

基于符号语义反向传播的熟料性能显式表征及反向优化

韩瑞琪¹, 张亮亮^{2,3}, 宁 帅¹, 王 琳^{2,3}, 袁景凌^{4,5}, 杨 波^{2,3}, 侯鹏坤¹, 李琴飞¹

(1. 济南大学, 山东省绿色与智能建筑材料重点实验室, 济南 250022;

2. 济南大学, 山东省泛在智能计算重点实验室, 济南 250022;

3. 泉城省实验室, 济南 250100; 4. 武汉理工大学, 计算机与人工智能学院, 武汉 430070;

5. 武汉理工大学, 硅酸盐科学与先进建材全国重点实验室, 武汉 430070)

摘 要: 水泥熟料的化学组成直接影响水泥的后期强度与质量稳定性, 是熟料质量判别与生产调控的关键依据。然而, 熟料化学组分与强度之间的显式定量关系尚不完全明确, 限制了性能成因显式解析。此外, 现有基于组成数据的强度预测研究多聚焦于影响因素到强度结果的单向预测, 难以同时实现在给定强度提升目标下对输入变量的反向调控。对此, 提出了基于符号语义反向传播的熟料性能显式表征及反向优化方法。首先, 构建涵盖不同来源熟料的化学组成-性能数据集, 然后基于遗传编程的符号回归方法建立熟料化学组成与 28 d 抗压强度之间的显式关系模型, 萃取出表达式树及其对应的显式方程。在候选表达式筛选过程中, 引入物理先验约束、模型复杂度与变量覆盖度联合引导的模型选择策略, 以增强模型的物理合理性与可解释性。进一步地, 引入目标输出区间驱动的语义反向传播, 将目标强度变化反推为输入变量的可行调节区间, 从而实现强度变化需求到熟料化学组分的反向优化。结果表明, 在 58 组熟料-性能数据上构建了可定量描述化学组分与强度关系的表达式树和显式方程, 并在测试集上表现出较好的预测精度, 测试集均方根误差为 1.409 MPa, 平均绝对误差为 1.126 MPa。反向优化结果表明, 模型可在给定目标强度提升条件下输出单变量可行调节区间和多变量协同优化方案, 例如期望输出强度由 55.4 MPa 提升 1.5~3.0 MPa, CaO 变动区间应增大 0.849%~1.640%等, 为水泥熟料强度预测、质量评价和配比反向调控提供了一种可解释的新方法。

关键词: 熟料性能预测; 显式表征; 反向优化; 符号回归; 物理先验约束

中图分类号: TP181; TQ172 文献标志码: A 文章编号: 0454-5648(2026)08-2644-17

网络出版时间: 2026-07-21



熟料是决定水泥强度形成与服役性能的核心中间产品, 其化学组成直接影响水化反应过程与后期力学性能^[1-2]。目前, 生产中普遍通过检测熟料化学成分, 并依据专家经验来调控原料配比, 以控制质量波动与生产成本。然而, 这种经验式调控模式存在人员依赖、粗放等问题, 最重要的是难以精确量化表征熟料质量的优劣, 影响了复用性以及精准控制^[3-4]。此外, 水泥行业是仅次于电力和钢铁的第三大碳排放源, 并且纵观水泥生产过程, 熟料生产在整个水泥制造过程中排放约 95%的 CO₂^[5-6]。减少对

熟料的使用并提升熟料利用效率, 是水泥行业低碳转型的重要方向^[7]。因此, 深入研究熟料化学组分与强度之间的定量表征, 赋能熟料质量评估与生产调控, 有望为水泥行业的提质、降本、减碳提供有力支撑。

围绕熟料组成与强度之间的关系, 已有研究主要从工艺指标、矿物组成计算和统计分析等角度展开^[8-10]。传统方法通常基于 CaO、SiO₂、Al₂O₃、Fe₂O₃ 等主要化学组分, 结合石灰饱和系数、硅率、铝率及 Bogue 公式推算熟料矿物组成^[11-12], 并进一步分

收稿日期: 2026-05-06。 修订日期: 2026-06-09。

基金项目: 国家自然科学基金青年项目(62403209); 硅酸盐科学与先进建材全国重点实验室开放基金原创项目(SYSJJ2025-02); 山东省自然科学基金(ZR2024QF021); 山东省高等学校青年创新团队计划(2024KJH104); 泉城省实验室科研项目(QCL20250304); 济南大学 2025 年青年教师学科交叉会聚建设项目(XKJC-202507); 全国建材行业重大科技攻关 揭榜挂帅 项目(2025JBGS07-01)。

第一作者: 韩瑞琪(1997), 男, 博士研究生。

通信作者: 张亮亮(1991), 男, 博士, 副教授。

Received date: 2026-05-06. Revised date: 2026-06-09.

First author: HAN Ruiqi (1997-), male, Doctoral candidate.

E-mail: 1251600672@qq.com

Correspondent author: ZHANG Liangliang (1991-), male, Ph.D., Associate Professor.

E-mail: ise_zhangll@ujn.edu.cn

析 C_3S 、 C_2S 、 C_3A 和 C_4AF 等矿物相对水泥强度发展的影响^[13]。这类方法物理意义明确, 便于工程应用, 能够在一定程度上揭示熟料组成、矿物形成与强度发展之间的内在联系。然而, 这些模型形式相对固定, 难以充分表征多组分耦合作用下强度形成的复杂非线性关系。

随着水泥生产过程数据的积累和人工智能方法的发展, 数据驱动模型逐渐被用于熟料及水泥基材料性能预测^[14-15]。模糊逻辑推理、支持向量机、随机森林、神经网络以及深度学习模型等方法^[16-19], 能够从复杂生产数据或化学组成数据中学习非线性映射关系, 并对抗压强度、凝结时间、流变性等性能指标进行预测^[20-22]。相比传统经验模型, 机器学习方法对非线性关系具有更强的拟合能力, 能够综合考虑多种组分之间的耦合效应^[23], 并在强度预测中取得了较高精度, 为熟料性能智能预测提供了高效的技术路径。

然而, 熟料化学组成与其抗压强度之间的显式定量关系仍未明晰^[3], 现有的黑盒模型难以获得可解释的显式数学表达式^[24], 制约了其对强度成因的直接解析。其次, 模型功能存在局限, 多数研究仅支持从组分到强度的正向预测^[25], 难以同时根据当前所推理强度以及给定强度提升目标(数值或范围), 反向推导熟料化学组分的合理调节方向和可行区间。因此, 如何同时实现强度精准预测、显式关系表达和组分反向优化, 仍是熟料性能建模中的关键问题。

针对上述问题, 本工作提出一种基于符号语义反向传播(Semantic Backpropagation, SBP)^[26]的熟料性能显式表征及反向优化方法。首先, 构建熟料化学组成-28 d 抗压强度数据集; 其次, 采用遗传编程符号回归^[27]建立化学组成与强度之间的显式关系模型, 过程中结合物理先验、模型复杂度和变量覆盖度筛选候选表达式, 得到最优树结构表达式方程及其对应数学公式; 最后, 引入目标输出区间驱动的语义反向传播方法, 将目标强度提升反推为化学组分的可行调节区间。该方法可为熟料强度预测、质量评价和配比反向调控提供可解释的新途径。

1 实验

1.1 原材料与样本来源

本研究收集 58 种不同区域、厂商及产线的熟料样品作为研究对象。各熟料样品分别用于后续 X 射线荧光光谱仪(XRF)化学组成表征及胶砂强度试

验。制备水泥样品所用石膏为二水硫酸钙, 购自山东科源生化有限公司, 纯度为 99%。胶砂实验所用标准砂购自山东宏盛源建材有限公司, 为符合国家标准要求的水泥强度检验标准砂^[28]。各熟料样品均在相同制样条件下制备水泥样品, 以尽可能保证不同样品之间的可比性。

1.2 胶砂制备与抗压强度测试方法

胶砂试样制备与强度测试参照 GB/T 17671—2021《水泥胶砂强度检验方法》进行^[28]。胶砂试样的灰砂(水泥和沙子)质量比为 1:3, 由于本研究将实际制样时采用的水灰(水和水泥)质量比作为建模输入变量之一, 因此各样本的水灰质量比(W/C)并不固定, 取值范围控制在 0.47~0.53 之间。试件尺寸为 40 mm×40 mm×160 mm。试件成型后置于标准养护条件下养护, 养护温度为 20 °C, 相对湿度不低于 90%, 成型 20 h 后脱模, 并继续养护至规定龄期。

28 d 龄期时进行抗压强度测试, 测试时间误差控制在±8 h 以内。每组试样先进行抗折试验, 再对抗折后的 6 个半截试件进行抗压试验, 以 6 个抗压结果的平均值作为该样品的 28 d 抗压强度。当单个测试值与平均值偏差超过 10%时, 按标准要求进行剔除处理。抗压强度单位为 MPa, 结果保留至 0.1 MPa。

1.3 熟料化学组成测试

为获取熟料样品的化学组分信息, 采用压片 XRF 方法对各熟料样品进行表征。熟料样品首先经 HBM-PLUS/II 型震动磨(济南海博实验室仪器有限公司)研磨, 研磨转速为 1300 r/min, 研磨时间为 90 s。随后采用 SY-24 型手动压片机(济南海博实验室仪器有限公司)进行压片, 模具直径为 95 mm, 样品面压为 21.5 MPa, 保压时间为 1 min, 对应压力机推力约 15 t。

压片完成后, 使用 HS-XT6 型 XRF 设备(山东海石仪器科技有限公司)对熟料样品进行测试。结合建模需求, 最终选取 CaO 、 SiO_2 、 Al_2O_3 、 Fe_2O_3 、 MgO 、 SO_3 、 K_2O 与 Na_2O 作为主要表征变量。

1.4 数据集构建与划分

基于熟料样品的 XRF 测试结果及对应胶砂 28 d 抗压强度, 构建熟料化学组成-性能数据集。考虑到模型复杂度控制与熟料质量评价的实际需求, 未将全部 XRF 化学组成测试结果作为输入, 选取熟料生产与质量分析中较为关注的 CaO 、 SiO_2 、 Al_2O_3 、 Fe_2O_3 、 MgO 、 SO_3 、 K_2O 、 Na_2O 以及 W/C 作为输入变量, 以 28 d 抗压强度作为输出变量, 上述氧化物变量均为 XRF 测得的质量分数。最终构建的数据

集共包含 58 条样本。

数据集按照 0.85 : 0.15 划分训练集和测试集，其中训练集 49 条，测试集 9 条。为尽可能保持训练集和测试集划分的数据保持独立同分布，即在目标变量分布上的一致，首先对 28 d 抗压强度进行等频分层，保持数据分层情况下训练集和测试集在每层均有代表性样本，再在各分层内按照比例随机抽样划分训练集和测试集，训练集内采用五折交叉验证，以提高样本量较少情况下的数据划分的稳定性和代表性^[29–30]。

