

# 有機合成における酵素番号予測のための 特徴選択とクラスタリングを用いた ケモインフォマティクス

Chemoinformatics Using Feature Selection and Clustering  
for Enzyme Commission Number Prediction  
in Organic Synthesis

武藤 克弥 (Katsuya Muto)  
t815070@st.pu-toyama.ac.jp

富山県立大学 工学部 電子・情報工学科 情報基盤工学講座

Teams, 14:20-14:35 Wednesday, February 16, 2022.

1. はじめに
2. 有機合成分野と情報分野の関わり
3. 背景
4. 提案手法
5. 実験結果並びに考察
6. おわりに

# 1. はじめに

2/24

※本研究を進めるにあたり、酵素分野に関して生物工学科酵素化学工学講座の浅野泰久教授、有機合成分野に関してくすりのシリコンバレーTOYAMA研究拠点化プロジェクトディレクター 岩崎源司博士(薬学)よりご指導いただいた。

1. はじめに
2. 有機合成分野と情報分野の関わり
4. 提案手法
5. 実験結果並びに考察
6. おわりに

近年、新型コロナウイルスなどの影響で新薬開発の需要が高まっており、効率的に目的物を生成するため、酵素を用いることが重要視されるようになった。

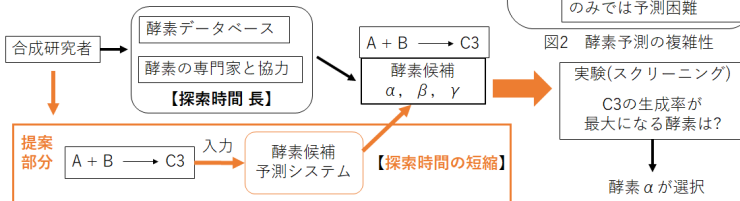
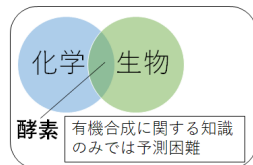
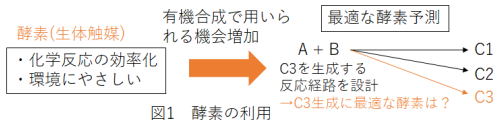


図3 酵素選択の流れ

## 2.1 有機合成と酵素

3/24

有機合成では生体内に存在する有機物などを人工的に生成し、医薬品や必要となる化合物を作り出す。

### 【酵素について】

- ・生体内の化学反応を触媒する生体触媒
- ・特定の化合物(基質)のみ作用し、反応を効率に進める。
- ・化学触媒に比べて、環境にやさしい。

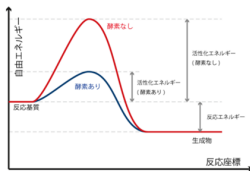


図4 酵素の有無による目的物生成に必要なエネルギー<sup>1</sup>

### 【有機合成で酵素を用いるメリット】

- ・目的物生成の合成ステップ短縮。
- 用いる試薬削減(環境面) + コストの減少

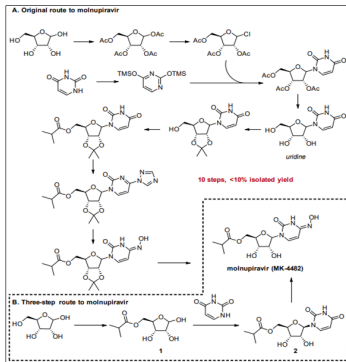


図5 モルヌピラビルの合成ステップ短縮化<sup>2</sup>

モルヌピラビル  
合成ステップ

従来：10ステップ

生体触媒の使用

3ステップに短縮

1 “酵素基質とは”, <https://bizcomjapan.co.jp/iris-biotech/knowledge/substrate/>, 閲覧日 2022.1.31.

2 Tamas Benkovics, John A. McIntosh, Steven M. Silverman, Jongrock Kong, Peter Maligres, Tetsuji Itoh, Hao Yang, et al.,

“Evolving to an Ideal Synthesis of Molnupiravir, an Investigational Treatment for COVID - 19”, *ChemRxiv*, 2020.

