

激光与光电子学进展

基于机器学习的光学功能玻璃研究进展

付丽丽^{1*}, 张志强¹, 徐慧敏¹, 任青颖¹, 郑锐林^{1,2***}, 韦玮^{1**}¹南京邮电大学电子与光学工程学院、柔性电子(未来技术)学院, 江苏 南京 210023;²南京邮电大学理学院, 江苏 南京 210023

摘要 光功能玻璃材料在研究过程中所涉及到的研发周期长和效率低的问题,极大阻碍了光学玻璃材料的发展。机器学习技术的出现对玻璃材料学发展起到了极大的推动作用,通过学习数据蕴含的规律,在庞大而复杂的玻璃数据中学习并预测新数据,加快了光学功能玻璃的研发进程。本文总结并展示了在光学玻璃预测中涉及到的几类机器学习算法并对其做了简要介绍,在此基础上重点介绍这些理论算法在玻璃研究工作中的重要应用,包括加速和改进传统玻璃研究方法、助力玻璃的成分-性质关联预测以及对光学玻璃的配方设计的建议等方面的内容,最后对机器学习在光功能玻璃研究中的应用前景以及未来发展趋势做了分析与展望。

关键词 神经网络; 机器学习; 光功能玻璃; 性质-成分关联; 性能预测

中图分类号 O439

文献标志码 A

DOI: 10.3788/LOP231278

Research Progress of Optical Functional Glass Based on Machine Learning

Fu Lili^{1*}, Zhang Zhiqiang¹, Xu Huimin¹, Ren Qingying¹, Zheng Ruilin^{1,2***}, Wei Wei^{1**}¹College of Electronic and Optical Engineering & College of Flexible Electronics (Future Technology), Nanjing University of Posts and Telecommunications, Nanjing 210023, Jiangsu, China;²School of Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, Jiangsu, China

Abstract The research process of optical functional glass materials involves long research and development cycles and low efficiency. Greatly hindered the development of optical glass materials. The emergence of machine learning technology has greatly promoted the development of glass materials science. By learning the laws contained in the data, learning and predicting new data from the huge and complex glass data has accelerated the research and development process of optical functional glass. This paper summarizes and demonstrates several types of machine learning algorithms involved in the prediction of optical glass and briefly introduces them. On this basis, it focuses on summarizing the important applications of these theoretical algorithms in glass research, including accelerating and improving traditional glass research methods, assisting glass composition-property correlation prediction, and suggestions for optical glass formulation design. Finally, the application prospects and future development trends of machine learning in optical functional glass research are analyzed and forecasted.

Key words neural network; machine learning; optical functional glass; property-composition correlation; performance prediction

1 引言

光学功能玻璃具有高度的光学透过性、化学及物性(结构和性能)上的高度均匀性,以及特定和精确的光学常数特性,在各类光学仪器、信息通信、发光显示等领域有着极其重要的应用,是光电信息技术产业的基础和重要组成部分。作为光电子基础材料的光学

玻璃在光传输、光储存和光电显示三大领域的应用十分广泛。玻璃是非结晶性的,它不受化学计量学要求的约束,再加上在玻璃转变温度以上的连续变化的黏度特性,意味着玻璃具有几乎无限的成分和加工可能性^[1]。正因如此,光学玻璃的分类多种多样:按功能不同可分为无色光学玻璃、有色光学玻璃、防辐照玻璃、耐辐照玻璃、紫外和红外玻璃、光学石英玻璃等种类;

收稿日期: 2023-05-11; 修回日期: 2023-06-07; 录用日期: 2023-06-15; 网络首发日期: 2023-07-12

基金项目: 国家自然科学基金(61905119)

通信作者: *fulili@njupt.edu.cn; **weiwei@njupt.edu.cn; ***ruilinzheng@hotmail.com

按用途可分为光导玻璃纤维、激光玻璃、光致变色玻璃、光的选择透过和反射玻璃以及非线性光学玻璃等。

随着光电子信息技术的发展,多种新型光学功能玻璃相继问世。为实现超长距离无中继的光纤通信,研究人员开发出了传输损耗低的卤化物玻璃,在后续的发展中又推出了掺铯光纤、掺镱光纤等具有特殊光学特性的光纤玻璃类型,进一步增强了光纤传输性能^[2-3]。在固体激光器的研究中,为实现高效率的激光倍频转换,研究人员提出了铽玻璃激光放大器,增强了激光器的系统输出特性^[4]。在某些场景下,为了实现特定功能,在玻璃中进行离子掺杂等技术操作,可以实现令人意外的效果,如在磷酸盐玻璃中添加稀土类金属,可以开发出可遮挡近红外线的玻璃,供照相机使用^[5]。此外,在太阳能电池、照明显示、检测防伪等领域应用新型钙钛矿纳米晶复合玻璃技术,进一步扩大了玻璃材料的应用范围,并为光电领域提供了新的机遇^[6]。尽管对于光学玻璃的研究已经十分深入,但因玻璃组成复杂多样,光学玻璃的发展进步十分缓慢,传统的设计方法的速度和准确性已不能满足当前对于新玻璃材料的需求,因此亟需寻求一种新的研究方式以加速新型光学玻璃的开发进程。

机器学习(ML)无疑是一种解决这些问题的捷径。目前,大多数的材料研究仍采用试错的方法,不仅程序复杂,而且耗时费力。第一性原理计算、有限元分析等手段虽然可以进行材料结构和性能方面的计算,但计算量较大、耗费资源、效率不高^[7]。密度泛函理论(DFT)是一种不错的计算建模方法,可计算原子结构的能量并用其估算材料性能^[8]。但是DFT计算速度慢且计算量非常大,哥伦比亚大学的科学家Nongnuch Artrith表示,DFT限制在皮秒量级的时间范围内,只能模拟约200个原子。2011年,在美国材料基因组(MGI)计划推动下,信息学战略在材料科学领域已初具规模,加速了机器学习的应用。目前,ML已广泛应用于材料科学研究,旨在利用现有材料的特性和功能去发现具有更理想功能的新材料,ML已成为用于新材料研发的重要研究工具^[9]。

目前,基于ML的新型玻璃材料的研究重点在于玻璃的性质预测上,例如,Bishnoi等^[10]使用高斯过程回归(GPR),预测了具有稀疏数据集的硅酸盐玻璃的杨氏模量,研究表明,在稀疏数据集的应用上GPR模型明显优于神经网络(NN),同时确保没有过度拟合,为预测的可靠性提供了定量界限。美国康宁公司的研究人员利用积累的玻璃数据对预测给定成分的氧化物玻璃密度、杨氏模量、剪切模量和泊松比进行了全面的机器学习研究,结果表明,ML方法中的随机森林、K最近邻、神经网络提供了良好的预测性能,为未来玻璃开发提供了合理的方法基准^[11]。大块金属玻璃(BMGs)的玻璃形成能力(GFA)是一个决定性的参

数,Ghorbani等^[12]进行了大量研究,利用ML方法,通过消除合金玻璃设计中的障碍来预测BMG的性能,测量值和预测值的比较结果表明,该模型是可靠的,可以真正应用于预测BMG的GFA中。研究结果表明,ML在玻璃科学中可以发挥其独特优势,无论是大数据集还是稀疏数据集,都可以通过相关的算法进行研究,对玻璃科学的发展提供了有力的支持。利用机器学习,我们可以对玻璃的性质进行更加精致的探索,反过来,利用ML所预测到的结果则可以为新玻璃的合成提供一个可信的配方支持,极大地减少实验盲目性和配方设计复杂性,这无疑为研究新型光学功能玻璃提供了全面且便捷的设计方法。本文首先对光学玻璃预测中涉及到的几类机器学习算法做了简要介绍,在此基础上介绍这些理论算法在玻璃研究工作中的重要应用,包括加速和改进传统玻璃研究方法、助力玻璃的成分-性质关联预测以及对光学玻璃的配方设计的建议等方面内容。

2 机器学习方法

ML解决数据问题需依赖不同的算法。正如数据科学家所指出的那样,没有一种特定的算法可以完整地解决所有问题,所采用的算法类型取决于希望解决的问题类型、变量的数量以及最适合它的模型等^[13]。ML利用算法模型对现有的数据进行学习和预测、对数据进行一系列操作和理解,这些算法一般可分为三大类:监督式学习、非监督式学习、强化学习。

