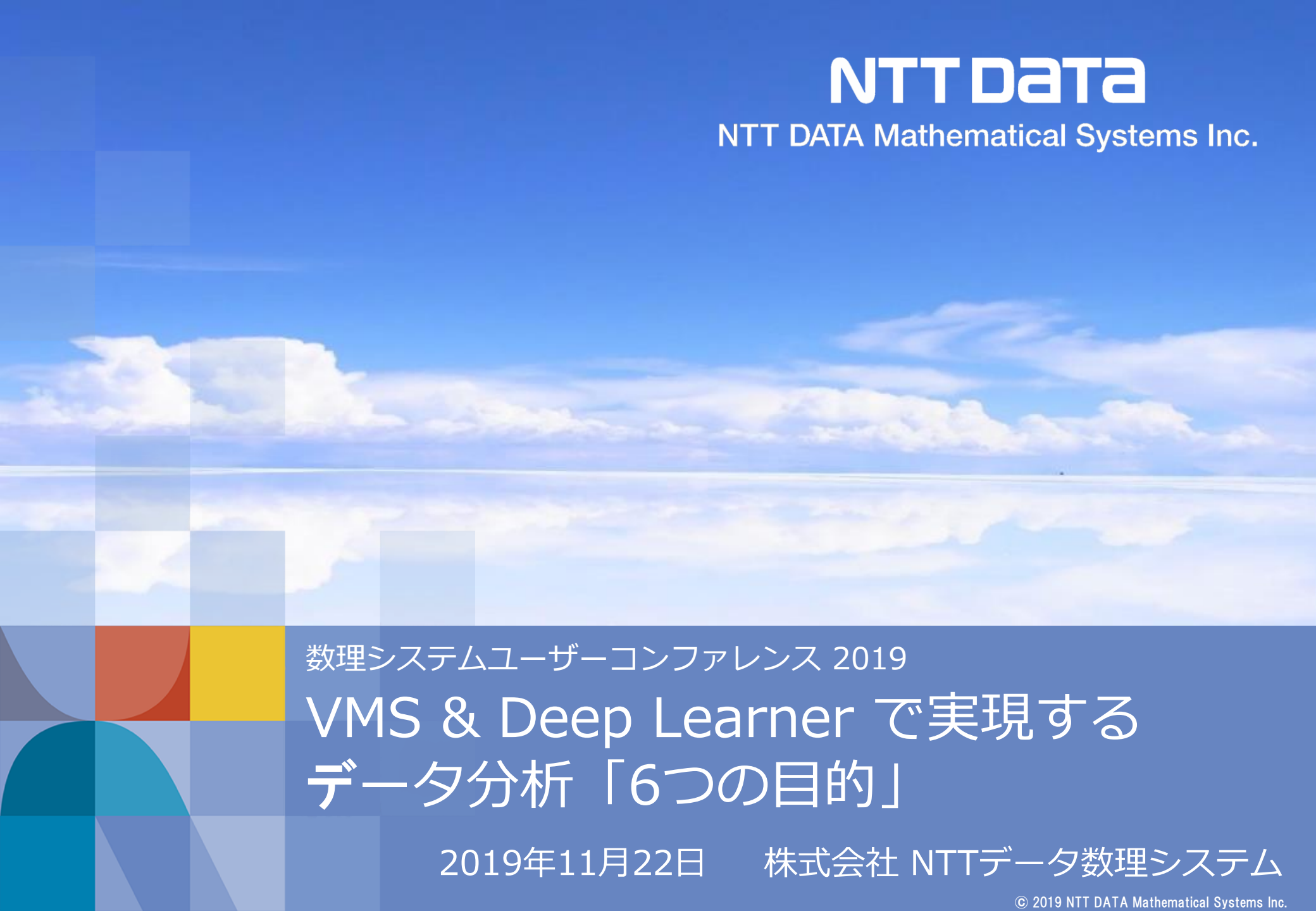




NTT data

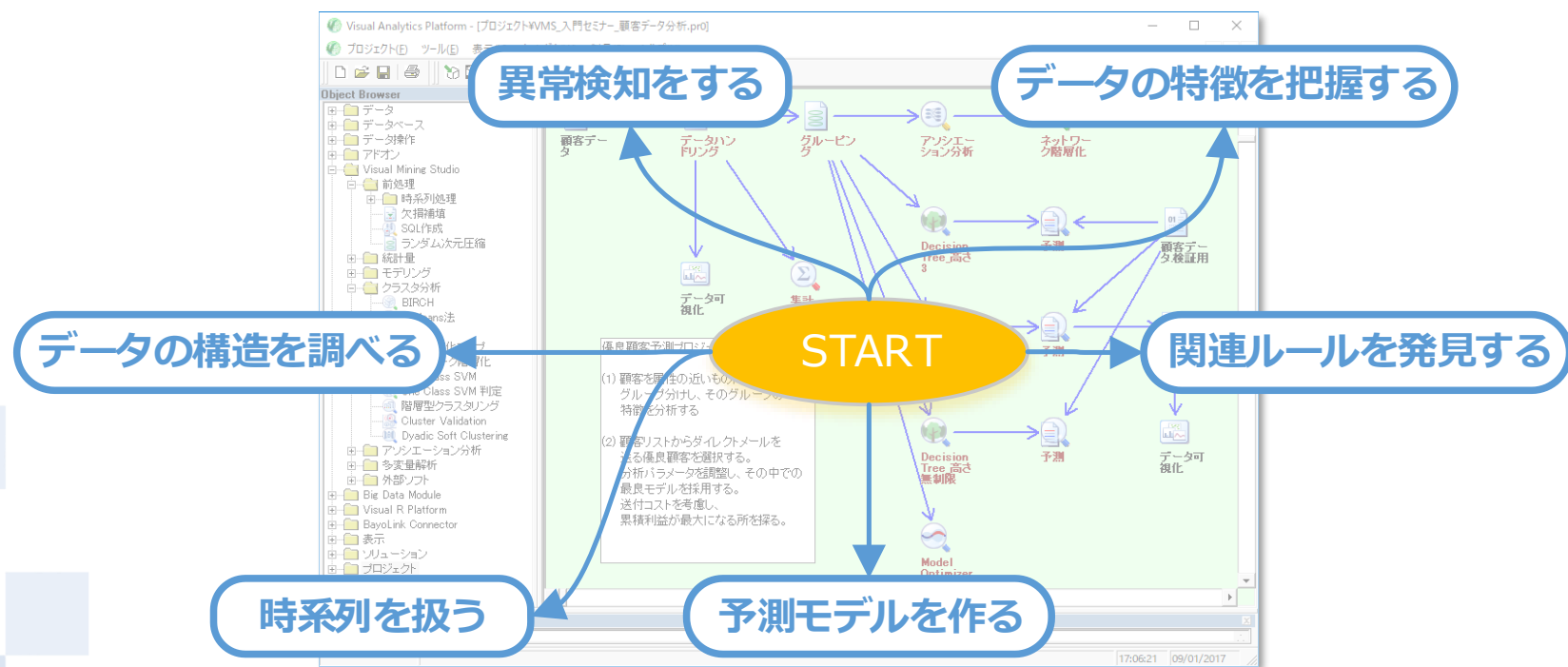
NTT DATA Mathematical Systems Inc.

数理システムユーザーコンファレンス 2019

VMS & Deep Learner で実現する データ分析「6つの目的」

2019年11月22日 株式会社 NTTデータ数理システム

長年皆様にご愛顧頂いております データマイニングツール
