# belief revision アルゴリズムとは

2010-03-25

産業技術総合研究所 脳神経情報研究部門
一杉 裕志

# 内容

- max と argmax
- MPE(most probable explanation)
- 脳とMPE
- belief revision アルゴリズム
- 確率伝播アルゴリズムと belief revision アルゴリズムの一般形

# 前提知識

- 下記資料をお読みください。
  「確率伝播アルゴリズムとは」
  http://staff.aist.go.jp/y-ichisugi/besom/20100226what-is-bp.pdf
  - 以下の概念を解説しています。
    - 確率論の基礎
      - 同時確率、条件付確率、周辺化、ベイズの定理
    - ベイジアンネット
    - 確率伝播アルゴリズム

# 概要

- 脳が行う認識アルゴリズムの候補として、確率伝播アルゴリズムよりも、MPE を計算する belief revision アルゴリズムの方が有望に思えるので、それについて解説する。
- belief revision アルゴリズムについては 書籍 [Pearl 1988] に書かれているが、分かりにくい点があるので、自分なりに補足して説明する。
（間違いがあればご指摘ください。）

y-ichisugi@aist.go.jp

# max と argmax

# max

- max(a,b) は、a と b のうち大きいほうの値。
  例：max(3,5)=5
- $\max_x f(x)$ は、x の定義域の中で最大の f(x) の値。
  例：x∈{1,2,3,4} の時、 $\max_x(-(x-2)^2)$
  $= \max(-1,0,-1,-4) = 0$
- 本文書ではΣと同様の文法に従うものとする。
  - [Pearl 1988] も同様。

$$\max_A B \max_C D \max_E F = \max_A (B \max_C (D \max_E F))$$

# argmax

- $\mathrm{argmax}_x$ f(x) は、f(x) を最大にする時の x の値。
  - 例: x ∈{1,2,3,4} の時、

$$\arg\max_x(-(x-2)^2) = 2$$

- [Pearl 1988] p.255 では $\max_x^{-1}$ f(x) と表記しているが、同じもの。

# 多変数の場合

下記の関係が成り立つ。

$$\max_{a,b} f(a,b) = \max_a(\max_b f(a,b)) = \max_b(\max_a f(a,b))$$

$$\arg\max_{(a,b)} f(a,b) = (\arg\max_a(\max_b f(a,b)),\ \arg\max_b(\max_a f(a,b)))$$

例：

| | | | | |
|---|---|---|---|---|
| f(1,1)=1 | f(2,1)=4 | f(3,1)=6 | f(4,1)=4 | $\max_a$f(a,1)=6 |
| f(1,2)=2 | f(2,2)=2 | f(3,2)=2 | f(4,2)=5 | $\max_a$f(a,2)=5 |
| f(1,3)=5 | f(2,3)=3 | f(3,3)=1 | f(4,3)=1 | $\max_a$f(a,3)=5 |
| f(1,4)=3 | f(2,4)=7 | f(3,4)=3 | f(4,4)=2 | $\max_a$f(a,4)=7 |
| $\max_b$f(1,b)=5 | $\max_b$f(2,b)=7 | $\max_b$f(3,b)=6 | $\max_b$f(4,b)=5 | $\max_{a,b}$f(a,b)=7 |

# MPE (most probable explanation)

# MPE (most probable explanation)

- ベイジアンネット内すべての変数の集合を **W** とし、その一部の変数の観測値 **e** が与えられた時、**W** の値の組 **w** のうち P(**w**|**e**) を最大にするものを MPE と呼ぶ。

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} P(\mathbf{w} \mid \mathbf{e})$$