算法是ML的关键,监督式学习应用广泛,常用来处理分类问题和回归问题^[14]。其中,常用的ML算法包括支持向量机、决策树、K-近邻、人工神经网络以及深度学习等。对于非监督式学习,训练模型是为了找出数据样本之间存在的内在关系^[15]。非监督式学习包括K均值聚类算法、均值漂移算法等,常用于关联规则的学习和聚类。与监督式学习不同,强化学习的输入数据直接反馈到模型,模型将会立刻做出调整,强化学习常用于动态系统和机器人的控制^[16]。下面将对本文涉及的机器学习算法或模型展开介绍。

2.1 线性回归

线性回归(LR)是一种用来模拟一个因变量和多个解释变量之间关系的线性方法^[17]。一个解释变量的情况称为简单线性回归模型,多个解释变量的情况称为多元线性回归模型。作为一种常用的统计方法,其原理明确,模型简单易用。其中,简单的一元线性模型可以表示为

$$y = \beta_0 + \beta_1 x + b, \quad (1)$$

式中: β_0 和 β_1 为决定该条拟合直线走势的参数, β 可以用最小二乘法来进行估计计算^[18]; b 为误差,代表数据偏离真实值的程度。单一元线性回归的思想为寻找一条直线可以尽可能多地通过这些数据点。

然而在常用的机器学习算法中,应用更多的是多元线性回归,将 β 用 \mathbf{w} 替换可得

$$y = w_1x_1 + w_2x_2 + \dots + w_nx_n + b = [w_0w_1w_2\dots w_n] \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} + b_0 \quad (2)$$

为了计算出 \mathbf{w} ,引入损失函数,并采用梯度下降法^[19]进行求解

$$\frac{\partial L(\mathbf{w})}{\partial w_i} = -\sum_{j=1}^m \left[y^{(j)} - \sum_{i=1}^n w_i x_i^{(j)} - b \right] \cdot x_i^{(j)}, \quad (3)$$

$$w_{i+1} = w_i + \alpha \left[\sum_{j=1}^m (y^{(j)} - \sum_{i=1}^n w_i x_i^{(j)} - b) \cdot x_i^{(j)} \right], \quad (4)$$

式中: α 为学习率,需手动设定。令公式不断迭代直至收敛,在线性回归算法中,有两类常用的算法:Lasso回归^[20]和岭回归^[21]。这两者的区别在于损失函数 J 中加入的正则化系数不同。Lasso回归中加入L1正则化,

$$J = \frac{1}{n} \sum_{i=1}^n [f(x_i) - y_i]^2 + \lambda \|\mathbf{w}\|_1. \quad (5)$$

岭回归中加入L2正则化,

$$J = \frac{1}{n} \sum_{i=1}^n [f(x_i) - y_i]^2 + \lambda \|\mathbf{w}\|_2^2. \quad (6)$$

线性回归模型拥有很好的可解释性,可以从权重 \mathbf{w} 中直接看出每个特征对结果 y 的影响程度,通过绘制散点图来观察是否存在线性关系。总的来说,线性回归模型思想简单且容易实现,对于小数据集或简单的关系效果明显,虽然对于非线性数据不能达到理想的效果,但面对大多数数据环境,仍是机器学习中很受欢迎的建模手段。

2.2 遗传算法

遗传算法(GA)来自于生物的起源与进化问题,遵循达尔文的适者生存理论,GA由Holland提出,通过模拟一个人工种群的进化过程,包括选择、交叉以及变异等机制,在每次迭代中都保留一组候选个体,不断重复此过程并经过若干代进化后,最终达到最佳理想的适应状态^[22]。

遗传算法是建立在自然选择和遗传概念基础上的基于搜索的算法,它的工作流程如图1^[23]所示。具体过程^[24]如下:1)选择。从群体中选择优胜的个体,淘汰劣质个体,这里的选择并不等价于排序并截取的过程,而是概率性的选择,也即适应度越高的群体被选择到的概率越大。应用轮盘赌选择法,根据不同特征的概率分布做出随机选择。2)交叉。交叉的方式有多种,如一点交叉、多点交叉、统一交叉、戴维斯命令交叉等,交叉的过程需要从上一代的染色体中寻找两条染色体,一条是父染色体另一条是母染色体,然后将这两条染色体的某一个位置切断并拼接在一起,从而生成一条新的染色体。这条新染色体上包含了一定数量的

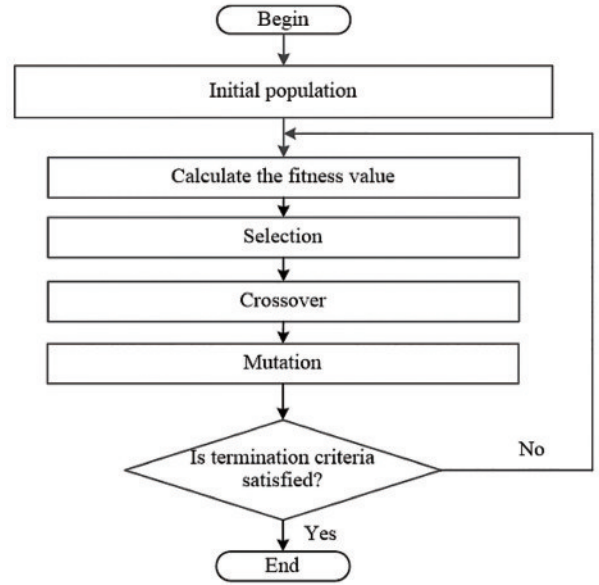


图1 遗传算法工作流程^[23]

Fig. 1 Genetic algorithm workflow^[23]

上一代基因片段。3)变异。当通过交叉生成了一条新的染色体后,需要在新染色体上随机选择若干个基因,然后随机修改基因的值,从而给现有的染色体引入新的基因,这便突破了当前搜索的限制,更有利于算法寻找到全局最优解。经过上述过程筛选过后,如符合条件则输出最佳结果,否则,便重复上述过程。

GA算法通过变异机制避免算法陷入局部最优,搜索能力较强,引入自然选择中的概率思想,个体的选择具有随机性,可拓展性也强,易于与其他算法进行结合使用。但同时遗传算法编程较为复杂,涉及到基因编码与解码,算法内包含的交叉率、变异率等参数的设定,需要依靠经验来确定,尤其是对于初始种群的优劣依赖性较强,这也是GA最大的缺点之一。通常来说GA算法在实际中并不常单独使用,结合其他算法或许是更好的应用方式。

2.3 人工神经网络

人工神经网络(ANN),通常简称为神经网络(NNs),是一种受生物神经网络启发而发明的数据处理算法^[25]。人工神经网络基于人工神经元的连接单元或节点,这些单元或节点松散地模拟生物大脑中的神经元。每个连接,就像生物大脑中的突触一样,都可以将信号传递给其他神经元。人工神经元接收信号然后处理它们,并可以向连接到它的神经元发出信号。连接处的“信号”是一个实数,每个神经元的输出由其输入之和的某种非线性函数计算。神经元和边缘通常具有随着学习而调整的权重,权重增加或减少可以改变连接处信号的强度。通常,人工神经元被聚合成层,不同的层可能会对其输入执行不同的转换。输入信号从输入层传播到输出层,最终完成对数据的复杂转换与处理。

神经网络可分为单层神经网络模型与多层神经网络模型,其中单层神经网络又被称为单层感知器(SLP)。神经网络由许多相互连接的神经元组成,每个节点执行一个独立的计算,通过被施加权重系数的连线连接在一起。一般而言,人工神经网络可以分为三层神经元:输入层(接收信息)、隐藏层(负责提取模式,执行大部分内部处理)和输出层(产生并呈现最终的网络输出),一个 ANN 可以有多个级别的隐藏层^[26]。输入层由用于实现目标变量估计的回归变量组成。层之间的连接被加权,权重可以通过添加非线性优化器来完成。训练的目标是在给定一组已知输入和目标值的情况下,最小化 ANN 输出和目标输出之间的误差。在 ANN 训练阶段,每个连接的权重根据提供的数据进行调整。然而,值得注意的是,基于数据和应用存在不同的激活函数^[27]。

典型的预测 ANN 如图 2 所示。各神经元分别属于不同的层,每一层的神经元可以接收前一层神经元的信号,并产生信号输出到下一层。第 0 层叫输入层,最后一层叫输出层,其他中间层叫作隐藏层。ANN 的工作过程如下:输入层接收外部输入,将其送入隐层中进行处理,隐层通过激活函数将处理结果传递给下一层隐层或输出层。输出层产生神经网络的输出结果。该神经网络的训练过程包括前向传播和反向传播两个阶段,前向传播时,神经网络将前面层的输出传递给后面层进行计算,计算结果与真实值进行比较,然后使用反向传播算法来更新网络的权重和偏置,使得网络能够更准确地预测输出结果。通过多次迭代训练,神经网络可以逐渐提高准确性,实现有效的数据分析和预测。