Visual Mining Studio に加え、2018年度には 深層学習デザインツール
Deep Learner をリリースし、当社は、真にお客様のデータ活用を
サポートし続けるべく製品に磨きをかけ進化を続けております。

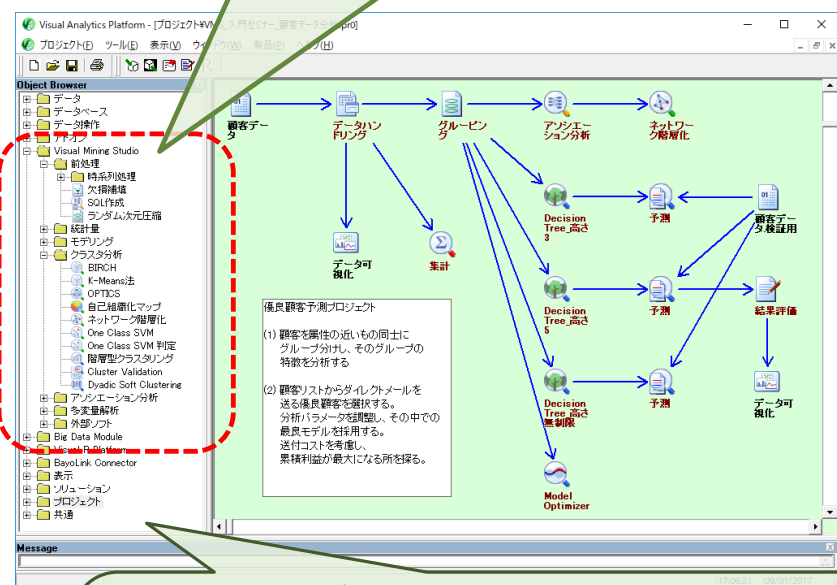


これらのツールが備える豊富な分析機能をご活用
いただくべく、上図のような **6つの目的別** に **VMS** および
Deep Learner の機能を総覧し、紹介させていただきます。
皆様にとって何か新たな発見がありましたら幸いです。

