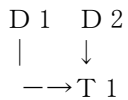


ベイジアン・ネットワークとその応用

○ ベイジアン・ネットワークによる推定

下の図は、病気D1~2が血液検査の結果T1に影響を及ぼすことを表しています。



一般に、1)ノードが確率変数で表され、2)ノード間の矢印が作用経路を表し、3)その作用の強さが条件付確率で定量化されている、有向非循環グラフをベイジアン・ネットワークと云います。

相関図との違いは、作用の強さが条件付確率で表現されることです。相関係数は、どんな場合でも係数として適切とは限りませんが、条件確率は、どんな場合でも成り立ちます。

但し、要因間の作用が循環する経路がある場合には、ベイジアン・ネットワークは使えません。現実には、要因間の作用は循環しますので、ベイジアン・ネットワークは現実妥当性がありません。相関係数から出発するパス解析も、循環する作用を理論的には許容しますが、組合せ爆発などの計算可能性の制約で、循環作用を事前に「ある程度」整理する必要があります。つまり、実際的には条件付確率のグラフでもパス係数のグラフでも、循環グラフは扱えません。

条件付確率 $P(T1 | D1, D2)$ は確率変数間の関係を表しています。ネットワークを構成する全ての確率変数の同時分布がわかれば、各ノードに付随する条件付確率あるいは事前確率を求めることができます。一般的に、ベイジアン・ネットワークを構成するのノードの数が n 個であれば、 $2^n - 1$ 個の確率が必要になります。

確率変数の同時分布式

$$P(D1, D2, T1) = P(D1) P(D2) P(T1 | D1, D2)$$

実際的には、未だ観測していない値を推定するのが、ベイジアン・ネットワークの用途になります。

確信度 $BEL(x) = P(x | e)$ e : 証拠(エビデンス)

上の例でいえば、T1のとき、その原因がD1の発症あるいはD2の発症である確信度(事後確率)は、

$$BEL(D1=1) = P(D1=1 | T1=1)$$

$$BEL(D2=1) = P(D2=1 | T1=1) \quad \text{で与えられます。}$$

ベイジアン・ネットワークは、1)ノードとなる確率変数、2)グラフ構造、3)ノードに付随する条件付確率を観測結果から決定します。データが不足して最適なグラフ構造が決定できない場合には、EMアルゴリズムなどで最尤推定を行うことができます。

* 上記は、「MEDICAL IMAGING TECHNOLOGY VOL21」の須鎌弘樹氏の論文を参考にさせていただきました。

○ マーケティングへの応用

ベイジアン・ネットワーク解析は、一人一人のお客様別のアプローチを、自動化してマス展開するプロモーションの基礎となる技術です。例えば「購入率を高めるDMの送り先の抽出」を行うことができます。アメリカでは、実用段階に入っている技術です。

日本の通販市場は、参入する会社数は増えて、人口は減少。会社数が増えたから全体の売上額が増加しただけで、一社当たりの売上高は減少しています。メディアも構造変化しています。テレビは、リーチの高さでは他のメディアを圧倒していますが、「新聞からテレビへ」という通販メディアの経験則はもう通用しないかも知れません。

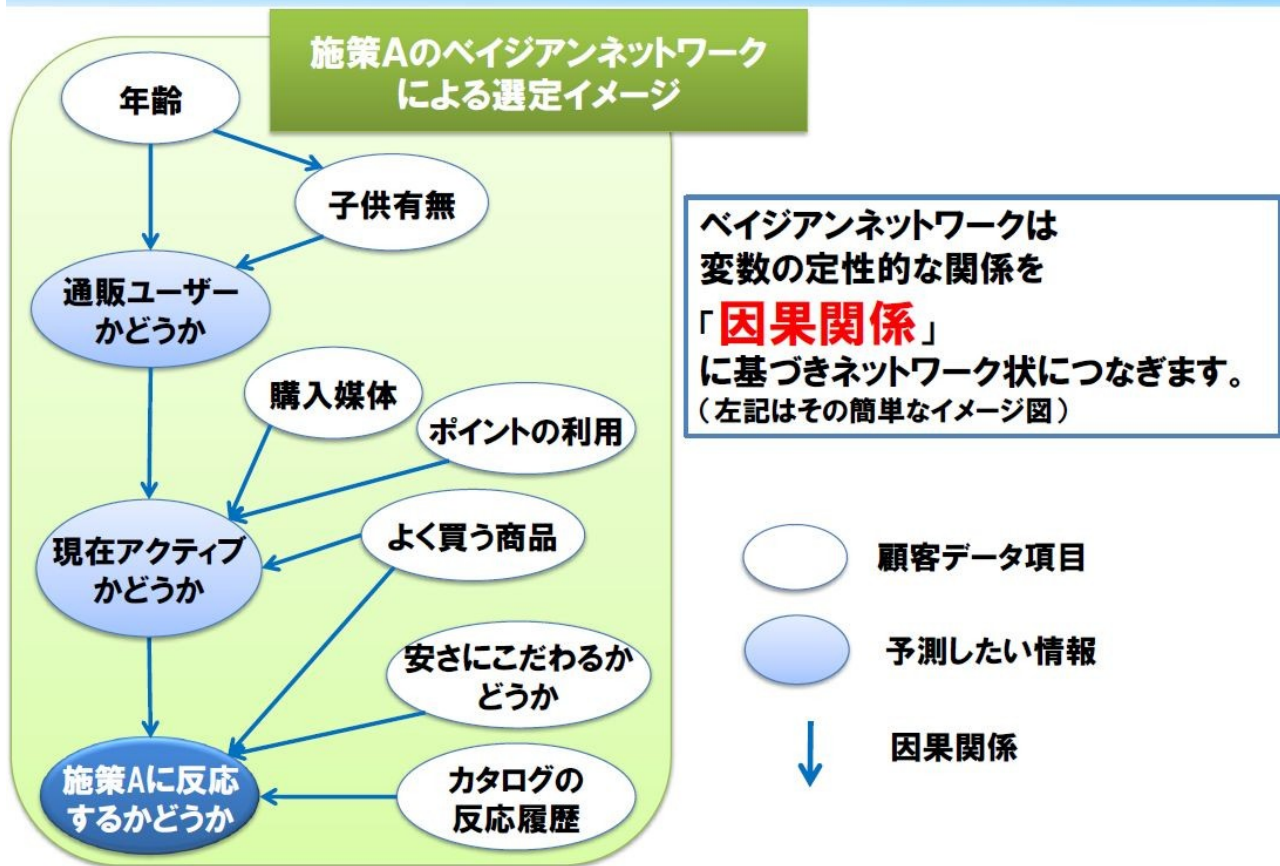
媒体費の収益に与える影響は大きく、自社リストの運用が課題となっています。ですが、ハウスリストと比べて、「休眠客」の方が多いのが現状。送付対象の絞込みによるコスト効率の追求が必要です。