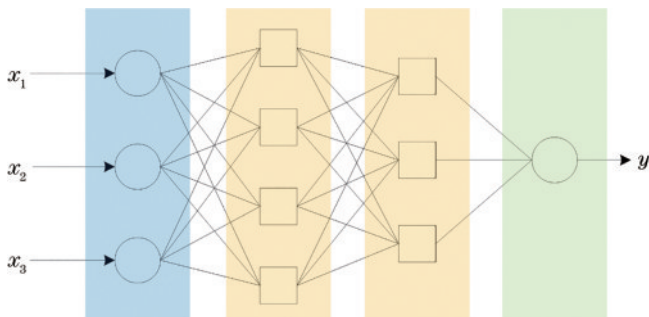


图 2 人工神经网络模型

Fig. 2 Artificial neural network mode

虽然 ANN 模型在数据并行处理能力、学习性、容错性和泛化能力上优势明显,但它们也有一些局限性^[28]。首先,与所有数据驱动模型一样,ANN 在其训练范围之外表现不佳。此外,可能影响 ANN 模型性能的第二个限制是过拟合,当模型在训练数据中学习了太多噪声时,就会发生过度拟合,此时便失去了预测的一般性。虽然这可能不会显著影响具有单步提前预测范围的模型,但这在预测更长时间范围的模型中可

能会有问题。第三个限制是 ANN 模型本身是基于黑盒的模型,因此其内部是未知的,尽管它们可以提供足够的预测方法,但是它们缺乏对能源消耗及其行为的潜在参数的理解。除了 ANN 之外,图神经网络算法(GNN)也逐渐被用于材料计算之中^[29]。因为图数据有复杂、多样化的属性类型,而 GNN 可以模拟多种算法场景,因此该算法解决了传统算法在处理非欧氏空间数据时表现差的问题,极大地拓宽了机器学习在材料结构与性能发现中的应用范围^[30-31]。

3 机器学习在光学玻璃研究中的应用

由于玻璃组成成分的复杂性以及原子构成的无序性,人们对于玻璃的研究一直处于传统试错法和效率极低的模拟方法之中,这对于实现新的玻璃突破极为不利,随着光电信息科学的快速发展,特种光学玻璃的需求正逐步上升,现亟需一种新的研究方法用以加速玻璃科学的发展。玻璃研制过程中出现的问题,大都集中于玻璃的某一种特性优化以及新型功能玻璃配方研发,除此之外,探究玻璃形成的机制对于新玻璃的发现也极为重要。目前的玻璃研究工作经历了数十年的发展已积累了大量的实验数据,这为新数据驱动方式的加入提供了契机。将 ML 技术引入新玻璃的研究之中,极大地提升研究效率,减少因试错法带来的随机性和盲目性,为玻璃科学的进一步发展提供强大的驱动工具。本小节将对机器学习在光学玻璃研究应用中的三个重要研究方向展开介绍:1)传统研究方法和理论的加速与改进;2)对于特种光学玻璃材料的相关性质,开展玻璃成分与性质之间的关联预测;3)以机器学习驱动配方设计与优化问题。

3.1 加速、改进传统研究方法

基于量子计算和传统方法的材料学发展迅速,在过去的半个多世纪里,密度泛函理论(DFT)成为了凝聚态物理计算材料学和计算化学领域最常用的方法之一^[32]。除此之外,分子动力学(MD)模拟同样为研究分子体系的结构与性质提供了重要帮助。在最近的研究中,基于 ML 的辅助方法在传统玻璃材料领域引起极大的震动,包括 ML 在辅助加速 MD 模拟计算以及利用 ML 解决传统方法存在的问题中效果显著,成为了当前玻璃材料研究领域十分热门的话题。

将经典研究方法与 ML 相结合可以加快传统方法在材料性质预测中的速度,把数据筛选效率提升到了新的阶段。在 Urata 等^[33]的研究中,为了解释超低损耗玻璃光纤氟掺杂之谜,在已知少量氟掺杂可以通过降低二氧化硅玻璃的密度波动来进一步降低瑞利散射的原理下,通过 ML 优化了精确的力匹配势(FMP),以便了解氟元素如何抑制玻璃光纤中的瑞利散射。这项研究试图在经典 MD 模拟中揭示氟掺杂抑制密度波动的机制,利用了基于 DFT 计算的精确 FMP,为了实现这一目标,采用了原子间势的泛函类型,并使用贝叶斯

优化机器学习算法对参数进行优化,使用 ML 第 2500 次迭代的参数作为 MD 模拟的计算数据,这里的 MD 模拟使用 LAMMPS 包进行。最终结果表明,在 ML 优化辅助下,获得的精确 FMP 能够使用相对大规模的经典 MD 模拟,缩短了昂贵的量子计算周期。类似地,在预测硅酸盐玻璃的杨氏模量研究中,因玻璃的组成成分十分复杂,试错法已不能胜任实验中庞大而繁杂的数据,而已知的 Makishima-Mackenzie (MM) 模型并不能出色地完成玻璃非线性的关系预测,因此 MD 模拟在这里就凸显出其优势所在,但作为一种使用蛮力的计算方式,MD 模拟在大成分范围计算中十分费力。随着 ML 的快速发展, Yang 等^[34]将 MD 模拟与 ML 相结合,利用 MD 模拟所生成的玻璃数据,分别在 PR、Lasso、RF、ANN 中对数据做了理论实验,结果表明高通量分子动力学模拟和机器学习相结合可以提供一种稳健的方法来预测硅酸盐玻璃的杨氏模量,极大提升了传统方法的预测速度。Hu 等^[35]同样把 MD 模拟与 ML 相结合,使用 MD 模拟的力场势和元素摩尔分数构建数据集(其中包括物理和成分信息),随后通过梯度增强机(GBM-LASSO)实现了最小绝对收缩和选择算子,开发了统计学习预测模型。通过大量模拟和实验数据验证表明,GBM-LASSO 模型不仅在训练集的合成范围内,而且在训练集以外的高维合成空间中,都具有良好的密度和弹性模量预测能力。这表明 ML 在加速传统研究方式(如试错法和 MD 模拟)中起到了重要作用,特别是通过“相结合”的方式按下了玻璃材料科学研究的加速键。

除了对传统研究方法的加速之外,ML 还在改进

数据完整性、进行材料数据参数优化中起到了其他方式无可比拟的作用。Liu 等^[36]提出了一种基于从头算分子动力学模拟(AIMD)、高斯过程回归(GPR)和贝叶斯优化(BO)的新的力场参数化方法,该项工程旨在建立硅酸盐玻璃建模中精确、可转移和高效计算的原子间力场。然而,大量需要优化的力场参数使得传统的参数化方法效率低下或可能受到偏差的影响。因为传统的基于梯度的优化方法[如最陡下降或共轭梯度(CG)]在探索粗糙函数时效率非常低,并且很可能产生局部最小值而不是全局最小值,因此,传统的优化方法往往是有偏见的,作为传统的“基于直觉的”力场参数化的替代途径,人工智能和 ML 技术有潜力提供一些有效的、无偏见的优化方案。

图 3 给出了参数化过程的概述。首先,创建一个包含一些“已知点”的初始数据集,即代价函数 R_x 选择参数集的值 E 。其次,使用 GPR 对已知点进行插值,并评估整个参数空间上的插值不确定性。再次,使用 BO 方法预测一组最佳参数 E ,即探索参数空间、降低模型不确定性和找到成本函数的全局最小值之间的最佳平衡。与 BO 预测的参数集相关的代价函数 $R_x\{E\}$ 随后通过经典的 MD 模拟并将模拟液体的结构与参考 AIMD 配置的结构进行比较来计算,这个新的数据点 $R_x\{E\}$ 将被添加到数据集中。然后,利用新的数据集对基于 GPR 的插值进行优化,并通过 BO 预测新的最优参数集。这个循环迭代重复,直到在代价函数中获得一个令人满意的最小值。最后,通过 CG 优化进一步细化 BO 预测的全局最小值,自此便得到了基于 ML 优化的新的力场参数。

