



基于机器学习的无机磁性材料磁性基态分类与磁矩预测

黎威 龙连春 刘静毅 杨洋

Classification of magnetic ground states and prediction of magnetic moments of inorganic magnetic materials based on machine learning

Li Wei Long Lian-Chun Liu Jing-Yi Yang Yang

引用信息 Citation: *Acta Physica Sinica*, 71, 060202 (2022) DOI: 10.7498/aps.71.20211625

在线阅读 View online: <https://doi.org/10.7498/aps.71.20211625>

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

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>

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>

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>

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>

基于机器学习和器件模拟对Cu(In,Ga)Se₂电池中Ga含量梯度的优化分析

Optimization of Ga content gradient in Cu(In,Ga)Se₂ solar cells through machine learning and device simulation

物理学报. 2021, 70(23): 238802 <https://doi.org/10.7498/aps.70.20211234>

基于深度学习的光学表面杂质检测

Deep-learning-assisted micro impurity detection on an optical surface

物理学报. 2021, 70(16): 168702 <https://doi.org/10.7498/aps.70.20210403>

基于机器学习的无机磁性材料磁性 基态分类与磁矩预测^{*}

黎威¹⁾ 龙连春^{1)†} 刘静毅¹⁾ 杨洋²⁾

1) (北京工业大学材料与制造学部, 北京 100124)

2) (中国科学院物理研究所, 北京 100190)

(2021 年 9 月 1 日收到; 2021 年 11 月 22 日收到修改稿)

磁性材料是信息时代重要的基础材料, 不同的磁性基态是磁性材料广泛应用的前提, 其中铁磁基态是高性能磁性材料的关键要求. 本文针对材料项目数据库中的无机磁性材料数据, 采用机器学习技术实现无机磁性材料铁磁、反铁磁、亚铁磁和顺磁基态的分类以及无机铁磁性材料磁矩的预测. 提取了材料的元素和结构属性特征, 通过两步式特征选择方法分别为磁性基态分类和磁矩预测筛选了 20 个材料特征, 发现材料特征中的电负性、原子磁矩和原子外围轨道未充满电子数对两种磁性性能具有重要贡献. 基于机器学习的随机森林算法, 构建了磁性基态分类模型和磁矩预测模型, 采用 10 折交叉验证的方法对模型进行定量评估, 结果表明所构建的模型具有足够的精度和泛化能力. 在测试检验中, 磁性基态分类模型的准确率为 85.23%, 精确率为 85.18%, 召回率为 85.04%, F1 分数为 85.24%; 磁矩预测模型的拟合优度为 91.58%, 平均绝对误差为 $0.098 \mu_B/\text{atom}$. 本研究为无机铁磁性材料的高通量分类筛选与磁矩预测提供了新的方法和选择, 可为新型无机磁性材料的设计研发提供参考.

关键词: 机器学习, 随机森林, 磁性基态, 磁矩

PACS: 02.60.Pn, 75.50.Gg, 32.10.Dk

DOI: 10.7498/aps.71.20211625

1 引言

磁性材料种类繁多、用途广泛, 在磁记录、磁致伸缩、磁润滑、磁致冷和磁传感器等领域具有重要应用^[1-4]. 按照磁性物理学的划分, 材料的磁性基态可分为磁有序的铁磁基态、反铁磁基态、亚铁磁基态和磁无序的顺磁基态. 其中, 铁磁基态是未来高性能磁性材料的关键要求之一^[5], 根据铁磁基态产生的铁磁性材料在传感与控制、信息处理、信息存储等技术领域具有重要的应用价值^[6-8]. 从宏观角度来说, 长程有序排列的磁矩是材料铁磁性产生的原因^[9], 通过磁矩可以得到材料的磁化强度、

磁滞回线, 反映材料的磁化性能^[10]. 因此, 研究铁磁性材料的磁矩十分重要.

由于材料的实验测试既昂贵又耗时^[11], 材料科学家们在实验制备测试材料的同时, 使用密度泛函理论 (density functional theory, DFT) 进行材料性能计算. 然而, 对于复杂的材料结构, 精确的 DFT 计算也会消耗大量时间. 多年来, 通过 DFT 计算产生了一些大型材料数据库, 如无机晶体结构数据库^[12]、开放量子材料数据库^[13]和材料项目数据库^[14]等. 面对海量材料数据, 如何通过数据技术手段从中提取知识, 避免重复运算, 浪费算力, 成为关注的热点^[15]. 近年来, 许多研究者将机器学习 (machine learning, ML) 技术应用于材料研究,

^{*} 国家重点研发计划 (批准号: 2018YFB0703500) 资助的课题.

