

研究種目：基盤研究 (A)

研究期間：2007 ~ 2009

課題番号：19200022

研究課題名 (和文)

グラフ理論とカーネル法の融合による化学構造設計法

研究課題名 (英文)

New Methods for Designing Chemical Structures Using Graph Theory and Kernel Methods

研究代表者

阿久津 達也 (AKUTSU TATSUYA)

京都大学・化学研究所・教授

研究者番号： 90261859

研究分野：バイオインフォマティクス

科研費の分科・細目：情報学・生体生命情報学

キーワード：カーネル法、グラフアルゴリズム、特徴ベクトル

## 1. 研究計画の概要

従来の化合物に対するカーネル法では、個々の化学構造 (グラフ構造) を特徴ベクトル空間上の点に写像し、その空間上で分類や予測を行う。当該研究では従来手法の逆を行うことにより新規化学構造を計算機により導き出す計算手法を開発する。具体的には「特徴ベクトルからもとの化学構造を推定する」ことにより新規な化学構造を導き出す方法について、理論基盤構築および実用的アルゴリズムの開発を行う。推定や列挙の計算効率について理論的研究を行うとともに、実際の化学構造データに適用可能で様々な制約に対応可能な推定アルゴリズムを開発する。また、望ましいと思われる特徴ベクトルを選びだすためのスコア関数を開発する。さらに、WEB サーバーを開発して成果を利用可能とする。

## 2. 研究の進捗状況

### (1) 木状の化学構造の数え上げ

与えられた特徴ベクトルの制約を満たす木状の化学構造を数え上げるアルゴリズムを、木の数え上げ手法、および、**Detachment Cut** と呼ばれる枝刈り手法などを組み合わせることにより開発した。さらに実装を行い既存手法と比較することにより、その優位性を示した。一方、外平面的グラフの効率的な数え上げアルゴリズムなどに関する理論的成果も得た。

### (2) 化学構造数え上げサーバ開発

上で示したアルゴリズムを実装した化学構造数え上げサーバの第一版を完成させ **Enumol** を名づけ公開した。このサーバでは与えられた特徴ベクトルを満たす木状の化

学構造を列挙するものであり、結果はグラフィクス形式で見ることができ、また、ファイル形式でダウンロードすることもできる。

### (3) 立体異性体の数え上げ

木構造および外平面グラフ構造を持つ化合物に対して立体異性体を高速に数え上げるアルゴリズムを開発した。さらにこのアルゴリズムを実装し、既知の結果と比較することにより、その正当性を確認した。

### (4) 立体異性体を区別するカーネルの開発

光学異性体を区別できるカーネル関数を開発した。この手法は、既存の木パターンンググラフカーネルとよばれる、部分木の出現頻度に基づくカーネル関数を利用している。そして、部分木の出現頻度を計算する際に立体異性に関する情報をチェックすることにより立体異性体を区別する。光学異性体を含む構造活性相関データを用いて計算機実験を行い、その有効性を確認した。

## 3. 現在までの達成度

② おおむね順調に進展している。

(理由)

数え上げアルゴリズムの効率化やそれを実装した WEB サーバーの構築・公開を行うなど、おおむね順調に達成している。これらの成果は国際的にも認知されている。特に、ケモインフォマティクス分野における著名な研究者とある会議で 2010 年 5 月に面会した際に、制約付きで数え上げができる実用的アルゴリズムを開発しているのは、世界でもその研究者のグループとドイツの **MOLGEN** を開発しているグループと本研究グループくらいしかいないと言われた。

しかし、特徴ベクトルを指定するためのス

コア関数の開発や実用に向けての改良には長期的な研究が必要であることがわかり、計画を完全に達成するには至らなかった。

#### 4. 今後の研究の推進方策

上に示したように長期的な研究の必要性が判明したため、最終年度の前年度に新たな基盤研究A「離散的手法とカーネル法の融合による構造設計法」を申請し採択されたため、この課題においてスコア関数の学習法の開発や特徴ベクトルからの推定・列挙法の改良を行うことになった。

#### 5. 代表的な研究成果

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計4件)

- ① Y. Ishida, Y. Kato, L. Zhao, H. Nagamochi and T. Akutsu: Branch-and-bound algorithms for enumerating treelike chemical graphs with given path frequency using detachment-cut, Journal of Chemical Information and Modeling, in press.
- ② T. Imada, S. Ota, H. Nagamochi and T. Akutsu: Enumerating stereoisomers of tree structured molecules using dynamic programming, Lecture Notes in Computer Science, No. 5878, pp. 14-23, 2009.
- ③ H. Fujiwara, J. Wang, L. Zhao, H. Nagamochi and T. Akutsu: Enumerating tree-like chemical graphs with given path frequency, Journal of Chemical Information and Modeling, Vol. 48, pp. 1345-1357, 2008.
- ④ Y. Kato, T. Akutsu and H. Seki: A grammatical approach to RNA-RNA interaction prediction, Pattern Recognition, Vol. 42, pp. 531-538, 2008.

[学会発表] (計3件)

- ① 阿久津達也: 木構造および化学構造に対する特徴ベクトル: 埋め込み、検索、構造推定, 第12回情報論的学習理論ワークショップ, 2009年10月20日, 九州大学.
- ② Y. Ishida, L. Zhao, H. Nagamochi and T. Akutsu: Improved algorithms for enumerating tree-like chemical graphs with given path frequency, The 19th Int. Conference on Genome Informatics, 2008年12月1日, Gold Coast, Australia.
- ③ T. Urata, J.B. Brown, T. Tamura, T. Kawabata and T. Akutsu: A graph kernel

method incorporating chirality, 日本バイオインフォマティクス学会年会, 2007年12月18日, 日本未来館(東京).

[図書] (計0件)

[産業財産権]

- 出願状況 (計0件)
- 取得状況 (計0件)

[その他]

木状化学構造の数え上げサーバー EnuMol  
(<http://sunflower.kuicr.kyoto-u.ac.jp/tools/enumol/>)  
を開発・公開した。