

ILPを用いた植物の免疫活性を誘導する化合物の構造予測

Prediction of compounds leading to immune activity of plants using ILP

松本 淳志*¹
Atsushi Matsumoto

金盛 克俊*²
Katsutoshi Kanamori

大和田 勇人*³
Hayato Ohwada

*¹*²*³ 東京理科大学 理工学部 経営工学科
Department of Industrial Administration, Tokyo University of Science

Recent years, plant immune activator has been noted. But kind of plant immune activator is less because it is difficult to find a compound that contributes to the immune activity. In this study, we try to predict of compounds leading to immune activity of plants using ILP. By using the proposed method, it is possible to predict of compounds leading to immune activity of plants using Structural information of the compound. Predicted structure rule also includes structures in the known plant immune activator. The relationship of unknown structure rules of the plant immune activity, there is a need for further investigation

1. はじめに

病気や害虫被害による農作物の生産性低下の問題は未だ解決されていない重大な問題である。この問題に対して生産者は殺菌剤や殺虫剤といった化学農薬での対処を行ってきた。これらは直接病原菌や害虫を殺すため、圃場の生物相の破壊やヒトの健康被害の原因となる恐れがある。そこで近年、菌や害虫を殺すのではなく植物自身の免疫力を高めるという発想から、植物免疫活性剤を用いる手法が注目されている。しかし植物免疫活性剤の候補化合物を発見することは非常に時間やコストがかかるという問題がある。現在市販されている植物免疫活性剤は非常に種類が少なく、植物の免疫活性化のメカニズムも未知の部分が多い[1]。

そこで本研究では ILP を用いて植物免疫活性剤となる化合物を構造から予測し、ルールとして示す手法を提案する。ILP は述語論理を用いた機械学習の手法の一つであり、データ間の関係的、論理的なパターンを発見出来るという特徴から、化合物の構造を表現することに適している。また予測した化合物の構造をルールとして得ることが出来るのが ILP を用いるメリットである。近年、化合物データに対して ILP を用いて構造予測を行う研究が行われており、高い予測精度を示している[2]。しかしこれは化合物が結合する標的タンパク質が既知の場合であるため、植物の免疫活性化のようにメカニズムの解明が行われていないケースへの適用は行われていない。そこで本研究では標的タンパク質が未知な場合でも適用出来る ILP による構造予測の手法を提案する。

2. 提案手法

本研究の提案手法について述べる。本手法では背景知識として各化合物データに関する情報を述語として定義し、ILP システムに取り入れることで学習を行う。定義した述語は次の通りである。カッコ内は引数を表す。

1. `atom(compound_name,atom_id,element)`
化合物中の原子の種類(C, H など)を表現する
2. `bond(Compound_name,atom_id,atom_id,bondtype)`
どの原子と原子が結合しているのかを表し、その結合の種類を表現する。

3. `Num_AromaticRings(compound_name,Num_AromaticRing)`
化合物中に含まれる芳香環構造の数を表現する。
4. `Num_Rings(compound_name,Num_Ring)`
化合物中に含まれる環状構造の数を表現する。
5. `LogP98(compound_name,value)`
化合物の脂溶性を表現する。
6. `LogD(compound_name,value)`
Ph 値の変化によって脂溶性がどのように変化するかを指標であり、薬の投与した用量と作用の関係性を表す指標である。
7. `ring(compound_name,ring_id,atom_id,ringsize,ringtype)`
各化合物に含まれる原子が、その化合物のどのような環状構造に含まれているのかを表現する。環状構造に含まれる原子と他の構造との繋がりを表現出来る他、2 つ以上の環状構造が存在する化合物に対して、それら 2 つの環状構造の関係性を表現出来るようになる。

以上 7 つの化合物に関する述語からいくつかを選択し、組み合わせ背景知識として ILP システムに取り入れることで化合物の構造を学習結果として得ることが出来る。背景知識は各述語の原子式の集合である。化合物の構造を表現する上で `atom` と `bond` は必ず必要であるため、この 2 つの述語は必ず選択することとする。

3. 実験

植物免疫活性剤に関する化合物データを用いて構造予測を行う。

3.1 データセット

本実験では朽津ら[3]が実際に実験を行ったシロイヌナズナの植物免疫活性剤に関する 10000 の化合物データである。活性化を示した 271 の化合物を正事例とし、残りの 9729 の化合物を負事例とするが、データが不均衡の場合、学習精度の低下を招くので負事例をアンダーサンプリングして 813 に減らした。よって正事例 271、負事例 813 の合計 1084 化合物を実験に用いる。