- 意味は「観測事実に対する一番ありそうな解釈」、「一番もっともらしい説明」。

参考: [Pearl 1988] p.250

# MPEと同時確率

- 観測値が与えられていない変数の集合の値の組を **h** とすると、MPE は下記のようにも書ける。

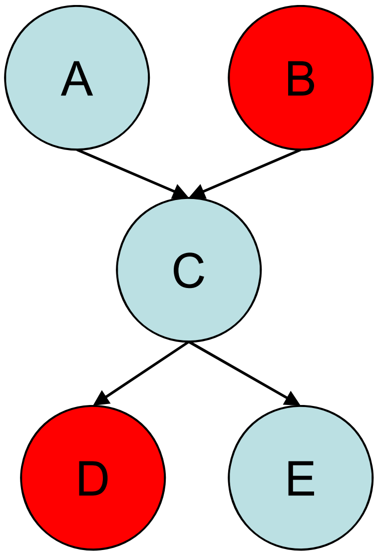
$$
\begin{aligned}
&\mathbf{h}^* = \arg\max_{\mathbf{h}} P(\mathbf{h} \mid \mathbf{e}) \\
&= \arg\max_{\mathbf{h}} P(\mathbf{h}, \mathbf{e}) / P(\mathbf{e}) = \arg\max_{\mathbf{h}} P(\mathbf{h}, \mathbf{e}) \\
&\mathbf{w}^* = \mathbf{h}^* \cup \mathbf{e}
\end{aligned}
$$

- **h*** をMPEと呼ぶ場合もある。

# 具体例

- 下記ネットワークにおいて B, D の観測値 b, d, が与えられたとする。MPE となる値の組 A=a*,C=c*,E=e* は下記のように定義される値。

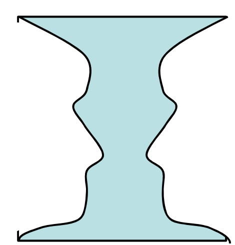
$$
\begin{aligned}
&(a^*, c^*, e^*) \\
&= \arg\max_{(a,c,e)} P(a, c, e \mid b, d) \\
&= \arg\max_{(a,c,e)} P(a, b, c, d, e)
\end{aligned}
$$



# 脳とMPE

# 曖昧図形とMPE

- 人間は壷にも横顔にも解釈できる曖昧図形を、ある瞬間にはそのどちらか１つだと認識し、決して壷５１％横顔４９％というような認識はしない。
  つまり、脳の認識の結果は、確率伝播アルゴリズムが計算する「ノードごとの周辺事後確率」よりも、MPEの方に似ている。

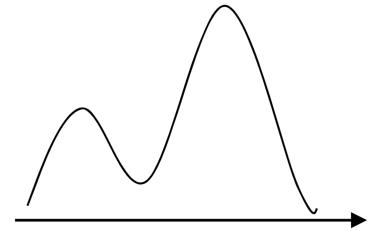


# EMアルゴリズムとMPE

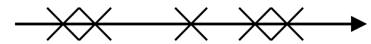
- 脳の学習アルゴリズムとして、EMアルゴリズムは有望な候補だろう。
  - 感覚連合野のニューロンを隠れ変数と考える。
- 隠れ変数の確率分布の代わりに、k-mean のように1つの値(=MPE)を使うことについて[Neal and Hinton 1998] ６章で少し論じられている。
  - 計算量が少なくて済むという利点がある。
  - 理論的には必ずしも収束しないが、全く ad hoc で無意味、というわけでもない。
- なおk-meanはEMのある極限と見なせる。(ビショップ本下 p.159 参照)

# モンテカルロEMと近似MPE

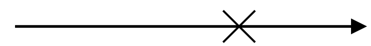
- 隠れ変数の確率分布の代わりに、確率分布から得られたサンプルを使う**モンテカルロEMアルゴリズム**がある。サンプルが1つだけの場合、**確率的EM**と呼ぶ。(ビショップ本 下 p.250-251)
- 近似MPE計算は局所解に陥ることがあるが、それにより厳密MPEよりもサンプル抽出のよい近似になる、かもしれない？



