

ベイジアンネットワークへの誘い

～データマイニングツールとの関係～

(株)NTT データ数理システム データマイニング部
石富 妙

1. はじめに

ベイジアンネットワーク(以下ベイジアンネットと略)は様々な事象間の因果関係(厳密には確率的な依存関係)をグラフ構造で表現するモデリング手法の一つで、故障診断、気象予測、医療的意思決定支援、マーケティング、推薦システムなど様々な分野で利用や研究が行われています。

本発表ではベイジアンネットについて概要を説明した後、**BAYONET**の機能や**Visual Mining Studio**との関係の強化についてデモを交えながらご紹介します。

2. ベイジアンネットワークについて

ベイジアンネットはベイズの定理を基本としています。ベイズの定理は二つの事象について直接の依存関係を表していますが、これをネットワークへと拡張したのがベイジアンネットです。ベイジアンネットは確率的に依存しているノードを矢印で結びます。また、依存関係の大きさは条件付き確率で定量的に表現します。

ベイジアンネットは次のような特徴があります。

- ネットワークで表現されるため、構造の理解が容易
ベイジアンネットは、確率的に依存関係にあるノードを矢印で結んだ有向グラフで表現されます。ノード間の依存関係を視覚的にわかりやすく表現できます。
- 予測において、説明変数の入力に欠損があってもよい
ニューラルネットワークなどと異なり、ベイジアンネットは予測時に全ての説明変数に観測値を入力する必要はありません。観測値のない変数については条件に応じた適切な確率分布が割り当てられます。
- モデルの利用用途が限定されない
ベイジアンネットは矢印の順方向だけでなく、逆方向にも推論が行えます。よって観測を入力する変数と予測対象となる変数がモデルで限定されることなく、自由に選択できます。

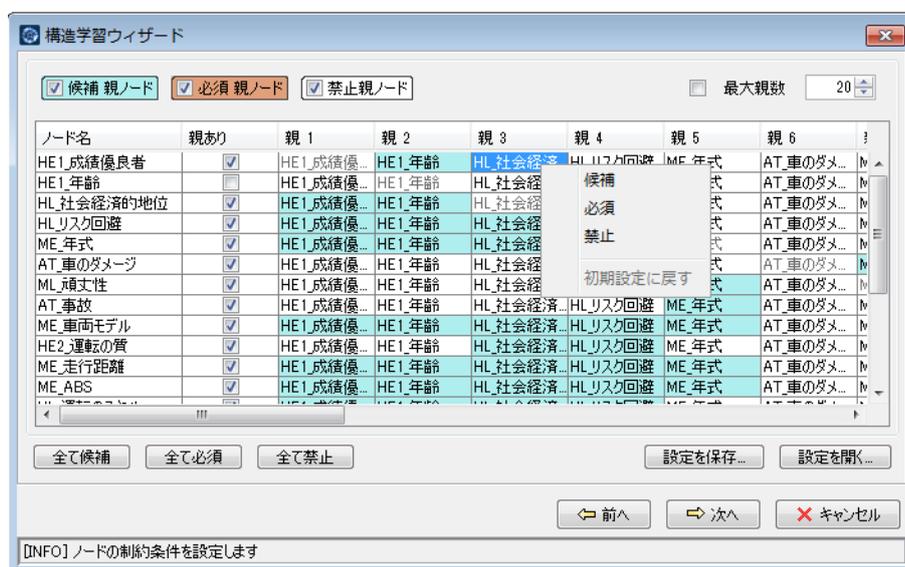
3. BAYONET について

BAYONET は(独)産業技術総合研究所によって開発され、NTT データ数理システムにおいて開発・販売を行っているソフトウェアです。BAYONET は、ベイジアンネットの構築、検証、モデルを使った予測など、モデリングに必要な機能を提供しております。

- 構造学習

学習データを使って、ベイジアンネットのネットワーク構造を機械学習のアルゴリズムより推定します。構造学習では確率的に依存しているノードを抽出し矢印で結びます。構造の推定はAIC などの評価規準により行います。

構造学習により因果構造を表現するモデルを作るには、ノードの親子関係に制約条件を指定しますが、操作しやすいインターフェースにて設定できます。



BAYONET の構造学習はマルチスレッドで実装されており、マルチコア環境を活かした高速な学習を実現します。

- モデルインポート機能

ベイジアンネットを構築するための学習データを BAYONET に取り込むための機能です。データインポートでは、データのカテゴリ化や数値データに対する離散化などの前処理機能も提供しております。

- モデル検証

構築したモデルについて、検証用データで推論を行い予測対象とする変数(目的変数)について正解率などの指標を算出します。目的変数は複数数を指定できます。指

標は正解率の他、目的変数の値ごとに適合率や再現率を算出します。

- 推論ツール(エクセルアドイン)