## 2.2 酵素番号 (EC 番号)

4/24

酵素を 4 組の数字 (EC ○. ○. ○. ○) の組み合わせで分類したもの。

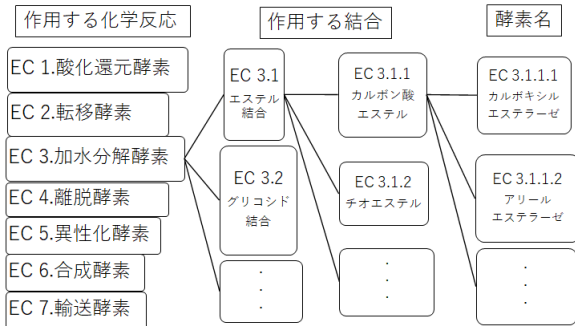


図6 EC番号による酵素分類

EC 3.1.1.1  
(カルボキシルエステラーゼ)  
を用いた代表的な反応式  
=自然界で観測される反応

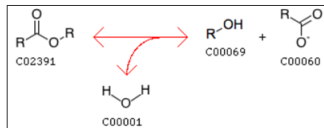


図7 EC 3.1.1.1の代表的な反応式

## 2.3 使用する化学データベース

5/24

Kyoto Encyclopedia of Genes and Genomes(KEGG) と PubChem から必要となる情報を取得する。

**KEGG: 酵素情報(遺伝子情報, タンパク質相互作用), 酵素の化学反応情報などを記載,**

**PubChem: 化合物の化学・物理特性, 毒性情報, 文献情報等を収録,**

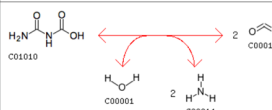
Entry	R00005	Reaction
Name	urea-1-carboxylate amidohydrolase	
Definition	Urea-1-carboxylate + H2O <=> 2 CO2 + 2 Ammonia	
Equation	C01010 + C00001 <=> 2 C00011 + 2 C00014	
		
Comment	The yeast enzyme (but not that from green algae) also reaction of EC 6.3.4.6 urea carboxylase, thus bringin hydrolysis of urea to CO2 and NH3 in the presence of bicarbonate. R00774 (6.3.4.6)	
Reaction class	RC02756 C00011_C01010	
Enzyme	3.5.1.54	

図8 KEGGのWebサイト<sup>3</sup>


PubChem SID	3669
Structure	
Source	KEGG
External ID	C00379
Source Category	Curation Efforts Research and Development
Version	11 <a href="#">Revision History</a>
Status	Live
Related Compounds	PubChem CID CID 6912 (Xylitol)

図9 PubChemのWebサイト<sup>4</sup>

それぞれのデータベースで実装されているAPIを用いてデータを取得

3 "KEGG: Kyoto Encyclopedia of Genes and Genome", [https://www.genome.jp/kegg/kegg\\_ja.html](https://www.genome.jp/kegg/kegg_ja.html), 閲覧日 2022.1.17.

4 "PubChem", <https://pubchem.ncbi.nlm.nih.gov/>, 閲覧日 2022.1.17.

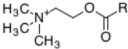
## 3.2 計算機での化学構造表現, EC 番号予測

6/24

化学構造を計算機上で扱うための, 様々な表現法が開発されている。

1. はじめに
2. 有機成分分野と情報分野の関わり
3. 提案手法
4. 提案手法
5. 実験結果並びに考察
6. おわりに

構造式



SMILES

\*C(=O)OCC[N+](C)(C)C

フィンガープリント

100001011 . . .

—特徴的な構造の有無を判定

—化合物の類似度比較

物理・化学的特性値

・分子量, 電荷, 疎水性など

・化合物A: (100, 8.32, -0.23, . . .)

—特性値の特徴ベクトルで構造を表現

特徴量(記述子)として機械学習に適用

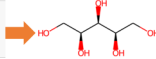
図10 計算機上における化学構造表現

### RDKit

プログラム上で化学構造を扱うためのPythonライブラリ

- ・ SMILES
- ・ フィンガープリント(構造記述子)
- ・ 特性値を表現する記述子(208種類実装)

```
from rdkit.Chem
Xylitol = Chem.MolFromMolFile('Xylitol.mol')
Xylitol
```



```
from rdkit.Chem import Descriptors
print("SMILES: " + Chem.MolToSmiles(Xylitol))
print("分子量: " + str(Descriptors.MolWt(Xylitol)))
print("LogP: " + str(Descriptors.MolLogP(Xylitol)))
print("TPSA: " + str(Descriptors.TPSA(Xylitol)))
```

