

代表的な拡張機能の紹介

—Operator Toolbox—

株式会社KSKアナリティクス



1. Operator Toolboxとは

Operator Toolboxの概要

以前、拡張機能の概要について[ご紹介](#)しましたが、今回からは代表的な拡張機能をご紹介します。

第1回目はOperator Toolboxです。この拡張機能はマーケットプレイスの「よくダウンロードされているもの」の項でも、常に上位に来ている人気の拡張機能です。

「工具箱」という名前の通り、この拡張機能では様々な場面で使える便利なオペレータが収録されています。ここからRapidMinerの標準オペレータに採用されるものも多いので、いち早く使いこなすことも出来ます。

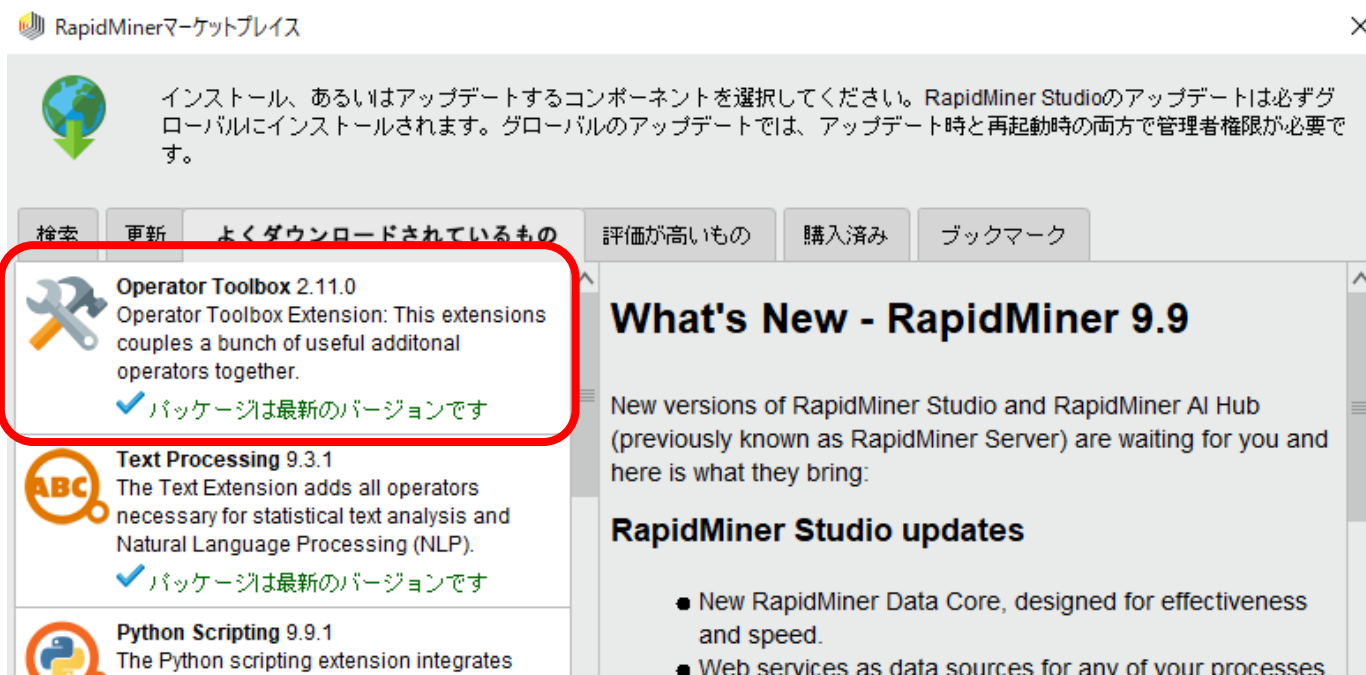
各オペレータにはチュートリアルも付いていますので、使う前にまずはヘルプからチュートリアルプロセスへ飛んで、使い方を試してから使うことをお勧めします。

1. Operator Toolboxとは

Operator Toolboxの準備

以前ご紹介した通り、マーケットプレイスからインストールか、下記WEBページからダウンロードし手動で適用します。

https://marketplace.rapidminer.com/UpdateServer/faces/product_details.xhtml?productId=rmx_operator_toolbox



2. オペレータ紹介（一部抜粋）

Blending

Append (Superset)



- Append (Superset)

このオペレータは標準搭載のAppendとほぼ同じ機能ですが、追加されるExampleSetsの属性が同じでなくとも良く、同じタイプである必要もありません。

Merge



- Merge Attributes

こちらも2つ以上のExampleSetを1つにまとめるオペレータですが、特徴量（縦の列）を行ごとに追加し、各ExampleSetの行は1行目は1行目同士、2行目は2行目同士、といった形で結合されることに注意してください。

