

簡単・低コストですぐに使える
データマイニングツール
“シンプルデータマイナライト”

- VMStudio決定木モジュール採用 -

平成19年11月22日

NTTアドバンステクノロジー株式会社

国分 利直

(toshinao.kokubun@ntt-at.co.jp)
(<http://www.sotomiru.net/DM/>)

開発の経緯 - 手法的な要件 -

NTT通信設備の需要予測 投資計画、納期短縮化、等
...ライフサイクルが長い、ダイナミックな動きが少ない
お客様のダイナミックな動きを多種 / 膨大データの中から
捉えることが必要
...時間軸、空間軸
新規加入者の予測
...予測手法としてデータマイニング

開発の経緯 - 業務的な要件 -

現場の営業・企画部門が簡単に使用したい
自社内の業務で完結したい
効果の評価手法が不明確であった



- ・専門ツールしか無かった...高機能 / 高価 / 高スキル
- ・他社への委託...守秘性 / 機密性での課題、高コスト
- ・業務フローに沿えるツールが無かった

開発の経緯

- 汎用ツールの開発 -

精度を落とさず、機能限定で簡易化

...決定木分析手法の採用

...モデルの生成経緯を読むことができる

分析業務フローへの適合化(オールインワン化)

...データ加工～分析～ターゲティング～評価

本ツールの特長

予測と**分類**を決定木分析手法により簡易に実行できる。

ウィザード形式に沿った流れで分析用データの作成、分析、ターゲティングの一連の操作を実行できる。

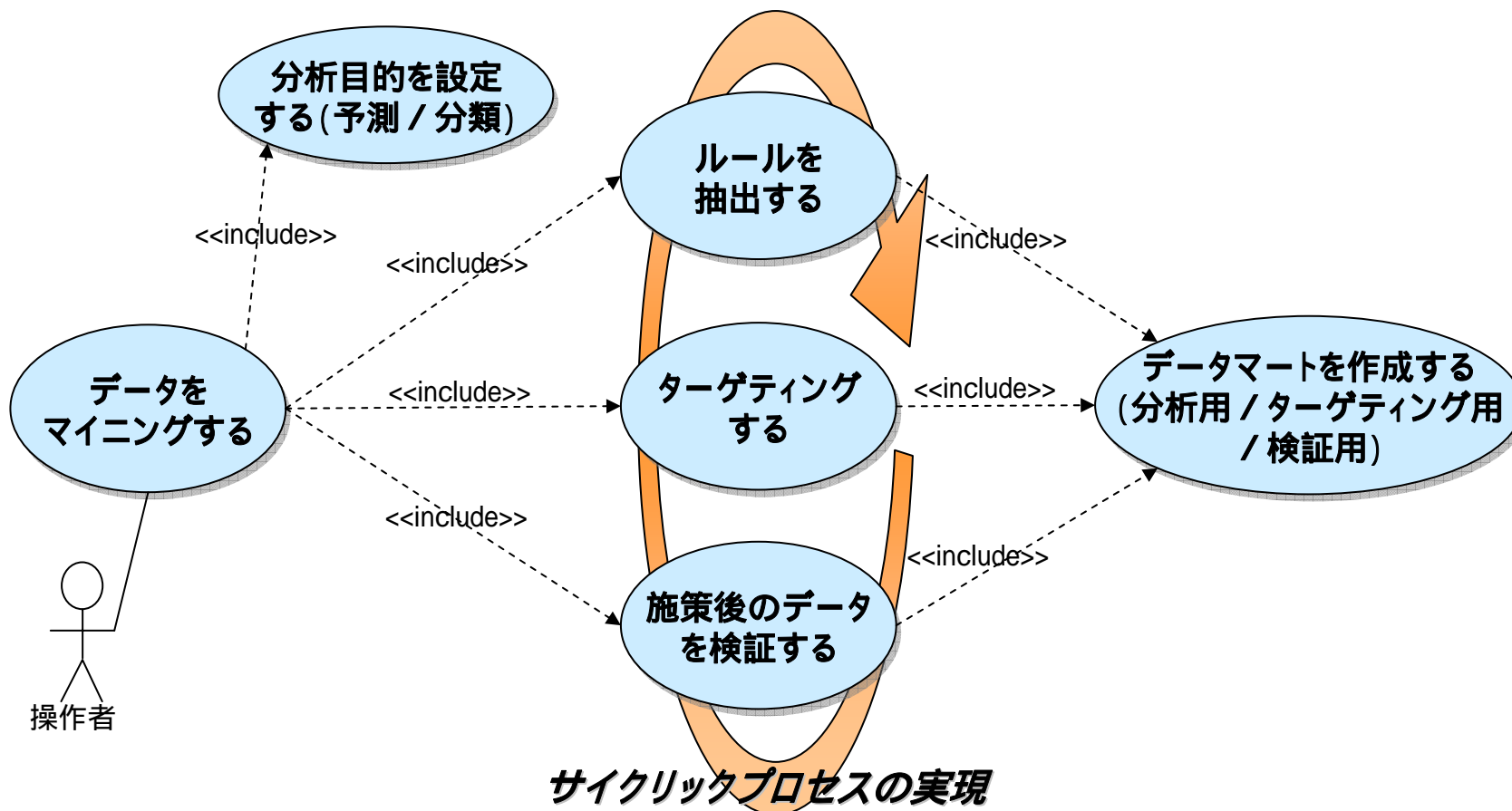
タブ形式を用いた画面の移動により簡易に分析やターゲティング条件を変更し、繰り返し実施できる。

