

机器学习加速搜寻新型双钙钛矿氧化物光催化剂*

万新阳 章烨辉 陆帅华 吴艺蕾 周楚桦[†] 王金兰[‡]

(东南大学物理学院, 南京 211189)

(2022 年 4 月 1 日收到; 2022 年 5 月 11 日收到修改稿)

$A_2BB'O_6$ 型双钙钛矿氧化物材料, 相比于 ABO_3 型单钙钛矿氧化物材料, 具有更好的稳定性和更宽泛的能带选择范围, 在光催化全解水领域具有良好的应用前景. 然而, 由于晶体结构和组成元素的多样性, 实验和理论上快速、准确搜寻高催化活性的 $A_2BB'O_6$ 型双钙钛矿氧化物材料具有相当大的挑战性. 本文由材料数据库的带隙值数据出发, 采用机器学习与第一性原理相结合的方法, 从 50000 多种 $A_2BB'O_6$ 型双钙钛矿氧化物材料中筛选出近 8000 种可能适用于光催化全解水的材料. 对筛选结果的统计分析表明, B/B' 位均为 d^{10} 金属离子的双钙钛矿氧化物, 更有可能成为全解水光催化剂. 随后通过进一步的第一性原理计算挑选出 Sr_2GaSbO_6 , Sr_2InSbO_6 和 K_2NbTaO_6 这 3 种带边位置合适且不含铅、汞离子的 $A_2BB'O_6$ 型双钙钛矿氧化物材料作为候选的全解水光催化剂.

关键词: 机器学习, 双钙钛矿氧化物, 光催化, 全解水, 第一性原理计算

PACS: 71.15.Mb, 71.20.-b, 71.20.Nr, 81.16.Hc

DOI: 10.7498/aps.71.20220601

1 引言

如何减少化石燃料的使用以及降低二氧化碳的排放, 并寻找可持续的清洁能源, 一直都是人们广泛关注的问题^[1-3]. 太阳能是目前最直接、最广泛的清洁能源, 如何将光能有效地转化为电能或化学能^[4,5]是寻找清洁能源的关键, 而光催化全解水被认为是最理想、最方便的手段^[6-8]. 在高效光催化全解水反应中, 光催化剂的选取则是光催化反应的核心. 相比于价格昂贵、含量稀少的贵金属催化剂, 钙钛矿氧化物材料因其结构稳定、种类多样、易合成、价格低廉等优点近些年受到大家广泛的关注^[9-11]. 然而 ABO_3 型单钙钛矿材料普遍带隙值过大, 导致其对可见光的吸收效率较低. 最近, 新型 $A_2BB'O_6$ 双钙钛矿氧化物在实验上被成功合成并展

现出更好的稳定性和光催化活性^[12,13]. 在 $A_2BB'O_6$ 双钙钛矿氧化物中, B/B' 位通常由高价元素组成, 其组分相较于单钙钛矿更为复杂多样^[14], 同时结构中的 BO_6 与 $B'O_6$ 八面体结构交替排列, 形成的 $B-O-B'$ 结构产生超交换作用^[15], 影响其能带结构, 使其具有更宽泛的带隙选择范围. 因此, $A_2BB'O_6$ 型氧化物双钙钛矿材料被认为能够更好地满足光催化全解水的要求, 有着很广阔的应用前景.

然而, $A_2BB'O_6$ 型双钙钛矿氧化物多样的元素组合在带来多样化的结构和性质的同时, 也对从不同组合中找到适用于光催化全解水的候选材料造成了不小的困难. 基于第一性原理计算的材料设计方法在钙钛矿材料设计中具有不错的准确性, 但对 10^4 数量级的搜索空间来说, 依然需要一种更快速的筛选方法. 随着材料数据库的不断完善, 机器学习作为一种数据驱动的方法^[16], 能够在极短时

* 国家重点研发计划 (批准号: 2017YFA0204800, 2021YFA1500700)、国家自然科学基金 (批准号: 22033002, 22003009) 和江苏省研究生科研与实践创新计划 (批准号: KYCX20_0075) 资助的课题.

[†] 通信作者. E-mail: qh.zhou@seu.edu.cn

[‡] 通信作者. E-mail: jlwang@seu.edu.cn

间内对大量材料进行快速筛选, 已经在新材料设计中展现出了巨大的潜力, 例如二维铁磁材料^[17,18]、双钙钛矿卤化物光伏材料^[19]、多相催化剂^[20]等. 此外, 传统的机器学习模型虽然能够比较成功地预测出新型材料, 但往往局限于某一单一体系, 同时对预测材料之间关联性的挖掘也不足. 因此, 提升机器学习模型对不同结构的适用性、挖掘目标材料的构效关系也是该领域研究关注的重点问题.

本工作的整体研究流程如图 1 所示. 通过机器学习和第一性原理结合的方法, 构建了新型 $A_2BB'O_6$ 型双钙钛矿氧化物 (后文简称双钙钛矿) 材料快速筛选的框架, 包含数据准备、特征选择、带隙分类模型和回归模型的构建, 以及对预测结果构效关系的统计分析和计算结果验证. 首先, 通过整合钙钛矿材料数据库和高通量计算, 获得了近 2500 种钙钛矿氧化物材料数据. 其次, 建立了两步机器学习模型框架, 第一步通过机器学习分类模型, 挑选出带隙值大于 0.2 eV 的双钙钛矿材料, 第二步建立机器学习回归模型, 对分类模型结果中带隙大于 0.2 eV 的双钙钛矿材料进行预测, 得到准确带隙值. 然后, 将机器学习模型用于预测集, 对 50000 多组新双钙钛矿材料带隙进行了快速预测和筛选, 得到近 8000 种新型光催化材料. 在此基础上, 进一步对预测结果进行统计分析, 总结了 B/B' 位金属离子组成以及钙钛矿结构的对称性对双钙钛矿带隙大小的影响规律. 最后, 挑选了 29 个带隙值在 0.2 eV 以上的双钙钛矿材料进行 DFT 的验证, 并最终挑选出 3 种不含铅汞离子、带隙大小合适、带边位置合适的双钙钛矿材料作为全

解水光催化剂.

2 计算方法

2.1 XGBoost 算法

本文使用 XGBoost (eXtreme gradient boosting)^[21] 算法训练分类和回归模型. 不同于梯度提升 (gradient boosting)^[22] 算法, XGBoost 算法在损失函数上使用了二阶泰勒近似, 使得其与 Newton Raphson 方法相联系.

