

平成 30 年 5 月 22 日現在

機関番号：33917

研究種目：若手研究(B)

研究期間：2015～2017

課題番号：15K15944

研究課題名(和文) 効率的な離散最適化手法を用いた化合物の自動立体構造推定法の開発

研究課題名(英文) The application of efficient discrete optimization algorithms to the development of a computer automated method for the elucidation of chemical structures

研究代表者

小市 俊悟 (KOICHI, Shungo)

南山大学・理工学部・准教授

研究者番号：50513602

交付決定額(研究期間全体)：(直接経費) 2,200,000円

研究成果の概要(和文)：本研究では、構造とNMRスペクトルが既知の化合物データベースを利用して、新規化合物の構造を、そのNMRスペクトルから計算機により自動で推定する方法の確立を最終目的として、その実現のために解くことが必要な様々な組合せ最適化問題を、効率的な離散最適化手法を中心に応用することで解決することを具体的な課題とした。はじめに、グラフ推定問題の効率的な離散最適化解法を応用して、推定法の適用範囲を高い対称性を持つ構造にまで広げることに成功した。また、検索された部分構造について、それらを推定構造として重ねていく際に必要となる順番の生成法の改良に取り組み、最大木問題の解法を応用して新たな生成法を開発した。

研究成果の概要(英文)：The aim of this project is to establish a computer automated method for the elucidation of chemical structures using an NMR spectrum-structure database. For this purpose, we used several efficient discrete optimization algorithms to solve some combinatorial problems arising in the development of the method. We first applied an algorithm for the graph inference problem to handle and elucidate highly symmetric structures, which successfully extended the applicability of the method. Next, we addressed a problem in generating orders of retrieved substructures. The generation has to be exhaustive because, for each of the orders, the substructures are merged in the order to form a whole structure, which may depend on the order. This generation problem was resolved with the help of an algorithm for maximum weight spanning tree problem.

研究分野：情報化学

キーワード：離散最適化 化合物構造推定 データベース 核磁気共鳴分析法

## 1. 研究開始当初の背景

化合物の構造決定に用いられる主要な分析方法の一つに核磁気共鳴分光法(以下、NMR)がある。NMRによって得られるNMRスペクトルを解析することで構造を決定するが、構造を推定するためには経験も必要とされ、専門家であっても構造決定は容易なことではない。実際、誤った構造が報告されることもあり、それがデータベースに登録されていることもある。

このようなNMRスペクトルに基づく構造決定をデータベースと計算機を利用して実現しようという試みは、化学に計算機が導入された1960年代から取り組まれている。しかし、対象とする構造を限定しない汎用的かつ実用的な自動構造推定法は、未だ実現していないと言えよう。その主たる原因は、その実現のために解決すべき課題の多くが、組合せ最適化問題と呼ばれる、現代の計算機をもってしても、一般に解くのが難しい問題となるからである。例えば、図1にあるように、推定構造を構築する際には、部分構造間の共通部分を見つけ出さなければならないが、この問題は計算量理論において難しい問題とされており、解くのは容易ではない。

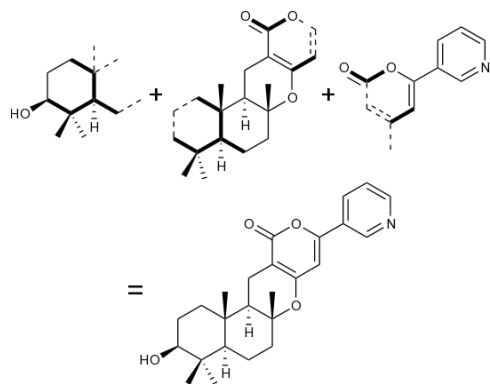


図1. データベースから検索された部分構造(上段の構造)の共通部分(太線部)とそれに基づいて構築される推定構造(下段の構造)。

一方で、化合物のデータベースは着実に大きくなっているため、それを利用した計算機による自動構造決定は、以前にも増して、その確立が望まれている。また、組合せ最適化問題についても、この20~30年の間に研究が進み、効率的な解法が開発された問題も多くなってきた。このような状況を踏まえて、組合せ最適化問題に対する最新の解法を応用し、化合物の自動構造推定法を確立しようという研究を、ここ10年ほど進めてきた。これまでもすでに一定の成果をあげてきたが、その実用化のためには、解決しなければならない問題が残っている。

