

RapidMinerの基本操作

株式会社KSKアナリティクス



RAPIDMINER

はじめに

はじめに

この資料では、自身で分析を進められるように
RapidMinerの基本的な使い方を習得することを目指します。

基本編とTips編に分かれているので、
自身が興味のある範囲を選択してご覧いただけますと幸いです。

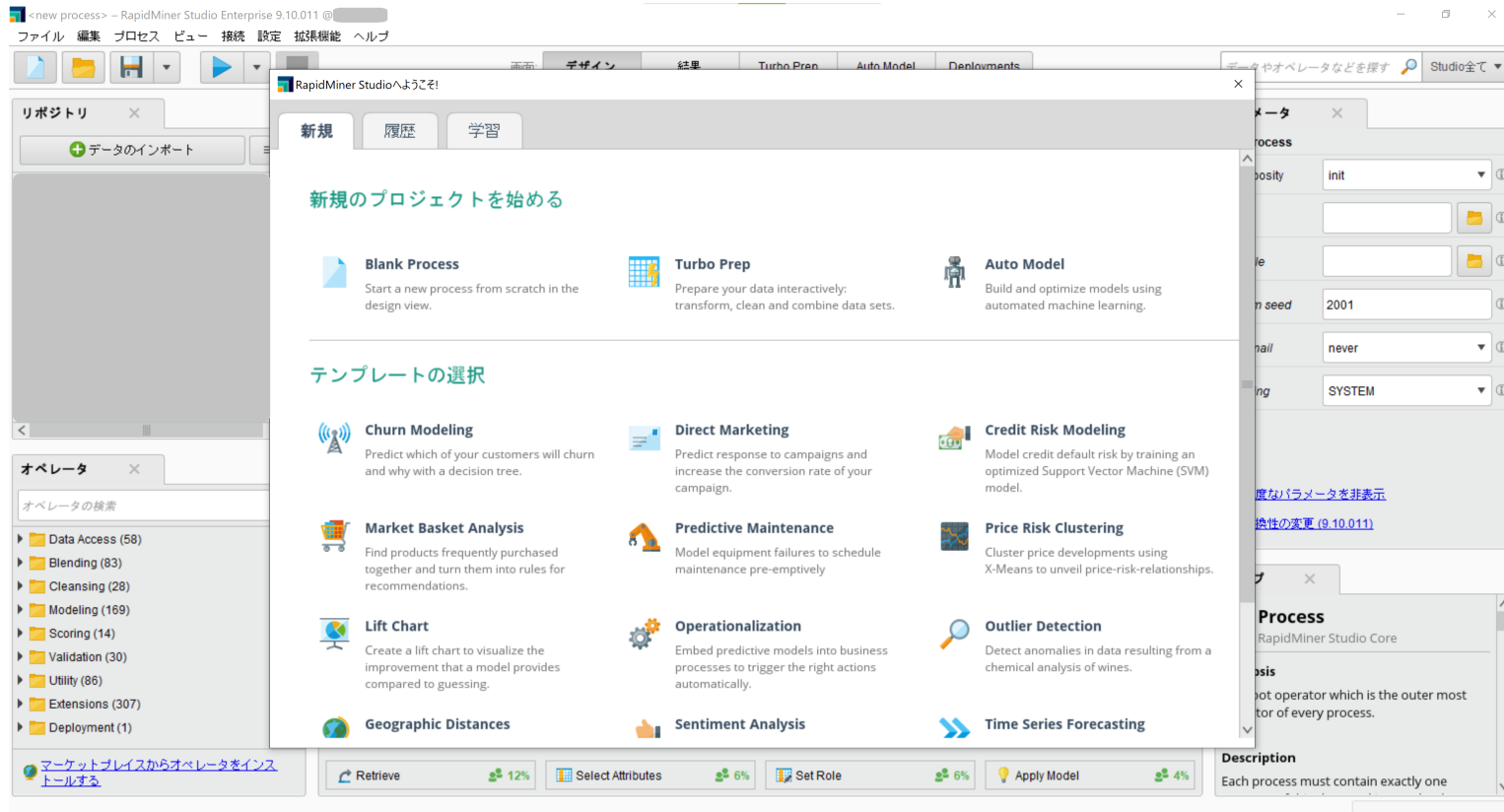
はじめに

目次

はじめに	<u>2</u>
準備	<u>4</u>
基本編	<u>7</u>
Tips	<u>45</u>
トラブルシューティング	<u>67</u>
おわりに	<u>68</u>

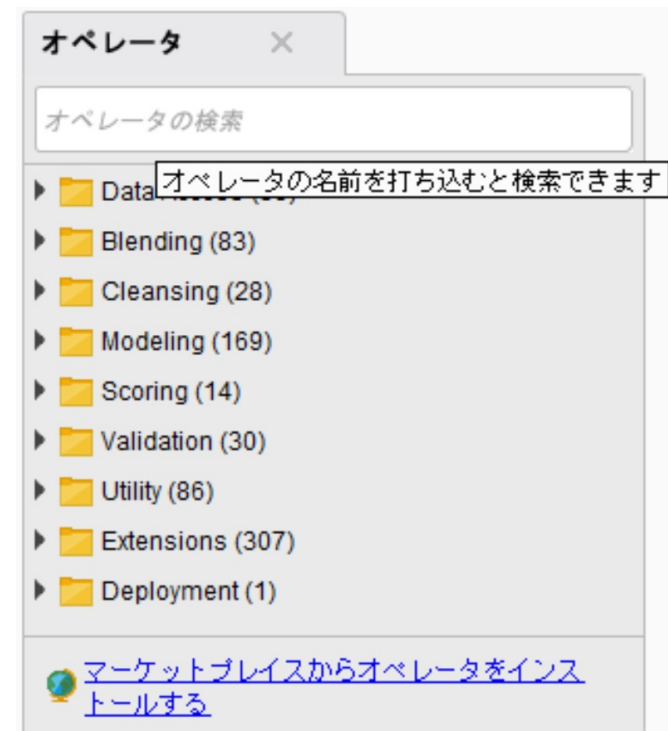
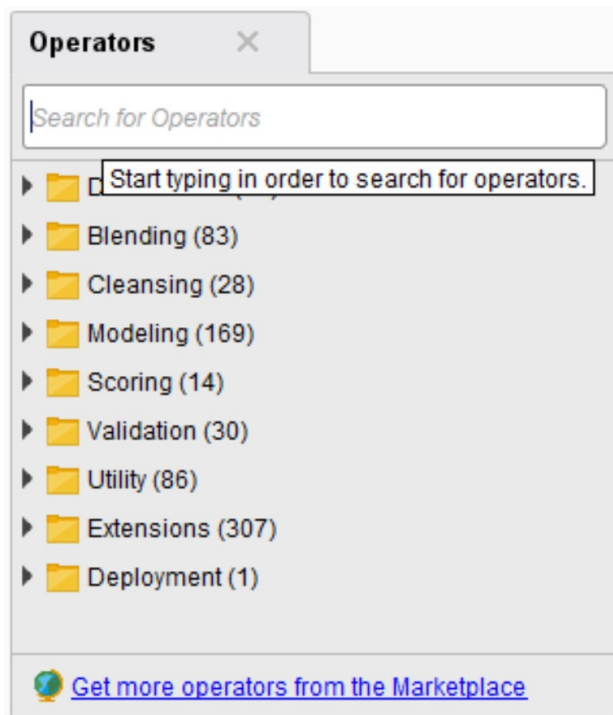
前提

- RapidMiner Studioをインストール済み
- RapidMiner Studioを起動できる



RapidMiner Studioの日本語化

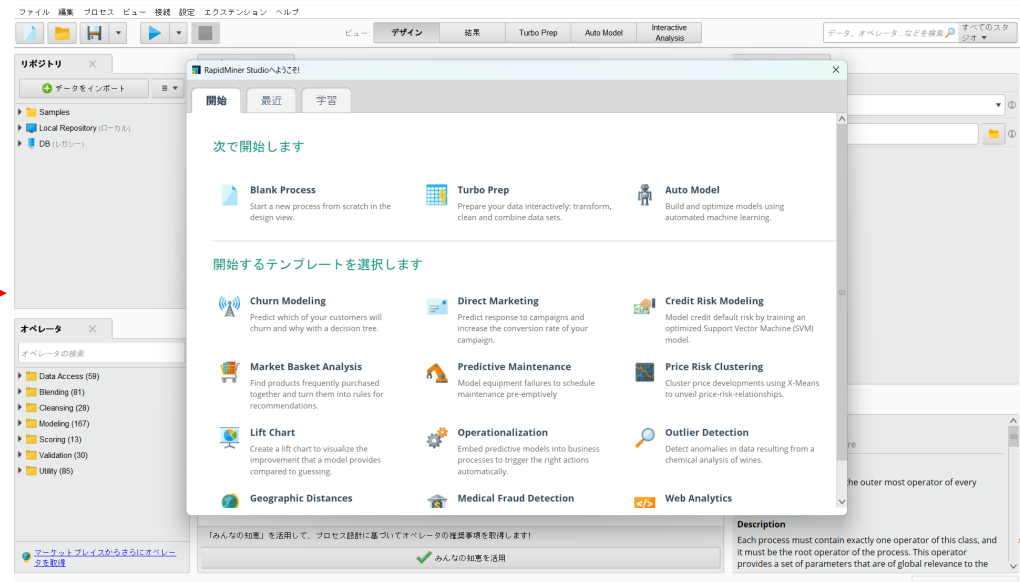
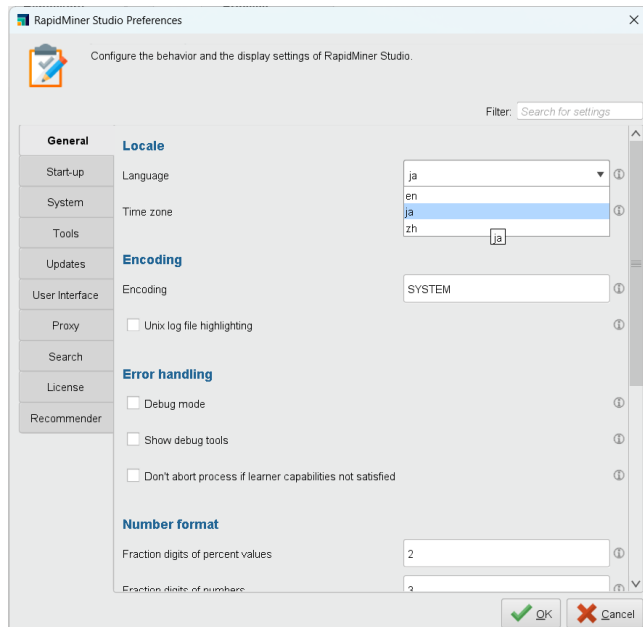
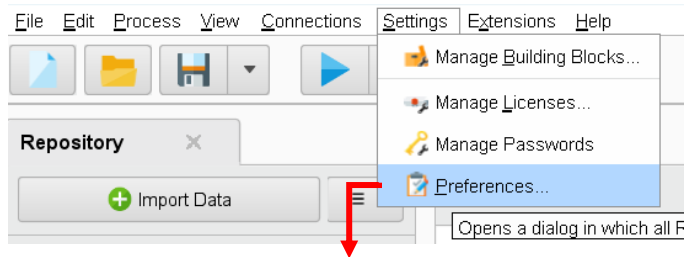
RapidMiner Studioのメニューの一部を日本語にすることができます。



RapidMiner Studioの日本語化（最新Verの場合）

画面上部のメニューの**Settings > Preference > Generalタブ** より
Localeを"ja"に変更してください。

RapidMinerを再起動させると、RapidMinerが日本語化されます。



※Ver10.2より、日本語化ファイルが同梱されるようになりました。

RapidMiner Studioの日本語化 (Ver10.2より前の場合)

RapidMiner Studio Ver10.2より前のバージョンを使用される場合は、**.jarファイルを指定の場所に配置してから** RapidMiner上で日本語表示へ切り替えます。

<手順>

1. 日本語化ファイルの[ダウンロード](#)
2. .jarファイルを以下に配置
¥.RapidMiner¥extensions
例) C:¥Users¥<ユーザー名>¥.RapidMiner¥extensions
3. RapidMiner Studioを起動
4. Settings > Preference… を選択
5. Language を en から ja に変更
6. RapidMiner Studioの再起動

※ファイルの置き場所がわからない場合は、Preferenceの Start-up > Extension directory を参照してください。

スタートページ

The screenshot displays the RapidMiner Studio Enterprise 9.10.011 interface. The main window shows a '新規' (New) dialog box with two sections: '新規のプロジェクトを始める' (Start a new project) and 'テンプレートの選択' (Select a template). The 'Start a new project' section includes 'Blank Process', 'Turbo Prep', and 'Auto Model'. The 'Select a template' section includes 'Churn Modeling', 'Direct Marketing', 'Credit Risk Modeling', 'Market Basket Analysis', 'Predictive Maintenance', 'Price Risk Clustering', 'Lift Chart', 'Operationalization', 'Outlier Detection', 'Geographic Distances', 'Sentiment Analysis', and 'Time Series Forecasting'. A red dashed box highlights these options, with a red arrow pointing to a callout box that says '各機能にジャンプ' (Jump to each function). Another red arrow points to a callout box that says '各テンプレートにジャンプ' (Jump to each template). The interface also features a 'リポジトリ' (Repository) on the left, an 'オペレータ' (Operators) search bar, and a 'Description' panel at the bottom right.

各機能にジャンプ

各テンプレートにジャンプ

スタートページ

The screenshot shows the RapidMiner Studio Enterprise 9.10.011 start page. The main navigation bar includes '新規' (New), '履歴' (History), and '学習' (Learn). The '学習' button is highlighted with a red dashed box and labeled 'クリック' (Click). Below the navigation bar, there are two main sections: 'その他の学習素材' (Other Learning Materials) and 'チュートリアル' (Tutorials). The 'チュートリアル' section is highlighted with a red box and labeled '付属のチュートリアルへ (日本語訳)' (To the attached tutorials (Japanese translation)).

Annotations on the left side of the screenshot:

- '無料の動画教材へ' (To free video lessons) points to 'Explore RapidMiner Academy'.
- '英語ドキュメントへ (日本語はこちら)' (To English documentation (Japanese is here)) points to 'Go to Documentation'.
- 'ユーザーコミュニティへ' (To user community) points to 'Visit Community'.

The 'チュートリアル' section contains a list of topics and a detailed description of the 'Get started' tutorial:

- Get started (1/8)
- Prepare data (0/6)
- Build a model (0/5)
- Collaborate and scale (0/1)
- Use Hadoop (0/1)

The 'Get started' tutorial description includes the following steps:

- 1. Operators and Processes**
Retrieve data and inspect it
- 2. Modeling**
Combine operators to build a statistical model
- 3. Accessing Data**
Import data to the repository, add it to your process, visualize it
- 4. Filtering and Sorting**
Determine the highest passenger fare women paid on the Titanic
- 5. Merging and Grouping**
Join two data sets and aggregate to find the most purchased product
- 6. Creating and Removing Columns**
Calculate total sales based on number of items sold and price. Keep only the interesting columns
- 7. Changing Times and Roles**

スタートページ

Blank Process

Start a new process from scratch in the design view.

テンプレートの選択

Churn Modeling

Predict which of your customers will churn and why with a decision tree.

Direct Marketing

Predict response to campaigns and increase the conversion rate of your campaign.

Credit Risk Modeling

Model credit default risk by training an optimized Support Vector Machine (SVM) model.