2.2 密度泛函理论计算

本文采用基于密度泛函理论 (density functional theory, DFT) 的第一性原理计算方法, 通过应用软件 VASP (Vienna ab-initio simulation package, VASP)^[23] 来计算材料的电子结构. 计算采用了广义梯度近似 (generalized gradient approximation, GGA) 下的 PBE (Perdew-Burke-Ernzerhof) 泛函^[24], 并考虑了铁磁与 A 型反铁磁的情况^[25]. 采用投影缀加平面波 (projected augmented wave, PAW) 方法来描述离子实和价电子之间的相互作用^[26]. 平面波的截断能设置为 500 eV, 并采用 Γ 中心的 Monk horst-Pack 方法^[27] 进行第一布里渊区采样, k 点网格选取为 $3 \times 3 \times 3$, 离子弛豫的能量收敛标准设置为 10^{-6} eV, 原子之间相互作用力的收敛标准为 0.02 eV/Å. 对于过渡金属元素, 参考 Materials Project 的数据来源, 引入了相应的 Hubbard U^[28] 修正进行相关计算, 修正值如附录表 A1 所示. 对于最终预测的 3 个新型双钙钛矿材料, 采用 HSE06

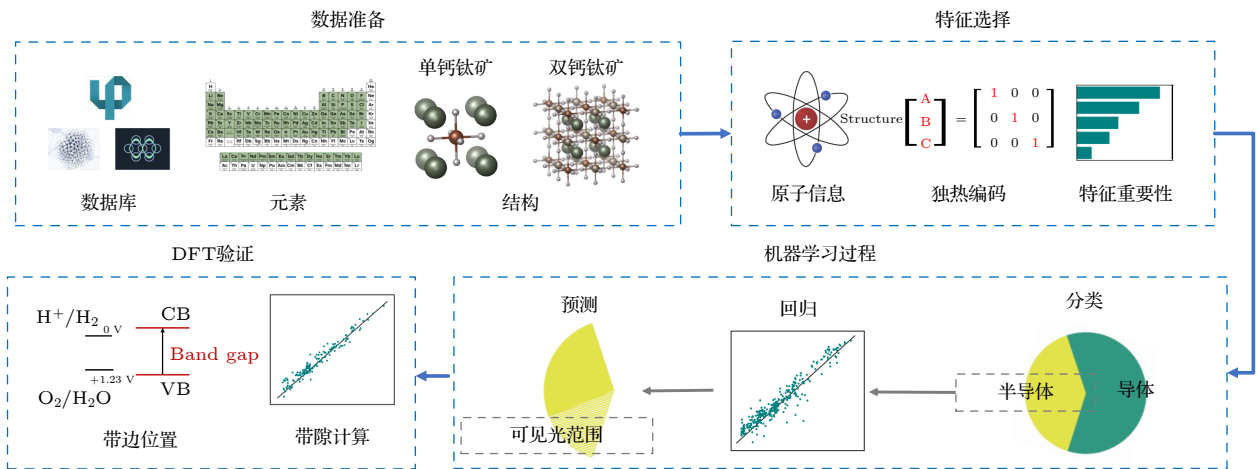


图 1 基于两步机器学习算法的双钙钛矿氧化物筛选框架. 包括数据准备、特征选择、机器学习过程和 DFT 验证 4 个步骤

Fig. 1. Multistep machine learning-based screening framework for double perovskite oxides. There are four steps including data collection, feature selection, machine learning process and DFT verification.

(hybrid Heyd-Scuseria-Ernzerhof, HSE) 杂化泛函^[29]进行了进一步的计算。

3 结果讨论

3.1 数据准备

首先, 从 AFLOW^[30], OQMD (the open quantum materials database)^[31] 和 Materials Project^[32] 等常用数据库中挑选出钙钛矿结构的氧化物数据. 单钙钛矿氧化物和双钙钛矿氧化物的化学式分别表示为 ABO_3 和 $A_2BB'O_6$, 但并非所有满足这一元素组成和化学式的材料都是钙钛矿结构. 经典的钙钛矿结构判定方式是 Goldschmidt 提出的容忍因子 t ^[33], 然而这种判定方式原本是基于单钙钛矿材料提出的, 且在很多情况下不够准确. 很多研究团队因此对钙钛矿判定条件进行了更进一步的研究, 如 Sun 等^[34] 提出的 $(\mu + t)^n$, Bartel 等^[35] 提出的新容忍因子 $\tau = \frac{r_X}{r_B} - n_A \left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)} \right)$, Weng 等^[36] 提出的 μ/t 等. 考虑分类准确率, 采用了 Filip 等^[37] 提出的基于八面体、倾斜度等限制的多重判定条件. 利用该判定条件对前述材料数据库进行数据清洗, 并通过结构相似度分析^[38] 剔除重复数据后, 得到单钙钛矿 (ABO_3)1027 组和双钙钛矿 ($A_2BB'O_6$)463 组. 根据结构相似度分类结果, 保留主要的 3 种类别, 分别命名为“立方类结构”“正

交类结构”和“菱方类结构”, 结构如图 2(a) 所示.

众所周知, 机器学习模型构建需要足够多的数据量作为支撑. 根据以上筛选方法得到的 463 组双钙钛矿数据较少, 为了保证训练数据的完备性, 将数据库中满足判定条件的 1027 组单钙钛矿数据加入训练集中 (等价于 B' 位与 B 位相同的双钙钛矿). 同时, 从立方类结构的双钙钛矿 (结构简单, 便于大量获取数据) 中随机挑选数据库中不存在的 1005 组进行第一性原理高通量计算, 得到相应的 PBE 带隙值. 在双钙钛矿结构中, 会因为等效的 B 位和 B' 位阳离子不同的排布方式, 对带隙产生不同的影响. 在本工作的数据中并未对每种结构不同的排布一一计算, 而采用最常见的 F 型阳离子排布方式. 其中计算数据皆为 F 型排布, 数据库主句中绝大部分为 F 型排布, 这样的排布类型能量往往要远低于其他排布类型^[39]. 最终合计得到 2495 组钙钛矿氧化物的带隙值, 这足以支持机器学习模型训练^[19,40]. 预测集则由满足钙钛矿结构判定条件, 且并未在训练集中出现的双钙钛矿构成, 共 56894 组数据. 从图 1(b) 中可以看出训练集双钙钛矿 A 位和 B 位 (包含 B' 位) 元素分布情况, 即 A 位元素主要出现在碱金属、碱土金属、类金属及过渡金属区域的 I B, II B 和 III B 副族, 而 B 位元素主要集中在过渡金属区域. 其原因在于 B 与 B' 位 6 配位的环境决定了 B 位元素离子半径较小 (过渡金属离子等), 而 A 位离子半径较大 (稀土或碱土

