

ヘテロ原子に注目したパスフラグメントによる化学構造データマイニング

Chemical Structure Data Mining using Path Fragments with Terminal Hetero Atoms

藤島悟志*¹
Satoshi Fujishima

高橋由雅*¹
Yoshimasa Takahashi

岡田孝*²
Takashi Okada

*¹ 豊橋技術科学大学 工学部 知識情報工学系
Department of Knowledge-based Information Engineering, Toyohashi University of Technology

*² 関西学院大学 理工学部 情報科学科
Department of Informatics, School of Science and Technology, Kwansai Gakuin University

In structure data mining of chemicals, many of the method mine and use linear fragments from their chemical structures as knowledge on their structure-activity relationships. However, in most cases, many trivial fragments that are removed in the mining process are mined and examined for the efficiency. Then, the extraction and the synthesis of characteristic feature, which described in the rules derived by a mining method, are performed by the hand of an expert and it took too long to reach the final characteristic structure. In the present work, we propose a new approach to structure data mining of chemicals using path fragments that both of their terminal atoms are hetero atoms. It also allows us to do rational refinement of the seed (a fragment) according to the same restriction when the fragment is extended. The approach has been computerized and tested. The details of the system will be discussed with illustrative examples.

1. はじめに

近年、様々な分野においてデータマイニングあるいはチャンス発見と呼ばれる、大量のデータから有用な知識を発掘するための新たな技術の確立に多くの期待が寄せられている。構造活性相関への応用についても様々な研究が行われているが、部分構造などの構造記述子を用いたマイニング結果は、共通して、最終的な知識(ファーマコフォア)そのものは得られないという問題がある。最終的なファーマコフォアの抽出には、マイニング結果を利用した専門家の手作業による解析作業が必要となる。著者らはこれまで、より効率的な構造特徴の抽出および知識発見を目的とし、マイニングによって得られた部分構造を種として、対象化学構造群により特徴的な構造に精錬していくシステムの開発を行ってきた[藤島 06]。本研究では、新たなアプローチとして、パスフラグメントを用いた精錬過程を考案し、実データによる解析を行ったので報告する。

2. 方法

2.1 パスフラグメント

与えられた化学グラフ中の任意の二つの原子と、それらの間の最短パスの長さで表される部分グラフをパスフラグメントと定義する。図 1 にパスフラグメントの例を示す。図 1 左の化学グラ

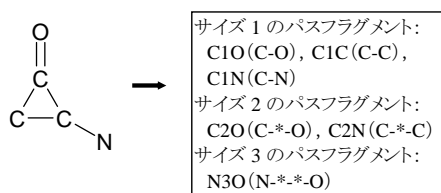


図 1. パスフラグメントの例

フから得られるパスフラグメントをサイズごとに列挙すると右のようになる。本研究ではパスの途中に位置する原子の種類や結合の多重度は考慮しない。また、末端原子がヘテロ原子のみで構成されるパスフラグメントのみを抽出する。これは、抽出したフラグメントの末端原子が炭素であるものは、積極的に化学構造の特性や活性に関与しないためである。またこれにより、生成フラグメントの数を抑えることもできる。よって、図 1 から最終的に得られるパスフラグメントは N3O のみとなる。

2.2 パスフラグメントの精錬

先行研究で開発したシステムでは、入力された種フラグメントを核として、原子種、結合種を付加させていく。つまり、種フラグメントの構成自体に変化はない(図 2)。本研究では、新たなアプローチとして、種フラグメント(パスフラグメント)の末端原子はそのまま、末端原子間の幅を変化させる(フラグメントの伸縮を行う)ことで、緩やかな精錬過程を試みる。図 3 にその概念図を示す。

例として仮想クラス A に対して解析を行った例を示す。クラス A に属する 50 構造中 20 構造が種フラグメント(O6N)を支持している。精錬によって、O7N, O10N では新たに、7 件、5 件の構造がそれぞれ支持されるとすると、最終的に得られる知識は、O{6,7,10}N となり、32 件の構造が支持していることになる。

