

目的変数最適化へのアプローチ

ーRapidMiner Prescriptive Optimizationの実装ー

株式会社KSKアナリティクス



1. 予測モデル作成と最適化

はじめに

予測モデルの作成は、次に何が起きようとしているのかを予測するのに役立ちます。適切な方法を用いれば、なぜそれが起きるのか原因を理解することもできる場合があります。

しかし、何が起こるのか、なぜ起こるのかだけでは、意思決定を行うには十分とは言えません。意思決定を行うためには、結果を変更するためにどのパラメーターをどれくらいにすれば所望の値に近づけるのかを知る必要があります。それを実現できるのがRapidMinerの最適化オペレーター（Prescriptive Optimizer）です。

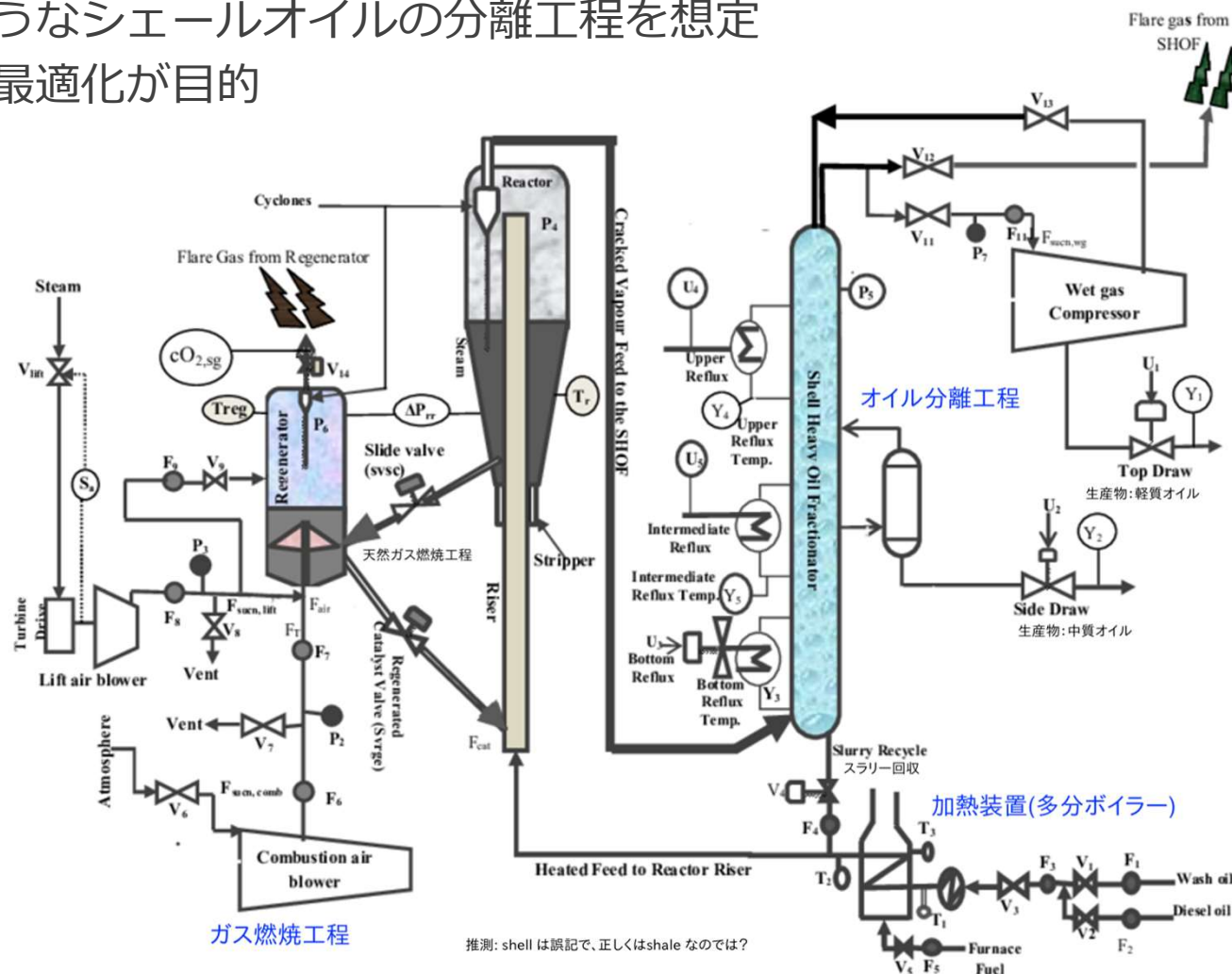
本資料では、目的変数と制御可能な変数、制約がある変数を組み込んだモデルを作成し、最適化を実行します。

※使用するデータと作成したサンプルプロセスをダウンロード頂ければ
お手元のRapidMinerで再現して頂くことが可能です。

2. データ理解と前処理

分析課題

下記のようなシェールオイルの分離工程を想定
生産量の最適化が目的



2. データ理解と前処理

データセット

時間間隔も内容も異なる 3 つのデータセットがあります。
目的変数は生産量（Yield）です。

【sensor.xlsx】

- preheated air temperature : 空気余熱
- steam flow : 蒸気流量
- Condenser temperature : コンデンサの温度
→ 30秒間隔、9,360レコード

【weather.xlsx】

- Outside_press : 外の気圧
- Outside_temp : 外の温度
→ 2時間間隔、39レコード

【yield.xlsx】

- **Yield : 生産量**
- margin : マージン
- batch_id : バッチID
- reactant degree : 反応温度
→ 5分間隔、911レコード