SMILES: OC[C@H](O)[C@@H](O)[C@H](O)CO  
 分子量: 152.14600000000002  
 LogP: -2.9462999999999995  
 TPSA: 101.15

図11 RDKitを用いたSMILES・特性値の算出

## 4.1 提案手法 (EC 番号の予測)

7/24

ターゲット反応式と EC 反応式の構造変化を比較し、最も類似する EC 反応式の EC 番号を最適な酵素候補として予測する。

反応物から生成物になる際の特性値変化量N種類 = 化合物の構造変化の指標

【仮定】

- ・ターゲットとEC反応式の構造変化が類似  
→ そのEC番号の酵素 = 目的物生成に最適な酵素候補(類似性の概念)

提案部分

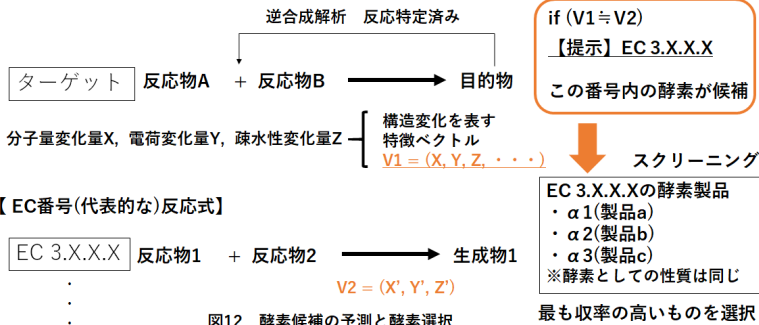


図12 酵素候補の予測と酵素選択

1. はじめに
2. 有機合成分野と情報分野の関わり
4. 提案手法
5. 実験結果並びに考察
6. おわりに

## 4.1 構造変化を捉える特徴

8/24

反応物から生成物に変化する際の「特性値変化量」で類似性比較

先行研究: 反応差分フィンガープリント(RFP)

EC 番号反応式: 反応物 1 + 反応物 2 → 生成物 1 + 生成物 2

構造変化の表現

1種のFPでは限界あり(=多種のFPが開発)

$$\rightarrow RFP = FP_{\text{生成物 1+生成物 2}} - FP_{\text{反応物 1+反応物 2}}$$

→多数の「物理・化学的特性値の変化量」を用いる  
(RDKit: 特性値を表現する記述子208種)

提案手法: 特性値変化量

各反応式に対する記述子( $n$ 種)の特性値:  $cv_j (j = 1, 2, \dots, n)$

$$cv_j = (PD_1 + PD_2) - (RT_1 + RT_2)$$

( $RT_i$ : 反応物  $i$  の特性値,  $PD_i$ : 生成物  $i$  の特性値)

各反応式の特徴ベクトル:  $DF_i (i = 1, 2, \dots, m)$

$$DF_i = (cv_{i1}, cv_{i2}, \dots, cv_{ij}, \dots, cv_{in})$$

	記述子 1	記述子 2	...	記述子 $n$
$DF_1$	$cv_{11}$	$cv_{12}$	...	$cv_{1n}$
$DF_2$	$cv_{21}$	$cv_{22}$	...	$cv_{2n}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$DF_m$	$cv_{m1}$	$cv_{m2}$	...	$cv_{mn}$

表1 各反応式の特性値変化量

$DF_T$  (ターゲット)に最も類似する  $DF_i$  のEC番号を提示

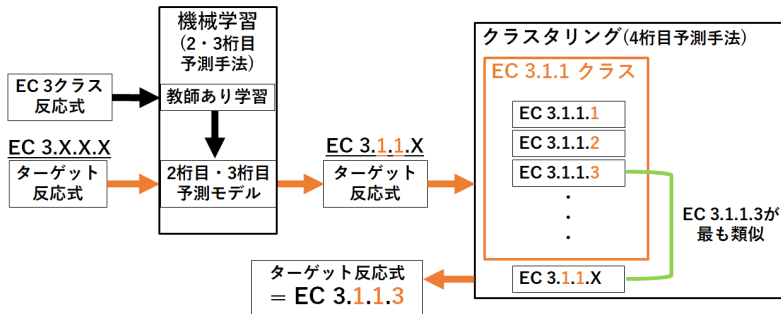


## 4.1 EC 番号予測の流れ

9/24

EC 3 クラスのみを対象とし、2・3 桁目の予測を行ったのち、4 桁目の予測を行う。

4桁目：KEGG学習データ少→クラスターリング(教師なし学習)