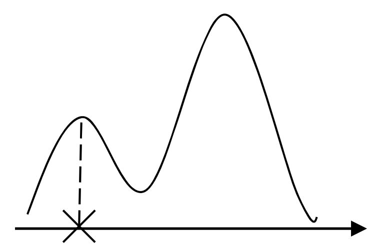
確率分布



サンプル5個



サンプル1個



MPEの局所解

# belief revisioin アルゴリズム

# belief revision アルゴリズムとは

- loop のないベイジアンネットにおいて、MPEの厳密解を効率的に計算するアルゴリズム。
  - 確率伝播アルゴリズムほど知名度がないので、定訳はない？
    - 信念改訂アルゴリズムと呼ぶのが適切か？
  - 人工知能の分野で belief revision という用語があるようだが、まったく別のもの。
- 隠れマルコフモデル(HMM)専用の belief revision アルゴリズムは、**ビタビ(Viterbi )アルゴリズム**としてよく知られている。

# 原理

- a >=0 のとき

$$\max(ab, ac) = a \max(b, c)$$

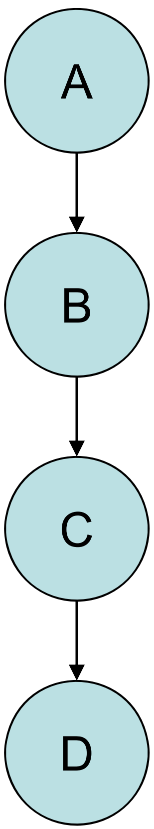
  - 同様に

$$\max_x(a f(x)) = a \max_x f(x)$$

- したがって確率伝播アルゴリズムの導出と同様の最適化が行える。

# 簡単なベイジアンネットの例

- このネットワークを例に、D の観測値 d が与えられた時のMPE ：
  $(a^*,b^*,c^*)=\mathrm{argmax}_{(a,b,c)}\ P(A=a,B=b,C=c,D=d)$
  を計算するアルゴリズムを導出してみる。




$$P(A,B,C,D)$$
$$=P(D|C)P(C|B)P(B|A)P(A)$$

# MPEの計算式

$$c^* = \arg\max_{c} \max_{a,b} P(a,b,c,d)$$

$$= \arg\max_{c} (\max_{a,b} (P(d \mid c) P(c \mid b) P(b \mid a) P(a)))$$

$$= \arg\max_{c} ( \quad \underline{P(d \mid c)} \max_{b} ( \quad \underline{P(c \mid b)} \max_{a} (P(b \mid a) P(a)) \quad ) \quad )$$

同様に

$$b^* = \arg\max_{b} \max_{a,c} P(a,b,c,d)$$

$$= \arg\max_{b} \quad (\max_{c} P(d \mid c) P(c \mid b)) \quad (\max_{a} P(b \mid a) P(a))$$

$$a^* = \arg\max_{a} \max_{b,c} P(a,b,c,d)$$

$$= \arg\max_{a} \quad (\max_{b} \quad (\max_{c} P(d \mid c) P(c \mid b)) \quad P(b \mid a)) \quad P(a)$$

# メッセージ

- 関数 $m_{AB}, m_{BC}, m_{DC}, m_{CB}, m_{BA}$ を下記のように定義。確率伝播アルゴリズムの時と同様に $m_{XY}$ を「XからYへのメッセージ」と呼ぶ。

