

チュートリアル：ベイジアンネット実践編

Tutorial on Bayesian Networks: Practice

本村 陽一*

産業技術総合研究所 情報処理研究部門

情報技術の適用範囲が飛躍的に拡大している現在、これまでコンピュータにとっては扱いにくかった不確実で非明示的な情報を取り扱うことが必要になってきている。例えばインターネットショッピングにおいて、顧客のニーズを以前の取引履歴や顧客管理情報から読み取り、目の前の顧客に最適な情報を推奨するリコメンデーションシステムの場合を考えてみよう。顧客リストの中から目の前の顧客にもっとも近い購買パターンを見つけるとしても、年齢や性別が同じだからといって同じ嗜好性を持っているとは限らない。また顧客が、登録時に申請したアンケートと同じ嗜好性をいつまでも持ち続けるというわけでもないだろう。そこで顧客のアクション (WWW ブラウジング履歴など) や属性、アンケート情報から総合的に判断して嗜好性を予測することが必要になるわけだが、必ずしも一つの答として決定できるとは限らない。その対象となる候補は同時に複数存在することもありえる。そしてそれぞれの候補についても「これまで見ていた WWW ページから、ユーザはおそらくサッカー関連の情報に興味がある可能性が高い」といった漠然とした不確実性を伴ったものとして扱うことが自然だろう。

こうした問題においては確率的な枠組みが有効である。サッカー関連の商品に興味のある可能性が 60%、旅行に興味のある可能性が 30% などというように複数の候補に確信度をつけ、不確実性を含んだ状態で取り扱うことができる。そして、直前に見ていたページが例えば韓国に関するものであった場合に、ワールドカップサッカー関連の情報、旅行関連情報、料理関連情報のそれぞれに興味のある確率を計算して、目の前の顧客が最も高い確率で関心を持つと思われる対象を提供することが考えられる。またこの確率を計算するには様々な多くの要因 (例えばアンケートで答えた趣味など) を考慮に入れて、その間の依存関係 (スポーツが趣味ならサッカーに興味がある可能性も高いなど) を利用することでより精度の高い予測が可能になる。

このような複数の要因の依存関係に基づいて確率計算を行う情報処理モデルとしてベイジアンネットが最近、様々な分野で注目を集めている。とくに最近、一般に広く利用できるソフトウェアとして入手できるものも増え、先端的な企業が製品開発や研究に活用している例もある。

そこで本チュートリアルでは、今後ますます重要になると思われる、顧客管理システムや対話的情報システムの中で必要となるユーザモデリング、大量のデータから隠れた依存構造を発見して様々な予測を行うデータマイニングとしての応用に焦点を当て、ベイジアンネットを適用する場合に必要な知識とノウハウをできる限り実践的な視点から解説する。

とくに他のデータマイニングや統計的手法である、多変量解析、共分散構造分析、因子分析、クロス集計表の検定、決定木、ニューラルネットワークとの関連と相違を見ることで、ベイジアンネットの特長を浮彫りにする。

ベイジアンネットは国内でもようやく日本語の文献 (人工知能学会誌 vol.17, No.5, 2002 年 9 月号など) を目にするできるようになったがまだまだ海外と比べてその認知度は高いとは言えない。しかし技術レベルとしては理論研究から実験評価の段階を経て、今やソフトウェアとしての実装や応用、さらには特許化¹、事業化へと確実に社会・産業へと浸透する段階に進みつつある。

本チュートリアルを通じて既存の各手法について知識のある研究者や、ベイジアンネットが適用できる問題領域に強いモチベーションを持った学生、技術者、企業関係者の方に、少しでもベイジアンネット技術の実際と可能性を知っていただきたい。現実的なツールとしてベイジアンネットを利用する機会が増えることで、国内の学術研究分野への貢献はもちろんのこと、産業応用への展開も進め、日本の情報技術の発展にわずかも寄与できれば幸いである。

* 〒 305-8568 茨城県つくば市梅園 1-1-1 つくば中央第二, tel: 0298-61-5415, e-mail: y.motomura@aist.go.jp

¹ 複数の外資系企業がすでに日本においてもベイジアンネット関連の応用技術の特許を取得している。