



NTT data

NTT DATA Mathematical Systems Inc.

ベイジアンネットワークへの誘い ～BayoLinkS 2019 年度の展望～

2019年11月22日
(株)NTTデータ数理システム

Trusted Global Innovator

NTT DATA Group

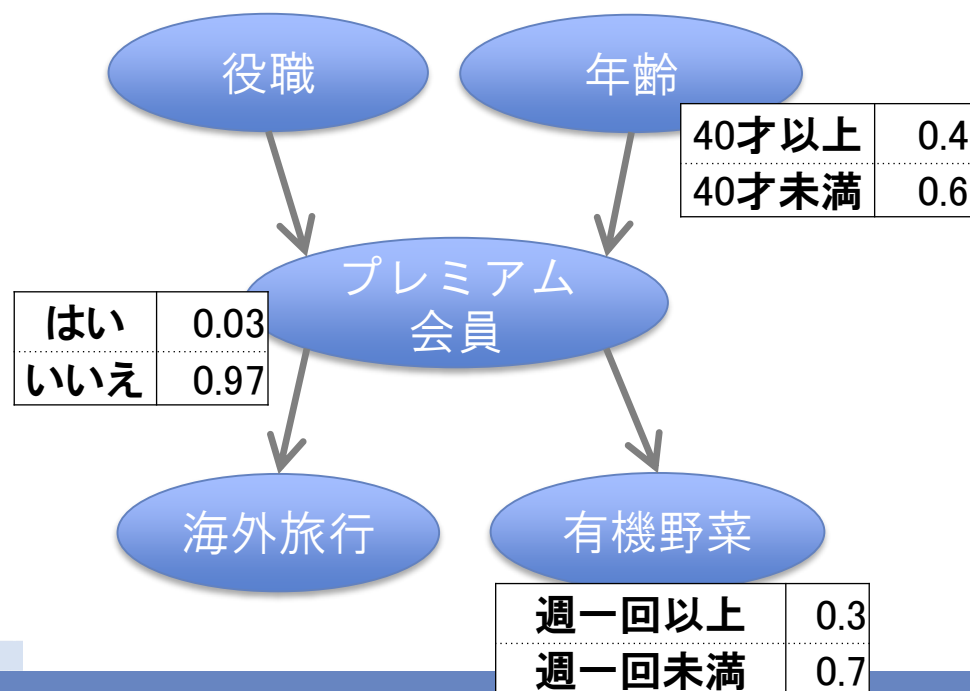
NTT data

© 2018 NTT DATA Mathematical Systems Inc.

ベイジアンネットワークとは？

- ✓ 予測モデルの一つ
- ✓ 事象間の確率的な関係をネットワークで表現する
- ✓ 観測された情報をモデルに入力し、確率を計算する
- ✓ 未知の状況の予測や分析を行う

□ ベイジアンネットの具体例：プレミアムカード会員モデル



『有機野菜』の条件付き確率

有機野菜 (購入頻度)	プレミアム = はい	プレミアム = いいえ
週一回以上	0.65	0.3
週一回未満	0.35	0.7

ベイジアンネットの特徴

✓ 自由度の高いモデル

- モデルに観測を入力し推論すると、**観測のない全てのノード**で事後確率が更新される
- 矢印の**逆方向**でも推論ができる
- 1つのモデルを使って、様々な方向で分析や予測ができる

