

論文内容の要旨

博士論文題目 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

氏名 田中健一

(論文内容の要旨)

植物界における構造決定されている代謝物は約 50,000 種と報告されており、KEGG では 16,021 代謝物、KNApSAcK では 40,957 代謝物がデータベース化されている。一方で、代謝経路に関しては、構造決定されている代謝物の約 10%程度に相当する情報がデータベース化されるに留まっている。代謝経路を推定することは生体による有用代謝物の高収率合成などの可能性があり非常に重要課題である。いままでに、フィンガープリントや最大共通部分グラフを用いて化学構造間の類似性を評価し、代謝経路を予測する手法が提案されているが、予測精度の低さや NP 困難なアプローチである等の課題があり、本研究では、化学構造から頻出連結部分グラフを抽出し、部分グラフの包含関係に着目した代謝経路予測を行うことで高速かつ予測精度の高いアルゴリズムの研究開発を進めた。このことを達成するために、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の包含関係に着目した。この特徴をヒューリスティックとして導入することで、頻出連結部分グラフ抽出処理の高速化を実現した。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。また、予測結果から多くの既知代謝経路や新規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。

氏名	田中 健一
----	-------

(論文審査結果の要旨)

平成 21 年 12 月 22 日に開催した公聴会の結果を参考に、平成 22 年 2 月 20 日に本博士論文の審査を実施した。以下に述べる通り、本博士論文は、本学位申請者が、独立した研究者としてバイオインフォマティクスを中心とした分野で研究開発活動が続けていくために必要な素養を備えていることを示すものである。

田中健一は、本論文において、生体内における代謝経路を予測することを目的に、代謝物の化学構造における環構造に着目し、構造式における側鎖をのぞくフレームワーク構造の包含関係を、原子を点、結合を辺であらわすグラフにより記述することによって得られる化学グラフの包含関係から導くアルゴリズムを提案した。この包含関係による代謝予測の精度は、いままでに提案された方法に比べてはるかに高く、また、処理速度の高速化を図ることができた。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ものが 34%であったが、提案手法では、全てにおいて経路予測の計算ができた。また、代謝経路の視覚化を目的とした新たな二次元配置法を自己組織化写像マップを応用し提案した。これらのアルゴリズムを搭載したソフトウェアは、フリーで公開しており、また、すでに、一部のアルゴリズムは公共データベースでも使用されるに至っており、社会への普及も進みつつある。

本論文は、メタボロミクス研究におけるバイオサイエンスと情報科学の融合における代謝経路予測法についての新たな方法を提案し、実際に予測することをおこなっており、情報科学とバイオサイエンスの境界領域の発展に貢献するものである。よって、本論文は、博士(工学)の学位論文としての価値があるものと認める。