

分子行列の固有ベクトルを用いた構造特徴プロファイリング

Structure Characteristics Profiling using Eigenvector of Molecular Matrix

齋藤 佳介
Saito Keisuke

高橋 由雅
Yoshimasa Takahashi

豊橋技術科学大学大学院 工学研究科 情報・知能工学専攻
Department of Computer Science and Engineering, Toyohashi University of Technology

Structure profiling based on quantification and numerical characterization of a chemical structure provides us an important basis for similar structure searching, machine learning for classification of chemicals, data mining for structure-activity relationships etc. In the present work, we propose a structure profiling method which is based on eigenvector of molecular matrix. We also investigated the relationships between eigenvector elements and intra molecular environment of vertex atoms. The results suggest that eigenvector elements of a molecular matrix can be used to represent the structural environment of individual vertex atoms.

1. はじめに

化学構造情報を離散・数値化し、数値的に記述することは、データベースに対する類似性検索や機械学習に基づく薬物活性クラス分類、構造活性相関のためのデータマイニングなど、様々な分野で重要な基礎となる。当研究室では、化学物質の類似性に着目した構造プロファイリング手法として Topological Fragment Spectra(TFS)[Takahashi 98]法を提案し、機械学習による薬化合物の活性クラスの分類/予測などに対する有用性を示してきた[Takahashi 02] [Kawai 08]。しかし、TFS 法は構造全体の類似性の評価には良好な結果を与えるものの、個々のピークに対しては必ずしもユニークな部分構造が対応するものではなく、一つのピークに対して、しばしば、多数の異なる部分構造が対応する場合があることが知られている。このことから、構造活性相関知識の発見といった化学データマイニングの観点からは、TFS の特定のピークを具体的な部分構造に対応付けることは困難である。

そこで本研究では新たに、分子行列の固有ベクトルを用いた構造プロファイリング手法を提案するとともに、化学データマイニングへの応用の可能性について検討を行った。

2. 関連研究

化学構造式をコンピュータで扱うとき、化学構造式は一種のグラフとみなせることから、化学構造の数学的考察にはグラフ理論を適用することができる。本研究で使用する分子行列とは、グラフの隣接行列の考えを拡張し、グラフ中の点や辺の重み付けにより原子種や結合の種類を含めた原子結合行列のことを指す。先に宮脇らは、与えられた分子グラフに対し、対応する分子行列から固有値解析によって固有値を求め、元の分子グラフの構造記述子として用いることを提案するとともに、類似構造解析への応用を試みている[宮脇 12]。彼らは単純グラフ表現レベル(分子のトポジカルな単純骨格表現)で同型となる種々の分子グラフを用いた構造類似性評価に対する応用例の中で、分子行列から得られた固有値が頂点原子の種類や結合タイプの違いによる構造差異を反映していることを明らかにするとともに、最大固有値から降順に並べた固有値パターンベクトルによる類似構造検索の有用性を示した。しかしながら、固有値は構造全

体の特徴を表したものであり、現段階では、個々の値と特定の局所構造(部分構造)との関係を明らかにできていない。

3. 分子行列の固有ベクトルと構造プロファイリング

本研究ではこの分子行列から行列演算により固有ベクトルを計算し、その固有ベクトルを分子グラフの構造記述子として用い、分子グラフを構成する各頂点原子の分子内環境の数値的な記述と構造特徴解析に対する有用性について検討を行った。分子グラフの行列表現とその分子行列から導出される固有ベクトルの例を図1に示す。