2 方法

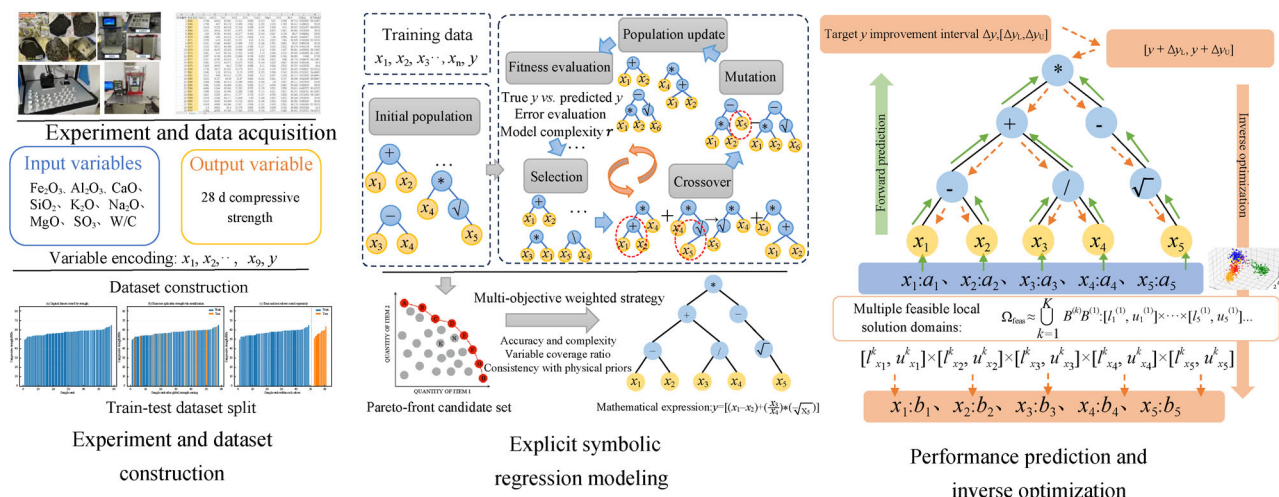
为建立基于化学组分测试数据的熟料强度显式关系，并进一步实现目标输出驱动的输入反向优化。方法部分由问题定义、显式建模、区间反推与优化的三部分构成。首先，将熟料化学组分信息及 W/C 与 28 d 强度之间关系建模问题形式化为符号回归任务。其次，利用遗传编程生成候选表达式，并结合物理先验约束、模型复杂度及变量覆盖度进行表达

式树筛选。最后，针对给定目标强度变化需求，构建输入变量的区间反推与优化方法。

2.1 问题定义与总体框架

本工作以熟料化学组成信息及 W/C 作为输入变量，以 28 d 抗压强度为输出变量，构建熟料表征信息–性能映射关系。设 $\mathbf{X}$ 表示全部样本输入矩阵，第 i 个样本的输入向量为 $\mathbf{X}_i$ ，输入强度为 y_i ，则建模任务可表达为由数据集 $\mathcal{D} = \{(\mathbf{X}_i, y_i)\}_{i=1}^N$ 学习显式函数 $f(\mathbf{X})$ ，使其能够较好地表征输入变量与 28 d 强度之间的关系。

在此基础上，本工作进一步考虑目标输出驱动的反向优化问题。对于当前输入 $\mathbf{X}^*$ 及其对应预测输出 $y^* = f(\mathbf{X}^*)$ ，当给定目标强度变化区间 Δy 时，目标是求得满足工艺边界约束的输入变量可行变化区间或者协同调整方案，使调整后的输入落入目标区间。基于上述目标，本工作构建了熟料化学组成–显式建模–目标驱动反向优化的总体研究框架，完整工作框架如图 1 所示。



Note: x_i is the i th input variable; x_1 – x_9 represent Fe₂O₃, Al₂O₃, CaO, SiO₂, K₂O, Na₂O, MgO, SO₃ and W/C, respectively; y is 28 d compressive strength; a_i and b_i denote the initial and optimized values of x_i ; Δy , Δy_L and Δy_U denote the target improvement interval and its lower and upper bounds, respectively; l_{xi} and u_{xi} denote the lower and upper bounds of the feasible interval of x_i ; $B^{(k)}$ and Ω_{feas} denote the k th feasible local solution box and the feasible solution domain, respectively.

图 1 基于符号回归与语义反向传播的熟料强度预测及反向优化总体流程

Fig. 1 Overall workflow of clinker strength prediction and inverse optimization based on symbolic regression and semantic backpropagation

2.2 强度显式建模

2.2.1 符号回归问题定义与目标函数 符号回归是一类从数据中自动发现显式数学表达式的机器学习方法，其目标不是在预设函数形式下仅估计参数，而是在由变量、常数和数学算子构成的表达式空间中同时搜索模型结构及其参数，使所得模型在预测精度与表达式复杂度之间折中，从而兼具可解释性与一定的反推能力。为获得可解释且可反推的熟料

表征信息–强度关系模型，本工作将建模任务形式化为符号回归问题，设候选表达式为 $f(\mathbf{X}; s, \theta)$ ，其中 s 为表达式结构， θ 表示待估参数，则可以将符号回归问题表示为式(1)。

$$\min_{s, \theta} \mathcal{L}[f(\mathbf{X}; s, \theta), y] + \lambda C(s) \quad (1)$$

其中： $\mathcal{L}$ 为预测误差项， $C(s)$ 为表达式复杂度项； λ 为复杂度权重系数。即在所有候选表达式结构 s

和对应参数 θ 中, 寻找一组使目标函数最小的最优结构和参数。

相较于黑盒模型, 基于符号回归方法构建显式表达式不仅能够给出输入变量与输出变量之间的定量关系, 而且便于后续开展变量灵敏度及关系分析与目标输出区间反推, 因此更适合本工作所关注的目标驱动反向优化问题。

2.2.2 基于遗传编程的表达式生成与实现 本工作采用基于遗传编程的符号回归方法进行候选表达式搜索, 工程实现基于 Operon/PyOperon^[31]。候选模型以表达式树形式表示, 内部节点为算子, 叶节点为输入变量或常数。本工作使用的算子集合包括加、减、乘、受保护除法及部分受保护一元算子, 以保证候选表达式在搜索与后续反推过程中的数值稳定性。

在搜索过程中, 通过选择、交叉、变异与精英保留等遗传操作, 不断产生新的候选表达式, 并在误差-复杂度多目标框架下形成帕累托前沿候选集。为提高参数拟合能力, 候选表达式中的常数项在搜索过程中进一步结合局部优化进行调整。最终得到一组在预测误差和模型复杂度之间具有不同权衡关系的候选表达式, 为后续模型筛选提供基础。

2.2.3 物理先验约束与变量覆盖度联合引导的模型筛选 在构建数据集小样本条件下, 仅依据预测误差和表达式复杂度进行模型筛选, 容易得到虽拟合性能较好但物理意义不足或关键输入覆盖不充分的候选表达式。考虑到本工作后续仍需基于显式模型开展输入区间反推, 因此最终模型不仅应具有较好预测能力, 还应尽可能满足基本材料规律并保持结构完备性。

基于此, 本工作在候选表示式筛选阶段, 引入由交叉验证误差、表达式长度、物理先验违例率及变量覆盖度共同构成的综合评价策略。具体而言, 物理先验约束主要用于刻画候选表达式是否符合水泥材料的基本认知, 如输出强度的合理范围、部分关键变量作用方向的一致性; 表达式长度用于约束模型复杂度; 变量覆盖度用于约束表达式构成的变量数量, 避免关键输入变量缺失带来的结构不完备问题。通过联合考虑上述指标, 对帕累托前沿候选表达式进行择优, 最终确定用于后续强度预测与反向优化的显式模型。综合评价策略表示如式(2)所示。

$$\mathcal{J}(f) = \text{CVErr}(f) + \alpha \mathcal{P}_{\text{phy}}(f) + \beta \mathcal{P}_{\text{len}}(f) + \gamma \mathcal{P}_{\text{cov}}(f) \quad (2)$$

其中: f 表示符号回归过程中生成的一个候选表达式函数; $\mathcal{J}(f)$ 为综合评价总评分; $\text{CVErr}(f)$ 为交叉验证误差项; $\mathcal{P}_{\text{phy}}(f)$ 为物理先验约束项, $\mathcal{P}_{\text{len}}(f)$ 为表达式长度约束项; $\mathcal{P}_{\text{cov}}(f)$ 为变量覆盖度约束项; α 、 β 、 γ 分别为对应物理约束项、表达式长度约束项、变量覆盖度约束项权重系数。

设候选表达式为 f , 当前折验证集为 $\mathcal{D}_v$, 其样本数为 m_{val} 。 $\text{CVErr}(f)$ 采用验证集均方误差 (MSE_{val}) 来作为主误差项以优化模型。 $\mathcal{P}_{\text{phy}}(f)$ 的计算包括 2 个主要过程:

1) 计算原始累计违例分数 (P_{raw})。 P_{raw} 由输出强度合理范围、W/C 负相关、CaO 局部正方向一致性及 MgO 过强正向响应限制共同组成。当候选表达式出现预测强度小于 0 或者大于 150 MPa, 即 $f(\mathbf{X}_i) < 0$ 或 $f(\mathbf{x}_i) > 150$ 记为 1 次加权物理先验违例。此外, 对变量 j 的局部响应采用有限差分 $\Delta_j f(\mathbf{X}_i) = f(\mathbf{X}_i + \varepsilon \mathbf{e}_j) - f(\mathbf{X}_i)$ 计算, 其中, $\varepsilon = 10^{-4}$ 。当 $\Delta_{\text{W/C}} f(\mathbf{X}_i) > 0$ 、 $\Delta_{\text{CaO}} f(\mathbf{X}_i) < 0$ 、 $\Delta_{\text{MgO}} f(\mathbf{X}_i) > 0.1$ 时, 即 W/C 增加导致强度异常升高、CaO 局部响应方向与材料学认识不一致、MgO 出现过强正向响应, 也分别记为 1 次加权物理先验违例。根据不可接受程度, 上述加权先验违例依次按照权重 10、5、2 和 1 计入惩罚并累计至 P_{raw} 。

2) 为避免因不同交叉验证折的验证集样本数目不同而造成物理惩罚不可比, 将原始物理惩罚除以该折验证集样本数, 得到单位样本物理惩罚, 即 $\mathcal{P}_{\text{phy}}(f) = P_{\text{raw}} / m_{\text{val}}$ 。

表达式长度惩罚项定义为 $\mathcal{P}_{\text{len}}(f) = \frac{L(f)}{L_{\text{max}}}$, 其中: $L(f)$ 为表达式长度, $L_{\text{max}} = 128$ 。变量覆盖度惩罚项定义为 $\mathcal{P}_{\text{cov}}(f) = 5(p - |V_f|)^2$, 其中: $p = 9$ 为输入变量总数, $|V_f|$ 为候选表达式实际包含的输入变量集合。最终式(2)可以进一步表示为式(3)。

$$\mathcal{J}(f) = \text{MSE}_{\text{val}}(f) + \alpha \frac{P_{\text{raw}}(f)}{m_{\text{val}}} + \beta \frac{L(f)}{L_{\text{max}}} + \gamma [5(p - |V_f|)^2] \quad (3)$$

其中: $\alpha = 5$ 、 $\beta = 1$ 、 $\gamma = 2$ 。评分越小表示候选模型综合性能越优, 该评分设计以交叉验证误差为主导, 同时对明显物理违例、表达式复杂度和关键变量缺失进行软惩罚, 以保证最终表达式兼具预测精度、物理合理性、结构简洁性和反向优化可用性。

