

# 多段GBIツールを利用した化学物質の部分構造情報の有効利用

○田中栄太郎、稲積宏誠

青山学院大学理工学部 (〒229-8558 相模原市淵野辺 5-10-1)

## 【目的】

我々は、グラフ構造表現された事例集合に共通する部分構造抽出を目的とし、グラフマイニング手法である GBI(Graph-Based Induction)法<sup>1</sup>に注目してきた。グラフマイニングについては多くの提案がなされ、多方面への応用が期待されている。特に、GBI 法についても改良や応用が提案されているが、我々はまず、GBI 法の改良とその実装方法について検討を始め、化学物質である生理活性物質を対象としてその適用可能性を確認した。次に、決定木分析を可能とする基礎情報として、部分構造とそれらの関係構造を抽出することを実現した。さらに、決定木分析の改良として、その基礎情報によりいかにして有効な知識を発見するかについて検討を行ってきた。

一連の GBI 法の改良や試みを通じて、GBI 法で抽出される多くの部分構造の意味づけや解釈をいかにして実現するかということが、特に重要であることがわかった。その結果、対象とする問題の領域知識やヒューリスティックに基づいて特定の部分構造を指定し、それを起点とした探索領域から得られる特徴的な部分構造の組み合わせを発見するための支援システムの開発を進めた。特に、多段的な GBI 法の適用を可能とすることで、その応用領域の拡大を目指した。多段的な GBI 法とは、指定した部分構造以外の部分構造を初期の結合関係に関する情報を保持した上でリセットすることによって、基盤となる部分構造を特定した後に、これに対する置換基との関係性を評価することを可能とするものである。

本発表では、生理活性物質や HIV-DATA に対してこの手法を適用することによって、支援システムとしての有効性を示すとともに、ツールとしての機能向上に向けての議論することを目的とする。

## 【概要】

本ツールにより可能な基本的な処理は、対象物質に含まれる部分構造の逐次的な抽出、ユーザが指定する部分構造に対する付加構造の活性度等の条件に基づいた抽出、ユーザが指定する複数の特定部分構造以外をリセットしながらの継続実行などである。

たとえば、フラボノイドを対象として、その構造上の特徴と活性度の関係について分析した結果、図に示すような共通構造(本システムでは基盤構造と呼ぶ)に対しては、図 1 に示すような異なる位置に同様の部分構造が付加されていることを示すことができた。また、AIS screening data として HIV 患者に投与された化合物の活性度に注目した分析の結果は、図 2 に示すような共通部分構造を持つ化合物であっても、付加構造の微妙な違いが活性度の違いと関連していることを示すことができた。特に、抗 HIV 薬として知られる逆転写阻害剤アジドチミジン(AZT)が抽出できたことは注目される。

これらの処理は、図 3 に示す GUI 環境で実現されている。今後は、化学分野の専門家であるユーザの立場からのシステム改良を行っていく予定である。

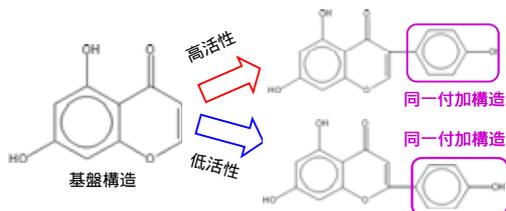


図 1 フラボノイドにおける活性度の異なる付加構造の例

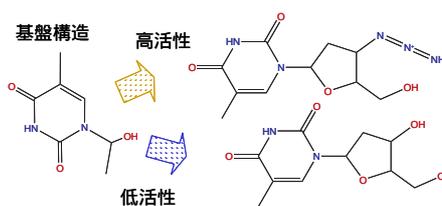


図 2 HIV-DATA における活性度の異なる付加構造の例

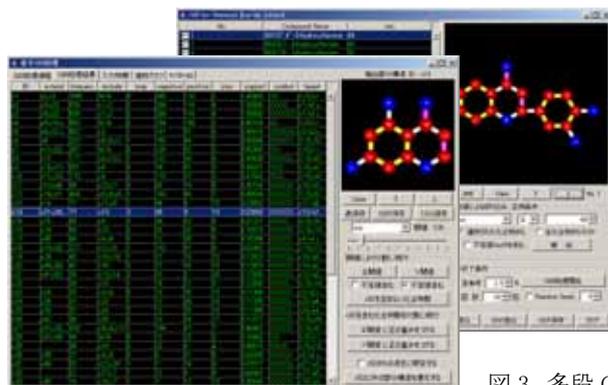


図 3 多段 GBI 処理を行うための GUI 環境

## 参考文献

- [1] 松田 喬, 元田 浩, 鷲尾 隆: "一般グラフ構造データに対する Graph-Based Induction とその応用", 人工知能学会論文誌, Vol. 16, No. 4, pp. 363-374(2001)
- [2] 田中栄太郎, 稲積宏誠: GBI(Graph-Based Induction)法の拡張と GUI によるグラフマイニング支援環境の構築 - 化学物質を対象として -, 人工知能学会 知識ベースシステム研究会, SIG-KBS-A405, pp. 1 - 6 (2005)