

アドイン: Toolkit for Materials Informatics について

1. このアドインの目的

マテリアルズ・インフォマティクス、ケモ・インフォマティクスのためのアドインです。
以下のようなことが可能です。

- ・ 分子記述子を生成する
- ・ フィンガープリントを生成する
- ・ 分子の描画
- ・ 予測モデルで記述子を使用するための前処理

分子記述子の生成、フィンガープリントの生成、分子の描画などは、RDKit (RDKit)を使用しています。

2. システム要件

- ・ JMP 18 以降
- ・ 表示言語は、英語および日本語のみ
- ・ Windows OS
- ・ 以下の Python Package がインストールされ、JMP と連携できること
- ・ RDKit
- ・ Numpy
- ・ Pandas

Python integration の機能(JMP 18 の新機能)を利用しています。

3. テスト環境

- ・ JMP 18.0
- ・ Windows 11
- ・ English, Japanese
- ・ Python 3.11.5
- ・ RDKit 2023.09.6

4. 事前の準備

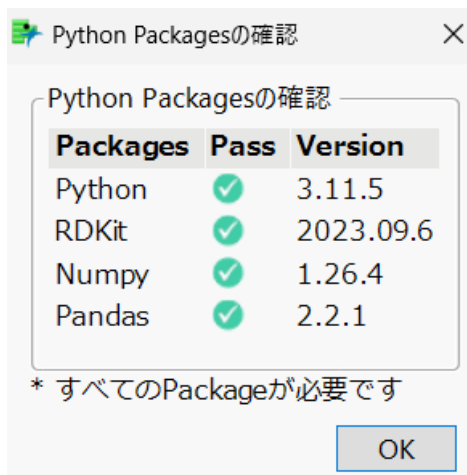
1) RDKit のインストール

1. [ファイル]>[新規作成]>[JSL スクリプト]を選択します
2. 以下のスクリプトを入力して[編集]>[スクリプトを実行]を選択します

```
Python Install Packages("numpy pandas rdkit");
```

2) アドインのインストール

1. 「ファイル」>「開く」で「Toolkit for Materials Informatics.jmpaddin」を選択して、インストールします。
2. 「アドイン」メニューにある「Toolkit for Materials Informatics」>「Check Python Packages」を選択します。以下のようなウィンドウが表示されれば、問題ありません。



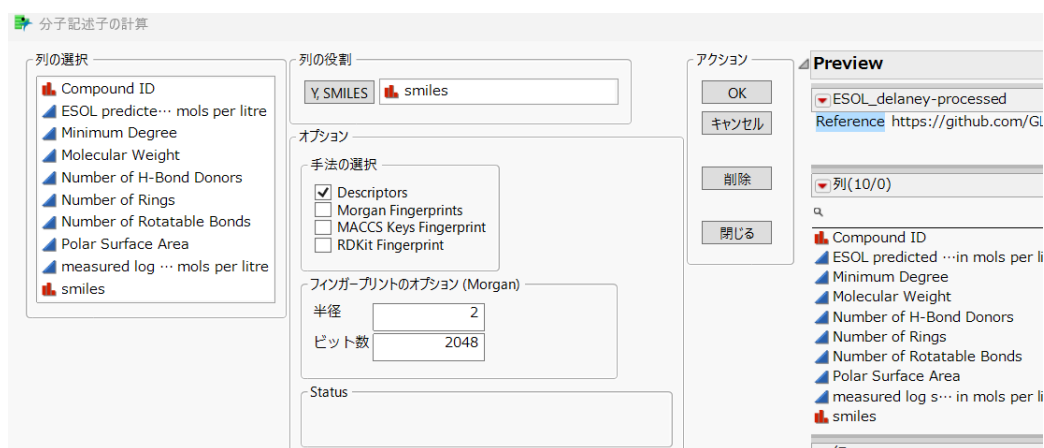
5. 使い方

このドキュメンテーションでは、ESOL_delaney-processed.csv (Lambard) を使用しています。

[1] 分子記述子/フィンガープリントの生成

SMILES から分子記述子/フィンガープリントを生成します。

- 1) 「アドイン」>「Toolkit for Materials Informatics」>「Descriptor Calculation」を選択します。
- 2) SMILES が入力されている列を「Y,SMILES」に指定します。手法の選択の一覧から生成したい、記述子、フィンガープリントにチェックを入れます。

