2008－BIO－12 (9)
2008／3／4

不斉炭素原子を考慮した化合物に対するグラフカーネル法

浦田 隆史[1], J.B. Brown[1], 田村 武幸[1], 川端 猛夫[2], 阿久津 達也[1]

概要 化合物を生物学的な特性に基づいて高精度に分類することは，創薬に適した化合物を探す過程において重要な役割を果たす．従来のグラフカーネル法は，このような化合物の分類に対してもある程度有効ではあるが，不斉炭素原子を含む化合物が異なる特性を示す場合に適切に区別することができなかった．そこで本研究では Tree-Pattern グラフカーネルを拡張することにより，トポロジーが同じでありながら立体配置が異なる化合物どうしを区別できる方法を提案する．我々は Ecdysteroids と Cramer's Steroids の２種類の化合物のデータ集合に対し提案手法を SVM と SVR で実装し，多くの場合で予測精度が改善されることを確認した．

Graph Kernels for Compounds Incorporating Chirality

Takashi Urata[1], J.B. Brown[1], Takeyuki Tamura[1], Takeo Kawabata[2], Tatsuya Akutsu[1]

Abstract In the process of drug discovery, maintaining high accuracy in predicting the biological characteristics of chemical compounds is paramount. Although existing graph kernel methods are efficient for this purpose to some extent, they cannot appropriately classify chiral compounds which have the same topology with different biological characteristics. Therefore in this paper, we propose a new method by extending the Tree-Pattern graph kernel so that chiral compounds which have the same topology can successfully be classified. We implemented SVMs and SVR which include the proposed method in their formulation, and validated an improvement in prediction accuracy using Ecdysteroids and Cramer's Steroids datasets.

1 はじめに

化合物の特性を予測することは重要である．例えば創薬研究におけるリード化合物の探索では，目的に合致した化合物を無数の化合物群の中から選出する．この過程は，コンビナトリアルケミストリーやハイスループットスクリーニングなどの技術を用いて，効率的に化合物を合成して評価をすることにより可能であるが，膨大な時間やコストを要するという問題がある．そこで，特性予測の技術により有望な化合物を絞った後に，実験による評価を行なうことが望ましい [1]．また，約 10 万種類もの化合物が市場で流通しているものの，毒性の有無が判明している化合物はきわめて少ない [2]．理想的には市場で流通する化合物すべてに対して毒性情報を検査するべきであるが，動物を用いる安全性試験で毒性評価を行なうことは，莫大な時間や費用等を要するためほぼ不可能である．よって，特性を予測する技術がやはり重要になる [2]．

化合物の特性予測問題では，特性が既知の化合物の構造情報と特性との間にある定量的な関係を発見し，化合物の構造情報から特性を求めるモデルを構築するのが基本的な方法である．これは定量的構造活性相関 (Quantitative Structure-Activity Relationships, QSAR)，あるいは定量的構造物性相関 (Quantitative Structure-Property Relationships, QSPR) と呼ばれる [3]．なかでも CoMFA(Comparative Molecular Field Analysis)[4], 4D-QSAR[5, 6] のように三次元立体構造を利用する方法は，予測精度が高いものの，計算時間が大きくなるという欠点がある．一方でグラフカーネルを用いる方法 [7, 8, 9] やトポロジカル記述子を用いる方法 [1, 10] などの二次元構造式を利用する方法は，計算時間も短く，データも扱いやすいという利点がある．しかしトポロジーが同じでありながら，立体配置が異なる化合物を区別できないという問題がある．

化合物を一意に特定するためには，化合物内の各原子の結合関係 (トポロジー) を表した二次元構造だけでなく，立体配置や立体配座まで指定する必要がある．立体配置とは，三次元空間における立体中心，立体軸，立体面に対して，原子が異なる位置を占める場合の配置を表す．例えば，異なる 4 種類の基が結合する原子，不斉原子が存在する化合物 (キラリティ化合物 [3]) は，2 種類の立体配置が存在する．また，二

[1]Bioinformatics Center, Institute for Chemical Research, Kyoto University.
E-mail:{urata, jbbrown, tamura, takutsu}@kuicr.kyoto-u.ac.jp
[2]Division of Synthetic Chemistry, Institute for Chemical Research, Kyoto University
E-mail: kawabata@scl.kyoto-u.ac.jp