$$m_{AB}(a) = P(a)$$

$$m_{BC}(b) = \max_a P(b|a)P(a) = \max_a P(b|a)m_{AB}(a)$$

$$m_{DC}(c) = P(d|c)$$

$$m_{CB}(b) = \max_c P(d|c)P(c|b) = \max_c m_{DC}(c)P(c|b)$$

$$m_{BA}(a) = \max_b \quad (\max_c P(d|c)P(c|b)) \quad P(b|a)$$

$$= \max_b m_{CB}(b)P(b|a)$$

# belief revisionアルゴリズム

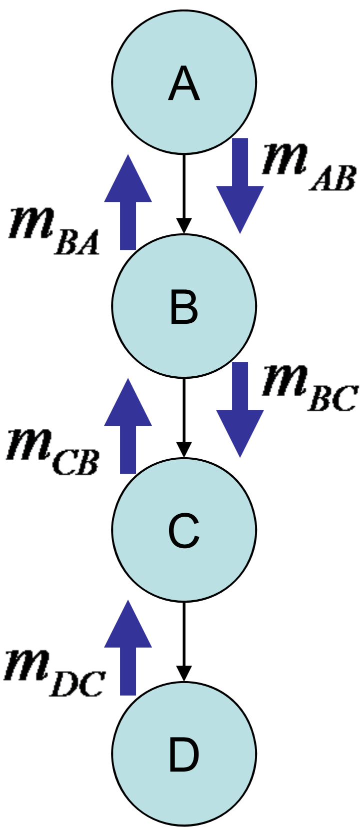
$$m_{AB}(a) = P(a)$$

$$m_{BC}(b) = \max_a P(b \mid a) m_{AB}(a)$$

$$m_{DC}(c) = P(d \mid c)$$

$$m_{CB}(b) = \max_c m_{DC}(c) P(c \mid b)$$

$$m_{BA}(a) = \max_b m_{CB}(b) P(b \mid a)$$




$$c^* = \arg\max_c m_{DC}(c) \max_b P(c \mid b) m_{BC}(b)$$

$$b^* = \arg\max_b m_{CB}(b) m_{BC}(b)$$

$$a^* = \arg\max_a m_{BA}(a) m_{AB}(a)$$

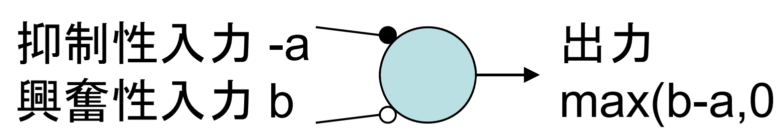
# belief revision アルゴリズムの性質

- belief revision アルゴリズムは、確率伝播アルゴリズムのΣをmaxに、正規化を argmax 変えたものに一致する。
- 確率伝播アルゴリズムと同様に、一般に下記の性質を持っている。
  - X が送るメッセージおよび $x^*$ は、X に送られてくるメッセージと X の条件付確率表という局所的な情報だけから計算できる。
  - どれが観測変数であっても、観測変数以外のノードが行う計算は変わらない。

# max を実行する神経回路

- $\max_a P(b|a)m_{AB}(a)$ とか $\max_b m_{CB}(b)P(b|a)$ という形の計算がたくさん出てくる。
  この計算は神経回路にとって、内積計算ほどは簡単ではない。
- しかしながら、大脳皮質内に存在しそうな比較的簡単な神経回路でmax が実現可能なことが示されている。[Litvak and Ullman 2009]

  a + max(b-a, 0) = max(a,b) という関係をうまく使う。
  ニューロンは非負の値しか出力しないので max(b-a,0) の演算は簡単に実現できる。



# 脳と belief revision

- ベイジアンネット（有向グラフ）ではなくマルコフ確率場（無向グラフ）で loopy belief revision を実行する spiking neuron の神経回路モデルが提案されている。[Litvak and Ullman 2009]
- 大脳皮質が確率伝播アルゴリズムを実行している証拠はいくつかある。これらの証拠は belief revision アルゴリズムの証拠にもなる可能性がある。
  - 今後、より詳細に検討していく予定。

