情報システム工学科
伊達 章

神経回路網特論
2006 年 7 月 6 日

確率的モデルによる情報処理

1 例

図 1 に示すようなマルコフ的情報源を考えよう．

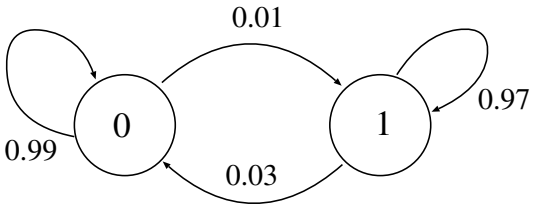

図 1: 状態遷移図

この情報源は 0, 1 の 2 つの状態をもち，状態間を時々刻々遷移する．状態は丸で囲われており，状態遷移の確率は状態間に引かれた矢印のついた線の上に書かれている．図では，状態 0 にあったときに，次もまた状態 0 のままでいる確率が 0.99，状態 1 に移る確率が 0.01 であることを示している．

このような確率的な状態遷移を 500 回 繰り返したときの一例を図 2 左に示す．おおよそ 100 回の状態遷移のうち 2,3 回，異なった状態に遷移する様子がわかる．これは図 1 に示したモデルより確率的に生成された結果という意味で，ランダムサンプル と呼ばれる．同じ事をも

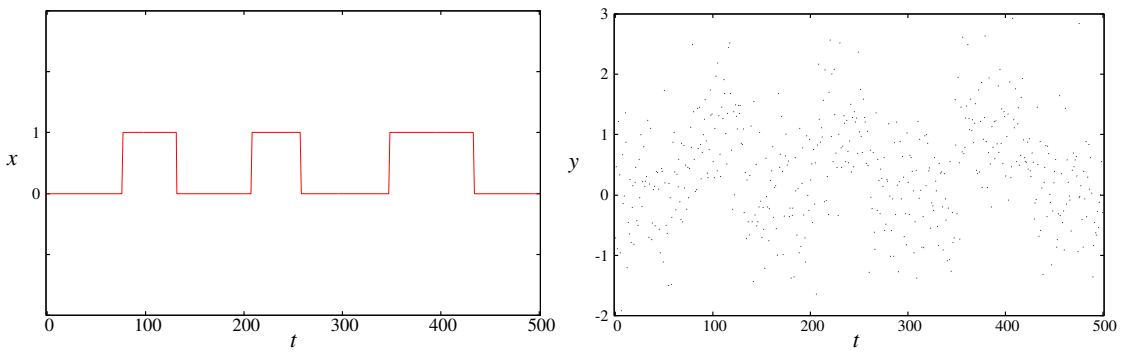

図 2: 左：ランダムサンプル．右：もとのデータにノイズが加わった観測データ

う一度おこなうと，100 回に数回，状態を遷移するという傾向は同じであるが，図 2 左とは異なった，もう一つのランダムサンプルが得られる．ここで，出現する可能性のある系列は 2^{500} 通りあり，その中には 00000000$\cdots$000 というもののから 0101010101$\cdots$0101 という滅多なことではみることができない系列も含まれる．

このように生成されたデータを遠く離れた地点に送ることを考えてみよう．ケーブルの性質が悪く，実際には，図 2 左にあるような信号のはずが，遠くから送られてくるデータを観測するため，もとのデータ（01 系列）にノイズが加わり，観測できたのは図 2 右に示すものだったとする．このデータを観測して，もとのデータがなんであったか．できるだけ正確に推定したい．これが問題である．

不幸中の幸いで，今，以下のことはわかっているとしよう．

- もとのデータは図 1 に示すモデルから生成されている．
- ノイズは平均 0，分散 $\sigma^2 = 0.7^2$ の正規分布にしたがっている．
- ノイズは各時間で独立（異なる時刻で観測されるノイズの相関はない）．

解決策の一つは，図 2 右をにらんで，どこで状態遷移がおこったか目で判断することである．それでいい結果が得られるならそれでよい．

図 3 には，これから紹介する動的計画法という手法を使って，もとのデータを推定した結果を示している．推定した結果をもとのデータと重ね合わせると見にくいので推定結果を上方向に 3 だけずらしている．