[†] 通信作者. E-mail: longlc@bjut.edu.cn

主要分为预测材料性能和发现新材料两个方面^[16]. 在预测材料性能方面, Isayev 等^[17]构建了一种材料特征计算工具, 使用 ML 中的梯度提升决策树算法预测了无机晶体材料的六种热机械性能; 寇雯博等^[18]提出了一种小波-ML 混合方法, 可用于混杂复合材料等效热传导性能的预测. 在发现新材料方面, 杨自欣等^[19]构建了用于预测铅基钙钛矿材料铁电居里温度的 ML 模型, 使用得到的模型筛选出了 2 种目标钙钛矿铁电材料; Lu 等^[20]采用 DFT 计算结合 ML 预测的方式, 从 5158 个无铅杂化钙钛矿中筛选了 3 个稳定的钙钛矿材料. 这些研究展现了 ML 在计算材料领域的可行性.

目前, 使用 ML 研究材料磁性基态主要关注铁磁和反铁磁基态的分类, 多种磁性基态分类的 ML 研究有待发展. 例如, Long 等^[5]从 AtomWork 数据库^[21]中收集了 1749 条铁磁基态和 1056 条反铁磁基态的金属间化合物数据, 使用 ML 中的随机森林算法实现了金属间化合物的铁磁/反铁磁基态分类, 最终的分类准确率为 87%; Frey 等^[22]针对材料项目数据库中的 3153 种过渡金属氧化物训练了 ML 模型, 实现了过渡金属氧化物的铁磁/反铁磁基态分类, 分类模型的精确率和召回率的调和平均数 (F1 score) 为 85%. 虽然已有研究者使用 ML 预测材料磁矩, 但面向的是多种磁性基态材料, 相关研究主要关注算法的准确性, 并没有分析材料特征, 缺少探讨单一磁性基态材料的磁矩. 铁磁性材料作为一种重要的磁性材料, 针对性的机器学习研究具有重要意义. 如 Yamamoto^[23]使用晶体图神经网络模型对开放量子材料数据库和材料项目数据库中的材料数据进行研究, 实现了非磁性材料/磁性材料的分类和磁矩的预测, 最终集成分类模型的感受性曲线下方面积 (AUC) 为 0.957, 集成预测模型的平均绝对误差为 $0.119 \mu_B/\text{atom}$.

本研究以无机磁性材料的四种磁性基态和无机铁磁性材料的磁矩为研究对象, 计算了材料的元素和结构特征, 并通过两步法筛选材料特征, 确定了磁性基态分类与磁矩预测的重要材料属性. 基于筛选后的材料特征, 训练了磁性基态分类和磁矩预测的 ML 模型, 使用多种评价指标对模型的 10 折交叉验证结果进行定量评估, 得到了性能最优的分类模型和预测模型. 最后, 将两个模型应用于测试集中材料的磁性基态分类和磁矩预测, 验证了模型的精度和泛化能力.

2 方法

2.1 数据集建立

通过材料项目数据库的应用程序接口^[24]获取了 98888 条无机磁性材料数据, 包含材料的化学式、CIF 文件、磁性基态和晶胞磁矩信息. 针对本文的研究对象, 对数据集中无机磁性材料的磁性基态和无机铁磁性材料的晶胞磁矩进行描述性统计, 如图 1 所示. 图 1(a) 是四种磁性基态的数量分布, 其中顺磁基态 (PM) 数据 58068 条, 铁磁基态 (FM) 数据 32248 条, 亚铁磁基态 (FiM) 数据 5733 条, 反铁磁基态 (AFM) 数据 2839 条. 图 1(b) 展示了无机铁磁性材料晶胞磁矩的密度分布, 范围为 0— $280 \mu_B$, 可以看到数据呈重尾分布, 类似于指数数据.

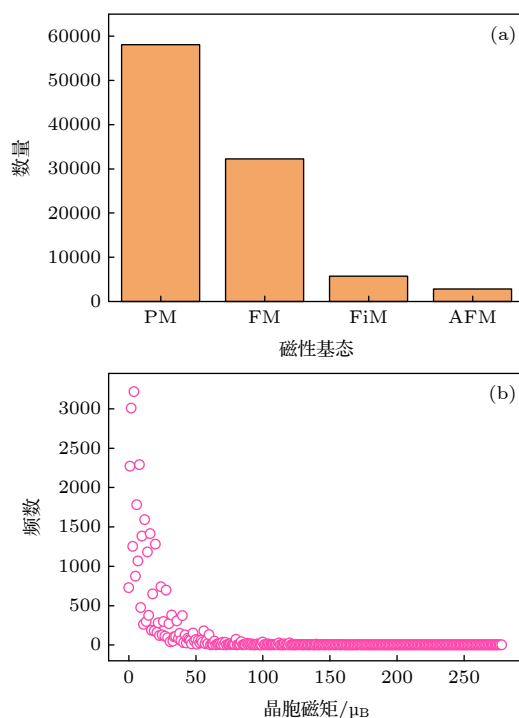


图 1 材料数据集的描述性统计 (a) 磁性基态分布直方图; (b) 晶胞磁矩频数分布图

Fig. 1. Descriptive statistics of material data set: (a) Distribution histogram of the magnetic ground state; (b) frequency distribution of the unit cell magnetic moment.