2.2.4 固定显式表达式的 Ridge 参数校准优化 为进一步提升最终显式模型的预测精度, 本工作在已筛选得到的符号回归表达式基础上, 引入固定树结构的岭回归(Ridge Regression)^[32]参数校准优化方法。针对符号回归寻优的显式表达式, 保持其表达式树结构不变, 在其输出基础上引入少量关键变量进行线性参数校准优化。考虑 CaO、SiO₂ 和 W/C 分别与熟料硅酸盐矿物形成及水泥浆体空隙结构密切相关, 本工作选取 $w(\text{CaO})$ 、 $w(\text{SiO}_2)$ 和 W/C 作为指定校准变量, 构造校准输入矩阵[式(4)]。

$$Z = [f_{\text{SR}}^*(x), w(\text{CaO}), w(\text{SiO}_2), (W/C)] \quad (4)$$

随后采用岭回归建立校准优化模型[式(5)]。

$$\hat{y} = b + af_{\text{SR}}^*(x) + \beta_1 w(\text{CaO}) + \beta_2 w(\text{SiO}_2) + \beta_3 (W/C) \quad (5)$$

式中: $\hat{y}$ 为经过 Ridge 参数校准优化后的最终预测强度值; $f_{\text{SR}}^*(x)$ 为符号回归筛选得到的固定显式表达式在输入样本 x 下的计算输出值; $w(\text{CaO})$ 与 $w(\text{SiO}_2)$ 分别为熟料中 CaO 和 SiO₂ 的质量分数; b 、 a 、 β_1 、 β_2 、 β_3 分别对应关键变量的待优化参数。

2.3 目标驱动的输出区间反推与优化

2.3.1 符号语义反向传播基本思想 为实现由目标强度变化反推输入变量调节范围, 本工作采用符号语义反向传播思想^[26], 在表达式结构上自顶向下传播目标输出区间。对于给定显式模型 $f(\mathbf{X})$ 和目标输出区间 $[y_L^*, y_U^*]$, 其中 y_L^* 和 y_U^* 分别表示目标输出区间的上界和下界, SBP 在每个算子节点处依据该算子的局部逆映射关系, 将父节点允许输出区间逐层映射为子节点可行取值区间, 直至叶节点变量层。

与仅依赖局部偏导方式的一阶近似方法不同, SBP 直接作用于表达式结构本身, 能够显式处理非线性算子及其定义域限制, 从而输出变量的可行区间集合, 而不仅仅是变化方向, 为目标输出驱动的输入区间反推生成提供了基础。

2.3.2 前向-反向区间收缩 前向-反向区间收缩^[33]可视为区间约束传播中的一种域收缩方法。其基本思想是先利用区间算术沿表达式树进行前向传播, 得到中间变量和输出变量的区间包络, 再根据等式或不等式约束进行反向传播, 从而收缩输入变量的可行区间且不剔除满足约束的解。考虑到单次区间反推在高阶、非单调或者变量重复出现的表达式中容易产生较宽的可行包络, 本工作在 SBP 基础上进一步构建前向-反向区间收缩过程。具体而言,

首先在当前输入盒约束下对表达式进行前向区间传播, 得到该输入盒可能对应的输出区间; 随后将其与目标输出区间求交, 保留有效输出部分; 再以该有效输出区间为根语义约束, 沿表达式树执行 SBP 反向传播, 对输入盒进行收缩。通过重复执行前向传播、目标求交、反向传播与收缩, 直至盒变化趋于稳定, 从而获得更为紧致的输入可行域。

2.3.3 条件分裂与联合可行域逼近 由于区间反推在复杂显式模型中仍可能保留较宽的盒状外逼近, 即求得可行区域盒内存在多数区域满足输出变化对应要求, 但是由于可行盒区域较宽, 内部仍包络部分子区域不满足或者超出变化要求。为进一步提高联合可行域定位精度, 本工作在前向-反向收缩基础上引入条件分裂策略。在当前输入盒多轮收缩后仍较宽且收缩停滞时, 沿圈定变量维度对该盒进行二分, 并对所得各个子盒重复执行前向-反向收缩与可行性判定。通过收缩-分裂-再收缩的递归过程, 逐步以多个局部小盒逼近满足目标输出区间的输入联合可行域。最终结果不再表示为各变量孤立区间, 而是表示为各变量存在联合关系的若干局部可行盒的并集。

2.3.4 单变量与多变量优化方案提取 在获得输入联合可行域后, 本工作进一步提取单变量和多变量条件下的调节建议。对于单变量情形, 固定其余变量当前点, 仅保留目标变量的工艺边界, 通过联合可行盒在该变量维度上的投影得到可行变化区间, 并结合局部灵敏度信息给出变化方向建议。对于多变量情形, 则在联合可行盒多轮迭代构造候选调整解集, 并依据最小改动原则, 在满足目标输出区间的前提下选择改变量较小的候选解。若引入权重, 可进一步通过加权范式反映不同变量在调节成本、可操作性或者工艺优先级上的差异, 从而得到更符合工程需求的协同优化方案。

3 结果与讨论

为了验证本工作所提出方法在熟料化学组成-强度显式建模及目标预测驱动反向优化中的有效性, 本章基于 58 组熟料样本信息, 从数据分布特征、候选表达式筛选、化学组成及 W/C-强度关系分析, 单变量与多变量反向优化结果等方面进行讨论。首先, 对样本的化学组成信息特征及强度数据分布进行统计分析; 随后, 对候选表达式筛选过程及最终模型进行展示, 并结合显式模型讨论输入变量与强度之间的关系; 在此基础上, 进一步给出目标驱动

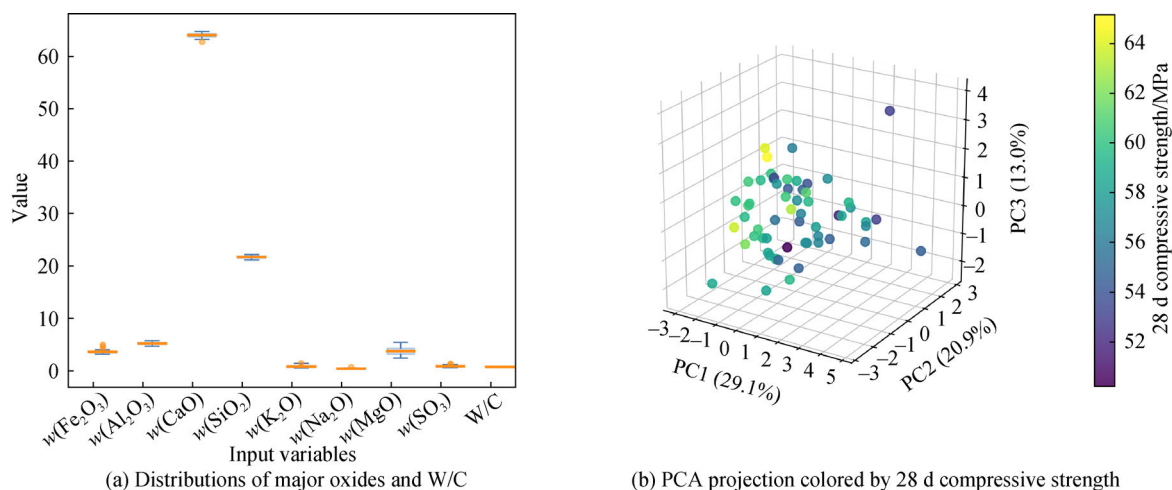
条件下的单变量和多变量反推结果, 并通过采样验证评估方法可靠性。

3.1 样本特征与强度分布

为分析样本的整体化学组成特征及其与强度分布的基本情况, 对 58 组熟料样本的主要氧化物组成、W/C 及 28 d 抗压强度进行可视化分析, 结果见图 2 与图 3。由图 2a 可见, 9 个输入变量在样本间存在一定波动范围, 样本在化学组成和工艺条件上具有必要差异。图 2b 的样本输入的主成分分析 (Principal Component Analysis, PCA)^[34] 三维投影结

果进一步表明, 58 组样本在输入空间中呈分散分布, 表明当前数据具有一定多样性, 可为后续显式建模提供基础。

图 3a 显示 28 d 抗压强度在样本间存在一定离散性, 能够为化学组成-性能关系学习提供必要的输出变化信息。图 3b 表明采用强度分层后训练集与测试集在各分层区间内均有分布, 且整体保持目标变量分布的一致性。综上, 当前数据集在输入和输出两方面均具备开展后续模型构建与反向优化的基本条件。



Note: In Fig. 2a, the orange horizontal lines in the boxes denote median values, and the orange dots denote outliers beyond the whiskers of the box plots; w is mass fraction; And in Fig. 2b, PC denotes principal component; PC1, PC2 and PC3 represent the first, second and third principal components, respectively, and the percentages in parentheses indicate the explained variance ratios.

图 2 熟料样本输入特征分布及 PCA 三维投影

Fig. 2 Distribution of clinker input features and three-dimensional PCA projection

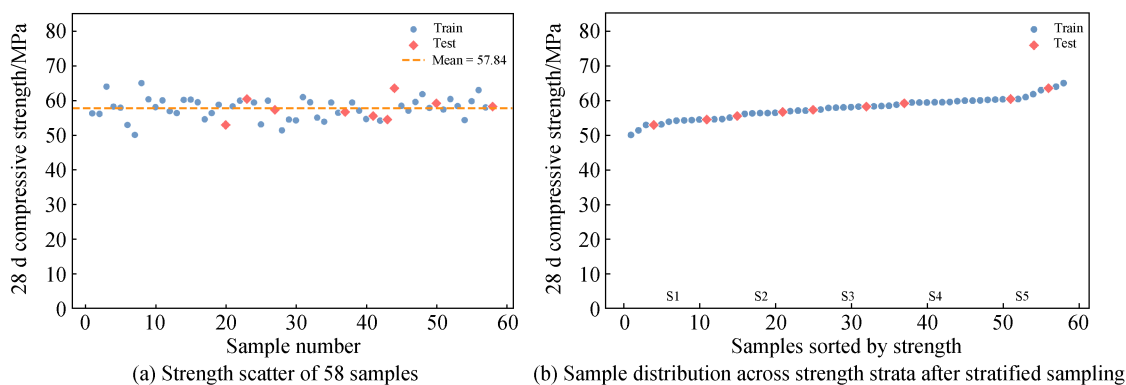


图 3 28 d 抗压强度分布及训练-测试集分层划分结果

Fig. 3 Distribution of 28 d compressive strength and stratified train-test split result

3.2 候选表达式筛选与最终模型确定

基于 2.2.2 节所述遗传编程符号回归框架, 本工作首先在训练集内部开展五折交叉验证建模, 以降低小样本条件下单次划分带来的偶然性影响。具体而言, 将训练集进一步划分为 5 个子集, 每次选取

其中 4 折作为训练子集、1 折作为验证子集, 重复进行 5 次训练与验证。对于每一折, 符号回归算法均可生成一组位于误差-复杂度帕累托前沿上的候选表达式。不同候选树在表达式长度、训练拟合能力、变量覆盖程度及验证集误差等方面存在明显差

异,因此最终模型的确定不能仅依据单一误差指标,而需要在多项性能之间进行综合权衡。

考虑到本工作研究目标不仅是建立熟料化学组成与 28 d 抗压强度之间的显式定量关系,而且还需依托所建模型进行目标驱动的输入区间反推与优化,因此最终模型应同时满足以下要求: 1) 具有较好的预测性能; 2) 保持相对可控的结构复杂度,以增强模型可解释性; 3) 尽可能符合材料学常识和物理先验约束; 4) 尽量覆盖全部关键输入变量,为后续单变量和多变量联合反向优化提供完整表达基础。由此可见,在本工作场景下,模型选择不能简单选择误差最小,而应综合考虑并将目标设计为预测–可解释–可反推的综合研究目标。