Visual Mining Studio

■ 特徴

定番から先進までを網羅
マイニング機能群



アイコン アイコン



と を矢印でつないで
処理の流れを定義する
ビジュアルプログラミング環境

業界最多 充実の分析メニュー

- 分類分析・クラスタ分析・
アソシエーション分析 を中心とした、
多彩で実用的な分析機能群

大規模データへの対応

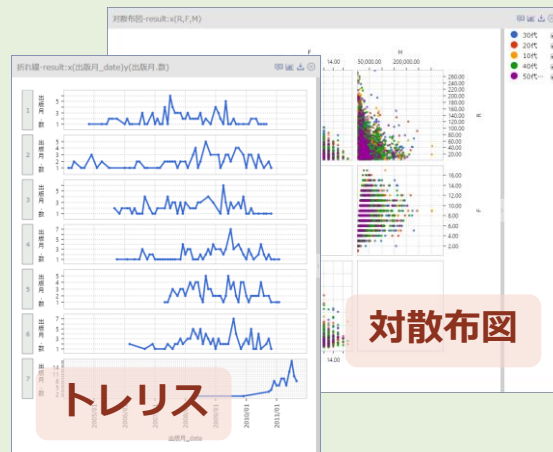
- 並列処理、64Bit対応 による
大量データの高速処理

マウス操作でデータ分析

- アイコンで処理の流れを定義
- 簡単な操作で試行錯誤が可能

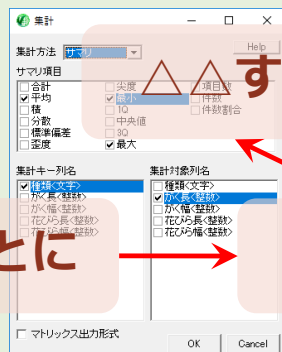
1. データの特徴を把握する

データ可視化機能が2017年にリニューアルしより強力に、よりモダンに生まれ変わりました。
表との連動機能も強化されており、強力なグラフ化機能がデータへの理解を助けます。



