1J3-5

命題論理に基づく確率モデルのためのベイズ推定

Bayesian inference for logic-based probabilistic modeling

石畠正和　　亀谷由隆　　佐藤泰介
Masakazu Ishihata　Yoshitaka Kameya　Taisuke Sato

東京工業大学 大学院情報理工学研究科
Graduate School of Information Science and Engineering, Tokyo Institute of Technology

We propose an MCMC and variational Bayes (VB) for Bayesian inference in propositional logic-based probabilistic models (PBPMs). The MCMC is an application of that in PCFGs to PBPMs and the VB is a generalization of the EM algorithm for PBPMs. These proposed methods are efficiently executed as dynamic programming algorithm on BDDs which compactly represent observations. We apply them for evaluating abductive hypotheses about metabolic pathway.

1. はじめに

近年，遺伝子データや関係データベースなど記号データ中の不確実性を扱う方法として，記号データ上の確率モデルが広く利用されている．記号データを扱う確率モデルの例として，HMM(hidden Markov model, 隠れマルコフモデル) や BN(Bayesian network, ベイジアンネットワーク), PCFG(probabilistic context free grammar, 確率的文脈自由文法) などが挙げられる．HMM は音声認識，BN は人工知能，そして PCFG は自然言語処理で得に利用されている．これらの記号データを扱う確率モデルを統合的に扱う方法として，論理に基づく確率モデルが注目されている [亀谷 11]．論理に基づく確率モデリングを実現する１つの方法として，PRISM[Sato 01] が提案されている．PRISM は Prolog の確率的拡張であり，HMM, BN, PCFG は PRISM プログラムにより表現可能である．更に PRISM の組み込み述語により，それらの確率モデルに対するパラメータ学習も容易に行えるようになっている．しかし，PRISM は確率計算を効率的に行うため，扱う論理式に排反性条件を課しており，この条件を取り除くため任意の命題論理式で表現される確率モデルに対する EM アルゴリズムも提案されている [Ishihata 10]．このような命題論理に基づく確率モデルの応用として，論理推論の一つであるアブダクションにより得られた仮説の統計的評価が挙げられる [Inoue 09]．アブダクションは論理式で表現される背景知識より，観測事実を論理的に説明する仮説を発見する．多くの場合，発見される仮説の数は膨大となり，それらをどう評価するかが問題となる．命題論理に基づく確率モデルを用いれば仮説上の確率分布を定義可能であり，最尤推定により仮説の確率を学習すればそれらを統計的に評価できる．

最尤推定は観測データが少ないときに過学習を起こしやすいことが知られている．この原因としてパラメータを観測確率を最大とする一点で推定する事が挙げられる．これに対してベイズ推定は，観測を得た後のパラメータの分布 (事後分布) を推定する．予測分布は事後分布の期待値によって表現されるため，観測が少ない場合においても高い汎化性能が得られる．また，パラメータの事前分布はパラメータの確かさを表現するため，知識としてパラメータの確からさが与えられる場合，その知識をデルに直接取り込むことが可能となる．つまり，論理に基づく確率モデルのベイズ推定では，規則などの決定的な知識は論理式として，パラメータの確かさなどの経験的な知識は事前分布として表現することが可能である．記号的なデータを扱う確率モデルに対するベイズ推定は既に提案されており，HMM であれば [MacKay 97], PCFG ならば [栗原 04, Johnson 07], PRISM ならば [Sato 09, Sato 11] などが挙げられる．しかし，任意の命題論理式で表現される確率モデルに対するベイズ推定はまだ提案されていない．そこで本研究では命題論理に基づく確率モデルに対するベイズ推定を提案する．ベイズ推定を実行する方法として，MCMC 法と変分ベイズ法が代表的である．本研究ではその双方を提案する．更に本論文では命題論理に基づく確率モデルのベイズ推定をアブダクションによって得られた仮説評価に適用する．ここでは専門家の決定的な知識は論理式として，経験的な知識は事前分布として表現されるとし，仮説の評価結果がそれらの知識を正しく反映していることを確認する．