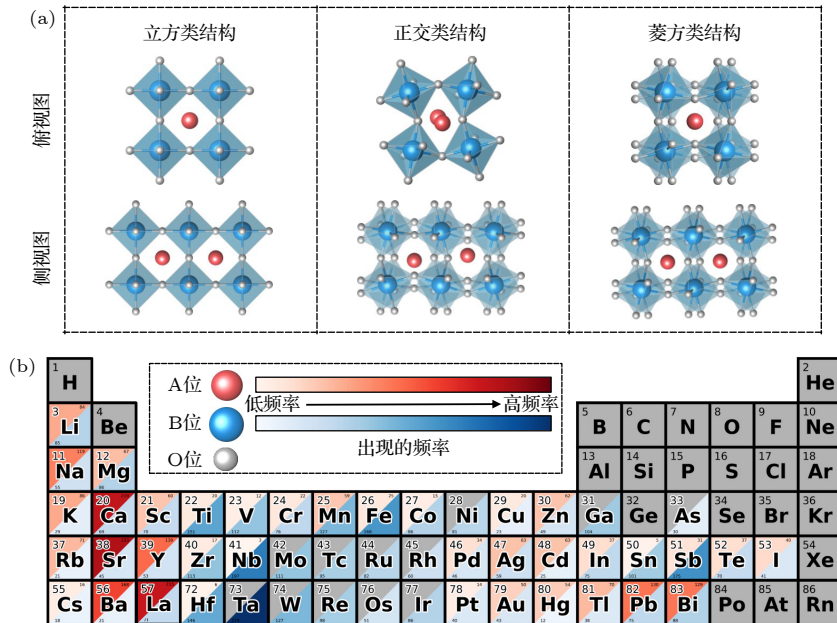


图 2 (a) 3 种不同的钙钛矿构型; (b) 训练集中 A 位和 B 位元素出现的频率

Fig. 2. (a) Three different perovskite structures; (b) occurrence frequency of A - and B -site elements in the training set.

金属元素). 因此导致了钙钛矿结构中元素多变的价态, 影响钙钛矿包括电子结构在内的多种性质^[41].

除数据量外, 影响机器学习模型准确性的另一个主要因素是特征 (feature) 与目标性质的关联性. 为了综合考虑元素组分因素和晶体结构因素, 本文选取构成双钙钛矿 A 位和 B/B' 位原子的物理性质 (单钙钛矿 B' 位同 B 位) 作为一部分特征, 将 3 种不同构型钙钛矿材料进行独热编码^[42] 作为另一部分特征, 剔除信息冗余和与目标性质关联不高的特征后, 最终获得 45 个特征来组成描述材料的特征向量, 特征符号表示如附录表 A2 所示.

3.2 机器学习模型对带隙的预测

为了提升双钙钛矿材料带隙预测的准确率, 本文采用了两步建模的方法. 第一步, 选择适用于结构化数据与中低维度特征的 XGBoost 算法训练一个分类模型, 用来筛选 PBE 带隙值大于 0.2 eV 的钙钛矿材料, 排除金属或光吸收明显不佳的窄带隙

材料. 在训练和测试集的数据中共有 757 组数据的带隙值大于 0.2 eV. 图 3(a) 列出了特征重要性分析中排在前 10 的特征, 可以看出结构独热编码 (structure) 的重要性要远高于其他特征, 意味着结构的对称性对钙钛矿材料的带隙值影响显著. 图 3(a) 中分类结果的混淆矩阵表明分类模型特异度为 94.6%, 表明了该模型可以有效筛选 0.2 eV 以下的双钙钛矿材料. 机器学习分类模型性能常用受试者操作曲线线下面积 (area under curve, AUC) 评估, 其值一般在 0.5—1.0 之间, 越接近于 1.0 模型效果越好. 由图 3(b) 可以看出, 带隙分类模型在测试集上的 AUC 值可达 0.86, 证明了模型的可靠性. 将训练好的分类模型应用到 56894 组预测数据上, 得到了 14407 组带隙值大于 0.2 eV 的双钙钛矿材料, 带隙大于 0.2 eV 的数据占预测数据总量的 1/4 左右. 分类过程作为第一步筛选, 剔除了大量金属性或窄带隙的双钙钛矿材料, 有助于回归模型的进一步构建.