複数の分析項目(説明変数)をグルーピングして“説明変数セット”とし、**セット単位**あるいは**セットを複数組み合わせ**て分析できる。

分析結果(樹形図やリフトカーブ)をエクセルファイルとして出力できる。

複数の**分析モデルを複合化**してターゲティングに用いることができる。
ターゲットリストをCSV形式で簡易に出力することができる。

構成コンポーネント



分析目的

ルール抽出ウィザード

新規プロジェクト
新しいプロジェクトを作成し、ルール抽出を実行します。

プロジェクト名

分析目的

- 予測
取得時期の異なる2つのデータを用い、データを予測する場合に指定します。目的変数は2値の名義値または連続値に対応しています。
- 分類
1つのデータを用い、データを分類する場合に指定します。目的変数は多値の名義値または連続値に対応しています。

キャンセル(C) 戻る(B) 次へ(N)

予測と分類の2種類の分析を実施することができます。

取得時期の異なる時系列データを用意することにより予測分析を行うことができます。

ウィザード操作

ルール抽出ウィザード

ルール抽出用データソース

ルールを抽出するために使用するデータの場所を指定します。

データソース(t0)

CSVファイル 1行目をヘッダとして読み込む

ODBC データソース名

ユーザーID パスワード

テーブル名

キャッシュを使う

データソース(t1)

CSVファイル 1行目をヘッダとして読み込む

ODBC データソース名

ユーザーID パスワード

テーブル名

キャッシュを使う

キャンセル(C) 戻る(B) 次へ(N)

ルール抽出ウィザード

データ詳細

データに含まれている変数の属性を定義します。

使用	ID	目的変数	変数名	データ	変数型	説明
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	ID	ID00001,ID00002...	名義	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	会員	(いはい)	名義	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	性別	女男	名義	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	年齢	25,27,29,29,31,32...	連続	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	電力消費量	10000,100000,110...	連続	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	住宅種別	マンション,一戸建...	名義	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	居住地	県内,市外,市内	名義	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	駅からの距離	0-1 km,10+ km,1...	名義	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	TV台数	0,1,2,3,4,5	連続	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	インターネット有無	(いはい)	名義	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ペット数	0,1,2,3,4	連続	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	購買有無	無有	名義	

すべて使用する

キャンセル(C) 戻る(B) 次へ(N)

