

単純グラフ同型なフレームワークを有する薬物分子の NTG ネットワーク

NTG network of drug molecules which involve the same NTG scaffold

野村 京平
Kyohei Nomura

大田 黒 空
Sora Ohtaguro

高橋 由雅
Yoshimasa Takahashi

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

In the preceding works, we proposed a basic skeletal feature representation of molecules with a graph called Non-Terminal vertex Graph (NTG) that has no vertices of degree of 1. We also reported that non-terminal vertex graph (NTG) is a useful feature representation of drug molecules in structure-activity analysis, and molecular framework represented by the NTGs is closely related to a particular drug action. In this paper, we propose a new approach to knowledge discovery for the NTGs in association with drug action. The method is based on making a network of NTG frameworks with their drug activity information. The NTG network can be used to get some association rules for NTG-based molecular features and drug activity. The paper describes the detail of the algorithm for making a NTG network and the utility of the network for knowledge acquisition.

1. はじめに

本研究の構造特徴解析に関する研究成果として、化学グラフの環構造に注目した部分グラフ、NTG(Non-Terminal vertex Graph)の類似する化合物はその活性も類似することが判明しており[大野 02],その特徴を活用したシステム開発が進められている。

本研究では、各々の NTG が単純グラフ表現の下で全て同型となる(同一のフレームワークを有する)薬物群を対象とした化学構造データマイニングのための新たな方法を提案する。ここでは、先に当研究室の大田黒らによって提案された、単純グラフ同型な分子グラフ間の構造的な近接性や相違を明示に表すための指標である構造差異[大田黒, 09]を利用した NTG ネットワークの生成と薬物分子グラフマイニングへの応用について議論する。

2. 方法

2.1 本研究で対象とする分子グラフと NTG

NTG とは、次数 1 の頂点および孤立頂点を持たない連結グラフをいう。また、各頂点や辺に対する重みづけの相違により、単純グラフ表現(NTG/SG)、頂点重み付きグラフ(NTG/VG)、辺重み付きグラフ(NTG/EG)、および化学グラフ表現(NTG/CG)などの異なる表現レベルが定義されている。本研究では、単純グラフ表現が同型となる NTG (NTG/CG) 群のみを解析の対象としている。単純グラフ表現において同一の骨格を有する分子グラフとその NTG の例を図1に示す。

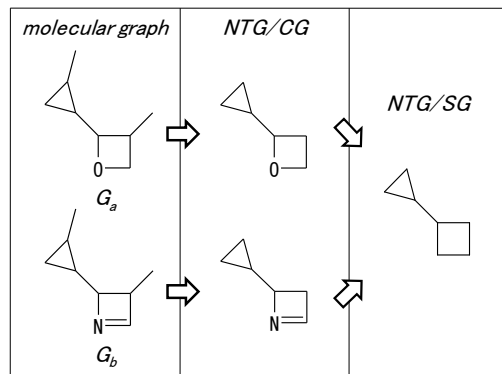


図1 同一単純グラフ骨格を有する分子グラフの例

2.2 NTG間の構造差異

大田黒らは、同一の単純グラフ骨格を有する任意の 2 つの分子グラフの構造差異を数値的に評価することを目的として、頂点差異、辺差異、化学構造差異の 3 つの異なる指標を提案している[大田黒 09]。本研究では、単純グラフ表現が同型となる任意の二つの NTG 間の構造差異の評価に対してこれらの指標を用いた。

頂点差異 Dif^V :

頂点ラベルの相異に注目した構造差異を表す。二つの NTG, G_a, G_b において、頂点差異 $Dif^V(G_a, G_b)$ は以下のように定義される。

$$Dif^V(G_a, G_b) = \sum_{i=1}^n S(G_a, G_b, i, j) \quad (1)$$

$$S(G_a, G_b, i, j) = \begin{cases} 1 & \text{if } C(G_a)_{ij} \neq C(G_b)_{ij} \\ 0 & \text{if } C(G_a)_{ij} = C(G_b)_{ij} \end{cases} \quad (2)$$

