チートシート：KNIMEワークフローの作り方　ビギナー向け

Open for Innovation KNIME

KNIME アナリティクス・プラットフォームのはじめ方

・インストールガイドを参考に導入手順を実施します(knime.com/installation)。

・KNIME Analytics Platform のインストール完了後、７つのチェック項目を確認します(knime.com/blog/seven-things)。

・Eラーニングコースの受講をおすすめします(knime.com/knime-introductory-course)。

・KNIME ハブからワークフロー、ノード、コンポーネントの例を見てみましょう。(hub.knime.com)

・信号機表示システムを理解しましょう。

未設定：ノードは未設定状態でそのままでは実行できません。

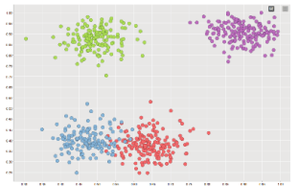
設定済：ノードは適切に設定されておりいつでも実行可能です。

実行済：ノードは正常に実行されており、結果はダウンストリームノードで閲覧可能です。

可視化

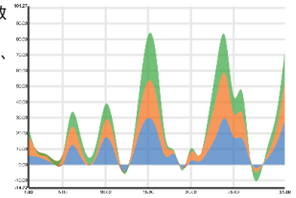
(KNIMEの可視化はすべてインタラクティブ形式です)

Scatter Plot：入力されたデータを二次元上の点で表現する散布図です。行・列や軸の選択、色や形などは設定ウィンドウで編集できるほか、ノードビューでも随時調整できます。



Sunburst Chart：リング状の階層構造によりカテゴリの内包関係を表現します。各リングは各変数の大きさに従って分割されて表示されます。多変量解析で特に効果的な可視化となります。

Stacked Area Chart：複数の数値データ系列を積み重ねて表示します。線で区切られた領域は、比較がしやすいように色で塗り分けられます。トレンドを表現するときにこのチャートがよく使われます。



Line Plot：Line Plot：数値データ点を線でつなぎ、X軸とY軸の関係を表現する線グラフです。例えばX軸に時間をおけば、グラフは時系列変化の表現となります。

Color Manager：Color Manager：入力データの値に応じて色属性を割当てます。色属性は後続の可視化表現に反映されます。

Pie Chart：Pie Chart：円を色つきの領域に分割する円グラフです。各部分の面積は各カテゴリの大きさに比例しています。

Data Explorer：Data Explorer：入力データの各種統計値とヒストグラムを用いたインタラクティブな可視化結果を出力します。

Box Plot：Box Plot：四分位数を使って数値を表現します。ヒゲの端の部分で外れ値の検討を行うことができます。

Bar Chart：Bar Chart：長方形の高さで各変数の大きさを表現する棒グラフです。複数の変数を同時に描画することができます。

分析

Decision Tree：Decision Tree：Learner ノードは C4.5 または CART 決定木の学習を行います。設定ウィンドウでは、枝刈り、打ち切りなどさまざまな設定ができます。Learner ノードと Predictor ノードはどちらもインタラクティブな画面を持ち、決定木に入力データが伝搬していく様子を可視化することができます。

k-Means：k-Means：k-Means クラスタリングアルゴリズムを実行します。事前にクラスタ数の設定が必要です。このノードはクラスタを生成します。Cluster Assigner ノードは入力データに最も近いクラスタを抽出します。

Logistic Regression：Logistic Regression：Learner ノードはロジスティック回帰モデルを学習し、予測値を計算します。設定ウィンドウではソルバや特徴量の選択、過学習の抑制などのさまざまな調整ができます。

Scorer：Scorer：適合率、F1値、カッパ係数など、分類器の精度を定量化するさまざまな指標値を計算します。

Numeric Scorer：Numeric Scorer：平均二乗誤差、平均絶対誤差、決定変数R^2など、予測値の精度を定量化するさまざまな誤差関数を計算します。

ROC Curve：ROC Curve：二値分類問題で受信者動作特性(ROC)曲線を描くために使用します。２つのクラスからランダムに一点を選び、確率とスコアを計算する操作を繰り返すことで ROC 曲線を生成します。

