

部分グラフの包含関係に着目した代謝経路予測手法の提案

田中健一

情報科学を専門としている私が、本研究室の様なウェットな研究室で働くことを希望した理由は、自分の持っている技術を医療や食糧問題の解決に役立てたいと考えたからです。私の得意分野は情報科学とソフトウェア開発ですが、本当に必要とされているソフトウェアを開発する為には、化学や生物学の専門家が何を重要視していて、どのようなシステムを提供する必要があるのかを正しく判断する必要があります。逆に、情報科学の視点から何が重要かを相手に伝えることも必要です。異分野の人と共同研究を行う際、専門用語や考え方の違いの壁が存在し、これらをクリアにする為には日常からディスカッションが可能な環境が必須でウェットな研究室に身を置くべきと考えました。今回は博士研究で開発したソフトウェアの紹介を通して、現時点で何が出来るかについて説明を行います。最後にソフトウェアのデモンストレーションを行うので、「この様な事はできないのか？」等のコメントを頂けると非常にうれしいです。

紹介論文

“Metabolic pathway prediction focusing on inclusive relation of subgraphs”,

K. Tanaka, *Doctoral Dissertation, NAIST*, NAIST-IS-DD0761018, March 17, 2010.

"Metabolic pathway prediction based on inclusive relation between cyclic substructures",

K. Tanaka, et al. *Plant Biotechnology*, **26**, 459-468 (2009)

要旨

構造決定されている代謝物は約 50,000 種と報告されているが、約 90%に関しては代謝経路が未解明な状態にある。先行研究では、(1)文献に登場する代謝経路情報を集約する。(2)化学構造類似性を目で判断し代謝経路を整理する。(3)フィンガープリントや最大共通部分グラフを用い化学構造類似性を計算することで代謝経路を予測する。等のアプローチがとられてきたが、(1)や(2)の手法は膨大なマンパワーが必要である、(3)の手法は予測精度の低さや計算コストが膨大である等の課題が残っており、高速かつ予測精度の高いアルゴリズム開発が必須な状況である。本研究では、最大共通部分グラフの抽出を行う際に、既知代謝経路の大部分において基質と生成物間の骨格構造に包含関係が認められる特徴を利用することで計算コストの大幅な低減を実現した。また、最大共通部分グラフから類似度を求める代わりに部分グラフ間の包含関係を計算することで予測精度を向上させた。先行研究では、15,050 代謝物間の経路予測において計算を完了できたのが 3.34%であったのに対し、提案手法では 34,653 代謝物間の経路予測が 2 週間程度で可能となった。予測結果から多くの既知代謝経路や新規(と思われる)代謝経路を確認することができた。さらに、代謝経路の視覚化を目的とし、自己組織化写像マップを応用した描画アルゴリズムを提案した。提案手法により、与えられた描画空間を有効に利用したネットワーク配置が可能となり、多くの化学構造を一枚の代謝マップ上に表示させることに成功した。開発したシステムは MetClassifier として WEB 上で公開している。また、一部のアルゴリズムは公共データベースである MassBank や NPEDIA の部分構造検索エンジンにも広く採用され現在稼働中である。