图 4 展示了候选表达式的筛选过程及其分布特征。图 4a 为最终入选折对应训练子集的帕累托前沿,横轴为表达式长度,纵轴为对应折的训练子集决定系数 R^2 。随着表达式长度增加,模型训练拟合能力总体提升,但复杂度增加并不一定带来同等幅度的性能改善,说明误差降低与结构复杂度之间可能存在折中关系。图 4b 汇总展示了五折交叉验证过程中通过预筛选的候选表达式分布情况,横轴为表达式长度,纵轴为交叉验证均方根误差(RMSE),点

颜色表示物理先验惩罚强度,星形标记表示最终选定模型。由图 4 可见,部分候选树虽然具有较低的验证误差,但物理惩罚较高,说明其存在一定的物理先验违背; 另一些候选树虽然结构较短,但在变量覆盖或综合性能方面不足,难以满足后续反向优化分析需求。因此,仅依据误差–复杂度两项指标难以完成本工作场景下的最终模型确定。

在帕累托前沿候选树基础评判标准上,进一步引入由交叉验证误差、物理先验惩罚、表达式长度惩罚及变量覆盖度惩罚构成的综合评分函数,对各折候选模型进行联合筛选。筛选过程中,首先优先保留变量覆盖程度较高的候选表达式,再在此基础上综合比较其验证误差、物理先验惩罚与复杂度水平,从而兼顾模型表达完整性与综合性能。每一折均据此选出一批代表性最优候选模型,随后再从 5 个折最优模型中确定综合评分最优者作为最终模型。表 1 列出了若干代表性候选模型的比较结果。结果表明,最终入选模型虽然并非全部性能最优,但其在预测性能、物理合理性、模型复杂度及变量覆盖完整性之间取得了更为均衡的折中,即表现出最低综合评分 15.889,尤其能够覆盖全部关键输入变量,为后续区间回推和联合优化分析提供了必要前提。

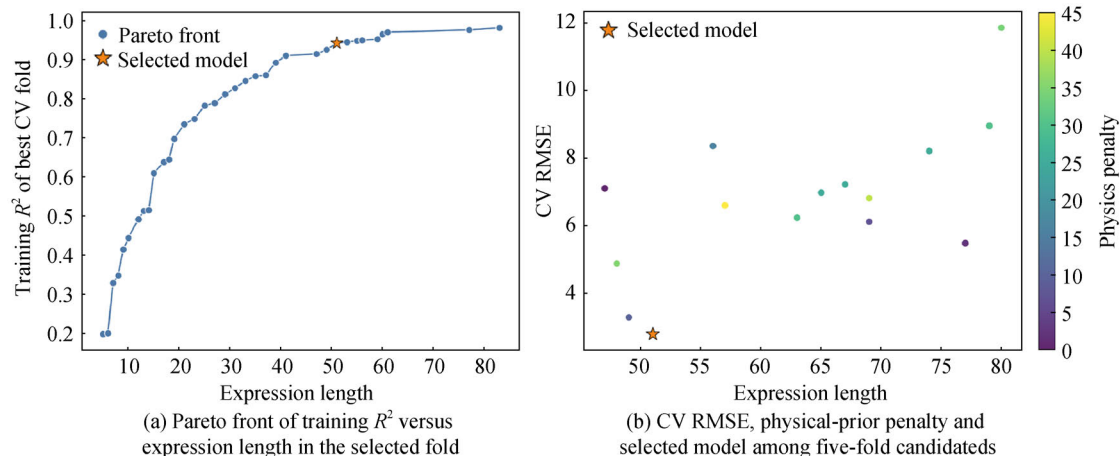


图 4 符号回归候选表达式的误差–复杂度分布与综合筛选结果

Fig. 4 Error-complexity distribution and integrated selection of candidate symbolic expressions

表 1 代表性候选符号回归模型的综合筛选指标对比

Table 1 Comparison of integrated selection metrics for representative candidate symbolic-regression models

ID	Variable coverage	Expression length	CV RMSE	Physical-prior penalty	Integrated score	Remarks
E1	9/9	51	2.777	14.0	15.889	Final selected model
E2	9/9	49	3.269	10.0	16.622	Second-best model
E3	9/9	77	5.472	2.0	31.656	Lowest physical penalty
E4	9/9	56	8.339	16.0	77.976	Representative full-variable model
E5	9/9	69	6.098	7.0	41.228	Representative full-variable model
E6	8/9	48	4.865	35.0	51.544	Insufficient variable coverage

最终选定的显式模型树结构如图 5 所示, 为了提高公式可读性, 本工作将符号回归所得原始长表达式按层次结构重写为式(6)~式(13)。由此可见最终模型并非简单线性组合, 而是由乘积项、比值项及平方根型非线性项共同组成, 能够反映熟料主要氧

化物组成及 W/C 对 28 d 抗压强度的耦合作用。模型同时包含 $w(\text{CaO})$ 、 $w(\text{SiO}_2)$ 、 Al_2O_3 质量分数 $[w(\text{Al}_2\text{O}_3)]$ 、 Fe_2O_3 质量分数 $[w(\text{Fe}_2\text{O}_3)]$ 、 MgO 质量分数 $[w(\text{MgO})]$ 、 SO_3 质量分数 $[w(\text{SO}_3)]$ 、 K_2O 质量分数 $[w(\text{K}_2\text{O})]$ 、 Na_2O 质量分数 $[w(\text{Na}_2\text{O})]$ 以及 W/C 等

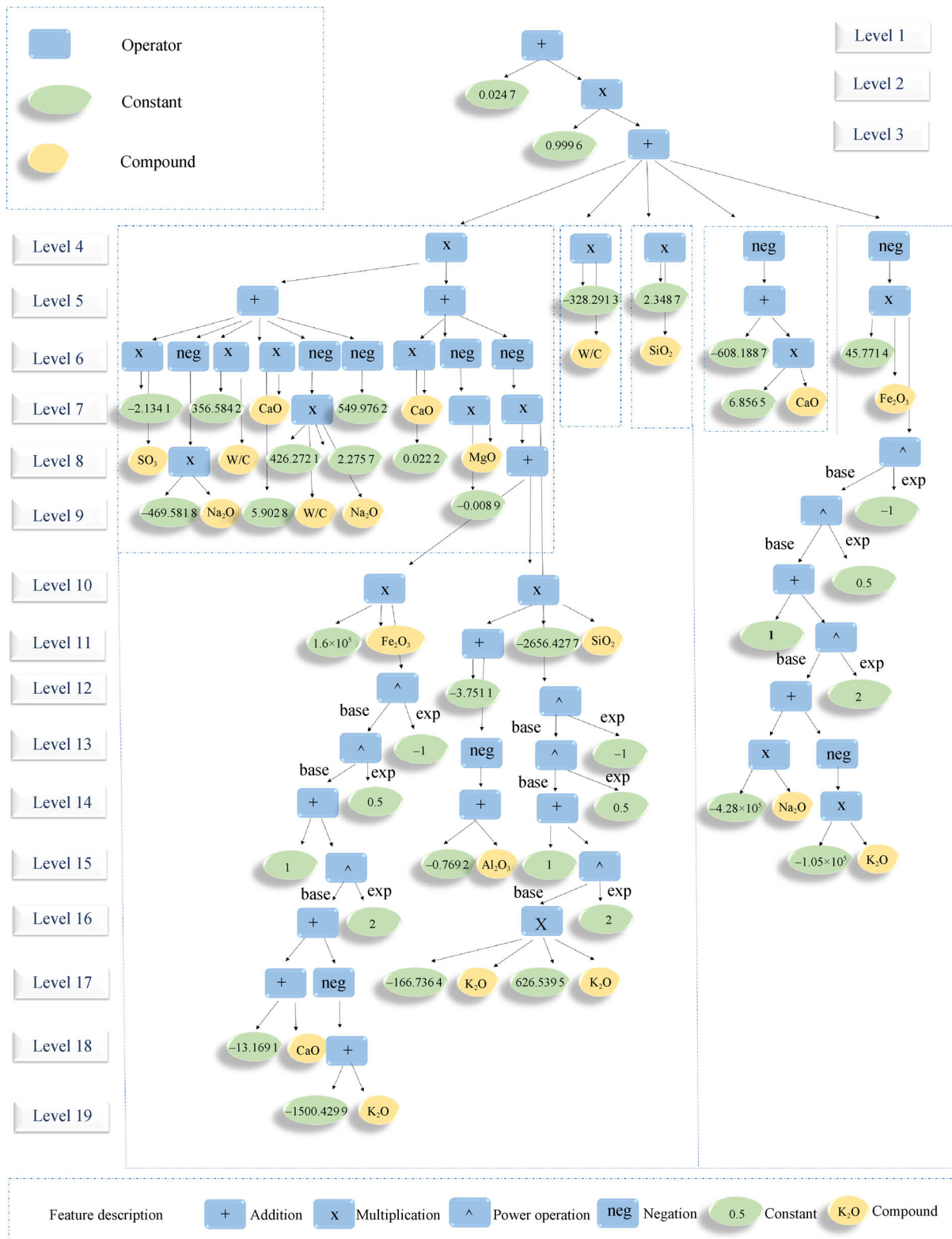


图 5 最终选择的符号回归模型的表达式树结构

Fig. 5 Expression tree structure of the selected symbolic-regression model

全部关键输入变量, 为后续输入区间回推与联合优化提供了显式表达基础。树结构数学表达式如式(6)所示。

$$\hat{f}_{28} = 0.0247 + 0.9996(M_1 M_2 + M_3 - M_4) \quad (6)$$

$$M_1 = -2.1341w(\text{SO}_3) + 469.5815w(\text{Na}_2\text{O}) + 356.5842(\text{W/C}) + 5.9028w(\text{CaO}) - 970.0669(\text{W/C})w(\text{Na}_2\text{O}) - 549.9764 \quad (7)$$

$$M_2 = 0.02224w(\text{CaO}) + 0.008916w(\text{MgO}) - \frac{M_{21} + M_{22}}{M_{23}} \quad (8)$$

$$M_{21} = \frac{[1.06496 \times 10^5 w(\text{Fe}_2\text{O}_3)]}{\sqrt{1 + [1500.4299w(\text{K}_2\text{O}) - 13.1691w(\text{CaO})]^2}} \quad (9)$$

$$M_{22} = [9964.6403 - 2043.3162w(\text{Al}_2\text{O}_3)]w(\text{SiO}_2) \quad (10)$$

$$M_{23} = \sqrt{1 + \{1.04467 \times 10^5 [w(\text{K}_2\text{O})]^2\}^2} \quad (11)$$

$$M_3 = -328.2913(\text{W/C}) + 2.3487w(\text{SiO}_2) + 608.1887 - 6.8565w(\text{CaO}) \quad (12)$$

$$M_4 = \frac{45.7714w(\text{Fe}_2\text{O}_3)}{\sqrt{1 + [105294.4531w(\text{K}_2\text{O}) - 428086.3125w(\text{Na}_2\text{O})]^2}} \quad (13)$$

其中: M_1 、 M_2 、 M_3 和 M_4 为更直观展示表达式的整体结构而引入的中间项子表达式, M_{21} 和 M_{22} 分别表达 M_2 分式中的 2 个分子项, M_{23} 表示其分母项。 $\hat{f}_{28}$ 为前述符号回归所得显式表达式。该校准优化过程仅调整输出层的线性组合参数, 不改变符号表达式树结构, 因此保持了模型原有的显式表达特征。

经 Ridge 参数校准优化后, 最终用于预测与后续反向优化分析的显式模型如下:

$$\hat{y} = 25.0071 + 1.0039\hat{f}_{28} + 0.1109w(\text{CaO}) - 1.4068w(\text{SiO}_2) - 4.3897(\text{W/C}) \quad (14)$$