重結合でつながった炭素原子両方において，結合する2つの基が異なる場合，2種類の立体配置が存在する (*cis/trans* 立体配置)．一方，立体配座とは，単結合周りの回転によって変換可能な，実際の三次元立体構造の配置を表す．トポロジーが同じで立体配置，または立体配座の異なる化合物どうしを立体異性体と呼ぶ．特に立体配置の異なる立体異性体を立体配置異性体と呼び，立体配座の異なる立体異性体を配座異性体と呼ぶ．また，立体配置異性体の中で，不斉原子により立体配置の異なる化合物どうしは，お互いに鏡に映した関係 (鏡像関係) にあることより，鏡像異性体 (エナンチオマー，Enantiomers) と呼ばれる．一方，不斉原子を2個以上持つ立体配置の異なる化合物どうしは，鏡像関係にない場合があり，これらをジアステレオマー (Diastereomer) と呼ぶ．

このような立体配置異性により生物学的特性（生物活性）に多様性が生じる [11, 12]．例えばジベレリン生合成の前駆体であるカウレノールは，2種の鏡像異性体のうち片方のみが植物成長を促進する効果がある．また昆虫幼若ホルモン類 **JH** I と **JH** III の両鏡像体については，実験の結果により，(+)- **JH** I が同 (−)-体の 12000 倍，(+)- **JH** III が同 (−)-体の 5000 倍の生物活性を発揮することが示されている．一方で両方の鏡像体が同様の活性を発揮する場合もある．ネコを誘引するネコハッカの成分として知られる (+)-ネペタラクトンは，ネコに対しては非天然型 (−)-体でも誘因性を示す．また，刺身のつまに用いるヤナギタデの辛味成分として単離された (−)-ポリゴジアールは強力な昆虫摂食阻害物質であるが，非天然型 (+)-体も強力な昆虫阻害活性を示す．他にも，植物が病原微生物に攻撃された際に作り出す抗菌性物質である，イネにおける (+)-オリザレキシン A およびマメにおける (+)-ピサチンについては，それら天然型に加え非天然型 (−)-体にも十分強い活性が見られる．

このように立体配置異性は化合物の生物学的な特性に重大な影響を与えるにも関わらず，従来のグラフカーネル法では区別することができなかった．そこで本稿において我々は，二次元構造式だけでなく立体配置の情報も取り扱える手法を提案する．この提案手法は従来のグラフカーネル法の拡張であるため，計算時間がトポロジー情報だけを利用する場合とほぼ同じであるという利点がある．また二次構造式で記述されたデータは3次元構造を利用する場合よりも，取り扱いやすいという利点もある．我々は Ecdysteroids と Cramer's Steroid の2種類の化合物のデータ集合に対し提案手法を SVM と SVR で実装し，多くの場合で予測精度が改善されることを確認した．

2 Tree-Pattern グラフカーネルの拡張

Tree-pattern の定義 [9] あるグラフ $G = (V_G, E_G, l_G)$ と木 $t = (V_t, E_t, l_t)$ が与えられたとする．ただし，l_G はそれぞれ V_G と E_G から原子と共有結合への写像とする．同様に，l_t はそれぞれ V_t と E_t から原子と共有結合への写像とする．木 t の頂点集合 V_t 内の頂点に任意の順番 $(n_1, n_2, ..., n_{|t|})$ を与えたとき，グラフ G の頂点から選択した $|t|$ 個の頂点の列 $(v_1, v_2, ..., v_{|t|})$ （重複して選択してもよい）が以下の条件を満たすならば，その頂点列 $(v_1, v_2, ..., v_{|t|})$ を，グラフ G における木 t (頂点列 $(n_1, n_2, ..., n_{|t|})$) に関する Tree-pattern と定義する．

$$\begin{cases} \forall i \in [1, |t|], & l_G(v_i) = l_t(n_i) \\ \forall (n_i, n_j) \in E_t, & (v_i, v_j) \in E_G \wedge l_G((v_i, v_j)) = l_t((n_i, n_j)) \\ \forall (n_i, n_j), (n_i, n_k) \in E_t, & j \neq k \Longleftrightarrow v_j \neq v_k \end{cases} \tag{1}$$