1. はじめに
2. 有機合成分野と情報分野の関わり
4. 提案手法
5. 実験結果並びに考察
6. おわりに

## 4.2 EC3 クラス 4 桁目予測手法 (前処理)

10/24

本処理：SOM によってターゲット反応式と各 EC 番号反応式の類似性を評価  
前処理：次元削減 (凝集型クラスタリング)

### 前処理で多重共線性の対処を行う

**多重共線性**：記述子間に高い相関があると発生  
= 同じ記述子があると同義 → 予測精度低下の原因

緑線：———

||

記述子間で最も相関が低い  
(距離が遠い)ペア  
→ クラスタ間距離

クラスタ1

クラスタ間の相関が最も高い  
||  
クラスタ間距離が最短

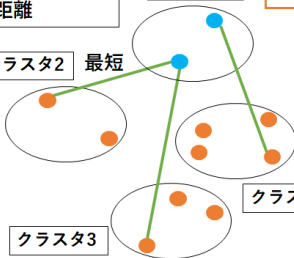
クラスタ2 最短

マージ

クラスタ 1+2

クラスタ4

クラスタ3



1. はじめに
2. 有機合成分野と情報分野の関わり
4. 提案手法
5. 実験結果並びに考察
6. おわりに

## 4.2 EC3 クラス 4 桁目予測手法 (本処理)

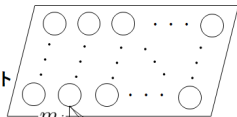
11/24

### SOM による反応式クラスタリング

自己組織化マップ (Self-Organizing Map: SOM) を用いて、次元削減後の特徴ベクトルを 2 次元上に可視化し、クラスタリングを行う

出力層

k個の  
ユニット



$$\xi_i = (\xi_{i1}, \xi_{i2}, \dots, \xi_{ip})$$

入力層  $x_j = (x_{j1}, x_{j2}, \dots, x_{jp}), (j = 1, 2, \dots, n)$

→ n個の反応式, p種の記述子からなる特徴ベクトル

ユニット  $m_i (i = 1, 2, \dots, k)$  ランダムに初期化

ユニット  $i$  の重心:  $r_i = (r_{i1}, r_{i2})$

ユニットの重みベクトル:  $\xi_i = (\xi_{i1}, \xi_{i2}, \dots, \xi_{ip})$

$x_j$  と  $\xi_i$  のユークリッド距離:  $\|x_j - \xi_i\|$

手順 (1):  $\|x_j - \xi_i\|$  が最小となる勝者ユニット  $m_c$  を求める

手順 (2):  $x_j$  に近くなるように  $m_c$  の重みを更新

$$\xi_i = \xi_i + \alpha(t)(x_j - \xi_i)$$

→  $\alpha(t)$ : 学習率係数 (学習回数  $t$  の増加で減少)

手順 (3):  $m_c$  の付近の  $m_i$  も多少  $x_j$  に近づくように更新

$$\xi_i = \xi_i + h(t)(x_j - \xi_i)$$

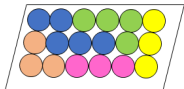
→  $h(t)$ : 近傍関数 (勝者ユニットから離れるほど影響力が弱まる)

$$h(t) = \alpha(t) \exp \left[ \frac{-\|r_c - r_i\|}{2\sigma^2(t)} \right] \quad (\sigma^2(t): \text{調整関数})$$