データは
どんな分布をしている？
項目間に関係はあるか？
などをまず、これらの機能群で確認し、掘り下げていきます。

集計は分析の基本であり最も利用頻度の高い機能です。



項目 **する**
対象列 **××を**
キー列 **○○ごとに**

データ可視化

集計

データの内容表示・
豊富なグラフ化機能

集計・
統計量算出

集計&グラフ化

変数間の相関を見る

顧客のグルーピング

群間の差が大きくなる
条件を抽出

有効な説明変数の探索

ビジュアル集計

相関

RFM分析

群間比較

Feature Selection

データの特徴を把握する

★は初心者の方にもおすすめ！

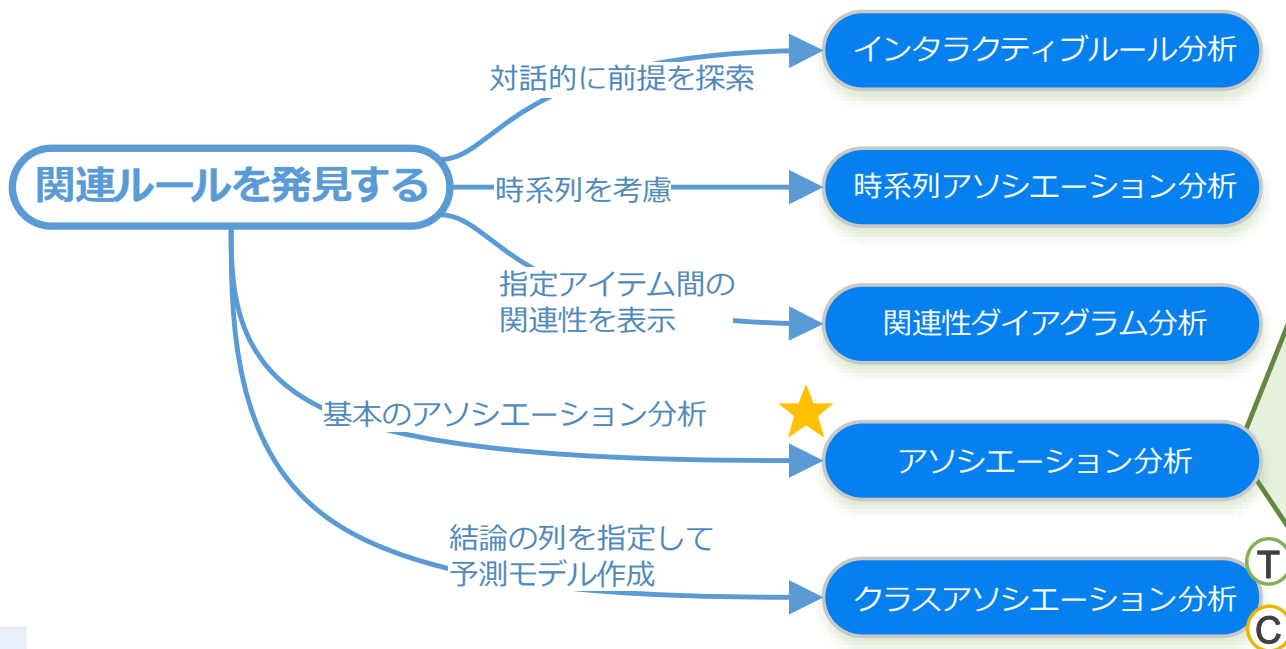