2. 問題設定

2.1 命題論理に基づく確率モデルと最尤推定

ここではまず命題論理に基づく確率モデルを定式化する．互いに独立な離散確率変数列を $\mathcal{X} \equiv \{X_i\}_{i=1}^{N_V}$，各変数が従う確率分布のパラメータ列を $\boldsymbol{\theta} \equiv \{\boldsymbol{\theta}_i\}_{i=1}^{N_P}$ とする．ここで $\boldsymbol{\theta}_i \equiv \{\theta_{ij}\}_{j=1}^{N_i}$ $(0 \le \theta_{ij} \le 1, \sum_{j=1}^{N_i} \theta_{ij} = 1)$ は N_i 面体のサイコロの分布 (カテゴリカル分布) のパラメータである．同じ分布に従う確率変数が複数存在するとき，$N_P < N_V$ となる．確率変数 X_i が従う分布の添字を $\pi(i)$ と表す．従って X_i がパラメータ $\boldsymbol{\theta}_{i'}$ で定義される分布に従うとき，$\pi(i) = i'$ となる．これより，確率変数 X_i がその j 番目の値 x_{ij} $(1 \le j \le N_{\pi(i)})$ を取る確率 $p(x_{ij} \mid \boldsymbol{\theta}_{\pi(i)})$ は $\theta_{\pi(i)j}$ となる．同様に，$\mathcal{X}$ が値 $x \equiv \{x_i\}_{i=1}^{N_V}$ を取る確率 $p(x \mid \boldsymbol{\theta})$ は以下の様に計算できる．

$$p(x \mid \boldsymbol{\theta}) = \prod_{i=1}^{N_P} \prod_{j=1}^{N_i} \theta_{ij}^{\sigma_{ij}(x)}, \quad \sigma_{ij}(x) \equiv |\{x_{i'} | \pi(i') = i, x_{i'} = x_{i'j}\}|$$

ここで $\Omega_{\mathcal{X}}$ を $\mathcal{X}$ に対する割り当てのすべての集合である．すると確率事象 $y \subseteq \Omega_{\mathcal{X}}$ の確率 $p(y \mid \boldsymbol{\theta})$ は以下により計算される．

$$p(y \mid \boldsymbol{\theta}) = \sum_{x \in y} p(x \mid \boldsymbol{\theta})$$

連絡先: ishihata@mi.cs.titech.ac.jp

今，T 個の確率事象 $y^{(1)},\dots,y^{(T)}$ を独立に観測したとする．すると観測列 $\boldsymbol{y}\equiv\{y^{(t)}\}_{t=1}^{T}$ を得る確率 $p(\boldsymbol{y}\mid\boldsymbol{\theta})$ は以下となる．

$$p(\boldsymbol{y}\mid\boldsymbol{\theta})=\prod_{t=1}^{T}p(y^{(t)}\mid\boldsymbol{\theta})$$

対数 $\log p(\boldsymbol{y}\mid\boldsymbol{\theta})$ を対数尤度と呼び，最尤推定では対数尤度を最大化するパラメータ $\hat{\boldsymbol{\theta}}$ を求める．ここで観測列 $\boldsymbol{y}$ に対応する $\mathcal{X}$ の値列 $\boldsymbol{x}\equiv\{x^{(t)}\}_{t=1}^{T}$ は観測不可能とする．完全データ $\boldsymbol{x}$ が与えられる場合，対数尤度 $\log p(\boldsymbol{x}\mid\boldsymbol{\theta})$ を最大化する $\hat{\boldsymbol{\theta}}$ は以下の様に容易に計算できる．

$$\hat{\theta}_{ij}=\frac{\sigma_{ij}(\boldsymbol{x})}{\sum_{j'=1}^{N_i}\sigma_{ij'}(\boldsymbol{x})}$$

ここで $\sigma_{ij}(\boldsymbol{x})\equiv\sum_{t=1}^{T}\sigma_{ij}(x^{(t)})$ である．これに対し，不完全データ $\boldsymbol{y}$ から $\hat{\boldsymbol{\theta}}$ を計算することは自明ではないが，EM アルゴリズムは以下の様に $\boldsymbol{\theta}^{(k)}$ に関する $\sigma_{ij}(\boldsymbol{x})$ の期待値を用い，パラメータを繰り返し更新することで対数尤度を極大化する．

$$\theta_{ij}^{(k+1)}=\frac{\mathrm{E}[\sigma_{ij}(\boldsymbol{x})]_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})}}{\sum_{j=1}^{N_i}\mathrm{E}[\sigma_{ij}(\boldsymbol{x})]_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})}}$$