2. データ理解と前処理

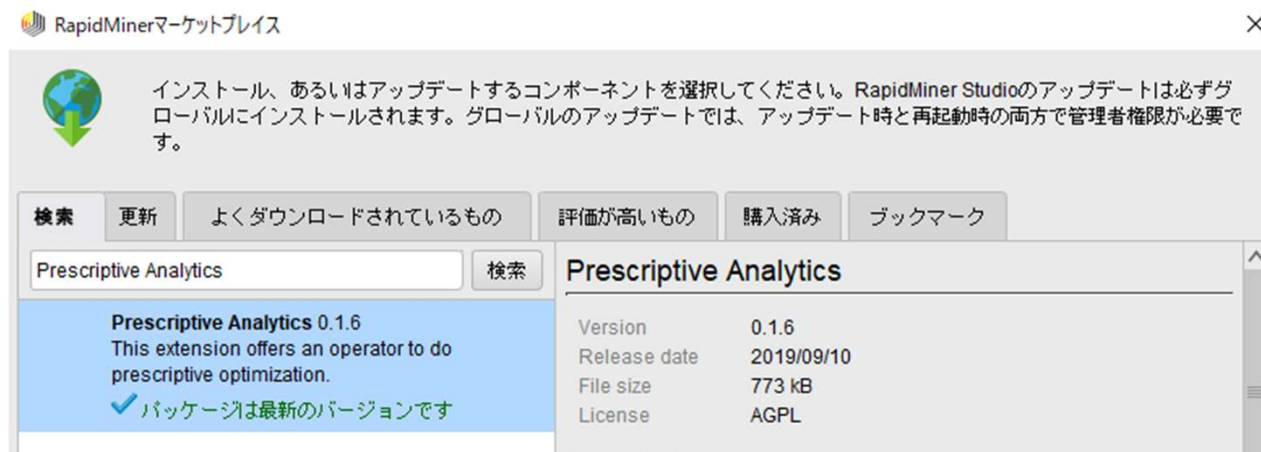
事前準備

RapidMinerのマーケットプレイスから

“Prescriptive Analytics”エクステンションをインストールします

- ・ ツールバー「拡張機能」→「マーケットプレイス」

“Prescriptive Analytics”で検索し、インストールを行ってください。



上記手順が出来ない場合は、下記URLよりファイルダウンロードして、マニュアルでエクステンションを適用します。

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

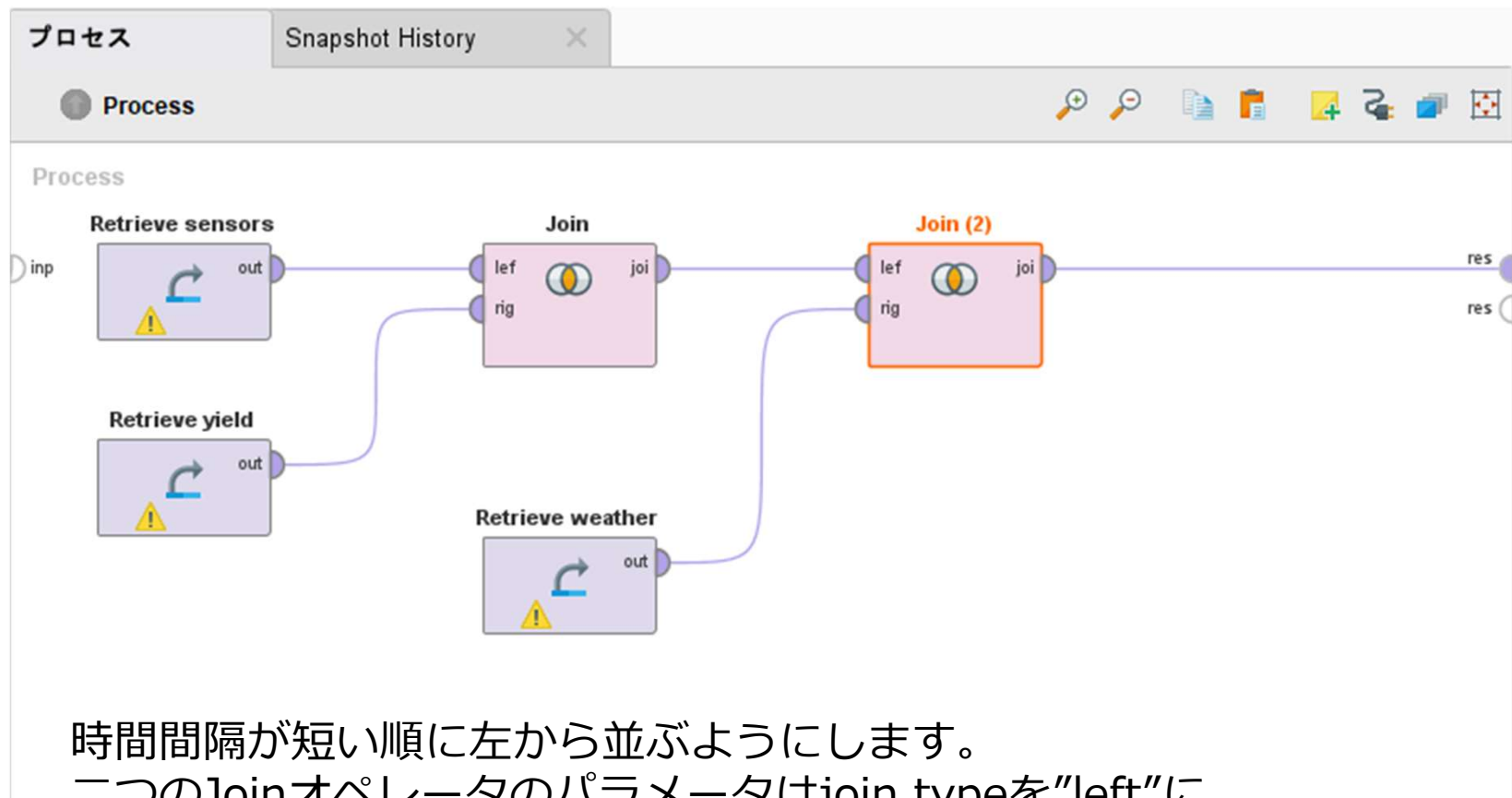
ダウンロードしたファイルは下記に格納し、RapidMinerを再起動します。

C:\¥Users¥<ユーザー名>\.RapidMiner¥extensions

2. データ理解と前処理

前処理① 結合

3つのデータセットを結合し、一つのデータセットにします。



時間間隔が短い順に左から並ぶようにします。
二つのJoinオペレータのパラメータはjoin typeを"left"に、
key attributesをleft、right共にdatetimeと設定します。