2.2 特征构建与特征选择

通过 ML 技术建立材料属性与材料性能之间的映射, 需要提取数字化的材料特征变量. 利用材料信息平台 Matminer^[25]计算了数据集中材料的元素和结构特征, 共产生了 582 种材料特征. 其中

元素特征主要包括材料的组成、元素信息和电子排布; 结构特征主要包括空间群数、晶胞体积和正弦库仑矩阵.

冗余特征会导致模型的低效或过度拟合, 为了克服此缺陷, 采用两步式特征选择方法去除冗余特征. 第 1 步使用交叉验证递归特征消除 (RFECV) 的方法, 逐一评估特征的重要性, 在保证模型精度基本不下降的前提下, 去除冗余特征; 第 2 步采用 ML 模型的特征排序, 进一步精简并选择出对模型最重要的特征. 通过两步式特征选择方法分别为磁性基态分类模型和磁矩预测模型选择了 20 个特征, 如表 1 所列. 表中展示了选择出的 34 个重要特征, 其中有 6 个特征为两个模型共用特征, 各个特征的物理意义见附录表 A1 所列.

2.3 机器学习模型构建

将选择出的材料特征数据与磁性性能数据组合为材料数据集, 按照 4:1 的比例随机分为训练集和测试集. 首先使用训练集中的材料数据进行 ML 模型的训练, 然后使用测试集中的材料数据检验 ML 模型的精度和泛化能力. ML 模型的训练和检验过程均采用 10 折交叉验证的方法进行定量评

估. 其中, 分类模型的评估指标为: 准确率 (accuracy)、精确率 (precision)、召回率 (recall) 和 F1 分数 (F1 score); 预测模型的评估指标为: 拟合优度 (R^2) 和平均绝对误差 (MAE).

为了捕获材料特征数据与材料磁性性能之间复杂的映射关系, 采用随机森林 (random forest, RF)^[26] 作为 ML 模型. RF 不需要对材料特征进行缩放, 能够直观地得出材料特征与材料性能之间的关系. ML 模型的超参数会影响模型的拟合能力, 通过网格搜索的方式分别对随机森林分类 (RFC) 模型和随机森林回归 (RFR) 模型进行超参数优化, 得到模型的超参数如表 2 所列. 本研究中 ML 模型的训练、评估和超参数优化均通过 python 库中的 scikit-learn^[27] 实现, ML 模型的构建流程如图 2.

3 结果与讨论

3.1 模型训练结果

3.1.1 分类模型训练

使用 RF 中的随机森林分类 (RFC) 构建了磁性基态分类模型, 完成了无机磁性材料铁磁、反铁磁、亚铁磁和顺磁基态的分类筛选. 在磁性基态分

表 1 基于两步式特征选择法获得的材料特征
Table 1. Material features obtained by the two-step feature selection method.

特征类型	特征		
元素	Mode Electronegativity*	Mean NdUnfilled*	Max MeltingT
	Min Electronegativity	Avg_dev NdUnfilled*	Mode Number
	Range NUnfilled	Max NdUnfilled	Max Number
	Avg_dev NUnfilled	Mean GSmagmom*	Min NValence
	Max NUnfilled	Range GSmagmom	Range NfValence
	Mode NfUnfilled	Avg_dev GSmagmom*	Avg_dev NfValence
	Mean NfUnfilled	Max GSmagmom	Avg_dev NdValence
	Range NfUnfilled*	Max AtomicWeight	Mode MendeleevNumber
	Avg_dev NfUnfilled	Mode AtomicWeight	Avg_dev MendeleevNumber
	Max NfUnfilled	Mean GSvolume_pa	Min MendeleevNumber
	Range NdUnfilled	Range MeltingT	
结构	Vpa	Sine coulomb matrix 0	

* 该特征同时用于磁性基态分类和磁矩预测.

表 2 本研究中机器学习模型的超参数
Table 2. Hyperparameters of the machine learning model in this study.