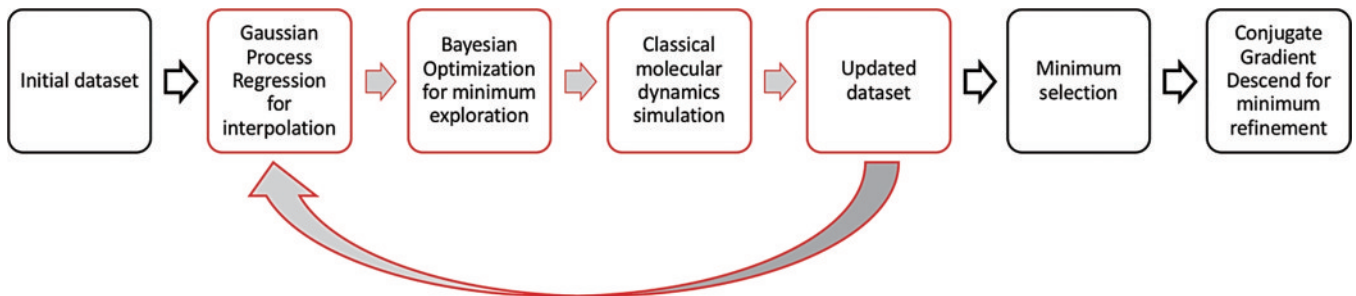


图 3 使用 ML 进行贝叶斯插值处理过程^[36]

Fig. 3 Bayesian interpolation process using ML^[36]

基于梯度计算和基于 ML 优化的比较如图 4(a) 所示,代价函数 R_x 对力场参数的依赖非常粗略,在放大精细细节后粗糙度水平似乎会增加[图 4(b)]. 在 (q_{Si}, A_{SiO}) 空间中基于 ML 和基于 CG 的优化所探索的路径(从相同的初始点开始)如图 4(a) 所示。从图 4(c) 可以看出,与 ML 优化方法相比,CG 优化很快就会“卡”在代价函数的局部最小值中,并且无法成功识别全局最小值。总的来说,这些结果证实了 ML 可以极大地提高参数化过程的效率,并且抑制了关于优化起点的任意选择所产生的偏差风险,显著降低了“个

人直觉”在参数化过程中所发挥的作用。同时也证明了 ML 在数据集处理和优化传统方法中的优势,这无疑是玻璃材料研究中的极大进步。

3.2 助力成分-性质关联预测

在常规研究工作中,预测硅酸盐玻璃的溶解速率十分困难,然而这项工作对于理解硅酸盐玻璃的成分与耐久性关系十分重要。虽然经历了几十年的发展,但人们对于玻璃溶解的内在与外在机制仍然知之甚少,对于各类条件下的玻璃损伤性溶解没有一个可靠的模型用以准确预测,因此阐明硅酸盐玻璃中的溶解

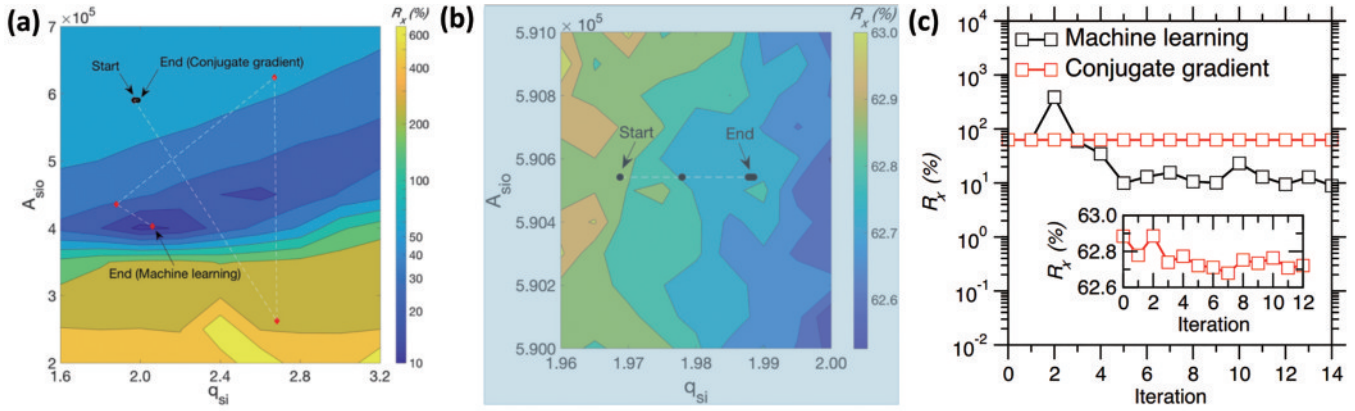


图 4 机器学习与共轭梯度优化的对比^[36]。(a)等高线图,显示成本函数 R_x 作为 q_{Si} 和 A_{SiO_2} 的函数,红圈和黑圈分别表示机器学习和共轭梯度优化所探索的路径;(b)图(a)数据的缩放;(c)代价函数 R_x 在机器学习和共轭梯度优化过程中的演化(插图:在共轭梯度优化的情况下获得的放大数据)

Fig. 4 Comparison between machine learning and conjugate gradient optimization^[36]. (a) Contour plot showing the cost function R_x as a function of q_{Si} and A_{SiO_2} , the red and black circles represent the paths explored by machine learning and conjugate gradient optimization, respectively; (b) scaling of the data in Fig. (a); (c) evolution of cost function R_x in machine learning and conjugate gradient optimization (inset: enlarged data obtained under conjugate gradient optimization)

机制对于深入了解玻璃损伤机制十分必要。在这里, Krishnan 等^[37]利用 ML 技术对玻璃的溶解动力学进行了数据模型上的解释,采用数据驱动的人工神经网络模型对一种铝硅酸盐玻璃溶解动力学进行了预测。为了证明人工神经网络可以提供可靠的预测结果,作者采用多种 ML 方法以及人工神经网络进行了对比预测,使用 8 种不同的钠铝硅酸盐玻璃样本数据和分别在多种酸性及碱性条件下的溶解动力学数据。所使用的人工神经网络算法通过优化网络的隐藏层个数来调整模型的精确程度。为了更加准确地理解和预测其溶解动力学,作者又加入了相关的物理特征输入,将物理约束与 ML 相结合,显著改善了预测结果,使得 ML 的能力得到了提升。

在这项工作中,玻璃数据输入来自 Hamilton^[38]报告的溶解率数据库,含有 299 个数据位点,其中:输入

变量包括玻璃组成、溶液的初始 PH 值、测量时的 PH 值;输出为 SiO_2 浸出率。Krishnan 等使用 ANN 与其他回归模型(线性回归、Lasso 回归、弹性网络回归、支持向量机回归以及随机森林回归)对数据进行训练并测试,结果表明,除了 ANN 模型,其他模型在数据集上的结果不尽如人意。为此在数据集中加入了包含玻璃材料溶解关系的先验物理信息,然后将数据集划分为两份,使用无“物理知情”的数据集和有“物理知情”的数据集进行训练。训练结果如图 5 所示,图中横坐标为溶液 PH,纵坐标为预测的浸出率,从图 5(a)中可以看到,除了 ANN,其他算法的预测结果明显不符合硅酸盐玻璃中溶解速率随 PH 的 V 型依赖关系,而图 5(b)中这种现象得到了改善,说明加入物理约束的机器学习能显著改善预测准确性。尽管如此,ANN 仍旧提供了最佳预测精度。

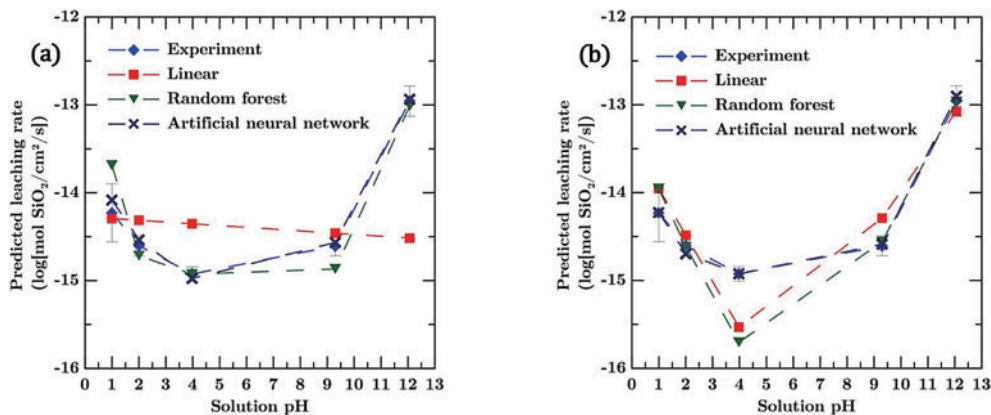


图 5 助力成分-性质关联预测^[37]。(a)使用线性回归、随机森林和人工神经网络方法进行的纯数据驱动预测;