他の多くのオープンソースのデータ分析ツールとの併用も可能です。例えば、H2O、Weka、Keras、Spark、MLLib などの GUI を備えた KNIME ノードが使用できます。また、R、Python、Java などを用いて開発やデバッグをするためのノードもサポートされています。

読込み

File Reader：File Reader：テキストファイル全般、特にCSVなど区切り文字のあるファイルの読み込みで使用します。テキストデータの読み込みで良く使われます。

Excel Reader (XLS)：Excel Reader(XLS)：Excelファイル(XLS, XLSX)の読み込みに使用します。読み込み対象のシートやセルは設定ウィンドウで制御します。

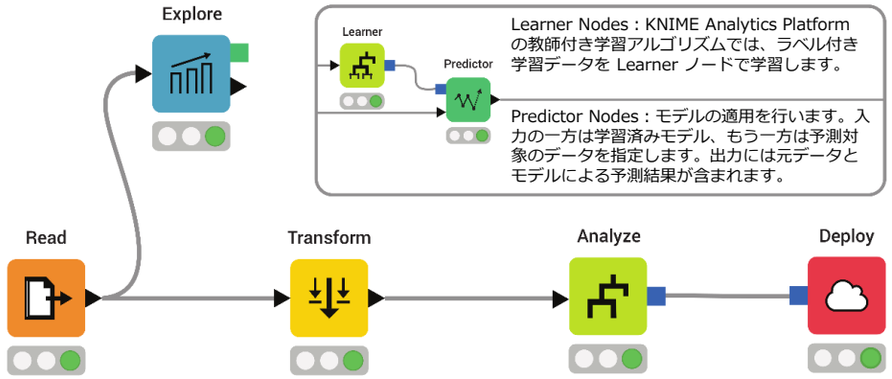
Table Creator：Table Creator：手動でデータテーブルを作成することができます。設定ウィンドウ上のデータシートでコピー＆ペーストなどの操作が可能です。小規模なデータ作成に便利です。

Model Reader：Model Reader：機械学習モデルの読込みを行います。どの Learner ノードで生成したモデルも読込み可能です。

Table Reader：Table Reader：「.table」ファイルの読込に使用します。「.table」は KNIME の独自フォーマットで、省メモリ・高速処理に最適化され簡単な設定で最大限のパフォーマンスが引き出せるよう考慮されています。

Google Sheets Reader：Google Sheets Reader：Google Sheet ファイルの読込みを行います。Google サーバ接続時に認証が必要です。ID・パスワードは KNIME ワークフローには保存されません。

knime:// プロトコル：KNIME インストールディレクトリからの相対パス等でファイルを参照します。「knime://knime.workflow/../<filename>」や「knime://<knime.server.mountpoint>/<path>/<filename>」のように使用します。



変換

GroupBy：GroupBy:選択された列の中で、ユニークな値を用いてテーブルの行をグループ化し、定義されたグループ単位で集計や統計量を計算します。シンプルなノード名ですが、強力な機能を有しさまざまな使い方ができます。例えば、行の重複削除などが可能です。

Pivoting：Pivoting:GroupByノードの集計機能を拡張し、列と選択された入力データ列の固有の値を持つ行で構成されたデータテーブルを生成します。(注:グループ化している列の固有の値は行に格納され、ピボット列の固有の値が列になります。

Rule Engine：Rule Engine:入力データテーブルのそれぞれの行に対し、ルールを適用します。本ノードで行えるすべての操作は、Column Expressions ノードでも使用できます。

Partitioning：Partitioning:指定した方法でデータテーブルを二分割します。このノードは一般的に、機械学習モデルを評価する際の学習・検証データ生成のために用いられます。

Row Filter：Row Filter:ルールに従って、入力データテーブルの行を含めたり除外したりします。フィルターのルールは、選択された列の値または指定した数値の範囲で記述します。

Math Formula：Math Formula:複数の入力列に対し、和や平均を計算したりlogやexpの変換のような数学的な処理を実装します。本ノードで行える処理は、Column Expressions ノードでも使用できます。