模型	超参数
RFC	$n = 400$, features = 'log2', samples_split = 2, samples_leaf = 1
RFR	$n = 300$, features = 'auto', samples_split = 2, samples_leaf = 1

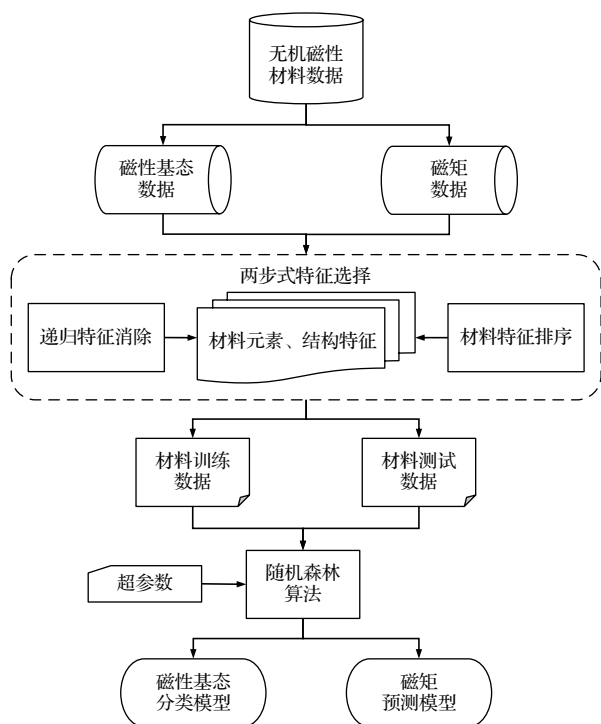


图2 机器学习模型的构建流程

Fig. 2. Construction process of the machine learning model.

类模型的两步式特征选择中,第1步通过 RFECV 方法将材料特征减少到了 82 个,模型的平均分类准确率为 89.46%;第2步通过 RFC 的特征排序选择了排名靠前的 20 个材料特征对模型进行训练,此时模型的平均分类准确率为 87.67%,如图 3(a) 所示.考虑到特征相关性,第2步被剔除的材料特征对磁性基态的分类仍有少量贡献,模型分类准确率的略微下降在可控且合理的范围内.图 3(b) 展示的是分类模型的材料特征排序结果,其中,材料原子磁矩、原子外围轨道未充满电子数等元素特征对磁性基态分类的贡献程度较大;结构特征中的正弦库仑矩阵、晶胞体积对磁性基态分类的贡献程度

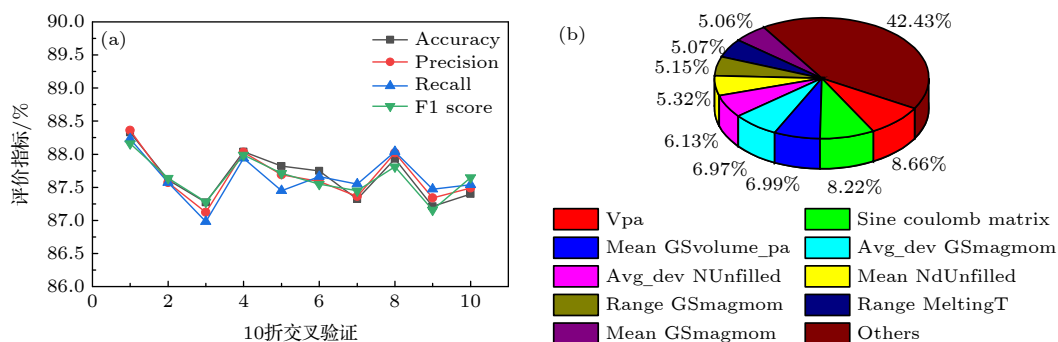


图3 磁性基态分类模型的训练结果 (a) 10折交叉验证; (b) 材料特征排序

Fig. 3. Training results of the magnetic ground state classification model: (a) 10-fold cross-validation; (b) ranking of material features.

虽小,但是排名靠前,可见结构特征在四种磁性基态分类时起到了关键作用.

3.1.2 预测模型训练

基于获得的分类模型,使用 RF 中的随机森林回归 (RFR) 对无机铁磁性材料的磁矩进行预测分析.为避免晶胞大小不同和晶胞中原子个数不同对结果产生影响,将磁矩值平均到晶胞中的每个原子.预测模型训练与分类模型训练相似,采用两步法去除了对磁矩预测无关或影响较小的冗余材料特征.第1步通过 RFECV 得到了 45 个材料特征,模型的平均拟合优度为 95.77%;第2步通过 RFR 的特征排序得到了 20 个材料特征,此时模型的平均拟合优度为 94.68%,如图 4(a) 所示.图 4(b) 展示的是预测模型的材料特征排序,结构特征在经过两步式的特征选择后被移除,而化合价、电负性、f 轨道未充满电子数和材料原子磁矩等元素材料特征对无机铁磁性材料的磁矩预测提供了较大的贡献.

3.2 模型检验结果

将材料数据集分为训练集和测试集,模型的训练和交叉验证均在训练集上完成,测试集数据在整个模型构建过程中没有参与,此时模型在训练集上的应用效果最能检验模型的精度和泛化能力.