(b)“基于物理”的预测

Fig. 5 Boosting Component Property Correlation Prediction^[37]. (a) Pure data-driven prediction using linear regression, random forest, and artificial neural network methods; (b) “physics-based” prediction

为了比较加入了物理约束下的人工神经网络与其他各类机器学习模型的差异和优势,作者将他们分别进行了新一轮的测试,测试结果如图 6 所示。相较于其他回归模型,加入底层物理信息的人工神经网络算法对玻璃溶解浸出速率的预测准确性比其他 5 种算法

的拟合程度都明显要高,这证明了人工神经网络算法具有很好的外推能力,具体表现如图 6(f)所示。总体而言,这种“物理知情”的机器学习方法一定程度上保证了硅酸盐玻璃溶解动力学预测的准确性和可行性,将这种方法扩展到预测其他玻璃性质上同样适用。

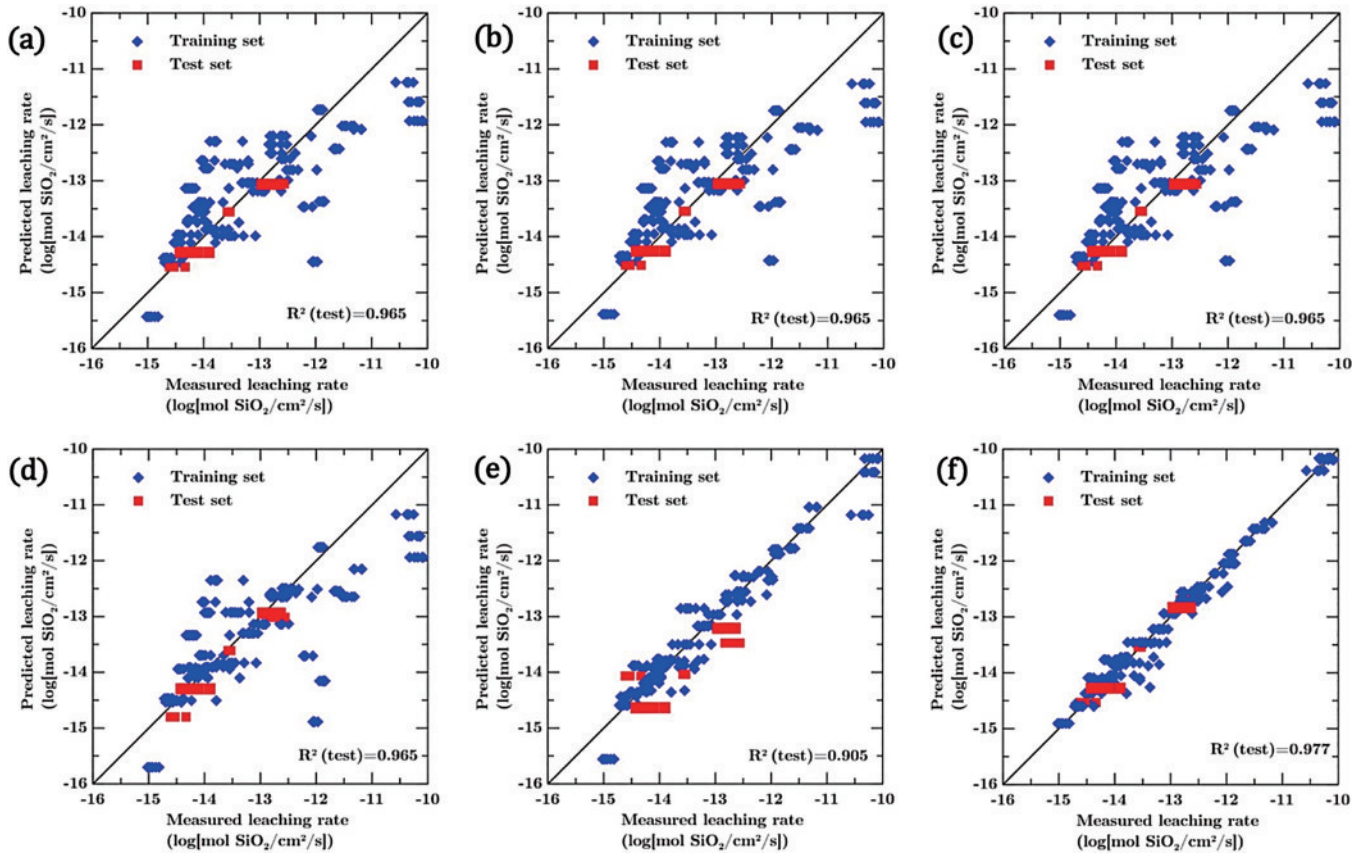


图 6 预测的浸出率散点图^[37]。(a)线性回归;(b)Lasso 回归;(c)弹性网络回归;(d)支持向量机回归;(e)随机森林;(f)人工神经网络算法

Fig. 6 Scatter plot of predicted leaching rate^[37]. (a) Linear regression; (b)Lasso regression; (c) elastic network regression; (d) support vector machine regression; (e) random forest; (f) artificial neural network algorithm

由于基于机器学习的回归预测方法大都属于黑盒模型,不具备可解释性,即对预测结果的运作过程或模型的具体学习情况缺乏实际解释,针对这个问题 Zaki 等^[39]提出了利用机器学习结合 SHAP (SHapley Additive ExPlanations)的解释方法,在成分-性质关联预测中发挥了极大作用。SHAP 方法用于解释机器学习模型的输出,它基于 Shapley 值,使用博弈论将模型预测的功劳分配给每个特征,用于表示每个特征对模型结果的贡献^[40]。它有助于以人类可以理解的方式解释机器学习模型的预测,通过为每个输入要素分配数值,可以显示各要素对最终预测结果的贡献方式和贡献程度。这样,我们就可以了解模型如何做出决定,进而从中识别出最重要的特征。

Zaki 团队开发了面向阿贝数(V_d)和折射率(N_d)的机器学习模型,利用包含两个隐层的神经网络对来自 INTERGLAD 中的数据展开训练,以氧化物成分作为输入,光学特征(V_d 、 N_d)作为输出,最终得到的模型对

测试集展开预测。为了捕捉成分-属性关系,利用 SHAP 对结果展开分析,如果该值较高,则该分量对模型预测的影响较大,反之亦然^[41]。然后,通过计算给定氧化物对应于其中存在氧化物的所有玻璃组合物的绝对值的平均值来获得平均绝对 SHAP 值。因此,如果氧化物的平均绝对 SHAP 值较高,则该成分在平均意义上对预测给定性质有更大的影响。图 7(a)为 SHAP 平均值对 V_d 的影响程度由上而下依次减弱。图 7(b)为各类氧化物对 V_d 的贡献程度,如有利于提高 V_d 的有 SiO_2 、 P_2O_5 、 B_2O_3 等,对提高 V_d 不利的成分有 TiO_2 、 Nb_2O_5 、 Bi_2O_3 等。对于 N_d 的 SHAP 解释同 V_d 。利用这种解释方式可以帮助我们了解各类化合物在光学玻璃中对相关参数的贡献,有助于在配方设计时合理搭配各要素成分比例,为玻璃配方设计提供了参考方向。此外,在玻璃的其他性质预测中也可参考上述方式,如预测成分与密度的关系、成分与玻璃化转变温度之间的关系等。由于机器学习本身为类似黑盒的纯数据规

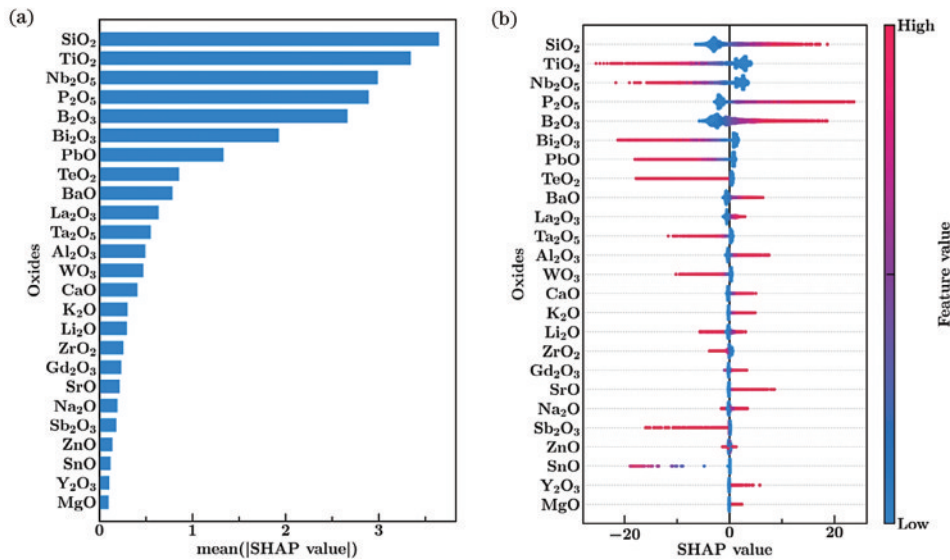


图7 预测结果^[39]。(a) SHAP平均值对模型输出的平均影响;(b) V_d 的 SHAP值的总结图

Fig. 7 Forecast results^[39]. (a) Average influence of SHAP average on model output; (b) summary of V_d SHAP values