① 目的変数(教師データ)が必要

② カテゴリ値の目的変数に対応

③ 数値の目的変数に対応

2. 関連ルールを発見する

データの傾向把握を推し進め
項目間にどんな関連があるか？
そこから何らかのルールが抽出できないか？
といった観点で、
有用な知見の発掘を行います。

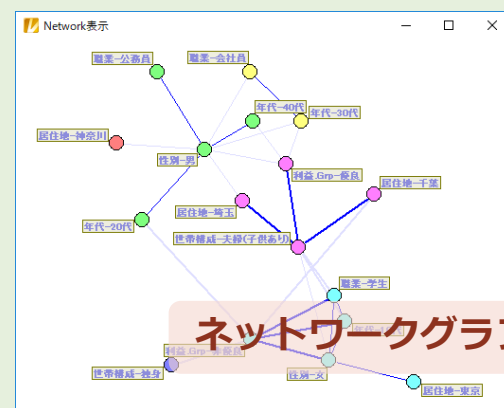


アソシエーション分析 は、

事象A が起きていた時に
事象B も起きている、
その確率や頻度が高い
事象A, B のペア情報

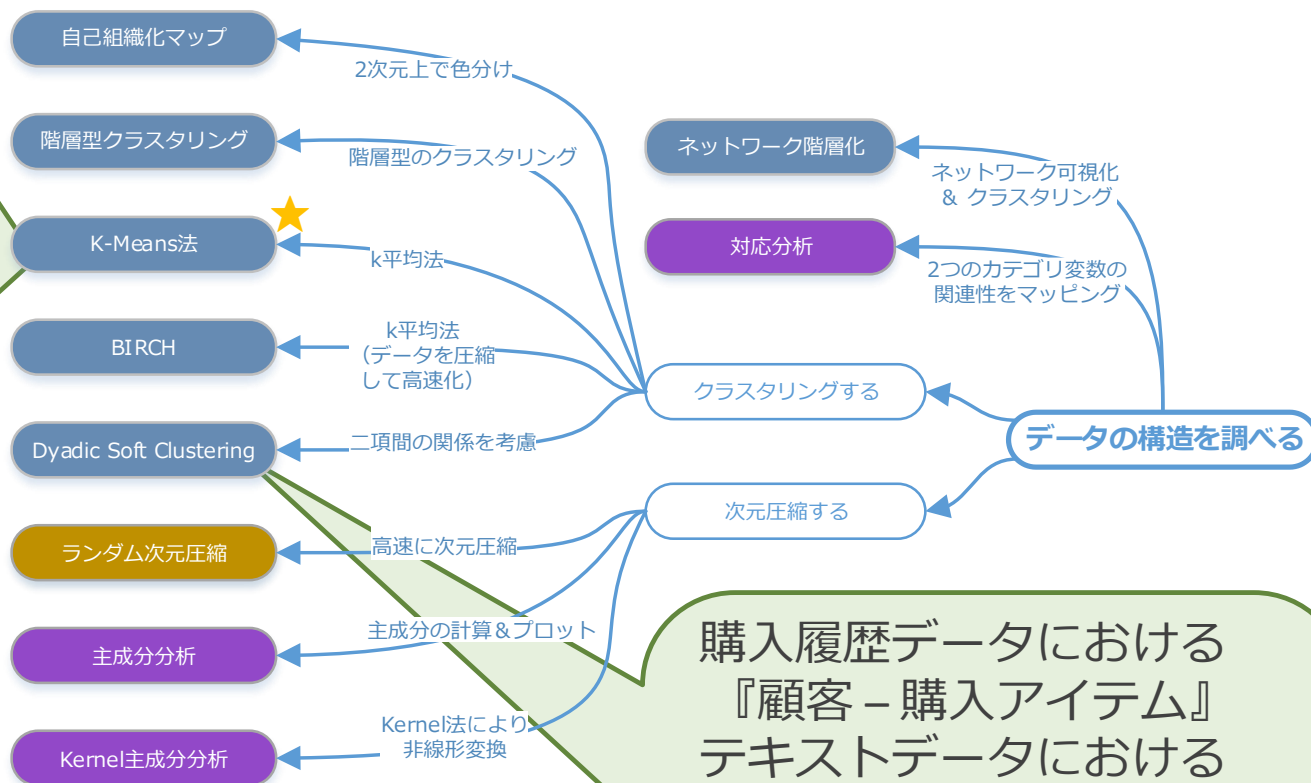
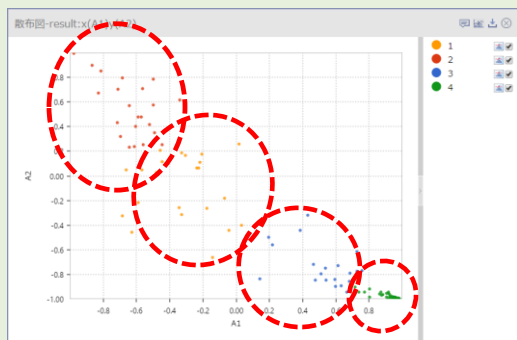
を抽出します。