3.2.1 分类检验

针对训练完成的磁性基态分类模型,使用测试集检验其对无机磁性材料四种磁性基态的分类能力.图 5(a) 是混淆矩阵,它反映了磁性基态分类模型在测试集上的精度.图 5(a) 中的数值表示行标签材料被预测为列标签材料的数目,可以看到 96.5% 的顺磁基态和 84.8% 的铁磁基态被正确分类,反铁磁和亚铁磁基态的分类效果较差,这可能是由于

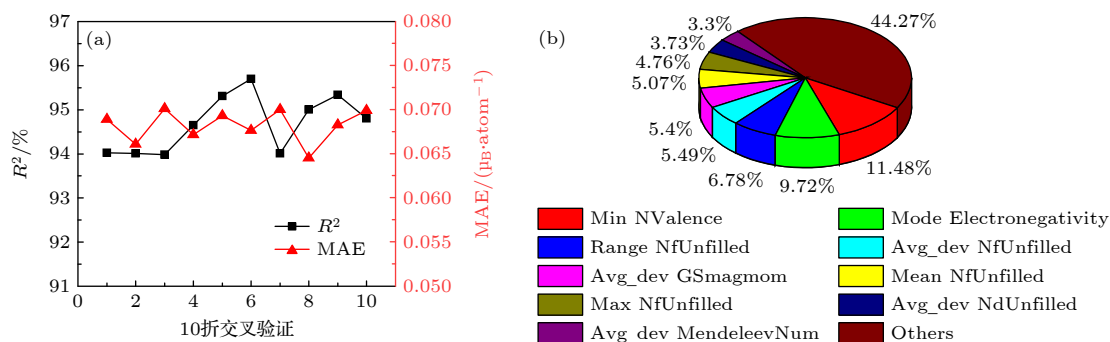


图 4 磁矩预测模型的训练结果 (a) 10 折交叉验证; (b) 材料特征排序

Fig. 4. Training results of the magnetic moment prediction model: (a) 10-fold cross-validation; (b) ranking of material features.

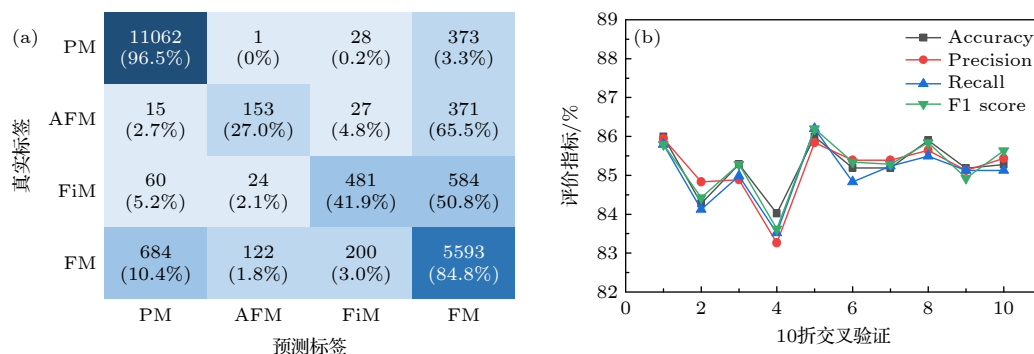


图 5 磁性基态分类模型的检验结果 (a) 混淆矩阵; (b) 10 折交叉验证

Fig. 5. Test results of the magnetic ground state classification model: (a) Confusion matrix; (b) 10-fold cross-validation.

数据库中反铁磁基态和亚铁磁基态的材料数量较少, 数据分布不平衡导致. 从图 5(b) 的结果可以看出, 4 种分类评价指标的 10 折交叉验证结果都很平均, 说明模型没有过拟合和偏向抽样, 能够真实地反映磁性基态分类模型对四种磁性基态的分类效果. 同时, 将本研究的磁性基态分类模型与其他研究磁性基态分类的机器学习模型进行了对比, 如表 3 所列. 这两个研究实现了铁磁和反铁磁基态的分类, 属于二分类问题. 本研究实现了铁磁、反铁磁、亚铁磁和顺磁基态的分类, 属于四分类问题, 而且进一步提升了分类效果.

表 3 本研究磁性基态分类模型与其他研究者工作的定量评估对比

Table 3. Quantitative evaluation of the magnetic ground state classification model in this study and in comparison with other works.