律查找方法,因此可先不必考虑各性质之间的差异,皆可利用机器学习结合各类算法模型实现预定目标的预测。机器学习具有广泛的预测能力、强大的数据关联与查找能力,这为实现成分-性质关联预测提供了强大的工具支撑。

3.3 驱动玻璃配方设计与优化

在光电信息产业的发展过程中,人们对以平面显示、光纤通信、医疗检测为代表的光学玻璃材料的需求快速增加,同时对于玻璃材料的光学特性要求也在不断提高。为了设计出满足特定需求的光学玻璃,需针对玻璃配方设计展开深入研究。到目前为止,所报道的大多数研究工具和机器学习模型都集中在预测给定玻璃成分的选定性能上。然而,对于新的玻璃开发者来说,最重要的是解决逆向设计问题,即找到与被预测属性有相关性的可能成分或可能的配方。

高折射率一直是光学玻璃设计中的重要参数,为了寻找一种安全无毒的元素作为提升玻璃折射率的成分,Menezes等^[42]将Nb₂O₅作为制作高折射率光学功能玻璃的氧化物成分,利用名为GLAS的机器学习算法对含有特定成分的玻璃配方进行了预测与设计。该项工作的目的是开发一种新的含有Nb元素的光学玻璃,使其具有相对较高的折射率($N_d > 1.65$),中度的阿贝数($35 < V_d < 55$),以及均匀的玻璃化转变温度 T_g 。为了开展该项工作,Menezes等采用了DEMa-UFSCar所提出的GLAS算法,该方法可以在玻璃形成能力、折射率、玻璃转变温度、光学透过率和耐腐蚀性等方面对玻璃性能做出预测,不过该算法的具体实现方式目前尚未披露。得益于强大的机器学习算法与硬件支持,Varshneya^[43]采用了Sciglass数据库所开放的玻璃数据信息,并结合INTERGLAD^[44]的数据作为补充,设计了用于开展该项工作的可用数据库。利用GLAS

开展机器学习,模拟了含Nb光学玻璃的配方,选出最佳成分范围中符合要求的参数,并对该配方进行了制备,作为对照组将该成分表中的Nb₂O₅替换为La₂O₃,各组玻璃在制备后将对相关折射率参数进行测定。

为了验证算法所建议配方的优劣,对表1中的玻璃进行制备。结果表明,Glass12具有良好的玻璃形成能力,但整体呈现出黄色。为了改善这种现象,研究者对Glass12进行了改造,向其中添加了更多的硝酸盐以维持Nb的氧化状态,得到的玻璃被命名为Glass14。接下来采用与Glass14相似的配方,但是降低了化合物中杂质的含量,得到Glass16样本。为了比较La₂O₃组分的玻璃,使用与Glass14相同含量的La₂O₃取代Nb₂O₅,得到Glass15,表2中GLAS为GLAS算法建议的La₂O₃玻璃。经过比较可以得出结论:Glass14比Glass15具有更高的折射率和密度,即氧化镧在达到高折射率方面不如氧化铌,同时还可以看到,GLAS程序提供的目标值与实测值非常接近,说明GLAS具有出色的预测能力。机器学习在玻璃特征学习与配方预测中起到了常规实验无法比拟的优势,即机器学习的加入大大简化了设计流程,极大减少了实验次数并且提升了配方可用性。

类似地,Cassar等^[45]提出了一种机器学习方法,将人工神经网络(ANN)与遗传算法(GA)相结合,解决了玻璃配方设计中的反向设计问题。该项工作通过训练玻璃转变温度(T_g)和折射率(N_d)的预测模型,在光学玻璃的设计趋势(高折射率和低玻璃转变温度)的指导下搜索相关的玻璃组成,最后在实验室中选择并生产了两种由组合算法建议的候选组合。首先在Sciglass数据库中收集氧化物玻璃数据,从中提取 T_g 以及 N_d 参数信息,然后经过数据预处理,利用包含两个隐层的深度神经网络对这些数据属性做出归纳。经

表 1 GLAS算法建议的配方和预测性质对比^[42]Table 1 Comparison of formula and prediction properties suggested by GLAS algorithm^[42]

Item	Glass 12	Glass 14	Glass 15	GLAS ¹	Glass 16	
Oxide / %	CaO	4.9	4.9	4.9	4.44	4.9
	K ₂ O	12.7	12.7*	12.7	12.59	12.7*
	Na ₂ O	2.4*	2.4*	2.4*	2.22	2.4*
	SiO ₂	60.0	60.0	60.0	60	60.0
	ZnO	3.6	3.6	3.6	3.7	4.25
	MgO	2.4	2.4	2.4	2.22	2.42
	SnO ₂	0.6	0.6	0.6	0.61	—
	Nb ₂ O ₅	12.1	12.1	—	—	12.1
	La ₂ O ₃	—	—	12.1	13.33	—
	SrO	1.2	1.2*	1.2*	1.48	1.2*
N_d target	—	1.68	—	1.63	1.68	
T_g target / °C	—	697	—	677	697	
N_d measured	—	1.70(±0.02)	—	—	1.70(±0.02)	
T_g measured / °C	685	682	670	—	680	

Notes: 1 means the properties of glass recommended by the GLAS algorithm are very similar to those of glass 15, * means source of nitrate.

过参数与模型优化后,可以得到预测精准的神经网络模型,最终使用GA对模型预处理后的数据进行训练。

训练结果如图8所示,可见大部分数据都均匀地分布于标识线两侧,且各类指标[决定系数(R^2)、相对偏差(RD)、均方根误差(RMSE)、平均绝对误差(MAE)、中位绝对误差(MedAE)]都准确地反映了该模型的拟合程度,折射率与化学成分具有合理的线性关系。事实上,一个简单的普通最小二乘线性回归(OLS)[46]的性能几乎与神经网络模型一样好,但OLS不足以捕捉玻璃化转变温度与成分的关系,所产生的模型明显比ANN差(R^2 为0.994,RD为6%,RMSE为62 K,MAE为45 K,MedAE为34 K),因此这里选择神经网络模

型可以提升准确性。使用的GA算法可以定义特定化学成分的玻璃基因表达,通过对特定化合物成分进行表示,便可以得到预测模型的输入,经过相关参数调节便可得到经遗传算法优化后的候选玻璃成分清单。图9是各化学元素预测残差的均值和标准差,可以看出,预测误差取决于玻璃中存在的化学元素,有些化学元素的数据容易比其他元素的数据产生更多的噪声。过渡金属(如V)和挥发性物质(如Pb)就是数据中可能有额外噪声的元素的例子,元素电荷或玻璃化学计量的未知变化会影响非晶结构,而非晶结构决定着玻璃的许多特性,一些化学元素的化学性质,如图9(a)中的Er和图9(b)中As,没有被各自的模型捕捉到。通过模型提

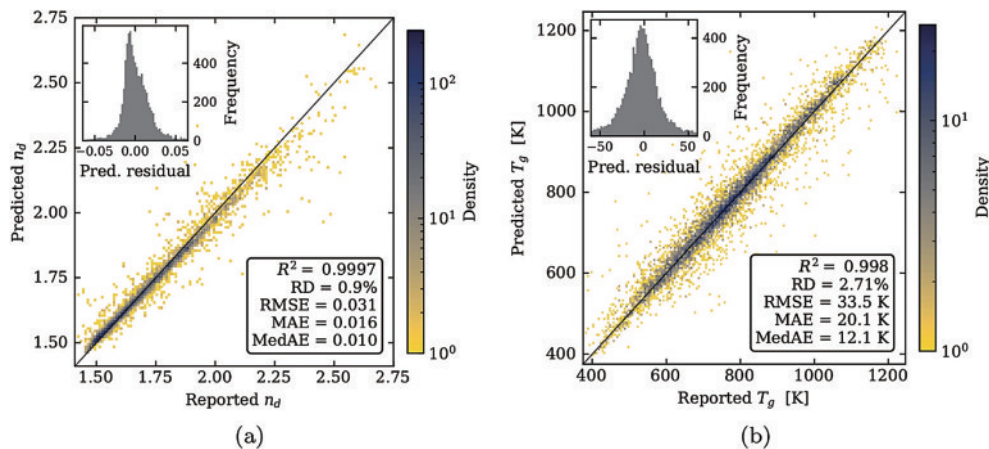


图 8 (a) 折射率和(b)玻璃化转变温度的预测值及其直方图(插图:预测残差的直方图,即报告值与预测值之间的差值,垂直的颜色条显示了数据点的密度)^[45]