一般に期待値 $\mathrm{E}[\sigma_{ij}(\boldsymbol{x})]_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})}$ の計算には指数時間掛かる．そこでこの計算を効率的に行うため，$\mathcal{X}$ 上の確率分布を命題化し，確率事象 y を命題論理によりコンパクトに表現する．

命題化された確率モデルでは確率事象 y は命題論式により表現される．今，確率変数 X_i の値域 $\{x_{ij}\}_{j=1}^{N_{\pi(i)}}$ に対して全順序 $x_{ij}<x_{ij'}$ $(j<j')$ を与える．すると，X_i の確率分布は確率的命題変数集合 $\mathcal{A}_i\equiv\{A_{ij}\mid A_{ij}\equiv\text{“}X_i\le x_{ij}\mid X_i>x_{i,j-1}\text{”},1\le j\le N_{\pi(i)}-1\}$ 上の確率分布として表現される．ここで確率的命題変数 A_{ij} は互いに独立して以下の分布に従うとする．

$$p(A_{ij}|\boldsymbol{\theta})\equiv\frac{\theta_{\pi(i)j}}{\phi_{\pi(i)j}},\quad p(\neg A_{ij}|\boldsymbol{\theta})\equiv\frac{\phi_{\pi(i),j+1}}{\phi_{\pi(i)j}},\quad \phi_{ij}\equiv\sum_{j'=j}^{N_i}\theta_{ij}$$

確率的命題変数集合 $\mathcal{A}_i$ の定義より，確率事象 “$X_i=x_{ij}$” は以下の論理式により表現できる．

$$\text{“}X_i=x_{ij}\text{”}\equiv\begin{cases}A_{ij}\wedge\bigwedge_{j'=1}^{j-1}\neg A_{ij'} & 1\le j<N_{\pi(i)}\\ \bigwedge_{j'=1}^{j-1}\neg A_{ij'} & j=N_{\pi(i)}\end{cases}$$

その結果，確率事象 “$X_i=x_{ij}$” $(1\le j<N_{\pi(i)})$ の確率は $\mathcal{A}_i$ 上の分布を用いて以下の様に計算される．

$$\begin{aligned}p(\text{“}X_i=x_{ij}\text{”}\mid\boldsymbol{\theta})&=p(\mathcal{A}_{ij}\mid\boldsymbol{\theta})\prod_{j'=1}^{j-1}p(\neg\mathcal{A}_{ij'}\mid\boldsymbol{\theta})\\&=\frac{\theta_{\pi(i)j}}{\phi_{\pi(i)j}}\prod_{j'=1}^{j-1}\frac{\phi_{\pi(i),j'+1}}{\phi_{\pi(i)j'}}\\&=\theta_{\pi(i)j}\quad(=p(X_i=x_{ij}\mid\boldsymbol{\theta}))\end{aligned}$$

任意の確率事象 y は確率的命題変数集合 $\mathcal{A}\equiv\{\mathcal{A}_i\mid 1\le i\le N_V\}$ の命題論理式 f_y として以下の様に表現できる．

$$f_y\equiv\bigvee_{x\in y}f_x,\qquad f_x\equiv\bigwedge_{x_i\in x}\text{“}X_i=x_i\text{”}$$

従って $\mathcal{X}$ 上の分布は命題化された分布により計算可能である．命題化された分布は任意の論理式に対して確率を与えることが可能であり，任意の確率事象は論理式で表現可能である．しかし論理式に沿った確率計算は指数的な時間が掛かる．そこでこの論理式を BDD (binary decision diagram, 二部決定図) を用いて圧縮し，確率計算を効率的に行う手法が提案されている [Ishihata 10]．この手法は EM アルゴリズムの期待値 $\mathrm{E}[\sigma_{ij}(\boldsymbol{x})]_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})}$ を BDD のサイズに比例する時間で計算可能である．

2.2 命題論理に基づく確率モデルとベイズ推定

最尤推定では対数尤度を最大化する $\hat{\boldsymbol{\theta}}$ を推定するのに対し，ベイズ推定は以下の事後分布 $p(\boldsymbol{\theta}\mid\boldsymbol{y},\boldsymbol{\alpha})$ を推定する．

$$p(\boldsymbol{\theta}\mid\boldsymbol{y},\boldsymbol{\alpha})=\frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\alpha})}{p(\boldsymbol{y}|\boldsymbol{\alpha})},\quad p(\boldsymbol{y}\mid\boldsymbol{\alpha})=\int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\alpha})d\boldsymbol{\theta}$$