评价指标(平均值)	本研究模型	文献[5]	文献[22]
Accuracy/%	85.23	81.10	—
Precision/%	85.18	84.29	—
Recall/%	85.04	85.51	—
F1 score/%	85.24	85.08	85.00

3.2.2 预测检验

为了检验磁矩预测模型对无机铁磁性材料磁矩的预测能力, 使用测试集中的磁矩数据进行预测, 观察预测值与真实值的拟合程度、误差范围. 图 6(a) 展示了预测模型对磁矩的预测值和真实值对比, 图中蓝色圆圈对应不同的无机铁磁性材料, 其横坐标为获得的预测模型对某材料磁矩的预测值, 纵坐标为该材料磁矩的真实值, 红色虚线表示预测值与真实值完全吻合时的情况. 从图 6(a) 可以看出, 对于测试集中 6450 条未参与训练的无机铁磁性材料, 预测模型对其磁矩的预测值都落在了真实值附近. 从图 6(b) 可以看出, 2 种评价指标的 10 折交叉验证结果都很平均, 较训练集结果仅有略微下降, 说明磁矩预测模型没有过拟合, 对无机铁磁性材料的磁矩具有较好的预测能力. 同时, 也将本研究的磁矩预测模型与其他研究磁矩预测的机器学习模型进行了对比, 如表 4 所列, 本研究的机器学习模型对磁矩的预测具有更低的平均绝对误差 (MAE).

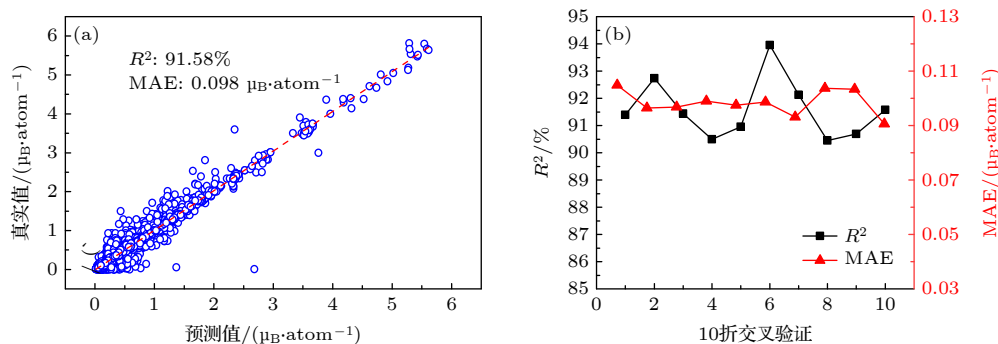


图 6 磁矩预测模型的检验结果 (a) 预测值与真实值的拟合情况; (b) 10 折交叉验证

Fig. 6. Test results of the magnetic moment prediction model: (a) Fitting degree between predicted value and real value; (b) 10-fold cross validation.

表 4 本研究磁矩预测模型与其他研究者工作的定量评估对比

Table 4. Quantitative evaluation of the magnetic moment prediction model in this study and in comparison with other works.

评价指标(平均值)	本研究模型	文献[23]
$R^2/\%$	91.58	—
$MAE/(\mu_B \cdot atom^{-1})$	0.098	0.119

4 结 论

本研究针对材料项目数据库中无机磁性材料的四种磁性基态和无机铁磁性材料的磁矩, 计算了统一的数字化材料特征, 这些特征包含材料的元素属性和结构属性描述, 使得构建的机器学习模型能够从中学习规律, 以低廉的计算成本实现磁性基态的分类和磁矩的预测:

1) 通过两步式的特征选择方法, 发现了对磁性基态分类和磁矩预测具有重要贡献的材料特征, 包括材料元素特征中的电负性、原子磁矩和原子外围轨道未充满电子数, 显示了材料的组成元素性质和电子排布与材料磁性性能之间的密切关系。

2) 基于随机森林算法, 建立了磁性基态分类模型和铁磁性材料磁矩预测模型. 在材料测试集的检验中, 分类模型对四种磁性基态的平均分类准确率达 85.23%, 预测模型对磁矩预测的平均绝对误差仅为 $0.098 \mu_B/atom$, 两个模型均展现了良好的精度和泛化能力。

参考文献

- [1] Zhang Z D 2015 *Acta Phys. Sin.* **64** 067503 (in Chinese) [张志东 2015 物理学报 **64** 067503]
- [2] Li L Z, Jiang J L, Wei R H, Li J P, Tian Y, Ding J N 2016