すなわち，Tree-pattern とは，グラフ G の中にラベルの一致した木 t を含むとき，それを表す頂点列のパターンであると言える．

このとき，Tree-pattern グラフカーネルを，以下のように定義する．ここで，T とは，木を各点とする空間（木空間）であり，実際に計算する場合は木空間 T を定義する．また $w(t)$ は，木 t に対する重みを表し，$\psi_t(G)$ は，グラフ G における木 t に関する Tree-pattern の出現回数を表す．

$$K(G_1, G_2) = \sum_{t \in T} w(t) \psi_t(G_1) \psi_t(G_2) \tag{2}$$

Tree-pattern グラフカーネルの計算は，グラフ G に対して木空間 $T = \{t_1, t_2, t_3, ...\}$ に関する特徴ベクトル $(\sqrt{w(t_1)}\psi_{t_1}(G), \sqrt{w(t_2)}\psi_{t_2}(G), \sqrt{w(t_3)}\psi_{t_3}(G), ...)$ への変換を行ない，各グラフ間で得られた特徴ベクトルの内積を計算していると解釈できる．例えば図1左上側のホルムアルデヒド CH_2O に対応するグラフに関して，図中左下側の木 t_1, t_2, t_3 に関する Tree-Pattern の出現回数をカウントする．Tree-pattern $t_1 = H \rightarrow C \rightarrow H$ は始点の H の選び方が2通り，終点の H の選び方も2通りあるので，4つの t_1 が出現することになる．このように $v_{i-1} \rightarrow v_i \rightarrow v_{i+1}$ と移動する際に，$v_{i-1} = v_{i+1}$ となるような移動を tottering と呼ぶ．本稿では tottering を許す場合と tottering を許さない場合の両方について解析を行っていく．

また文献 [9] では，Tree-pattern グラフカーネルの例として，Size-based tree-pattern グラフカーネル，Branching-based tree-pattern グラフカーネル，Until-N Branching-based tree-pattern グラフカーネルの

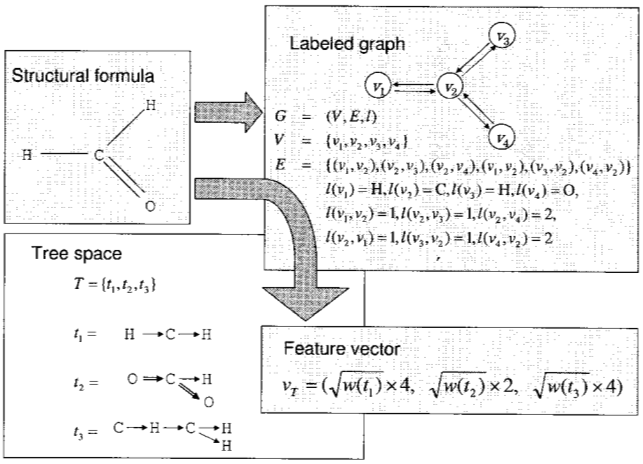


図 1: 分子化合物のデータ変換の例

3つが提案されている．これらのカーネルの実際の計算では，既に述べたグラフから特徴ベクトルへの変換は行なわず，動的計画法で各グラフ間のカーネルを直接計算する方法が提案されている．ここでは省略する．

立体異性体を考慮した化合物のためのグラフカーネル：

現在提案されているグラフカーネルの一般的な設計法は，構造データどうしのカーネルを，構造データの部分構造どうしのカーネルによって再帰的に定義するという畳み込みカーネルのアイディアによる．つまりグラフカーネルは，グラフの部分構造どうしのカーネルによって再帰的に定義される．ここで，立体異性を区別するグラフカーネルを設計するための簡素でかつ自然な拡張として，グラフに立体配置の情報を導入し，グラフの部分構造についてもその情報を記憶した上で，立体配置の情報を含めたグラフの部分構造どうしのカーネルを新たに定義するという方法が考えられる．本研究では，グラフの部分構造に木を用いる Tree-pattern グラフカーネルにおいて，立体配置の情報を含めた木どうしのカーネルを定義し，立体異性を区別するグラフカーネルを実現した．Tree-pattern グラフカーネル [9] において，グラフの部分構造である木どうしのカーネルは，その木どうしがラベルを含め一致する場合のみ正の値を持ち，それ以外は 0 となるものであった．一方本稿では，立体配置の情報を含めた木どうしのカーネルは，その木どうしがラベル，立体配置を含め一致する場合のみ正の値を持ち，それ以外は 0 となるものと定義する．