Fig. 8 Predicted values and their histograms values of (a) refractive index and (b) glass transition temperature (inset: histogram of forecast residuals, the difference between reported and predicted values, the vertical color bars show the density of the data points)^[45]

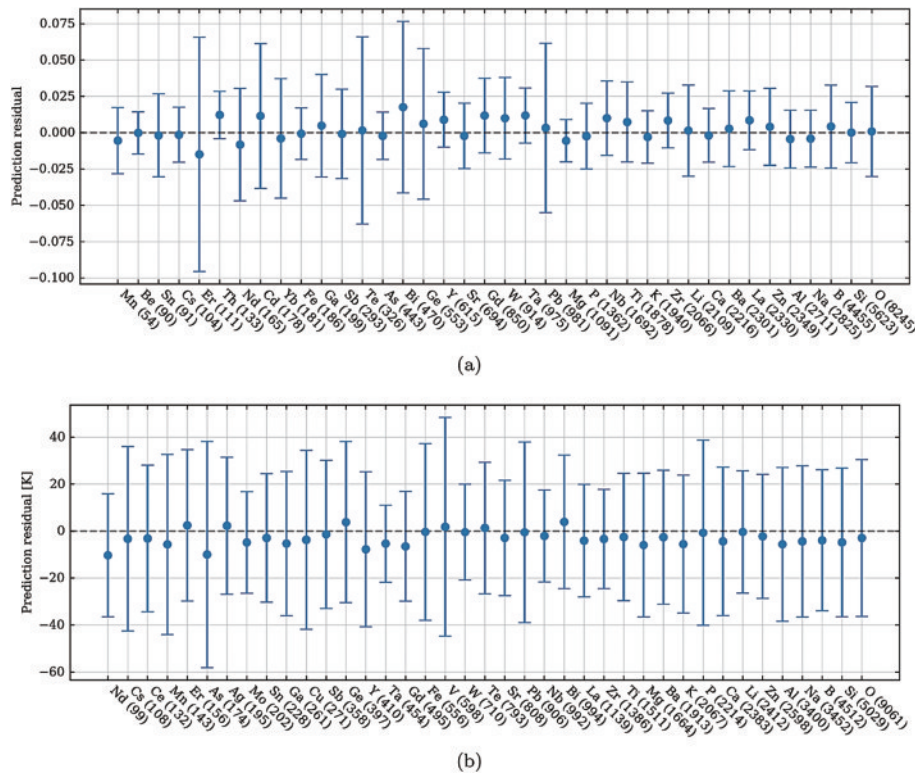


图9 数据集中每种化学元素的预测残差的均值和标准差^[45]。(a)折射率;(b)玻璃化转变温度(括号内的数字是保留数据集中含有该化学元素玻璃成分的数量,预测残差是报告值与预测值之间的差值,元素的顺序为从左到右,从少到多)

Fig. 9 Mean and standard deviation of predicted residuals for each chemical element in the data set^[45]. (a) Refractive index; (b) glass transition temperature (numbers in parentheses are the number of glass components containing the chemical element in the retained data set, and the predicted residual is the difference between the reported and predicted values, with the elements in order from left to right and from less to more)

供的信息,在实验室中就可以选择并生产由算法建议的候选组合,减少了实验的盲目性。

4 结束语

ML已成功应用于各类材料的研发之中,在光学玻璃制造领域也逐渐崭露头角,这意味着以数据驱动方式的材料学研究已经广泛展开,众多的研究实例再次证明其在材料研究领域的强大生命力。相较于传统的研究方法,基于ML技术的研究不仅在效率上得到了大幅提升,还有效加速了新材料的开发进程。这对于玻璃科学的发展同样具有重要意义。可以预见,在未来的研究中,将ML技术与玻璃性质研究相结合的前景将十分广阔。

随着ML技术在玻璃材料研究领域的应用不断开展,越来越多的研究实例证明了数据驱动方式在玻璃科学研究中的便利性与实用性。然而,尽管ML在材料学研究中的优势不断凸显,目前仍存在一些亟待解决的问题,如过拟合问题。由于数据驱动的研究方式依赖于独特的模式,ML算法不能很好地拟合实验数据,一方面是由于可用数据的缺乏,另一方面也与研究者对算法优化问题的研究欠缺有关。为解决过拟合问题,需要不断调整ML算法的参数,以适应不同的研究

需求^[47]。此外,好的模型预测精度不是一蹴而就的,仍需科研工作者的不断调试与探索。在模型构建中还存在一个大问题,即研究的算法模型缺乏可解释性。这是因为对于“ML+光学玻璃”的研究仍处于起步阶段,还没有一个可靠的理论可以解释算法的具体预测机制。因此,仅仅依靠纯数据驱动方式的ML模型无法得到合理的解释^[48]。此外,为提高模型预测精度,可以采取将基础研究理论与ML相结合的方式。这种“理论+数据”混合驱动的模式可以提高模型的预测精度,增强其泛化能力。同时,将基础研究理论与ML相结合,以提高模型预测精度和泛化能力的方法也将为该领域的发展提供重要的支持。

ML作为人工智能科学的重要组成部分,将其与材料学研究相结合无疑为玻璃开发工作提供了强有力的研究工具。在玻璃材料的研究中,ML可以被应用于光学玻璃材料的设计、优化和预测等方面。然而,当前的玻璃数据库的数据准确性和完整性有待提高,同时缺乏支持大型机器学习模型的数据集。此外,目前缺乏集成化的光学玻璃材料专用软件包,这种“散装”的代码组成形式妨碍了ML方法在玻璃研究中的推广。因此提出以下两点建议:1)建立高质量的光学玻璃数据库,这是ML在玻璃研究中发挥作用的先决条

件。数据库的开发可以在前人研究的基础上展开数据搜集与整理工作,将符合条件的数据进行分类与筛选,建议建立一个开源的数据库环境,以便各科研工作者进行数据分享与数据库更新。2) 开发一种集成化的预测和优化工具,可以推进 ML 在玻璃研究中的程序化发展,提高未来研究工作的便利性和准确性^[49]。目前已经产生了一些玻璃预测工具软件,但原理上大都基于传统的预测方式开展,在人工智能与数据科学发展的环境下,“ML+材料”是一种全新的玻璃开发模式,开发这样一种可被研究者广泛接受且容易使用的工具软件十分必要。在未来的研究中,需要解决玻璃数据库的数据准确性和完整性问题,同时开发集成化的光学玻璃材料专用集成软件,以便更好地发挥 ML 在玻璃研究中的作用。总的来说,ML 结合玻璃科学的研究工作已经成为未来材料研究领域的重要组成部分,对于 ML 在光学玻璃材料研究中出现的问题仍需进一步解决,对该项工作的研究还有很长的路要走。