2. データ理解と前処理

前処理① 結合

下図の9,360行のデータセットができます。

ExampleSet (Join (2))										
開く		Turbo Prep		Auto Model		フィルタ (9,360 / 9,360 行): all				
Row No.	datetime	Pre_heated...	Steam_flow	Condenser_...	Yield	margin	batch_id	reactant_de...	Outside_pre...	Outside_temp
1	2017/07/10 1...	102.900	94.058	69.200	4258.002	498.700	10388	825.001	1027	9.800
2	2017/07/10 1...	102.955	97.358	69.207	?	?	?	?	?	?
3	2017/07/10 1...	103.010	100.657	69.213	?	?	?	?	?	?
4	2017/07/10 1...	103.065	103.956	69.220	?	?	?	?	?	?
5	2017/07/10 1...	103.120	107.255	69.227	?	?	?	?	?	?
6	2017/07/10 1...	103.175	110.555	69.233	?	?	?	?	?	?
7	2017/07/10 1...	103.230	113.854	69.240	?	?	?	?	?	?
8	2017/07/10 1...	103.285	117.153	69.247	?	?	?	?	?	?
9	2017/07/10 1...	103.340	120.452	69.253	?	?	?	?	?	?
10	2017/07/10 1...	103.395	123.752	69.260	?	?	?	?	?	?
11	2017/07/10 1...	103.450	127.051	69.267	3746.959	521.857	10389	629.235	?	?
12	2017/07/10 1...	103.195	125.894	69.073	?	?	?	?	?	?
13	2017/07/10 1...	102.940	124.736	68.880	?	?	?	?	?	?
14	2017/07/10 1...	102.685	123.579	68.687	?	?	?	?	?	?
15	2017/07/10 1...	102.430	122.422	68.493	?	?	?	?	?	?

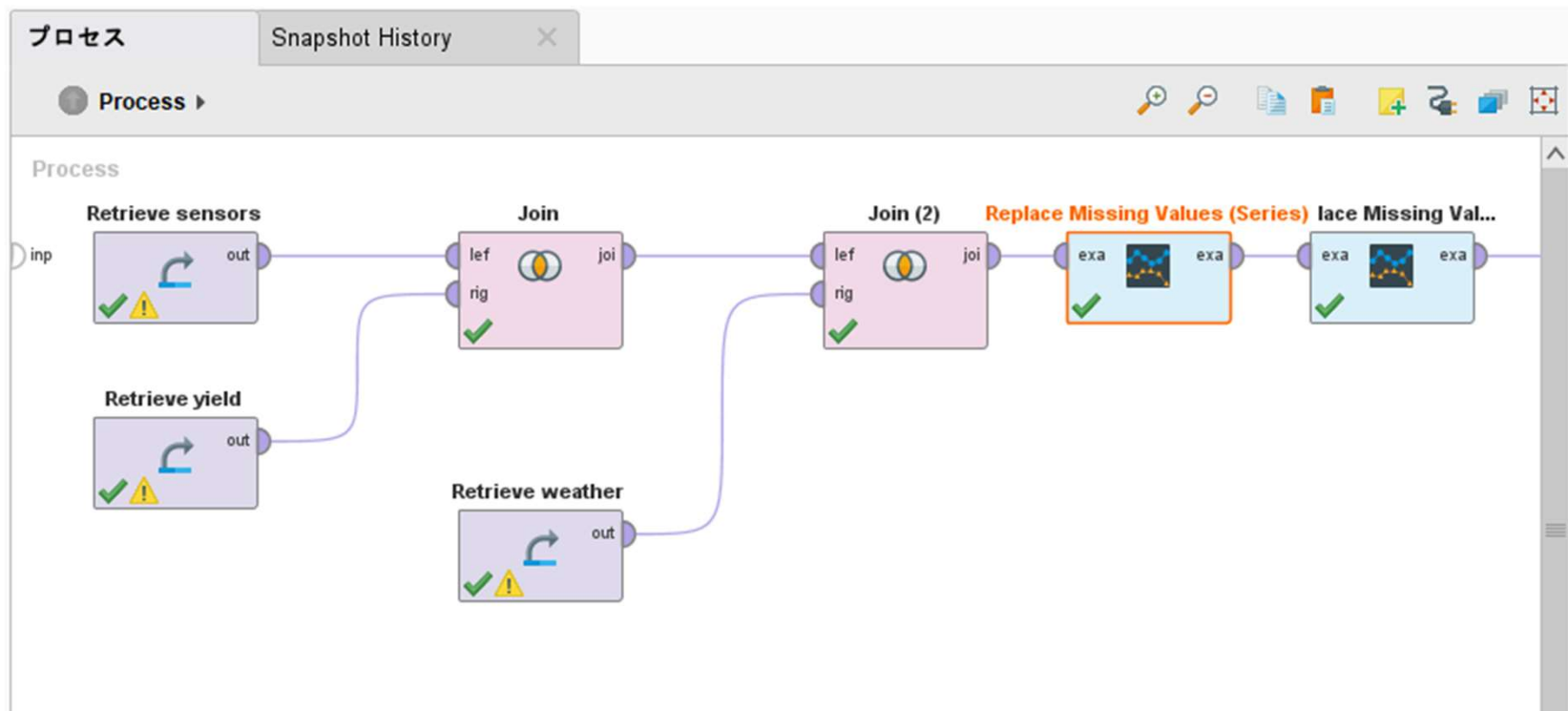
ExampleSet (9,360 行, 0 特別属性, 10 通常属性)

2. データ理解と前処理

前処理② 欠損値補間

“Replace Missing Values (Series)”を二つ、下図のように配置します。

- ・一つ目はattribute filter typeをsubsetにし、属性選択をクリックして、Outside_pressとOutside_tempを右側のボックスに移します。
replace type numericalをlinear interpolationに変更します。
- ・二つ目は初期設定から変更ありません。



2. データ理解と前処理

前処理② 欠損値補間

一つ目の“Replace Missing Values (Series)”のパラメータ

The image shows two overlapping windows from a data processing software interface. The background window is titled 'パラメータ' (Parameters) and contains the 'Replace Missing Values (Series)' configuration. The foreground window is titled '属性を選択: attributes' (Select Attributes: attributes) and shows a list of attributes with two selected items.