Market Basket Analysis

Find products frequently purchased together and turn them into rules for recommendations.

Predictive Maintenance

Model equipment failures to schedule maintenance pre-emptively.

Price Risk Clustering

Cluster price developments using X-Means to unveil price-risk-relationships.

Lift Chart

Create a lift chart to visualize the improvement that a model provides compared to guessing.

Operationalization

Embed predictive models into business processes to trigger the right actions automatically.

Outlier Detection

Detect anomalies in data resulting from a chemical analysis of wines.

Geographic Distances

Sentiment Analysis

Time Series Forecasting

Retrieve 12%

Select Attributes 6%

Set Role 6%

Apply Model 4%

基本UI

The screenshot shows the RapidMiner Studio Enterprise 9.10.011 interface. The main window is titled "プロセス" (Process) and contains a large empty area with the text "メインの作業場所" (Main workspace) and "プロセスが空です。データを追加し、オペレータと接続してください。" (Process is empty. Add data and connect to operators). The interface is annotated with red boxes and labels:

- リポジトリ (Repository):** Located on the left, it contains a "データのインポート" (Import Data) button and the text "データの保管場所" (Data storage location).
- オペレータ (Operators):** Located at the bottom left, it shows a search bar and a list of operator categories: Data Access (58), Blending (83), Cleansing (28), Modeling (169), Scoring (14), Validation (30), Utility (86), Extensions (307), and Deployment (1). A link "マーケットプレイスからオペレータをインストールする" (Install operators from marketplace) is also visible.
- パラメータ (Parameters):** Located on the right, it shows settings for the "Process" operator, including logverbosity (init), logfile, resultfile, random seed (2001), send mail (never), and encoding (SYSTEM). There are also links for "高度なパラメータを非表示" (Hide advanced parameters) and "互換性の変更 (9.10.011)" (Compatibility change).
- ヘルプ (Help):** Located at the bottom right, it shows the "Process" operator help, including a synopsis: "The root operator which is the outer most operator of every process." and a description: "Each process must contain exactly one".

At the bottom of the interface, there is a "Recommended Operators" section with the following items:

- Retrieve (12%)
- Select Attributes (6%)
- Set Role (6%)
- Apply Model (4%)

オペレータの調整

オペレーター一覧

オペレータのヘルプ

プロセス

メインの作業場所

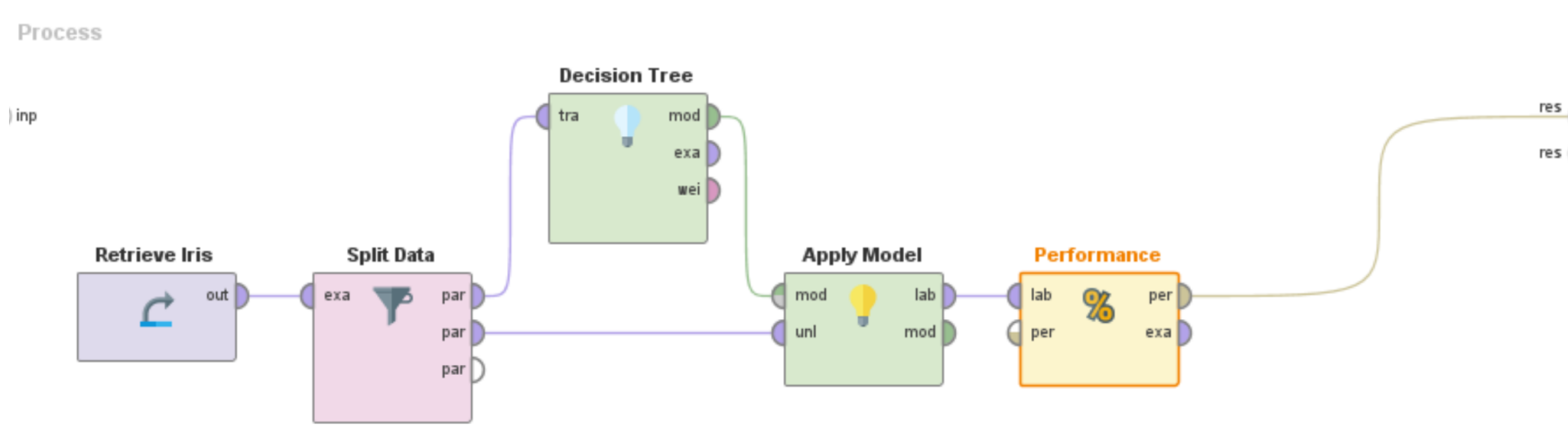
プロセスが空です。
データを追加し、
オペレータと接続してください。

Recommended Operators

- Retrieve 12%
- Select Attributes 6%
- Set Role 6%
- Apply Model 4%

プロセス

プロセス：データの変換や分析を行う、オペレータが繋がったセット
(他製品では、「フロー」「プログラム」「パイプライン」「ダイアグラム」とも)



RapidMinerを使用する目的…**プロセスを完成させて、結果を得ること**

リポジトリ

データの保管場所

プロセスが空です。
データを追加し、
オペレータと接続してください。

Recommended Operators

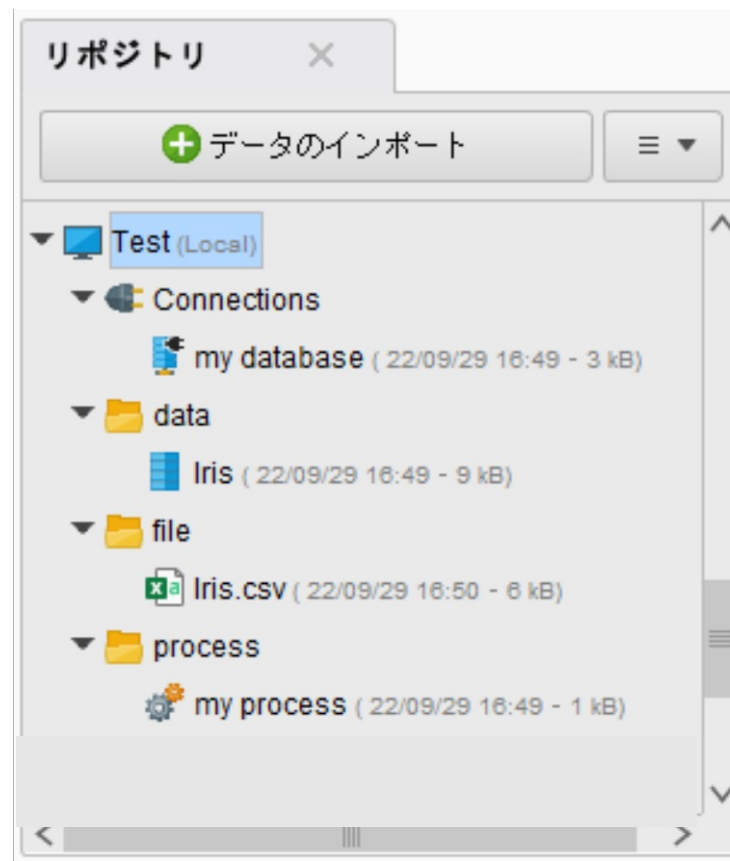
- Retrieve 12%
- Select Attributes 6%
- Set Role 6%
- Apply Model 4%

リポジトリ

リポジトリ：データやプロセスなど、作業内容を保存する場所。

リポジトリでは、以下を保存できます。

- データ
- プロセス
- 接続
- 結果
- 任意のファイル (ver9.7以降)



オペレータ

The screenshot shows the RapidMiner Studio Enterprise 9.10.011 interface. The main canvas is empty, displaying the text: "プロセスが空です。データを追加し、オペレータと接続してください。" (The process is empty. Add data and connect to operators.)

The left sidebar contains the "リポジトリ" (Repository) and "オペレータ" (Operators) panels. The "オペレータ" panel is highlighted with a red box and shows a list of operator categories:

- Data Access (58)
- Blending (83)
- Cleansing (28)
- Modeling (169)
- Scoring (14)
- Validation (30)
- Utility (86)
- Extensions (307)
- Deployment (1)

Below the list is a link: "マーケットプレイスからオペレータをインストールする" (Install operators from the marketplace).

The right sidebar shows the "パラメータ" (Parameters) panel for the "Process" operator, with settings for logverbosity, logfile, resultfile, random seed, send mail, and encoding. Below it is the "ヘルプ" (Help) panel for the "Process" operator, which includes a synopsis and description.

The bottom of the interface shows a "Recommended Operators" section with suggestions: Retrieve (12%), Select Attributes (6%), Set Role (6%), and Apply Model (4%).

オペレーター一覧

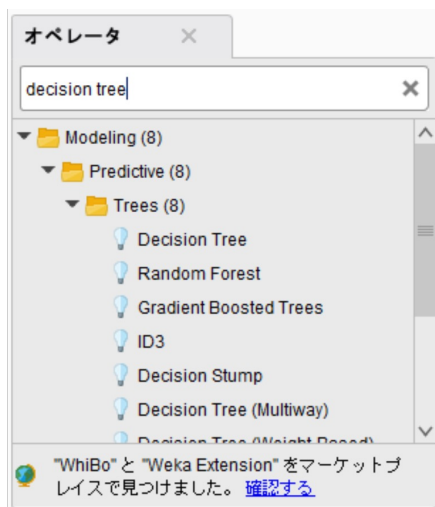
オペレータ

オペレータ：様々な動きをするオペレータの格納場所。

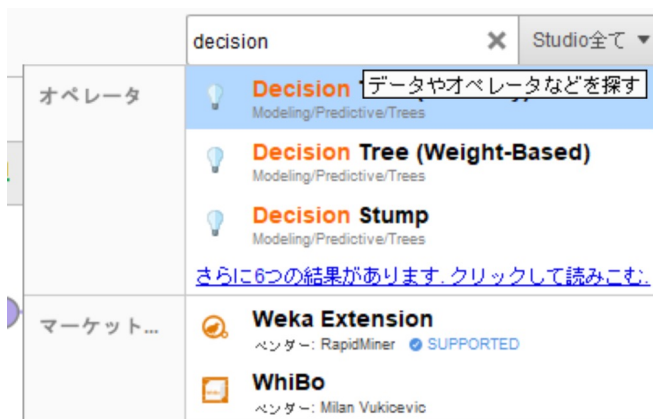


オペレータはプロセスの要素で、パラメータの選択に応じて動作が変わります。RapidMinerには数百ものオペレータがあるので、必要なオペレータを見つけることが作業の大部分を占めています。

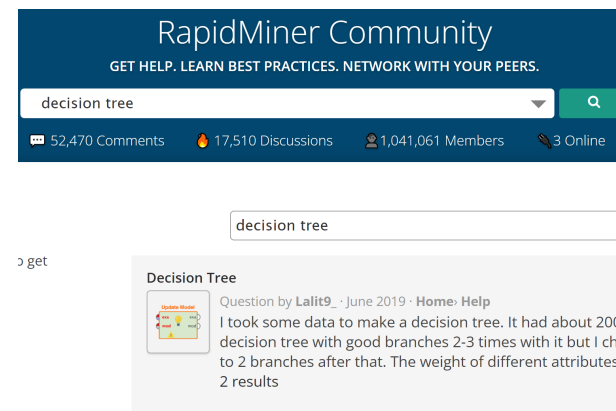
検索は、検索窓の使用や、グローバル検索、コミュニティなどを探してみてください。



オペレータの検索窓



グローバル検索



RapidMiner Community

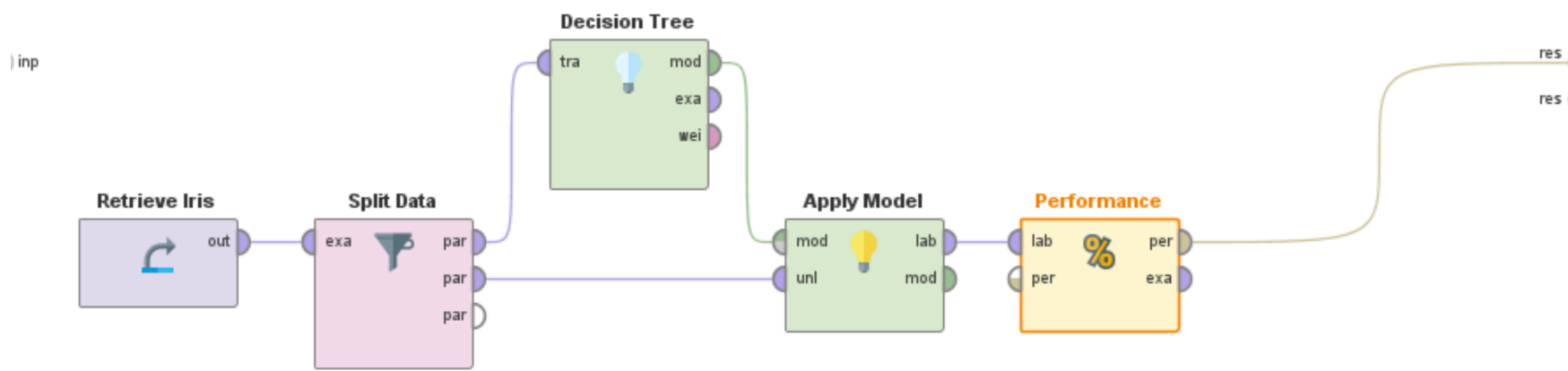
ポート

オペレータには**ポート**が付いています。
ポートを介してオペレータを繋ぎ、プロセスを構築します。

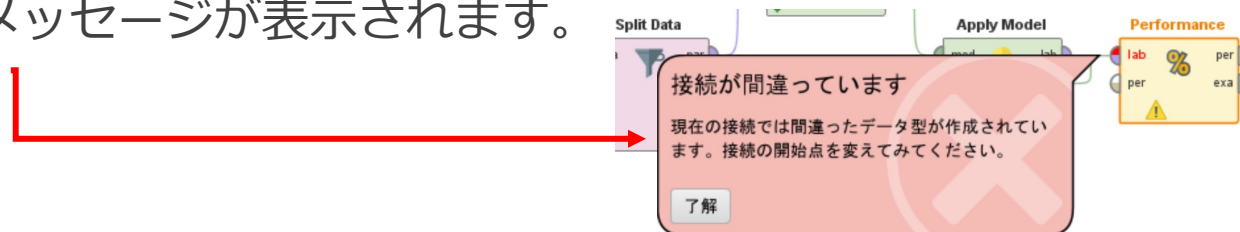
Retrieve Iris



Process



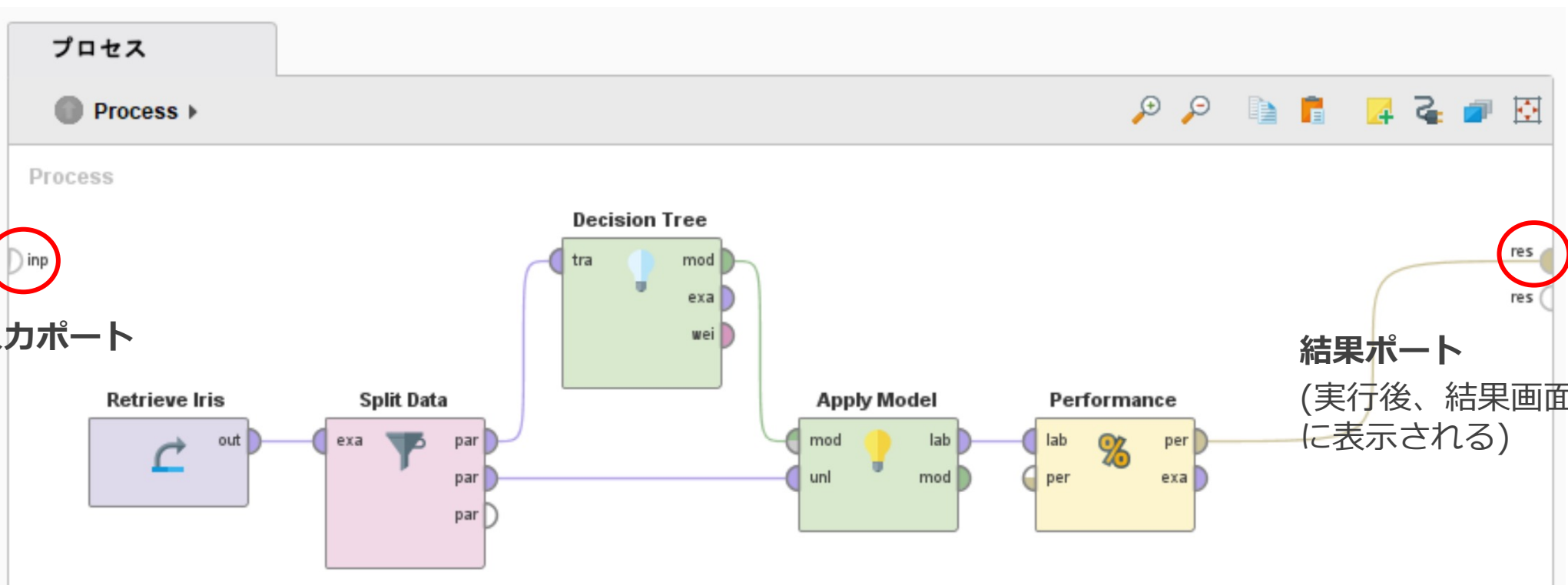
2つのオペレータを接続する場合は、それぞれのポートに**互換性**が必要です。
互換性がない場合は、エラーメッセージが表示されます。