本研究では，不斉炭素原子により生じる立体異性 (鏡像異性，ジアステレオマー)，炭素原子どうしの二重結合により生じる立体異性 (*cis/trans* 異性) を，区別できるグラフカーネルを提案する．例えば図 2 の c を中心に a, b, c, d がそれぞれ左回りと右回りに配置された立体異性な 2 つの化合物 (それぞれ c_1, c_2 と呼ぶ) は，図中左側のキラルを考慮しない Tree-pattern の t_1 と t_2 の両方を 1 つずつ部分構造として含む．よって t_1 と t_2 の出現回数を特徴ベクトルの要素として用いても SVM で c_1 と c_2 を区別することはできない．しかし，図中右側のように t_{c1} と t_{c2}，t_{c3} と t_{c4} の出現回数をそれぞれ別の要素として特徴ベクトルを算出すると，c_1 と c_2 における $(t_{c1}, t_{c2}, t_{c3}, t_{c4})$ の出現回数はそれぞれ (1,0,1,0) と (0,1,0,1) となり，c_1 と c_2 を区別することが可能となる．

同様に，二重結合によりつながった炭素原子ペアに対し a, b, c, d がそれぞれ図 3 のように配置された立体異性な 2 つの化合物 (それぞれ c_3，c_4 と呼ぶ) は，図中左側の *cis/trans* 異性を考慮しない Tree-pattern の t_3 と t_4 の両方を 1 つずつ部分構造として含むので，t_3 と t_4 の出現回数を特徴ベクトルとして用いても SVM で c_3 と c_4 を区別することはできない．しかし，図中右側のように t_{d1} と t_{d2}，t_{d3} と t_{d4} の出現回数をそれぞれ別の要素として特徴ベクトルを算出すると，c_3 と c_4 における $(t_{d1}, t_{d2}, t_{d3}, t_{d4})$ の出現回数はそれぞれ (1,0,1,0) と (0,1,0,1) となり，c_3 と c_4 を区別することが可能となる．このように，部分構造として木を用いることで，グラフの立体配置を木にそのまま記憶させることができ，立体配置情報を含めた木どうしのカーネルを効率よく定義することができる．

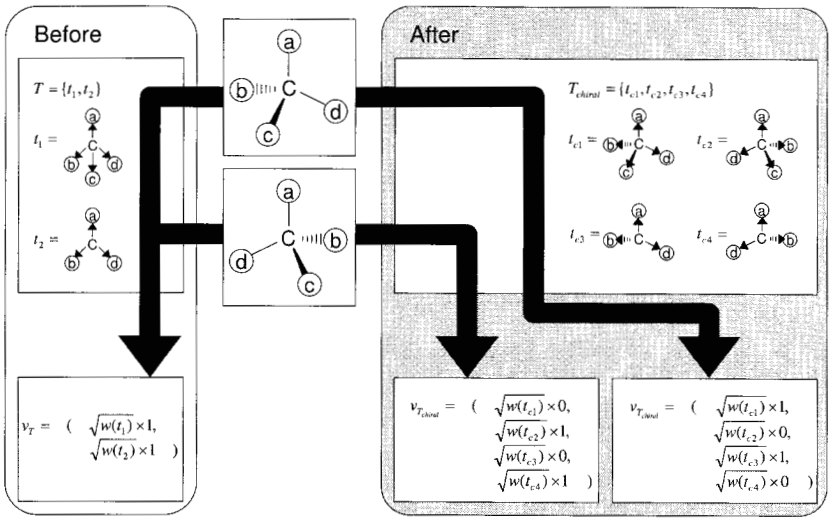


図 2: 不斉炭素原子の立体配置の情報の導入前と導入後の違い

3 コンピュータ実験

3.1 データセット

データセット 1 - Ecdysteroids は，昆虫の脱皮・変態を制御する 20-hydroxyecdysone(20E) というステロイドホルモンの関連化合物 (Ecdysteroid)104 件である．これらの脱皮ホルモンに対しては，脱皮ホルモン受容体とどれだけ結合するか (結合親和力) を示す bioassay により，脱皮に関する活性の強度を測定する方法が提案されており [13]，今回用いる 104 件の脱皮ホルモンはこの方法により活性強度が求められている．これら脱皮ホルモンを用いた活性予測のための研究は複数行なわれており [10, 14, 15, 16]，文献 [16] で用いられた訓練セット 71 件とテストセット 33 件の計 104 件の脱皮ホルモンを，今回の実験に用いる．
データセット 2 - Cramer's Steroids は，文献 [4] で紹介されたデータセットであり，活性予測研究のためのベンチマークとなっているステロイド群 31 件である．これらのステロイドに対しては，corticosteroid binding globulin(CBG) との結合親和力が利用可能であり，これらの特性を予測する研究が複数行なわれてきた [4, 17, 10]．本研究では文献 [4] で用いられたとおり，データセット 31 件を訓練セット 21 件とテストセット 10 件に分割する．