ここで $p(\boldsymbol{\theta}\mid\boldsymbol{\alpha})$ はパラメータの事前分布，$p(\boldsymbol{y}\mid\boldsymbol{\alpha})$ は周辺尤度である．事前分布は以下のディリクレ分布の積であり，$\boldsymbol{\alpha}\equiv\{\boldsymbol{\alpha}_i\}_{i=1}^{N_P}$, $\boldsymbol{\alpha}_i\equiv\{\alpha_{ij}\}_{j=1}^{N_i}$ をハイパーパラメータと呼ぶ．

$$p(\boldsymbol{\theta}\mid\boldsymbol{\alpha})\equiv\prod_{i=1}^{N_P}\frac{1}{Z(\boldsymbol{\alpha}_i)}\prod_{j=1}^{N_i}\theta_{ij}^{\alpha_{ij}-1},\quad Z(\boldsymbol{\alpha}_i)\equiv\frac{\prod_{j=1}^{N_i}\Gamma(\alpha_{ij})}{\Gamma\left(\sum_{j=1}^{N_i}\alpha_{ij}\right)}$$

ディリクレ分布はカテゴリカル分布の共役事前分布であり，完全データ $\boldsymbol{x}$ に対する事後分布 $p(\boldsymbol{\theta}\mid\boldsymbol{x},\boldsymbol{\alpha})$ も以下の様にディリクレ分布の積となる．

$$p(\boldsymbol{\theta}\mid\boldsymbol{x},\boldsymbol{\alpha})\propto p(\boldsymbol{x}\mid\boldsymbol{\theta})p(\boldsymbol{\theta}\mid\boldsymbol{\alpha})=\prod_{i=1}^{N_p}\frac{1}{Z(\boldsymbol{\alpha}_i)}\prod_{j=1}^{N_i}\theta_{ij}^{\alpha_{ij}+\sigma_{ij}(\boldsymbol{x})-1}$$

従って，$p(\boldsymbol{\theta}\mid\boldsymbol{x},\boldsymbol{\alpha})=p(\boldsymbol{\theta}\mid\boldsymbol{\alpha}+\sigma(\boldsymbol{x}))$ と書ける．ここで $\sigma(\boldsymbol{x})\equiv\{\sigma_i(\boldsymbol{x})\}_{i=1}^{N_P}$, $\sigma_i(\boldsymbol{x})\equiv\{\sigma_{ij}(\boldsymbol{x})\}_{j=1}^{N_i}$ である．これより完全データに対する周辺尤度 $p(\boldsymbol{x}\mid\boldsymbol{\alpha})$ は以下の様に計算される．

$$p(\boldsymbol{x}\mid\boldsymbol{\alpha})=\prod_{i=1}^{N_i}\frac{Z(\boldsymbol{\alpha}_i+\sigma_i(\boldsymbol{x}))}{Z(\boldsymbol{\alpha}_i)}$$

同様に不完全データ $\boldsymbol{y}$ に対する事後分布 $p(\boldsymbol{\theta}\mid\boldsymbol{y},\boldsymbol{\alpha})$ と周辺尤度 $p(\boldsymbol{y}\mid\boldsymbol{\alpha})$ は以下の様に計算される．

$$p(\boldsymbol{\theta}\mid\boldsymbol{y},\boldsymbol{\alpha})\propto\sum_{\boldsymbol{x}\in\boldsymbol{y}}\prod_{i=1}^{N_p}\frac{1}{Z(\boldsymbol{\alpha}_i)}\prod_{j=1}^{N_i}\theta_{ij}^{\alpha_{ij}+\sigma_{ij}(\boldsymbol{x})-1}$$

$$p(\boldsymbol{y}\mid\boldsymbol{\alpha})=\sum_{\boldsymbol{x}\in\boldsymbol{y}}\prod_{i=1}^{N_i}\frac{Z(\boldsymbol{\alpha}_i+\sigma_i(\boldsymbol{x}))}{Z(\boldsymbol{\alpha}_i)}$$

これらの計算は $\boldsymbol{y}$ に対する完全データ $\boldsymbol{x}$ の全組み合わせを考えるため，指数的な計算時間を要する．そこで事後分布の計算を近似的に行う方法として，MCMC に基づくサンプル近似と変分ベイズ法 (VB) による変分近似が利用されている．本論文では $\mathcal{X}$ の分布に対する MCMC と VB の双方を導出し，それらを命題化された分布を用いて効率的に実行する方法を提案する．