## 2. 研究の目的

本研究課題における研究の最終目標は、計算機による化合物の自動構造決定法を確立することである。その実現のためには、様々な組合せ最適化問題を解く必要があるが、一般に組合せ最適化問題を解くには多大な計算時間を要するため、効率性を重視した解法を採用しないと、実用的な時間で構造決定を遂行することができない。そこで、効率的な離散最適化手法を最大限に活用することを方針としてきた。本研究課題でも、効率性を重視した問題の解決を図る。

本研究課題では特に、次のような問題を解決することを具体的な課題とした。

- (1) 高い対称性を持つ構造を推定する場合の推定法の確立
- (2) 部分構造を重ね合わせる際に必要となる部分構造の順番を生成する方法の改良
- (3) データベースの巨大化を見据えたデータベースの再設計およびインデクシング

課題(1)は自動構造推定法の適用範囲を拡張するために解決を目指すものである。課題(2)は、これまで素朴な方法を実装していた部分について、実用性と信頼性を高めるために、効率性を維持しつつ網羅性が高い方法に置換することを目指すものである。課題(3)は、今後のさらなるデータベースの巨大化を見据え、同じ結果を得るにしても、より高速な処理が可能となるように、データベースを改良することを目指すものである。

## 3. 研究の方法

「2. 研究の目的」で挙げた課題のうち、まず、課題(1)について、その問題点から説明する。本研究では、高い対称性を持つ構造とは、同一部分構造を複数持つ化合物のことを意味する。NMRスペクトルを構成する化学シフト値とは、化合物中の各炭素が、それを取り巻く構造に依存して示す数値であるから、同一部分構造を複数持つ化合物は同じ化学シフト値を複数持つ。このような構造に対して、推定構造を構築する際に、本研究でこれまで採用してきた「データベースから検索した部分構造に対して、それらの共通部分を重ねることで推定構造を組み立てる」方法では構造推定ができない。同一の部分構造は一つの構造に重ね合わせられることになるため、同一の部分構造を複数用いて、それらを重ねないことではじめてできるような構造は構築できないのである。図2は、その一例を示している。全体構造Cを構築するためには部分構造Aを二つ使用する必要があるが、二つの部分構造Aは完全に重なるため、構造Cの青い部分のみが出力される。

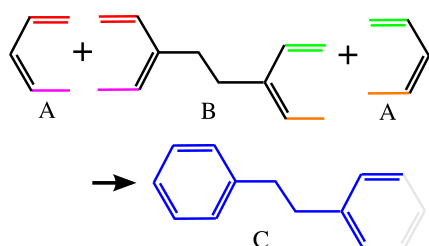


図2．対称性が高い構造の推定．

この問題に対して、グラフ推定問題に対する手法を応用することにした。図3は分子を対象とした場合のグラフ推定問題の一例である。図3の右側にあるような各原子の個数と各結合の個数を入力として、それらを満たすような左側の構造を出力する問題がグラフ推定問題である。化学シフト値の個数と、データベースから検索された部分構造中の炭素数から必要な原子や結合の個数を計算し、それを入力としたグラフ推定問題を解くことで推定構造を得ようというのが目指す手法である。グラフ推定問題は、効率的な離散最適化手法が存在する組合せ最適化問題の一つとして知られている。