ここで、 $C(G_a), C(G_b)$ は G_a, G_b の各々の結合行列、 n はグラフの頂点数、 $S(G_a, G_b, i, j)$ は結合行列の i 行 j 列の要素が等しいならば 0、そうでないならば 1 の値をとる関数を表す。頂点差異は、式(1)(2)に従い、結合行列の対角要素のうち、値の異なる要素の数を表す。

辺差異 Dif^E :

辺の重みの相異に注目した構造差異を表す。二つの NTG, G_a, G_b における辺差異 $Dif^E(G_a, G_b)$ は以下のように定義される。

$$Dif^E(G_a, G_b) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n S(G_a, G_b, i, j) \quad (3)$$

ここで、 $S(G_a, G_b, i, j)$ は(2)式で定義される関数である。辺差異は、式(2)(3)に従い、結合行列の対角要素を含まない上三角要素のうち、値の異なる要素の数を表す。

化学構造差異 Dif^C :

頂点ラベルおよび辺の重みの両方の相異に注目した構造差異を表す。二つの NTG, G_a, G_b の間の化学構造差異 $Dif^C(G_a, G_b)$ は以下のように定義される。

$$Dif^C(G_a, G_b) = \sum_{i=1}^n \sum_{j=i}^n S(G_a, G_b, i, j) \quad (4)$$

同じく、 $S(G_a, G_b, i, j)$ は(2)式の関数である。構造差異は、式(2)(4)に従い、結合行列の対角要素を含む上三角要素のうち、値が異なる要素の数を表す。

2.3 NTG ネットワーク

本研究で提案する NTG ネットワークとは以下の定義にもとづいて生成されるものをいう。(図 2)

- 基本フレームワークとなる NTG/SG を G_0 としたとき、NTG データベースから抽出された各 NTG/CG を G_1, G_2, \dots, G_n とする。
- G_1, \dots, G_n を頂点とするネットワークを考えたとき、2 つの頂点を結ぶリンクはその両端の頂点、すなわち双方の NTG の構造差異に伴う活性変化を表すものとする。

なお、リンクは全ての頂点間で生成可能であるが、本研究では事前に指定した構造差異の任意の値を生成のための閾値として用いることによりフィルタリングするものとする。図 2 に、単純グラフ同型フレームワークを有する NTG ネットワークの例を示す。

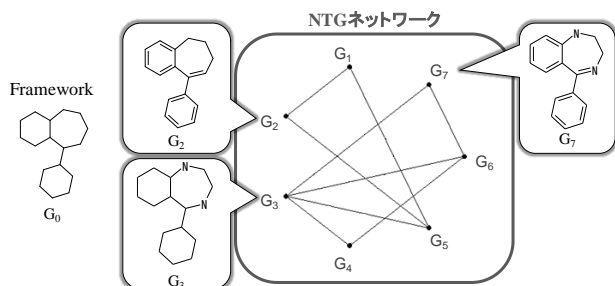


図 2 単純グラフ同型な分子グラフの NTG ネットワーク

3. NTG ネットワークの生成処理の流れ

本研究における NTG ネットワーク生成処理の具体的なアルゴリズムを以下に説明する。

NTG ネットワークの生成と描画処理の流れ:

step.1 同型なフレームワークを有する NTG/CG の抽出

ユーザがフレームワークとして使用する NTG/SG の ID 番号を入力として与えた場合、同型なフレームワークを有する全ての NTG/CG の結合表データをデータベースより抽出する。

step.2 結合表データから、構造差異の計算

step.1 で得られた結合表データを入力として与えた場合、全ての NTG 間の構造差異を計算し、リストしたものを出力する。構造差異のリストは、[NTG の ID 番号, NTG の ID 番号, 化学構造差異, 頂点差異, 辺差異] で構成されるものとする。