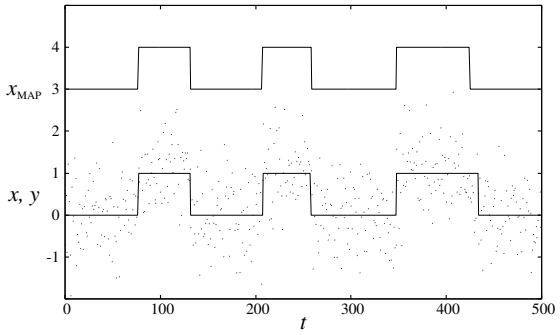

図 3: 事後確率を最大にする x の推定値 x_{MAP}

状態の変化 $0 \to 1$ や $1 \to 0$ が起った箇所を完全ではないもののの，かなり正確に当てていることがわかる．少なくとも人間の目で判断するのと同等以上の推定能力をもっていそうである．

2 マルコフチェーン

準備として状態 0 と状態 1 を遷移する単純なマルコフチェーンを考えよう．まずは初期状態であるが，時刻 $t = 0$ において，状態 0 にいる確率を $p_0 = 0.5$ する（確率は足し算すると 1 であるから $p_1 = 0.5$ であることもすぐにわかる）．時刻 t で $X_t = i$ という状態にいた場合に，時刻 $t + 1$ で $X_{t+1} = j$ という状態に遷移する確率 $\mathrm{Prob}\{X_{t+1} = j | X_t = i\}$ を p_{ij} と書く．この p_{ij} を並べてつくった $P = \{p_{ij}\}$ という行列を遷移行列といい，図 1 の場合，

$$\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 0.99 & 0.01 \\ 0.03 & 0.97 \end{bmatrix} \tag{1}$$

と書ける．

T 回遷移を繰り返すとして（上の例では $T = 500$ ），このマルコフチェーンから生成される最も尤もらしい 0 と 1 の系列は何だろうか．たとえば $11111\cdots111$ とすべて 1 がでる確率は $0.5 \times (0.97)^T$ である．可能な 01 の系列は 2^T 個ある．いろいろ考えるまでもなく，$00000\cdots000$ とすべて 0 である系列が最も高い確率 （$0.5 \times (0.99)^T$） で出現することに気づく．

この生成される最も尤もらしい 0 と 1 の系列を求めることを数学的に記述すると

$$\underset{x_0,x_1,\cdots,x_T}{\operatorname{argmax}} \ \mathrm{Prob}(X_0 = x_0, X_1 = x_1, \cdots, X_T = x_T) \tag{2}$$

となる．Prob の部分を確率変数 $X_0, X_1, \cdots, X_T$ の 同時確率分布 という．argmax という記号は，その同時確率 P を最大にする $x_0, x_1, \cdots, x_T$ の組み合わせ（0,1 の要素からなる T 次元ベクトル）という意味である．こういう風に書くと記述が長くなるので，これからは

$$\underset{x_0,x_1,\cdots,x_T}{\operatorname{argmax}} \ \mathrm{Prob}(x_0, x_1, \cdots, x_T) \tag{3}$$

と書くようにする．今の場合，これは $000000\cdots0$ である．この節では，当たり前のことをくどくど書くか，これはあとあと必要となる表現なので，辛抱していただきたい．

現在の状態に依存して，確率的に状態が遷移していく過程をマルコフ連鎖という．つまり過去の状態遷移履歴は，現在の状態にとりこまれているので，上式は

$$\underset{x_0,x_1,\cdots,x_T}{\operatorname{argmax}} \ \mathrm{Prob}(x_0)\mathrm{Prob}(x_1|x_0)\mathrm{Prob}(x_2|x_1)\cdots\mathrm{Prob}(x_T|x_{T-1}) \tag{4}$$

と条件付き確率を使って掛け算の形でかける．こう書けるところがミソでである．これももっと短く記述したいので，

$$\underset{x_0,x_1,\cdots,x_T}{\operatorname{argmax}} \ p_{x_0}p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{T-1}x_T} \tag{5}$$

と書く．例えば，p_0 は初期状態で状態 0 にいる確率，p_{01} は時刻 t で $X_t = 0$ という状態にいたと仮定した場合，時刻 $t+1$ で $X_{t+1} = 1$ という状態に遷移している確率である．ここでは簡単のため，この遷移確率は時間的には変動しないことにする．