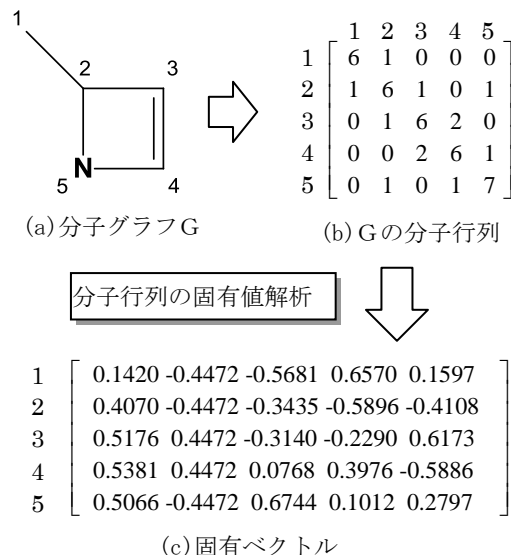


図1 分子グラフの行列表現と固有ベクトル

図1(c)の左端の数字は(a)の分子グラフ G の頂点原子の番号を表す。固有ベクトルは列ベクトルとして、対応する固有値の大きさに従い降順に示されている。ここで、図1(a)より、各頂点原子は1(C), 2(C), 3(C), 4(C), 5(N)であり、頂点5以外はすべてC原子であるが、これら1~4の4つの頂点原子Cはすべて分子内環境が異なり、非等価であることが分かる。一方、上記の固有ベクトルに注目すると、第1固有値に対応する固有ベクトルの各要素の値はすべて異なっていることが分かる。

次に、トポロジカルな対称性が存在する場合の例を図 2 に示す。グラフ自己同型となる頂点原子は、トポロジカルな意味において、すべての分子内環境が等価である。従って、これらの分子行列から導出される固有ベクトルの要素の値についても同様なことが言える。図 2 の例に示されるように、等価な頂点に対応する固有ベクトルの要素の値はすべて等しくなることが分かる。このように、固有ベクトルの各要素は、分子グラフの頂点と一対一に対応していることがわかる。このことから、固有ベクトルを用いることで分子グラフの特定の頂点原子に関する周辺環境の相違を数値情報として詳細に記述可能であることが示唆される。

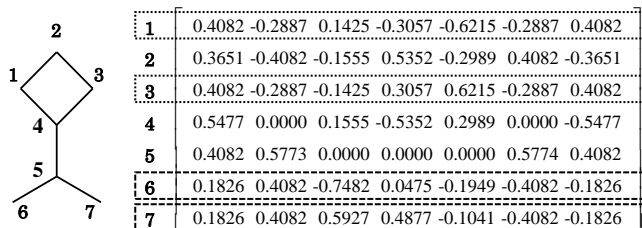


図 3 トポロジカルな対称性を有する分子グラフと固有ベクトル (固有ベクトルは対応する固有値の大きさに従い、左から降順に列ベクトルとして示されている)

4. 分子グラフの構造特徴解析への応用

ここでは鎖状飽和炭化水素類の構造異性体(ヘプタン異性体全 9 種, オクタン異性体全 18 種, ノナン異性体全 34 種)を例に、それぞれの分子行列から、各頂点に対応する第一固有値に対する固有ベクトルを求め、得られた固有ベクトルの要素を頂点原子の周辺環境の相違に基づいて分類し、各頂点原子の分子内環境と固有ベクトルの要素の値との関係を調べた。以下に、ヘプタン(C7H16)の構造異性体の分子グラフについて分子内環境に注目した構造特徴解析の例について示す。解析に用いたヘプタン異性体 9 種の分子グラフを図 3 に示す。

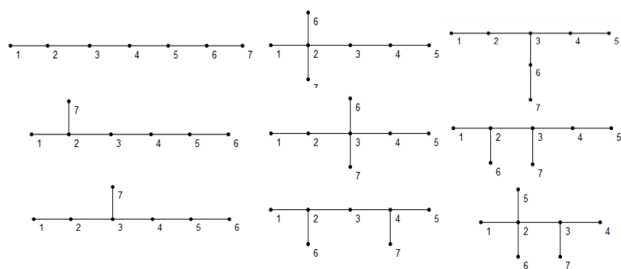


図 3 ヘプタンの構造異性体の分子グラフ

はじめに、ヘプタン異性体分子グラフの各頂点を分子内環境毎に分類し、各頂点に対応する第 1 固有ベクトルの各要素の値の度数分布を調べた。これらの結果を、結合水素原子を含む