3.2 SVM による評価実験

まず，立体異性を考慮に入れたグラフカーネルとサポートベクターマシン (SVM) を用いて，2 種類のデータセットに対して，特性の予測を 2 値分類の形で行なう．SVM の利用には，Gist[18] と呼ばれるソフトウェアツールを用いる．データセットの特性値は実数をとるため，ここでは，閾値を設定することで特性値を 2 値化する．今回の実験では，訓練セット，テストセットそれぞれに対してある程度均等に分割する閾値として，6，6.5，7 の 3 種類を用いた．各閾値におけるデータセットの分布を表 1 に示す．

今回提案したグラフカーネル法は，Tree-pattern グラフカーネルを拡張したものである．Tree-pattern グラフカーネルは深さ h や重み定数 λ などのパラメータを持ち，一つのパラメータ設定ごとに一つのカーネル関数を生成するので，今回の実験では，適当な範囲で各パラメータを変化させ，複数のパラメータ設定を定義し，各パラメータ設定におけるカーネル関数を用いて特性予測を行い，性能を評価する．設定が必要である各パラメータを以下のように設定し，各パラメータ設定による 60 種類のカーネル関数について，実験を行なう: カーネルタイプ = {Size-based, Branching-based, Until-N Branching-based}, 深さ $h = \{3, 4, 5, 6\}$, 重み定数 λ={0.1, 0.2, 0.3, 0.4, 0.5}.

今回の実験では，2 つの指標により評価を行う．まず訓練セットに対して，Leave-One-Out 交差検定 (LOO-CV) を行い，その予測値より ROC 曲線の Area Under the Curve (AUC) を算出し，性能を評価する．次に訓練セットを用いて予測モデルを構築し，その予測モデルを用いてテストセットの予測を行なう．同様