基本編

ポート

端にあるポートにも、名前が付いています。



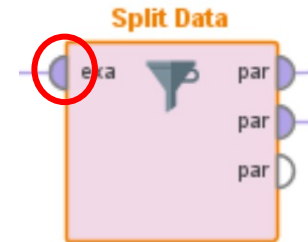
ポート (主なもの)

- **ExampleSet系** (紫色)

ExampleSetを意味するポート

(exa, tra, unl, ori, lab, lef, rig, joi, merなど)

オペレータの例: Split Data, Decision Tree, Apply Model, Performanceなど



- **Model** (緑色)

この出力ポートからのデフォルトモデル。機械学習モデルなど。

オペレータの例: Decision Tree, Linear Regression, Apply Modelなど

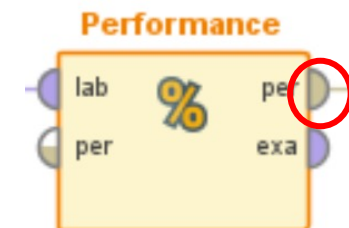


- **Performance** (黄色)

パフォーマンスベクトル。評価指標などパフォーマンスに関する情報が含まれます。

オペレータの例: Performance, Cross Validation, Performance to Dataなど

※ポートの一覧は[ドキュメント](#)をご覧ください



パラメータ

オペレータの調整

The screenshot displays the RapidMiner Studio Enterprise 9.10.011 interface. The main workspace shows an empty process with the text: "プロセスが空です。データを追加し、オペレータと接続してください。" (The process is empty. Add data and connect to operators.)

The **パラメータ** (Parameters) panel on the right is highlighted with a red box and contains the following settings:

- logverbosity: init
- logfile: [empty field]
- resultfile: [empty field]
- random seed: 2001
- send mail: never
- encoding: SYSTEM

Below the parameters, there are two status indicators: "高度なパラメータを非表示" (Hide advanced parameters) and "互換性の変更 (9.10.011)" (Compatibility change (9.10.011)).

The **ヘルプ** (Help) panel at the bottom right shows the **Process** operator details:

- Synopsis:** The root operator which is the outer most operator of every process.
- Description:** Each process must contain exactly one

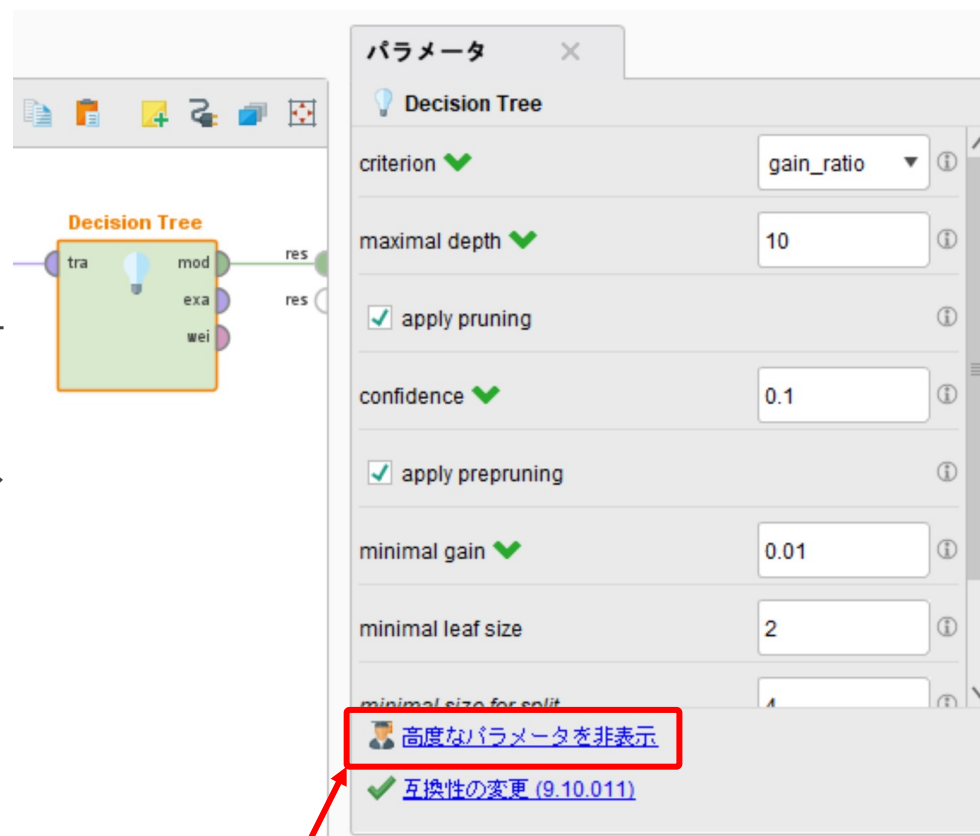
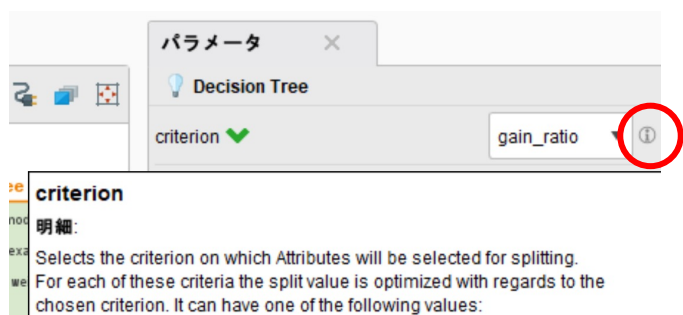
The **オペレータ** (Operators) panel on the left lists various operator categories such as Data Access (58), Blending (83), Cleansing (28), Modeling (169), Scoring (14), Validation (30), Utility (86), Extensions (307), and Deployment (1). The **Recommended Operators** section at the bottom lists: Retrieve (12%), Select Attributes (6%), Set Role (6%), and Apply Model (4%).

パラメータ

パラメータ：オペレータの動作を設定

パラメータの内容は、オペレータによって変わります。

動作の詳細は、パラメータ横のインフォメーションか、ヘルプを確認してください。



「高度なパラメータを表示/非表示」ですべてのパラメータを表示できます。
(高度なパラメータは斜体になっています。)

基本編

ヘルプ

ヘルプ

Process

プロセスが空です。
データを追加し、
オペレータと接続してください。

ヘルプ

Process
RapidMiner Studio Core

Synopsis
The root operator which is the outer most operator of every process.