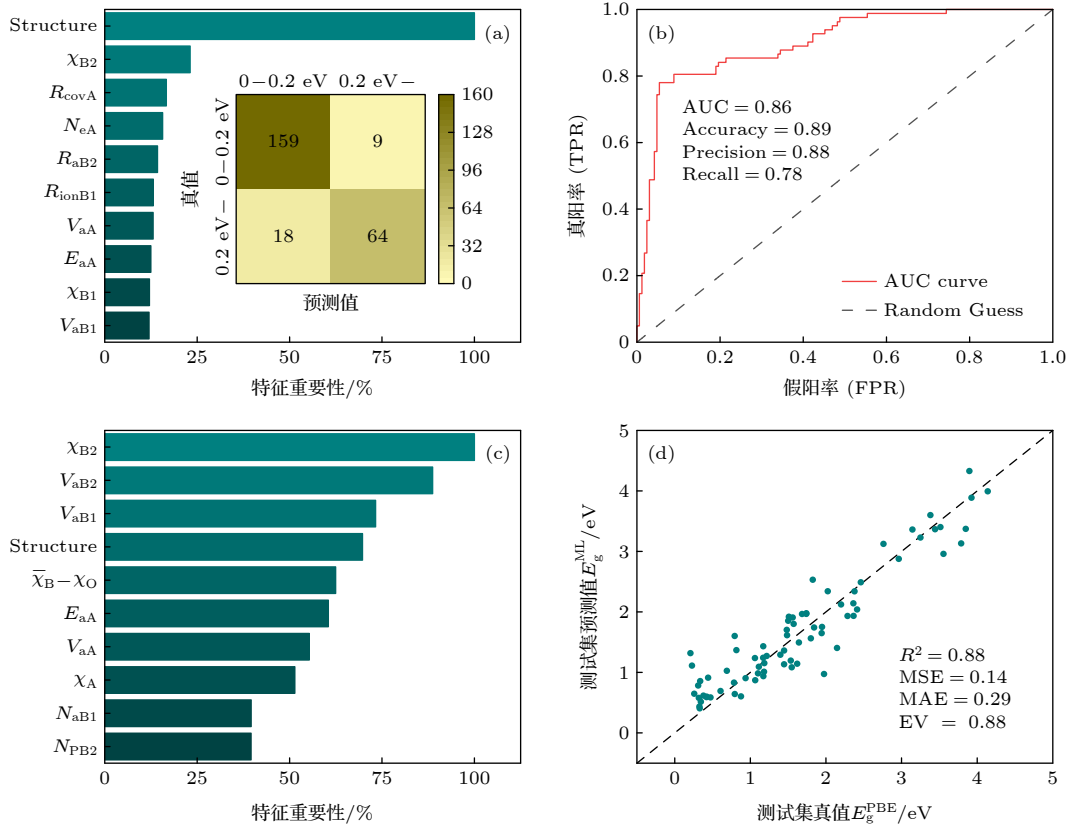


图 3 (a) 分类模型中重要性前十的特征和混淆矩阵; (b) 分类模型测试集 ROC 曲线和 AUC 值; (c) 回归模型中重要性前十的特征; (d) 回归模型测试集 R^2 , 均方误差, 平均绝对误差和解释方差

Fig. 3. (a) Relative feature importance of top 10 most important features and confusion matrix for bandgap classification; (b) receiver operating characteristic (ROC) curve for bandgap classification test set, area under the ROC curve (AUC) is provided; (c) relative feature importance of top 10 most important features for bandgap regression; (d) performance of bandgap regression model, coefficient of determination (R^2), mean square error (MSE), mean absolute error (MAE) and explained variance (EV) are provided.

第二步,为了得到准确的带隙值,继续利用 XGBoost 算法来建立带隙值的回归模型. 新的训练集和测试集是由第一步回归后带隙值大于 0.2 eV 的 757 组钙钛矿构成,包括 257 个 ABO_3 单钙钛矿和 500 个 $A_2BB'O_6$ 双钙钛矿. 从特征的重要性分析可知, B 位原子的电负性、 B 位原子体积和结构为重要的特征 (图 3(c)). 训练的回归模型在测试集上的准确度用决定系数 R^2 表示, 其值为 0.88, 均方误差 MSE 为 0.14 eV (图 3(d)). 这一结果说明此回归模型对于双钙钛矿氧化物带隙值的预测是可靠的. 最后, 利用训练好的回归模型应用到前一步得到的 14407 个双钙钛矿氧化物体系, 得到了其 PBE 带隙值. 值得注意的是, 在 PBE 计算中, 半导体氧化物的带隙, 尤其是钙钛矿氧化物的带隙, 相较实验值通常被低估了 40% 到 50%^[43,44]. 因此, 基于 PBE 数据预测的双钙钛矿氧化物材料带隙也会同样被低估. 对于可见光能量范围内 (1.6—3.1 eV) 的光催化全解水反应, 优先选择预测 PBE 带隙在 0.8—1.8 eV 的范围内的材料, 共有 7909 个, 占总体预测数据的 13.9%. 这些双钙钛矿氧化物均有可能成为光催化全水解反应的催化剂.

3.3 带隙数据的统计分析

为了进一步分析组分和结构对双钙钛矿氧化物带隙的影响, 对机器学习预测结果进行了统计分析. 众所周知, 具有八面体结构且含有 d^0 和 d^{10} 金属离子的氧化物被广泛应用于光催化全解水领域, 原因在于此类氧化物材料的导带是由金属原子的 d 轨道和 sp 轨道构成, 价带由 O 的 $2p$ 轨道构成^[45], 因而改变金属原子可以有效地调节电子结构, 进而提高催化活性. 一般地, d^0 金属离子为: Ti^{4+} , Zr^{4+} , Nb^{5+} , Ta^{5+} , Mo^{6+} 和 W^{6+} ; d^{10} 金属离子为: Ga^{3+} , In^{3+} , Sn^{4+} 和 Sb^{5+} ^[46–48]. 针对以上 d^0 与 d^{10} 金属离子, 主要关注 B 或 B' 位离子为 d^0 或 d^{10} 电子构型的双钙钛矿材料.

首先, 对比预测集中 B/B' 位为 d^0/d^{10} 金属离子与 B/B' 位离子不含 d^0/d^{10} 离子体系的带隙值是否存在区别. 如图 4(a) 所示, 带隙值越大, B/B' 位离子含 d^0/d^{10} 离子的双钙钛矿占比越大, 而 B/B' 位离子均不含 d^0/d^{10} 离子的双钙钛矿占比越小. 在 0.8—1.8 eV 范围内, B/B' 位均不含 d^0/d^{10} 离子的体系占 26%, B/B' 位中有一个为 d^0/d^{10} 离子的体系占 31%, 而 B/B' 位都是 d^0/d^{10} 离子的体系

占到 43%. 因而, 对于可见光吸收能量范围对应的 PBE 带隙值, 双钙钛矿材料中有 74% 的 B/B' 位含有 d^0/d^{10} 离子. 这进一步证实, B 或 B' 位离子为 d^0 或 d^{10} 电子构型的双钙钛矿材料, 更适合作为光催化候选材料.