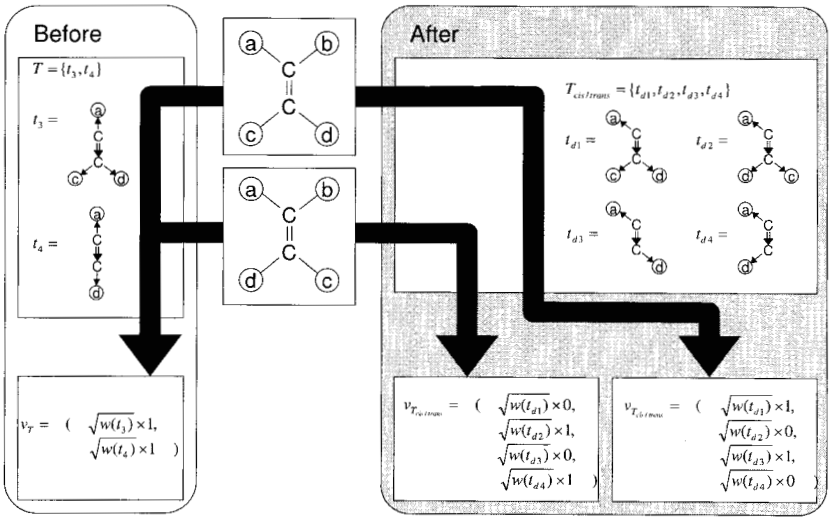


図 3: 炭素原子の二重結合により生じる立体異性の情報の導入前と導入後の違い

表 1: 各閾値における訓練セット，テストセットの分布

データセット 1 - Ecdysteroids

	訓練セット		テストセット	
閾値	+1	−1	+1	−1
6	42	29	15	18
6.5	33	38	10	23
7	23	48	6	27

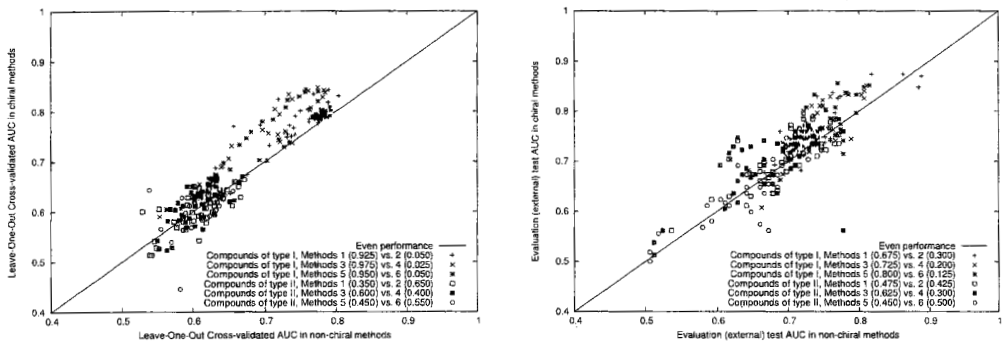
データセット 2 - Cramer's Steroids

	訓練セット		テストセット	
閾値	+1	−1	+1	−1
6	9	12	9	1
6.5	8	13	7	3
7	6	15	5	5

に，その予測値より ROC 曲線の AUC を算出し，性能を評価する．

キラリティを導入していない 3 種類のカーネル Size-based, Branching-based, Until-N Branching-based(それぞれ method 2, 4, 6) と，それらに対してキラリティを導入したカーネル (それぞれ method 1, 3, 5) の ROC 曲線の AUC における比較を図 4 に示す．図 4(a) は LOO-CV，図 4(b) は評価用テストデータセットに対する ROC 曲線の AUC の値を示している．各点は，カーネルパラメータの値を決定して生成された各モデルを表す．図中の凡例における括弧の中の値は，その方法のほうが比較対象の方法より AUC の値が大きかったモデルの数の割合である．お互いに同じ性能を持つことを示す直線 (Even performance) に対して，左上に多くの点がプロットされており，キラリティ導入による性能の向上が確認された．

また様々な設定のもとでの LOO-CV による AUC が最高になるモデルの予測精度を表 2 に示す．LOO-CV で最も予測精度が高いモデルを，テストセットに対する予測精度で評価する．例えばデータセット 1 に対し，既存の手法における LOO-CV での最高の予測精度は，閾値が 6.0, 6.5, 7.0 の場合にそれぞれ，0.804, 0.809, 0.741 であり，それらのモデルを採用した場合のテストセットに対する予測精度は 0.733, 0.691, 0.698 である．同様に我々の提案手法では，LOO-CV の 0.847, 0.842, 0.675 に対し，テストセットにおける結果はそれぞれ 0.796, 0.770, 0.741 であり，3 つの場合すべてで我々の提案手法が既存手法よりも高い予測精度を示した．



(a) Performance of graph kernels, LOO-CV (b) Performance of graph kernels on evaluation (test) data

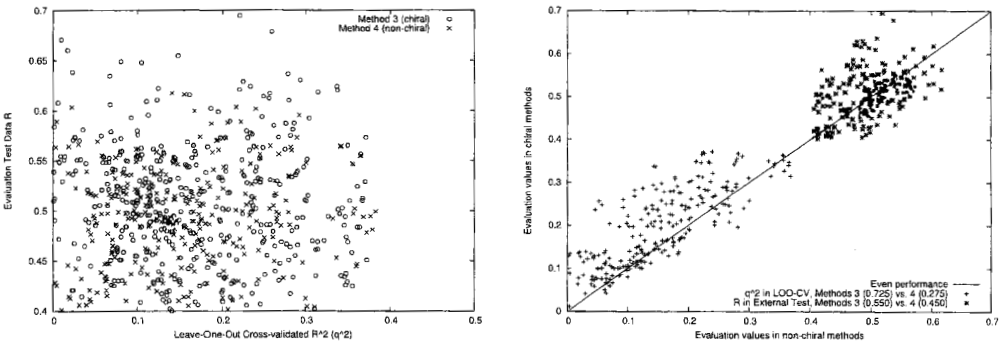
図 4: SVM による評価実験. Compounds of type I は閾値が 6.0 のデータセット 1 , Compounds of type II は閾値が 7.0 のデータセット 1 を表す.

表 2: SVM 分類実験において LOO-CV における AUC が最高となるモデルの結果.

閾値 : 6	拡張			AUC(データセット 1)		AUC(データセット 2)	
部分構造	no tottering	chiral	*cis/trans*	LOO-CV	Test	LOO-CV	Test
パス	×	×	×	0.778	0.819	**0.991**	0.778
パス	○	×	×	0.771	0.815	**0.991**	0.778
木	×	×	×	0.787	0.759	0.954	**1.000**
木	○	×	×	0.804	0.733	0.981	0.778
木	×	○	×	0.806	**0.833**	0.954	0.889
木	○	○	×	**0.846**	0.785	0.981	0.667
木	○	×	○	0.807	0.730	0.981	0.778
木	○	○	○	0.847	0.796	0.981	0.667