✓ 柔軟性の高いモデル

- **欠損**があっても推論ができる

✓ 因果関係の仮説をモデルの構造で表現できる

- 変数間の因果関係を矢印の向きで表現できる
- 事前知識を取り入れることで、**より自然なモデル**を表現できる
- 推論の構造を**直観的**に理解できる

BayoLink の機能

- ✓ ベイジアンネットワークの構築・分析支援システム
- ✓ モデルの構築から検証、分析などの各工程を強力にサポート

学習データからモデルを構築・分析するまでの工程

データの前処理

- 数値データの離散化
- カテゴリデータのグループ化

機械学習

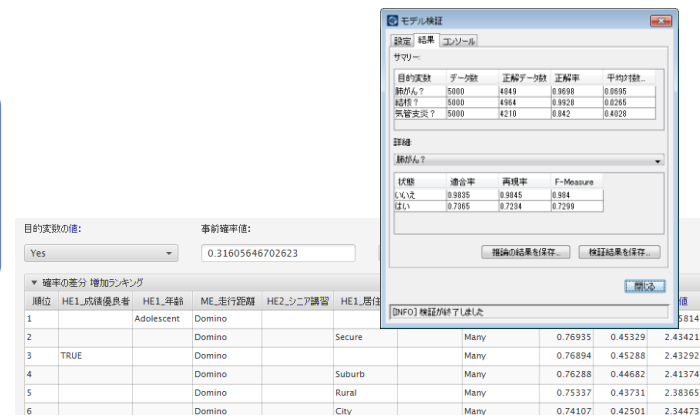
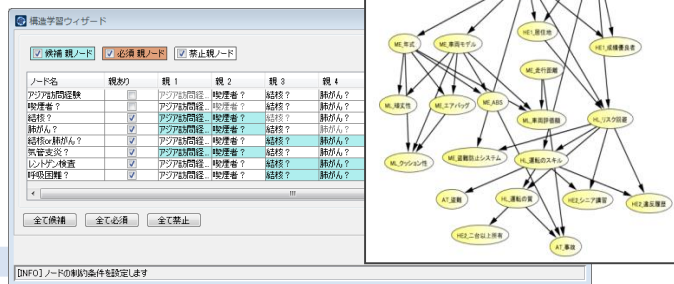
- 構造学習
- パラメータ学習

検証

- 予測精度などの指標値の算出

予測 / 分析

- 確率推論
- 感度分析



※2016年 BAYONET からBayoLink に名称変更

□ マーケティング分野

- アンケート分析を活用し ヒトの評価と行動をモデル化する
- 商品の関心が高いヒトの属性やライフスタイル、ターゲット層の絞り込みに活用。
- 従来のエクセルによるクロス集計分析より短時間での分析が可能。また膨大な質問項目の組合せの中から知見を得る。

□ 製造分野

- 製造機器のセンサデータや業務日報を紐付け、さらに現場のノウハウなどを取り込みモデル化する。
- 異常の発生から、どの部分に原因があるかを推論によって探る。

□ 医療/ ヘルスケア分野

- 症例データや生活習慣のデータを分析することで、疾病のリスクを推論。
- 診察結果や周辺データなどを階層構造でモデル化し、疾病の因果関係をわかりやすく表現する。
- 確率推論により、疾病のリスクを高める原因を探ることができ、リスクを回避するための具体的な行動を示すことが可能。

Visual Analytics Platform のご紹介

NTTデータ数理システムの分析ツールをシームレスに連携するプラットフォーム



各種分析ツールの機能をアイコンとして画面上に配置、矢印でつなぐだけで処理フローを完成。データ分析に必須な様々なデータ加工機能が搭載されています。また、データマイニング やテキストマイニングなどの他製品と組み合わせることで、分析の幅が広がります。

BayoLinkS

BayoLink

ベイジアンネットワークのモデル構築から確率推論まで

- ベイジアンネットワークモデルの自動構築
- 確率推論
- モデル検証
- モデル編集

開発元:(独)産業技術総合研究所

Visual Analytics Platform

データの前処理から、分析の実行まで

- 充実のデータ前処理機能・可視化機能
- 分析工程を俯瞰できるプロジェクトボード上での分析フロー作成
- テキストマイニング・データマイニングをはじめとする
NTTデータ数理システム他製品とのシームレス関係