进一步探究了 B/B' 位包含 d^0 或 d^{10} 离子的具体类别对回归模型预测带隙值的影响. 如图 4(b) 所示, B/B' 位如果只包含 d^{10} 离子 (蓝色), 带隙值分布的峰值所对应的带隙最小, 分布范围集中在 0.5—2.0 eV 之间. 对于 B/B' 位只包含 d^0 离子 (红色) 的双钙钛矿, 数量较少, 其带隙值要稍大于 B/B' 位只包含 d^{10} 离子的钙钛矿材料, 带隙值分布主要集中在 1.2—2.0 eV 之间. 而 B/B' 位同时包含 d^0/d^{10} 离子 (灰色) 的双钙钛矿, 数量最多, 带隙值的分布范围也最广, 涵盖了 1 eV 以上的广泛区间. 图 4(b) 中绿色区域为适合光催化的带隙范围, 从带隙分布上来看, 此区域主要包含了 B/B' 位只有 d^{10} 离子 (蓝色) 的双钙钛矿材料与 B/B' 位只有 d^0 离子 (红色) 的双钙钛矿材料. 可知 B/B' 位只有 d^{10} 离子 (蓝色) 或只有 d^0 离子 (红色) 的双钙钛矿材料更有望成为带隙合适的光催化材料.

然而上述结论并未考虑分类模型筛选的双钙钛矿材料 (带隙 0—0.2 eV). 因此, 想要在实验上指导双钙钛矿 B 与 B' 位点元素该如何选择, 还需综合分类与回归的预测结果, 整体分析带隙的分布与 B 或 B' 位所包含离子的关系. 在回归带隙统计分析的 3 种情况之外, 扩充选取 B/B' 位只有一个是 d^0 或 d^{10} 离子, 另一个非 d^0 或 d^{10} 离子的两种情形, 即 B/B' 位 5 种情况下的带隙分布 (图 4(c)). 从图 4(c) 可以看出, B/B' 位中只有一个是 d^0 或只有一个 d^{10} 离子两种情况带隙分布类似, 大部分数据集中在 0.2 eV 带隙值以下, 并不适合作为光催化候选材料. 而在 0.8—1.8 eV 的能量范围内 B/B' 位都为 d^{10} 离子的占比最大, 为 37%, 此带隙范围内 B/B' 位都为 d^0 离子占比其次, 达到 23%. 综合以上讨论, 可以得出: B/B' 位均为 d^{10} 离子的双钙钛矿氧化物材料最适合做光催化全解水的候选材料, 其次是 B/B' 位均为 d^0 离子的体系.

钙钛矿结构对称性对材料的带隙值有何种影响也是值得关心的一个问题. 从双钙钛矿材料回归模型预测的结果里挑选出 B 位为 d^0/d^{10} 离子的材料, 查看对于同样的化学式, 不同的晶型与带隙大小的对应关系. 从图 4(d) 中可以看出, 低对称性的

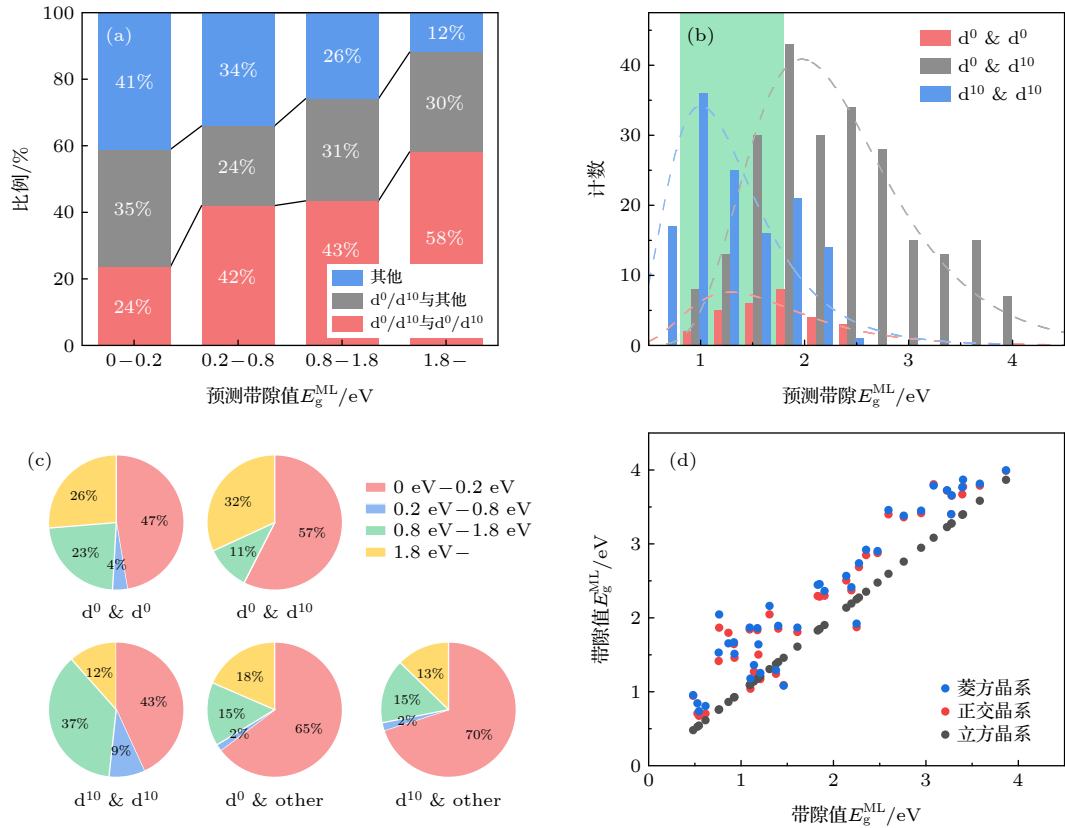


图 4 (a) 预测的钙钛矿带隙值百分比统计图, 红色区域的代表双钙钛矿的 B/B' 位点都是 d^0 或 d^{10} 金属离子, 灰色区域代表 B/B' 位点中只有一个是 d^0 或 d^{10} 金属离子, 蓝色区域代表 B/B' 位点不含 d^0 或 d^{10} 金属离子; (b) B/B' 位点都是 d^0 或 d^{10} 金属离子的双钙钛矿带隙分布图, 绿色区域为可见光能量范围; (c) 不同 B/B' 位组分下双钙钛矿带隙统计图; (d) 相同化学式下, 3 种晶系结构的双钙钛矿带隙值, 其中 B/B' 位点都是 d^0 或 d^{10} 金属离子

Fig. 4. (a) The percentage chart of predict set of bandgap values with the percentage of perovskites, red represents all B/B' sites are d^0 or d^{10} metal ions, grey represents only one of B/B' sites is d^0 or d^{10} metal ion, blue represents none of B/B' sites are d^0 or d^{10} metal ion; (b) the perovskite bandgap distribution diagram, colored area represents visible light energy range; (c) pie chart of the distribution ratio of different B/B' site ions; (d) comparison of bandgap values of 3 different structures, B/B' sites are all with d^0 or d^{10} metal ions.