ルール抽出ウィザード

サンプリング

データの中からルール抽出で使用するサンプルを選択します。

名義値サンプリング

2対8

名義値	割合 [%]	サンプル数
有	10	1000
無	90	9000

サンプル数の上限値

連続値サンプリング

件数
 割合 [%]

連続値の変化量

差分
 比

固定シード
 ランダムシード

キャンセル(C) 戻る(B) 次へ(N)

ウィザード操作に沿って分析を進めることができます。

タブ操作

【ルール抽出】

使用	ID	目的変数	変数名	データ	変数型	説明変数セット
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		ID	ID00001,ID00002...	名義	
<input checked="" type="checkbox"/>	<input type="checkbox"/>		会員	い/いえ/はい	名義	顧客情報
<input checked="" type="checkbox"/>	<input type="checkbox"/>		性別	女/男		
<input checked="" type="checkbox"/>	<input type="checkbox"/>		年齢	25,27,28,29,31,32...		
<input checked="" type="checkbox"/>	<input type="checkbox"/>		電力消費量	10000,100000,110...		
<input checked="" type="checkbox"/>	<input type="checkbox"/>		住宅種別	マンション,一戸建...		
<input checked="" type="checkbox"/>	<input type="checkbox"/>		居住地	県内,市外,市内		
<input checked="" type="checkbox"/>	<input type="checkbox"/>		駅からの距離	0-1 km,10+ km,1...		
<input checked="" type="checkbox"/>	<input type="checkbox"/>		TV台数	0,1,2,3,4,5		
<input checked="" type="checkbox"/>	<input type="checkbox"/>		インターネット有無	い/いえ/はい		
<input checked="" type="checkbox"/>	<input type="checkbox"/>		ペット数	0,1,2,3,4		
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		購買有無	無/有		

すべて使用する

タブ操作に沿って指定した画面で条件等を変更して分析することができます。

【ターゲットティング】

説明変数セットの設定と使用例

The image displays two windows from the 'ルール抽出ウィザード' (Rule Extraction Wizard) application.