Replace Missing Values (Series) Parameters:

- attribute filter type: subset
- attributes: 属性選択 (Attribute Selection)
- invert selection: ☐
- include special attributes: ☐
- has indices: ☐
- overwrite attributes: ☒
- replace type numerical: linear interpolation
- replace type nominal: previous value
- replace type date time: previous value

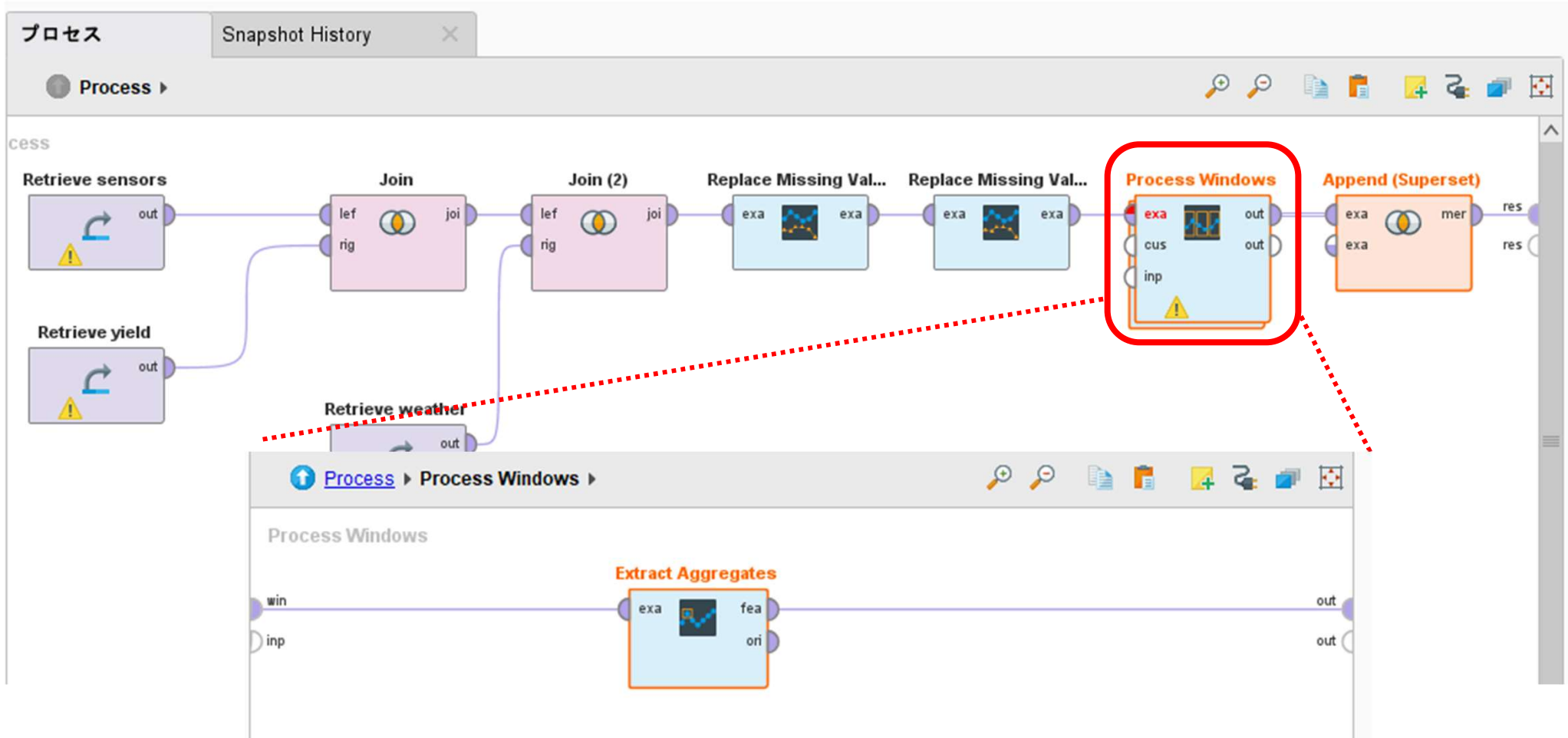
Select Attributes: attributes Dialog:

- 属性 (Attributes):
 - batch_id
 - Condenser_temp
 - datetime
 - margin
 - Pre_heated_air_temp
 - reactant_degree
 - Steam_flow
 - Yield
- 選択した属性 (Selected Attributes):
 - Outside_press
 - Outside_temp

2. データ理解と前処理

前処理③ 時系列データの特徴量抽出

下図の通り、三つのオペレータを配置します。



2. データ理解と前処理

前処理③ 時系列データの特徴量抽出

パラメータ設定は下図の通りです。

The 'Process Windows' dialog box is shown with the following settings highlighted by red boxes:

- ☒ has indices
- indices attribute: datetime
- ☒ expert settings
- unit: example based
- windows defined: from start
- window size: 10
- ☒ no overlapping windows

Process Windows

The 'Extract Aggregates' dialog box is shown with the following settings highlighted by red boxes:

- ☒ mean
- ☐ geometric mean
- ☐ first quartile
- ☐ median
- ☐ third quartile
- ☐ min
- ☐ max
- ☐ std deviation
- ☐ kurtosis
- ☐ skewness
- ☒ add time series name
- ☐ ignore invalid values

Extract Aggregates

2. データ理解と前処理

前処理④ 列名の置き換え

下図の二つのオペレータを配置し、特徴量名を整え、データセットを保存します。
保存先はお好みの先を指定してください。

The screenshot displays the KSK Analytics workflow editor. The main process area shows a sequence of operators: 'Join (2)', 'Replace Missing Val...', 'Replace Missing Val...', 'Process Windows', 'Append (Superset)', 'Rename by Replacing', and 'Store'. A red arrow points from the 'Rename by Replacing' operator in the workflow to its configuration dialog box. The dialog box has tabs for 'パラメータ' (Parameters) and 'ヘルプ' (Help). Under 'パラメータ', the 'attribute filter type' is set to 'all'. The 'replace what' field contains the text '\\.mean', which is highlighted with a red rectangle. The 'replace by' field is empty. To the right of the dialog box, there is a note in Japanese: '※meanの前に、.を入れてください' (Please add a dot before mean).

