

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 2 日現在

機関番号：14301

研究種目：基盤研究(A)

研究期間：2010～2014

課題番号：22240009

研究課題名(和文) 離散的手法とカーネル法の融合による構造設計法

研究課題名(英文) An Approach to Novel Structural Design by Combining Discrete Methods 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

交付決定額(研究期間全体)：(直接経費) 32,700,000円

研究成果の概要(和文)：本研究ではグラフアルゴリズムなどの離散的手法と機械学習手法の一つであるカーネル法を組み合わせることにより新規構造、特に新規化学構造を設計するための方法について研究を行った。その結果、パスの頻度からなる特徴ベクトルの上下限を入力として与えた場合の(i)木状化合物列挙アルゴリズム、(ii)ベンゼン環を含む木状化合物列挙アルゴリズム、(iii)任意の環1個を含む木状化合物列挙アルゴリズムの開発、そのいくつかを新たに実装したWebサーバー(Enumol Version 2)の再構築、化学構造比較の計算量解析などの成果を得た。

研究成果の概要(英文)：In this project, we studied a new approach for design of novel (chemical) structures by combining discrete methods (including graph theory/algorithms) and kernel methods. We developed algorithms for (i) enumerating tree-like chemical structures, (ii) enumerating tree-like chemical structures with benzene rings, (iii) enumerating tree-like chemical structures each having one additional ring, all from given upper and lower bounds of a feature vector consisting of frequency of labeled paths. Some of these algorithms were implemented in a significantly updated web server, Enumol Version 2. We also analyzed computational complexity of a chemical structure comparison problem.

研究分野：数理生物情報学

キーワード：特徴ベクトル カーネル法 化学構造 グラフ理論 立体異性体 構造列挙 木構造

1. 研究開始当初の背景

1990年代から2000年代にかけて機械学習分野などにおいて大きく発展をとげたカーネル法は、画像理解、自然言語処理、バイオインフォマティクスなど様々な分野に幅広く応用されてきた。カーネル法の基本的な考え方は、対象となるデータを、高次元（もしくは無限次元）空間上の点（特徴ベクトル）に写像し、写像した空間で判別や解析を行うというものである。逆に写像された空間の点からもとのデータに戻すことにより、新規化合物などを設計しようという研究が行われ、機械学習分野では Preimage 問題、化学分野では逆構造活性相関 (Inverse QSAR) の名のもとで研究が行われてきた。しかしながら、逆写像計算や列挙のためのヒューリスティックな計算手法はあるものの、厳密な計算手法はほとんど無いという状況にあった。そこで研究代表者らは平成19年度からの基盤研究(A)「グラフ理論とカーネル法の融合による化学構造設計法」において研究を進め、木状化学構造の高速列挙法の開発などの成果を得た。しかしながら、対象となる化学構造が限られていたため、より広いクラスの化学構造の高速な構造列挙を主目的として本基盤研究を充足した。

2. 研究の目的

本研究では上記の基盤研究(A)に引き続き、従来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について研究する。つまり、望ましい性質を持つと考えられる特徴ベクトルを何らかの方法で選びだした後で、「特徴ベクトルから、もとの構造を推定する」ことにより新規な構造を導き出す方法について、理論および実用アルゴリズムの両面から研究する。具体的には離散的手法（効率的列挙やグラフアルゴリズム）とカーネル法に基づき、以下を目標に研究を行う。

- (1) グラフの pre-image の推定問題、および、列挙問題の離散的側面、計算論的側面について現在の研究を深化させ、計算可能なクラスの拡大を図る。
- (2) 列挙と枝刈りに基づく実用的アルゴリズムを継続して開発する。ただし、付加情報なしに大きな構造を扱うのは困難であるので、適切な制約のもとで効率的列挙を行うアルゴリズムを開発する。
- (3) 望ましい性質を持つと考えられる特徴ベクトルを選択する手法を開発する。
- (4) これらを組み合わせた構造設計のプロトタイプを開発し、方法論の有効性を実証するとともに、それを実装した WEB サーバーを構築・公開する。

なお、本研究では具体性を重視するため化学構造設計に集中するものの、方法論自体は化合物以外の構造データにも適用可能なものであり、情報学分野全体の発展に大きく貢献すると考えている。

3. 研究の方法

本研究では新規構造設計手法の理論基盤および技術基盤を確立するために前基盤研究(A)に引き続き、異分野の研究者による共同研究を行う。具体的には、バイオインフォマティクスの専門家（阿久津、林田）、グラフ理論・アルゴリズムの専門家（永持）、実験的研究を行っている薬剤設計の専門家（川端）による協力体制で研究を行う。なお、実用的アルゴリズム、WEB サーバーの開発、計算機実験には多大な労力が必要となるため、ポスドク研究員を1名雇用する。さらに、計算機実験や結果の評価などのために大学院生の協力も得る。

