension under different operation modes (a, c) use operation mode 1; (b, d) use operation mode 2; (a, b) use the descriptor μ ; (c, d) use the descriptor $\mu^*/A/B$

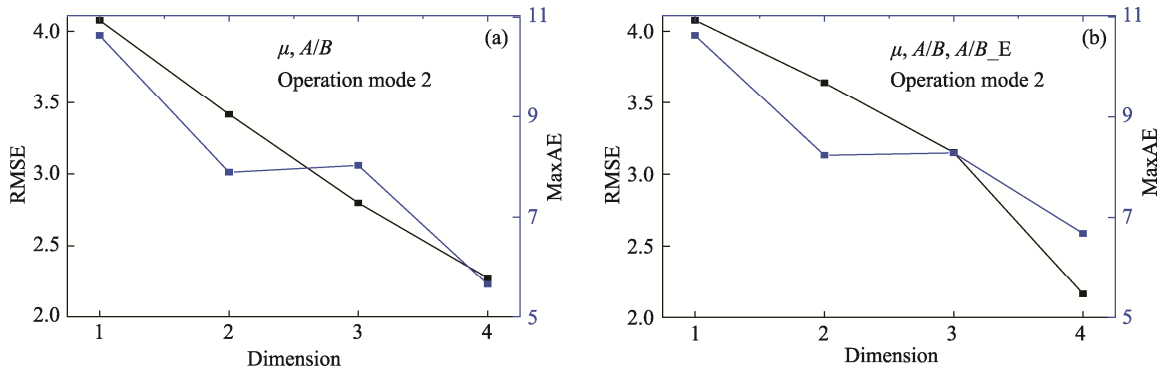


图 5 采用两个描述符(a)和三个描述符(b)拟合结果的 RMSE 和 MaxAE 随维度的变化

Fig. 5 RMSE and MaxAE of two descriptors (a) and three descriptors (b) as a function of dimension

小的。因此,除了 μ 和 A/B 外, A/B_NU 对材料 T_c 的影响较大,这与图 3 中 A/B_NU 描述符相关性系数较大是一致的。图 6(c)表明,从初始两个描述符逐步添加至 7 个描述符, RMSE 逐渐降低,添加第三个描述符,在描述符的交互作用下, MaxAE 有所提高。总之,选择多个相对独立的基础特征量为描述符有助于降低拟合结果的误差。7 个描述符时,运算时间延长了 1 个数量级,但拟合误差没有明显改善。最终,选择六个描述符来拟合模型是最优的。图 6(d) 显示了最终选择的六个基础描述符模型的 RMSE 和 MaxAE 随维度的变化。

3.3 模型的建立与验证

表 2 给出了正则化的外部测试集数据,其中, μ' 和 A/B' 分别为 $(\mu-4.7)\times 10$ 和 $(A/B-2.2)\times 10$ 。拟合得出的预测模型如下:

$$y = a \frac{(x_4 - x_1) / (x_3 - x_4)}{x_1^3 (x_5 + x_1)} + b \frac{(x_3 - x_1) - (x_2 - x_3)}{(x_2 - x_6) - (x_3 - x_5)} + c \sqrt[3]{\frac{x_1}{x_4}} - \frac{x_6 - x_2}{x_5 - x_1} + d \frac{x_2^2 / (x_2 - x_5)}{(x_3 - x_5) - |x_1 - x_2|} + e \quad (5)$$

其中, $a=1.977447895$, $b=2.964590758$, $c=1.778816427$, $d=0.063342758$, $e=647.6695156$, x_i 是