さて，さきの例で $T = 500$ のとき，上記の確率は $0.5 \times (0.99)^{500}$ になる．これは非常に小さな正の数である．どのくらい小さな数か実際にコンピュータを使って計算しようとすると，アンダーフローがおこり計算できない（どうなるかやってみるといい）．そういうときには log をとればいいというわけで，$\log 0.5 + T\log(0.99)$ を計算し，その後 e の片に載せる．つまり $0.5 \times (0.99)^T = \mathrm{e}^{\log 0.5 + T\log(0.99)}$ とすると計算できる．

式 (5) を計算するためには確率を最大にする系列をみつけることが必要である．ここで log が単調増加関数であることを思いおこせば，式 (5) の中身の log を最大化しても同じであることがわかる，したがって問題は

$$\underset{x_0,x_1,\cdots,x_T}{\operatorname{argmax}} \ \{\log p_{x_0} + \sum_{t=1}^{T} \log p_{x_{t-1}x_t}\} \tag{6}$$

と書ける．

ここでいきなりではあるが，

$$C_0(i) \ = \ \log(p_i) \tag{7}$$

$$C_n(i) \ = \ \max_{x_0,x_1,\cdots,x_{n-1}} \{\log p_{x_0} + \sum_{t=1}^{n-1} \log p_{x_{t-1}x_t} + \log p_{x_{n-1}i}\} \tag{8}$$

という量を定義しておこう．$C_n(i)$ は状態遷移を時刻 n まで繰り返し，値 $i \in \{0,1\}$ で終わっているときの最も尤もらしい系列 （best path と呼ぶ）の発生確率である．たとえば，時刻 $n = 1$ では

$$C_1(i) \ = \ \max_{x_0}\{\log p_{x_0} + \log p_{x_0 i}\} \tag{9}$$

となる．さらに，この best path のうち X_{n-1} の状態を $S_n(i)$ と書くことにする．時刻 n での $C_n(i), i \in \{0,1\}$ が与えられたとすると，

$$S_{n+1}(j) \quad = \quad \underset{i}{\mathrm{argmax}}\{C_n(i) + \log p_{ij}\} \tag{10}$$

と書け，

$$C_{n+1}(j) \quad = \quad C_n(S_{n+1}(j)) + \log p_{S_{n+1}(j)j} \tag{11}$$

と書ける．表現がややこしく見えるが，この式はおさえておこう．$S_n(j)$ は時刻 $n-1$ の状態なので $S_{n-1}(n,j)$ と書いた方がわかりやすいかもしれないが，つねに $n-1, n$ の関係なので，余計なインデックスをつけないほうがいいし，このまま計算機プログラムを書けるといういみで，このままにしておく．

$S_{n+1}(j)$ の意味するところは，時刻 $n+1$ で状態 j にあると仮定したとき，「最も尤もらしい系列の時刻 n における状態」である．例えば，

$$S_1(j) \quad = \quad \underset{i}{\mathrm{argmax}}\{C_0(i) + \log p_{ij}\} \tag{12}$$

は，時刻 1 で状態 j にいる（$x_1 = j$）と仮定したときの x_0 の最適な推定値となる．

今はまだ $x_1 = 0$ なのか $x_1 = 1$ なのか分からないので，4 通りの系列（$x_0x_1 = 00, 01, 10, 11$）のうち，時刻 1 でのとりうる 2 つの状態に対してそれぞれ最適な x_0 の値 $S_1(0), S_1(1)$ を覚えておく必要がある．この $S_1(0)$ と $S_1(1)$ の値は具体的に計算できる．

$x_1 = j$ と仮定した場合，最適な x_0 が $S_1(j)$ である．これを式で書くと

$$C_1(j) \quad = \quad C_0(S_1(j)) + \log p_{S_1(j)j} \tag{13}$$

となる．これも具体的に計算できる．

このようなことを一般的に書いたものが最初の式である．時刻 n での最適な状態がわかると，時刻 n までの系列の発生確率に，それぞれの状態に遷移する確率をかければ，時刻 $n+1$ で状態 j で にいる最も尤もらしい系列の発生確率を計算できる．

1. $x_n = 0, 1$ それぞれの場合について最適な経路とその発生確率が既に分かっている．
2. それをもとにして $x_{n+1} = j$ の場合の最適な x_n と全体の系列の発生確率を計算する．
3. これを続ける．