結果を **ネットワーク階層化**
を用いて可視化することも
理解の上で大変有効です。



3. データの構造を調べる

K-Means法は類似したデータのまとめ上げを自動的に作成する**クラスタリング**手法の代表的なものです。



さらに踏み込んで、次のような考え方でデータ内に潜む構造の抽出を試みます。

類似するデータのグループを自動的に作れないか？
複雑なデータをもっと単純に表せないか？

購入履歴データにおける『顧客 - 購入アイテム』
テキストデータにおける『発言者 - キーワード』
など、**2種類の項目の組**で特徴づけられるデータの把握には**二項ソフトクラスタリング**が大変有効です。

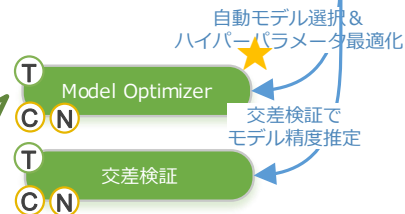
4. 予測モデルを作る

機械学習の手法群を用いて
予測モデルを作成します

例：正常/異常を判別（分類）

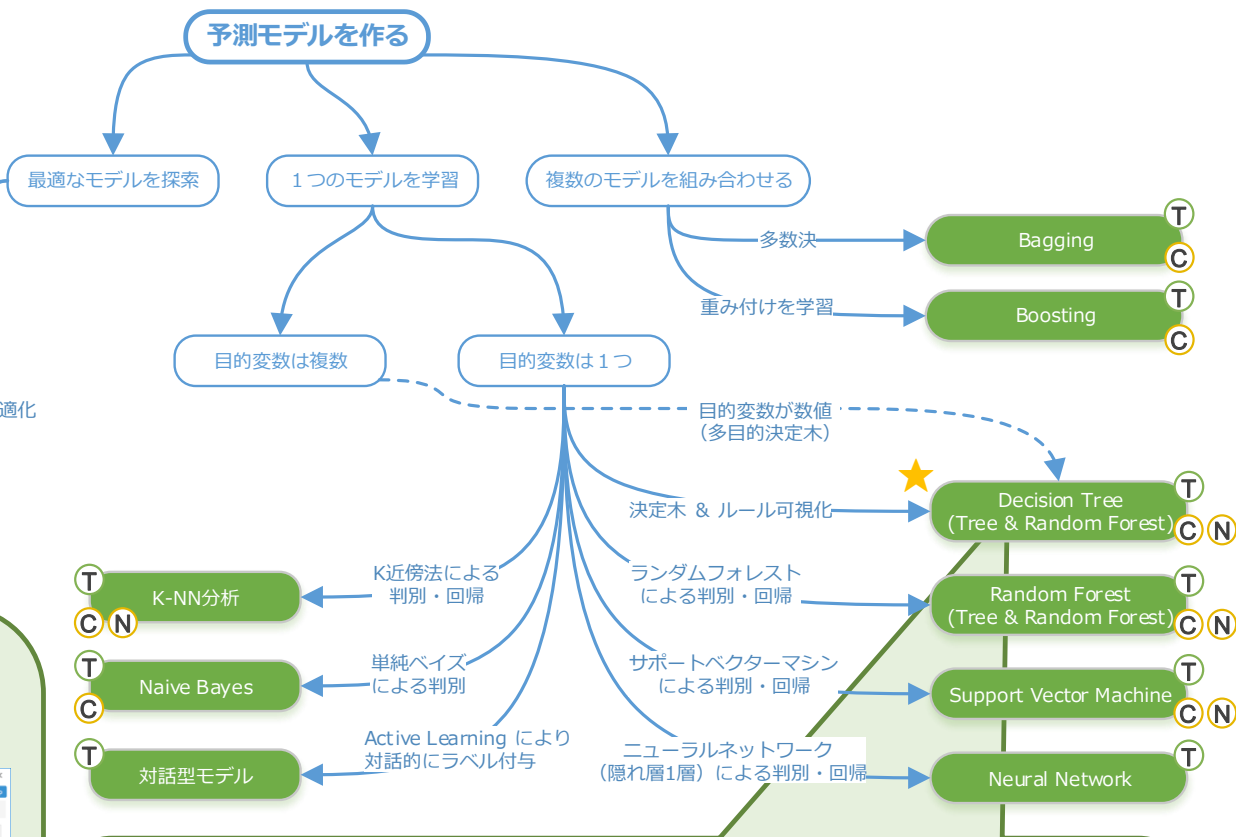
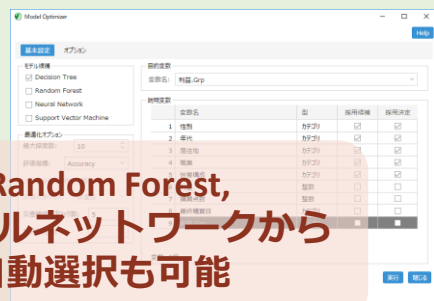
売上高を予測（回帰）

構築したモデルは
未知な状況の予測に
利用できます。

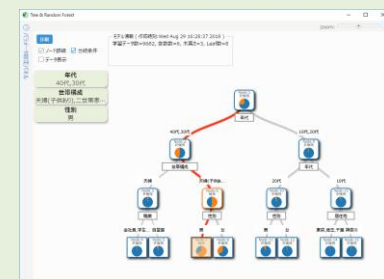


Model Optimizer では、
自動で効率的なパラメータ
探索を
実現します。

決定木, Random Forest,
SVM, ニューラルネットワークから
手法の自動選択も可能



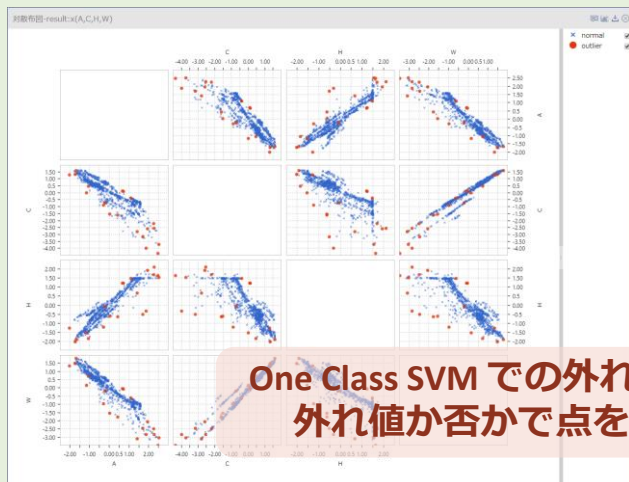
Tree & Random Forest では、
決定木 により 分類ルール の
抽出が可能、データの
傾向把握にも役立ちます。