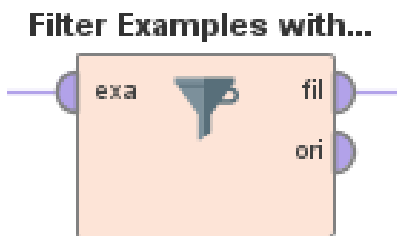
2. オペレータ紹介（一部抜粋）

Blending



- Fuzzy Matching

入力されている2つのExampleSetから1つずつ変数を選び、あいまい検索的な手法で似ている変数を結合させることができます。



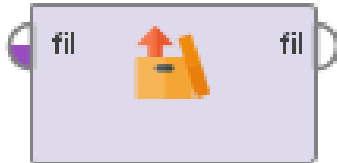
- Filter Examples with Missing Values

標準のFilter Examplesとほぼ同様の機能ですが、こちらは欠損値がある列もフィルタリングをかけることができます。

2. オペレータ紹介（一部抜粋）

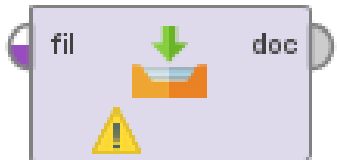
Data Access

Un-gzip



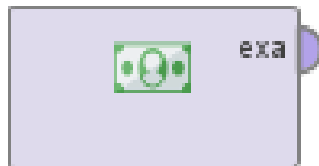
- Un-gzip
gzipで圧縮されたファイルを解凍できます。

Read Office File



- Read Office File
ワードやパワーポイントのファイルを読み込み、ドキュメントファイルに変換します。

Read Yahoo Finance



- Read Yahoo Finance
Yahoo!ファイナンスのデータを読み取ることが出来るオペレータです。

2. オペレータ紹介（一部抜粋）

Macros

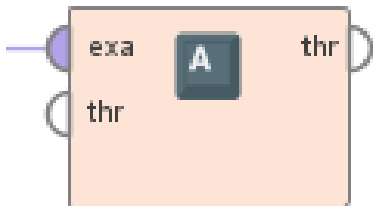
Extract Macro (Form...



- Extract Macro (Format)

標準のExtract Macroと同等の機能を持っていますが、更に日付や数値のフォーマットを指定することも出来ます。

Set Macros from Ex...



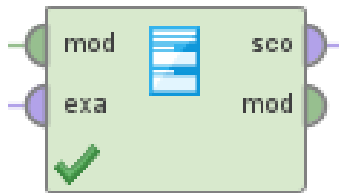
- Set Macros from Example set

マクロ名とマクロ値を含む属性を持つ ExampleSet を入力することができます。このオペレータは、入力されたそれぞれの値でマクロを設定します。

2. オペレーター紹介（一部抜粋）

Models

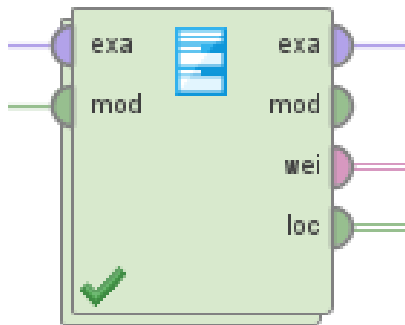
GLM Contribution



- GLM Contribution

GLMモデルの予測に対する個々の変数の貢献度を出力できます。単に予測するだけでなく、結果の説明性を向上させます。

Get Local Interpretation



- Local Interpretation (LIME)

LIMEに近いアルゴリズムを使って、解釈が容易な局所的な特徴の重み（「解釈」）を生成することで、複雑なモデルの決定の「理由」を例ごとに理解するのに役立ちます。

2. オペレータ紹介（一部抜粋）

Outliers/Performance

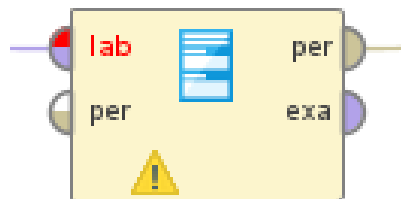
Isolation Forest



- Isolation Forest

Isolation Forestを用いて外れ値検知を行うオペレータです。通常のランダムフォレストと違い、ラベルが必要ありません。

Performance (AUPRC)

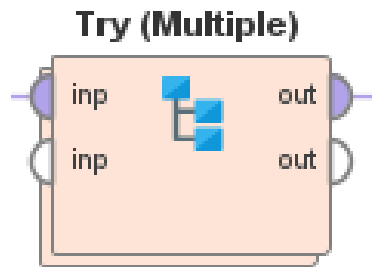


- Performance (AUPRC)

AUCだけでなく、AUPRCも使って精度測定が出来ます。AUCがROC下面積（真陽性率対偽陽性率）を測定するのに対し、AUPRCはPrecision - Recallグラフ下面積を測定します。

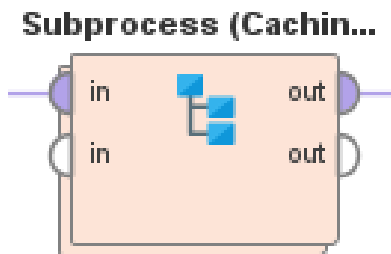
2. オペレータ紹介（一部抜粋）

Utility



- Try (Multiple)

1つのサブプロセスがエラーなしで終了するまで、他のサブプロセスを試行します。入力に対し、様々な処理を試せますので、例えばファイルの種類がわからない場合に、複数のreadオペレータを試すことができます。



- Subprocess (Caching)

このオペレータは、Subprocessオペレータを改良したもので、Subprocessの結果をキャッシュして次の実行時に再利用することができます。

3. まとめ

更なる活用へ向けて

Operator Toolboxには2021年7月現在で約60ものオペレータが収録されておりますので、今回紹介したのはその一部となります。拡張機能ではありますが、ここで実用性が認められたものは標準オペレータに採用されることもあり、標準のオペレータとは別の使い方があるもの、標準よりも進化したものなど、多種多様です。

名前の通り、数多くの道具が入っておりますので、是非まずはインストールしてご自身のプロセスで活用が出来るかお試しください！