# 確率伝播アルゴリズムと belief revision アルゴリズムの一般形

# loop のないベイジアンネットに対する確率伝播アルゴリズム [Pearl 1988]

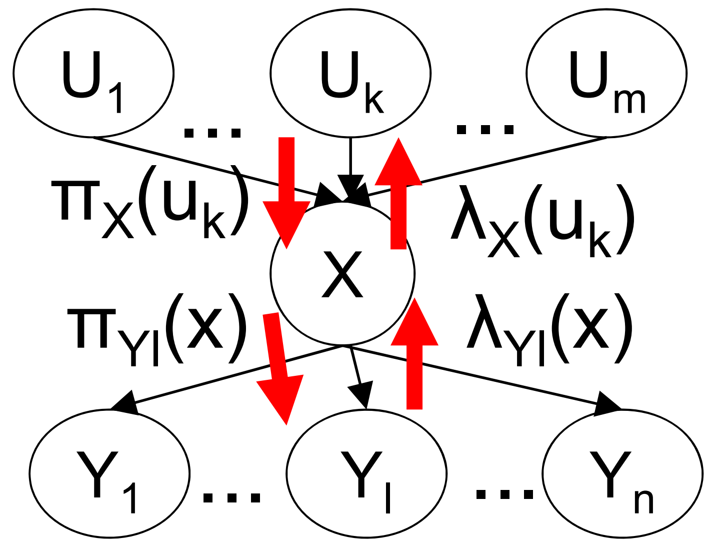
$$BEL(x) = \alpha\lambda(x)\pi(x)$$

$$\pi(x) = \sum_{u_1,\cdots,u_m} P(x|u_1,\cdots,u_m)\prod_k \pi_X(u_k)$$

$$\lambda(x) = \prod_l \lambda_{Y_l}(x)$$

$$\pi_{Y_l}(x) = \beta_1 \pi(x) \prod_{j\neq l} \lambda_{Y_j}(x)$$

$$\lambda_X(u_k) = \beta_2 \sum_x \lambda(x) \sum_{u_1,\cdots,u_m/u_k} P(x|u_1,\cdots,u_m)\prod_{i\neq k} \pi_X(u_i)$$