最终显式模型在训练集和测试集上的预测性能见表 2。模型在训练集和测试集上均取得了较低的 RMSE 和平均绝对误差(MAE), 并表现出较大的决定系数(R^2), 且训练集和测试集两者 R^2 差异较小, 表明所建显式模型在保持较好预测能力的同时具有一定泛化性能。训练集与测试集评价结果较为接近, 也说明了所建显式模型在当前数据条件下具有较好的拟合能力和一定的泛化性能。

表 2 最终显式模型在训练集和测试集上的预测性能评价
Table 2 Prediction performance of the final explicit model on the training and test sets

	MSE	RMSE	R^2	MAE
Train set	1.697	1.303	0.811	0.903
Test set	1.986	1.409	0.817	1.126

图 6 对比了最终模型在训练集和测试集上的实测值与预测值。大多数样本点分布于对角线 $y = x$ 附近, 表明模型预测值与实测值具有较好一致性。训练集和测试集样本均未表现出明显偏离趋势, 说明所建显式模型能够较好刻画熟料化学组分特征与抗压强度之间的定量关系。

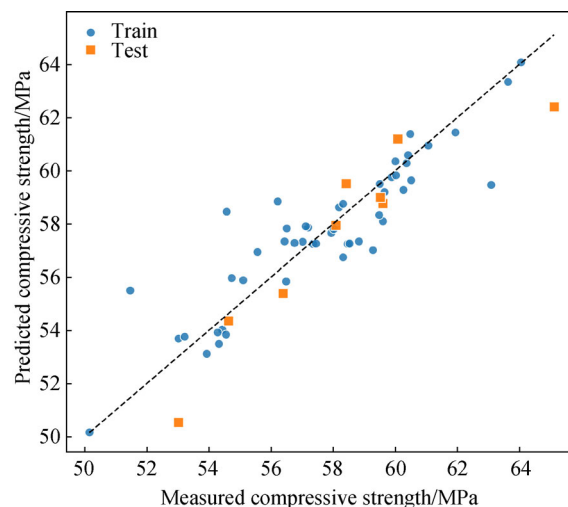


图 6 最终显式模型在训练集与测试集上的 28 d 抗压强度预测结果对比

Fig. 6 Comparison of prediction results of the final explicit model for 28 d compressive strength on the training and test sets

为进一步验证方法优势, 本工作将所提方法与多元线性回归(MLR)、岭回归、随机森林(RF)、支持向量回归(SVR)、XGBoost 以及未引入物理先验约束、变量覆盖度约束与长度惩罚的符号回归模型(SR)进行对比。所有模型均采用一致的训练集与测试集划分, 预测性能对比如表 3 所示。

由表 3 可见, 本工作所提方法的性能优于所有其他对比模型。虽然线性模型与非线模型也表现出较高的强度预测精度, 但其预测误差明显偏大, 且其模型结构难以支持显式关系解析与目标驱动的输入区间反推。SR 虽能生成显式表达式且在测试集上表现优于部分常规模型, 但其因缺乏物理先验约束、变量覆盖度约束及后校准优化, 其预测性能与反向优化适用性仍存在不足, 可能存在如: 表达式仅涵盖部分输入变量、表达式过长导致模型过于复杂、

表 3 不同模型在训练集和测试集上的预测性能对比

Table 3 Comparison of prediction performance of different models on the training and test sets

Model	Train				Test				Description	Support for SBP-based inverse optimization
	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2		
MLR	4.501	2.122	1.638	0.499	4.329	2.081	1.788	0.602	Linear model	Limited
Ridge Regression	4.968	2.229	1.697	0.448	4.180	2.044	1.469	0.616	Linear model	Limited
RF	3.647	1.910	1.456	0.594	5.076	2.253	1.470	0.533	Nonlinear black-box model	No
SVR-RBF	4.675	2.162	1.448	0.480	5.166	2.273	1.756	0.525	Nonlinear black-box model	No
XGBoost	4.118	2.029	1.524	0.542	5.083	2.254	1.708	0.533	Nonlinear black-box model	No
SR	2.257	1.502	1.104	0.749	3.959	1.990	1.459	0.636	Conventional symbolic regression	Limited
Ours	1.697	1.303	0.903	0.811	1.986	1.409	1.126	0.817	Proposed method	Yes

部分表达式违背先验物理规律等。相比之下，本工作方法在保持显式结构的同时，引入物理先验约束、变量覆盖筛选与 Ridge 参数矫正优化，实现了更优的测试集性能，并为后续分析与 SBP 反向优化提供了可直观使用的显式模型。

3.3 化学组成及 W/C-强度关系分析

为进一步解释最终显式模型所反映的熟料化学组分及 W/C 信息-强度关系，本工作基于所选符号回归模型对各输入变量开展了总体敏感性分析。具体地，采用局部扰动方法计算模型输出对各输入变量变化的响应，并以归一化平均绝对敏感度来分析

变量对 28 d 抗压强度预测结果的总体影响程度，同时统计各变量局部导数为正与为负样本占比，以分析其作用方向的稳定性。本分析仅表征变量在现有样本区间内的局部影响特征，不等同于组分全局单调作用规律。结果如图 7 所示，不同变量对强度预测的影响程度存在明显差异，其中 K_2O 、 Na_2O 、W/C、CaO 与 Al_2O_3 的敏感性相对较高，而 Fe_2O_3 和 MgO 的直接局部影响较弱。总体来看，最终模型表明水泥熟料 28 d 强度并非由单一氧化物独立决定，而是受 W/C、主要氧化物组成及其交互作用共同控制。

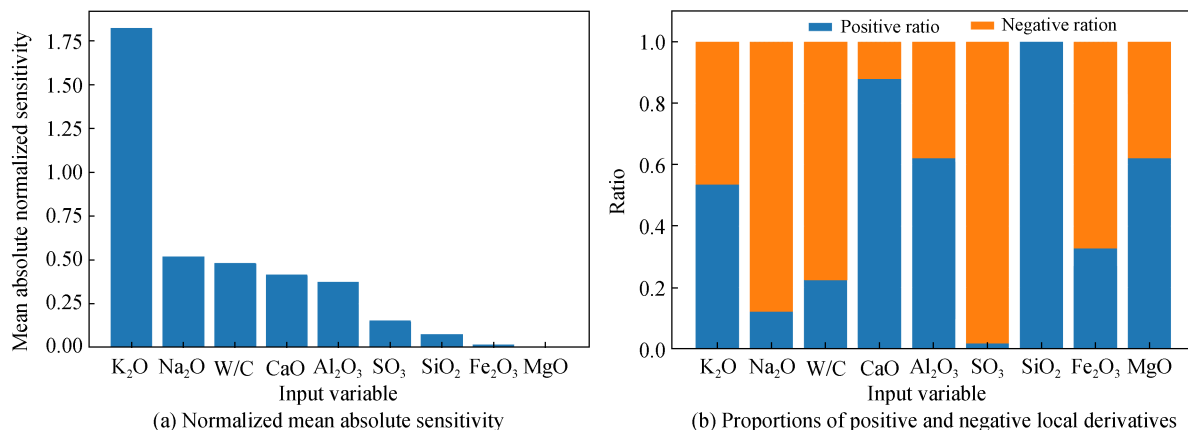


图 7 最终显式模型的输入变量敏感性与作用方向分析

Fig. 7 Sensitivity and directional effect of input variables in the final explicit model

从作用方向来看，W/C 的敏感性排序位居前列，且负向占比明显高于正向占比，说明在当前样本的大多工况下，W/C 增大时模型预测强度倾向于下降。这与 W/C 增大导致硬化浆体毛细孔隙率升高，结构致密性下降，进而不利于后期强度发展的基本认识一致^[13]。

CaO 则表现为较高敏感性与较高正向占比，说明在多数局部样本条件下 CaO 对强度具有正向贡

献。从熟料矿物形成机理看，CaO 在合理范围内提高有利于提升石灰饱和程度并促进 C_3S 形成，符合 28 d 强度发展机理^[6, 12]；但须注意，CaO 的作用并非严格单调，其局部影响受 SiO_2 、 Al_2O_3 、 Fe_2O_3 及碱性氧化物水平的耦合调剂，这也与熟料多组分协同决定强度的认识相符。 SiO_2 是 C_3S 和 C_2S 形成的基础组分，在模型中既出现在主效应项，也参与了非线性耦合项。其对强度的贡献取决于石灰饱和程

度、硅率以及 C_3S 与 C_2S 比例等, 因此不宜简单理解为全局单调正效应, 这与熟料多组分协同作用的机理是一致的^[2, 6, 12]。

K_2O 和 Na_2O 的高敏感性主要源于它们在模型中参与了大量非线性与耦合项, 而非单变量的独立强化项。从模型结构来分析, K_2O 多次出现在平方根型饱和修正项及高系数耦合项中, 微小变化即可改变相关修正幅度, 因此表现出较强局部敏感性; Na_2O 既参与了主效应项, 又与 W/C 形成显著的负交互项, 同时还与 K_2O 共同进入非线性修正结构, 使其作用方向受到 W/C 与碱含量组合状态控制。结合正负项占比结果, Na_2O 的局部敏感性以负向为主导, 表明在本次试验样本覆盖区间内, Na_2O 含量升高时, 模型预测强度倾向于下降。该规律与材料学基础机理相符, 即碱含量对水泥水化进程存在双向调控特征, 体系碱含量偏高时可能扰乱水化产物生长演化, 劣化硬化浆体微观结构, 可能不利于后期水化产物增长^[35–36], 且模型中表现为与 W/C 、 K_2O 的耦合后响应增强。

相较而言, Al_2O_3 、 Fe_2O_3 、 SO_3 和 MgO 在模型中更多体现为耦合修正变量, 而非独立的强影响因素。 Al_2O_3 和 Fe_2O_3 分别与铝率、液相形成以及 C_3A 、 C_4AF 等潜在矿物组成有关, 其对强度贡献通常会通过影响烧成过程和水化反应间接体现^[6, 8, 12], 因此局

部敏感性低于主变量。 SO_3 与碱性氧化物、铝酸盐相的反应存在强耦合关系, 其局部敏感性结果应理解为当前样本范围内的模型响应, 而非普遍机理结论, 实际生产中, SO_3 需要与碱含量匹配才能发挥最优作用^[8, 13]。 Fe_2O_3 和 MgO 的归一化敏感度较低, 说明在当前模型与样本范围内, 二者对强度的直接局部影响相对有限, 这与它们在熟料中主要通过改善烧成、调节液相性质^[6, 8]间接影响强度的机理是相符的。

总体看来, 最终显式模型中的乘积项、比值项和平方根型非线性项, 本质上是对石灰饱和系数、硅率、铝率及潜在矿物组成等多组分耦合关系的数学表征^[6, 12]。模型结果清晰表明, 28 d 抗压强度并非由单一氧化物独立决定, 而是受硅酸盐矿物形成潜力、碱性氧化物对烧成与水化的调制作用以及 W/C 主导的硬化浆体孔结构致密性共同影响, 这与水泥材料强度形成的多因素协同机理一致, 也验证了模型在材料学意义上的合理性。

3.4 目标驱动的单变量区间回推结果

为验证显式模型在目标驱动调控中的单变量可用性, 本工作选取测试集中的一代表性样本作为分析对象。同时, 设置较为宽松的工艺边界上下限, 如表 4 所示, 以约束回推优化结果区间取值范围。