3. MCMC ベイズ推定

MCMC(Markov chain Monte Calro, マルコフ連鎖モンテカルロ) 法は目的分布からの直接サンプリングが困難である場

合，サンプリングが容易な提案分布を用い，目的分布を定常分布に持つようなマルコフ連鎖を構成することで目的のサンプリングを行う．MCMC の特殊形である Metropolis-Hastings (M-H) サンプリングは，提案分布から得られた候補を確率 A で受理/棄却を繰り替えすことで目的のサンプリングを実現する．PCFG に対する M-H アルゴリズムは既に知られており [Johnson 07]，$\mathcal{X}$ の事後分布 $p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\alpha})$ からの M-H サンプリングは [Johnson 07] と同様に以下の手順で実効可能である．

1. 区間 $[1, T]$ から t をランダムに得る．
2. 提案分布 $p(x \mid y^{(t)}, \boldsymbol{\theta}^{\mathrm{old}})$ から候補 x' を得る．
3. 候補 x' を確率 $A(x^{(t)}, x')$ で受理する．
4. 候補 x' を受理したなら $\boldsymbol{\theta}^{\mathrm{old}}$ を $\boldsymbol{\theta}^{\mathrm{new}}$ に更新する．

ここで $A(x^{(t)}, x')$ と $\boldsymbol{\theta}^{\mathrm{new}}$ はそれぞれ以下である．

$$A(x^{(t)}, x') \equiv \min\left\{1, \frac{p(x' \mid \boldsymbol{x}^{(-t)}, \boldsymbol{\alpha})p(x^{(t)} \mid y^{(t)}, \boldsymbol{\theta}^{\mathrm{old}})}{p(x^{(t)} \mid \boldsymbol{x}^{(-t)}, \boldsymbol{\alpha})p(x' \mid y^{(t)}, \boldsymbol{\theta}^{\mathrm{old}})}\right\}$$

$$\theta_{ij}^{\mathrm{new}} \equiv \frac{\sigma_{ij}(\boldsymbol{x}^{(-t)}) + \alpha_{ij}}{\sum_{j'=1}^{N_i} \sigma_{ij'}(\boldsymbol{x}^{(-t)}) + \alpha_{ij'}}$$

なお $\boldsymbol{x}^{(-t)} \equiv \boldsymbol{x} \backslash x^{(t)}$ であり，確率 $p(x \mid \boldsymbol{x}^{(-t)}, \boldsymbol{\alpha})$ は以下の様に計算できる．

$$\begin{aligned} p(x \mid \boldsymbol{x}^{(-t)}, \boldsymbol{\alpha}) &= \frac{p(\boldsymbol{x} \mid \boldsymbol{\alpha})}{p(\boldsymbol{x}^{(-t)} \mid \mathcal{A})} \\ &= \prod_{i=1}^{N_P} \frac{Z(\alpha_i + \sigma_i(\boldsymbol{x}))}{Z(\alpha_i + \sigma_i(\boldsymbol{x}^{(-t)}))} \end{aligned}$$

ここで提案分布 $p(x \mid y^{(t)}, \boldsymbol{\theta}^{\mathrm{old}})$ からのサンプリングは $p(X_1 \mid y^{(t)}, \boldsymbol{\theta}^{\mathrm{old}}), p(X_2 \mid x_1, y^{(t)}, \boldsymbol{\theta}^{\mathrm{old}}), \ldots$ と順次サンプリングを実行することで実現できる．このサンプリングは命題化された布用いて $p(A_{11} \mid y^{(t)}, \boldsymbol{\theta}^{\mathrm{old}}), p(A_{12} \mid a_{11}, y^{(t)}, \boldsymbol{\theta}^{\mathrm{old}}), \ldots$ のサンプリングとして実現可能である．ここで $y^{(t)}$ を表現する BDD を用いれば，この命題された分布からのサンプリングは BDD 上の動的計画法として効率的に実行できる．