String to Date&Time：String to Date&Time:文字列の値をDate&Time値に変換します。文字列に含まれるDate&Timeの形式は、手動で定義することも、自動で判別させることもできます。

Cell Splitter：Cell Splitter:選択された列の値を二つ以上の部分文字列に、定義された区切り文字で分割します。区切り文字には、カンマやスペース、その他文字、文字列が使えます。

Column Filter：Column Filter:ルールに従って、入力データテーブルの列を含めたり除外したりします。保持する列は、手動で選択したり型や列名に対する正規表現で選択したりできます。

Column Rename：Column Rename:ダイアログの設定に基づき、選択された列の名前や型を新しく割り当てます。

Joiner：Joiner:一つ以上のキー列の共通の値に基づいて、二つのテーブルの行を結合します。内部結合、左・右外部結合、完全外部結合といった一般的な結合方法を使うことができます。

Sorter：Sorter:選択された列の値によって、昇順または降順にテーブルを並び替えます。並び替えに使用する列は複数選択することもできます。

Concatenate：Concatenate:同じ名前の列同士で、二つのテーブルを垂直方向に結合します。どちらかのテーブルにしかない列は、その値が欠損値で埋められます。結合するデータテーブルは一度に４つまで指定できます。

Missing Value：Missing Value:入力データテーブルの欠損値を処理する方法を定義するノードです。全列に対して同一の処理を定義することもできれば、それぞれの列に個々の処理を定義することもできます。

String Manipulation：String Manipulation:文字列型の列に対して、二つ以上の文字列連結、一つ以上の文字列抽出や空白の削除などができます。全ての操作は、Column Expressions ノードでも使用できます。

出力

Data to Report：Data to Report:BIRT(KNIMEに組み込まれているOSSのレポートツール)に出力するデータテーブルを選択するノードです。KNIMEからBIRTへ切り替える際に、選択されたデータセットはBIRTに入力されます。Image To Report ノードは、BIRTに出力する入力画像を選択します。

Excel Writer (XLS)：Excel Writer(XLS):入力データテーブルをExcel形式のファイル(XLS、XLSX)に書き込むためのノードです。

Table Writer：Table Writer:入力データテーブルをKNIME専用の .table 形式でファイルに書き込みます。この形式はファイルの全構造情報を持ち、メモリや処理速度の点で最適化されています。ファイル内にテーブル構造を持たせておくことは、ユーザー間でデータファイルをやり取りする際の利点となります。

CSV Writer：CSV Writer:入力データテーブルをcsvファイルに書き込みます。

Google Sheets Writer：Google Sheets Writer:入力データテーブルをGoogle Sheetファイルに書き込みます。Google サーバ接続時に認証が必要です。ID・パスワードは KNIME ワークフローには保存されません。

Send to Tableau Server：Send to Tableau Server:入力データテーブルをTableauファイルやレポート生成サーバーに出力します。

その他の情報

・KNIME Forum:グローバルなコミュニティに参加し、forum.knime.com で積極的に会話をしましょう。

・ KNIME Books:knime.com/knimepressで、より多くのヒント、アイディアや学びを得ましょう。

・ KNIME Event:学習コース、ワークショップ、ミートアップに参加しましょう(knime.com/events)。

・ KNIME Blog:話題のトピックス、技術的課題、業界のニュース、ノウハウなどを共有しましょう(knime.com/blog)。

・ Workflow Hub:ワークフローの例を閲覧したり、あなたが作成したワークフローを共有しましょう。workflow.knime.comのページで、他の人が作成したワークフローに評価やコメントを書きましょう。

・ More Guides:まだSASやExcelを使っていますか？knime.com/knimepress にある便利なガイドを参照して、KNIME Analytics Platform へ移行しましょう。

・ KNIME Server:チーム内の共同作業や自動化、管理、デプロイについては、knime.com/server でKNIME Serverに関する情報を確認しましょう。

KNIME社の承諾を得て、日立産業制御ソリューションズが内容を翻訳・掲載するものです。