

# 変数選択機能を持つベイジアン・ネットワーク・ソフトの開発

## A development of Bayesian network software with variables selection function

植野 真臣  
Maomi Ueno

電気通信大学大学院 情報システム学研究所

Graduate School of Information Systems, The University of Electro-Communications

This paper develops a new Bayesian network software, which is called "Bayesian Discovery". The unique features of this software are as follows: (1) The conditional probabilities parameters are estimated based on the Dirichlet-multinomial distribution which is the most general form of the posterior distribution from Bayesian approach. The Dirichlet-multinomial distribution can represent various statistical model by changing the value of the hyperparameters. (2) The software has the following causal model selection criteria : 1. The exact predictive distribution of the Dirichlet-multinomial distribution which can includes Cooper's criterion, marginal likelihood, and so on. 2. The asymptotic distribution of the predictive distribution. 3. AIC. 4. MDL. 5. ICOMP. (3) The software has the following various searching algorithms of the network structure: 1. complete search, 2. CH Procedure, 3. Greedy algorithm without restrictions, 4. Greedy algorithm with some restrictions, 5. Genetic Algorithm. (4) The software has the variables selection function based on the statistical decision approach.

### 1. はじめに

本稿では、著者により開発された新しいベイジアン・ネットワーク・ソフトウェア「Bayesian Discovery」を紹介する。その特徴は、

(1) 条件付確率パラメータは、最も一般的な表現であるディレクレイ多項分布に基づくベイズ推定値をとり、事前分布のハイパーパラメータの設定により、様々な推定値を表現することができる。デフォルトは、一様分布を表現するハイパーパラメータが設定されている。

(2) 評価基準として、[Cooper90]や ML を含むディレクレイ多項分布より直接導かれた予測分布、およびその漸近分布、AIC, MDL, ICOMP など多様な因果モデル構築基準を持つ。特に、特定のアークにもハイパーパラメータを与えることができ、特定のアークを事前に設定してしまうことや事前にアークを引かないように設定してしまうことも可能である。また、漸近解析により、ハイパーパラメータの設定と予測分布の振る舞いも多く研究されており、これらを利用した用途に応じたハイパーパラメータの設定も可能である。

(3) 構造のサーチアルゴリズムは、以下を用意している。

1. 全数探索
2. CH Procedure[Cooper90]
3. Greedy Algorithm(A) 全ノードの中で、いくつかのノードに分け、局所的なネットワーク構造を最適化させた後に、各局所的ネットワーク同士を統合する。
4. Greedy Algorithm(B) 全ノードの中で、いくつかのノードに分け、局所的なネットワーク構造を最適化させた後に、各局所的ネットワーク同士を CH Procedure[Cooper90] で用いられるノードの順序情報を用いて統合する。
5. 各アークのある・なしを遺伝子とみた GA(遺伝アルゴリズム)

(4) 統計的意思決定アプローチによるノードの価値(Key Node) の計算と表示により、予測効率の観点から変数選択が可能になり、より頑健な因果モデルを構築でき、また、重要なキーノードを探索することができる。

### 2. ベイジアン・ネットワーク・ソフトウェア

これまでも、多くの有用なベイジアン・ネットワーク・ソフトウェアが開発されてきたが、大別して、(1) 確率推論のためのソフトウェア、(2) 因果モデルのデータからの学習のための(因果発見)ソフトウェアの二つに分類できる。(1)では、複雑なネットワーク構造を多重木に変換することで、確率伝播の計算量を減少させる Junction Tree アルゴリズム[Jensen96]を搭載している Hugin (<http://www.hugin.com>), Junction Tree アルゴリズムを標準として、モデルを変えて推論結果を比較できる機能や Gibbs サンプリングの機能を持つ BayesianLab(<http://www.bayesia.com>)、通常のベイジアン・ネットワークのみでなく、ダイナミック・ベイジアン・ネットワークの推論をさまざまなアルゴリズムで行える Bayes Net Toolbox(BNT) (<http://bnt.sourceforge.net/>)などが有名であるし、(2)では、[Cooper90]のアルゴリズムを搭載した商用ソフトの Bayesware Discover (<http://www.bayesware.com>)、MML(Minimum Message Language)アプローチによる因果発見ソフトで、GA(遺伝アルゴリズム)、Metropolis search などの確率的探索手法を搭載したフリーソフト CaMML(Causal Discovery vis MML) (<http://www.datamining.monash.edu.au/software/camml/>)、PC アルゴリズムと GA(遺伝アルゴリズム)探索機能を持つ TETRAD