BayoLinkS ～ VAP関係によるメリット①

BayoLink 単独より 高度な分析が可能に！

ビジュアルプログラミング環境

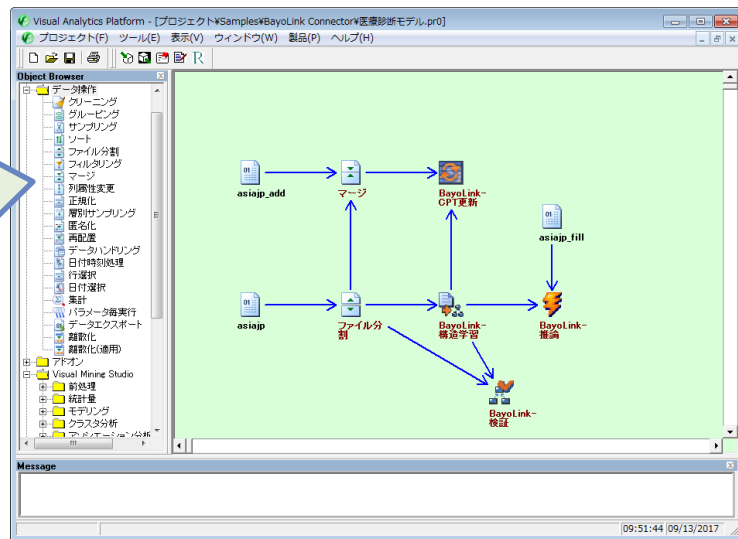
- ・ ベイジアンネット分析の全工程を一つの画面で！
- ・ モデル構築や分析フローを、マウスクリックで簡単に再現できる

豊富なデータ加工メニュー

- ・ マージ、クリーニングなどの基本的な加工機能の他、
層別サンプリングなどデータマイニングの手法を取り入れた加工も搭載。
- ・ ベイジアンネット分析には欠かせない、数値データの離散化を搭載