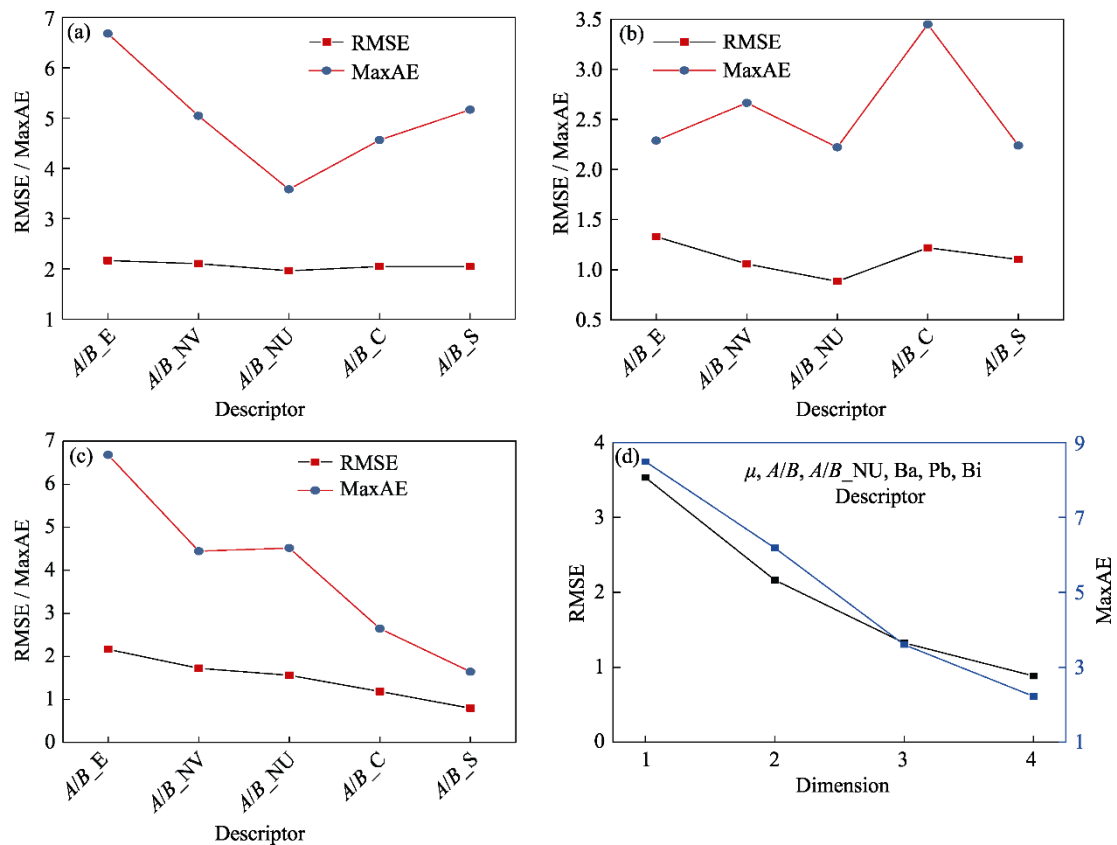


图 6 描述符的变化和模型维度对拟合结果 RMSE 和 MaxAE 的影响

Fig. 6 Effects of the changes of descriptor and model dimension on the RMSE and MaxAE

(a) μ and A/B are the first two descriptors, and the third descriptor changes; (b) μ , A/B , Ba, Pb, and Bi are the first five descriptors, and the sixth descriptor changes; (c) μ and A/B are the first two descriptors and the third to seventh descriptors are introduced one by one. (d) Effect of the change of the dimension of the model fitted by six finally selected descriptors (μ , A/B , A/B_NU , Ba, Pb, and Bi) on RMSE and MaxAE

表 2 调整描述符参数后的测试集

Table 2 The test sets after adjusting descriptor parameters

Sample	$T_c/^\circ\text{C}$ (y)	$\mu'(x_1)$	$A/B'(x_2)$	$A/B_NU(x_3)$	Ba (x_4)	Pb (x_5)	Bi (x_6)
$\text{Bi}_{0.62}\text{Pb}_{0.23}\text{Ba}_{0.15}\text{Fe}_{0.62}\text{Ti}_{0.38}\text{O}_3$	547	0.2821	0.3882	0.5036	0.0750	0.1150	0.3100
$\text{Bi}_{0.6}\text{Pb}_{0.25}\text{Ba}_{0.15}\text{Fe}_{0.6}\text{Ti}_{0.4}\text{O}_3$	540	0.2690	0.4606	0.5000	0.0750	0.1250	0.3000
$\text{Bi}_{0.54}\text{Pb}_{0.31}\text{Ba}_{0.15}\text{Fe}_{0.54}\text{Ti}_{0.46}\text{O}_3$	502	0.2295	0.6789	0.4897	0.0750	0.1550	0.2700

基础描述符。该预测模型的 RMSE 为 0.59°C , MaxAE 为 1.38°C 。

模型中出现次数最多的为 x_1 、 x_2 和 x_3 (μ 、 A/B 和 A/B_NU), 分别出现了 7、6 和 5 次, 说明 μ 、 A/B 和 A/B_NU 的确对 T_c 影响较大, 这与相关性和 SISSO 的误差分析结果一致。图 7 为训练集和测试集实验值和预测 T_c 值。表 3 给出测试集的外部验证结果, 平均相对误差(MRE)为 1.00%。测试集与总数据集的预测 T_c 与实验 T_c 的相关系数 R 分别为 0.99998 和 0.99995。其中 MRE 的定义如下:

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{p_i - e_i}{e_i} \right| \times 100\% \quad (6)$$

4 结论

从热力学角度, T_c 与约合质量 μ 的二次多项式关系表明, T_c 随 μ 增加而上升, 但拟合结果误差较大。选取六个相对独立的基础描述符, 通过 SISSO 复杂运算构建得到的四维预测模型显著提升了预测精度。其中, μ 、 A/B 和 A/B_NU 对 T_c 影响最大。预测模型的 RMSE 为 0.59°C , MaxAE 为 1.38°C , 验证集的 MRE 为 1.00%。描述符的选取(拆分、比值、约合)和修正方式还可以进一步降低预测误差。本课题构建的四维预测模型既可以用于预测其它陶瓷体系的居里温度, 还可以进一步用于构建主动学习

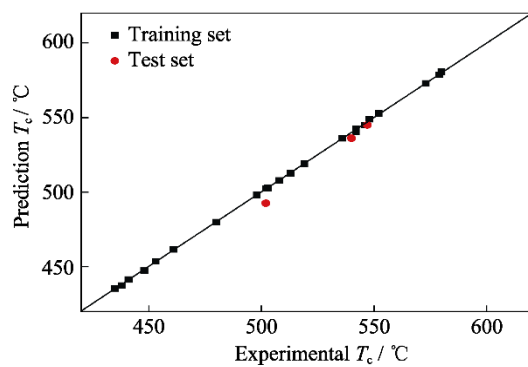


图 7 训练集(黑色方块)和测试集(红色圆形)的实验 T_c 与预测 T_c

Fig 7 Experimental T_c and prediction T_c for training set (black square) and test set (red circle)

表 3 外部验证集结果

Table 3 Results of external verification set

Sample	Experimental $T_c / ^\circ\text{C}$	Prediction $T_c / ^\circ\text{C}$	Absolute error/ $^\circ\text{C}$	Relative error/%
Bi _{0.62} Pb _{0.23} Ba _{0.15} Fe _{0.62} Ti _{0.38} O ₃	547	544.91	2.09	0.38
Bi _{0.6} Pb _{0.25} Ba _{0.15} Fe _{0.6} Ti _{0.4} O ₃	540	536.12	3.88	0.72
Bi _{0.54} Pb _{0.31} Ba _{0.15} Fe _{0.54} Ti _{0.46} O ₃	502	492.52	9.48	1.89

框架, 能够快速、简便地预测钙钛矿型压电陶瓷的居里温度。

参考文献:

- [1] CHEN J, QI Y, SHI G, *et al.* A high temperature piezoelectric ceramic: $(1-x)(\text{Bi}_{0.9}\text{La}_{0.1})\text{FeO}_3-x\text{PbTiO}_3$ crystalline solutions. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control.*, 2009, **56(9)**: 1820–1825.
- [2] CHEN J, JIN D, CHENG J. Impedance spectroscopy studies of $0.7\text{Bi}(\text{Fe}_{1-x}\text{Ga}_x)\text{O}_3-0.3\text{PbTiO}_3$ high temperature piezoelectric ceramics. *J. Alloys Compd.*, 2013, **580**: 67–71.
- [3] CHENG J, YU S, CHEN J, *et al.* Dielectric and magnetic enhancements in $\text{BiFeO}_3\text{-PbTiO}_3$ solid solutions with La doping. *Appl. Phys. Lett.*, 2006, **89(12)**: 122911.
- [4] CHEN J, CHENG J. Enhanced high-field strain and reduced high-temperature dielectric loss in $0.6(\text{Bi}_{0.9}\text{La}_{0.1})(\text{Fe}_{1-x}\text{Ti}_x)\text{O}_3-0.4\text{PbTiO}_3$ piezoelectric. *Ceram. Int.*, 2015, **41(1)**: 1617–1621.
- [5] 谢颖. ABO₃型钙钛矿的相变机理表面稳定性和电子结构的理论研究. 哈尔滨: 黑龙江大学出版社, 2015.
- [6] MURPHY K. Machine learning: a probabilistic perspective. Cambridge: The MIT Press, 2012, **58(8)**: 27–71.
- [7] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, *et al.* Scikit-learn: machine learning in python. *Journal of Machine Learning Research*, 2011, **12(10)**: 2825–2830.
- [8] RUPP M, TKATCHENKO A, MÜLLER K, *et al.* Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.*, 2012, **108(5)**: 058301.
- [9] JORDAN M, MITCHELL T. Machine learning: trends, perspectives, and prospects. *Science*, 2015, **349(6245)**: 255–260.
- [10] ZHONG M, TRAN K, MIN Y, *et al.* Accelerated discovery of CO₂ electrocatalysts using active machine learning. *Nature*, 2020, **581(7807)**: 178–183.
- [11] BATRA R. Accurate machine learning in materials science facilitated by using diverse data sources. *Nature*, 2021, **589(7843)**: 524–525.
- [12] RANDHAWA G, HILL K, KARI L. ML-DSP: Machine learning with digital signal processing for ultrafast, accurate, and scalable genome classification at all taxonomic levels. *BioMed Central*, 2019, **20(1)**: 267–23.
- [13] CHENG Z, ZHU E, CHEN N. Application of orthogonal expansion to mapping and modelling. *Chemometr.*, 1993, **7(4)**: 243–253.
- [14] CHEN N, LI C, QIN P. Chemical pattern recognition applied to materials optimal design and industry optimization. *Chin. Sci. Bull.*, 1997, **42(10)**: 793–799.
- [15] CHEN N, LU W, CHEN R, *et al.* Chemometric methods applied to industrial optimization and materials optimal design. *Chemom. Intel. Lab. Syst.*, 1999, **45(1)**: 329–333.
- [16] CHEN N, ZHU D, WANG W. Intelligent materials processing by hyperspace data mining. *Eng. Appl. Artif. Intel.*, 2000, **13(5)**: 527–532.
- [17] RESTA R. Macroscopic polarization in crystalline dielectrics: the geometric phase approach. *Review Modern Physics*, 1994, **66(3)**: 899–916.
- [18] ARMIENTO R, KOZINSKY B, FORNARI M, *et al.* Screening for high-performance piezoelectrics using high-throughput density functional theory. *Physics Review B*, 2011, **84(1)**: 014103.
- [19] PRASANNA V, BENJAMIN K, ALP S, *et al.* Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning. *Nature Communications*, 2018, **9**: 1668.
- [20] EVAN M, SUHAS Y, ILYA G. Prediction of the Curie temperatures of ferroelectric solid solutions using machine learning methods. *Computational Materials Science*, 2021, **199(7061)**: 110730.
- [21] OUYANG R, CURTAROLO S, AHMETCIK E, *et al.* SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.*, 2018, **2(8)**: 083802.
- [22] OUYANG R, AHMETCIK E, CARBOGNO C, *et al.* Simultaneous learning of several materials properties from incomplete databases with multi-task SISSO. *J. Phys.: Mater.*, 2019, **2(2)**: 024002.
- [23] OUYANG R. Exploiting ionic radii for rational design of halide perovskites. *Chem. Mater.*, 2020, **32(1)**: 595–604.
- [24] NING Z. Dielectric, ferroelectric, piezoelectric and aging properties of $\text{BiFeO}_3\text{-PbTiO}_3\text{-BaTiO}_3$ high temperature piezoelectric ceramics. Shanghai: Master Thesis, Shanghai University, 2020.
- [25] TU T. Fabrication of BF-PT-BT high temperature piezoelectric ceramics and sensors. Shanghai: Master Thesis, Shanghai University, 2020.

- sity, 2017.
- [26] CHRISTOPHER J, CHRISTOPHER S, BRYAN R, *et al.* New tolerance factor to predict the stability of perovskite oxides and halides. *Science Advances*, 2019, **5(2)**: eaav0693.
- [27] BUTLER K, FROST J, SKELTON J, *et al.* Computational materials design of crystalline solids. *Che. Soc. Rev.*, 2016, **45(22)**: 6138–6146.
- [28] YU J, ITOH M. Physics-guided data-mining driven design of room-temperature multiferroic perovskite oxides. *Phys. Status Solidi RRL*, 2019, **13(6)**: 1900028.
- [29] UUSI E, MALM J, IMAMURA N, *et al.* Characterization of RMnO_3 (R=Sc, Y, Dy-Lu): high-pressure synthesized metastable perovskites and their hexagonal precursor phases. *Materials Chemistry & Physics*, 2008, **112(3)**: 1029–1034.
- [30] SCHOBER P, BOER C, SCHWARTE L. Correlation coefficients: appropriate use and interpretation. *Anesthesia and Analgesia*, 2018, **126(5)**: 1763–1768.
- [31] YANG X, LI M, SU Q, *et al.* QSAR studies on pyrrolidine amides derivatives as DPP-IV inhibitors for type 2 diabetes. *Medicinal Chemistry Research*, 2013, **22(11)**: 5274–5283.