具体的には以下のトピックについて研究を進める。

- (1) サポートベクター回帰などの回帰手法を用いて化合物に対するスコア関数学習法を開発し、さらに学習された関数から適切な特徴ベクトルを選択する手法を開発する。
- (2) これまで行ってきた立体異性体列挙のアルゴリズムを発展、完成させる。
- (3) これまで開発してきた木状構造を持つ化合物に対する列挙アルゴリズムを発展させ、より広いクラスに適用可能なものとする。
- (4) より柔軟な制約のもとで列挙できるようなこれまでのアルゴリズムを拡張する。
- (5) これまでに開発・公開してきた WEB サーバー EnuMol の機能強化・再構築を行う。

4. 研究成果

上記で述べた目的や方法に従い研究を行い、以下の成果を得た。しかしながら、特徴ベクトルの選択法については検討は行ったものの具体的な成果を得るには至らず今後の課題として残された。

- (1) 立体異性体を区別可能なカーネル関数
前基盤研究(A)で着手していた立体異性体を区別可能なカーネル関数、および、それを用いたサポートベクター回帰による化合物活性予測法を完成させ、論文としてまとめた。
- (2) 外平面グラフ構造を持つ化学構造の立体異性体列挙
前基盤研究(A)で着手していた木状の化学構造の立体異性体列挙アルゴリズムを完成させるとともに、そこでの手法を拡張することにより外平面グラフ構造を持つ化学構造の立体異性体列挙アルゴリズムを開発した。
- (3) 特徴ベクトルの上下限制約のもとでの構造列挙

これまでの定式化では制約が強く解が存在しない場合が数多くあるため、特徴ベクトルの各要素の上下限を与えるというより弱い制約のもとで、その制約を満たす木状化学構造を列挙するアルゴリズムを開発した。さらに、二重結合を含む場合の同様の列挙を高速化するために二段階法を開発した。そして、これらの有用

- 性を計算機実験により確認した。
- (4) ベンゼン環を含む木状の化学構造の列挙
ベンゼン環を持つ場合には、オルト、メタ、パラなどの異性体を区別する必要がある。立体異性体列挙の際に用いた動的計画法に基づく手法を適用することにより、上記異性体を区別可能してベンゼン環を含む木状の化学構造を列挙するアルゴリズムを開発した。なお、平成 26 年度よりの新たな基盤研究(A)において、より洗練され拡張性の高い列挙アルゴリズムが開発中となっている。
 - (5) 環を 1 個含む化学構造の列挙
パス頻度に基づく特徴ベクトルの上下限制約のもとで、(ベンゼン環以外にも)環を 1 個含む構造まで許した化学構造を列挙するアルゴリズムを開発した。計算機実験の結果、出力構造 1 個あたりの計算時間は木状化合物の場合と同程度であることがわかり、開発手法の有用性を確認することができた。
 - (6) 幅優先探索法に基づく化学構造の列挙
上記(3)-(5)の研究では深さ優先探索に基づき木状化合物を数え上げるという方法論を用いてきたが、別のアプローチも検討するために、幅優先探索に基づく木状化合物の列挙手法を開発した。深さ優先探索の場合と異なり計算量に関する理論的な解析はできておらず、また、特徴ベクトルによる制約への対応もできていないが、分子式のみを与えての列挙については計算機実験の結果、深さ優先探索に基づく手法と同程度以上の速度で動作することを確認することができた。
 - (7) 化学構造の類似性判定の計算量解析
2 個の化合物の構造比較の定式化の一つとして、最大共通部分グラフ検出問題がある。この問題は一般には NP 困難であるが、木に類似した構造(次数制約つき almost k-tree)については多項式時間で計算可能であることが知られていた。しかしながら、他のクラス、特に多くの化合物を含むグラフのクラスについては、その計算複雑度が知られていなかった。研究の結果、この問題は、次数制約のある外平面的グラフについて多項式時間で解け、一方、k=11 の部分 k 木については次数制約があっても NP 困難であることが判明した(なお、その後の研究で k=4 に改良)。
 - (8) WEB サーバ第 2 版の開発
以前より開発してきた化学構造の数え上げを行う WEB サーバのユーザインターフェイスを一新するとともに、上下限を与えた場合の木状化学構造の列挙、ベンゼン環を含む木状化合物の列挙の両機能を追加した。これらの大幅な改良を行ったため、WEB サーバを EnuMol Version 2 として新たに公開した(図 1、図 2)。