DM送付先の抽出では、顧客属性と購買履歴を用いて、お客ごとの特定のキャンペーンに対する反応確率を予想します。この反応確率から顧客を選定します。ある事例では、ベイジアン・ネットワーク解析は、RFMによる抽出に比べ、休眠顧客の復活率を2倍にしました。

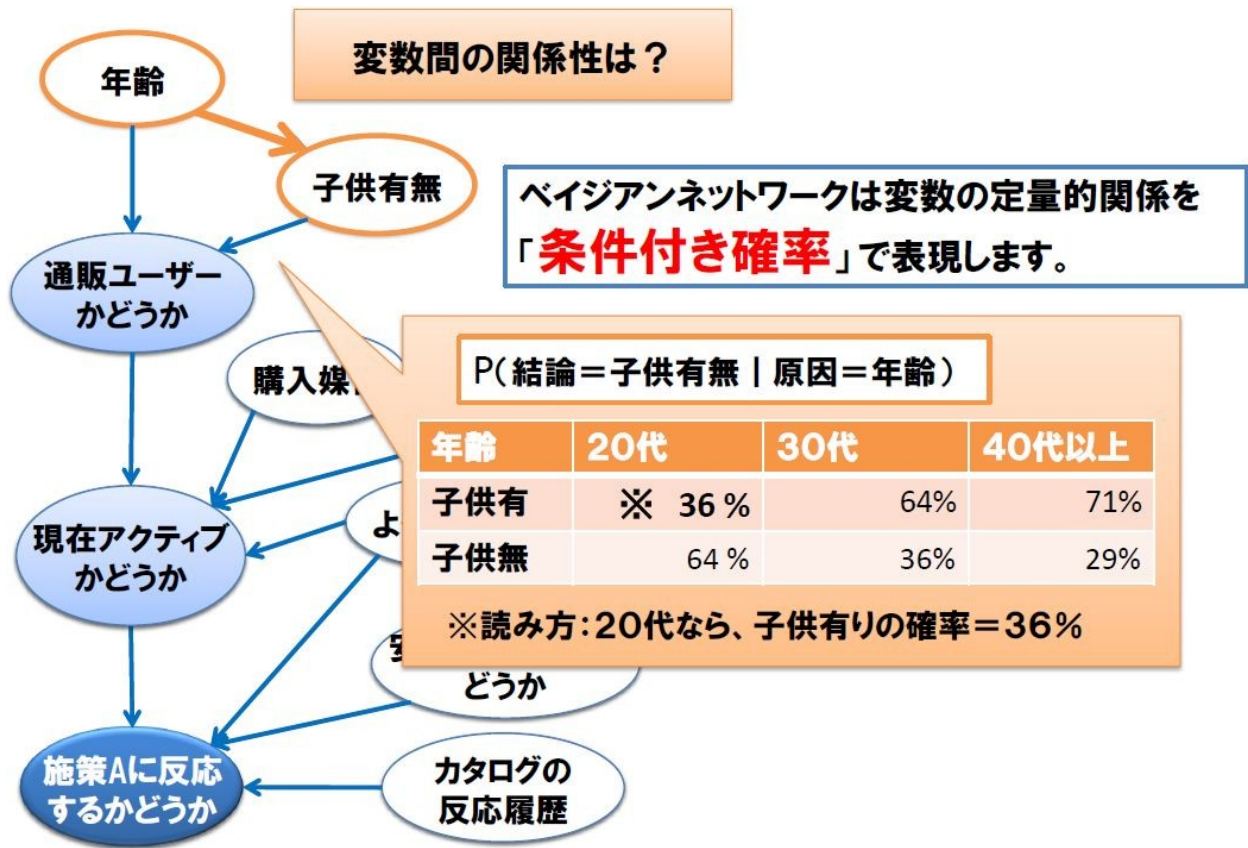
ベイジアン・ネットワーク解析の特徴は、1) 一人一人の確率を算出する、2) 要因経路を整理する、3) 主観的知見も利用できる、ことです。要因経路は、お客様がアクティブかどうか、特定のキャンペーンに反応するかどうか、というように階層的な目的変数(結果)を設定することができ、要因についての理解力を高めます。

日本での応用は未だ少ないようですが、マーケティング意思決定を支援する手法として大いに期待されます。

反応率予測(1): 変数の定性的関係 = 因果関係



反応率予測(2): 変数の定量的関係 = 条件付き確率



*上掲の資料は、モデライズ株式会社のセミナー資料を利用させて頂きました。