- Acta Phys. Sin.* **65** 018103 (in Chinese) [李绿洲, 蒋继乐, 卫荣汉, 李俊鹏, 田煜, 丁建宁 2016 物理学报 **65** 018103]
- [3] Sander D, Valenzuela S O, Makarov D, Marrows C H, Fullerton E E, Fischer P, McCord J, Vavassori P, Mangin S, Pirro P, Hillebrands B, Kent A D, Jungwirth T, Gutfleisch O, Kim C G, Berger A 2017 *J. Phys. D: Appl. Phys.* **50** 363001
- [4] Vedmedenko E Y, Kawakami R K, Sheka D D, Gambardella P, Kirilyuk A, Hirohata A, Binek C, Chubykalo F O, Sanvito S, Kirby B J, Grollier J, Everschor S K, Kampfrath T, You C Y, Berger A 2020 *J. Phys. D: Appl. Phys.* **53** 453001
- [5] Long T, Fortunato N M, Zhang Y X, Gutfleisch O, Zhang H B 2021 *Mater. Res. Lett.* **9** 169
- [6] Yamada Y, Ueno K, Fukumura T, Yuan H T, Shimotani H, Iwasa Y, Gu L, Tsukimoto S, Ikuhara Y, Kawasaki M 2011 *Science* **332** 1065
- [7] Yao Q S, Lu M, Du Y P, Wu F, Deng K M, Kan E J 2018 *J. Mater. Chem. C* **6** 1709
- [8] He C L, Xu H J, Tang J, Wang X, Wei J W, Shen S P, Chen Q Q, Shao Q M, Yu G Q, Zhang G Y, Wang S G 2021 *Acta Phys. Sin.* **70** 127501 (in Chinese) [何聪丽, 许洪军, 汤建, 王潇, 魏晋武, 申世鹏, 陈庆强, 邵启明, 于国强, 张广宇, 王守国 2021 物理学报 **70** 127501]
- [9] Gong C, Zhang X 2019 *Science* **363** eaav4450
- [10] Wang H Y, Liu H Y, Sun J F, Gu N 2018 *Sci. Sin. Technol.* **48** 921 (in Chinese) [王皓瑶, 刘海祎, 孙剑飞, 顾宁 2018 中国科学: 技术科学 **48** 921]
- [11] Jha D, Choudhary K, Tavazza F, Liao W K, Choudhary A, Campbell C, Agrawal A 2020 *Nat. Commun.* **11** 3643
- [12] Belsky A, Hellenbrandt M, Karen V L, Luksch P 2002 *Acta Crystallogr., Sect. B: Struct. Sci.* **58** 364
- [13] Kirklin S, Saal J E, Meredig B, Thompson A, Doak J W, Aykol M, Ruhl S, Wolverton C 2015 *NPJ Comput. Mater.* **1** 15010
- [14] 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
- [15] Schleder G R, Padilha A C M, Acosta C M, Costa M, Fazzio A 2019 *J. Phys. Mater.* **2** 032001
- [16] Liu Y, Zhao T L, Ju W W, Shi S Q 2017 *J. Materialomics* **3** 159
- [17] Isayev O, Oses C, Toher C, Gossett E, Curtarolo S, Tropsha A 2017 *Nat. Commun.* **8** 15679
- [18] Kou W B, Dong H, Zou M Q, Han J Y, Jia X X 2021 *Acta Phys. Sin.* **70** 030701 (in Chinese) [寇雯博, 董瀚, 邹岷强, 韩均言, 贾西西 2021 物理学报 **70** 030701]
- [19] Yang Z X, Gao Z R, Sun X F, Cai H L, Zhang F M, Wu X S 2019 *Acta Phys. Sin.* **68** 210502 (in Chinese) [杨自欣, 高章然,

- 孙晓帆, 蔡宏灵, 张凤鸣, 吴小山 2019 *物理学报* **68** 210502]
- [20] Lu S H, Zhou Q H, Ouyang Y X, Guo Y L, Li Q, Wang J L 2018 *Nat. Commun.* **9** 3405
- [21] Xu Y B, Yamazaki M, Villars P 2011 *Jpn. J. Appl. Phys.* **50** 11RH02
- [22] Frey N C, Horton M K, Munro J M, Griffin S M, Persson K A, Shenoy V B 2020 *Sci. Adv.* **6** eabd1076
- [23] Yamamoto T https://storage.googleapis.com/rimcs_cgmn/cgmn_matsci_May_27_2019.pdf [2021-8-10]
- [24] Materials Project API <https://materialsproject.org/open> [2021-8-10]
- [25] Ward L, Dunn A, Faghaninia A, Zimmermann N E R, Bajaj S, Wang Q, Montoya J, Chen J M, Bystrom K, Dylla M, Chard K, Asta M, Persson K A, Snyder G J, Foster I, Jain A 2018 *Comput. Mater. Sci.* **152** 60
- [26] Breiman L 2001 *Mach. Learn.* **45** 5
- [27] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E 2011 *J. Mach. Learn. Res.* **12** 2825

附录

表 A1 基于两步式特征选择法获得的材料特征及其物理含义

Table A1. Material features and their physical meanings obtained by the two-step feature selection method.

