

ベイジアンネットワーク実務応用家からの手紙

～ マーケティングにおけるベイズモデルの必要性と活用の実際～

女子栄養大学 芳賀麻誉美

1. 「ベイズモデル」を語る難しさ

マーケティング分野におけるベイズモデルの活用は、非常に多岐に渡るが、専門家でさえその実態を正確に知る手段は多くなく、どの部分に何が使われているとすべて言い当てることは難しい。

まして実務応用家にとっては、マーケティングにおけるベイズモデルの必要性と活用の実際を語ることは非常に困難である。一般的なビジネスパーソンに至ってはお手上げというのが実情ではないだろうか。

難しさの理由は複数ある。

1つ目は、手法と用語の関係という非常に単純な問題である。

ベイズモデルと一言に言っても、「ベイズ×××」と呼ばれる手法は非常に多岐・多種類に渡るが、それぞれの概要を示し、体系化、違いを一般向けに解説した文献・書籍がほとんどないのである。専門家から見れば、大きく異なる手法であっても、一般のビジネスパーソンから見れば、「ベイズという言葉がついているし同じものじゃないの?」、程度の認識であることを、まずは指摘しておきたい。

たとえば、従来からある「ベイズ統計」と呼ばれるものと、「ベイジアンネットワーク」の違いについて、果たしてどれだけの人が理解し、言葉を使い分けられているだろうか。「ベイジアン・クラシファーズ」もわかりである。

もちろん、大きな目で見れば「ベイズモデル」という形でくくれるとしても、これらは立派に異なる手法であるといえる。

実務家の混乱を招いている要因の1つが、この「ベイズ×××」という手法の多さと言葉の問題なのである。

2つ目は、現実の応用と理論が、必ずしも有機的につながっておらず、ビジネス上の利用と研究が切り離されていることがあげられよう。特に日本ではこの傾向が強いと考えている。

商用的にベイズモデルを利用する場合、どういったデータに基づき、どのような事前処理をした上で、モデルの構築をし、システム化するかといった、すべての段階に経験的知見が活かされる。他方、それらの知見は、企業が優位性を保つために重要であるため公表されることは少なく、事例論文も多いとは言えないのが実情である。

たとえば、あるシステムでベイジアンネットワークを利用しているらしい、という話を聞くことはあっても、具体的にどのように利用したかは定かではないといった事例が非常に多い。また、そういった話を全く聞くことがなくとも、日本企業の優秀な技術者は、秘かに手法利用を模索していることも多い。

この問題は実際にはかなり根深い。これは実務応用の立場と理論研究の立場では、問題解決の基準のプライオリティが大きく異なることに起因する。たとえば、企業の応用上の目的のプライオリティは、理論的な正しさがいつも高いわけではないのである。効率性、速さ、構築費や保守費、など総合的な視点で最適な方法を選択していくのが実情だ。

以上の2つを越えて、内容と言葉を正確に理解し使いこなし、さらに情報を共有化した上で見えてくるコラボレーションがあつてこそ、マーケティングにおけるベイズモデルの活用に、劇的な進歩があると考えられる。

以下、本報告では、著者の理解の範囲で「ベイジアンネットワーク」に関する議論を進めていく。

2. ベイジアンネットワーク活用の実際

ベイジアンネットワーク活用事例として世界的に有名なのは、NASA のスペースシャトルの推進システムの保全、医療診断、さらに、Microsoft Office アシスタントによるユーザー補助などである。

現在、日本で取り込まれているマーケティング的課題としては、購買履歴や対象者情報、商品情報等を用いたリコメンデーションシステムの構築、機器の故障診断、流通価格予測、広告効果測定、プロモーション戦略、顧客ターゲティング、商品・サービス開発、消費者回遊行動のモデル化、電話応答における顧客ニーズ判定、Web ページの個人最適化など、多岐に渡る検討が行われている。

ユーザーの目から見たベイジアンネットワークの魅力は、ネットワーク構造が可視的なので、予測のみならず記述説明が可能であること。構造に対して知見を導入できること。予測効率も高いこと。シミュレーションが可能であること。1つのモデルを目的に合わせて多様に活用できること。