図 1 WEB サーバ: EnuMol Version 2



図 2 EnuMol Version 2 によるベンゼン環を含む構造の列挙

- (9) タンパク質複合体予測および代謝ネットワーク解析
関連する研究として、タンパク質複合体予測のためのカーネル関数の開発、代謝流束解析に基づく代謝ネットワークにおける遺伝子ノックアウトの影響指数の開発などを行った。

なお、本研究で数多くの成果が得られたことと更なる進展と新たな展開を図るため、期間を 1 年短縮して終了し、平成 26 年度より新たな基盤研究(A)を発足させた。

5. 主な発表論文等

(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 27 件、すべて査読有)

Y. Zhao, M. Hayashida, J. Jindalertudomdee, H. Nagamochi, T. Akutsu: Breadth first search approach to enumeration of tree-like chemical compounds, Journal of Bioinformatics and Computational Biology, 4, 1343007 (19 pages), 2013.

DOI: 10.1142/S0219720013430075

<http://hdl.handle.net/2433/196770>

T. Akutsu, T. Tamura: A polynomial-time algorithm for computing the maximum common connected edge subgraph of outerplanar graphs of bounded degree, Algorithms, 6, 119-135, 2013.

DOI: 10.3390/a6010119

<http://hdl.handle.net/2433/172448>

T. Akutsu, H. Nagamochi: Comparison and enumeration of chemical graphs, Computational and Structural Biotechnology Journal, 5, e201302004 (9 pages), 2013.

DOI: 10.5936/csbj.201302004

<http://hdl.handle.net/2433/172449>

M. Shimizu, H. Nagamochi, T. Akutsu: Enumerating tree-like chemical graphs with given upper and lower bounds on path frequencies, BMC Bioinformatics, 12, Suppl 14, S4 (9 pages), 2011.

DOI: 10.1186/1471-2105-12-S14-S3

<http://hdl.handle.net/2433/159708>

T. Imada, S. Ota, H. Nagamochi, T. Akutsu: Efficient enumeration of stereoisomers of outerplanar chemical graphs using dynamic programming, Journal of Chemical Information and Modeling, 51, 2788-2807, 2011.

DOI: 10.1021/ci200084b

<http://hdl.handle.net/2433/139571>

J.B. Brown, T. Urata, T. Tamura, M. A. Arai, T. Kawabata, T. Akutsu: Compound analysis via graph kernels incorporating chirality, Journal of Bioinformatics and Computational Biology, 8, Suppl 1, 63-81, 2010.

DOI: 10.1142/S0219720010005117

[学会発表](計 19 件)

Y. Zhao, M. Hayashida, J. Jindalertudomdee, H. Nagamochi, T. Akutsu: Breadth first search approach to enumeration of tree-like chemical compounds, The 24th International Conference on Genome Informatics, 2013/12/18, Biopolis (Singapore).

Jira Jindalertudomdee, 林田 守広, 阿

久津 達也: ベンゼン環を持つ木状化学構造の幅優先探索による列挙手法, 第 36 回情報化学討論会, 2013/11/7, 筑波大学(つくば市).

鈴木政喜, 永持仁, 阿久津達也: 上下限パス頻度に基づいた環構造を持つ化合物の列挙法, 第 35 回情報化学討論会, 2012/10/4, 広島大学(東広島市).

M. Shimizu, H. Nagamochi, T. Akutsu: Enumerating tree-like chemical graphs with given upper and lower bounds on path frequencies, The 22nd International Conference on Genome Informatics, 2011/12/5, Busan (Korea).

T. Akutsu: Inference and enumeration of chemical structures from feature vectors, IPAM Workshop II: Optimization, Search and Graph-Theoretical Algorithms for Chemical Compound Space, 2011/4/15, UCLA (LA, USA).

J.B. Brown, T. Urata, T. Tamura, M. A. Arai, T. Kawabata, T. Akutsu: Compound analysis via graph kernels incorporating chirality, The 21st International Conference on Genome Informatics, 2010/12/16, Hangzhou, (China).

[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]

木状化合物列挙 WEB サーバ EnuMol :
<http://sunflower.kuicr.kyoto-u.ac.jp/tols/enumol/>

6. 研究組織

(1) 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: 9 0 2 6 1 8 5 9

(2) 研究分担者

永持 仁 (NAGAMOCHI, Hiroshi)

京都大学・情報学研究科・教授

研究者番号: 7 0 2 0 2 2 3 1

川端 猛夫 (KAWABATA, Takeo)

京都大学・化学研究所・教授

研究者番号: 5 0 2 1 4 6 8 0

林田 守広 (HAYASHIDA, Morihiro)

京都大学・化学研究所・助教

研究者番号: 4 0 4 0 2 9 2 9