菱方类结构 (蓝色) 和正交类结构 (红色) 数据点整体位于高对称性的立方类结构 (黑色), 对比图 1(a) 中结构对称性结果, 发现双钙钛矿材料的对称性越低, 其带隙值越大。

从上述统计结果可以看出, 对于适合光催化材料的寻找可以从 B/B' 为 d^{10} 的钙钛矿入手, 还可以通过替换 B/B' 位 d^{10} 离子为 d^0 离子来进一步扩大带隙值选择范围, 同时钙钛矿结构对称性也可以用来进一步调节带隙的大小, 这为实验上寻找新型全解水光催化材料提供了理论指导。

3.4 DFT 验证和全解水光催化剂预测

在讨论元素组成与构型对于双钙钛矿带隙的影响后, 还需要对预测的结果进行进一步验证, 并通过计算电子能带结构, 获得带边位置满足光催化全解水的双钙钛矿材料。选用预测集中 B/B' 位为

d^0/d^{10} 金属离子且结构稳定 (容忍因子 $t > 0.85$) 的菱方类结构材料进行 DFT 验证。对结构保持稳定未发生畸变的 29 种材料计算了能带结构, DFT 所得带隙与机器学习预测带隙符合得较好, $R^2 > 0.8$ (见图 5(a))。

对上述 29 种材料 PBE 带隙分析后, 选出带隙值处在 0.8—1.8 eV 之间, 且不含铅、汞离子的 $\text{Sr}_2\text{GaSbO}_6$, $\text{Sr}_2\text{InSbO}_6$ 和 K_2NbTaO_6 这 3 种双钙钛矿结构进行 HSE 能带的计算。从表 1 中可以看出, 这 3 种双钙钛矿候选材料的 HSE 带隙都在可见光范围内, 且 HSE 带隙比其 PBE 带隙值要大 40%—50%。其中 $\text{Sr}_2\text{GaSbO}_6$ 和 $\text{Sr}_2\text{InSbO}_6$ 为间接带隙半导体材料, 而 K_2NbTaO_6 为直接带隙半导体材料。一般来说, 具有直接带隙的催化剂通常具有较高的电子空穴传输能力, 它们能更有效地进行光子收集^[49]。因此, 从带隙类型来看 K_2NbTaO_6 双

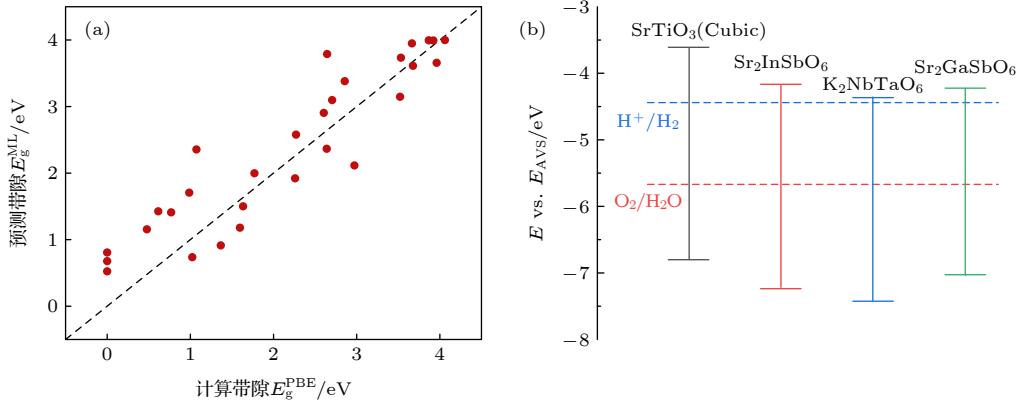


图 5 (a) 29 种菱方类结构双钙钛矿材料 DFT 带隙与机器学习预测带隙的比较; (b) 3 种候选双钙钛矿相对于水的氧化还原势的 HSE 带边位置, 以及作为比较基准的 SrTiO₃ (立方相) 带边位置

Fig. 5. (a) DFT bandgap verification of 29 rhombohedral double perovskites in the prediction set; (b) the HSE band edge positions with respect to the water reduction and oxidation potential levels of selected double perovskites. SrTiO₃ (cubic) is listed as a benchmark.

钙钛矿材料相较于另外两种材料更适合作为光催化剂. 另一方面, 要实现可见光范围内的光催化全解水, 除了其带隙值需要在可见光能量范围内, 半导体材料的价带顶还要低于氧化电势 ($V_{H_2O/O_2} = -5.67$ eV), 导带底也要高于还原电势 ($V_{H_2/H^+} = -4.44$ eV). 参照立方相 SrTiO₃ 的能带结构, 将 3 种双钙钛矿 O 的深能级与前者的对齐^[50–52], 可以得到 K₂NbTaO₆, Sr₂InSbO₆ 和 Sr₂GaSbO₆ 相对于水的氧化还原势的带边位置, 如图 5(b) 所示. 此 3 种双钙钛矿氧化物带隙值在 3 eV 左右, 对应可见光吸收范围虽然较小, 但这 3 种材料的带边位置均满足光催化全解水的带边位置要求, 合成材料价格低廉, 不含重金属元素. 因此, 综合考虑了以上几点, 它们很有可能是好的光催化剂材料.

表 1 3 种候选双钙钛矿材料的 PBE 带隙、HSE 带隙及带隙类型

Table 1. PBE and HSE bandgap of three kinds of double perovskite candidates and their bandgap categories.

Formula	E_{g-PBE} /eV	E_{g-HSE} /eV	Bandgap
Sr ₂ GaSbO ₆	1.37	2.80	indirect
Sr ₂ InSbO ₆	1.63	3.07	indirect
K ₂ NbTaO ₆	1.77	3.06	direct

4 总结

本文通过结合机器学习方法与第一性原理计算, 设计了一套快速筛选功能材料框架, 成功从 56894 种候选材料中筛选出 7909 种带隙符合光催

化全解水需求的双钙钛矿氧化物材料. 机器学习分类和回归的模型的准确率均接近 90%. 此外, 对预测结果的统计分析表明, 从 B/B' 位选择含有 d¹⁰ 金属离子的双钙钛矿材料, 更有可能找到符合光催化全解水带隙要求的材料, 同时, 将 B/B' 位的 d¹⁰ 金属离子替换为 d⁰ 金属离子可以调节带隙大小. 最后, 通过比对 HSE 带边位置和全解水所需的氧化还原电势位置, 挑选出 Sr₂GaSbO₆, Sr₂InSbO₆ 和 K₂NbTaO₆ 三种适合作为全解水光催化剂的双钙钛矿氧化物候选材料. 本工作作为快速、准确筛选设计双钙钛矿光催化剂提供了新的思路.