表 4 单变量反向优化中各输入变量采用的工艺边界范围

Table 4 Process bounds of input variables used in single-variable inverse optimization

Fe_2O_3	Al_2O_3	CaO	SiO_2	K_2O	Na_2O	MgO	SO_3	W/C
2.0–6.0	3.0–6.0	50.0–70.0	18.0–30.0	0.1–1.5	0.05–0.50	0.5–5.0	0.2–3.5	0.47–0.53

以该样本为基准, 固定其余输入变量为当前值, 仅保留单个目标变量为开放变量, 在其工艺边界内, 基于显式模型表达式开展前向—反向区间收缩; 当区间收缩停滞但可行域仍较宽时, 进一步沿该变量维度进行条件分裂与细化搜索。最终由验证后的局部可行盒在该变量维度上的投影得到单变量可行区间, 并提取其中满足目标窗口的最小改变量点作为单变量调节建议。代表性样本当前预测强度 y 为 55.42 MPa, 设目标提升区间 Δy 为 1.5–3.0 MPa, 则对应目标 $y + \Delta y$ 强度期望窗口为 56.92–58.42 MPa。

单变量反向优化结果见表 5 所示, 不同变量在当前样本附近的单变量调控能力存在明显差异。 Fe_2O_3 与 MgO 在给定边界内单独变化时均无法使预测强度进入目标窗口, 说明其在当前点附近单独调

节不足以支撑目标强度提升。相比之下, Al_2O_3 、 CaO 、 SiO_2 、 K_2O 、 Na_2O 、 W/C 均存在单变量可行区间, 表明这些变量在当前局部条件下具备一定单独调控潜力。进一步比较最小改变量可见, W/C 、 CaO 与 SiO_2 达到目标窗口所需的相对调整幅度较小, 表现出较好的单变量调控可操作性, 说明模型能够在保持较小改动的前提下生成可行调节建议。而 Al_2O_3 、 K_2O 与 SO_3 虽数学上可达, 但其可行区间更贴近变量边界, 说明对应调节方案对工艺边界可能依赖较强, 工程实施弹性相对有限。

此外, 从最小改变量结果来看, W/C 达到目标窗口所需改动最小, 对应最小改变量约为–0.012 5; Na_2O 次之, 约为–0.041 8; SiO_2 和 CaO 的最小改变量分别约为+1.574 7 和+0.851 5; Al_2O_3 最小改变量约为+1.008 1。进一步结合当前样本值考察相对变化

表 5 目标强度提升 $\Delta y \in [1.5, 3.0]$ MPa 条件下的单变量反向优化结果

Variable	Current value	Feasible	Absolute feasible interval	Relative change interval/%	Recommended direction	Minimum adjustment	Relative adjustment/%	Recommended value	Predicted strength	Remarks
Fe ₂ O ₃	3.758 0	No			Infeasible					Single-variable infeasible
Al ₂ O ₃	4.914 0	Yes	5.922 0–6.000 0	20.51–22.10	Increase	1.008 1	20.514 9	5.922 1	56.922 0	Close to upper bound
CaO	63.801 0	Yes	64.649 8–65.441 1	1.330–2.571	Increase	0.851 5	1.334 6	64.652 5	56.922 3	Preferred adjustment
SiO ₂	21.312 0	Yes	22.886 5–24.461 5	7.388–14.780	Increase	1.574 7	7.388 8	22.886 7	56.921 9	-
K ₂ O	0.656 0	Yes	0.1000 0–0.102 9	-84.76 – -84.32	Decrease	-0.553 4	84.363 0	0.102 6	56.922 0	Close to lower bound; Large adjustment required
Na ₂ O	0.213 0	Yes	(0.1310–0.1607) $\cup$ (0.1621–0.1712)	(-38.48 – -24.53) $\cup$ (-23.92 – -19.62)	Decrease	-0.041 8	19.624 6	0.171 2	56.921 9	Multiple feasible intervals
MgO	3.270 0	No			Infeasible					Single-variable infeasible
SO ₃	0.789 0	Yes	0.200 0–0.296 6	-74.65 – -62.41	Decrease	-0.493 2	62.509 5	0.295 8	56.924 2	Close to lower bound; Large adjustment required
W/C	0.510 6	Yes	0.485 5–0.498 1	-4.915 – -2.447	Decrease	-0.012 5	2.453 0	0.498 0	56.921 8	Preferred adjustment

幅度，则 CaO 与 W/C 的相对改动比例较低，更适合作为当前样本附近的优先单变量调节对象，SiO₂ 改变量适中，可作为其次考虑调节对象。相比之下，Al₂O₃ 虽然存在可行区间，但其可行解已明显接近上界；K₂O 与 SO₃ 的可行解区间则靠近下界，说明其达到目标窗口往往需要较大幅度或边界驱动的调整，因此实际应用中可能需要谨慎采用。

总体来看，单变量区间回推结果表明，目标强度提升并不意味着所有输入变量都可作为独立调控对象；不同变量在当前样本附近的单变量可达性、调节方向及可操作性存在明显差异。该结果一方面说明显式模型能够为目标驱动调控提供具体的区间级建议，另一方面也表明单变量调节能力具有明显的局部性和边界依赖性，这为后续多变量协同优化分析提供了必要基础。需要说明的是，表 5 中的单变量反向优化结果主要反映当前样本附近各输入变量的局部可达性和调节方向，并不等同于实际生产中可独立调节某一氧化物，实际工程应用时仍需结合生产约束进行筛选和修正。

3.5 多变量联合回推与优化结果

实际熟料质量调控中很难将变化严格限制在某一变量上。受原料组成、工艺波动及配料平衡关系影响，多个变量往往呈现联动变化特征，因此更

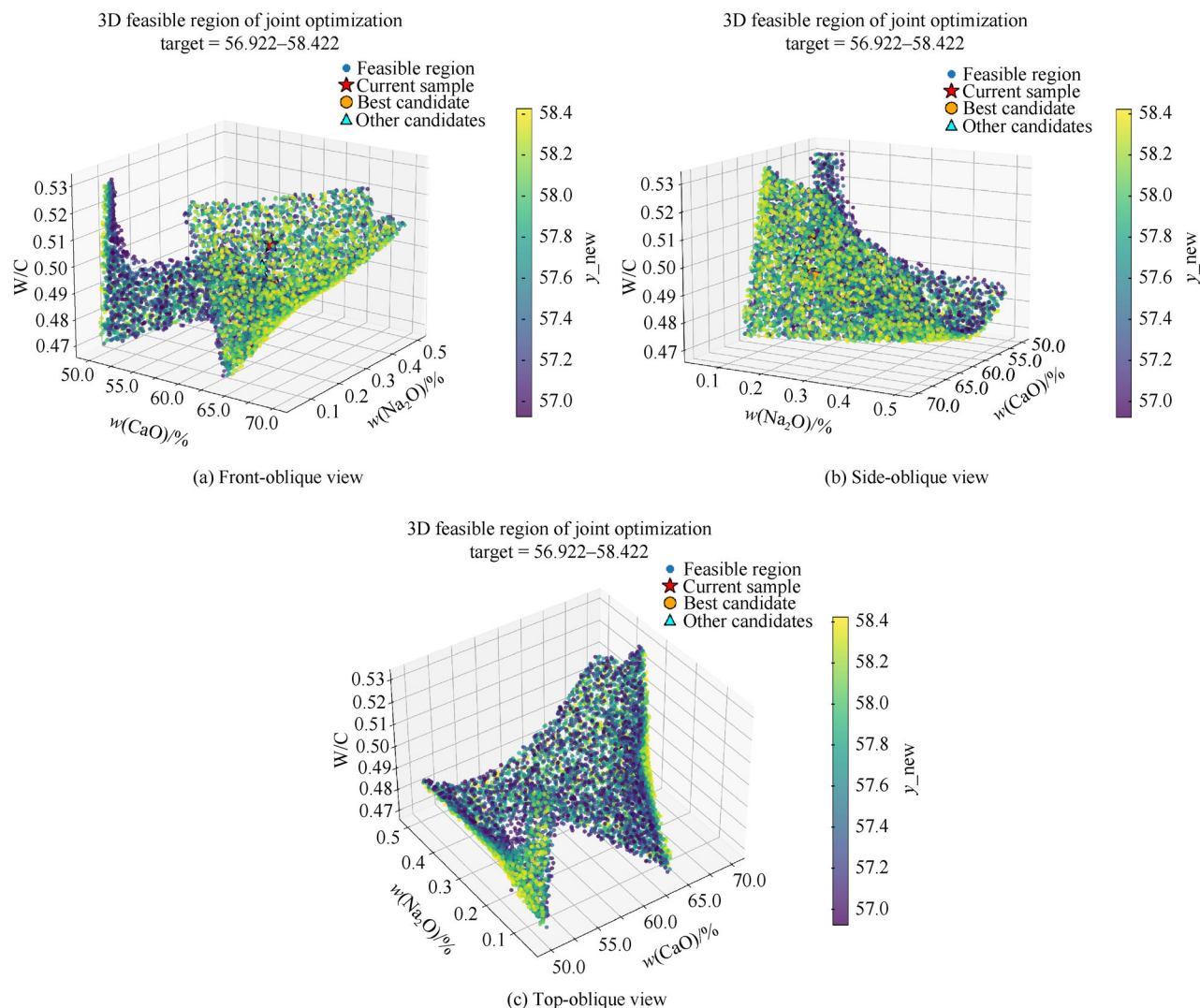
符合工程实际的调控方式，通常是多个变量的小幅协同调整。基于此，本工作进一步在与 3.4 节相同的代表性样本和目标强度窗口下，即相同测试样本 [$w(\text{Fe}_2\text{O}_3)=3.758\%$ 、 $w(\text{Al}_2\text{O}_3)=4.914\%$ 、 $w(\text{CaO})=63.801\%$ 、 $w(\text{SiO}_2)=21.312\%$ 、 $w(\text{K}_2\text{O})=0.656\%$ 、 $w(\text{Na}_2\text{O})=0.213\%$ 、 $w(\text{MgO})=3.270\%$ 、 $w(\text{SO}_3)=0.789\%$ 、 $W/C=0.5106$]，当前预测强度 y 为 55.42 MPa，设目标提升区间 Δy 为 1.5~3.0 MPa，则对应期望提升强度 $y + \Delta y$ 强度窗口约为 56.92~58.42 MPa，开展多变量联合回推与优化分析，以获得满足目标要求的联合可行域及协同调节方案。

与单变量结果不同，多变量联合回推得到不是各变量独立可行区间的简单拼接，而是由若干局部可行盒构成的联合可行域。该结果不仅给出各变量可能的变化范围，而且保留了变量之间的配对关系和协同约束，即某一变量的可行变化通常依赖于其他变量处于特定区间内，随机取少量具有明显差异代表性可行域展示如表 6 所示。因此，多变量联合优化能够更真实地反映目标强度提升条件下的输入耦合调控特征，并为候选方案提取和最优解选择提供基础。

图 8 给出了 CaO、Na₂O、W/C 的变量联合可行域的三维点云分布。从不同视角可以看出，满足目标

表 6 CaO、Na₂O 与 W/C 联合反向优化得到的代表性局部可行盒及采样验证结果Table 6 Representative local feasible boxes and sampled validation results for joint inverse optimization of CaO, Na₂O and W/C

Scheme ID	CaO interval	Na ₂ O interval	W/C interval	Sampled CaO	Sampled Na ₂ O	Sampled W/C	Predicted strength
R1	62.50–63.75	0.106 3–0.162 5	0.477 5–0.485 0	62.634 6	0.156 9	0.477 8	57.057 2
R2	60.00–61.25	0.331 3–0.387 5	0.470 0–0.477 5	60.026 5	0.382 5	0.474 3	57.492 9
R3	57.50–58.75	0.387 5–0.443 8	0.470 0–0.477 5	57.929 3	0.432 6	0.471 6	58.155 1
R4	65.00–66.25	0.331 3–0.387 5	0.492 5–0.500 0	65.942 7	0.341 5	0.498 7	58.153 2
R5	67.5–70.0	0.443 8–0.500 0	0.503 8–0.507 5	68.342 3	0.465 8	0.504 0	58.037 3



Note: Colors denote predicted strength within the target window, and different views show the feasible-region geometry and candidate solutions. W/C denotes water-to-cement ratio. Three subfigures show the same joint inverse-optimization result from different view.