4. 変分ベイズ法

MCMC ベイズ推定はサンプリングであるため非常に低速である．そこでより高速な近似手法として VB(variational Bayes, 変分ベイズ) 法が知られている．変分ベイズ法は計算したい事後分布 $p(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\alpha})$ をよく近似するテスト分布 $q(\boldsymbol{x}, \boldsymbol{\theta}) \equiv q(\boldsymbol{x})q(\boldsymbol{\theta})$ を変分法により発見する．今，周辺対数尤度 $\log p(\boldsymbol{y} \mid \boldsymbol{\alpha})$ を Jensen の不等式を用いて変形すると以下を得る．

$$\begin{aligned} \log p(\boldsymbol{y} \mid \boldsymbol{\alpha}) &= \log \sum_{\boldsymbol{x}} \int p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta} \mid \boldsymbol{\alpha}) d\boldsymbol{\theta} \\ &= \log \sum_{\boldsymbol{x}} \int q(\boldsymbol{x}, \boldsymbol{\theta}) \frac{p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta} \mid \boldsymbol{\alpha})}{q(\boldsymbol{x}, \boldsymbol{\theta})} d\boldsymbol{\theta} \\ &\geq \sum_{\boldsymbol{x}} \int q(\boldsymbol{x}, \boldsymbol{\theta}) \log \frac{p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta} \mid \boldsymbol{\alpha})}{q(\boldsymbol{x}, \boldsymbol{\theta})} d\boldsymbol{\theta} \\ &\equiv F[q] \end{aligned}$$

ここで得られた $F[q]$ と対数周辺尤度の差を取るとテスト分布 q と真の分布 p の KL ダイバージェンスが得られる．

$$\begin{aligned} \log p(\boldsymbol{y} \mid \boldsymbol{\alpha}) - F[q] &= \sum_{\boldsymbol{x}} \int q(\boldsymbol{x}, \boldsymbol{\theta}) \log \frac{q(\boldsymbol{x}, \boldsymbol{\theta})p(\boldsymbol{y} \mid \boldsymbol{\alpha})}{p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta} \mid \boldsymbol{\alpha})} d\boldsymbol{\theta} \\ &= KL(q \| p) \end{aligned}$$

つまり $F[q]$ を q に関して最大化することは q と p の距離を最小化することになり，q は良い p の近似と言える．この $F[q]$ の最大化問題は汎関数に対する微分方程式を用いることで以下の 2 式を繰り返し計算することで解ける．

$$\begin{cases} q^{(k+1)}(\boldsymbol{x}) & \propto \exp\left(\mathrm{E}[\log p(\boldsymbol{x}, \boldsymbol{y} \mid \boldsymbol{\theta})]_{q^{(k)}(\boldsymbol{\theta})}\right) \\ q^{(k+1)}(\boldsymbol{\theta}) & \propto p(\boldsymbol{\theta} \mid \boldsymbol{\alpha}) \exp\left(\mathrm{E}[\log p(\boldsymbol{x}, \boldsymbol{y} \mid \boldsymbol{\theta})]_{q^{(k+1)}(\boldsymbol{x})}\right) \end{cases}$$

上式に $\mathcal{X}$ の分布を代入すれば，$q^{(k+1)}(\boldsymbol{x})$ の更新は以下となる．

$$q^{(k+1)}(\boldsymbol{x}) = \prod_{i=1}^{N_P} \prod_{j=1}^{N_i} \left(\psi_{ij}^{(k)}\right)^{\sigma_{ij}(\boldsymbol{x})}$$

$$\psi_{ij}^{(k)} \equiv \exp\left(\Psi\left(\alpha_{ij}^{(k)}\right) - \Psi\left(\sum_{j'=1}^{N_i} \alpha_{ij'}^{(k)}\right)\right)$$

ここで $\Psi(\alpha)$ は digamma 関数，$\alpha_{ij}^{(k)}$ はテスト分布 $q^{(k)}(\theta)$ のハイパーパラメータである．更新された $q^{(k+1)}(\boldsymbol{x})$ を用いてハイパーパラメータ $\alpha_{ij}^{(k+1)}$ は以下の様に更新される．

$$\alpha_{ij}^{(k+1)} = \alpha_{ij} + \mathrm{E}[\sigma_{ij}(\boldsymbol{x})]_{q^{(k+1)}(\boldsymbol{x})}$$

期待値 $\mathrm{E}[\sigma_{ij}(\boldsymbol{x})]_{q^{(k+1)}(\boldsymbol{x})}$ は [Ishihata 10] における期待値計算において，θ_{ij} の代わりに $\psi_{ij}^{(k)}$ を用いることで計算可能である．従って VB 法も BDD 上で効率的に実効可能である．