2. データ理解と前処理

前処理④ 列名の置き換え

結果画面

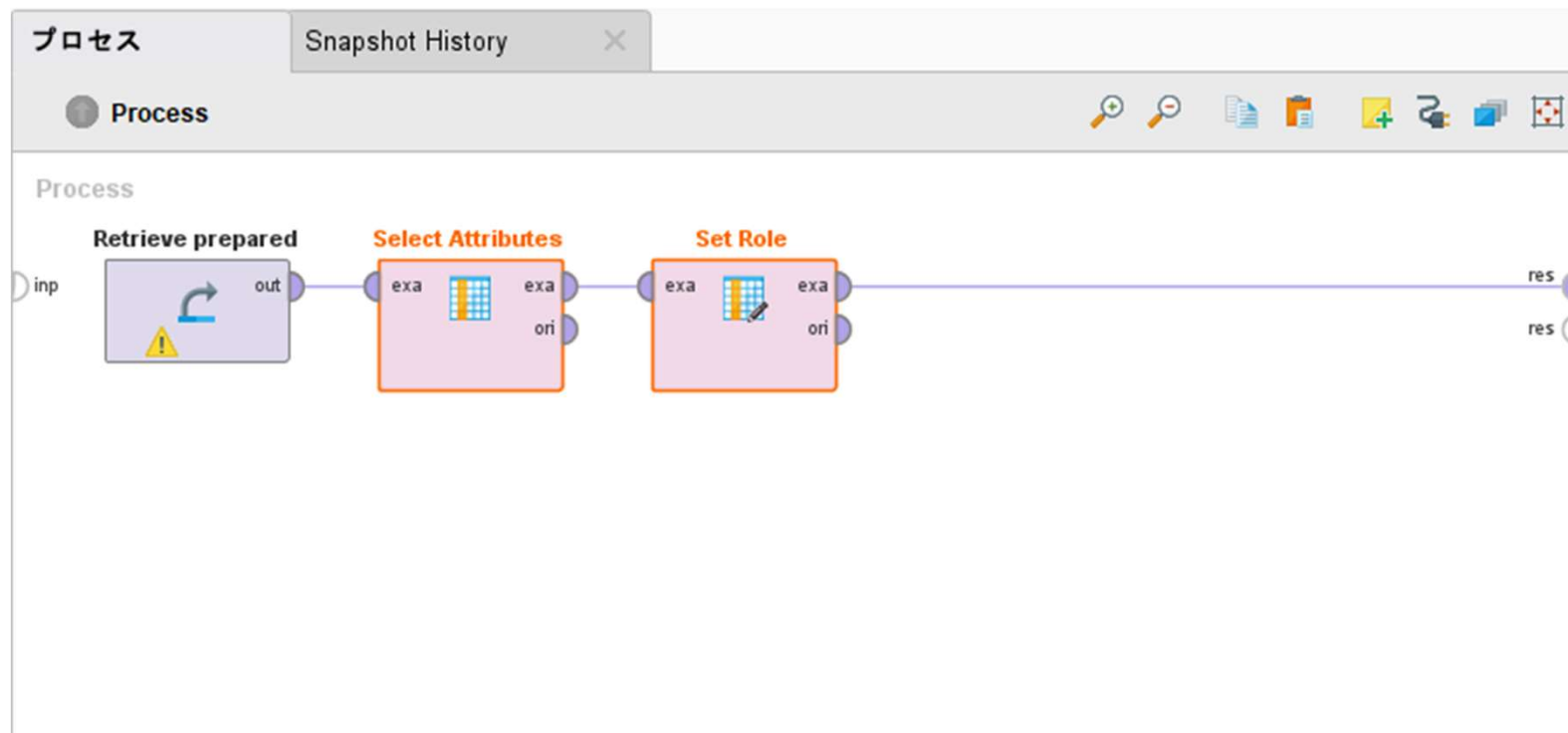
Row No.	Pre_heated...	Steam_flow	Condenser_...	Yield	margin	batch_id	reactant_de...	Outside_pre...	Outside_temp	Last datetim...
1	103.148	108.905	69.230	4258.002	498.700	10388	825.001	1026.940	9.826	2017/07/10 1...
2	102.303	121.843	68.397	3746.959	521.857	10389	629.235	1026.807	9.885	2017/07/10 1...
3	100.765	109.040	67.498	3801.912	495.091	10390	668.967	1026.673	9.943	2017/07/10 1...
4	100.780	109.996	67.468	4021.988	484.645	10391	753.745	1026.540	10.001	2017/07/10 1...
5	100.213	123.817	66.951	4021.988	484.645	10391	753.745	1026.407	10.060	2017/07/10 1...
6	101.230	123.075	67.957	3588.880	503.384	10392	601.455	1026.273	10.118	2017/07/10 1...
7	103.403	115.227	69.458	3876.510	510.702	10393	676.316	1026.140	10.176	2017/07/10 1...
8	103.302	105.455	69.517	3934.296	502.992	10394	702.111	1026.007	10.235	2017/07/10 1...
9	102.380	104.350	68.905	4220.698	501.035	10395	806.137	1025.873	10.293	2017/07/10 1...
10	101.377	110.572	68.218	3838.119	493.075	10396	682.170	1025.740	10.351	2017/07/10 1...
11	102.215	115.355	68.567	3965.948	493.614	10397	722.981	1025.607	10.410	2017/07/10 1...
12	103.905	117.015	69.452	3738.744	515.136	10398	632.529	1025.473	10.468	2017/07/10 1...
13	106.130	116.850	71.148	4110.280	512.151	10399	750.105	1025.340	10.526	2017/07/10 1...
14	107.305	120.288	72.020	3852.676	540.262	10400	643.004	1025.207	10.585	2017/07/10 1...
15	107.215	113.541	71.807	4152.456	525.295	10401	748.671	1025.073	10.643	2017/07/10 1...