	特征	物理含义
1	Mode Electronegativity*	材料组成元素电负性的众数
2	Min Electronegativity	材料组成元素电负性的最小值
3	Range NUnfilled	材料组成元素外围未充满电子数的范围
4	Avg_dev NUnfilled	材料组成元素外围未充满电子数的平均偏差
5	Max NUnfilled	材料组成元素外围未充满电子数的最大值
6	Mode NfUnfilled	材料组成元素f轨道未充满电子数的众数
7	Mean NfUnfilled	材料组成元素f轨道未充满电子数的平均值
8	Range NfUnfilled*	材料组成元素f轨道未充满电子数的范围
9	Avg_dev NfUnfilled	材料组成元素f轨道未充满电子数的平均偏差
10	Max NfUnfilled	材料组成元素f轨道未充满电子数的最大值
11	Range NdUnfilled	材料组成元素d轨道未充满电子数的范围
12	Mean NdUnfilled*	材料组成元素d轨道未充满电子数的平均值
13	Avg_dev NdUnfilled*	材料组成元素d轨道未充满电子数的平均偏差
14	Max NdUnfilled	材料组成元素d轨道未充满电子数的最大值
15	Mean GSmagmom*	材料组成元素磁矩的平均值
16	Range GSmagmom	材料组成元素磁矩的范围
17	Avg_dev GSmagmom*	材料组成元素磁矩的平均偏差
18	Max GSmagmom	材料组成元素磁矩的最大值
19	Max AtomicWeight	材料组成元素重量的最大值
20	Mode AtomicWeight	材料组成元素重量的众数
21	Mean GSvolume_pa	材料组成元素体积的平均值
22	Range MeltingT	材料组成元素熔点的范围
23	Max MeltingT	材料组成元素熔点的最大值
24	Mode Number	材料组成元素原子序数的众数
25	Max Number	材料组成元素原子序数的最大值
26	Min NValence	材料组成元素价电子的最小值
27	Range NfValence	材料组成元素f轨道价电子的范围
28	Avg_dev NfValence	材料组成元素f轨道价电子的平均偏差
29	Avg_dev NdValence	材料组成元素d轨道价电子的平均偏差
30	Mode MendeleevNumber	材料组成元素门捷列夫数的众数
31	Avg_dev MendeleevNumber	材料组成元素门捷列夫数的平均偏差
32	Min MendeleevNumber	材料组成元素门捷列夫数的最小值
33	Vpa	材料的晶胞体积
34	Sine coulomb matrix 0	正弦库仑矩阵的第0个特征值

* 该特征同时用于磁性基态分类和磁矩预测.

Classification of magnetic ground states and prediction of magnetic moments of inorganic magnetic materials based on machine learning^{*}

Li Wei¹⁾ Long Lian-Chun^{1)†} Liu Jing-Yi¹⁾ Yang Yang²⁾

¹⁾ (*Faculty of Materials and Manufacturing, Beijing University of Technology, Beijing 100124, China*)

²⁾ (*Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*)

(Received 1 September 2021; revised manuscript received 22 November 2021)

Abstract

Magnetic materials are important basic materials in the information age. Different magnetic ground states are the prerequisite for the wide application of magnetic materials, among which the ferromagnetic ground state is a key requirement for future high-performance magnetic materials. In this paper, machine learning is used to study the classification of ferromagnetic, antiferromagnetic, ferrimagnetic and paramagnetic ground states of inorganic magnetic materials and the prediction of magnetic moments of inorganic ferromagnetic materials. We obtain 98888 inorganic magnetic materials data from the Materials Project database, containing material ids, chemical formulae, CIF files, magnetic ground states and magnetic moments, and extract 582 elemental and structural features for the inorganic magnetic materials by using Matminer. We design a two-step feature selection method. In the first step, RFECV is used to evaluate material features one by one to remove redundant features without degrading the model accuracy. In the second step, we rank the material features to further refine and select the most important material features for the model, and 20 material features are selected for the classification of magnetic ground states and the prediction of magnetic moments, respectively. Among the selected material features, it is found that the electronegativity, the atomic own magnetic moment and the number of unfilled electrons in the atomic peripheral orbitals all make important contributions to the classification of magnetic ground states and the prediction of magnetic moments. We build a magnetic ground state classification model and a magnetic moment prediction model by using the random forest, and quantitatively evaluate the machine learning models by using the 10-fold cross-validation approach, and the results show that the constructed machine learning models has sufficient accuracy and generalization capability. In the test set, the magnetic ground state classification model has an accuracy of 85.23%, a precision of 85.18%, a recall of 85.04%, and an F1 score of 85.24%; the magnetic moment prediction model has a goodness-of-fit of 91.58% and an average absolute error of 0.098 μ_B per atom. This study provides a new method and choice for high-throughput classification and screening of magnetic ground states of inorganic magnetic materials and predicting the magnetic moment of ferromagnetic materials.

Keywords: machine learning, random forest, magnetic ground state, magnetic moment

PACS: 02.60.Pn, 75.50.Gg, 32.10.Dk

DOI: 10.7498/aps.71.20211625

^{*} Project supported by the National Key R&D Program of China (Grant No. 2018YFB0703500).

[†] Corresponding author. E-mail: longlc@bjut.edu.cn