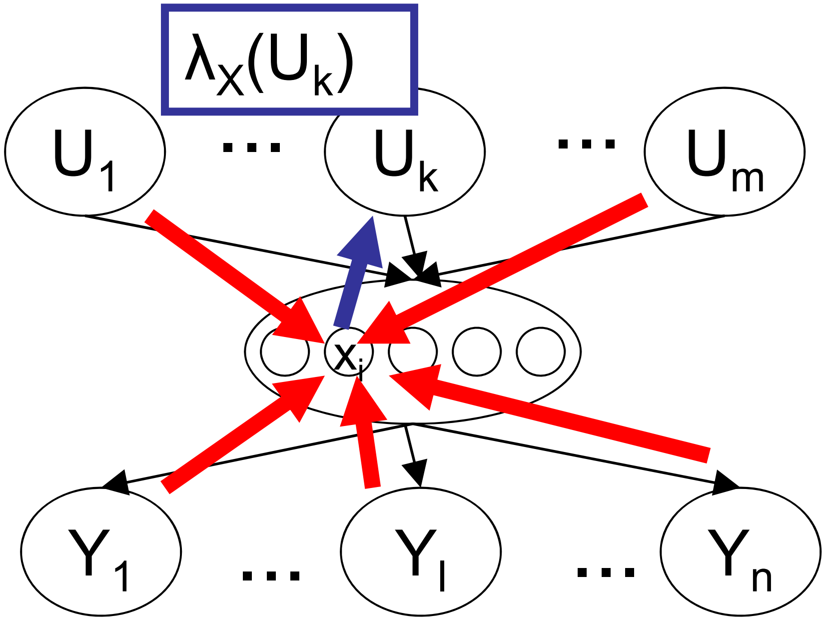

[Pearl 1988] p.182 。少し表記は変えてある。
なお、最上端、最下端ノードなどの境界条件については [Pearl 1988] p.181 参照

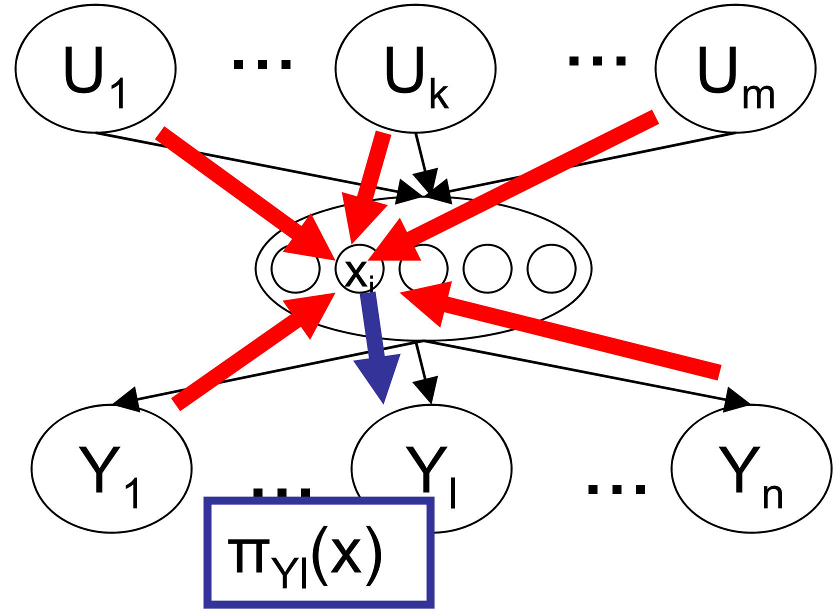
# 補足

- BEL(x) が変数 x の周辺事後確率。αは正規化定数。 $\alpha = 1/\sum_x \lambda(x)\pi(x)$
- β $_1$,β$_2$はメッセージの正規化定数だが、なくてもかまわない。(p.168)
- π(x),λ(x)はX内部の一時変数。他のノードに送られるメッセージではない。
- $\sum_{u_1,\cdots,u_m/u_k}$ は $\sum_{u_1,\cdots,u_{k-1},u_{k+1},\cdots,u_m}$ という意味。$u_k$ 以外の全ての $u_i$ の値の組についての総和。
- $\prod_{i\neq k}$ は $\prod_{i=1,\cdots,k-1,k+1,\cdots,n}$ という意味。k 以外の全ての i についての総積。

# メッセージに含まれる情報の性質

- メッセージ送信相手からの情報を除外している。




$$\lambda_X(u_k) = \beta_2 \sum_x \lambda(x) \sum_{u_1,\cdots,u_m/u_k} P(x|u_1,\cdots,u_m) \prod_{i\neq k} \pi_X(u_i)$$

$$\pi_{Y_l}(x) = \beta_1 \pi(x) \prod_{j\neq l} \lambda_{Y_j}(x)$$

# loop のないベイジアンネットに対する belief revision アルゴリズム [Pearl 1988]

- [Pearl 1988] p.255 では下記のように表現されている。
  - ただし関数名につく＊（アスタリスク）は省略した。使用する記号も少し変えた。
  - 誤りと思われる点を2点修正した。

$$F(x,u_1,\cdots,u_m) = (\prod_l \lambda_{Y_l}(x))P(x|u_1,\cdots,u_m)\prod_k \pi_X(u_k)$$

かっこを追加

$$x^* = \arg\max_x BEL(x)$$

$$BEL(x) = \max_{u_1,\cdots,u_m} F(x,u_1,\cdots,u_m)$$

βは不要なので削除

$$\lambda_X(u_k) = \max_{x,u_1,\cdots,u_m/u_k} (F(x,u_1,\cdots,u_m)/\pi_X(u_k))$$

$$\pi_{Y_l}(x) = \beta \frac{BEL(x)}{\lambda_{Y_l}(x)}$$

# belief revision アルゴリズムの[Pearl 1988]の表現を変形してみる

$$BEL(x) = \max_{u_1,\cdots,u_m} F(x, u_1, \cdots, u_m)$$

$$= \max_{u_1,\cdots,u_m} (\prod_l \lambda_{Y_l}(x)) P(x \mid u_1, \cdots, u_m) \prod_k \pi_X(u_k)$$

$$= (\prod_l \lambda_{Y_l}(x)) \max_{u_1,\cdots,u_m} P(x \mid u_1, \cdots, u_m) \prod_k \pi_X(u_k) = \lambda(x)\pi(x)$$

$$\lambda_X(u_k) = \max_{x, u_1,\cdots,u_m / u_k} (F(x, u_1, \cdots, u_m) / \pi_X(u_k))$$

$$= \max_{x, u_1,\cdots,u_m / u_k} (\prod_l \lambda_{Y_l}(x)) P(x \mid u_1, \cdots, u_m) \prod_{i \neq k} \pi_X(u_i)$$

$$= \max_x \lambda(x) \max_{u_1,\cdots,u_m / u_k} P(x \mid u_1, \cdots, u_m) \prod_{i \neq k} \pi_X(u_i)$$

$$\pi_{Y_l}(x) = \beta \frac{BEL(x)}{\lambda_{Y_l}(x)} = \beta \pi(x) \prod_{j \neq l} \lambda_{Y_j}(x)$$