图 8 CaO–Na₂O–W/C 三变量联合反向优化可行域的三维点云分布Fig. 8 Three-dimensional point clouds of the feasible region for joint inverse optimization of CaO, Na₂O and W/C

强度窗口的可行解并非分布在规则长方体区域内，而是表现为若干弯曲、片状和条带状的局部区域。这说明变量之间存在明显的协同与补偿关系：某一变量并不能在独立意义下任意变化，而需要与其他变量的变化共同匹配，才能使模型输出落入目标强度窗口。换言之，多变量联合可行域保留了输入变

量之间的耦合配对结构，而这类信息在单变量区间回推结果中是无法直接体现的。

在联合可行域基础上，本工作进一步从经采样验证的可行盒中提取代表性候选方案，并依据加权最小改动原则确定最终推荐解，期望目标提升区间 Δy 为 1.5~3.0 MPa，则对应期望提升强度 $y + \Delta y$ 强

度窗口约为 56.92~58.42 MPa, 联合候选方案前 3 项如表 7 所示。结果表明, 满足目标窗口的候选解并非唯一, 不同候选方案在变量分担方式上存在差异; 总体上, W/C 在候选方案表现为下降调节趋势, 而 CaO 与 Na₂O 则承担局部补偿作用。综合比较各方

案的改动幅度、目标命中情况及加权代价后, 最终推荐方案表现为多个变量的小幅协同变化, 而不是单一变量的大幅独立调整。这说明多变量联合优化不仅能够提供可行解, 而且能够给出更符合工程调控逻辑的协同调整建议。

表 7 CaO、Na₂O 与 W/C 三变量联合优化的候选调节方案及预测强度
Table 7 Candidate adjustment schemes and predicted strength for joint optimization of CaO, Na₂O and W/C

Scheme ID	CaO adjustment	Updated CaO	Na ₂ O adjustment	Updated Na ₂ O	W/C adjustment	Updated W/C	Total weighted cost	Predicted strength
C1	Original: 63.801 0 $\Delta = -0.015\ 6$	63.785 4	Original: 0.213 0 $\Delta = +0.003\ 4$	0.216 4	Original: 0.510 6 $\Delta = -0.015\ 0$	0.495 6	0.046 0	57.138 8
C2	Original: 63.801 0 $\Delta = -0.007\ 5$	63.793 5	Original: 0.213 0 $\Delta = -0.030\ 3$	0.182 7	Original: 0.510 6 $\Delta = -0.006\ 4$	0.504 1	0.047 6	57.010 3
C3	Original: 63.801 0 $\Delta = +0.014\ 4$	63.815 4	Original: 0.213 0 $\Delta = -0.014\ 8$	0.198 2	Original: 0.510 6 $\Delta = -0.014\ 3$	0.496 3	0.049 4	57.861 5

与 3.4 节单变量结果相比, 多变量联合优化显示出更明显的工程优势。单变量方案虽然能够识别潜在有效变量, 但在部分情况下存在贴近边界、改动幅度较大或单独不可达等问题; 相比之下, 多变量联合方案能够通过多个变量分担调节幅度, 在保持总体改动较小的同时达到目标强度窗口。因此, 多变量联合优化更符合实际生产中多因素联动调控的特点, 也更有利于形成可实施的调节建议。值得注意的是, 多变量联合优化虽比单变量调节更接近实际生产中的组分联动特征, 但表 7 所列方案仍属于显式模型约束下的数值可行解。工程应用时, 需将模型层面可行建议变化量进一步转化为原料配比调整方案, 并校核组分闭合、率值范围、f-CaO 控制及烧成条件等生产约束。如可通过带条件约束的多种熟料加权掺和求解方法来逼近成分占比, 乃至其它有约优化手段也可以被专项研究, 并于实际应用时结合生产约束进行筛选和修正。因而, 本工作反向优化结果更适合作为候选方案生成和趋势判断依据, 实际实施仍需结合生产约束进行筛选与修正。

3.6 方法局限性

本工作方法仍存在一定局限性。首先, 受样本数量限制, 当前显式模型及其反向优化结果主要反映现有数据范围内的局部规律, 模型在更大样本规模、不同原料体系及不同工况条件下的泛化能力仍需进一步扩充试验量并验证。其次, 关于构建的化学组分到强度的表达式树及其对应显式方程, 其内在机理需引入领域专家进一步解读。此外, 本工作输入变量主要基于 XRF 表征获取的化学组成信息及 W/C, 尚未进一步引入矿物相组成、烧成制度、冷却条件等深层影响因素, 因此当前模型对熟料形

成机理的刻画仍不够全面, 后续工作需将矿物相、工艺制度及更多过程变量纳入建模体系, 提升模型对熟料强度形成机理的表征能力。最后, 目标驱动反向优化结果目前主要完成了模型层面的数值验证, 需进一步结合真实场景进行生产数据验证与反馈。后续工作可结合矿物相组成、烧成制度等信息扩充输入变量及数据样本数量。未来将在扩充样本规模、输入维度、引入更多机理约束及开展闭环试验等方面进行进一步深入研究。

4 结论

1) 针对熟料化学组成与 28 d 抗压强度之间显式定量关系尚不完全清晰、现有模型多侧重正向预测且难以同时支撑目标强度驱动反向调控问题, 提出了基于符号语义反向传播的熟料性能显式表征及反向优化方法。通过试验与数据收集、符号回归显式建模、语义反向传播区间回推与多变量协同优化的熟料性能表征建模框架, 成功实现了正向预测到反向优化的闭环。试验结果表明, 该方法能够同时实现熟料强度预测、显式关系表达和目标驱动的组分调节, 为熟料性能建模与配方调控提供了全新的可行方案。

2) 基于所搜集 58 组不同来源熟料样品 XRF 化学组成、W/C 及对应 28 d 抗压强度数据, 构建了熟料表征-性能数据集, 并提出物理先验约束、模型复杂度和变量覆盖度联合筛选策略, 所得符号回归模型在测试集上的 RMSE 为 1.409 MPa, MAE 为 1.126 MPa。验证了提出选择策略的有效性与所选模型具备较好表征熟料化学组成及 W/C-强度性能之间非线性映射关系的能力。

3) 目标变化驱动的区间回推与反向优化方面,

基于显式表达式树与语义反向传播方法, 结合前向–反向区间收缩和条件分裂策略, 可获得满足目标强度变化要求的输入变量可行域。结果表明, 当设定强度提升目标后, 模型可在给定目标强度提升条件下, 反向优化回推并输出单变量可行调节区间和多变量协同优化方案, 例如期望输出强度由 55.4 MPa 提升 1.5~3.0 MPa, 单变量 CaO 变动区间应增大 0.849~1.640; 或多变量联合优化时, 取其中一方案作为示例: CaO、Na₂O、W/C 联合小幅度变动{-0.015 6, +0.003 4, -0.015 0}, 即可调增目标强度至 57.138 8 MPa 等。可实现由目标强度需求向熟料化学组成等调节建议的反向推导, 验证了所提出方法在熟料配比反向优化中的可行性。

4) 基于符号语义反向传播的熟料性能显式表征及反向优化方法为熟料性能预测、显式关系解析和目标驱动组分调控提供了一种新的可解释建模思路。相比单纯黑盒预测模型, 该方法兼具预测精度、表达式可解释性和反向优化能力, 可为熟料质量判别、反向优化与生产调控提供方法支撑。

参考文献:

- [1] ANDRADE NETO J S, CARVALHO I C, SANTANA H A, et al. The role of clinker mineralogy in cement properties: An analysis using statistical mixture design[J]. *Cem Concr Res*, 2026, 201: 108102.
- [2] MECHLING J M, LECOMTE A, DILIBERTO C. Relation between cement composition and compressive strength of pure pastes[J]. *Cem Concr Compos*, 2009, 31(4): 255–262.
- [3] STUTZMAN P, HECKERT A, TEBBE A, et al. Uncertainty in Bogue-calculated phase composition of hydraulic cements[J]. *Cem Concr Res*, 2014, 61: 40–48.
- [4] GONÇALVES J P, HAN T H, SANT G, et al. Toward smart and sustainable cement manufacturing process: Analysis and optimization of cement clinker quality using thermodynamic and data-informed approaches[J]. *Cem Concr Compos*, 2024, 147: 105436.
- [5] WILLIAMS F, YANG A D. Potential of reducing CO₂ emissions in cement production through altering clinker compositions[J]. *Ind Eng Chem Res*, 2024, 63(40): 17158–17167.
- [6] LUDWIG H M, ZHANG W S. Research review of cement clinker chemistry[J]. *Cem Concr Res*, 2015, 78: 24–37.
- [7] SCHNEIDER M, ROMER M, TSCHUDIN M, et al. Sustainable cement production: Present and future[J]. *Cem Concr Res*, 2011, 41(7): 642–650.
- [8] CHATTERJEE A K. Chemistry and engineering of the clinkerization process: Incremental advances and lack of breakthroughs[J]. *Cem Concr Res*, 2011, 41(7): 624–641.
- [9] SVINNING K, HØSKULDSSON A, JUSTNES H. Prediction of potential compressive strength of Portland clinker from its mineralogy[J]. *Cem Concr Compos*, 2010, 32(4): 300–311.
- [10] 郭向阳, 芦令超, 王守德, 等. 钡掺杂对高阿利特硅酸盐水泥熟料组成与性能的影响[J]. *硅酸盐学报*, 2009, 37(12): 2083–2089. GUO Xiangyang, LU Lingchao, WANG Shoude, et al. *J Chin Ceram Soc*, 2009, 37(12): 2083–2089.
- [11] COSTA A R D, COPPE M V, BIELEFELDT W V, et al. Thermodynamic modelling of cements clinkering process as a tool for optimising the proportioning of raw meals containing alternative materials[J]. *Sci Rep*, 2023, 13: 17589.
- [12] 沈 威, 黄文熙, 闵盘荣. 水泥工艺学[M]. 武汉: 武汉工业大学出版社, 1991.
- [13] BULLARD J W, JENNINGS H M, LIVINGSTON R A, et al. Mechanisms of cement hydration[J]. *Cem Concr Res*, 2011, 41(12): 1208–1223.
- [14] 刘晓, 王思迈, 卢磊, 等. 机器学习预测混凝土材料耐久性的研究进展[J]. *硅酸盐学报*, 2023, 51(8): 2062–2073. LIU Xiao, WANG Simai, LU Lei, et al. *J Chin Ceram Soc*, 2023, 51(8): 2062–2073.
- [15] LI Z Z, YOON J, ZHANG R, et al. Machine learning in concrete science: Applications, challenges, and best practices[J]. *npj Comput Mater*, 2022, 8: 127.
- [16] ZHANG L L, OH S K, PEDRYCZ W, et al. Building fuzzy relationships between compressive strength and 3D microstructural image features for cement hydration using Gaussian mixture model-based polynomial radial basis function neural networks[J]. *Appl Soft Comput*, 2021, 112: 107766.
- [17] NGUYEN H, VU T, VO T P, et al. Efficient machine learning models for prediction of concrete strengths[J]. *Constr Build Mater*, 2021, 266: 120950.
- [18] MARANI A, NEHDI M L. Machine learning prediction of compressive strength for phase change materials integrated cementitious composites[J]. *Constr Build Mater*, 2020, 265: 120286.
- [19] ISLAM M, LIU S R, WANG L, et al. Predicting cement strength as probability density: Resolving partial observability and sample scarcity for industrial quality control[J]. *IEEE Trans Ind Inform*, 2026, 99: 1–12.
- [20] 余睿, 陈飞翔, 范定强, 等. 基于机器学习技术的超高性能混凝土性能预测与精细化设计[J]. *硅酸盐学报*, 2025, 53(05): 1165–1176. YU Rui, CHEN Feixiang, FAN Dingqiang, et al. *J Chin Ceram Soc*, 2025, 53(05): 1165–1176.
- [21] OEY T, JONES S, BULLARD J W, et al. Machine learning can predict setting behavior and strength evolution of hydrating cement systems[J]. *J Am Ceram Soc*, 2020, 103(1): 480–490.
- [22] CHENG B Y, MEI L, LONG W J, et al. Ai-guided proportioning and evaluating of self-compacting concrete based on rheological approach[J]. *Constr Build Mater*, 2023, 399: 132522.
- [23] CYBENKO G. Approximation by superpositions of a sigmoidal function[J]. *Math Control Signal Systems*, 1989, 2(4): 303–314.
- [24] RUDIN C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead[J]. *Nat Mach Intell*, 2019, 1(5): 206–215.
- [25] TAGADE P M, ADIGA S P, PANDIAN S, et al. Attribute driven inverse materials design using deep learning Bayesian framework[J]. *npj Comput Mater*, 2019, 5: 127.
- [26] PAWLAK T P, WIELOCH B, KRAWIEC K. Semantic backpropagation for designing search operators in genetic programming[J]. *IEEE Trans Evol Comput*, 2015, 19(3): 326–340.
- [27] LA CAVA W, BURLACU B, VIRGOLIN M, et al. Contemporary symbolic regression methods and their relative performance[J]. *Advances in neural information processing systems*, 2021, 2021(DB1): 1–16.
- [28] 中国建筑材料科学研究总院有限公司、厦门艾思欧标准砂有限公