(<http://www.phil.cmu.edu/projects/tetrad/>)、決定木の出力を同時に行えるフリーソフト WinMine (<http://research.microsoft.com/~dmax/winmine/tool.doc.htm>)、さらに、[Cooper90]、C4.5, AIC, MDL, ML を評価基準として持ち、ユーザーが新しい評価基準を定義でき、全探索、Greedy Search を探索機能として持ち、ニューラルネットにより条件付確率を学習できる国産のソフト BayoNet [本村 03] (<http://www.msi.co.jp/BAYONET/index.html>)などが挙げられる。

今回、開発されたソフトウェア「Bayesian Discovery」の特徴は、様々な技術の寄せ集めではなく、ベイズ統計的基礎に忠実であり、一貫した理論によって説明できること、また因果モデル構築に最も重要であると考えられる変数発見を支援する変数評価(Key Node)機能である。

### 3. Bayesian Discovery

連絡先: 植野真臣、電気通信大学大学院情報システム学研究所, [ueno@ai.is.ucc.ac.jp](mailto:ueno@ai.is.ucc.ac.jp), 〒182-8585 東京都調布市調布ヶ丘 1-5-1 TEL : 0424-43-5621

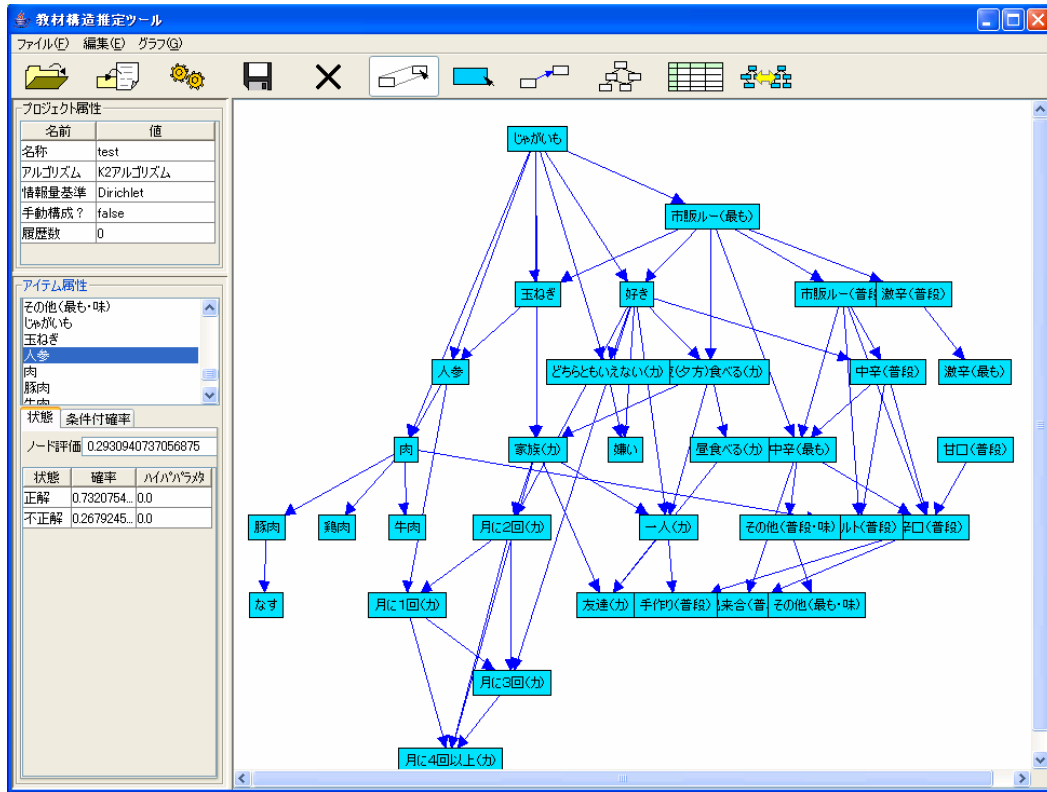


図 1. Bayesian Discovery のインターフェース

紙面の都合上、最も本論の特徴となる部分について記述する。

### 3.1 基本モデル

本システムでは、 $\theta_{ijk}$  を親ノード変数集合  $\Pi_i$  が  $j$  番目のパターンをとったときの  $x_i = k$  となる条件付確率を示すパラメータとする。このとき、データ  $X$  を得たときの尤度は、以下のとおりである。

$$p(\mathbf{X} | \Theta, S, S) = \prod_{i=1}^N \prod_{j=1}^{q_i} \frac{\sum_{k=0}^{r_i-1} n_{ijk}!}{\prod_{k=0}^{r_i-1} n_{ijk}!} \prod_{k=0}^{r_i-1} \theta_{ijk}^{n_{ijk}} \quad (1)$$

共益自然事前分布である以下のディレクレ分布を事前分布に考えれば、

$$p(\Theta | S) = \prod_{i=1}^N \prod_{j=1}^{q_i} \frac{\Gamma(\sum_{k=0}^{r_i-1} n'_{ijk})}{\prod_{k=0}^{r_i-1} \Gamma(n'_{ijk})} \prod_{k=0}^{r_i-1} \theta_{ijk}^{n'_{ijk}-1} \quad (2)$$