リアルタイムで収集されたデータを使い、機械学習的に、動的に構造構築と推論を行うことが可能なこと。(常に最新の状態に保てること)などである。

3. ベイジアンネットワーク応用上のポイント

多くの分野での検討が行われているといっても、実際の実務応用に至る事例が 100%というわけではない。また、検討の段階で多くの問題がユーザーの前に立ちはだかる。ここでは、検討チェックポイントの列挙という意味も含めて、これを述べる。

(1) データの問題

最も本質的な問題として、どういったデータを用意し、どの部分を利用するかということが、何よりも結果に与える影響が大きい問題である。当たり前のことのようにあるが、履歴データを使ったモデル化が目的で、事前にデータが用意されている場合などは、検討項目に挙げられていないこともあるため、考慮を忘れないことが重要である。

また、分析法とデータの収集やその量や質の問題は区別され議論されることが多いが、実務的な成果につなげるためには、「データと分析手法の組み合わせ」にも、より注目していく必要がある。データによって、最適なモデル構築の方法、推論計算、システム化といった結果が大きくことなることに、ユーザーはより注目しなければならない。一般化・共通化できる知見ももちろん重要であるが、それとともに、独自性の強い経験的知見にも目を配る必要がある。

こういった知見に目を向けることが、実は理論研究にも貢献するという側面もある。

基本的にマーケティング分野でベイジアンネットワークによる分析に利用されるデータは、大きく入力系データと出力系データにわけることができる。

入力系データは、以下の3つである。

- 「モノ・コト」データ-商品情報や価格リストなど対象刺激に対する情報、実験的な提示条件など-
- 「ヒト」データ-アンケートなどを利用し採取した対象者属性データやあるモノやコトに対する関与度や価値感などのデータ-
- 「バ」データ-出力を得た際の時間・場所・空間系のデータ。位置情報や気象データなどもここに入る。-

一方、出力系データも同じ観点で3分類できる。

- 「モノ・コト」データ-売上げ、閲覧数など-
- 「ヒト」データ

-A.あるバでのモノやコトに対する人間の行動履歴データ。POS や WEB 閲覧行動履歴、車のドライビングレコードなど。

-B.あるバでのモノやコトに対する人間の態度や意見、感想、評価データ。(モノコトの属性評価、情緒的ベネフィット、機能的ベネフィットといった形で階層化できることが多い)

- 「バ」データ-時間・場所・空間系のデータ。回遊行動、位置情報、などである。

具体的な分析、モデル化の前に、手持ちのデータそのものを以上の6分類に照らし合わせ、その後、使用データと範囲を検討後、分析に入る方が望ましい。

(2) 予備解析

統計学的知識に基づくモニタリングは、実際にはデータマイニングを効率的に行うために欠かせない技術である。いくら探索的な方法であったとしても、何百万件というデータをいきなり分析するのは非常に効率が悪い。良い結果を導くには、事象そのものと手元にあるデータに対する深い理解である。分析結果をある程度予見ほどに、サンプリングしたデータの挙動を鋭くモニタリングしておくことが、成功への鍵となる。

(3) 構造同定

ベイジアンネットワークに限らず、何らかの構造分析を行う手法の場合、変数選択と構造仮説をどのように持つかが非常に重要である。

これをベイジアンネットワーク流の言葉で言い換えれば、構造へ対しての事前知識をネットワーク構造の探索・同定に使用するかであり、探索の範囲を限定し、効率化すること、ともいえる。

ただし、市販されているベイジアンネットワークの分析ソフトには、これらの検討機能がないものも多いのが、実施上の大きな問題でもある。

なお、本シンポジウムで紹介した、一般的なマーケティングリサーチデータを使った事例と、商品開発や、より積極的なマーケティング戦略の立案に利用される「実験的調査データ」の事例であるが、通常、筆者はモデル化の際に、変数選択 変数のグループ化・層化 変数グループ間構造の探索範囲設定 変数グループ内構造の探索範囲設定、全体の構造化、という段階でモデル構築の探索範囲を限定しており、これが、実務上重要な、解釈の容易性と予測効率とに良い結果をもたらすという経験的知見を得ている。