つまり，4 通りの系列のうち，時刻 n でのとりうる 2 つの状態に対して最適な経路を常に 2 つをキープしておくことになる．時刻 T まで観測し，そのときの最も尤もらしい系列は

$$\begin{aligned}
\hat{x}_T &= \underset{i}{\mathrm{argmax}}\, C_T(i) && (14)\\
\hat{x}_{T-1} &= S_T(\hat{x}_T) && (15)\\
&\cdots \\
\hat{x}_{n-1} &= S_n(\hat{x}_n) \\
&\cdots \\
\hat{x}_1 &= S_2(\hat{x}_2) \\
\hat{x}_0 &= S_1(\hat{x}_1)
\end{aligned}$$

であると，順々に計算できる．

確率変数の依存性をはっきりさせるため，今後，確率変数間の依存性を図 4 に示すように表示する．線でつながっていないのは，依存性がないことを示している．たとえば X_5 のとりうる値は X_4 と X_6 にのみ依存している，といった具合いである．このような確率変数間の依存関係は図 1 のような状態遷移図をみていただけではわからない．図 4 のようなグラフ

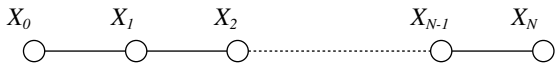

図 4: 確率変数の依存性を描いた図

を依存性グラフ，マルコフ確率場，状態空間モデルなどと呼ぶことがある．

以上の計算は，マルコフチェーンという，現在の状態は，古い過去には依存せず，直前にどの状態にいたかにだけで確率的に決まるという性質があるからこういうふうに書けた．普通はこんなにうまくはいかない．

3　隠れマルコフモデル

前節では，あたりまえのことをなぜ面倒な式を使って説明するのかと思ったかもしれない．本節では図 2 右のような状況を考える．観測できるのは $x_0, x_1, \cdots$ ではなく，$x_0, x_1, \cdots$ にノ

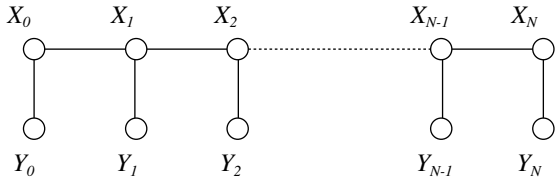

図 5: 確率変数の依存性：1 次元の隠れマルコフモデル

イズが加わった $y_0, y_1, \cdots$ である．

さて $x_0, x_1, \cdots$ は前節で考えたマルコフ性をもつ系列である．観測する $y_0, y_1, \cdots$ は図 5 に示すように，$Y_i = f(X_i) = X_i + Z_i$ という関係を満たしているものと考える．ここで Z_i は各時刻 i に独立に，確率に平均 0，標準偏差 $\sigma = 0.7$ の正規分布にしたがっているとする．このような状況を

$$Y_n \quad = \quad X_n + Z_n,\ Z_n \sim \mathcal{N}(0, (0.7)^2) \tag{16}$$

とかく．

観測データ $y_0, y_1, \cdots$ は $x_0, x_1, \cdots$ に依存して決まる．データを生成する源の $\{X_i\}$ はマルコフ性をもつが，$\{X_i\}$ の関数である $\{Y_i\}$ はマルコフ性をもたない．このようなモデルを隠れマルコフモデルという（x と y を関係付ける関数が 1 対 1 の非確率的な関数であれば $\{Y_i\}$ もマルコフ性をもつが，それは特殊な場合）．

目的は $y_0, y_1, \cdots$ を観測し，もとの $x_0, x_1, \cdots$ を推定することである．つまり $y_0, y_1, \cdots$ を観測し，最も尤もらしい $\hat{x}_0, \hat{x}_1, \cdots$ を求めることである．これを式で書くと

$$\underset{x_0, x_1, \cdots, x_T}{\operatorname{argmax}}\ \operatorname{Prob}(x_0, x_1, \cdots, x_T \mid y_0, y_1, \cdots, y_T) \tag{17}$$