step.3 結合表から、各 NTG の ID 番号、活性の切り出し

step.1 で得られた結合表データを入力として与えた場合、データから各 NTG/CG の ID 番号と活性を切り出す。

step.4 活性変化を調べる

step.3 で生成された各 NTG/CG の ID 番号と活性のパラメータを使用して活性変化、すなわち構造変化に伴って発現した活性と消失した活性を調べる。

step.5 構造差異、活性変化等のパラメータを NTG ネットワーク表示用のプログラムに渡す

step.2, step.3, step.4 で各々生成された構造差異や NTG, 活性のパラメータを NTG ネットワークの頂点、エッジのパラメータを引数として保存・引き渡しをする。頂点、エッジのパラメータは NTG ネットワークを描画するために用いられる。

step.6 NTG ネットワークの描画

step.5 で生成された頂点、エッジのパラメータを使用して、NTG ネットワークを描画する。NTG ネットワークを描画するためには、頂点の座標を設定する必要がある。頂点の配置は、予め指定した大きさの円周上に等間隔に配置する。円の中心座標は (X_0, Y_0) 、半径は r とする(図 3)。

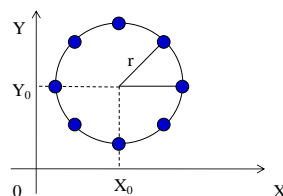


図 3 NTG ネットワーク描画空間の頂点座標

この時、各頂点の座標は以下のような処理により得られる。

```
for (int i=0; i<n; i++) {
    double x,y;
    x=X0+r*cos(2π/n*i);
    y=Y0+r*sin(2π/n*i);
    vertex.get(i).setX(x);
    vertex.get(i).setY(y);
}
```

ここで、 n は NTG ネットワークの頂点数、`vertex.get(i).setX(x)` は、頂点 i の X 座標を x に設定するメソッド、`vertex.get(i).setY(y)` は、頂点 i の Y 座標を y に設定するメソッドを表す。

4. 計算機実験の結果

実験に際しては、能登らが作成した NTG/薬理活性関係データベース[能登, 08]を用いた。ここでは、図 4 左に示すフレームワークと単純グラフ同型となる NTG (NTG/CG) 群に対するネットワークの生成を試みた。抽出された NTG/CG の件数は 12 件であった。これら 12 件の NTG に対して生成されたネットワークの出力結果を図 4 右に示す。描画されたネットワークでは各頂点 (NTG) が同円周上に等間隔に配置されている。ここでは、NTG ネットワークのリンクは、頂点差異が 1 となるもののみ表示するようフィルタリングした。この結果からは、他の頂点とリンクしない頂点が 3 件ある一方で、多いものでは 4 つのリンクをもつ頂点もあり、各 NTG 構造間の近接関係やその構造差異が明確に表現されていることがわかる。

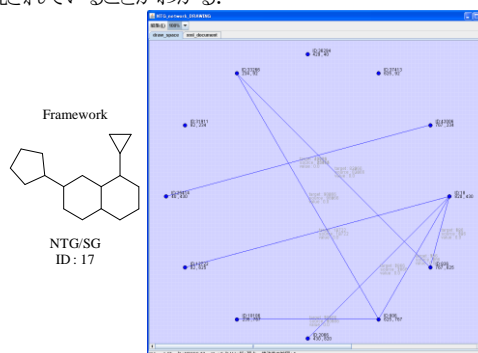


図 4 実データにおける NTG ネットワーク

5. まとめ

本研究で示した NTG ネットワークを活用し、各々のリンクの両端の頂点 NTG における構造差異と活性変化を調べることにより、薬物構造活性相関の新たな知識の獲得が期待できる。引き続き、より大規模な薬物分子の構造データに対して NTG ネットワーク作成と、そのデータベース化を進めていきたい。

参考文献

- [大野 02] 大野貴生, 高橋由雅, 第 30 回構造活性相関シンポジウム講演要旨集, 41-42, 2002.
- [大田黒 09] 大田黒空, 高橋由雅, 第 9 回データマイニングと統計数理研究会, 121-124, (2009.3).
- [能登 08] 能登上総, 高橋由雅, 2008 年度人工知能学会全国大会 (第 22 回) 論文集, 2P2-13 (2008.6).