5. 異常検知をする

特にデータに 正常・異常 のラベルが与えられていない
教師なし異常検知 の場合は、

データ全体の分布と照らし合わせて、外れているデータ
をいかに抽出するかという課題になります。



One Class SVM での外れ値抽出例
外れ値か否かで点を色付け

主要データと外れ値の2群に分ける
ことに特化したクラスタリング手法
であると考えすることもできます。

Isolation Forest

木構造ベースで高速処理

One Class SVM

教師なしSVM

OPTICS

密度ベース

異常検知をする

6. 時系列を扱う

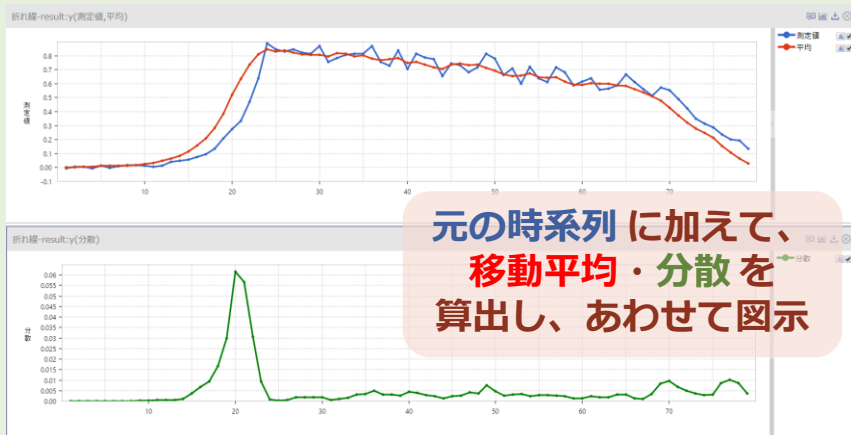
時系列データは、

各データ点同士が独立ではなく、その推移にも意味がある

という点に特徴があります。

時系列データの前処理・扱いに特化した機能を備えています。

時系列基本操作 機能で、移動平均や変動率などの情報を算出することができます。



時系列処理 / 基本操作

時系列処理 / 線分表現

時系列処理 / 距離計算

HSMM

推移列の展開・
移動平均等の計算

区分線形化

線分パターン
との距離

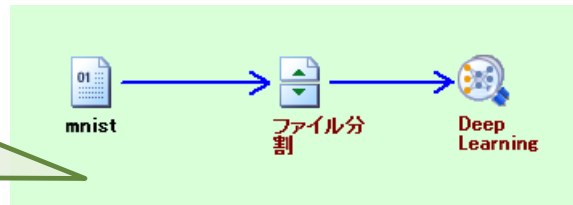
隠れセミマルコフモデル

時系列を扱う

Deep Learner

■ 特徴

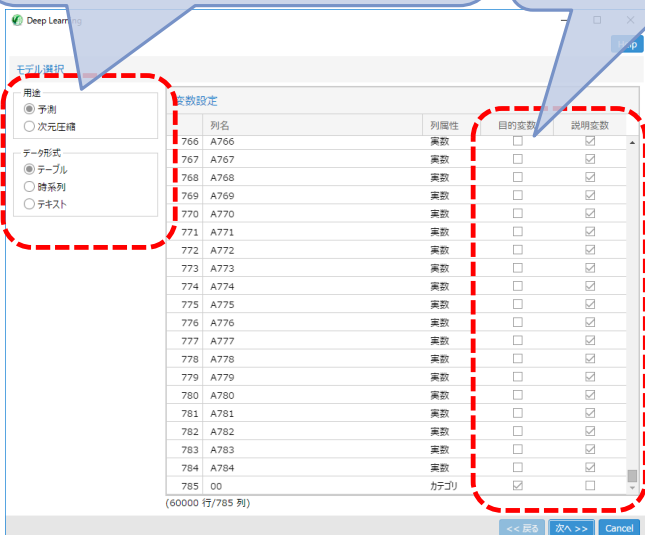
3ステップのマウス操作のみで
Deep Learning が利用可能



利用目的：
予測（教師あり学習）
次元圧縮（教師なし学習）
及びデータ形式を指定

説明変数・
目的変数を指定

パラメタ設定
Model Optimizer
自動チューニングも可能

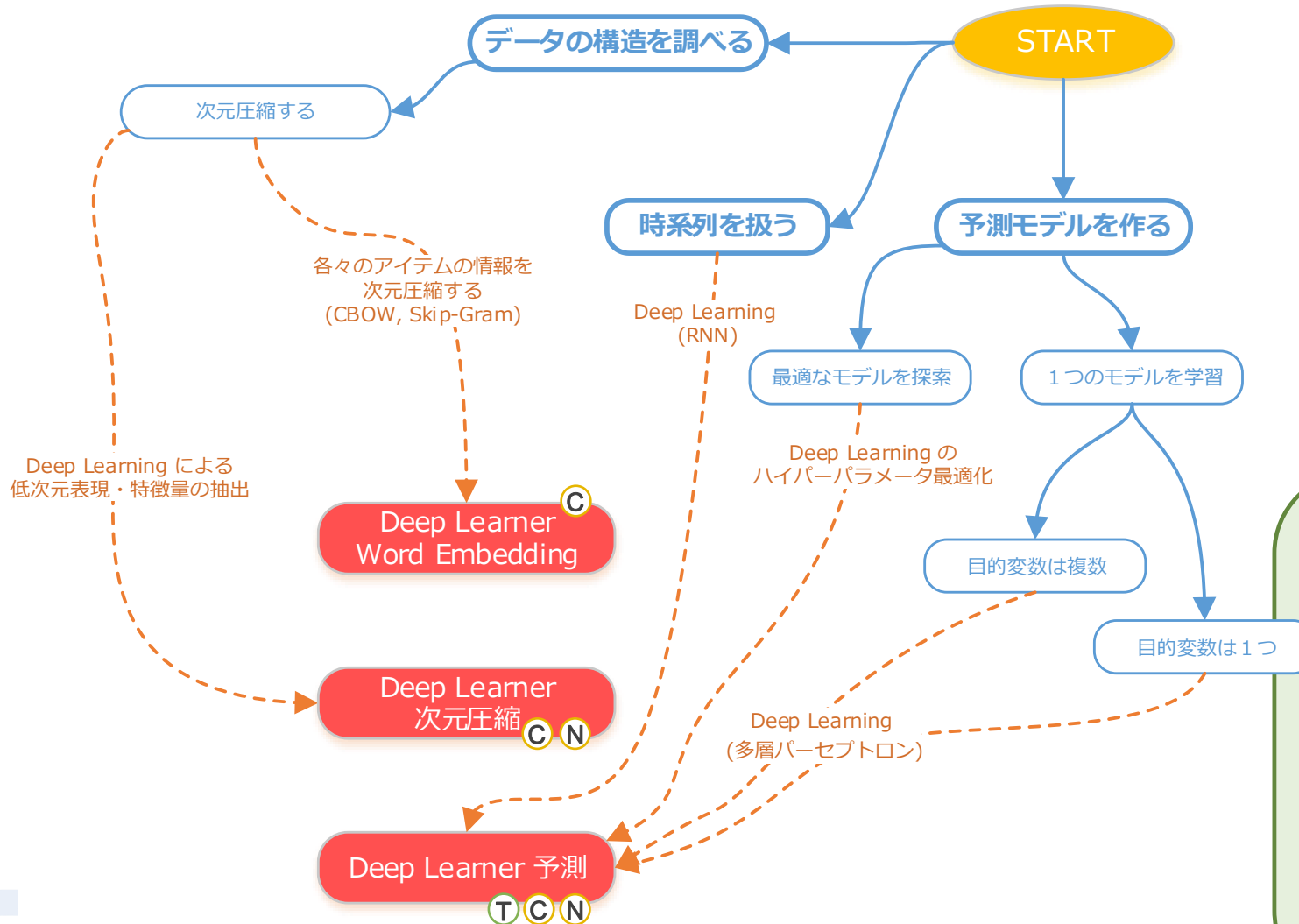


step1. モデル選択

step2. モデルデザイン

step3. 学習設定

Deep Learner がカバーする機能



学習の進捗状況は