ExampleSet (936 行, 0 特別属性, 10 通常属性)

3. モデル作成

モデル作成の準備

新規プロセスで、先ほど保存したデータセットを呼び出します（ここではprepared）
“Select Attributes”で特徴量選択を行い、
“Set Role”を配置し、Yieldを目的変数を設定します。



3. モデル作成

モデル作成の準備

Select Attributes

パラメータ × ヘルプ ×

Select Attributes

attribute filter type: subset

attributes: 属性選択

☐ invert selection

☐ include s

Set Role

パラメータ × ヘルプ ×

Set Role

attribute name: Yield

target role: label

set additional roles: Edit List (0)...

属性を選択: attributes

The attribute which should be chosen.

属性

検索

- # batch_id
- Last datetime in window
- # margin

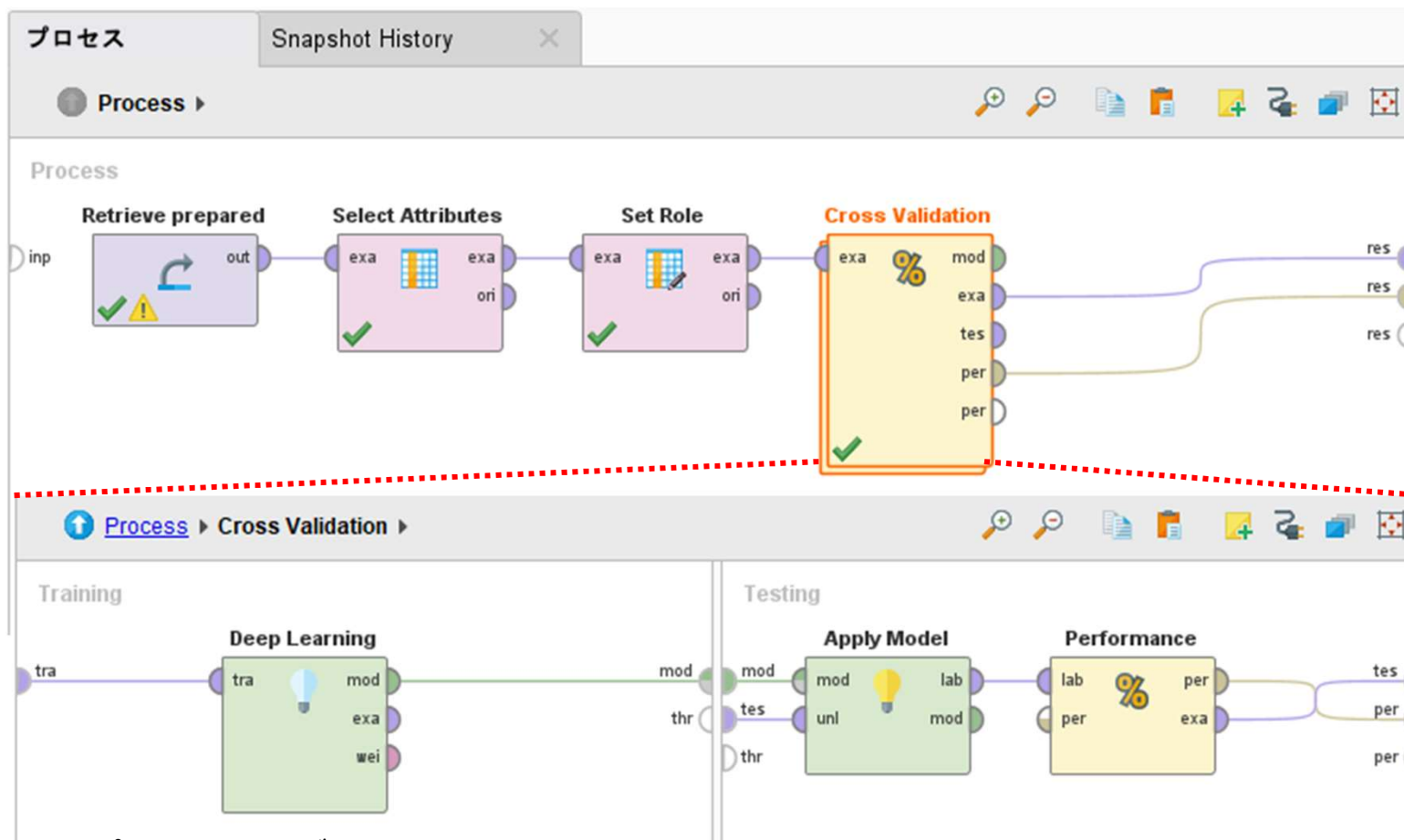
選択した属性

- # Condenser_temp
- # Outside_press
- # Outside_temp
- # Pre_heated_air_temp
- # reactant_degree
- # Steam_flow
- # Yield