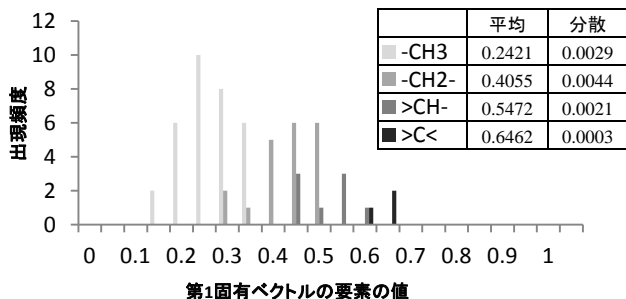


図 4 分子行列の固有ベクトルを用いたヘプタン異性体の構造特徴解析

各頂点原子の度数分布の統計量(平均値と分散)を合わせて図 4 に示す。ここでは各頂点原子(炭素原子)の分子内環境を結合水素数の異なる 4 種類(-CH3, -CH2-, >CH-, >C<)に分類して解析を試みた。その結果、対応する頂点原子ごとに固有ベクトルの要素の値(平均値)が明らかに異なっていることが分る。また、その分散も非常に小さな値を示している。

次に、上記の結果をもとに、最も度数の大きな(頻出の)-CH3 グループについて、隣接する頂点(非水素原子)までを考慮した分子内環境をもとに各頂点原子(炭素原子)を分類し、各グループごとに同様な度数分布と統計量を調べた。これらの結果を図 5 に示す。

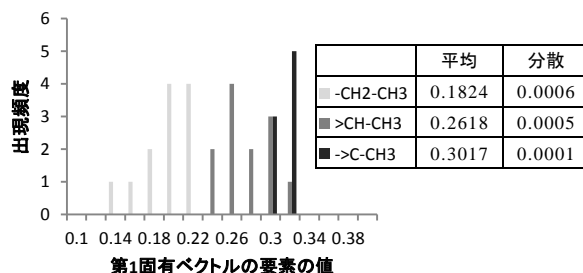


図 5 固有ベクトルを用いたヘプタン異性体の構造特徴解析 (図 4 の-CH3 グループの隣接環境と固有ベクトル要素の値)

この場合、前述のメチル基(-CH3)の炭素はその隣接環境により CH2-CH3, CH-CH3, C-CH3 の 3 つのサブグループに分けられる。各サブグループごとの分子内環境を有するメチル基炭素(頂点原子)に注目した第 1 固有ベクトル要素の平均値も明らかに異なり、分散も小さな値を示していることが分る。このことは、オクタンの 19 種の異性体、ノナンの 34 種の異性体を用いた場合にも同様な結果が観測された。以上のことから、分子行列の固有ベクトルは各頂点周りの環境を反映していることはあきらからかである。

5. まとめ

上記の結果は、分子行列の固有ベクトルを用いた分子内環境の数値的な記述表現の有用性と化学構造データマイニングへの応用の可能性を示唆するものである。引き続き、多重結合やヘテロ原子、さらには環を含むより複雑な構造に対する分子内環境の数値的な記述の可能性について引き続き検討を進めていきたい。

参考文献

[Takahashi 98] Y. Takahashi, H. Ohoka, Y. Ishiyama, Structural Similarity Analysis Based on Topological Fragment Spectra, *Advances in Molecular Similarity*, **2**, 93-104(1998)

[Takahashi 05] Y. Takahashi, K. Nishikoori, S. Fujishima, Classification of Pharmacological Activity of Drugs Using Support Vector Machine. In *Active Mining, Second International Workshop, Lecture Notes in Computer Science*; Tsumoto, S., Yamaguchi, T., Numao, M., Motoda, H., Eds.; Springer: Berlin, Germany, 2005; Vol. 3430, pp 303-311.

[Kawai 08] K. Kawai, S. Fujishima, and Y. Takahashi, Predictive Activity Profiling of Drugs by Topological-Fragment-Spectra-Based Support Vector Machines, *J. Chem. Inf. Model.*, **48**, 1152-1160 (2008).

[宮脇 12] 宮脇康司, 高橋由雅, 第 30 回構造活性相関シンポジウム講演要旨集, pp.103-104 (2011).