このようにして、注目活性に特徴的なフラグメントを探る。精錬過程で得られたフラグメントや各フラグメントを支持する注目活性構造の数などを統合して、最終的な精錬フラグメント(知識)を提供する。

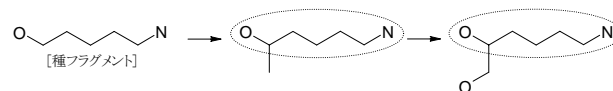


図 2. 先行研究の精錬過程。

連絡先: 藤島悟志, 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘
1-1 豊橋技術科学大学 知識情報工学系, Tel: 0532-44-6876, fujisima@mis.tutkie.tut.ac.jp

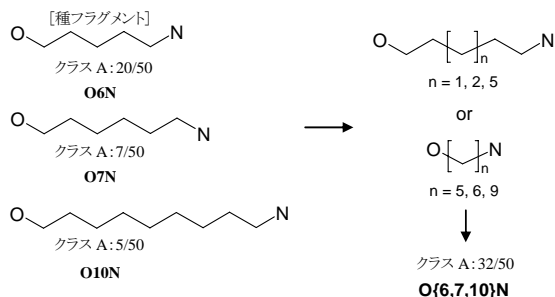


図 3. 提案する精練過程.

3. 実データによる特徴構造の抽出

提案手法の有用性を計るために、実データを用いて解析実験を行った。データセットは、米国 MDL 社の開発医薬品データベース MDDR (MDL Drug Data Report) [MDL 01] に記載されている、3 種の異なる受容体 (D1, D2, Autoreceptor) に作用する Dopamine Agonist (370 件) を用いた。

ここでは、Dopamine D2 Agonist に対して解析を行った結果を示す。D2 Agonist に特徴的なパスフラグメントの抽出実験として、S1O を種フラグメントとした解析を行った。図 4 に解析時に抽出されたパスフラグメントとその支持構造の例、そして最終的に得られた知識表現を示す。

O1S を支持する事例数は全体で 17 件あり、その内 14 件が D2 Agonist であることが分かり、支持事例数は少ないが O1S が D2 Agonist において特徴的であると考えられる。これを種フラグメントして解析を行うと、まず O2S が得られた。この支持事例数は全体で 17 件であるが、そのすべてが D2 Agonist であった。またその中の 6 件が O1S の支持事例とは異なる新しいものであることが分かった。同様に解析を進めると、O10S が得られ、支持事例 21 件中 17 件が D2 Agonist であり、新規構造として 5 件存在することが分かった。O11S では新たに 3 件が新規構造として得られた。これらの結果から、パスフラグメント O{1,2,10,11}S

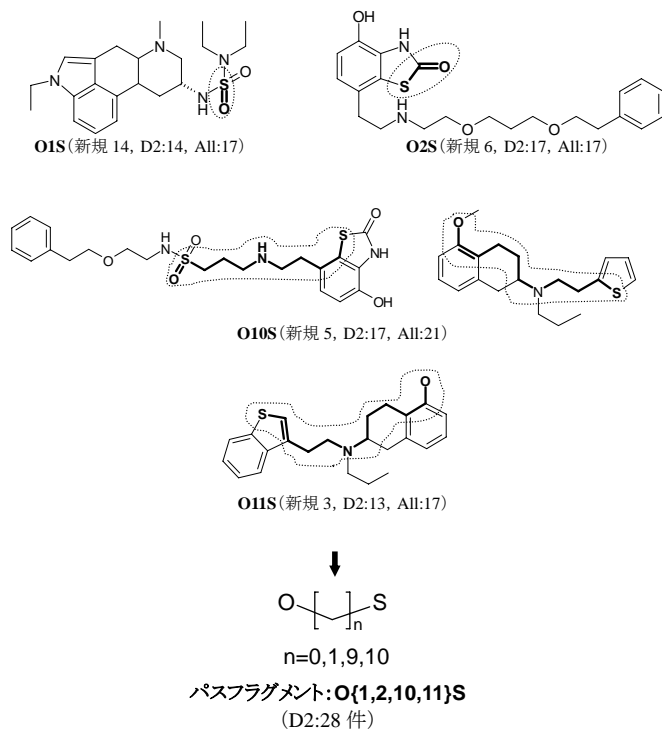


図 4. Dopamine D2 Agonist の解析結果 (一例)