5. 実験

ここではまず，MCMC 法と VB 法の性質を確認するため両者をクラスタリングにおけるモデル選択に応用する．次に，これらをアブダクションによって得られた仮説評価に応用する．

5.1 モデル選択への応用

ベイズ推定の応用としてモデル選択が知られている．例えばクラスタリングにおいてクラスタ数が未知である場合，クラスタ数をどのように決定するかは問題となる．これを客観的に決定する方法の 1 つとして，ベイズ推定を用いたモデル選択が挙げられる．ベイズ的なモデルでは，モデルの事後分布をモデルの選択基準に利用できる．モデルの事前分布を一様分布と仮定した場合，モデルの事後分布は周辺対数尤度 $\ln p(\boldsymbol{y} \mid \boldsymbol{\alpha})$ に比例する．通常，周辺対数尤度の計算は困難であるため，MCMC 法や VB 法によりその近似値を計算する．MCMC 法では対数周辺尤度のサンプル近似 (MLL) をモデル選択に用いる．VB 法では最大化の対象である $F[q]$(VFE) が周辺対数尤度の下限値となることからこれをモデル選択に用いる．

ここでは UCI リポジトリの Zoo データを naive Bayes モデルによってクラスタリングを行い，クラスタ数を変化させた場合の MLL, VFE, BIC (ベイズ情報料基準) を計算する．図 5.1 はクラスタ数を変化させたときの MLL, VFE, BIC の変化を表すグラフである．グラフより MCMC 法により得られた周辺対数尤度の近似値はクラス数が増加するに連れて非常に緩やかに減少していくのに対し，VB 法によって得られた周辺対数尤度の下限値 (VFE) と BIC は急速に減少していくことが分かる．これはモデルが複雑になるに連れて BIC や VB 法で仮定する近似分布が真の分布とかけ離れる事で近似精度が低くなることが原因と考えられる．この結果では MLL, VFE ともにクラスタ数 6 付近でピークが現れており，真のクラスタ数 7 に近い値が選ばれている．しかし，これらの値は事前分布に非常に大きく影響するため常に正しいモデル数が得られるわけではない．