参 考 文 献

- [1] Ballato J, Seddon A, Clare A, et al. Future of optical glass education[J]. *Optical Materials Express*, 2022, 12(7): 2626-2634.
- [2] Walasik W, Traoré D, Amavigan A, et al. 2- μm narrow linewidth all-fiber DFB fiber Bragg grating lasers for Ho- and Tm-doped fiber-amplifier applications[J]. *Journal of Lightwave Technology*, 2021, 39(15): 5096-5102.
- [3] Pincemin E, Jauffrit J, Disez P Y, et al. 12-core erbium/ytterbium-doped fiber amplifier for 200G/400G long-haul, metro-regional, DCI transmission applications with ROADMs[C]//2021 European Conference on Optical Communication (ECOC), September 13-16, 2021, Bordeaux, France. New York: IEEE Press, 2021.
- [4] Pan L, Ji S Z, Huang W F, et al. Joule-level twelve-pass LD end-pumped bonded neodymium glass laser amplifier[J]. *Photonics*, 2021, 8(4): 96.
- [5] Elisa M, Iordache S M, Iordache A M, et al. Peculiarities of the structural and optical properties of rare-earth-doped phosphate glasses for temperature sensing applications[J]. *Journal of Non-Crystalline Solids*, 2021, 556: 120569.
- [6] Weng K B, Long N B, Guo Y Q, et al. Nanocrystallization of $\alpha\text{-CsPbI}_3$ perovskite nanocrystals in $\text{GeS}_2\text{-Sb}_2\text{S}_3$ based chalcogenide glass[J]. *Journal of the European Ceramic Society*, 2020, 40(12): 4148-4152.
- [7] Calle-Vallejo F, Koper M T M. First-principles computational electrochemistry: achievements and challenges[J]. *Electrochimica Acta*, 2012, 84: 3-11.
- [8] Schleder G R, Padilha A C M, Acosta C M, et al. From DFT to machine learning: recent approaches to materials science-a review[J]. *Journal of Physics: Materials*, 2019, 2(3): 032001.
- [9] Zhou Q H, Lu S H, Wu Y L, et al. Property-oriented material design based on a data-driven machine learning technique[J]. *The Journal of Physical Chemistry Letters*, 2020, 11(10): 3920-3927.
- [10] Bishnoi S, Singh S, Ravinder R, et al. Predicting Young's modulus of oxide glasses with sparse datasets using machine learning[J]. *Journal of Non-Crystalline Solids*, 2019, 524: 119643.
- [11] Onbaşlı M C, Tandia A, Mauro J C. Mechanical and compositional design of high-strength coming gorilla® glass[M]//Andreoni W, Yip S. *Handbook of materials modeling: applications: current and emerging materials*. Cham: Springer, 2020: 1997-2019.
- [12] Ghorbani A, Askari A, Malekan M, et al. Thermodynamically-guided machine learning modelling for predicting the glass-forming ability of bulk metallic glasses[J]. *Scientific Reports*, 2022, 12: 11754.
- [13] Mahesh B. Machine learning algorithms-a review[J]. *International Journal of Science and Research*, 2020, 9: 381-386.
- [14] Osisanwo F Y, Akinsola J E T, Awodele O, et al. Supervised machine learning algorithms: classification and comparison[J]. *International Journal of Computer Trends and Technology*, 2017, 48(3): 128-138.
- [15] Glielmo A, Husic B E, Rodriguez A, et al. Unsupervised learning methods for molecular simulation data[J]. *Chemical Reviews*, 2021, 121(16): 9722-9758.
- [16] Hu Y, Bai X, Yang W J, et al. Concatenated dynamic reinforcement learning for multi-staged tasks (MST)[C]//Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering, October 22-24, 2021, Xiamen, China. New York: ACM, 2021: 781-788.
- [17] Pal M, Bharati P. Introduction to correlation and linear regression analysis[M]//Applications of regression techniques. Singapore: Springer, 2019: 1-18.
- [18] Yang Y M, Sun L L, Guo C R. Aero-material consumption prediction based on linear regression model[J]. *Procedia Computer Science*, 2018, 131: 825-831.
- [19] Maulud D, Abdulazeez A M. A review on linear regression comprehensive in machine learning[J]. *Journal of Applied Science and Technology Trends*, 2020, 1(4): 140-147.
- [20] Ranstam J, Cook J A. LASSO regression[J]. *British Journal of Surgery*, 2018, 105(10): 1348.
- [21] Hoerl R W. Ridge regression: a historical context[J]. *Technometrics*, 2020, 62(4): 420-425.
- [22] Katoch S, Chauhan S S, Kumar V. A review on genetic algorithm: past, present, and future[J]. *Multimedia Tools and Applications*, 2021, 80(5): 8091-8126.
- [23] Albadr M A, Tiun S, Ayob M, et al. Genetic algorithm based on natural selection theory for optimization problems[J]. *Symmetry*, 2020, 12(11): 1758.
- [24] Lambora A, Gupta K, Chopra K. Genetic algorithm-a literature review[C]//2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), February 14-16, 2019, Faridabad, India. New York: IEEE Press, 2019: 380-384.
- [25] Zhang Q M, Yu H Y, Barbiero M, et al. Artificial neural networks enabled by nanophotonics[J]. *Light: Science & Applications*, 2019, 8: 42.

- [26] Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: a scoping review[J]. *PLoS One*, 2019, 14(2): e0212356.
- [27] Feng J L, Lu S N. Performance analysis of various activation functions in artificial neural networks[J]. *Journal of Physics: Conference Series*, 2019, 1237(2): 022030.
- [28] Runge J, Zmeureanu R. Forecasting energy use in buildings using artificial neural networks: a review[J]. *Energies*, 2019, 12(17): 3254.
- [29] Wu Z H, Pan S R, Chen F W, et al. A comprehensive survey on graph neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(1): 4-24.
- [30] Karamad M, Magar R, Shi Y T, et al. Orbital graph convolutional neural network for material property prediction [J]. *Physical Review Materials*, 2020, 4(9): 093801.
- [31] Maurizi M, Gao C, Berto F. Predicting stress, strain and deformation fields in materials and structures with graph neural networks[J]. *Scientific Reports*, 2022, 12: 21834.
- [32] Louie S G, Chan Y H, da Jornada F H, et al. Discovering and understanding materials through computation[J]. *Nature Materials*, 2021, 20(6): 728-735.
- [33] Urata S, Nakamura N, Aiba K, et al. How fluorine minimizes density fluctuations of silica glass: molecular dynamics study with machine-learning assisted force-matching potential[J]. *Materials & Design*, 2021, 197: 109210.
- [34] Yang K, Xu X Y, Yang B, et al. Predicting the Young's modulus of silicate glasses using high-throughput molecular dynamics simulations and machine learning[J]. *Scientific Reports*, 2019, 9: 8739.
- [35] Hu Y J, Zhao G, Zhang M F, et al. Predicting densities and elastic moduli of SiO₂-based glasses by machine learning[J]. *NPJ Computational Materials*, 2020, 6: 25.
- [36] Liu H, Fu Z P, Li Y P, et al. Machine learning forcefield for silicate glasses[EB/OL]. (2019-02-09) [2023-02-06]. <https://arxiv.org/abs/1902.03486>.
- [37] Krishnan N M A, Mangalathu S, Smedskjaer M M, et al. Predicting the dissolution kinetics of silicate glasses using machine learning[J]. *Journal of Non-Crystalline Solids*, 2018, 487: 37-45.
- [38] Hamilton J P. Corrosion behavior of sodium aluminosilicate glasses and crystals[M]. Pennsylvania: The Pennsylvania State University, 1999.
- [39] Zaki M, Venugopal V, Bhattoo R, et al. Interpreting the optical properties of oxide glasses with machine learning and Shapely additive explanations[J]. *Journal of the American Ceramic Society*, 2022, 105(6): 4046-4057.
- [40] Mangalathu S, Hwang S H, Jeon J S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach[J]. *Engineering Structures*, 2020, 219: 110927.
- [41] 董国平, 万天择, 吴敏波, 等. 玻璃基因工程在激光玻璃等光功能玻璃领域的研究进展[J]. *激光与光电子学进展*, 2022, 59(15): 1516002.
- Dong G P, Wan T Z, Wu M B, et al. Recent applications of glass genetic engineering in laser glasses and other advanced optical glasses[J]. *Laser & Optoelectronics Progress*, 2022, 59(15): 1516002.
- [42] Menezes A D, Teixeira E P, Finzer J R D, et al. Machine learning-driven development of niobium-containing optical glasses[J]. *Research, Society and Development*, 2022, 11(9): e13811931290.
- [43] Varshneya A K. Review of 'SciGlass' database[J]. *American Ceramic Society Bulletin*, 1997, 76(5): 82-83.
- [44] Nishioka T. Glass fact database "interglad"[J]. *Ceramics Japan*, 1993, 28(8): 755-758.
- [45] Cassar D R, Santos G G, Zanotto E D. Designing optical glasses by machine learning coupled with a genetic algorithm[J]. *Ceramics International*, 2021, 47(8): 10555-10564.
- [46] Raposo F. Evaluation of analytical calibration based on least-squares linear regression for instrumental techniques: a tutorial review[J]. *TrAC Trends in Analytical Chemistry*, 2016, 77: 167-185.
- [47] Ying X E. An overview of overfitting and its solutions[J]. *Journal of Physics: Conference Series*, 2019, 1168: 022022.
- [48] Holzinger A, Kieseberg P, Weippl E, et al. Current advances, trends and challenges of machine learning and knowledge extraction: from machine learning to explainable AI[M]//International cross-domain conference for machine learning and knowledge extraction. Cham: Springer, 2018, 11015: 1-8.
- [49] Morgan D, Jacobs R. Opportunities and challenges for machine learning in materials science[J]. *Annual Review of Materials Research*, 2020, 50: 71-103.