

低分子創薬のためのインフォマティクス
Informatics tools for small molecule drug discovery

宮尾 知幸
Tomoyuki Miyao

奈良先端科学技術大学院大学
データ駆動型サイエンス創造センター/先端科学技術研究科
Nara Institute of Science and Technology (NAIST)

Artificial Intelligence (AI)は、もともと人間の認知や判断（やその過程）を真似るコンピュータを意味するが、化学や生物学などの科学の分野で利用されるAIには、人間では気づかなかつた知見や“あるかどうかも定かではない”法則などを発見することが期待されている。知見や法則そのものを実験データから自動抽出し仮説として提案することは、モデルを構築することと同義と考えることができる。例えば、化学構造と阻害活性値との間で構築する回帰モデル（定量的構造活性相関モデル）は、化学構造における特定の“特徴“またはその”組み合わせ“が活性にどのように影響を与えているかを表現した仮説とみなすことができる。

コンピュータにモデルを構築させるメリットとして、**1**:人間による偏見（バイアス）が入らないこと、**2**:大量のデータを効率的に処理できること、**3**:複雑な特徴および特徴間の関係を表現できること、**4**:モデルが取得したデータセットを最も説明するように最適化されていること、などが挙げられる。特に、**2**および**3**を従来から改善した学習方法として、ニューラルネットワークにおける深層学習があげられる。深層学習により大量のサンプルを効率的に学習することが可能になり（**2**）、マルチタスク学習などでは複数の目的変数を同時に予測するモデルを効率的に構築することができる（**3**）。これらの点を除けば現在行われている深層学習によるモデルは、従来から利用・開発されていた機械学習モデルと大きな違いはない。

本発表では、特に上記の**3**の項目に分類される以下の二つのテーマA, Bについて、定量的構造活性相関モデルを対象として、我々の研究成果を中心に研究紹介したい。また、実際に実験科学者にとって役立つモデルについてディスカッションをし、私が感じている課題を共有できればと思う。

A: 複数のデータセット（文献データなど）を利用して、自らが着目しているデータセットを説明するための、より高精度なモデル構築

Matsumoto et al., *ACS Omega* **2021**, 6, 18, 11964–11973

B: 分子構造のコンピュータにおける表現方法とモデリング手法の検討

Sato et al., *JCAMD* **2021**, 35, 179–193