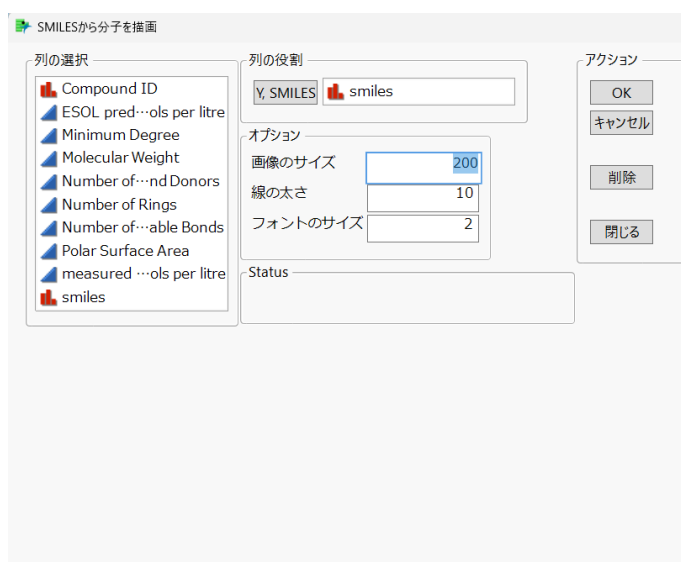


- 記述子、フィンガープリントは、RDKit を使用して生成しています。
<https://www.rdkit.org/docs/GettingStartedInPython.html#list-of-available-descriptors>
 - 記述子 (Descriptor) を選択した場合、上記のページの「List of Available Descriptors」の記述子が生成されます。
 - Morgan Fingerprints、MACCS Keys Fingerprint、RDKit Fingerprint については、上記のページの「List of Available Fingerprints」をご覧ください。
 - Morgan Fingerprints については、半径とビット数を指定してください。
- 3) 「OK」ボタンをクリックすると、選択した特徴量が生成されます。
 - 4) 終わりましたら、「閉じる」ボタンをクリックします。

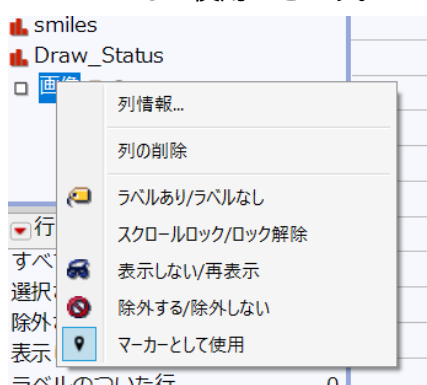
[2] 分子の描画

SMILES から分子を描画し、データテーブルに追加します。

- 1) 「アドイン」>「Toolkit for Materials Informatics」>「Drawing Molecules」を選択します。
- 2) SMILES が入力されている列を「Y,SMILES」に指定します。生成する画像のサイズを指定します。
- 3) 「OK」ボタンをクリックすると、データテーブルに生成された画像が入力された列が作成されます。



- 4) RDKit を利用して、画像を生成しています。
<https://www.rdkit.org/docs/source/rdkit.Chem.Draw.html#module-rdkit.Chem.Draw>
- 5) 列名の一覧から画像の列を右クリックして、「マーカーとして使用」を選択すると、画像をマーカーとして使用できます。



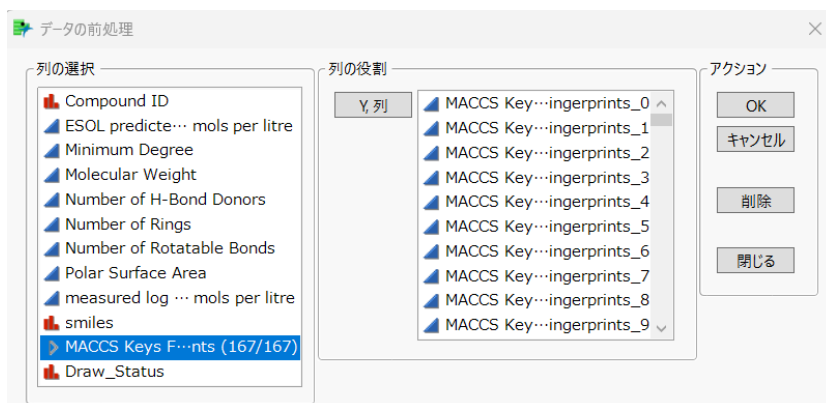
- 6) 画像をマーカーとして使用する場合、画像が見つからない場合は、マーカーのサイズを大きくするか、画像を生成する際の線の太さを調整してみてください。

[3] 前処理

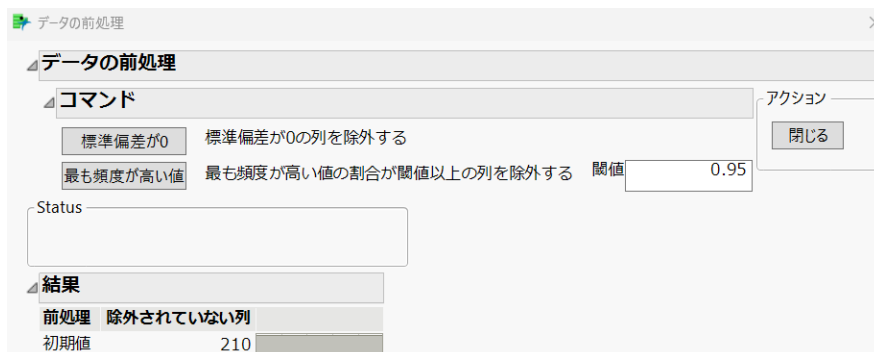
以下のような前処理を行うことができます。

- ・ 標準偏差が 0 である列を除外する
- ・ 最も頻度が高い値の割合が閾値以上である

- 1) 「アドイン」>「Toolkit for Materials Informatics」>「Preprocessing」を選択します。
- 2) 記述子、フィンガープリント等が入力されている列を「Y,列」に指定します。「OK」ボタンをクリックします。