Description
Each process must contain exactly one

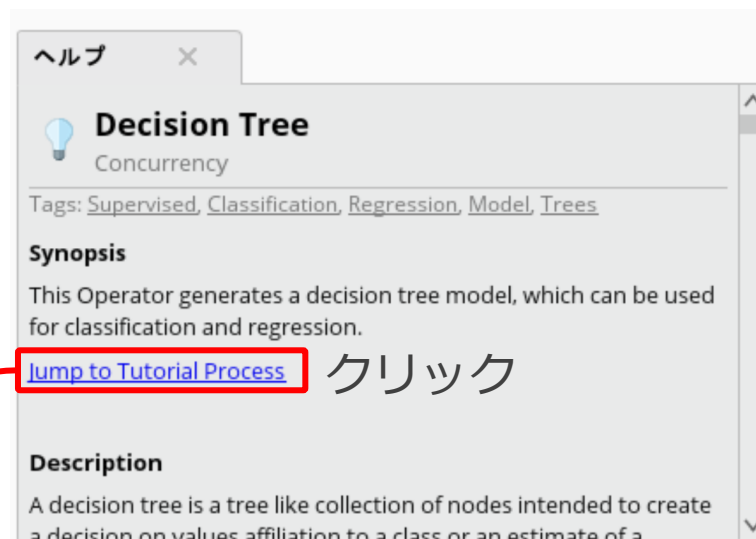
オペレータのヘルプ

ヘルプ

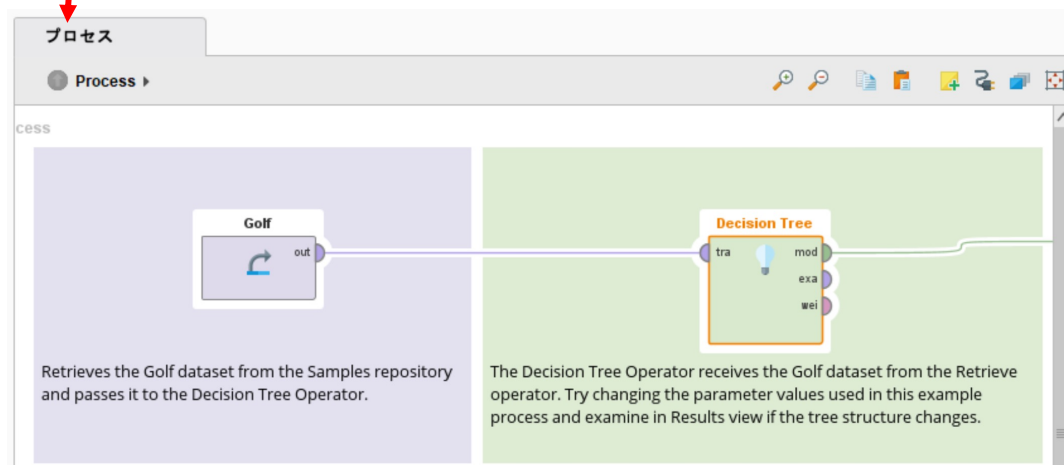
ヘルプ：選択したオペレータのヘルプ、説明

ヘルプの内容も、オペレータによって変わります。
ヘルプには、以下の内容が掲載されています。

- オペレータの概要、目的
- オペレータのポートについて
- パラメータについて
- チュートリアル

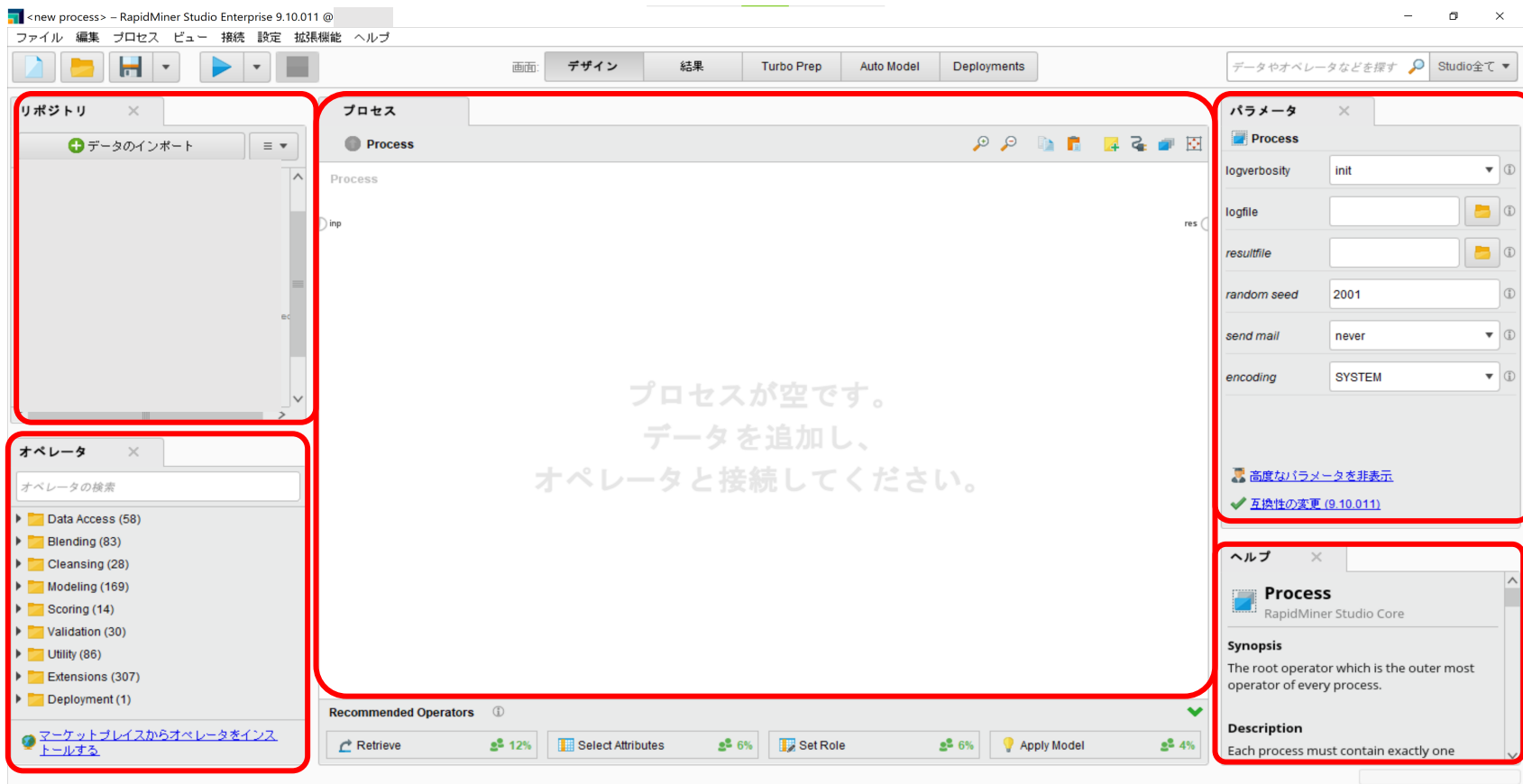


開く



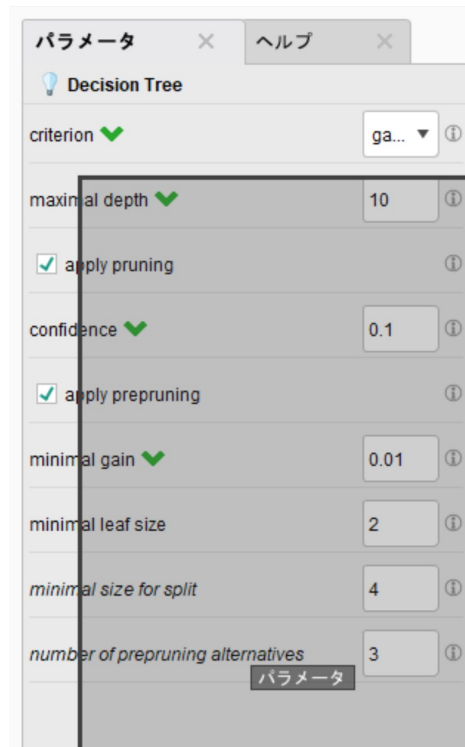
パネル

見てきたこれら全てを、**パネル**と呼びます。



パネル

各パネルは、ドラッグ&ドロップで自由に動かすことができます。
また右上のバツ印をクリックすると、パネルを閉じることができます。



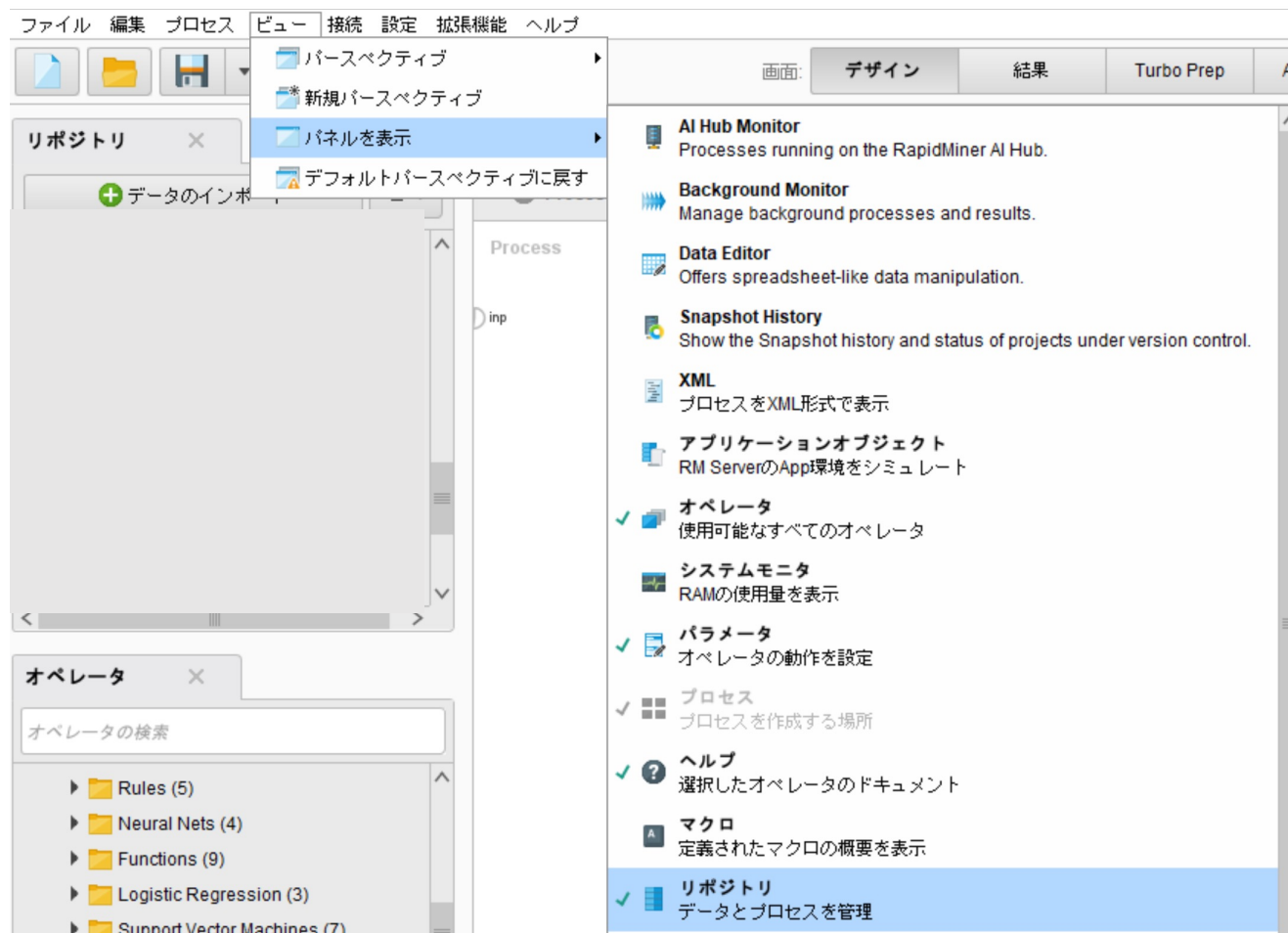
ドラッグ&ドロップで自由に配置



バツ印でパネルを閉じる

パネルの追加

ビュー > パネルを表示 より、パネルを追加することができます。



パネル一覧 1/2

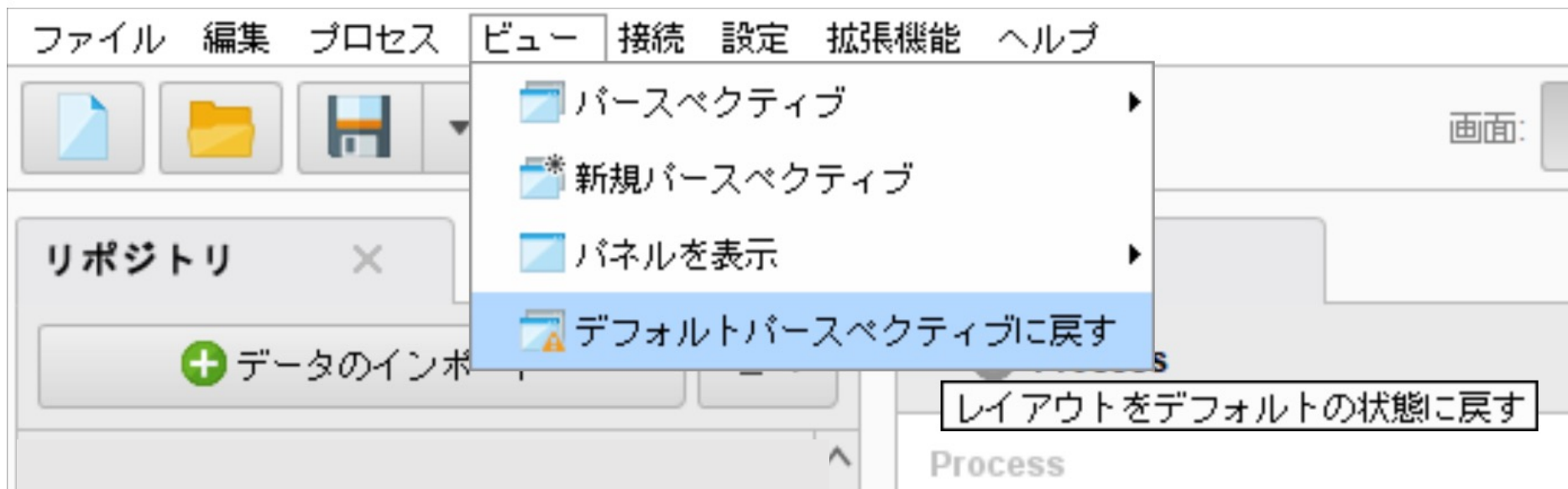
パネル (太字はデフォルトであるもの)	概要
アプリケーションオブジェクト	RapidMiner AI Hub Appをシミュレート
Background Monitor	バックグラウンドプロセスと結果の管理
環境	プロセスとマクロの高度な設定
Data editor	スプレッドシートのようなデータ操作を提供
ヘルプ	選択したオペレータの概要を表示
ログ	記録したイベントを表示
マクロ	定義されたマクロの概要を表示
オペレータ	プロセスに追加できるすべてのオペレータを表示
概要	大きなプロセスを縮小して表示
パラメータ	プロセス内のオペレータの動作を設定
問題	プロセス内の潜在的な問題を表示
プロセス	プロセスの作成およびデザイン
リポジトリ	データとプロセスを管理

パネル一覧 2/2

パネル	概要
システムモニタ	RAMの使用量を表示
履歴	このセッションのすべてのプロセス結果の履歴を表示
AI Hub Monitor	RapidMiner AI Hub上で実行しているプロセスを表示
Snapshot History	バージョン管理下のプロジェクトのスナップショット履歴とステータスを表示
階層構造	プロセスの階層構造を表示
XML	プロセスをXML形式で表示

パネルをデフォルトの状態に戻す

パネルをデフォルトの状態に戻したい場合は、**ビュー > デフォルトパースペクティブに戻す** をクリックします。



画面

今まで見てきたものが、「デザイン画面」です。
画面は中央上部で切り替えることができます。

画面: デザイン 結果 Turbo Prep Auto Model Deployments

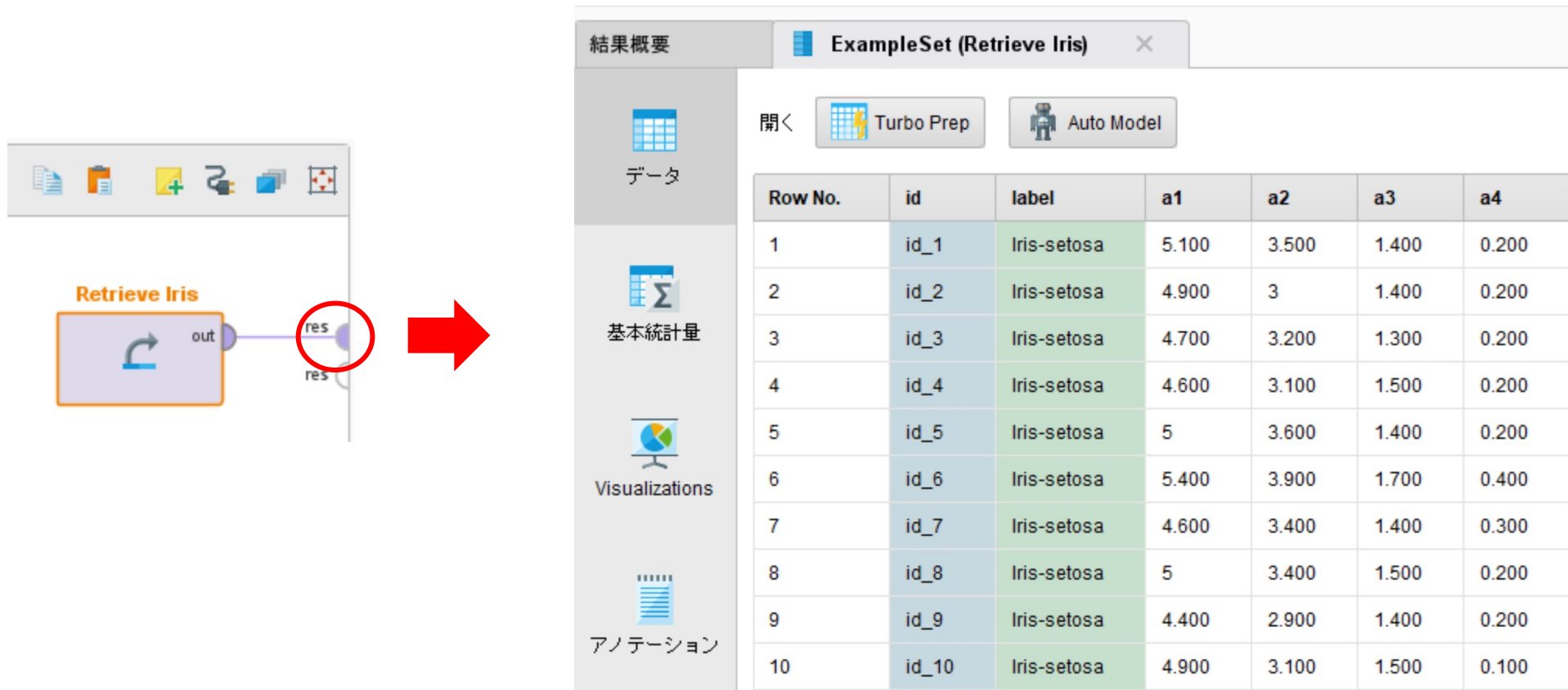
プロセスが空です。
データを追加し、
オペレータと接続してください。

Recommended Operators

- Retrieve 12%
- Select Attributes 6%
- Set Role 6%
- Apply Model 4%

結果画面

プロセスを実行すると、自動で「結果画面」に切り替わります。
結果ポートに繋がったものが、「結果画面」に表示されます。



The screenshot illustrates the workflow from a process to its results. On the left, a 'Retrieve Iris' process is shown with an 'out' port connected to a 'res' port. A red circle highlights the 'res' port, and a red arrow points to the right, indicating the transition to the results view.