としたときの予測分布を基準としている。

$$p(\mathbf{X} | S) = \prod_{i=1}^N \prod_{j=1}^{q_i} \frac{\prod_{k=0}^{r_i-1} \Gamma(n'_{ijk} + n_{ijk})}{\Gamma(\sum_{k=0}^{r_i-1} (n'_{ijk} + n_{ijk}))} \quad (3)$$

この基準は、ハイパーパラメータ  $n'_{ijk}$  の値を変えることにより、MDL や AIC など様々なモデル選択基準に収束する。また、これより漸近展開することにより、独自の基準も得ることができる。本システムでは、すべてのハイパーパラメータを自由に設定でき、この理論から間接的に導かれる情報量基準、BIC、MDL などすべての基準を用意しており、理論的矛盾もない。

### 3.2 変数評価機能: KeyNode

[植野 99]では、ある 2 ノード間のアークのあるなしは、他のノ

ード集合の組み合わせによって決定することを例示し、頑健な因果モデルを構築するためには、重要ノードの発見が必要であると考えている。結果として、以下のノードの値を意志決定モデルから定義している。

$$\sum_{l=1}^n 2^{q_l} p(x_1, \dots, x_n | x_l) \log p(x_1, \dots, x_n | x_l) \quad (4)$$

$$- \sum_{l=1}^n p(x_1, \dots, x_n) \log p(x_1, \dots, x_n)$$

一つのノードとネットワーク全体との相互情報量として解釈できるが、この値の大きなノードは、それを加えることにより既存のネットワークの構造を大きく変化させてしまう変数である。この値の大きなノードのみで作成されたネットワークは、頑健で他の発見によってダイナミックにパラダイム変換されることが少ない。本システムでは、ノードをクリックすることによってこの値を提示する機能と、自動的に値の低いノードを削除する機能を持っている。

### 参考文献

[Cooper90] Cooper, G.F. and Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data, *Machine Learning*, **9**, pp.54-62, (1990)  
 [Jensen96] Jensen, F.V.: An introduction to Bayesian Networks, University Clooeege London Press (1996)  
 [本村 03] 本村 陽一: ペイジアンネットソフトウェア BayoNet, 計測と制御, 42-8, 693-694, (2003)  
 [植野 99] 植野真臣, 事例に基づくエキスパートシステムの変数評価, 人工知能学会全国大会論文集, PP.561-562, (1999)