ベイジアンネットでの予測をエクセル上で実行する機能です。

エクセルシートに記入した説明変数の値で推論を行い、目的変数について事後確率や予測値をシートに出力します。

推論ツールにより、例えば次のような分析が可能です。

- 説明変数への入力の組み合わせにより、目的変数の確率分布がどのように変化するかを確認する。
- 目的変数に入力を設定し、説明変数の分布の変化を見ることにより有効な説明変数を絞り込む

4. データマイニングツールとの関係

BAYONET バージョン 6.1 (2013 年 9 月リリース) より **Visual Mining Studio** と関係ができるようになりました。**Visual Mining Studio** はデータマイニングにおける分析やデータ加工機能が数多く搭載されているソフトウェアです。

Visual Mining Studio はビジュアルプログラミング環境を特徴としており、データ加工や分析などの処理フローをマウス操作で簡単に作成することが可能です。

Visual Mining Studio では、分析処理をアイコンで定義しますが、**BAYONET** のアイコンとしては、構造学習、CPT 更新、推論、検証 の 4 つが実装されています。

- 構造学習

学習データを入力として、**BAYONET** の構造学習ウィザードで学習を行います。

- CPT 更新

構造学習アイコンで構築したモデルを、別の学習データで CPT を更新します。

また、完全データだけでなく不完全データを使って CPT を更新することができます。

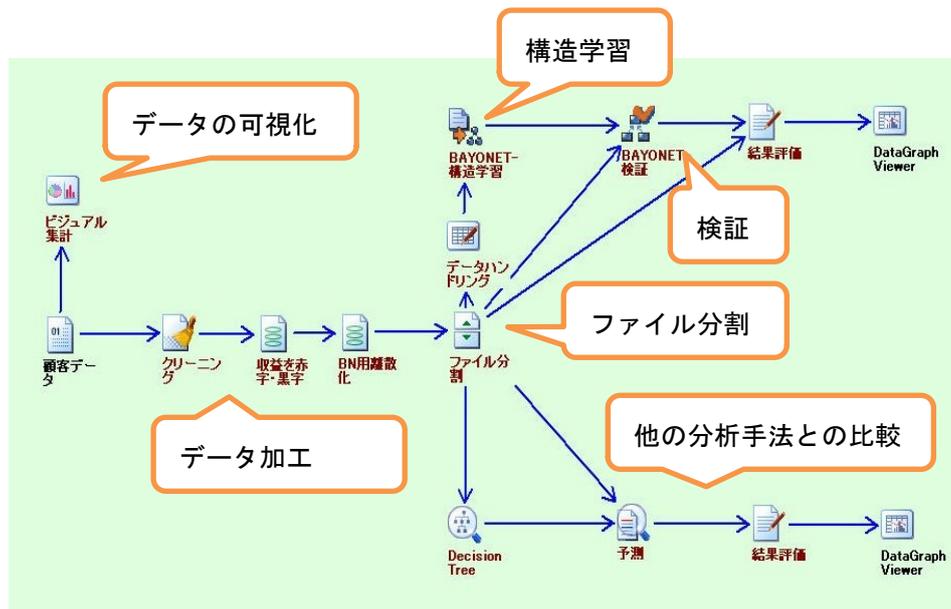
不完全データの場合は EM アルゴリズムを用いてパラメータを推定します。

- 推論

構造学習アイコンで構築したモデルを使って、入力データの欠損を推定値で補完します。複数のカラムを同時に補完することができます。

- 検証

構造学習アイコンで構築したモデルの検証を行います。目的変数の適合率、再現率などを算出します



今回のバージョンアップでは関係機能を強化するために新しいアイコンの実装や、既存アイコンについても機能の見直しを行いました。

- データインポート
Visual Mining Studio の外部のモデルファイルをインポートする機能です。
 スタンドアローン版の **BAYONET** で作ったモデル、または過去に作成したモデルを **Visual Mining Studio** に取り込み利用することが可能です。
- データエクスポート
Visual Mining Studio の構造学習や CPT 更新で作ったモデルを BIF 形式で外部に出力する機能です。スタンドアローン版の **BAYONET** に実装されていない EM アルゴリズムによる CPT 更新を **Visual Mining Studio** で行い、スタンドアローン版で再利用することができます。
- 構造学習
 アイコンをクリックすると、まず構造学習ダイアログが表示されます。設定を入力すると学習が始まり、終わると **BAYONET** に構築したモデルが表示されます。
 また、アイコンの実行後、メニューの[編集]で **BAYONET** を起動しモデルを確認することができます。ここでは、リンクの編集などの操作が可能です。