The results view, titled 'ExampleSet (Retrieve Iris)', displays a table of data. The table has columns for 'Row No.', 'id', 'label', 'a1', 'a2', 'a3', and 'a4'. The data is as follows:

Row No.	id	label	a1	a2	a3	a4
1	id_1	Iris-setosa	5.100	3.500	1.400	0.200
2	id_2	Iris-setosa	4.900	3	1.400	0.200
3	id_3	Iris-setosa	4.700	3.200	1.300	0.200
4	id_4	Iris-setosa	4.600	3.100	1.500	0.200
5	id_5	Iris-setosa	5	3.600	1.400	0.200
6	id_6	Iris-setosa	5.400	3.900	1.700	0.400
7	id_7	Iris-setosa	4.600	3.400	1.400	0.300
8	id_8	Iris-setosa	5	3.400	1.500	0.200
9	id_9	Iris-setosa	4.400	2.900	1.400	0.200
10	id_10	Iris-setosa	4.900	3.100	1.500	0.100

結果画面の用語

結果概要 ExampleSet (Retrieve Iris) ×

開く Turbo Prep Auto Model

Attribute (属性) = 列

Row No.	id	label	a1	a2	a3	a4
1	id_1	Iris-setosa	5.100	3.500	1.400	0.200
2	id_2	Iris-setosa	4.900	3	1.400	0.200
3	id_3	Iris-setosa	4.700	3.200	1.300	0.200
4	id_4	Iris-setosa	4.600	3.100	1.500	0.200
5	id_5	Iris-setosa	5	3.600	1.400	0.200
6	id_6	Iris-setosa	5.400	3.900	1.700	0.400
7	id_7	Iris-setosa	4.600	3.400	1.400	0.300
8	id_8	Iris-setosa	5	3.400	1.500	0.200
9	id_9	Iris-setosa	4.400	2.900	1.400	0.200
10	id_10	Iris-setosa	4.900	3.100	1.500	0.100

Example
= 行

結果画面の用語

結果概要 ExampleSet (Retrieve Iris) ×

開く Turbo Prep Auto Model

ExampleSet = データセット

Row No.	id	label	a1	a2	a3	a4
1	id_1	Iris-setosa	5.100	3.500	1.400	0.200
2	id_2	Iris-setosa	4.900	3	1.400	0.200
3	id_3	Iris-setosa	4.700	3.200	1.300	0.200
4	id_4	Iris-setosa	4.600	3.100	1.500	0.200
5	id_5	Iris-setosa	5	3.600	1.400	0.200
6	id_6	Iris-setosa	5.400	3.900	1.700	0.400
7	id_7	Iris-setosa	4.600	3.400	1.400	0.300
8	id_8	Iris-setosa	5	3.400	1.500	0.200
9	id_9	Iris-setosa	4.400	2.900	1.400	0.200
10	id_10	Iris-setosa	4.900	3.100	1.500	0.100

結果画面の用語

結果概要 ExampleSet (Retrieve Iris) ×

開く Turbo Prep Auto Model

Row No.	id	label	a1	a2	a3	a4
1	id_1	Iris-setosa	5.100	3.500	1.400	0.200
2	id_2	Iris-setosa	4.900	3	1.400	0.200
3	id_3	Iris-setosa	4.700	3.200	1.300	0.200
4	id_4	Iris-setosa	4.600	3.100	1.500	0.200
5	id_5	Iris-setosa	5	3.600	1.400	0.200
6	id_6	Iris-setosa	5.400	3.900	1.700	0.400
7	id_7	Iris-setosa	4.600	3.400	1.400	0.300
8	id_8	Iris-setosa	5	3.400	1.500	0.200
9	id_9	Iris-setosa	4.400	2.900	1.400	0.200
10	id_10	Iris-setosa	4.900	3.100	1.500	0.100

特別属性 (特別な意味を持つ属性)

結果画面の用語

結果概要 ExampleSet (Retrieve Iris) ×

開く Turbo Prep Auto Model

Row No.	id	label	a1	a2	a3	a4
1	id_1	Iris-setosa	5.100	3.500	1.400	0.200
2	id_2	Iris-setosa	4.900	3	1.400	0.200
3	id_3	Iris-setosa	4.700	3.200	1.300	0.200
4	id_4	Iris-setosa	4.600	3.100	1.500	0.200
5	id_5	Iris-setosa	5	3.600	1.400	0.200
6	id_6	Iris-setosa	5.400	3.900	1.700	0.400
7	id_7	Iris-setosa	4.600	3.400	1.400	0.300
8	id_8	Iris-setosa	5	3.400	1.500	0.200
9	id_9	Iris-setosa	4.400	2.900	1.400	0.200
10	id_10	Iris-setosa	4.900	3.100	1.500	0.100

通常属性 (学習に使用)

ルール

ルール：属性にRapidMinerでの特別な意味や扱いを伝える
RapidMinerには事前に定義されているルールがいくつかあります (例: Id, label)。
それら以外にも、独自で新たにルールを作成することも可能です。

ルールが割り当てられると、属性の背景に色がつきます。
属性に割り当てられているルールは、「基本統計量」タブで確認可能です。

The diagram illustrates how rules are applied to attributes. On the left, two attribute cards are shown: a blue card for 'id' and a green card for 'label'. Red boxes highlight the rule names 'Id' and 'Label' on these cards. On the right, a screenshot of the 'ExampleSet (Retrieve Iris)' window shows the 'Basic Statistics' tab. The table below shows the attributes and their assigned rules, with the 'id' and 'label' rows highlighted in blue and green respectively, matching the attribute cards.

属性名	データ型	欠損値
id	Nominal	0
label	Nominal	0
a1	Real	0
a2	Real	0

ルールについて

ルールは、**Set Role**オペレータで付与・調整できます。

ルールの種類については、Set Roleオペレータのヘルプをご覧ください。

The image shows a data workflow interface. On the left, a 'Retrieve Titanic' operator (purple box) is connected to a 'Set Role' operator (orange box). The 'Retrieve Titanic' operator has an 'out' port. The 'Set Role' operator has 'exa' ports on both sides and an 'ori' port at the bottom. The output of the 'Set Role' operator is connected to a 'res' port. On the right, a 'パラメータ' (Parameters) dialog box is open for the 'Set Role' operator. It contains the following settings:

- attribute name**: Survived
- target role**: label
- set additional roles**: Edit List (0)...

ロールの種類 (主なもの)

- **Regular (白色)**

特別な属性を持たない属性。学習に使用されます。

- **Id (青色)**

行を識別させたいときに付ける属性。すべての行がユニークである必要があります。学習には使用されず、Joinなどデータセットを結合をさせたいときにもよく用いられます。

- **Label (緑色)**

学習への目的変数(ターゲット属性)に使用する属性。
RapidMinerでは、このロールが付いたものを予測します。

- **Prediction (緑色)**

データセットを学習モデルに適用した結果、得た予測に付くロール。
モデルの評価を行う際は、LabelロールとPredictionロールの属性が必要です。

基本統計量

「基本統計量」タブに移動すると、データ型や欠損値の数、最小値など基本的な統計量を一覧で確認できます。

属性名	データ型	欠損値	基本統計量	フィルタ (6/6 属性):
id	Nominal	0	<p>最小頻度値: id_99 (1) 最頻値: id_1 (1)</p>	項目値: id_1 (1), id_10 (1), id_100 (1), id_101 (1), ...[146 さらに] 詳細...
label	Nominal	0	<p>最小頻度値: Iris-virginica (50) 最頻値: Iris-setosa (50)</p>	項目値: Iris-setosa (50), Iris-versicolor (50), Iris-virginica (50) 詳細...
a1	Real	0	<p>最小値: 4.300 最大値: 7.900 平均値: 5.843 標準偏差: 0.828</p>	
a2	Real	0	最小値: 2 最大値: 4.400	平均値: 3.054
a3	Real	0	最小値: 1 最大値: 6.900	平均値: 3.759
a4	Real	0	最小値: 0.100 最大値: 2.500	平均値: 1.199

データ型

結果概要 ExampleSet (Retrieve Iris) x

属性名 データ型 欠損値 基本統計量 フィルタ (6/6 属性): 属性の検索

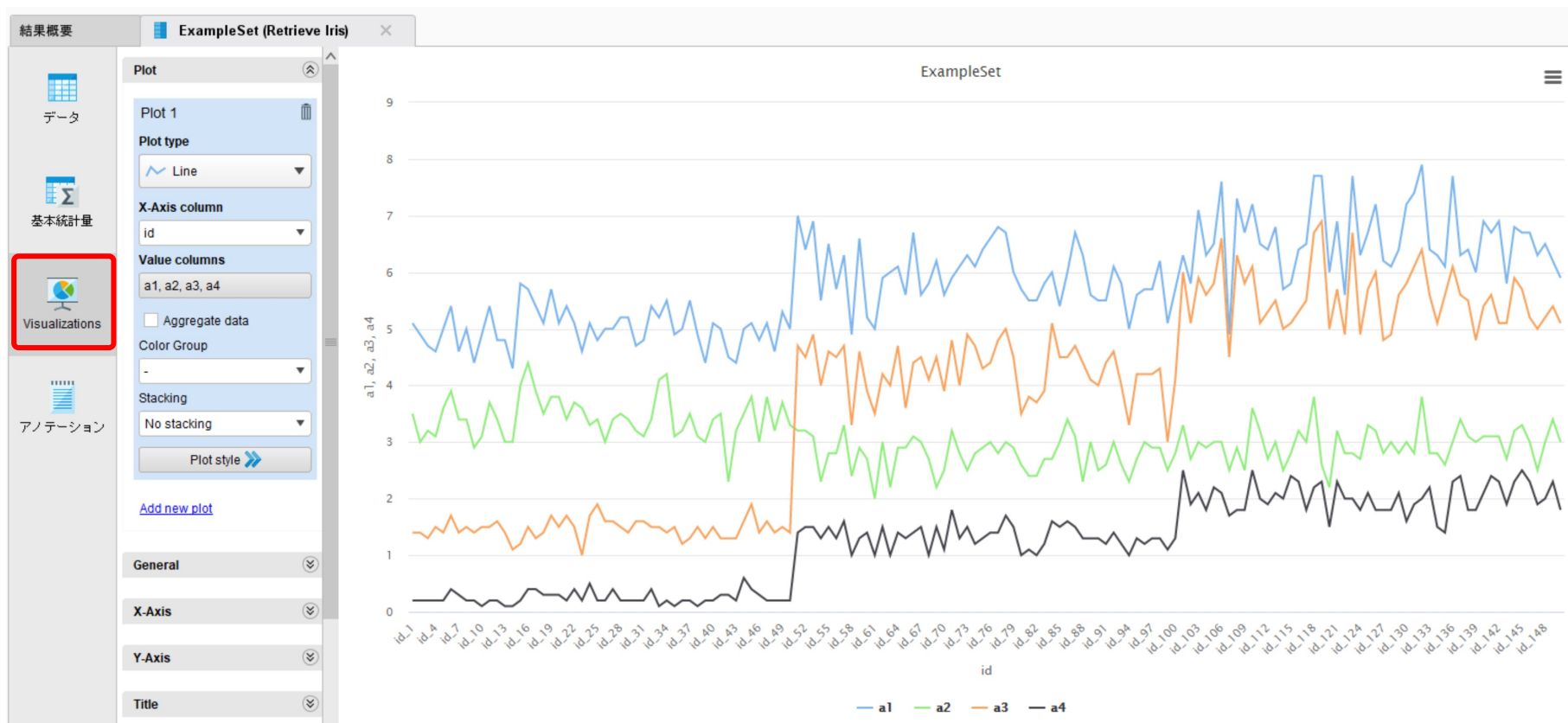
属性名	データ型	欠損値	基本統計量	項目値		
id id	Nominal	0	 <p>最小頻度値 id_99 (1)</p> <p>最頻値 id_1 (1)</p> <p>チャートの表示</p>	id_1 (1), id_10 (1), id_100 (1), id_101 (1), ...[146 さらに] 詳細...		
Label label	Nominal	0	 <p>最小頻度値 Iris-virginica (50)</p> <p>最頻値 Iris-setosa (50)</p> <p>チャートの表示</p>	Iris-setosa (50), Iris-versicolor (50), Iris-virginica (50) 詳細...		
a1	Real	0	 <p>最小値 4.300</p> <p>最大値 7.900</p> <p>平均値 5.843</p> <p>標準偏差 0.828</p> <p>チャートの表示</p>			
a2	Real	0	最小値 2	最大値 4.400	平均値 3.054	
a3	Real	0	最小値 1	最大値 6.900	平均値 3.759	
a4	Real	0	最小値 0.100	最大値 2.500	平均値 1.199	

データ型一覧

系列	データ型	概要
すべて	attribute	すべての可能なタイプ (「すべてのタイプ」) の親
日時	date_time	日付と時間の両方 (例 : 23.12.2014 17:59)
日時	date	時間のない日付 (例 : 23.12.2014)
日時	time	日付のない時間 (例 : 17:59)
数値	numeric	あらゆる種類の数値 (date、time、integer、real を含む)
数値	integer	整数 (例 : 23、-5、11,024,768)
数値	real	小数 (例 : 11.23や-0.0001など)
テキスト	nominal	あらゆる種類のテキスト値 (polynomialとbinomialを含む)
テキスト	binominal	2つの値のみ (例 : true/falseまたはyes/no)
テキスト	polynominal	多くの異なる文字列の値 (例 : 赤、緑、青、黄など)
テキスト	text	(polynomialと区別するために) より詳細な区別を可能にする nominalデータ型
テキスト	file_path	より詳細な区別を可能にする nominalデータ型です (まれに使用されます)。列を「ファイルパスのみを含む」としてマークするために使用できます。

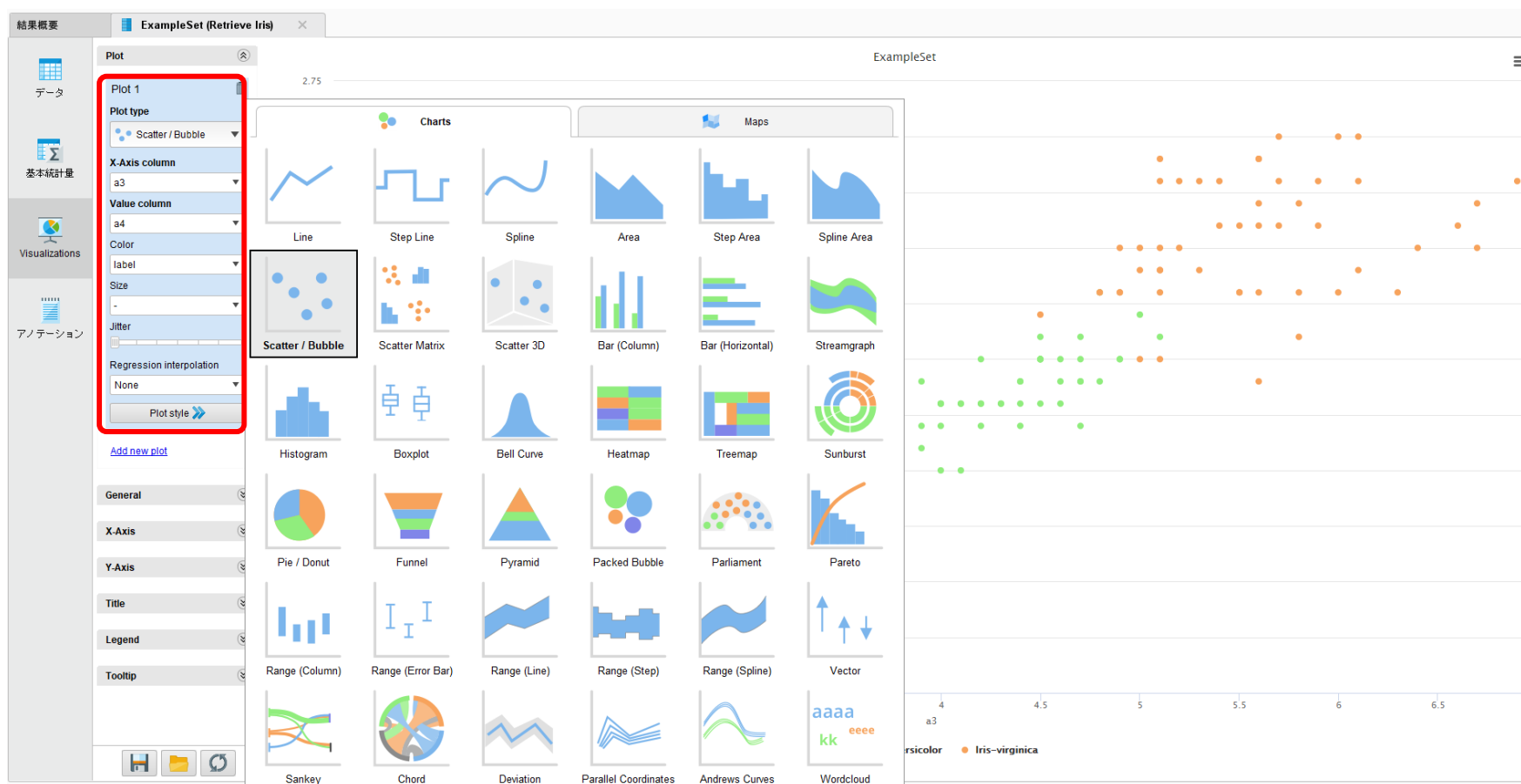
Visualization

Visualizationタブをクリックすると、データを可視化することができます。




Visualization

グラフは左の部分で設定でき、RapidMinerには様々なグラフが用意されています。データを可視化させて、データの理解を進めていきます。



基本の流れ

1. **リポジトリパネル**からデータを取得
↓
2. **オペレータパネル**からオペレータの追加
↓
3. **パラメータパネル**でパラメータの調整 (**ヘルプパネル**を参照しながら)
↓
4. 手順2と3を繰り返す
↓
5. プロセスの実行 
↓
6. **結果画面**で結果の確認
↓
7. 手順1,2,3のいずれかに戻る