**結局、確率伝播アルゴリズムのΣをmaxに変えたものに一致する。(p.256にもそう書いてある。)**

# loop のないベイジアンネットに対する belief revision アルゴリズム

$$x^* = \arg\max_x BEL(x)$$

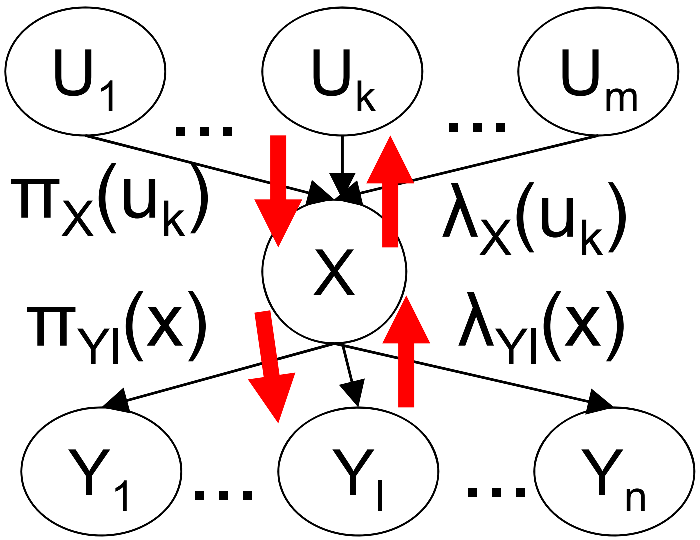
$$BEL(x) = \lambda(x)\pi(x)$$

$$\pi(x) = \max_{u_1,\cdots,u_m} P(x \mid u_1,\cdots,u_m)\prod_k \pi_X(u_k)$$

$$\lambda(x) = \prod_l \lambda_{Y_l}(x)$$

$$\pi_{Y_l}(x) = \beta_1 \pi(x) \prod_{j \neq l} \lambda_{Y_j}(x)$$

$$\lambda_X(u_k) = \beta_2 \max_x \lambda(x) \max_{u_1,\cdots,u_m / u_k} P(x \mid u_1,\cdots,u_m)\prod_{i \neq k} \pi_X(u_i)$$




最上端、最下端ノードなどの境界条件は確率伝播アルゴリズムと同じ。
[Pearl 1988] p.256 参照。

# 参考文献


- [Pearl 1988]
  Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference
  Judea Pearl: Morgan Kaufmann Pub ; ISBN: 1558604790
- 「ビショップ本」
  ・パターン認識と機械学習 上 - ベイズ理論による統計的予測
  C. M. ビショップ (著), 元田 浩 (翻訳), 栗田 多喜夫 (翻訳), 樋口 知之 (翻訳), 松本 裕治 (翻訳), 村田 昇 (翻訳) 出版社: シュプリンガー･ジャパン株式会社 (2007/12/10) ISBN-13: 978-4431100133
  ・パターン認識と機械学習 下 - ベイズ理論による統計的予測
  C. M. ビショップ (著), 元田 浩 (翻訳), 栗田 多喜夫 (翻訳), 樋口 知之 (翻訳), 松本 裕治 (翻訳), 村田 昇 (翻訳) 出版社: シュプリンガー･ジャパン株式会社 (2008/7/11) ISBN-13: 978-4431100317
- [Neal and Hinton 1998] Radford Neal, Geoffrey E. Hinton, A View of the EM algorithm that justifies incremental, sparse, and other variants, In Learning in Graphical Models (1998), pp. 355-368.
  http://www.cs.toronto.edu/pub/radford/emk.pdf
- [Litvak and Ullman 2009] Shai Litvak, Shimon Ullman: Cortical Circuitry Implementing Graphical Models, Neural Computation 21, 3010.3056 (2009)
  http://www.mitpressjournals.org/doi/abs/10.1162/neco.2009.05-08-783