が得られ、それを支持する事例数は 28 件である最終的な知識が抽出される。また、新規の支持構造が得られないため図 4 に示さなかったが、他のパスフラグメントとして O8S や O12S も支持事例数の約 9 割が D2 Agonist である結果も得られた。これらも含めた場合のパスフラグメント表記は、O{1,2,8,10,11,12}S と記述できる。解析結果から末端原子が O (酸素) と S (硫黄) で、その原子間が図 4 で示した距離で構成されるフラグメントは高い確率で D2 Agonist 活性を保持することが考えられる。また図 4 に示した構造は、先行研究 [岡田 04] で抽出された D2 Agonist に特徴的な構造と同一のものも得られていることから、本手法によって有益な情報が得られていると考える。

精練過程において、フラグメント中のヘテロ原子を考慮した解析も試みた。図 5 にその一例を示す。種フラグメントとして、O6N5S を与えた場合に、O と N 間、N と S 間の距離を伸縮させて精練を行った。N と S 間の距離は 5 のままで、O と N の距離が 6,7,8 の場合で D2 Agonist に特徴的あることが分かった。このように、種フラグメントで与えた 3 ヘテロ原子間の距離を解析することで図 4 の結果とは異なる視点からの解析が行えることは注目に値すると考える。

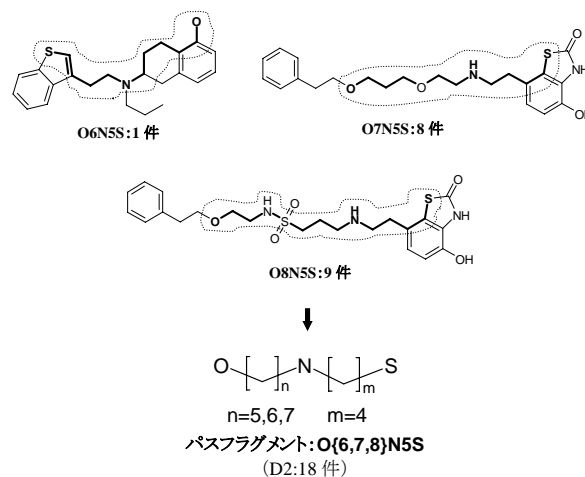


図 5. フラグメント中のヘテロ原子を考慮した Dopamine D2 Agonist の解析結果 (一例)

4. まとめ

化学構造の特徴的リード構造洗練における新たなアプローチとして、パスフラグメントの末端原子はそのまま、末端原子間の幅を変化させる手法を考案し、緩やかな精練過程を試みた。本手法の観点でも精練が行え、注目活性に特徴的なフラグメントが得られる結果得られた。精練過程において、本手法も併用することで、部分構造などの構造記述子を用いた構造活性相関研究での効果的な知識獲得に大きく寄与するものと考えられる。

参考文献

- [藤島 06] 藤島悟志, 岡田孝: 化合物の線形フラグメントに基づいた特徴的リード構造抽出システムの開発, 第 20 回人工知能学会全国大会論文集, 3A3-01 (2006).
- [MDL 01] MDL: Drug Data Report, MDL, ver. 2001.1, (2001).
- [岡田 04] 岡田孝, 高橋由雅, 加藤博明, 山川真透, アクティブマイニングによる化学物質群からのリスク分子発見, 特定領域研究 (情報洪水時代におけるアクティブマイニングの実現) 研究成果報告書, pp.411-422, (2004).