Tips

先ほどまで、RapidMinerの基本操作を見てきました。

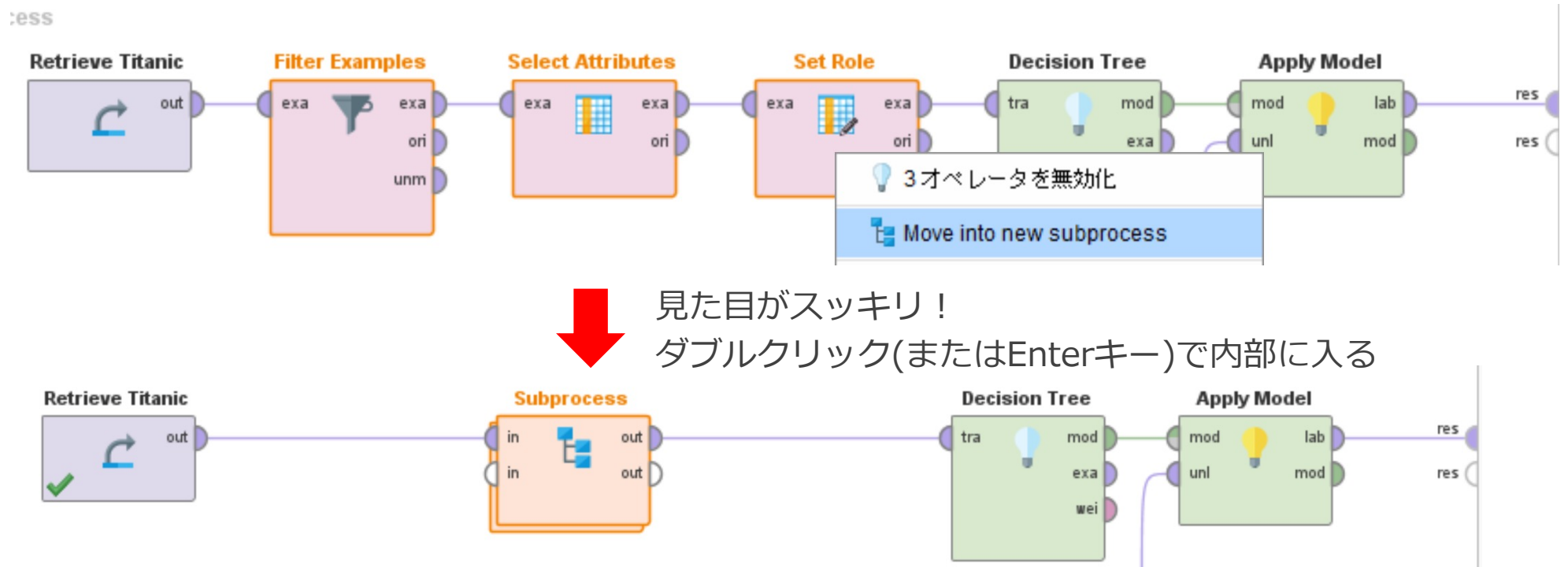
ここからは、実際にRapidMinerを使用する際のTipsについて記載していきます。

- サブプロセス
- メモ
- マクロ
- マーケットプレイス
- オペレータの繋ぎ方のコツ 4選
- オペレータの実行順
- リポジトリの実データの場所

サブプロセス

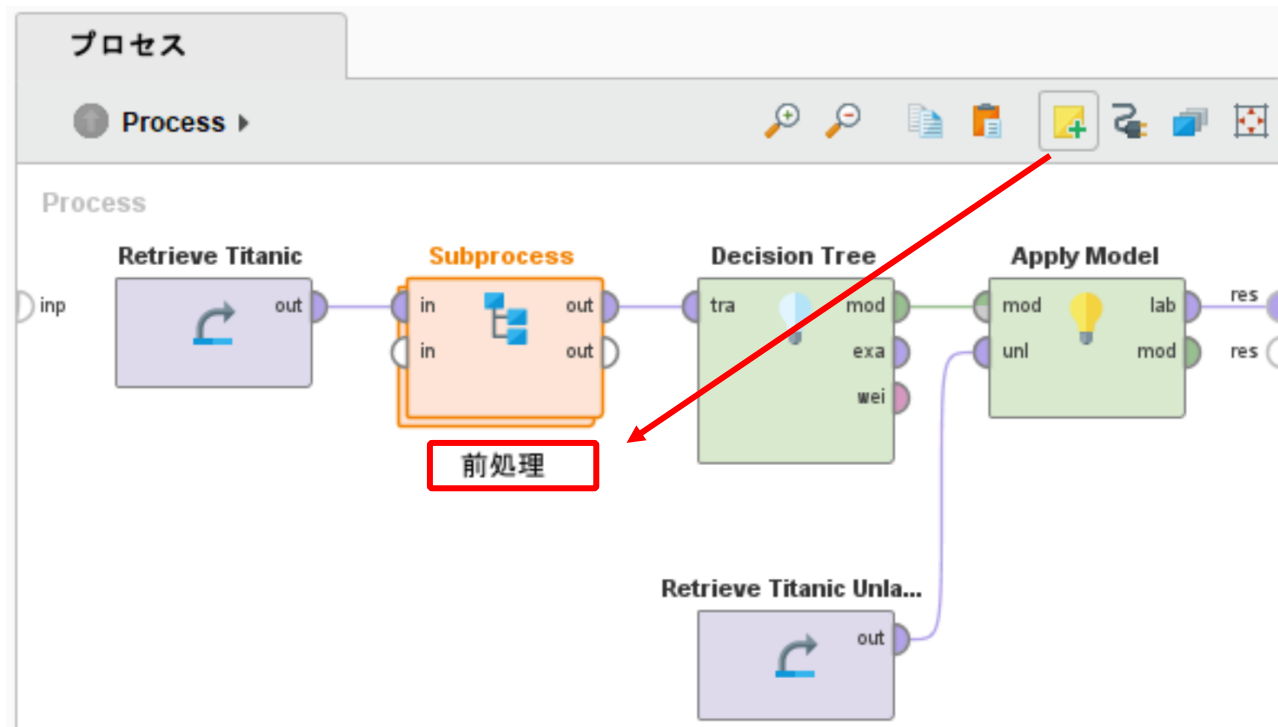
プロセスにオペレータを配置していくと、オペレータの数が増えて何をどこで行っているのか把握しづらくなります。

複数のオペレータを選択して **右クリック** > **Move into new subprocess** を選択すると、オペレータをまとめることができます。



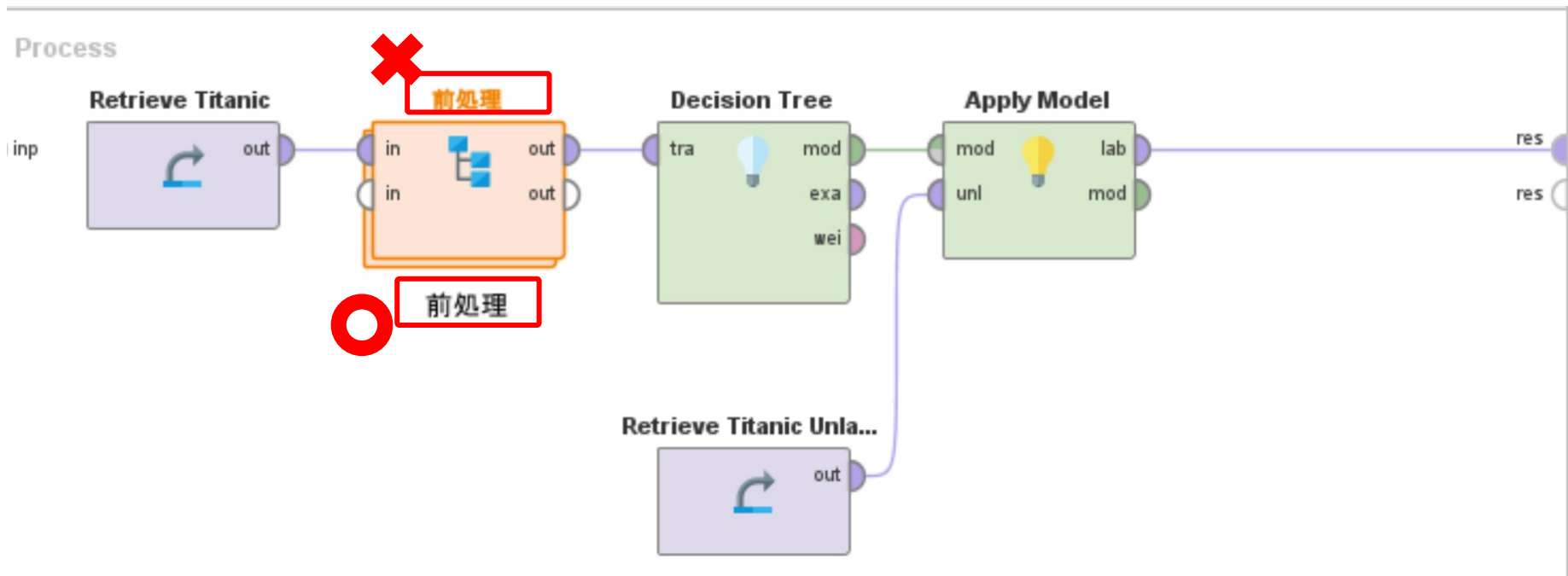
メモ

プロセスを選択して、プロセス右上のメモのマークをクリックすると、オペレータにメモを追加することができます。



メモの注意

オペレータ名を変更して、**オペレータ名を日本語にするのは避けてください**。
オペレータ名に日本語を多用すると、結果画面が表示されなくなることがあります。
日本語での処理の説明は、原則**メモ**で行うようにしてください。

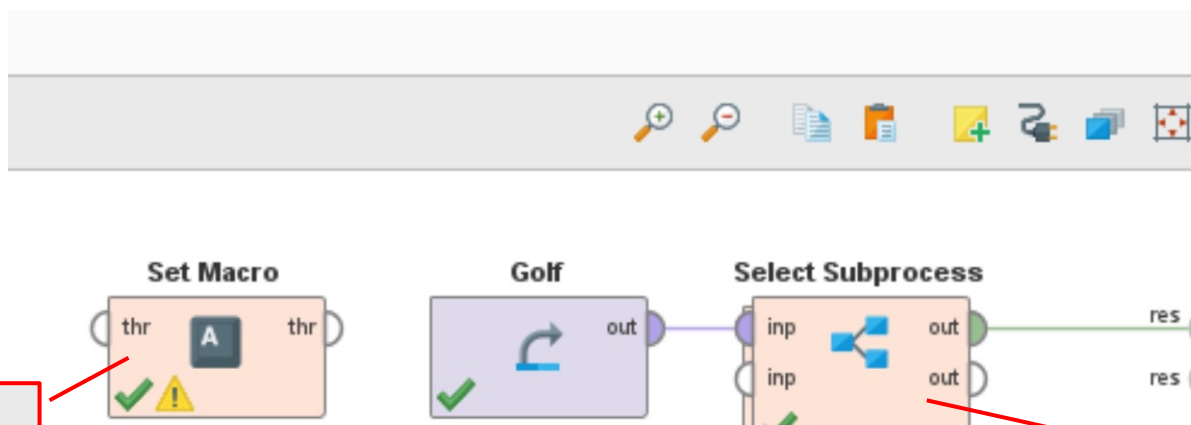


Tips

マクロ

RapidMinerには、**マクロ**というものがあります。
これはいわゆる「変数」で、**キーと値のペア**になっています。

例)



Set Macro

macro	number
value	1

"number"マクロの
定義

%{number}で
"number"マクロに
アクセス

Select Subprocess

select which	%{number}
--------------	-----------

number : 1

numberマクロに「1」が入っている状態

%{number}で呼び出し

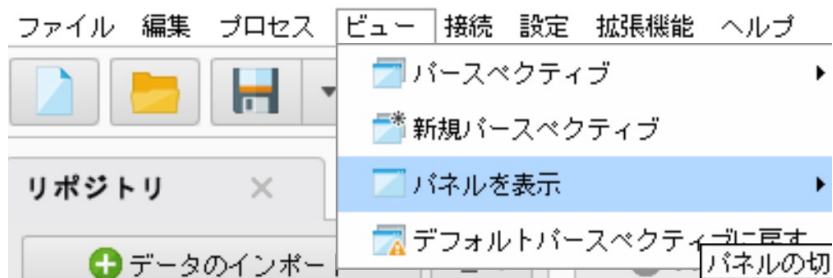
%{number}でnumberマクロを呼び出す。
今は、%{number} = 1 になります。

1

Tips

マクロ

定義したマクロは、マクロパネルで確認できます。
(ビュー > パネルを表示 > マクロ より)



The image shows a window titled 'マクロ' (Macro) with a close button. It contains a table with the following data:

Macro	Value
number	1

Tips

マクロ

マクロを使用することで、より複雑なプロセスを組んでいくことができます。
(例: BranchオペレータやLoop系のオペレータとの組み合わせなど)
以下は、Loopオペレータのチュートリアルです。

The image shows a workflow editor interface. At the top, a 'Retrieve Deals' operator is connected to a 'Loop' operator. The 'Loop' operator has a 'number of iterations' set to 10 and an 'iteration macro' set to 'iteration'. A red box highlights the 'iteration' macro name. To the right, a 'パラメータ' (Parameters) panel shows the 'Loop' configuration with the 'iteration' macro name highlighted in a red box. Below the main workflow, a detailed view of the 'Loop' operator's internal macro is shown. It contains a 'Delay' operator (1 second pause) and a 'Generate Attributes' operator. A red box highlights the 'Generate Attributes' operator's configuration table:

attribute name	function expressions
<code>%{iteration}</code>	<code>"Loop " + eval(%{iteration})</code>

Loopオペレータ
でマクロの定義

マクロの利用

Tips

マクロ

結果概要

IObjectCollection (Loop) ×

IObjectCollection

- ExampleSet
- ExampleSet
- ExampleSet
- ExampleSet
- ExampleSet
- ExampleSet
- ExampleSet
- ExampleSet
- ExampleSet
- ExampleSet

データ

基本統計量

Visualizations

開く Turbo Prep Auto Model

Row No.	Future Cust...	Age	Gender	Payment Me...	1
1	yes	64	male	credit card	Loop 1
2	yes	35	male	cheque	Loop 1
3	yes	25	female	credit card	Loop 1
4	no	39	female	credit card	Loop 1
5	yes	39	male	credit card	Loop 1
6	no	28	female	cheque	Loop 1

Loopオペレータで定義したマクロを利用している

マクロ

さらに応用的な使い方ですが、**環境パネル**を利用することで、オペレータを使用せずとも、マクロを定義することができます。

The screenshot displays a workflow editor with three main components: a 'Set Macro' block, a 'Golf' block, and a 'Select Subprocess' block. The 'Set Macro' block is highlighted with a yellow box and labeled '"number"マクロの定義'. The 'Select Subprocess' block is also highlighted with a yellow box and labeled '%{number}で"number"マクロにアクセス'. The 'Golf' block is positioned between them. To the right, the '環境' (Environment) panel is open, showing a table of macro definitions. The 'number' macro is highlighted with a red box, showing its value as '1'.