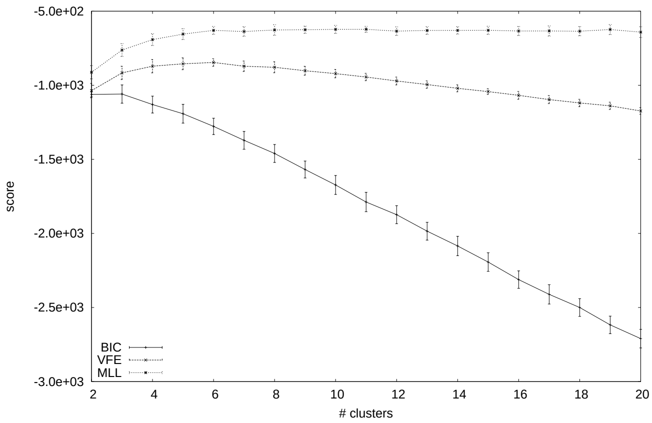

図 1: ベイズ推定に基づくモデルスコア

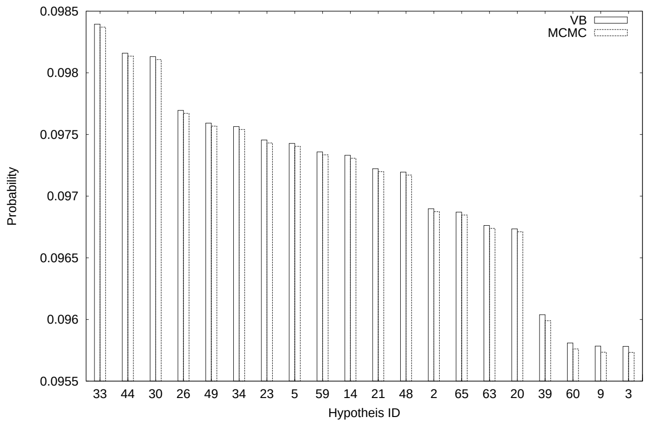

図 2: 仮説のランキング結果

5.2 統計的アブダクションへの適用

次に提案法をアブダクションによって得られた仮説の統計的評価に適用する．アブダクションは論理推論の 1 つであり，背景知識 KB から観測 O を論理的に説明するような仮説 H を発見する推論である．一般にアブダクションによって得られる仮説 H の数は膨大であり，これらをどう評価するかが問題となる．統計的アブダクションは仮説を構成する命題変数上に確率分布を導入し，その分布を用いて仮説 H の確率を計算することで，それらを統計的に評価する．統計的アブダクションの応用として，代謝経路に関する仮説発見などが挙げられる [Inoue 09]．しかし先行研究では，化学反応の有無などの決定的な知識は論理式として表現可能であるのに対し，個々の反応の起こりやすさといった経験的な知識はうまく表現出来ない．これに対してベイズ推定を用いれば，個々の反応の起こりやすさをパラメータの事前分布として自然にモデルに導入できる．

ここでは [Inoue 09] で得られた 66 個の仮説をベイズ推定を用いてランキングを行う．従来法では仮説の確率を最尤推定で学習した．しかしこのデータでは，ある 2 つの化学反応の有無が仮説を評価する上で重要であり，それらの反応は経験的に阻害されやすいことが知られている．本実験ではこの「阻害されやすい」という経験的知識をパラメータの事前分布として表現し，ベイズ推定を用いて各仮説の事後確率を計算する．図 5.2 は事後分布より計算された各仮説の事後確率のうち上位 20 個を表示したグラフである．ここで仮説の事後確率は MCMC 法と VB 法の 2 種類の方法で計算できるが，図 5.2 では VB 法による事後確率で仮説ランキングした．グラフより MCMC 法と VB 法の事後確率はほぼ等しくなっていることが分かる．更に [Inoue 09] で述べられた好ましい 22 個の仮説に注目すると，ランキングの上位 16 仮説がそれに該当することが分かった．これより事前分布を用いることで，経験的な知識を考慮した仮説のランキングが可能であることが確認できた．

6. まとめ

本研究では離散確率変数上確率分布を命題化することで，任意の論理式を扱える確率分布を定義し，その命題化された確率分布に対するベイズ推定を提案した．ベイズ推定を実行する方法として MCMC 法と VB 法を提案し，簡単な確率モデルに対して双方を実行し，それらの性質を比較した．更に，ベイズ推定をアブダクションによって得られた仮説の統計的な評価に適用し，経験的な知識を表現する事前分布を用いることで好ましいランキングが得られることを確認した．

参考文献

[Inoue 09] Inoue, K., Sato, T., Ishihata, M., Kameya, Y., and Nabeshima, H.: Evaluating abductive hypotheses using an EM algorithm on BDDs, in *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*, pp. 810–815 (2009)

[Ishihata 10] Ishihata, M., Kameya, Y., Sato, T., and Minato, S.: An EM algorithm on BDDs with order encoding for logic-based probabilistic models, in *Proceedings of the 2nd Asian Conference on Machine Learning (ACML'10)* (2010)

[Johnson 07] Johnson, M. and Griffiths, T. L.: Bayesian inference for PCFGs via Markov chain Monte Carlo, in *Proceedings of the North American Conference on Computational Linguistics (NAACL'07)* (2007)

[MacKay 97] MacKay, D. J.: Ensemble learning for hidden Markov models, in *Technical report, Cavendish Laboratory, University of Cambridge* (1997)

[Sato 01] Sato, T. and Kameya, Y.: Parameter Learning of Logic Programs for Symbolic-statistical Modeling, *Journal of Artificial Intelligence Research*, Vol. 15, pp. 391–454 (2001)

[Sato 09] Sato, T., Kameya, Y., and Kurihara, ichi K.: Variational Bayes via propositionalized probability computation in PRISM, *Annals of Mathematics and Artificial Inteligence*, Vol. 54, No. 1-3, pp. 135–158 (2009)

[Sato 11] Sato, T.: A General MCMC Method for Bayesian Inference in Logic-based Probabilistic Modeling, in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11)* (2011)

[亀谷 11] 亀谷由隆：論理に基づく確率モデリングのこれまで，これから．, 第 4 回情報論的学習理論と機械学習 (IBISML) 研究会 招待講演 (2011)

[栗原 04] 栗原賢一, 亀谷由隆, 佐藤泰介：動的計画法に基づく確率文脈自由文法の変分ベイズ法, 情報処理学会 自然言語処理研究会 (NL) 159-29, pp. 209–214 (2004)