閾値 : 6.5	拡張			AUC(データセット 1)		AUC(データセット 2)	
部分構造	no tottering	chiral	*cis/trans*	LOO-CV	Test	LOO-CV	Test
パス	×	×	×	0.801	0.730	0.981	0.619
パス	○	×	×	0.801	0.730	**0.990**	0.667
木	×	×	×	0.766	0.709	0.942	**0.952**
木	○	×	×	0.809	0.691	**0.990**	0.857
木	×	○	×	0.801	0.709	0.942	0.905
木	○	○	×	0.841	0.761	**0.990**	0.810
木	○	×	○	0.809	0.704	**0.990**	0.857
木	○	○	○	**0.842**	**0.770**	**0.990**	0.810

閾値 : 7	拡張			AUC(データセット 1)		AUC(データセット 2)	
部分構造	no tottering	chiral	*cis/trans*	LOO-CV	Test	LOO-CV	Test
パス	×	×	×	**0.741**	0.698	**1.000**	**0.760**
パス	○	×	×	0.739	0.691	0.989	0.720
木	×	×	×	0.658	**0.759**	0.978	0.720
木	○	×	×	0.671	0.722	**1.000**	0.680
木	×	○	×	0.671	0.679	0.967	0.720
木	○	○	×	0.670	0.710	**1.000**	0.640
木	○	×	○	0.674	0.728	**1.000**	0.680
木	○	○	○	0.675	0.741	**1.000**	0.640



(a) Performance of Branching-based Graph Kernel (1) (b) Performance of Branching-based Graph Kernel (2)

図 5: SVR による評価実験.

表 3: SVR 実験において q^2 が最高となるモデルの結果.

部分構造	拡張			データセット 1		データセット 2	
	no tottering	chiral	*cis/trans*	q^2	R	q^2	R
パス	×	×	×	0.364	0.473	0.758	0.485
パス	○	×	×	0.363	0.474	0.725	0.495
木	×	×	×	0.368	0.551	0.792	0.729
木	○	×	×	0.383	0.499	0.862	0.517
木	×	○	×	0.374	0.432	0.767	**0.740**
木	○	○	×	0.429	0.518	**0.865**	0.513
木	○	×	○	0.387	0.507	0.860	0.516
木	○	○	○	**0.433**	**0.553**	0.863	0.512

3.3 SVR による評価実験

次に，Support Vector Regression (SVR) による回帰予測の実験により，キラリティを考慮に入れたグラフカーネルを評価する．SVR の利用には， SVM^{light} [19] と呼ばれるソフトウェアツールを用いる．本実験では，グラフカーネルの持つパラメータ設定に加えて，SVR におけるパラメータ設定も行なうことで，6 つの各カーネルに対してそれぞれ 3120 個のモデルを作成し，それぞれの性能を比較した．

まず LOO-CV を行い，その予測値より，予測残差平方和 (Prediction Residual Error Sum of Squares; PRESS) にもとづいた評価指標 $1 - \frac{PRESS}{Variance\ of\ Actual\ Values}$ ($Cross\text{-}Varidated\ R^2$ (q^2)) を算出し，比較を行なう．次に評価用テストデータに対して，訓練データをもとに作成した予測モデルを用いて予測を行い，評価指標として実測値と予測値の相関係数 R を算出し，比較を行なう．

図 5(a) は Branching based グラフカーネルにおける性能を示す．各点は，それぞれパラメータ設定により得られた各モデルを表す．キラリティ導入前の方法 (×) と導入後の方法 (○) の各モデルに対して，そのモデルから得られた 2 つの評価指標の値を示す．キラリティを考慮した場合の方が，2 つの評価指標ともに値の良いモデルが得られていることがわかる．図 5(b) はキラリティの導入前後の性能を比較したものである．ここでも，各点は各モデルを表し，2 つの評価指標それぞれ (+，＊) に対して，キラリティ導入前後の値を比較している．ここでもキラリティ導入による性能の向上が見られた．Size-based カーネルや Until-N Branching based カーネルでも同様の結果が得られたが，ページ数の都合により省略する．

また表 2 と同様の SVR に関するデータを表 3 に示す．q^2 に対する最高の予測精度を示したモデルに対し，R を用いて評価すると，既存手法の 0.499 に対し我々の提案手法は 0.553 であり，ここでも予測精度の改善が見られた．

4 まとめ

本稿では従来の Tree-Pattern グラフカーネルを拡張し，トポロジーが同じでありながら立体配置の異なる化合物を区別可能な方法を提案した．我々は Ecdysteroids と Cramer's Steroids のデータ集合に対し，SVM と SVR を用いて提案手法を実装した．その結果キラルや cis/trans を考慮した場合の方が，それらを考慮しない場合よりも予測精度が向上することがデータ 1 に関しては確認された．しかしデータ 2 に関しては

予測精度は改善されないことが多かった．その理由としては，データ２はトポロジーが同じで立体配置のみ異なるようなデータのペアが少ないことと，データの数そのものが少ないことが考えられる．

参考文献

[1] 橋田充: 薬物動態特性の *In Silico* 予測に関する研究, *YAKUGAKU ZASSHI*, Vol. 125, No. 11, pp. 853–861 (2005).