- 司、安徽海螺水泥股份有限公司、浙江礼显试验仪器制造有限公司、北京国建联信认证中心有限公司、广西建科建设工程质量检测与评价有限责任公司、黄河勘测规划设计研究院有限公司、河南豫美建设工程检测有限公司等. GB/T 17671—2021《水泥胶砂强度检验方法》[S]. 北京: 北京标准出版社, 2021.
- [29] KOHAVI R. A study of cross-validation and bootstrap for accuracy estimation and model selection[C]//Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 1995: 1137–1143.
- [30] ARLLOT S, CELISSE A. A survey of cross-validation procedures for model selection[J]. *Statist Surv*, 2010, 4: 40–79.
- [31] BURLACU B, KRONBERGER G, KOMMENDA M, et al. Parsimony measures in multi-objective genetic programming for symbolic regression[C]//Proceedings of the Genetic and Evolutionary Computation Conference Companion. ACM, 2019: 338–339.
- [32] HOERL A E, KENNARD R W. Ridge regression: Applications to nonorthogonal problems[J]. *Technometrics*, 1970, 12(1): 69–82.
- [33] MOORE R E, KEARFOTT R B, CLOUD M J. Introduction to Interval Analysis[M]. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2009.
- [34] ABDI H, WILLIAMS L J. Principal component analysis[J]. *Wires Comput Stat*, 2010, 2(4): 433–459.
- [35] INAM J, JAN S. Alkalies in cement: A review: II. Effects of alkalies on hydration and performance of Portland cement[J]. *Cem Concr Res*, 1978, 8(1): 37–51.
- [36] MOTA B, MATSCHEI T, SCRIVENER K. Impact of NaOH and Na₂SO₄ on the kinetics and microstructural development of white cement hydration[J]. *Cem Concr Res*, 2018, 108: 172–185.

作者贡献声明:

韩瑞琪: 论文构思、设计研究方法、研究方案和模型算法、撰写初稿、调查研究、分析数据并推导结论、验证、绘制图表、审阅与修订

张亮亮: 论文构思、设计研究方案和模型算法、撰写初稿、数据管理、分析数据并推导结论、监督指导、审阅与修订

宁 帅: 论文构思、设计研究方法

王 琳: 调查研究、验证、项目管理、提供资源、获取资助、监督指导、审阅与修订

袁景凌: 调查研究、验证、监督指导、审阅与修订

杨 波: 项目管理、监督指导、审阅与修订

侯鹏坤: 监督指导、审阅与修订

李琴飞: 审阅与修订

Explicit Representation and Inverse Optimization of Clinker Performance Based on Symbolic Semantic Backpropagation

HAN Ruiqi¹, ZHANG Liangliang^{2,3}, NING Shuai¹, WANG Lin^{2,3}, YUAN Jingling^{4,5}, YANG Bo^{2,3}, HOU Pengkun¹, LI Qinfei¹

(1. Shandong Provincial Key Laboratory of Green and Intelligent Building Materials, University of Jinan, Jinan 250022, China;

2. Shandong Key Laboratory of Ubiquitous Intelligent Computing, University of Jinan, Jinan 250022, China; 3. Quancheng Laboratory, Jinan 250100, China; 4. School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, China; 5. State Key Laboratory of Silicate Materials for Architectures, Wuhan University of Technology, Wuhan 430070, China)

Extended Abstract

Introduction The chemical composition of cement clinker directly affects cement hydration, later-age strength development, and quality stability, and therefore serves as a key basis for clinker quality evaluation and production regulation. Conventional clinker performance analysis mainly relies on chemical composition testing, empirical rules, and mineral-phase estimation methods such as the Bogue calculation. Although these approaches have the physical meanings and practical engineering applicability, their predefined model forms are insufficient to fully capture the complex nonlinear strength relationships arising from the coupled effects of multiple oxide components and the water-to-cement ratio.

In recent years, machine learning methods have been widely applied to predict the performance of clinker and cement-based materials, providing some opportunities for modeling complex composition-property relationships. However, the existing black box models usually do not give clear mathematical formulas, which makes it difficult to directly explain how the strength develops. In addition, most studies mainly predict the strength from chemical composition. They are less able to work in the opposite direction, that is, to identify suitable adjustment directions and feasible ranges of clinker chemical components when a strength-improvement target is given, either as a specific value or as an interval. To solve these problems, this study was to develop an explicit modelling and inverse optimization method for clinker performance *via* combining symbolic regression and semantic backpropagation. This method could provide an interpretable prediction of 28 d compressive strength and target driven inverse regulation of clinker composition.

Methods This study proposed an explicit representation and inverse optimization method for clinker performance by combining symbolic regression with semantic backpropagation. This method was designed to address the limited explicit characterization of clinker composition-strength relationships and the difficulty of regulating clinker composition according to the strength requirements. First, clinker samples from different sources were collected, and their chemical compositions and 28 d compressive strengths were used to build a clinker composition performance dataset. Then, a symbolic regression based on genetic programming was used to search for explicit relationships between clinker chemical composition and 28 d compressive strength in an expression space formed

by variables, constants, and mathematical operators. Expression trees and their corresponding equations were obtained from the candidate models. To avoid physically unreasonable models or the loss of important variables caused by relying only on prediction error, a model selection strategy was introduced *via* considering physical prior constraints, model complexity, and variable coverage. This strategy could balance prediction accuracy, physical rationality, structural simplicity, and interpretability. Furthermore, a semantic backpropagation method driven by target strength variation was used to propagate the required strength improvement backward through the explicit expression structure to the input variables. Feasible adjustment intervals for individual chemical components and collaborative optimization schemes for multiple variables were then derived. In this way, the proposed method could transform target strength changes into practical suggestions for inverse regulation of clinker chemical composition.

Results and discussion The results show that the symbolic regression model selected by considering physical prior constraints, model complexity, and variable coverage achieves a good balance between prediction accuracy and interpretability. The final model includes all nine key input variables, including CaO, SiO₂, Al₂O₃, Fe₂O₃, MgO, SO₃, K₂O, Na₂O, and W/C. Its expression is not a simple linear combination, but is composed of product terms, ratio terms, and nonlinear terms. This structure allows the coupled effects of the main clinker oxides and W/C on the 28 d compressive strength to be explicitly described. The sensitivity analysis further shows that K₂O, Na₂O, W/C, CaO, and Al₂O₃ have relatively great effects on the model output. Among these variables, W/C shows a mainly negative effect around the most samples, while CaO makes a positive contribution under most local conditions. These results indicate that the obtained explicit model can capture the basic characteristics of strength development controlled by the combined effects of multiple clinker components.

The final explicit model demonstrates a satisfactory predictive performance on both the training and test datasets. For the training set, the RMSE, MAE, and R^2 are 1.303 MPa, 0.903 MPa, and 0.811, respectively. For the test set, the corresponding values are 1.409 MPa, 1.126 MPa, and 0.817. Compared with multiple linear regression, Ridge regression, random forest, support vector regression, XGBoost, and conventional symbolic regression, the proposed method achieves a better prediction accuracy on the test set. Meanwhile, in contrast to black-box machine learning models, the proposed model retains an explicit expression structure, which can be directly used for variable-relationship analysis and subsequent target-driven input interval back-propagation.

The inverse optimization results show that the proposed model can translate requirements for strength improvement into interpretable suggestions for composition adjustment. For a representative sample with a predicted strength of 55.42 MPa, when the target strength increase is set as [1.5, 3.0] MPa, single variable backpropagation identifies feasible adjustment intervals for CaO, SiO₂, Na₂O, W/C, and other variables. Among them, CaO and W/C require smaller minimum adjustment magnitudes, indicating that they may be more suitable for single variable regulation. When CaO, Na₂O, and W/C are backpropagated simultaneously, the feasible solutions are not simple combinations of independent adjustment intervals, but appear as several local feasible regions. This result reveals clear synergistic and compensatory effects among clinker components. Sampling validation further confirms that multiple joint adjustment schemes can shift the predicted strength into the target range. These findings indicate that the proposed method can provide practical candidate solutions for clinker strength improvement, quality evaluation, and inverse mix proportion regulation.

Conclusions This study demonstrated a potential of symbolic semantic backpropagation in clinker performance prediction, explicit relationship analysis, and strength-target-driven composition regulation. The proposed method established an explicit quantitative relationship between clinker chemical composition and compressive strength at 28 d *via* combining symbolic regression with semantic backpropagation as well as incorporating physical prior constraints, variable coverage screening and target-interval-based backpropagation. It could also transform strength improvement targets into feasible adjustment intervals for individual variables and collaborative optimization schemes for multiple variables. Compared with conventional black-box models, this method could offer a greater interpretability for the relationship between composition and strength and provides more direct inverse guidance for clinker composition regulation. Overall, it could provide an interpretable and practical approach for clinker strength prediction, quality assessment, and inverse mix proportion regulation.

Keywords clinker performance prediction; explicit representation; inverse optimization; symbolic regression; physical prior constraints