感谢东南大学大数据中心和天津国家超级计算中心的计算资源.

附录

表 A1 过渡金属计算时附加的 Hubbard U 值
Table A1. Hubbard U value for the transition metal elements.

元素	U 值
V	3.25
Mo	4.38
W	6.2
Ni	6.2
Mn	3.9
Fe	5.3
Cr	3.7
Co	3.32

表 A2 特征符号及含义

Table A2. The symbol of features and their corresponding meanings.

符号	含义
N_a	原子数
R_a	原子半径
V_a	原子体积
R_{cov}	共价半径
E_a	电子亲合能
N_e	电子数
χ	鲍林电负性
H_f	形成热
N_m	门捷列夫数
N_p	周期
R_{ion}	离子半径
t	容忍因子

参考文献

- [1] Dorian J P, Franssen H T, Simbeck D R 2006 *Energy Policy* **34** 1984
- [2] Omer A M 2008 *Renew. Sust. Energ. Rev.* **12** 2265
- [3] Pfenninger S, Hawkes A, Keirstead J 2014 *Renew. Sust. Energ. Rev.* **33** 74
- [4] Liang K, Huang T, Yang K, Si Y, Wu H Y, Lian J C, Huang W Q, Hu W Y, Huang G F 2021 *Phys. Rev. Appl.* **16** 054043
- [5] Ameen S, Rub M A, Kosa S A, Alamry K A, Akhtar M S, Shin H S, Seo H K, Asiri A M, Nazeeruddin M K 2016 *ChemSusChem* **9** 10
- [6] Chen S, Takata T, Domen K 2017 *Nat. Rev. Mater.* **2** 17050
- [7] Hisatomi T, Kubota J, Domen K 2014 *Chem. Soc. Rev.* **43** 7520
- [8] Maeda K, Domen K 2010 *J. Phys. Chem. Lett.* **1** 2655
- [9] Kumar A, Kumar A, Krishnan V 2020 *ACS Catal.* **10** 10253
- [10] Peña M A, Fierro J L G 2001 *Chem. Rev.* **101** 1981
- [11] Ouyang Y, Li Y, Zhu P, Li Q, Gao Y, Tong J, Shi L, Zhou Q, Ling C, Chen Q, Deng Z, Tan H, Deng W, Wang J 2019 *J. Mater. Chem. A* **7** 2275
- [12] Grimaud A, May K J, Carlton C E, Lee Y L, Risch M, Hong W T, Zhou J, Shao-Horn Y 2013 *Nat. Commun.* **4** 2439
- [13] Yin W, Weng B, Ge J, Sun Q, Li Z, Yan Y 2019 *Energy Environ. Sci.* **12** 442
- [14] Sun H, Xu X, Song Y, Zhou W, Shao Z 2021 *Adv. Funct. Mater.* **31** 2009779
- [15] Aczel A A, Bugaris D E, Li L, Yan J, de la Cruz C, zur Loye H C, Nagler S E 2013 *Phys. Rev. B* **87** 014435
- [16] Zhou Q, Lu S, Wu Y, Wang J 2020 *J. Phys. Chem. Lett.* **11** 3920
- [17] Lu S, Zhou Q, Guo Y, Wang J 2022 *Chem* **8** 769
- [18] Lu S, Zhou Q, Guo Y, Zhang Y, Wu Y, Wang J 2020 *Adv. Mater.* **32** 2002658
- [19] Wu Y, Lu S, Ju M, Zhou Q, Wang J 2021 *Nanoscale* **13** 12250
- [20] Goldsmith B R, Esterhuizen J, Liu J, Bartel C J, Sutton C 2018 *AIChE J.* **64** 2311
- [21] Chen T, Guestrin C 2016 *XGBoost: A Scalable Tree Boosting System* (Association for Computing Machinery) pp785–794
- [22] Natekin A, Knoll A 2013 *Front. Neurobot.* **7**
- [23] Hafner J 2008 *J. Comput. Chem.* **29** 2044
- [24] Perdew J P, Burke K, Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865
- [25] Cai B, Chen X, Xie M, Zhang S, Liu X, Yang J, Zhou W, Guo S, Zeng H 2018 *Mater. Horiz.* **5** 961
- [26] Blöchl P E 1994 *Phys. Rev. B* **50** 17953
- [27] Monkhorst H J, Pack J D 1976 *Phys. Rev. B* **13** 5188
- [28] Aryasetiawan F, Karlsson K, Jepsen O, Schönberger U 2006 *Phys. Rev. B* **74** 125106
- [29] Becke A D 1993 *J. Chem. Phys.* **98** 1372
- [30] Curtarolo S, Setyawan W, Hart G L W, Jahnatek M, Chepulskii R V, Taylor R H, Wang S, Xue J, Yang K, Levy O, Mehl M J, Stokes H T, Demchenko D O, Morgan D 2012 *Com. Mat. Sci.* **58** 218
- [31] Saal J E, Kirklin S, Aykol M, Meredig B, Wolverton C 2013 *JOM* **65** 1501
- [32] Jain A, Ong S P, Hautier G, Chen W, Richards W D, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson K A 2013 *APL Mater.* **1** 011002
- [33] Goldschmidt V M 1926 *Naturwissenschaften* **14** 477
- [34] Sun Q, Yin W 2017 *J. Am. Chem. Soc.* **139** 14905
- [35] Bartel C J, Sutton C, Goldsmith B R, Ouyang R, Musgrave C B, Ghiringhelli L M, Scheffler M 2019 *Sci. Adv.* **5** eaav0693
- [36] Weng B, Song Z, Zhu R, Yan Q, Sun Q, Grice C G, Yan Y, Yin W 2020 *Nat. Commun.* **11** 3513
- [37] Filip-Marina R, Giustino F 2018 *Proc. Natl. Acad. Sci. U. S. A.* **115** 5397
- [38] Ye W, Chen C, Dwaraknath S, Jain A, Ong S P, Persson K A 2018 *MRS Bull.* **43** 664
- [39] Zhao X G, Yang J H, Fu Y, Yang D, Xu Q, Yu L, Wei S H, Zhang L 2017 *J. Am. Chem. Soc.* **139** 2630
- [40] Lu S, Zhou Q, Ma L, Guo Y, Wang J 2019 *Small Methods* **3** 1900360
- [41] Goodenough J B 2004 *Rep. Prog. Phys.* **67** 1915
- [42] Okada S, Ohzeki M, Taguchi S 2019 *Sci. Rep.* **9** 13036
- [43] Wahl R, Vogtenhuber D, Kresse G 2008 *Phys. Rev. B* **78** 104116
- [44] Liu P, Nisar J, Pathak B, Ahuja R 2012 *Int. J. Hydrogen Energy* **37** 11611
- [45] Chou H, Hwang B, Sun C 2013 *New and Future Developments in Catalysis* (Amsterdam: Elsevier) pp217–270
- [46] Inoue Y 2009 *Energy Environ. Sci.* **2** 364
- [47] Kudo A, Hijii S 1999 *Chem. Lett.* **28** 1103
- [48] Kudo A, Miseki Y 2009 *Chem. Soc. Rev.* **38** 253
- [49] Acar C, Dincer I, Naterer G F 2016 *Int. J. Energy Res.* **40** 1449
- [50] Kaspar T C, Sushko P V, Spurgeon S R, Bowden M E, Keavney D J, Comes R B, Saremi S, Martin L, Chambers S A 2019 *Adv. Mater. Interfaces* **6** 1801428
- [51] Greiner M T, Helander M G, Tang W, Wang Z B, Qiu J, Lu Z 2012 *Nat. Mater.* **11** 76
- [52] El-Sayed A, Borghetti P, Goiri E, Rogero C, Floreano L, Lovat G, Mowbray D J, Cabellos J L, Wakayama Y, Rubio A, Ortega J E, de Oteyza D G 2013 *ACS Nano* **7** 6914