データ操作アイコン

- ・ クリーニング
- ・ マージ
- ・ ファイル分割
- ・ 匿名化
- ・ 層別サンプリング
など



離散化アイコン

等数分割、等区間分割、 χ^2 乗
分割などで数値データをカテ
ゴリデータに変換する



離散化

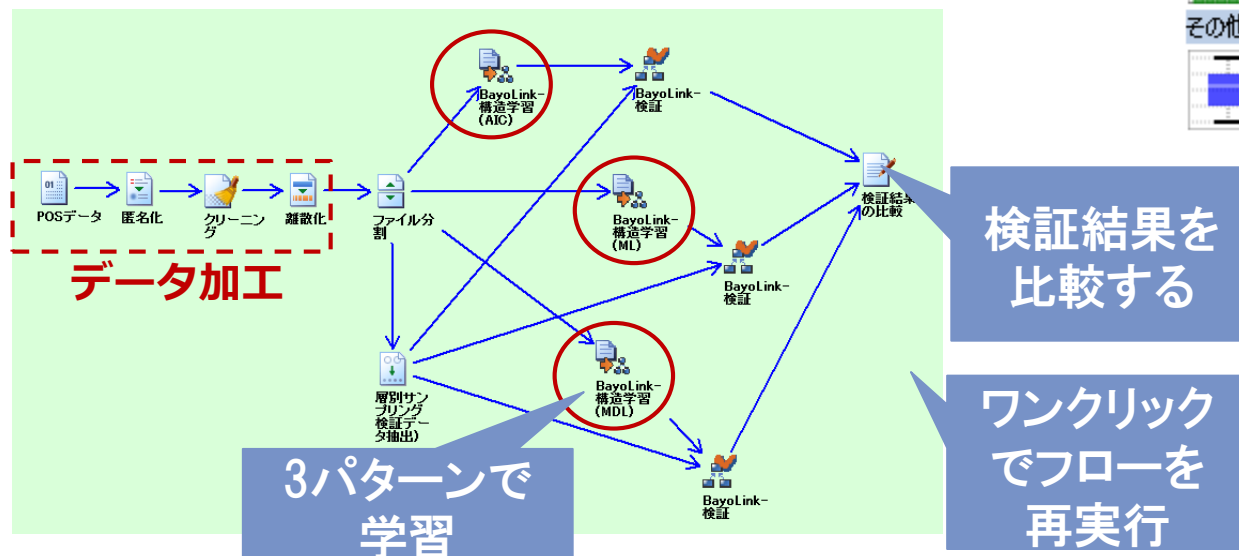
BayoLinkS ～ VAP関係によるメリット②

データ可視化

- 基本的なグラフの他、バブルチャートや箱ひげ図など多種多彩なグラフを描画できる。
- 集計アイコンで学習データに含まれる変数の分布を確認。

バッチ実行による分析の自動化

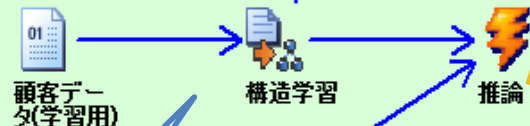
- 設定済みのアイコンやフローをコピーして、別の分析で再利用。例えば複数のパターンでベイジアンネットの構造学習を行い、それぞれのモデルを検証するようなフローも簡単に作成できる。
- また入力条件を変更しつつ、分析フローの繰り返し実行も可。



感度分析機能をVAPのアイコンとして新規搭載!!

また推論アイコンをリニューアル

- 目的変数を選択可能に
- アルゴリズム見直しで高速化



既存の顧客データから構造学習でベイジアンネットワークモデルを構築する。

新規顧客(予測)

新規顧客の属性から『収益』を予測する。また欠損している項目を予測して穴埋め。

	目的変数	説明変数	相互情報量
1	収益	職業	0.0751
2	収益	月収	0.0001
3	収益	年齢	0.0001

『収益』に影響する説明変数のランキング

	目的変数	種別	事前確率	事後確率	差分	リフト	職業	年齢	月収
1	収益	黒字	0.6228	0.8452	0.2224	1.3573	会社員	20~40代	25~45
2	収益	黒字	0.6228	0.8437	0.2209	1.3547	会社員	40~60代	25~45
3	収益	黒字	0.62			1.3535	会社員		25~45
4	収益	黒字	0.62			1.3519	会社員	20~40代	
5	収益	黒字	0.62			1.3493	会社員	20~40代	~25
6	収益	黒字	0.62			1.3352	会社員		
7	収益	黒字	0.62			1.3160	会社員	20~40代	45~
8	収益	黒字	0.62			1.3106	会社員		~25

『収益が黒字』となる可能性の高い変数の組合せ

実行速度が
従来版用より
大幅UP!!

説明変数の組合せ数	従来版	新規版
3	13 (分)	2 (分)
4	20 (分)	5 (分)

	顧客ID	性別	職業	年齢	月収	クレーム回数	利用時間	収益
1	1001	男	会社員	20~40代	~25	1~4回	5~50	黒字
2	1002	男	会社員	40~60代	45~	なし	5~50	黒字
3	1003	女	その他	20~40代	45~	なし	5~50	黒字
4	1004	男	公務員	20~40代	25~45	なし	5未満	赤字
5	1005	女	会社員	40~60代	25~45	1~4回	5未満	赤字
6	1006	男	会社員	40~60代	25~45	1~4回	5~50	黒字
7	1007	男	公務員	40~60代	~25	1~4回	5~50	赤字
8	1008	男	その他	20~40代	45~	1~4回	5未満	赤字
9	1009	男	会社員	40~60代	45~	なし	5~50	黒字
10	1010	女	会社員	20~40代	45~	なし	5~50	黒字

欠損しているセルを予測値で穴埋め

『収益』の予測値を追加

□ 2020 年 3月
不具合修正版リリース

□ 2020 年 秋
BayoLink ver.7.3 リリース
- 構造学習に新しい機能を搭載
(詳細はコンファレンスでご紹介します)



NTT DATA

NTT DATA Mathematical Systems Inc.

Trusted Global Innovator

NTT DATA Group

NTT DATA