3. モデル作成

モデル作成

“Cross Validation”を配置します。
事前の比較検討により、ここではDeep Learningを採用しています。



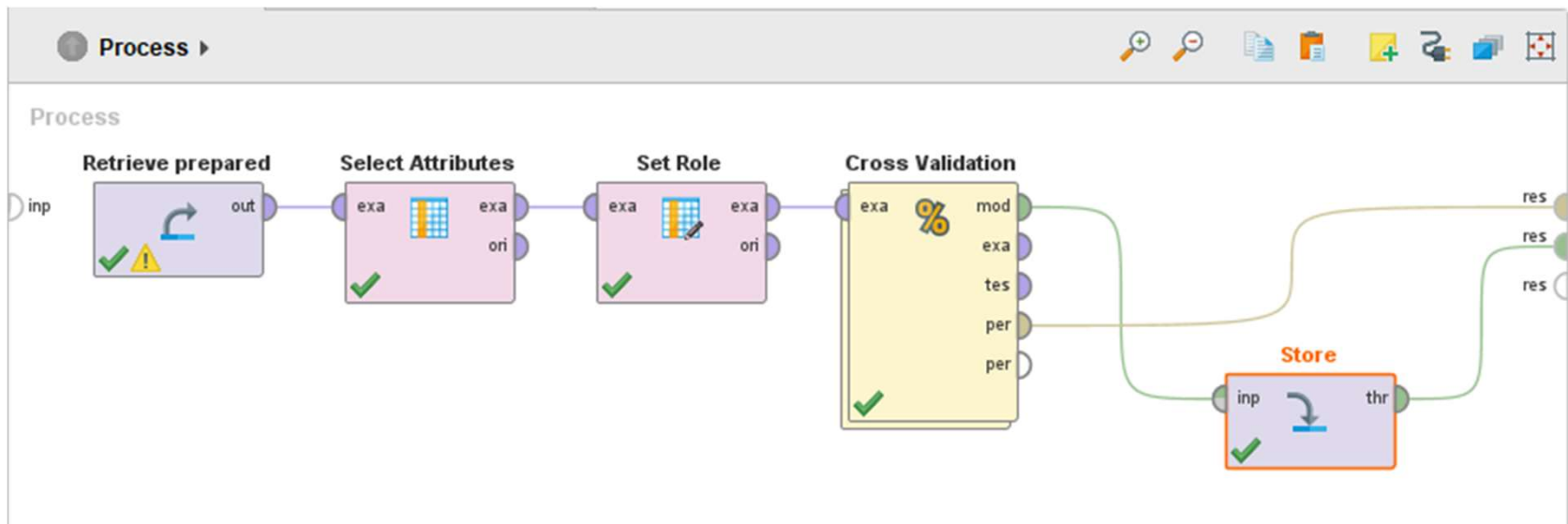
※パラメータはデフォルトです。

3. モデル作成

モデル作成と結果



作成したモデルを“Store”で任意の先に保存します。



4. 最適化

予測結果の出力

“Create ExampleSet”オペレータで下記設定のサンプルデータを作り、先ほど保存したモデル(ここではModel)を適用して、予測値を出してみます。

The screenshot displays the Orange3 software interface. The main workspace shows a workflow with three operators: 'Retrieve Model', 'Create ExampleSet', and 'Apply Model'. The 'Create ExampleSet' operator is highlighted with a red box, and a red arrow points from it to a pop-up window titled 'パラメータリストを編集: function descriptions'. This window contains a table with attribute names and function expressions.

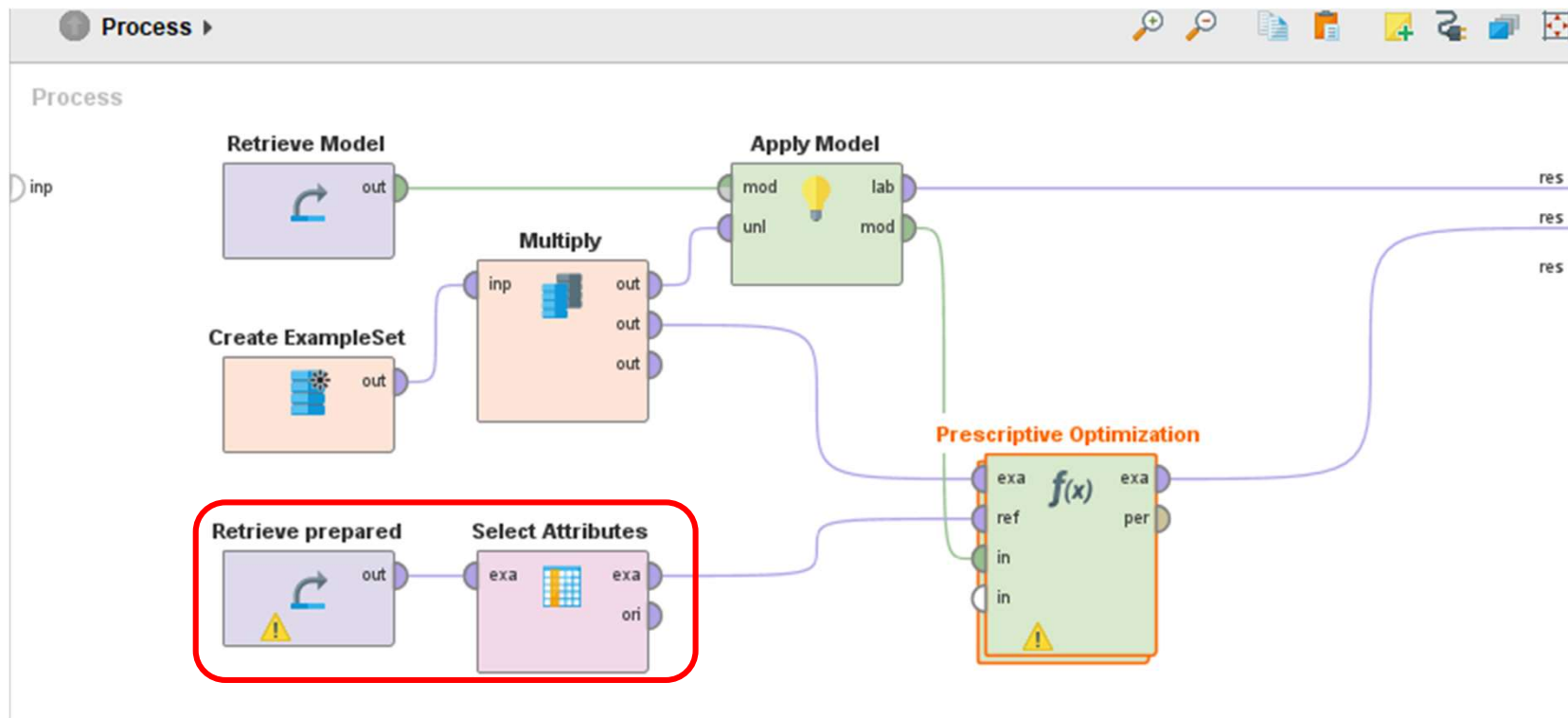
attribute name	function expressions
Condenser_temp	70
reactant_degree	300
Outside_press	1010
Outside_temp	20
Pre_heated_air_temp	80
Steam_flow	70

※予測結果はここでは割愛します。

4. 最適化

最適化

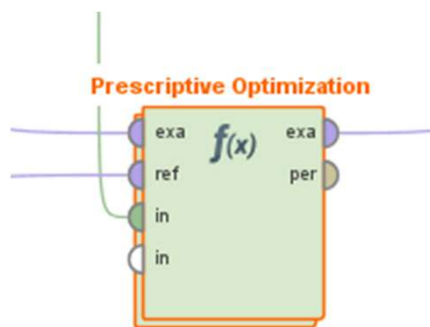
“Prescriptive Optimization”を配置し、“Multiply”で複製したサンプルを接続、モデル作成の際に配置した赤枠部をコピー&ペースト、参照例としてrefに繋がります。



4. 最適化

最適化

パラメータ設定



パラメータ ヘルプ

f(x) Prescriptive Optimization

attribute filter type: subset

attributes: 属性選択

☐ invert selection

☐ include special attributes

optimization method: BYOBA

number of candidates: 1

method for start values: average

configure bounds: Edit List (0)...