ネットワークモデルの構築段階で、ユーザーが迷うのは、構造探索手法と推論方法に多様な手法が提案されているものの、決定的な方法がないことである。経験的には、手法の選択方法と基準の取り方によって、確率推論段階での予測効率が最大化される構造とはならないこともあるし、解釈容易性が高いモデ

ルとも限らないという問題点もある。

実務的には、構造化、確率推論と、さらに解釈という3つの基準があることを認識した上でモデル構築しなければならないが、これらは一般的な書籍ではまったく触れられていないため、初心者に対しての大きな壁となっているといえる。

(4) 確率推論

まず、ある特定の目的変数がある場合には、予測効率がどの程度であるか、検討する必要がある。

ただし、マーケティング分野においては、ある特定の目的変数があるというよりは、この構造を使って、どれだけ沢山のマーケティングシナリオに合わせたシミュレーションが出来るかがポイントとなる。そのためには、想定できるシナリオを、十分に準備する必要がある。これらが揃ってはじめて、1つのモデルを多様に利用できるというベイジアンネットワークの利点が最大限に生かせることになる。

4. トラブルシューティング

ベイジアンネットワークを使ってみようという段階になって、困難にぶつかることがないとは言えない。そういった場合に、分析上のトラブルを以下のフローに沿って、その性格を把握しておくことが必要である。

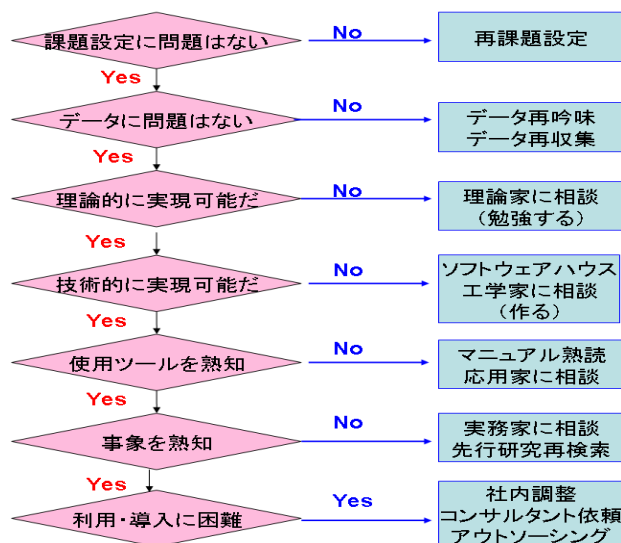


図1. 問題解決のフロー

犯しがちなミスとして、技術的にソフトウェアに搭載されていないから出来ないといった問題を、理論家に相談しても解決されることは少なく、また、純粋にツールが使いこなせていないだけの問題を、工学家やプログラマーに相談しても仕方がないのである。

自分の陥っている問題は、何が障害となって生じているのかを、客観的に判定する必要がある。

ただし、完全な初心者には、いったい何が問題でそれが一体技術的な問題なのか、理論的な問題なのかも判定がつかないこともあるのも現状であり、自分の問題を分析できる程度の知識を身につけておくこと、または、問題判定のアドバイスがもらえる知人を持つことなどが、最低限、ベイジアンネットワークに取り組む時に必要なことであるともいえる。

5. 今後の課題

現在、日本で望まれているのは、教育の充実によるベイジアンネットワークの一般への理論普及と、専門家による高度な研究の両方である。一般化と高度化が進むと、理論は進む。これにより、実務応用はご利益を受ける。

その一方で、現実の社会での実務応用が進めば進むほど、問題点や課題が発見され、理論研究も進む。

こういった、理論と応用の双方の進展こそが、社会に求められているものとする。

芳賀麻誉美

女子栄養大学 フードマーケティング研究室 助教

慶応義塾大学 総合政策学部 非常勤講師

立教大学 現代心理学部 非常勤講師

ヤフーバリューインサイト株式会社 マーケティング顧問

株式会社 Albert 顧問

< 著者連絡先 >

女子栄養大学 フードマーケティング研究室
〒350-0288 埼玉県坂戸市千代田 3-9-21

e-mail : haga@eiyo.ac.jp

<http://www.mayomi.org>