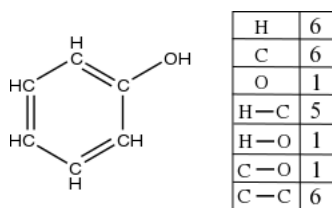


図3．分子を対象としたグラフ推定問題の一例。右側にある各原子の個数と各枝の個数から、それを満たすような左側の構造を出力する問題。

次に、「2．研究の目的」で挙げた課題のうち、課題(2)について説明する。上でも述べたように、推定構造は、データベースから検索された複数ある部分構造を共通部分に基づいて重ね合わせることで基本的には構築する。その際、最終的に構築される推定構造は、それらの部分構造を重ね合わせていく順番に依存し得るため、正解構造を逃さないためには可能な順番を網羅する必要がある。しかし、その一方で部分構造の個数が増えると、指数的に可能な順番の個数が増えるため、可能な順番の全列挙は実践的ではない。正解を逃さない程度に限定し、かつ網羅的な順番の生成が肝要となる。このような順番の生成について、これまでは全列挙の方法を基本的に生成される順番において連続する部分構造間に一定以上の共通部分があるものに限って生成するようにしていた。この方法は、実装が容易という利点があったが、対象となる部分構造によっては、順番の網羅性を欠く可能性があった。この問題を解決しつつ、効率性を確保することを本研究課題では目指した。そのために、データベースから検索された部分構造のネットワークを作成し、そのネ

ットワークの最大木を求め、さらに、その最大木を利用した順番の生成を行う方法を考案した。ここで部分構造のネットワークとは、各部分構造がネットワークの頂点に対応し、各枝(リンク)には端点となる部分構造間の共通度が重みとして与えられたものである。したがって、そのネットワークの最大木は、共通度に基づく部分構造間の関係を抽出したものと考えられ、それを利用すれば、十分な網羅性を持った順番の生成が可能となると考えた。最大木を求める問題もまた効率的な離散最適化手法が存在する組合せ最適化問題として知られている。

「2．研究の目的」で挙げた課題のうち、残る課題(3)について説明する。現状のデータベースは非常に素朴なものであり、1つの化合物データが1つのディレクトリに保存されたものである。このようなデータベースに検索をかける場合、並列化するにしても基本的には総当たりの検索が必要となる。しかし、データベースは着実に巨大化していくので、このような方法では、実践的な時間で検索を完了することが難しくなる。そこで、問い合わせが多い部分について、インデクシングにより、検索効率を高めることを計画した。また、XMLおよびXMLに基づく検索方法の採用などの現代的なデータベース設計を活用することを目指した。XMLは多くのデータベースで採用されているマークアップ言語であり、XMLデータベースのために開発された多様な既存ソフトウェア等が利用できるという利点もある。

#### 4．研究成果

「2．研究の目的」で挙げた課題のうち、まず、課題(1)について成果を述べると、グラフ推定問題に対する効率的な手法を応用することで、高い対称性を持つ化合物についても構造を推定することが可能となった。これにより、推定が可能な構造の範囲を拡張することができた。グラフ推定問題を応用する一つの利点として、対称性を陽に扱わないことで処理効率を上げることを想定していたが、この点については、対称性に関する情報を追加する必要があることが研究を進める上で判明した。しかし、そのための処理については、以前に別の研究課題で対称性を扱った際に利用した方法を流用することが可能であるとわかり、この問題も解決することができた。この成果については、日本コンピュータ化学会 2015 秋季年会において発表し、また、論文としても雑誌 Journal of Computer Chemistry, Japan に掲載されるに至った。

次に、「2．研究の目的」で挙げた課題のうち、課題(2)について述べる。最大木については既存の効率的な離散最適化手法を用いて容易に求めることができたが、求めた最大木をもとに部分構造の順番を生成する手法については、新たに考案・実装する必要があった。これについては、深さ優先探索と

呼ばれるネットワークの標準的な探索方法に、順列生成の手法を組み込むことで、必要な順番を網羅的に生成できるようになった。これにより効率性を維持しつつ、網羅性を高めることができたと考える。考案した手法は、最大木のような木構造のネットワークに限らず、一般のネットワークに適用可能である。特に、すべての頂点間に枝がある場合、頂点のすべての順番を生成するので、一般的な順列生成法の自然な拡張にもなっている。この成果により、以前より大きな構造を推定することも可能になると考えている。