Name	Location
input 1 (disconnected)	

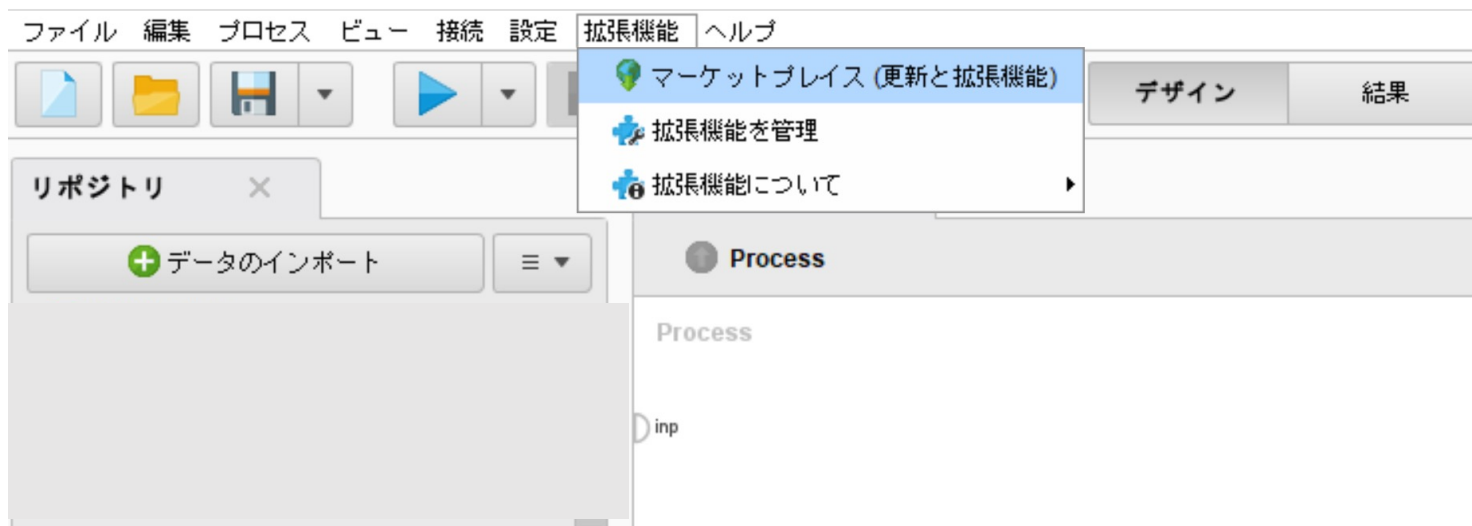
Name	Location
result 1 (Select Subprocess.output 1)	
result 2 (disconnected)	

Macro	Value ↓
number	1

Set Macroを使用せずとも、環境パネルでマクロを定義している。

マーケットプレイス

RapidMinerマーケットプレイスより、エクステンションをダウンロードできます。エクステンションを利用することで、RやPythonなどと連携できたり、より便利なデータ処理ライブラリを利用することができます。



マーケットプレイスへは、**拡張機能 > マーケットプレイス(更新と拡張機能)**よりアクセスできます。

※アクセスできない場合は、[こちら](#)を参照してください。

マーケットプレイス

ここで好みのエクステンションをダウンロードします。

The screenshot displays the RapidMiner Marketplace window. At the top, there is a header with the title "RapidMinerマーケットプレイス" and a close button. Below the header is a globe icon and a message: "インストール、あるいはアップデートするコンポーネントを選択してください。RapidMiner Studioのアップデートは必ずグローバルにインストールされます。グローバルのアップデートでは、アップデート時と再起動時の両方で管理者権限が必要です。"

The main content area is divided into two columns. The left column lists several extensions:

- Text Processing 9.4.0**: The Text Extension adds all operators necessary for statistical text analysis and Natural Language Processing (NLP). Includes a checkmark and the text "パッケージは最新のバージョンです".
- Operator Toolbox 2.14.0**: Operator Toolbox Extension: This extensions couples a bunch of useful additional operators together. Includes a checkmark and the text "パッケージは最新のバージョンです".
- Web Mining 9.7.1**: The Web Mining Extension provides access to internet sources like web pages, RSS feeds, and web services. Status: "インストールされていません".
- Anomaly Detection 4.0.1**: The Anomaly Detection Extension comprises the most well know unsupervised anomaly detection algorithms, assigning individual anomaly scores to data rows of example sets. Includes a blue arrow icon and the text "インストールしたバージョン: 4.0.000".
- Python Scripting 9.10.2**: The Python scripting extension integrates RapidMiner with the data scientist-friendly and widely used programming language Python and allows to embed Python code into RapidMiner processes. Includes a checkmark and the text "パッケージは最新のバージョンです".

The right column shows the details for the selected "Text Processing" extension:

- Text Processing**
- Version: 9.4.0
- Release date: 2021/12/14
- File size: 19.5 MB
- License: RM_EULA
- Dependencies
- Description: "The Text Extensions supports several text formats including plain text, HTML, or PDF as well as other data sources. It provides standard filters for tokenization, stemming, stopword filtering, or n-gram generation to provide everything necessary for preparing and analyzing texts."

At the bottom of the window, there is a status bar with a checkmark icon and the text "0つのパッケージをインストール" and a red X icon with the text "閉じる". A link "拡張機能のホームページへ移動する" is also visible.

マーケットプレイス

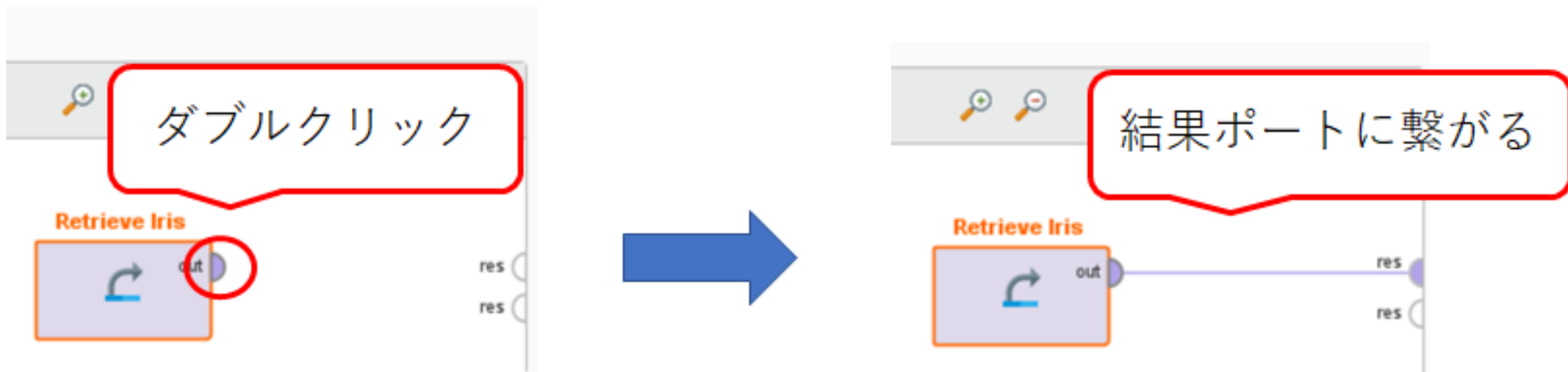
おすすめエクステンション

- **Python Scripting**
- **R Scripting**
- **Operator Toolbox** (※Text Processingも。依存関係があるため)
- **Converters**
- **Deep Learning** (※ND4J Back Endも。依存関係があるため)
- **Image Handling**
- **Sensor Link**
- **In-Database Processing**

Tips

オペレータの繋ぎ方のコツ①

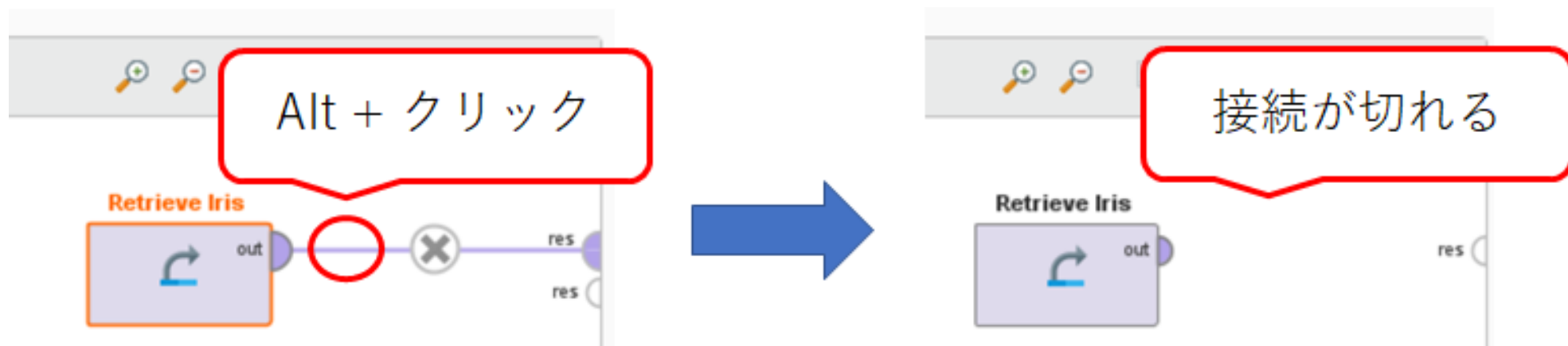
ポートをダブルクリックすると、結果ポートにつながります。



Tips

オペレータの繋ぎ方のコツ②

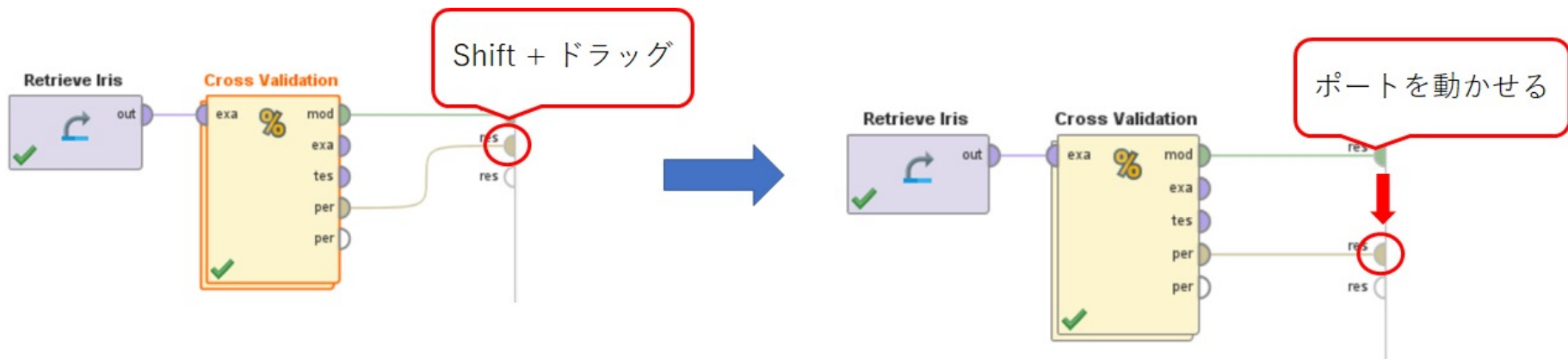
Altキー + クリックで接続を切れます。



Tips

オペレータの繋ぎ方のコツ③

Shiftキー + ドラッグでポートの位置を変えられます。



Tips

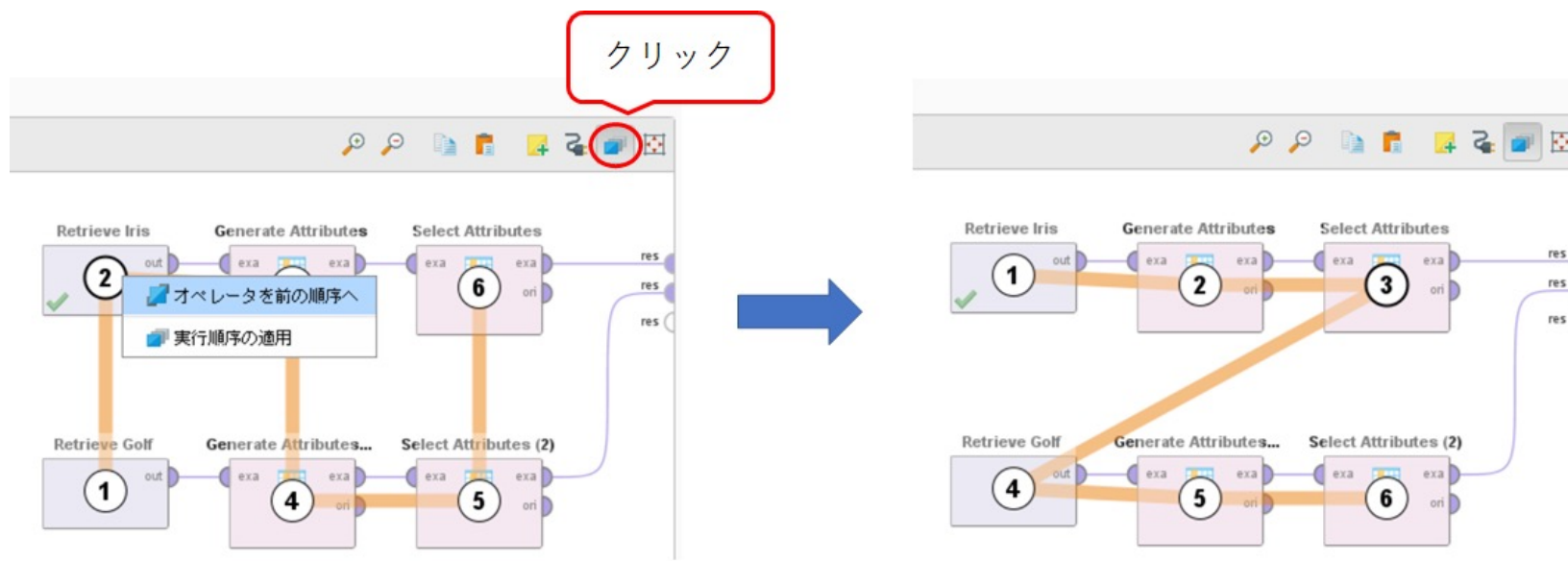
オペレータの繋ぎ方のコツ④

ポート同士が離れている場所にも、ポートを繋ぐことができます。
Remember、Recallオペレータを用いれば、たとえポート同士が離れていてもポートを繋ぎ、内容を送ることができます。



