

数据驱动模型对化学反应的定量预测

浙江省杭州第二中学 张一超

数据驱动模型在化学反应的定量预测中发挥着日渐明显的作用，尤其是在发现新型手性催化剂、理解反应机理和设计新分子材料等领域，体现了强大的应用价值。而这些模型通常基于定量构效关系（QSAR）和定量构-性质关系（QSPR）的原理构建。

QSAR（Quantitative Structure-Activity Relationship）是一种建模概念，用于预测化学物质的生物活性或反应活性。它通过统计方法关联分子的结构特征（如原子数量、官能团、立体化学等）与它们的生物活性或反应活性。QSAR 模型通常可用于药物设计、农用化学品、环境污染物的风险评估等领域。

QSPR（Quantitative Structure-Property Relationship）与 QSAR 类似，但关注的是化学物质的物理化学性质，如溶解度、沸点、反应速率等。QSPR 模型有助于理解分子结构与它们所表现的物理化学性质之间的关系，具有用实验数据的整合反哺理论的发展和完善的潜力，对更深层次认识物质有重要作用。

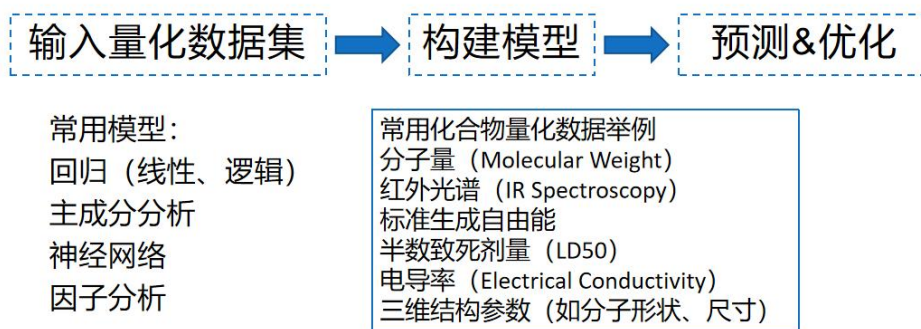


图 1 数据驱动模型建模的初级思路

数据驱动模型在化学领域发挥着至关重要的作用，它们通过分析大量的化学数据来预测和设计新药物，构建药物毒性或药性模型，从而加速药物发现过程并提高其安全性；在不对称反应催化剂的预测中，这些模型能够处理选择性与催化剂量化数据，识别影响反应选择性的关键特征点，进而优化催化剂设计，提高反应的立体选择性；数据驱动模型还能规划合成反应环境，通过输入合成效率与环境量化特征，计算出最高效率的环境条件，以实现绿色化学和可持续合成；在配体特征优化方面，模型可以指导配体结构的调整，优化对映体选择性，这对于手性药物的合成尤为重要；物理化学性质的机理研究也受益于数据驱动模型，它们能够通过模型对已有理论进行改良和完善，提供更深入的机理理解。

以一篇发表在 *Nature Reviews Chemistry* 的综述论文为例(Reid, J. P., & Sigman, M. S. (2018). Comparing Quantitative Prediction Methods for the Discovery of Small-Molecule Chiral Catalysts.)。这篇文章探讨了定量预测技术在小分子手性催化剂发现中的应用，评估了不同方法在预测催化剂性能方面的准确性和可靠性，充分证明了数据驱动模型在化学领域具有的重要作用。

同样，另一篇发表在 *Accounts of Chemical Research* 的综述论文（Crawford, J.

M., Kingston, C., Toste, F. D., & Sigman, M. S. (2021). Data Science Meets Physical Organic Chemistry.) 则探讨了数据驱动模型在设计新型催化剂的应用, 具有提高反应的效率和选择性的作用, 并且特别关注了数据驱动模型在手性合成方面的应用。

数据驱动模型已经成为化学研究中不可或缺的工具, 它们不仅已经在药物设计、催化剂发现、合成路径优化等领域展现出巨大的潜力, 也在物理化学研究方面提供了新的视角。随着算法的不断进步, 数据驱动模型将在化学领域的研究和应用中发挥更加重要的作用, 不断完善化学的信息化发展。

Reference:

- [1] Kulik, H. J., & Sigman, M. S. (2021). Advancing Discovery in Chemistry with Artificial Intelligence: From Reaction Outcomes to New Materials and Catalysts. *Accounts of Chemical Research*, 54(5), 2335–2336. <https://doi.org/10.1021/acs.accounts.1c00232>
- [2] Crawford, J. M., Kingston, C., Toste, F. D., & Sigman, M. S. (2021). Data Science Meets Physical Organic Chemistry. *Accounts of Chemical Research*, 54(6), 3136–3148. <https://doi.org/10.1021/acs.accounts.1c00285>
- [3] Robinson, S. G., & Sigman, M. S. (2020). Integrating Electrochemical and Statistical Analysis Tools for Molecular Design and Mechanistic Understanding. *Accounts of Chemical Research*, 53(2), 289–299. <https://doi.org/10.1021/acs.accounts.9b00527>
- [4] 江辰, 尤田耙, 等. (2006). 手性领域的定量构效关系研究. 中国科学技术大学.
- [5] Reid, J. P., & Sigman, M. S. (2018). Comparing Quantitative Prediction Methods for the Discovery of Small-Molecule Chiral Catalysts. *Nature Reviews Chemistry*, 2(10), 290–305. <https://doi.org/10.1038/s41570-018-0040-8>
- [6] Williams, W. L., Zeng, L., Gensch, T., Sigman, M. S., Doyle, A. G., & Anslyn, E. V. (2021). The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Central Science*, 7(7), 1622–1637. <https://doi.org/10.1021/acscentsci.1c00535>