となる．ここでベイズの公式を使い

$$\mathrm{Prob}(x_0,\cdots,x_T \mid y_0,\cdots,y_T) = \frac{\mathrm{Prob}(y_0,\cdots,y_T \mid x_0,\cdots,x_T)\mathrm{Prob}(x_0,\cdots,x_T)}{\mathrm{Prob}(y_0,y_1,\cdots,y_T)} \quad (18)$$

$$\mathrm{Prob}(x_0,\cdots,x_T \mid y_0,\cdots,y_T) = \frac{\mathrm{Prob}(x_0,\cdots,x_T,\ y_0,\cdots,y_T)}{\mathrm{Prob}(y_0,y_1,\cdots,y_T)} \quad (19)$$

となることがわかる．式 (19) の分母 $\mathrm{Prob}(y_0,y_1,\cdots,y_T)$ は $\mathrm{Prob}(x_0,\cdots,x_T \mid y_0,\cdots,y_T)$ の具体的な値をもとめたい場合には必要であるが，この確率を最大にする $x_0,\cdots,x_T$ を求めることについてだけであれば関係がない．確率変数間の依存性は図 5 に示すように，X_i はマルコフ性をもち，Y_i は X_i にのみ依存することがわかっているので，求めたい $x_0,x_1,\cdots$ は

$$\underset{x_0,x_1,\cdots,x_T}{\mathrm{argmax}}\ \mathrm{Prob}(x_0,x_1,\cdots,x_T,\ y_0,y_1,\cdots,y_T) \quad (20)$$

$$\underset{x_0,x_1,\cdots,x_T}{\mathrm{argmax}} \prod_{n=0}^{T} \mathrm{Prob}(y_n|x_n)\mathrm{Prob}(x_0) \prod_{n=1}^{T} \mathrm{Prob}(x_n|x_{n-1}) \quad (21)$$

となる．図 5 に示す依存性を示すグラフでは，各条件付確率が一つのリンクで表現されていることに注意したい．具体的にコンピュータを使って計算する場合は，前節と同じように式の中身の log をとり

$$\underset{x_0,x_1,\cdots,x_T}{\mathrm{argmax}}\ \{\log p_{x_0} + \sum_{t=1}^{T} \log p_{x_{t-1}x_t} + \sum_{t=0}^{T} \log q_{x_t y_t}\} \quad (22)$$

を求める．このような手法を事後確率最大化といい，推定した結果 $\hat{x}_0,\hat{x}_1,\cdots$ を事後確率最大推定値 (maximum a *posteriori* estimator, MAP) という．

前節と同様に

$$C_0(i) = \log(p_i) + \log(q_{iy_0}) \quad (23)$$

$$S_{n+1}(j) = \underset{i}{\mathrm{argmax}}\{C_n(i) + \log p_{ij} + \log q_{jy_{n+1}}\} \quad (24)$$

$$C_{n+1}(j) = C_n(S_{n+1}(j)) + \log p_{S_{n+1}(j)j} + \log q_{jy_{n+1}} \quad (25)$$

$$n = 0,1,\cdots,T-1$$

と書く．そうすると，時刻 T での x_T の推定値を

$$\hat{x}_T = \underset{i}{\mathrm{argmax}}\, C_T(i) \quad (26)$$

と求め，

$$\hat{x}_n = S_{n+1}(\hat{x}_{n+1}) \quad (27)$$

$$n = T-1,\cdots,0$$

と順々に計算できる．

以下の解説はは自由課題の際に参考にされたい．簡単にしか書いていないので，不明な点は随時，質問のこと．

4 パラメータ推定

モデルのパラメータ $p_{00} = 0.99, p_0 = 0.5$ などはわかっているものとして話を進めてきた．実は，パラメータの値がわかっていない場合でも，パラメータの値を観測データから推定し，うまくもとのデータを復元できる方法がある．これは EM (Expectation-Maximiziation) アルゴリズムと呼ばれている．

1. パラメータの初期値 $\{p_{ij}\}$ を乱数に設定（ただし，$\sum_j p_{ij} = 1,\ 0 \leq p_{ij} \leq 1$ に注意）．
2. 前節のとおり MAP 解を求める．
3. MAP 解 $\hat{x}_0, \hat{x}_1, \cdots, \hat{x}_T$ をもとにパラメータを推定する．その仕方はラグランジェの未定係数法を使って導き出せる．ここでは，パラメータの値をどう更新すればよいか，その結果だけを示す．

   X が状態 i にあるとした場合，次時刻に状態 j に遷移することは何回あったか，データ $\hat{x}_0, \hat{x}_1, \cdots, \hat{x}_T$ から経験的に計算する．これを n_{ij} と書く．

$$n_{ij} \quad = \quad \#\text{times make a transition from } i \text{ to } j \tag{28}$$