手順 (4):  $x_{j+1}$  から  $x_n$  まで 1,2 を繰り返す

指定した学習回数に達するまで, (1)~(4) を繰り返す

→ 最終的に確定した各勝者ユニットに反応式の特徴ベクトルをマッピング



各ユニットを色分けでクラスタリング

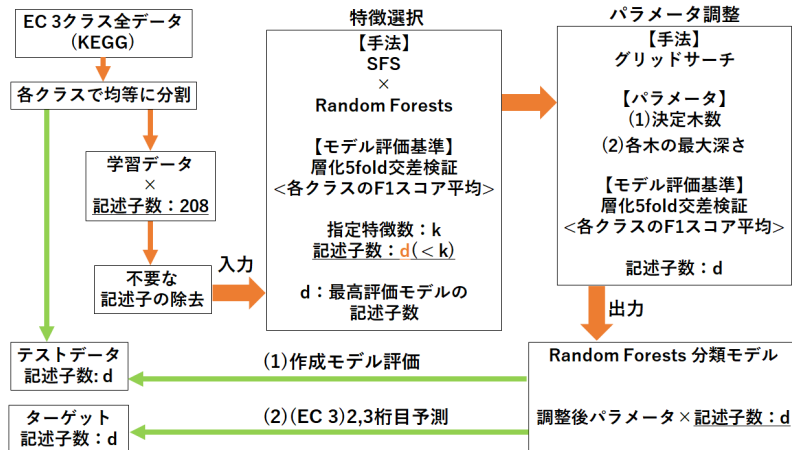
→ 各ユニット間のマージ:  $\xi_i$  どちらのユークリッド距離  
各クラス間のマージ: ウォード法

図14 各ユニットのクラスタリング

## 4.3 EC 3 クラス 2・3 桁目予測手法 (概要)

12/24

2・3 桁目までの範囲で EC 反応式を機械学習→分類モデルを作成  
→ターゲット反応式が分類されたクラス=予測結果



- はじめに
- 有機合成分野と情報分野の関わり
- 提案手法
- 実験結果並びに考察
- おわりに

## 4.3 EC 3 クラス 2・3 桁目予測手法 (前処理)

13/24

### ラッパー法による記述子選択

分類モデルの予測精度を評価し、最高評価となる組み合わせの記述子を選択する

用いる手法: Step Forward 法

- ①  $n$  個の記述子から 1 つ選択し、 $n$  種類の分類モデルを作成
- ② 最も評価の高いモデルに用いられている、記述子を選択する.
- ③  $n - 1$  個の記述子から 1 つ選択し、先ほど選択されたモデルに追加することで、新たな分類モデルを作成する.
- ④  $n - 1$  個のモデルで最も評価の高いものに用いられている、記述子の組み合わせを選択する.
- ⑤ 指定した特徴数になるまで 3 と 4 を繰り返す.

モデルの評価基準としては、各 EC 番号クラス分類時の F1 スコア平均 (+層化 5-fold 交差検証) を用いる.

1. はじめに
2. 有機合成分野と  
情報分野の関わり
4. 提案手法
5. 実験結果並びに  
考察
6. おわりに

## 4.3 EC 3 クラス 2・3 桁目予測手法 (本処理)

14/24

### ランダムフォレストによる EC 番号分類

複数の決定木を用いた機械学習モデル

下図の決定木を複数組み合わせたものが Random Forests

用いる決定木の数, 木の最大深さ, 用いる特徴量数などをハイパーパラメータとして調整する

各クラスの全データ数

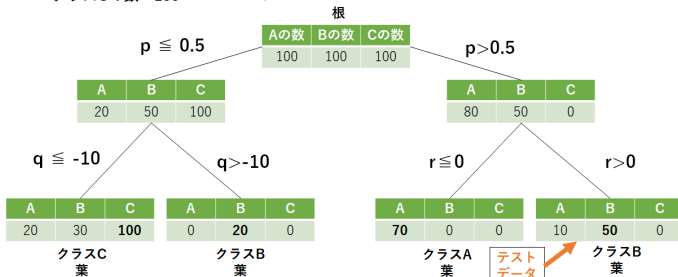
クラスAの数: 100  
クラスBの数: 100  
クラスCの数: 100

特徴量

p  
q  
r

テストデータ

p	q	r
0.7	-5	2



## 4.3 EC 3 クラス 2・3 桁目予測手法 (本処理)

15/24

### ランダムフォレストによる EC 番号分類

ノードの分割基準と用いる記述子は、情報利得  $IG$  で決まる

$$IG(D_P, f) = I(D_P) - \frac{N_{left}}{N_P} I(D_{left}) - \frac{N_{right}}{N_P} I(D_{right}) \quad (1)$$

$f$ : 分割時に用いられる特徴量,  $D_P$ : 上位ノード内のデータ,  
 $D_{left}$ ,  $D_{right}$ : 分割先の下位ノード内のデータ,  $N_P$ ,  $N_{left}$ ,  
 $N_{right}$ : 上位ノード, 下位ノードのデータ数,  $I_G(t)$ : ジニ不純度

$$I_G(t) = \sum_{i=1}^c p(i|t)(1 - p(i|t)) = 1 - \sum_{i=1}^c p(i|t)^2 \quad (2)$$