[2] 田辺和俊, 大森紀人, 小野修一郎, 松本高利, 長嶋雲兵, 上坂博亨, 鈴木孝弘: 化学物質の毒性情報と構造活性相関予測, 情報知識学会誌, Vol. 16, No. 3, pp. 63–84 (2006).

[3] Gasteiger, J., Engel, T.: Chemoinformatics, Wiley-VCH, (2003).

[4] Cramer, R. D., Patterson, D. E. and Bunce, J. D.: Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, *J. Am. Chem. Soc.*, Vol. 110, No. 18, pp. 5959–5967 (1988).

[5] Hopfinger, A. J., Wang, S., Tokarski, J. S., Jin, B., Albuquerque, M., Madhav, P. J. and Duraiswami, C.: Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism, *J. Am. Chem. Soc.*, Vol. 119, pp. 10509–10524 (1997).

[6] Hopfinger, A. J., Reaka, A., Venkatarangan, P., Duca, J. S. and Wang, S.: Construction of a Virtual High Throughput Screen by 4D-QSAR Analysis: Application to a Combinatorial Library of Glucose Inhibitors of Glycogen Phosphorylase *b*, *J. Chem. Inf. Comput. Sci.*, Vol. 39, pp. 1151–1160 (1999).

[7] Ralaivola, L., Swamidass, S. J., Saigo, H. and Baldi, P.: Graph kernels for chemical informatics, *Neural Networks*, Vol. 18, pp. 1093–1110 (2005).

[8] Mahé, P., Ueda, N., Akutsu, T., Perret, J.-L. and Vert, J.-P.: Graph Kernels for Molecular Structure-Activity Relationship Analysis with Support Vector Machines, *J. Chem. Inf. Model.*, Vol. 45, pp. 939–951 (2005).

[9] Mahé, P. and Vert, J.-P.: Graph kernels based on tree patterns for molecules, Technical Report HAL-00095488, Ecole des Mines de Paris (2006).

[10] Golbraikh, A. and Tropsha, A.: QSAR Modeling Using Chirality Descriptors Derived from Molecular Topology, *J. Chem. Inf. Comput. Sci.*, Vol. 43, pp. 144–154 (2003).

[11] 森謙治: 生物活性物質の化学-有機合成の考え方を学ぶ, 化学同人 (2002).

[12] 日本化学会 (編): 光学異性体の分離, 季刊 化学総説, No. 6, 学会出版センター, chapter 2 (1989).

[13] Clément, C. Y., Bradbrook, D. A., Lafont, R. and Dinan, L.: Assessment of a microplate-based bioassay for the detection of ecdysteroid-like or antiecdysteroid activities, *Insect Biochem. Molec. Biol*, Vol. 23, pp. 187–193 (1993).

[14] Dinan, L., Hormann, R. E. and Fujimoto, T.: An extensive ecdysteroid CoMFA, *Journal of Computer-Aided Molecular Design*, Vol. 13, pp. 185–207 (1999).

[15] Ravi, M., Hopfinger, A. J., Hormann, R. E. and Dinan, L.: 4D-QSAR Analysis of a Set of Ecdysteroids and a Comparison to CoMFA Modeling, *J. Chem. Inf. Comput. Sci.*, Vol. 41, pp. 1587–1604 (2001).

[16] Hormann, R. E., Dinan, L. and Whiting, P.: Superimposition evaluation of ecdysteroid agonist chemotypes through multidimensional QSAR, *Journal of Computer-Aided Molecular Design*, Vol. 17, pp. 135–153 (2003).

[17] Silverman, B. D.: The Thirty-one Benchmark Steroids Revisited: Comparative Molecular Moment Analysis (CoMMA) with Principal Component Regression, *Quantitative Structure-Activity Relationships*, Vol. 19, pp. 237–246 (2000).

[18] Pavlidis, P., Wapinski, I. and Noble, W. S.: Support vector machine classification on the web, *Bioinformatics*, Vol. 20, pp. 586–587 (2004).

[19] Joachims, T.: Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning, B. Scholkopf and C. Burges and A. Smola (ed.), MIT-Press, (1999).