オペレータの実行順

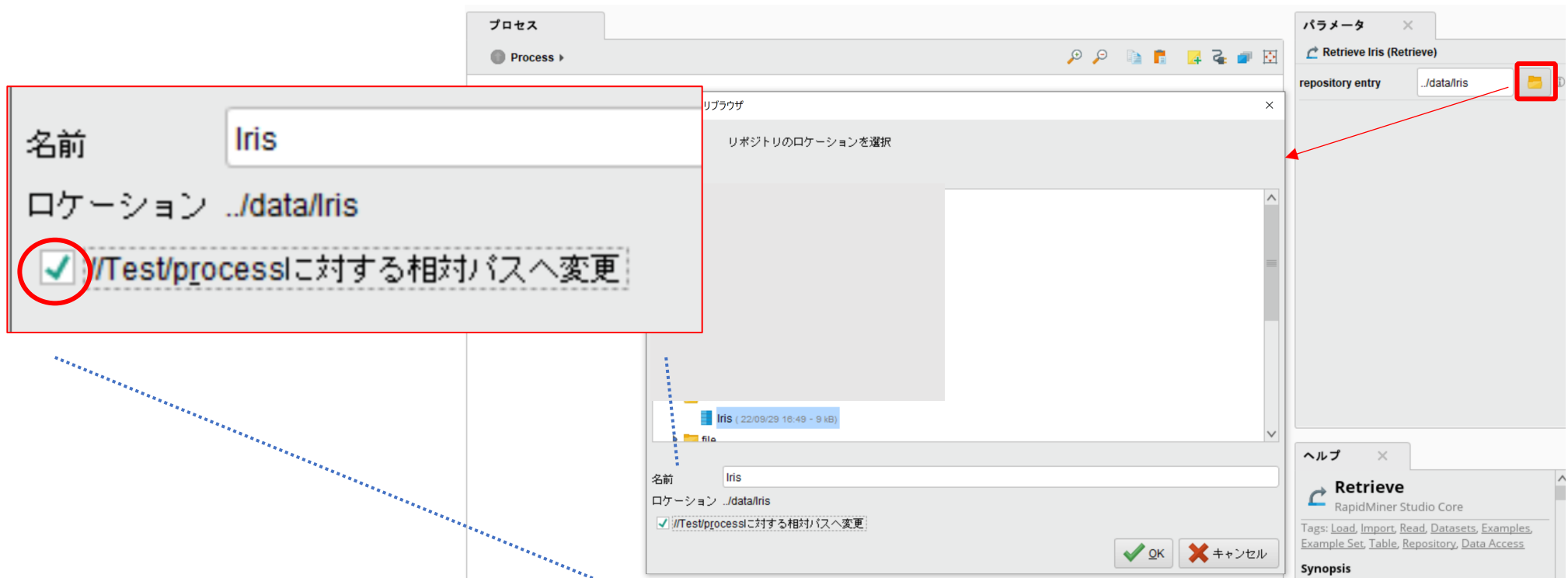
プロセスの右上にある四角形のマークをクリックすると、実行順を確認できます。さらに、オペレータを右クリックし、**オペレータを前の順序へ**を選択すると、オペレータの実行順を変更することが可能です。プロセスの作成に戻るには、もう一度四角形のマークをクリックします。この機能を使用して、オペレータが適切な順で実行されるよう調整できます。



相対パス

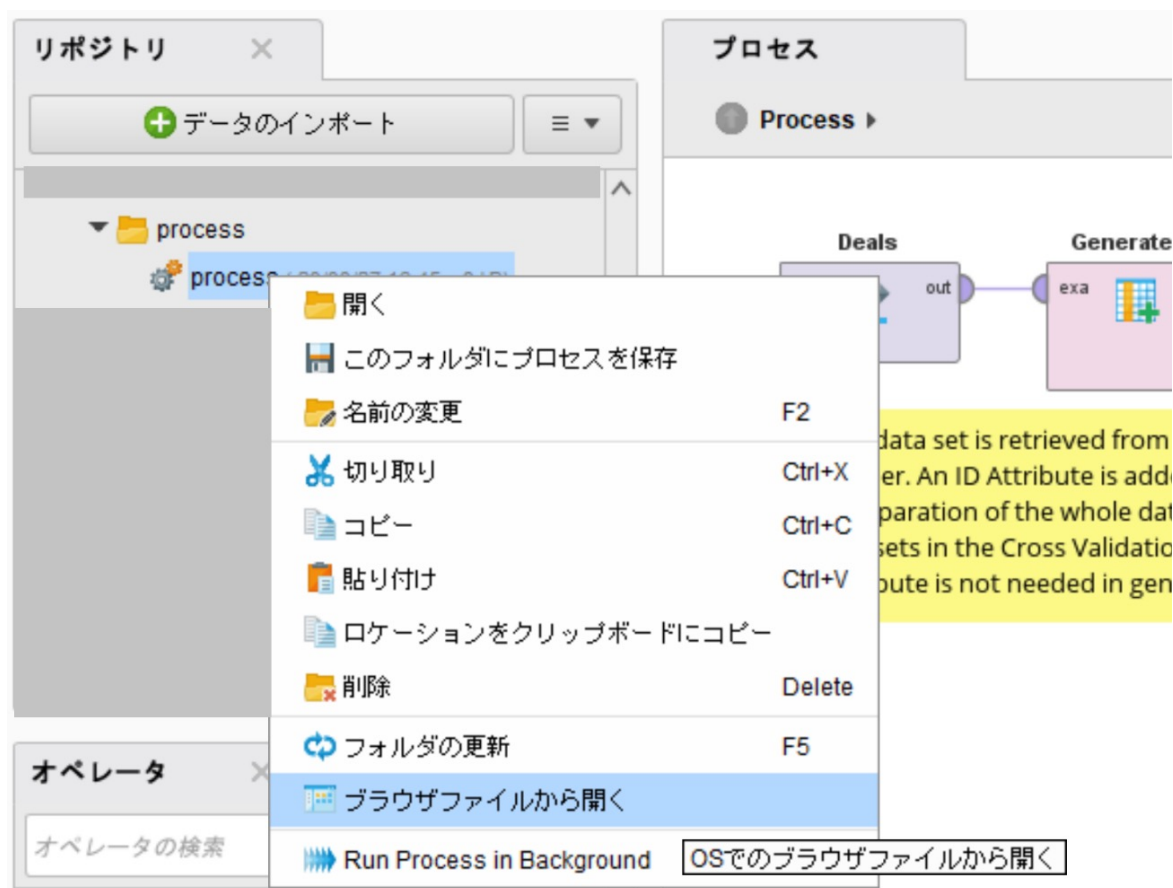
データを参照するときは、可能な限り**相対パスで取得**することをおすすめします。理由は、別の人へプロセスを渡した際も、そのまま実行できるようにするためです。

相対パスは、**～に対する相対パスへ変更** をクリックすると、相対パスに変わります。



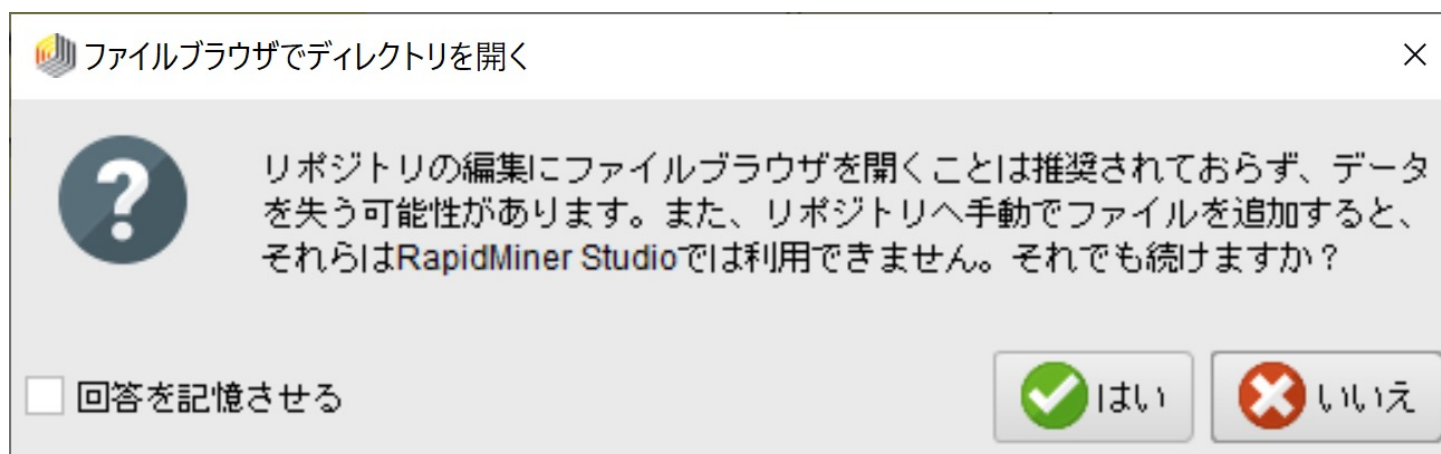
リポジトリの実データの場所

リポジトリを選択して右クリック > **ブラウザファイルから開く** よりリポジトリの実データの場所を開くことができます。



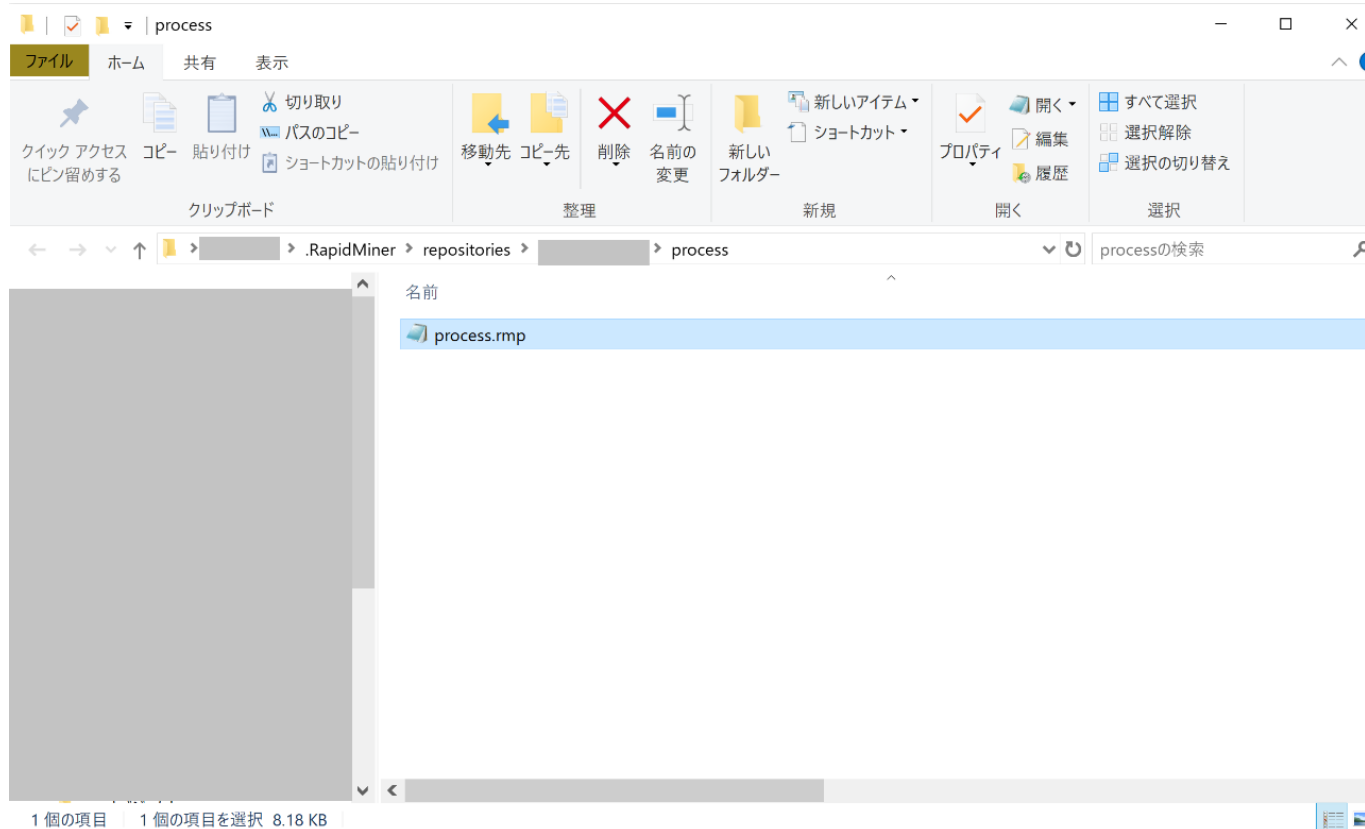
リポジトリの実データの場所

ファイルを開く際、以下の警告が出ますが、**はい** をクリックします。



リポジトリの実データの場所

開いた場所の内容とリポジトリの内容は連動しているので、
誤ってデータを消さないよう取り扱いには注意してください。



リポジトリの実データの場所

また、この場所に別の実データを移し、リポジトリを更新すると、リポジトリの内容をそのまま移行させることができます。

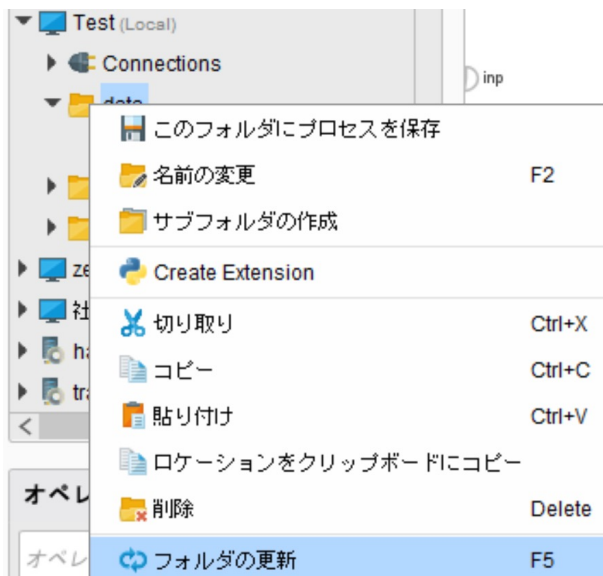
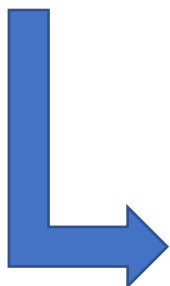
ト > RapidMiner > Repositories > Test > data

名前

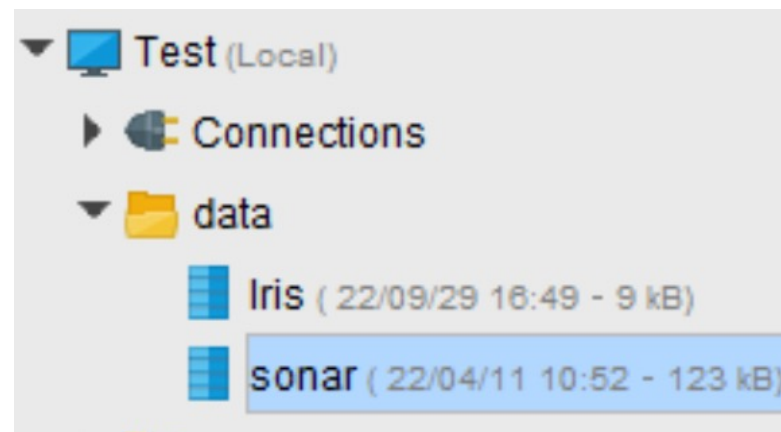
Iris.rmhdf5table

sonar.rmhdf5table

データを直接追加



リポジトリの内容が連動



リポジトリを右クリック
> フォルダの更新

トラブルシューティング

トラブルシューティング

RapidMinerを初めて操作する際の、よくあるお問い合わせをまとめました。
自身の問題に合わせて、該当のページをご覧ください。

- [RapidMinerが起動しない](#)
- [Could not contact license server.のエラー](#)

おわりに

まとめ

今回はRapidMinerの基本操作についてみてきました。

RapidMinerの操作について、最初は戸惑うことが多いと思います。
しかし、操作に慣れるには、その中でも実際に操作することが重要です。

この後は[チュートリアル](#)なども試して、ぜひ実際に動かしてみてください。