3.2 データ処理

化合物データの情報を提案手法で定義した述語の形式に変換し、ILP システムに取り入れられるようにする。提案手法で定

表 1 背景知識に用いる述語の組み合わせ

ILP1	atom,bond
ILP2	atom,bond,Num_AromaticRings
ILP3	atom,bond,Num_AromaticRings,Num_Rings
ILP4	atom,bond,AlogP98
ILP5	atom,bond,Num_AromaticRings,Num_Rings,AlogP98,LogD
ILP6	atom,bond,Num_AromaticRings,Num_Rings,LogD
ILP7	atom,bond,LogD,ring
ILP8	atom,bond,ring

義した述語からいくつかを選択し、8つの組み合わせで実験を行った。述語の組み合わせを表1に示す

3.3 ILPシステムによる学習

背景知識を ILP システムに取り込んで学習を行う。本研究では ILP システムとして GKS[4]を用いる。GKS は背景知識データ、正事例データ、負事例データの3つのデータを入力することで、学習結果として正事例に多く共通する構造ルールを出力することが出来る。10分割交差検定を行い、予測精度を検証する。

4. 結果

4.1 予測精度

提案手法による予測精度を表2に示す。表2中の tp,fn,tn,fp はそれぞれ true positive,false negative,true negative,false positiveを表す。atom,bond,ring の3つの述語からなる背景知識を用いた ILP8 が F 値が一番高いという結果が得られた。化合物の環状構造に関する述語を背景知識に用いた場合が用いなかった場合に比べて高い精度となっている。

4.2 出力された化合物の構造

F 値が一番高かった ILP8 において、特にスコアの高かったルールを示す。ルールのスコア(Score)はそのルールが導く正事例の数(T)からそのルールが導く負事例の数(F)を引いた値である。

$$\text{Score} = T - F \quad (1)$$

出力された構造のルールを次に示す。

1. 原子 C は原子 6 つから成る芳香環と単結合している。
{正事例:27, 負事例:10}
2. 原子 S が芳香環内に存在し、その他に何かの原子と二重結合している C 原子が同一化合物内に存在する。
{正事例:20, 負事例:8}
3. 二つの芳香環が単結合しており、それぞれの芳香環から少なくとも一つずつ単結合している原子が存在する。
{正事例:22, 負事例:10}

表 2 背景知識に用いた述語の組み合わせごとの予測精度

	ILP1	ILP2	ILP3	ILP4	ILP5	ILP6	ILP7	ILP8
tp	92	116	127	88	131	139	165	165
fn	179	155	144	183	140	132	106	106
tn	699	644	605	712	572	568	523	542
fp	114	169	208	101	241	245	290	271
Accuracy	0.730	0.701	0.675	0.738	0.649	0.652	0.635	0.652
Precision	0.447	0.407	0.379	0.466	0.352	0.362	0.363	0.378
Recall	0.339	0.428	0.469	0.325	0.483	0.513	0.609	0.609
F値	0.386	0.417	0.419	0.383	0.407	0.424	0.455	0.467



図 1 既知の植物免疫活性剤

4. 原子 N を含む芳香環が、原子 5 つから成る芳香環と単結合している。
{正事例:15, 負事例:3}
5. 原子 S を含む芳香環がもう一つの芳香環と単結合している。
{正事例:14, 負事例:2}

5. まとめと今後の課題

表2より、環状構造の情報を学習に取り入れた場合、予測精度が高まることがわかった。出力されたルールと既知の植物免疫活性剤の構造を見比べてみると、チアジニルとイソチアニルはルール2に、プロペナゾールはルール1, 2, 5に当てはまることが分かった。植物の免疫活性化との関係性が未知なルールに関しては、実際に実験を行い調査していく必要がある。

参考文献

- [1] Yoshiteru Noutoshi, Masateru Okazaki, Tatsuya Kida, Yuta Nishina, Yoshihiko Morishita, Takumi Ogawa, Hideyuki Suzuki, Daisuke Shibata, Yusuke Jikumar, Atsushi Hamada, Yuji Kamiya, Ken Shirasu, Novel Plant Immune-Priming Compounds Identified via High-Throughput Chemical Screening Target Salicylic Acid Glucosyltransferases in Arabidopsis. The Plant Cell, vol.24:3795-3804, 2012
- [2] Jose C A Santos, Houssam Nassif, David Page, Stephen H Muggleton, Michael J E Sternberg, Automated identification of protein-ligand interaction features using Inductive Logic Programming:a hexose binding case study. Santos et al. BMC Bioinformatics 2012, 13:162, 2012
- [3] T Higashi, T Kurusu, S Hasegawa, K Kuchitsu, Dynamic intracellular reorganization of cytoskeletons and the vacuole in defense responses and hypersensitive cell death in plants. Journal of Plant Research, Volume 124, Issue 3, pp315-324, 2011
- [4] Hayato Ohwada, Hiroyuki Nishiyama, Fumio Mizoguchi, Concurrent execution of optimal hypothesis search for inverse entailment. Lecture Notes in Artificial Intelligence, Springer-Verlag, No.1866, Vol.4, pp.165-173, 2000