- 3) 「標準偏差が 0」、「最も頻度が高い値」のボタンをクリックすると計算が行われ、列が除外されます。



- 4) 例えば、「標準偏差が 0」のボタンをクリックすると、各列に対して標準偏差が計算されます。標準偏差が 0 であった列に対して、除外の設定が行われます。どの列が除外されたかについては、レポート下の「結果」のアウトラインに表示されます。

結果

前処理 除外されていない列

初期値 210

標準偏差が0 191

標準偏差が0

データテーブルに出力

除外した列に関する情報

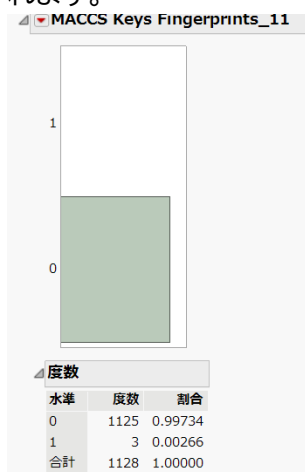
標準偏差が0

除外された列	残った列
Descriptor_NumRadicalElectrons	Descriptor_MaxAbsEStateIndex
Descriptor_SMR_VSA8	Descriptor_MaxEStateIndex
Descriptor_SlogP_VSA9	Descriptor_MinAbsEStateIndex

標準偏差が0

Column	SD
Descriptor_MaxAbsEStateIndex	3.7076812735
Descriptor_MaxEStateIndex	3.7076812735
Descriptor_MinAbsEStateIndex	0.4904189476
Descriptor_MinEStateIndex	1.2949130572

- 5) 「最も頻度が高い値」をクリックすると、例えばある列に、0 と 1 の値が入力されており、0 の頻度が 99%など、「閾値」に設定した値以上の場合(デフォルトは 95%)、その列が除外されます。



「閉じる」ボタンをクリックすると、ウィンドウが閉じます。

0.95

アクション

閉じる

[4] 部分構造の検索

指定した SMILES が部分構造に含まれるか否かでデータを検索、サブセットできます。

- 1) 「アドイン」>「Toolkit for Materials Informatics」>「Substructure Searching」を選択します。
- 2) SMILES が入力されている列を「Y,SMILES」に指定します。部分構造の描画で検索したい構造を描画します。SMILES が表示されるので、「Copy」ボタンをクリックして、「クリップボードから貼り付け」ボタンをクリックします。
- 3) 「OK」ボタンをクリックすると、データテーブルに「Substructure: <SMILES>」という列が追加されます。True という値が入力されているのが、部分構造がマッチした行です。「サブセット」ボタンをクリックすると、マッチした行のみがサブセットされます。

The screenshot shows the 'Substructure Searching' interface. On the left, a list of descriptors is visible, with 'smiles' selected. The central drawing area shows a benzene ring structure. Below the drawing, the SMILES string 'C1=CC=CC=C1' is entered, and a 'Copy' button is present. On the right, a 'Preview' table displays a list of compounds and their corresponding ESOL solubility values. The table is as follows:

Compound ID	ESOL solubility	
1	Amigdelin	
2	Fenfuram	
3	cbrel	
4	Picene	
5	Thiophene	
6	benzothiazole	
7	2,2',4,4',6'-PCB	
8	Estradiol	
9	Dieldrin	
10	Rotenone	
11	2-pyrrolidone	
12	2-Chloronaphthalene	
13	1-Pentene	
14	Primidone	
15	Tetradecane	
16	2-Chloropropane	
17	2-Methylbutanol	
18	Benzonitrile	
19	Oxazolin	
20	Undecanol	
21	2,2',3,4,6'-PCB	
22	Lenacil	
23	Phorate	
24	Phenacetin	
25	Dinitramine	
26	1-Heptanol	
27	Theophylline	
28	Butekethal	
29	p,p'-DDE	
30	Methyl octanoate	
31	1,4-Diethylbenzene	
32	Terbufos	
33	Phenmedipham	

使用方法については、以下のページをご覧ください。

[JSME Help \(jsme-editor.github.io\)](https://jsme-editor.github.io)

引用文献

RDKit: Open-Source Cheminformatics Software. [Online] [Cited: 9 7, 2023.]
<https://rdkit.org/>.

Lambard, Guillaume. [Online] [Cited: 9 7, 2023.]
https://github.com/GLambard/Molecules_Dataset_Collection.