$p(i|t)$ : ノード  $t$  のクラス  $i$  のデータ割合,  $c$ : ノード  $t$  内のクラス数

→各ノードで  $IG$  が最大となるように、特徴量  $f$  とデータを分割する閾値が決定される。

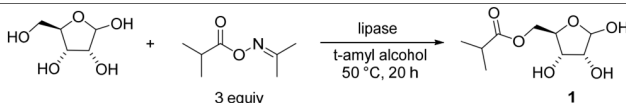
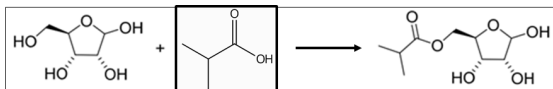
1. はじめに
2. 有機合成分野と情報分野の関わり
4. 提案手法
5. 実験結果並びに考察
6. おわりに

## 5.1 4 桁目予測 (準備)

16/24

ターゲット (2 つ) : 正解が EC3.1.1.3 だと分かっている EC 反応式 (比較対象) :  
EC3.1.1 クラスに属する反応式

target1



target2

Enzyme Name	5-isobutyryl ribose assay yield%
Novozym 435 (Candida antarctica lipase B)	65
IMMTLL-T2-150 (Thermomyces lanuginosus lipase)	40
IMMRES-T2-150 (Resinase HT lipase)	38
IMMLIPX-T2-150 (Lipex 100 L lipase)	56
IMML51-T2-150 (Novozymes 51032)	61
IMMP6-T2-250 (protease from Bacillus licheniformis)	11
Lipozyme RM IM	10
CDX IMB-103	33

1. はじめに
2. 有機合成分野と情報分野の関わり
3. 課題
4. 提案手法
5. 実験結果並びに考察
6. おわりに



## 5.2 4 桁目予測結果 (次元削減)

17/24

ターゲット 1(ターゲット 2) では 16(15) 個のクラスが形成され, 73(76) 次元特徴ベクトルが生成された.

	cluster0	cluster1	cluster2	cluster3	cluster4	cluster5	cluster6	cluster7	cluster8	cluster9	...	PEOE_VSA10	SMR_VSA6	SlogP_VSA10
<b>T</b>	-0.358029	-0.842985	0.204247	4.571328	0.622585	0.207469	-1.696273	2.341995	0.173683	-0.133043	...	0.796801	0.043121	-0.1269
<b>33</b>	-0.223667	-0.842985	0.219299	-0.103504	0.568775	0.207469	0.045521	-0.141173	0.173683	-0.133043	...	0.796801	0.043121	-0.1269
<b>6</b>	3.495807	-0.842985	0.161480	-0.103504	-0.360529	0.207469	0.004312	-0.141173	0.173683	-0.133043	...	0.046341	0.043121	-0.1269
<b>1</b>	3.518113	1.079686	0.207055	-0.103504	-0.497256	0.207469	0.009803	-0.141173	0.173683	-0.133043	...	0.046341	0.043121	-0.1269
<b>7_8</b>	-0.214983	-0.842985	0.219101	-0.103504	0.596503	0.207469	0.036318	-0.141173	0.173683	-0.133043	...	0.796801	0.043121	-0.1269
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
<b>106.1</b>	-0.233937	-0.842985	0.213692	-0.103504	0.310413	0.207469	0.083963	-0.141173	0.173683	-0.133043	...	-0.646994	0.043121	-0.1269
<b>113</b>	-0.251825	-0.842985	0.217973	-0.103504	0.596561	0.207469	0.012651	-0.141173	0.173683	-0.133043	...	0.046341	0.043121	-0.1269
<b>112</b>	-0.182730	-0.842985	0.219101	-0.103504	0.501004	0.207469	0.031504	-0.141173	0.173683	-0.133043	...	0.046341	0.043121	-0.1269
<b>111</b>	-0.280504	1.079686	0.215345	-0.103504	-0.970931	0.207469	0.055048	-0.141173	0.173683	-0.133043	...	-1.333272	0.043121	-0.1269
<b>118</b>	-0.266600	1.079686	0.223487	-0.103504	-1.377476	0.207469	0.075928	-0.141173	0.173683	-0.133043	...	0.046341	0.043121	-0.1269

