
主动学习助力高居里温度二维铁磁材料设计

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/progress/16785.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

主动学习助力高居里温度二维铁磁材料设计。

2021年12月3日，东南大学王金兰教授团队在Chem上发表了一篇题为On-the-fly Interpretable Machine Learning for Rapid Discovery of Two-dimensional Ferromagnets with High Curie Temperature的新研究。

课题组通过开发一种集成了特征工程、模型学习、数据采样、第一性原理计算和模型解释的主动学习框架，实现了高居里温度二维铁磁材料的快速准确预测，突破了复杂材料性质空间搜索的难题。论文唯一通讯作者是王金兰教授；第一作者是博士生陆帅华、周堃桦副教授。

近年来，人工智能技术在材料科学领域掀起了一股技术变革浪潮。将先进的机器学习技术与传统的实验或理论计算相结合，不仅显著加速了材料研发进程，也进一步提供了对材料构效关系的新认知。尽管这种全新的科学研究范式取得了举世瞩目的成就，但其在复杂体系和性质中的应用依然处于萌芽阶段，仍面临着诸多困难和挑战。

二维材料的晶体结构多样，影响铁磁性质的因素复杂（结构、组分、电子、自旋等），将机器学习应用于二维磁性体系困难重重。我们在先前的工作中，结合机器学习和第一性原理高通量计算，初步从现有的二维材料数据库中找到了一定数量的二维铁磁材料。但是实际可能的化学空间规模远远超过了现有的数据库。在更大范围的化学空间中，更有可能找到性能优异的二维铁磁材料。然而，现有的数据匮乏、描述符缺乏深层的物理见解、机器学习模型缺乏可解释性，这些因素制约了磁性材料化学空间的搜索。具体而言，为了搜索如此庞大的化学空间，机器学习算法需要对大量的高质量数据进行训练。但是对于二维磁性材料而言，磁基态、居里温度等关键性质计算成本高昂，导致现有的二维磁性材料数据严重匮乏。在实际机器学习材料设计任务中，如何在模型达到目标性能的前提下，尽可能地减少标注成本是一项亟需解决的挑战。为了解决这一机器学习材料设计领域普遍存在的问题，我们提出了一个具有反馈迭代功能的主动学习框架，将梯度提升算法和边缘采样算法相结合，来动态地构建完备的高质量数据集（图1）。

图1：基于梯度提升算法和边缘采样算法的可解释性主动学习框架。

该主动学习框架集成了特征工程、模型学习、数据采样、第一性原理计算和模型解释等功能，可以使我们在数据匮乏和特征维度较高的情况下，高效地搜索化学空间。与已有的主动学习框架不同，我们的框架为了成功探索二维铁磁材料空间，进行了一系列的技术改进：（1）集成了特征工程功能，包括特征组合和特征选择，在尽量保留材料信息的同时，避免高维特征集的产生；（2）对于主动学习而言，采样函数的构建至关重要，因为模型性能并不是随着标注样本的数量增加而无限增长的，机器学习模型需要优先学习最能够提升当前模型性能的样本。我们在框架中采用了两步边缘采样算法，该算法在采样过程中不仅仅考虑了决策边界的样本，而且考虑了材料结构与组分的多样性；（3）采用梯度提升算法作为代理模型，实现了模型的高效学习和高精度预测；（4）模型解释性的应用，打开了机器学习二维铁磁材料模型的黑盒子。值得一提的是，我们的框架是一个自动化程序，因此可以很容易地应用于其他材料的发现和化学空间探索，而不需要化学家或材料科学家有太多的机器学习知识。

图2：集成了晶体场理论、原子磁性和最近邻/次近邻局域原子化环境的磁性材料描述符。

除了数据，材料描述符的构建对于机器学习模型性能的提升也是十分重要的。我们上个工作基于晶体图提出的描述符，尽管在铁磁-反铁磁态分类中取得了良好的表现，但是在铁磁-反铁磁能量差回归中效果不尽如人意。其主要原因是该描述符只包含了原子的最近邻原子化环境信息，缺失了一些影响铁磁-反铁磁能量差的重要因素。为了提升机器学习模型性能，我们首先提出了一个新的局域原子化模型，改进了已有的晶体图描述符，在最近邻原子化环境信息的基础上，新增了次近邻原子化环境信息。其次，我们将新的晶体图描述符和晶体场理论以及磁性理论相结合，最终构建出混合磁性材料描述符。该描述符只依赖于原子的相对位置和基本性质，不需要进行繁琐的计算，因此具有良好的可移植性。

随后，我们从文献和数据库中挑选了20种具有母体材料的二维材料初始构型，其中大部分是二元金属材料。然后通过对这些初始结构进行简单的元素替换，我们构建了包含超过20万种候选化合物的化学空间，这些候选化合物都未被报道过。最后，我们将主动学习框架和磁性描述符应用到该化学空间的搜索中，模型在有限的迭代步数后，在材料热稳定性和磁性预测上都取得了0.9以上的精度（图3）。

图4：二维铁磁材料筛选和铁磁/反铁磁态能量差的机器学习预测结果

经过筛选，我们最终从20万种候选化合物中找到了9622种二维铁磁候选材料。其中，有722种材料的铁磁-反铁磁能量差大于0.5eV，这些材料很可能具有高居里温度（图4），为二维铁磁材料实验和理论研究提供了丰富的候选材料。此外，我们还利用SHAP分析，揭示了磁性原子配位环境、磁性原子周期数和配位构型对铁磁性的影响。结果显示，配体场弱、磁性原子周期数低的四配位铁磁材料更容易出现高居里温度。

这项工作突破了机器学习技术在复杂体系与性质应用中的瓶颈，为快速精准探索庞大化学空间提供了一种极具潜力的新策略，预测了一系列潜在的高温二维铁磁材料，发展的新型主动学习框架和通用材料描述符将极大地加速其他功能材料的研发进程。（来源：科学网）

相关论文信息：<https://doi.org/10.1016/j.chempr.2021.11.009>

作者：王金兰等 来源：《化学》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发