☒ use interpolation points default

max evaluations: 10000

initial radius: 10.0

属性を選択: attributes

属性を選択: attributes
The attribute which should be chosen.

属性: 検索

- # Outside_press
- # Outside_temp
- # reactant_degree

選択した属性: 検索

- # Condenser_temp
- # Pre_heated_air_temp
- # Steam_flow

適用 キャンセル

パラメータをどのように変えるべきか知りたいので、人が変更できる変数を選択します。
※外気温や気圧、反応温度は人が操作できません。

4. 最適化

最適化

パラメータ設定

Process > Prescriptive Optimization >

Prescriptive Optimization

tra in in

Apply Model (2)

mod unl lab mod

Generate Attributes

exa exa ori

Performance

exa per exa

Per

パラメータリストを編集: function descriptions

パラメータリストを編集: function descriptions
List of functions to generate.

attribute name	function expressions
taget	[prediction(Yield)]

エントリを追加 エントリを削除 適用

パラメータ ヘルプ

% Performance (Extract Performance)

performance type data_value

attribute name taget

example index 1

optimization direction maximize

4. 最適化

生産量最大化

上段が通常の予測モデルの結果、下段が処方的最適化の結果です。
各変数をどのようなパラメータにすべきか、一目でわかります。

The image displays two panels from the KSK ANALYTICS software interface, comparing a standard prediction model with a prescriptive optimization model. Red boxes and arrows indicate the mapping of variables between the two models.

ExampleSet (Apply Model)

Row No.	prediction(Yi...	Condenser_...	reactant_de...	Outside_pre...	Outside_temp	Pre_heated...	Steam_flow
1	3331.505	70	300	1010	20	80	70

ExampleSet (1行, 1 特別属性, 6 通常属性)

ExampleSet (Prescriptive Optimization)

Row No.	fitness	Condenser_...	reactant_de...	Outside_pre...	Outside_temp	Pre_heated...	Steam_flow
1	3433.423	79.660	300	1010	20	119.250	126.589

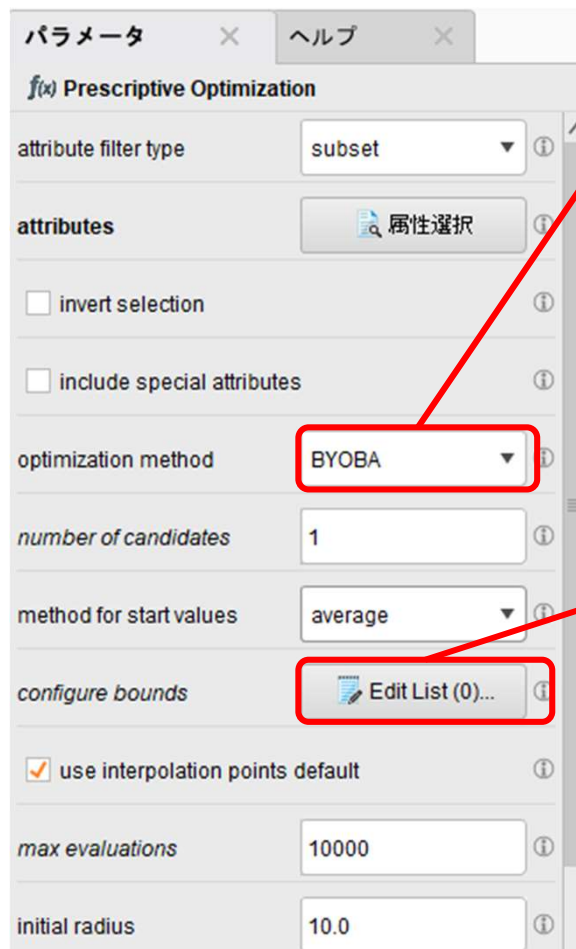
ExampleSet (1行, 0 特別属性, 7 通常属性)

Red boxes and arrows highlight the following mappings:

- prediction(Yi...)** (top) maps to **fitness** (bottom).
- Condenser_...** (top) maps to **Condenser_...** (bottom).
- Pre_heated...** (top) maps to **Pre_heated...** (bottom).
- Steam_flow** (top) maps to **Steam_flow** (bottom).

4. 最適化

Prescriptive Optimization Tips



パラメータ × ヘルプ ×

$f(x)$ Prescriptive Optimization

attribute filter type: subset

attributes: 属性選択

☐ invert selection

☐ include special attributes

optimization method: BYOBA

number of candidates: 1

method for start values: average

configure bounds: Edit List (0)...

☒ use interpolation points default

max evaluations: 10000

initial radius: 10.0

BYOBA
Evolutionary Optimization
Grid
CMA-ES
Powell

最適化メソッドは進化的アルゴリズムなど複数の方法から選択可能です。



パラメータリストを編集: configure bounds

パラメータリストを編集: configure bounds
Advanced parameters that can be set.

attribute name	values (Min, Max)
Steam_flow	70.0 110.0

エントリを追加 エントリを削除 適用

変数の値の上限、下限を設定可能
実工程での制約を反映できます。

5. おわりに

最適化の可能性

本デモで使用した“Prescriptive Optimization”をご活用頂ければ、より応用的な最適化、例えば多目的最適化といった問題にも取り組むことができるようになり、課題解決の幅が一気に広がります。

今回、ご紹介したPrescriptive Optimizerの方法はあくまで一例です。RapidMinerでは、ただ予測モデルを作成するだけでなく、最適化やモデルをディプロイし運用することもできます。