- はじめに
- 有機合成分野と情報分野の関わり
- 提案手法
- 実験結果並びに考察
- おわりに

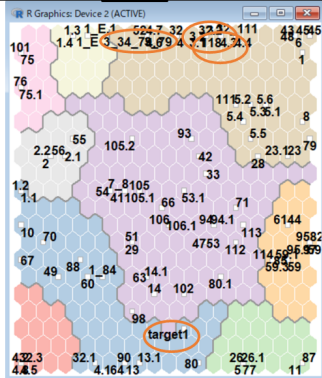
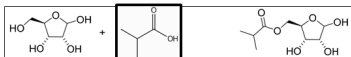
## 5.2 4 桁目予測結果 (SOM)

18/24

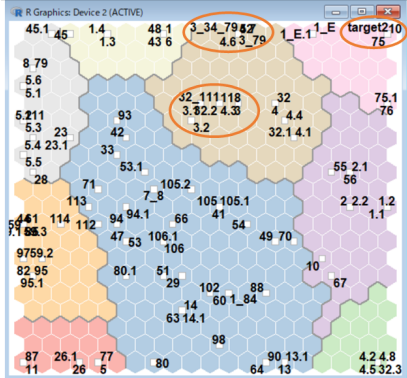
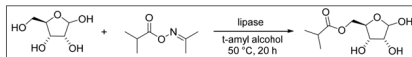
- ・ユニット数: 400(20×20)
- ・学習回数: 1,000(1段階目)→大まかな順序付け  
200,000(2段階目)→収束段階

T: ターゲット, E: EC3.1.1以外のEC番号,  
ピリオド: EC代表反応式が複数ある場合の区別,  
アンダーバー: EC番号重複

target 1



target 2



1. はじめに
2. 有機成分分野と情報分野の関わり
4. 提案手法
5. 実験結果並びに考察
6. おわりに

## 5.1 2・3桁目予測 (準備)

19/24

EC 反応式クラス：20, 反応式数が 6 以上のクラスのみ採用

表 5.5: EC 番号クラスと全データ数

クラス名	反応式数	クラス名	反応式数	クラス名	反応式数
3.1.1	122	3.2.2	24	3.5.5	12
3.1.2	59	3.3.2	6	3.5.99	11
3.1.3	152	3.4.13	6	3.6.1	94
3.1.4	29	3.4.19	7	3.7.1	35
3.1.6	14	3.5.1.	155	3.8.1	16
3.1.7	8	3.5.3	25	3.13.1	9
3.2.1	131	3.5.4	47	合計	962

学習データ：テストデータ = 4:1

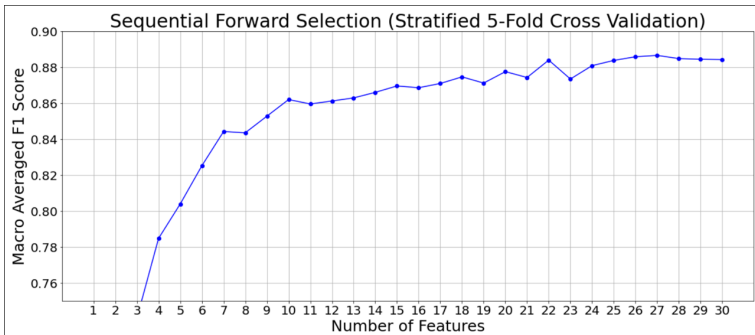
学習データ(769) → SFS × グリッドサーチ  
→ (記述子選択済み) Random Forest 分類モデル

テストデータ(193) → RFモデル → モデル評価

パラメータ調整時のモデル評価基準：

RandomForests 層化 5-fold 交差検証 (各クラスの F1 スコア平均)