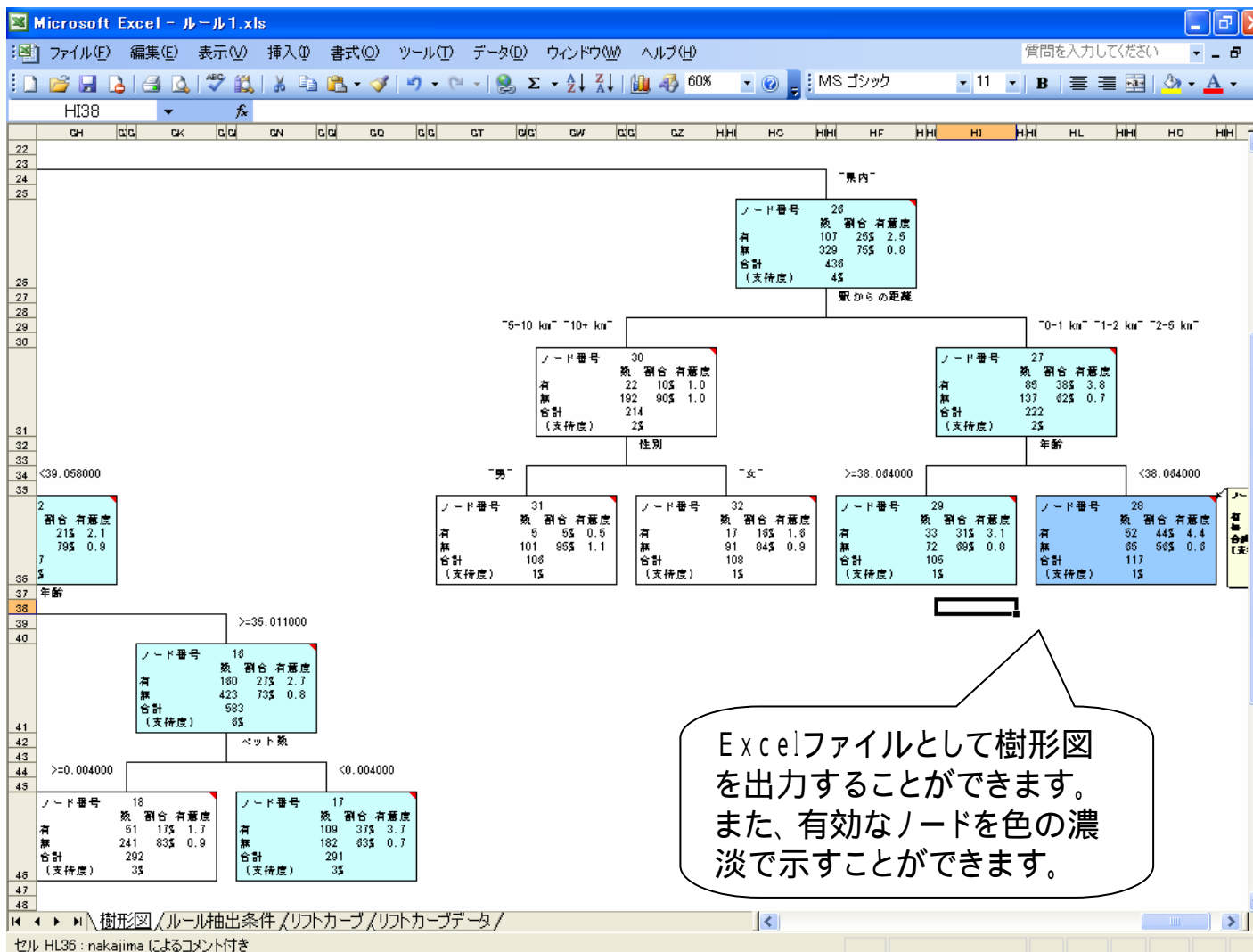
Left Window: 説明変数セット (Explanatory Variable Set)
 Title: 説明変数セット
 Instruction: 一まとまりの説明変数に名前を付けます。
 Field: 説明変数セット名 [エリア情報]
 List '含まない項目' (Excluded items): TV台数, インターネット有無, ペット数, 会員, 性別, 電力消費量, 年齢.
 List '含む項目' (Included items): 駅からの距離, 居住地, 住宅種別.

Right Window: ルール抽出条件 (Rule Extraction Conditions)
 Title: ルール抽出条件
 Instruction: 抽出するルールに名前を付け、抽出条件を設定します。
 Field: ルール名 [ルール(顧客)]
 Field: 備考 []
 Radio buttons: 説明変数, 説明変数セット
 List '左側': エリア情報
 List '右側': 顧客情報
 Fields: 子ノード最小数 [1.0] (数), 親ノード最小数 [2.0] (割合 [%]), アルゴリズム [Gini係数], 樹の深さ [6].
 Buttons: キャンセル(C), 戻る(B), 次へ(N)

Bottom Window: 説明変数セット (Explanatory Variable Set)
 Title: 説明変数セット
 Instruction: 一まとまりの説明変数に名前を付けます。
 Field: 説明変数セット名 [顧客情報]
 List '含まない項目' (Excluded items): 駅からの距離, 居住地, 住宅種別.
 List '含む項目' (Included items): TV台数, インターネット有無, ペット数, 会員, 性別, 電力消費量, 年齢.

Callout Box:
 説明変数のカテゴリ毎に複数の説明変数セットを設定し、ルール抽出の時に任意に組み合わせることができます。

樹形図出力例



まとめ

データマイニングソフトは高機能なツールであることは重要である一方、業務プロセスを完結させることがより重要である。

シンプルデータマイナライト は、多機能ではないが、営業前線ですぐに使える簡易操作性を有するツールである。

シンプルデータマイナライト は、マイニング分析業務プロセスの完結をサポートする低コストのツールである。

分析モジュールには、信頼性・柔軟性を有するVMStudio
決定木モジュールを採用している。