Machine learning accelerated search for new double perovskite oxide photocatalysis*

Wan Xin-Yang Zhang Ye-Hui Lu Shuai-Hua Wu Yi-Lei
Zhou Qiong-Hua[†] Wang Jin-Lan[‡]

(School of Physics, Southeast University, Nanjing 211189, China)

(Received 1 April 2022; revised manuscript received 11 May 2022)

Abstract

Double perovskite oxide $A_2BB'O_6$ has better stability and wider bandgap range than ABO_3 -type oxide, and exhibits great prospects in photocatalytic overall water splitting. However, owing to the diversity of crystal structure and constituents of perovskite oxide, rapidly and accurately searching for $A_2BB'O_6$ for photocatalyst is still a big challenge, both experimentally and theoretically. In this work, in order to screen out suitable double perovskite oxide photocatalysts, a multi-step framework combined with machine learning technique and first-principles calculations is proposed. Nearly 8000 candidates with proper bandgaps for water splitting are screened out from among more than 50000 $A_2BB'O_6$ -type double perovskite oxides. Statistical analysis of the results shows that double perovskite oxides with d^{10} metal ions at B/B' sites are more likely to have good absorption of visible light, and the structural symmetry of double perovskite also has influence on the bandgap value. Furthermore, first-principles calculations demonstrate that Sr_2GaSbO_6 , Sr_2InSbO_6 and K_2NbTaO_6 are non-toxic photocatalyst candidates with proper band edges for overall water splitting.

Keywords: machine learning, double perovskite oxides, photocatalysis, overall water splitting, first-principle method

PACS: 71.15.Mb, 71.20.-b, 71.20.Nr, 81.16.Hc

DOI: 10.7498/aps.71.20220601

* Project supported by the National Key Research and Development Program of China (Grant Nos. 2017YFA0204800, 2021YFA1500700), the Natural Science Foundation of China (Grant Nos. 22033002, 22003009), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. KYCX20_0075).

[†] Corresponding author. E-mail: qh.zhou@seu.edu.cn

[‡] Corresponding author. E-mail: jlwang@seu.edu.cn

机器学习加速搜寻新型双钙钛矿氧化物光催化剂

万新阳 章烨辉 陆帅华 吴艺蕾 周颢桦 王金兰

Machine learning accelerated search for new double perovskite oxide photocatalysis

Wan Xin-Yang Zhang Ye-Hui Lu Shuai-Hua Wu Yi-Lei Zhou Qiong-Hua Wang Jin-Lan

引用信息 Citation: *Acta Physica Sinica*, 71, 177101 (2022) DOI: 10.7498/aps.71.20220601

在线阅读 View online: <https://doi.org/10.7498/aps.71.20220601>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

铅基钙钛矿铁电晶体高临界转变温度的机器学习研究

High critical transition temperature of lead-based perovskite ferroelectric crystals: A machine learning study

物理学报. 2019, 68(21): 210502 <https://doi.org/10.7498/aps.68.20190942>

基于波动与扩散物理系统的机器学习

Machine learning based on wave and diffusion physical systems

物理学报. 2021, 70(14): 144204 <https://doi.org/10.7498/aps.70.20210879>

机器学习辅助绝热量子算法设计

Machine learning assisted quantum adiabatic algorithm design

物理学报. 2021, 70(14): 140306 <https://doi.org/10.7498/aps.70.20210831>

钙钛矿铁电半导体的光催化研究现状及其展望

Research progress and prospects of photocatalytic devices with perovskite ferroelectric semiconductors

物理学报. 2020, 69(12): 127706 <https://doi.org/10.7498/aps.69.20200287>

基于机器学习的非线性局部Lyapunov向量集合预报订正

Machine learning based method of correcting nonlinear local Lyapunov vectors ensemble forecasting

物理学报. 2022, 71(8): 080503 <https://doi.org/10.7498/aps.71.20212260>

基于机器学习 J_1 - J_2 反铁磁海森伯自旋链相变点的识别方法

Identifying phase transition point of J_1 - J_2 antiferromagnetic Heisenberg spin chain by machine learning

物理学报. 2021, 70(23): 230701 <https://doi.org/10.7498/aps.70.20210711>