→記述子数：27, 各決定木の最大深さ：15, 決定木数 100 が選択



1. はじめに
2. 有機合成分野と情報分野の関わり
3. 提案手法
4. 提案手法
5. 実験結果並びに考察
6. おわりに

# 5.1 2・3桁目予測結果 (生成モデル評価)

21/24

表 5.7: 作成モデルの各 EC クラスにおける分類精度

EC	要素数	適合率	再現率	F1 値	EC	要素数	適合率	再現率	F1 値
3.1.1	25	0.96	0.96	0.96	3.4.19	1	1.00	1.00	1.00
3.1.2	12	0.92	1.00	0.96	3.5.1	31	0.94	0.97	0.95
3.1.3	31	0.91	1.00	0.96	3.5.3	5	0.83	1.00	0.91
3.1.4	6	0.86	1.00	0.92	3.5.4	9	0.89	0.89	0.89
3.1.6	3	1.00	1.00	1.00	3.5.5	2	1.00	1.00	1.00
3.1.7	2	0.00	0.00	0.00	3.5.99	2	1.00	0.50	0.67
3.2.1	26	0.96	0.96	0.96	3.6.1	19	0.86	0.95	0.90
3.2.2	5	0.83	1.00	0.91	3.7.1	7	1.00	0.71	0.83
3.3.2	1	1.00	1.00	1.00	3.8.1	3	1.00	0.67	0.80
3.4.13	1	0.00	0.00	0.00	3.13.1	2	1.00	0.50	0.67
					合計	193			
					平均		0.85	0.80	0.81
					正解率				0.92

- はじめに
- 有機合成分野と情報分野の関わり
- 提案手法
- 実験結果並びに考察
- おわりに

## 5. 2・3桁目予測結果(ターゲット予測)

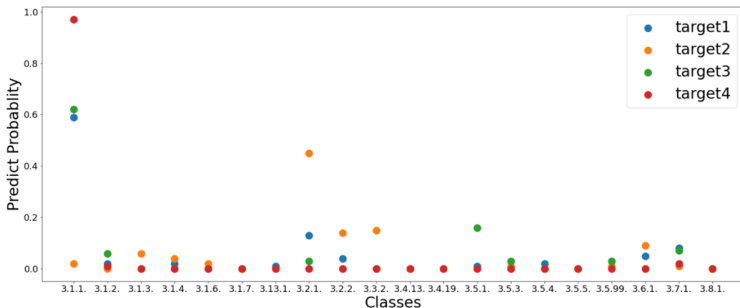
22/24

正解クラス EC3.1.1 のターゲット 4 つの予測

(1 と 2 : 4 桁目予測のターゲット, 3 と 4 : 別の文献の EC 反応式)

予測確率：各決定木の葉にあるデータの、クラス分布を平均したもの

→target2以外は正解クラスに最も高い確率



## 5. 2・3桁目予測結果(ターゲット予測2)

23/24

正解クラス EC3.5.3, 3.7.1 のターゲット 4 つでも予測  
(BRENDA データベースの登録論文のデータを使用)

ターゲット(正解ラベルEC 3.5.3)

	第1位	第2位	第3位	第4位	第5位
target1	3.5.3.	3.1.6.	3.5.99.	3.5.4.	3.1.1.
"	0.96	0.02	0.01	0.01	0.0
target2	3.5.3.	3.5.4.	3.5.99.	3.1.1.	3.1.2.
"	0.95	0.03	0.02	0.0	0.0
target3	3.5.3.	3.5.4.	3.5.99.	3.2.2.	3.1.1.
"	0.89	0.06	0.03	0.02	0.0
target4	3.5.3.	3.2.1.	3.1.1.	3.4.19.	3.7.1.
"	0.99	0.01	0.0	0.0	0.0

ターゲット(正解ラベルEC 3.7.1)

	第1位	第2位	第3位	第4位	第5位
target1	3.7.1.	3.1.2.	3.5.1.	3.1.1.	3.5.5.
"	0.31	0.17	0.152	0.1	0.08
target2	3.2.2.	3.13.1.	3.5.4.	3.5.1.	3.1.2.
"	0.3495	0.210526	0.060526	0.052526	0.05
target3	3.1.1.	3.7.1.	3.6.1.	3.1.2.	3.5.1.
"	0.29	0.22	0.11	0.106	0.1
target4	3.1.1.	3.5.1.	3.7.1.	3.1.2.	3.5.5.
"	0.28	0.24	0.13	0.12	0.1

3.5.3は全て高い確率で正解

3.7.1は target1 以外は別クラスが1位の予測

1. はじめに
2. 有機合成分野と情報分野の関わり
3. 実験結果並びに考察
4. 提案手法
5. 実験結果並びに考察
6. おわりに

### 考察

- EC3.1.1 ターゲット 2 が別のクラスと予想  
→実験に使われた試薬や温度が影響した？  
→それらを考慮した特徴量の作成が必要
- 別のクラスに分類されるターゲットがいくつかみられた  
→全てのクラスをバランスよく分類できる，スタンダード的な特徴量の検討

### 今後の課題

- EC3 以外のクラスへの拡張
- 2・3 桁目予測から 4 桁目予測へのシステム接続
- EC 番号以外の分類法の追加 (特徴ベクトルクラスタリングなど)