「2. 研究の目的」で挙げた課題のうち、残る課題(3)について述べる。データベースの再設計については、本研究で開発を進めてきた自動構造推定法の公開に向けて、データベースの仕様等を現時点で変更することが難しい状況となったため、少し異なるデータとデータベースを用いて、XML データベースの活用について研究を進めた。それは、Togni 試薬と呼ばれる試薬とその亜種に関して、それらの安定性や反応性をデータ指向的な方法で分析するものである。データはかなり膨大なものであったが、XML の特徴を活用し、データ分析を効率的に進めることができた。その成果は、雑誌 *Physical Chemistry Chemical Physics* に掲載されるに至った。本研究課題で取り扱う分子構造のような数値データではない構造に関するデータをどのようにデータとして保持すべきかについては、まだ検討することが多いと考える。データの取得方法が多様化するのに合わせて、このような構造に関するデータの処理方法の確立は、本研究に限らず、データ分析の分野でも今後ますます重要になると考える。

以上まとめると、課題(3)については、まだ開発・改良の余地があるが、課題(1)、(2)については、当初予定していた成果が得られたと考える。

その他、入力に結合の情報が付加された場合の推定法の開発も行った。これによって、より信頼性の高い推定結果が得られる。また、データベースに誤った構造が登録されていることがあることを述べたが、部分構造の検索機能を利用することで、そのような誤ったデータを発見することが容易になった。つまり、同じ NMR スペクトルに対して異なる構造の化合物がデータベースに登録されている場合、一方は誤ったものである可能性が極めて高い。このような場合、構造訂正が行うが、それらを成果として5件ほど、関連学会にて報告することができた。

## 5. 主な発表論文等

〔雑誌論文〕(計2件)

Shungo Koichi, Benjamin Leuthold, Hans Peter Lüthi, Why do the Togni reagent and some of its derivatives exist in the high-energy hypervalent iodine

form? New insight into the origins of their kinetic stability, *Physical Chemistry Chemical Physics*, 査読有, 19 巻, 2017, 32179 - 32183.

DOI:10.1039/C7CP05943D

Shungo Koichi, Hiroyuki Koshino, Hiroko Satoh, Handling of Highly Symmetric Molecules for Chemical Structure Elucidation in a CAST/CNMR System, *Journal of Computer Chemistry, Japan*, 査読有, 14 巻, 2015, 193-195  
DOI: 10.2477/jccj.2015-0067

〔学会発表〕(計7件)

越野広雪, 栗澤尚暎, 木村賢一, 小市俊悟, 佐藤寛子, CAST/CNMR システムを用いた Clerodane 型ジテルペンの <sup>13</sup>C NMR の帰属と構造の評価, 第56回 NMR 討論会, 2017年11月14-16日, 首都大学東京・南大沢キャンパス

越野広雪, 小市俊悟, 佐藤寛子, CAST/CNMR システムを用いたプレニル化されたフェノール類の構造訂正, 第61回香料・テルペンおよび精油化学に関する討論会, 2017年9月9-11日, 金沢工業大学扇が丘キャンパス

Shungo Koichi, Hans Peter Lüthi, Development of a data-centric stability test for iodanes by using the support vector machine, Competence Center for Computational Chemistry (C4) Workshop, 2017年1月26日 Zurich  
越野広雪, 小市俊悟, 佐藤寛子, CAST/CNMR システムを用いた <sup>13</sup>C NMR シフト値による様々な有機化合物の構造訂正, 第55回 NMR 討論会, 2016年11月16-18日, 広島国際会議場

越野広雪, 小市俊悟, 佐藤寛子, CAST/CNMR システムを用いた cis-デカリン骨格を有するテルペノイドの立体化学の評価, 第60回香料・テルペンおよび精油化学に関する討論会, 2016年10月29-31日, 東京農業大学オホーツクキャンパス  
Hiroyuki Koshino, Shungo Koichi, Hiroko Satoh, Structural Revision of Pyrone-related Natural Products by Using CAST/CNMR System, EUROMAR 2016, 2016年7月3-7日, Aarhus

小市 俊悟, NMR 分子構造解析システム CAST/CNMR における対称性が高い構造の自動構造推定について, 日本コンピュータ化学会 2015 秋季年会, 2015年10月30-31日, 函館市地域交流まちづくりセンター

## 6. 研究組織

(1) 研究代表者

小市 俊悟 (KOICHI, Shungo)  
南山大学・理工学部・准教授  
研究者番号: 50513602