リアルタイムに
グラフ化されます



関連ソリューション **Big Data Module**



Big Data Module

大規模な専用
コンピュータ
がいるんで
しょう？

ビッグデータ分析に
こんなイメージを
持っていませんか？

プログラムが
書けないと
分析できない
んでしょう？

操作が
難しいん
でしょう？

いいえ。BigDataModuleなら、プログラミング不要。
マウスだけの簡単操作でビッグデータ分析ができます！



【使用例】 大規模データの数値予測



売上予測



株価予測



電力需要予測

予測精度を高めるには...

- データ数を増やす
- 説明変数を増やす



計算時間の
爆発的な増加で
計算不可能

Big Data Moduleなら...

オンライン線形回帰でビッグデータでも予測可能

- データ数の線形オーダーの計算時間→**超高速**
- データ数に依存しないメモリ使用量→**超省メモリ**



簡単な使い方でビッグデータ分析

オンラインアルゴリズムによる
ビッグデータ分析アイコン群

Visual Mining Studio 同様
アイコンを繋ぐだけで簡単分析

データベースやHadoopからはもちろん、
大量のCSV形式ファイルから
ビッグデータを読み込んでの分析も可能

Pythonで処理が
記述できます

Pythonスクリプトアイコン



NTT DATA

NTT DATA Mathematical Systems Inc.

**Visual Mining Studio / Deep Learner
Big Data Module 担当**

TEL : 03 - 3358 - 6681

FAX : 03 - 3358 - 1727

【URL】 <https://www.msi.co.jp/vmstudio/>

【E-mail】 vmstudio-info@msi.co.jp

Trusted Global Innovator

NTT DATA Group

NTT DATA