   具体的には $n_{00}, n_{01}, n_{10}, n_{11}$ を計算する．パラメータ p_{ij} の値を以下の $\tilde{p}_{ij}$ の値に更新する．

$$\tilde{p}_{ij} \quad = \quad \frac{n_{ij}}{\sum_{j=0}^{1} n_{ij}} = \frac{n_{ij}}{n_{i0} + n_{i1}} \tag{29}$$

4. 2 に戻り，パラメータの値が収束するまで，この作業を繰り返す．

5 周辺分布 Y を利用したデータのモデル化

$X_0, X_1, \cdots, X_T$ はマルコフという単純なものであるが，図 5 に示したモデル（線型グラフモデル）を使って，確率変数間 $Y_0, Y_1, \cdots, Y_T$ にどんな依存性があっても，その構造をいくらでも精度良くモデル化できることが知られている．具体的には，先の例では $Y_0, Y_1, \cdots, Y_T$ は単にノイズが含まれたデータであったが，音声データや画像の確率的な構造をモデル化できる．つまり，このタイプのモデルは任意の $\mathrm{Prob}(y_0, y_1, \cdots, y_T)$ を表現できる能力がある．もちろん x_i のとりうる値の数や $x_i \to y_i$ の関数をどのように設計するか考える必要がある．図 6 では，音声データをモデル化する具体例を示している．具体的には $X_i \in \{1, \cdots, 20\}$ として $Y_i = 1 + X_i \bmod 8$ としている．学習したパラメータは X の状態遷移の確率 $\{p_{ij}\}$ だけである．その学習方法は上に示してあるものと同じ方法を用いた．図 6 の左側には「ア」という音声をサンプリング周波数 8kHz，量子化ビット数 8 (256 値) で記録したものを，400 個（100msec）の 8 値データにまで粗くサンプリングしなおしたものである．図 6 の右側には，このデータだけから，状態遷移確率を上記の EM アルゴリズムによりもとめ，その後，その学習済のモデルからランダムサンプルをおこなった結果を示している．

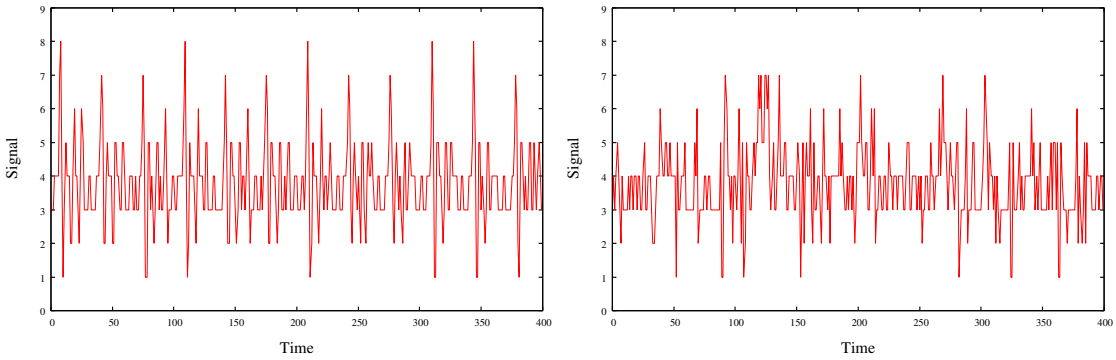

図 6: 左：粗くサンプリングした音声データ．右：状態数 20 でモデル化し，そのモデルからのランダムサンプル

6 画像のモデル化

画像をモデル化する場合，2 次元格子モデルがよく使われる．この場合，線型グラフの際に用いた動的計画法がを使えなくなる．ただし，並列処理で，事後確率を最大にする MAP 推定値を近似的に計算する Gibbs Sampler という方法が使える．なお，ツリー型のモデルを使えば，簡単な画像であれば，事後確率を最大にする推定値を動的計画法を使って求めることができる．

7 Gibbs Sampler

動的計画法を使えば，$\vec{x}_{\mathrm{MAP}}$ を正確に計算できることを既に紹介した．ここでは動的計画法を使わず，$\vec{x}_{\mathrm{MAP}}$ の近似解を計算する手法を紹介する．

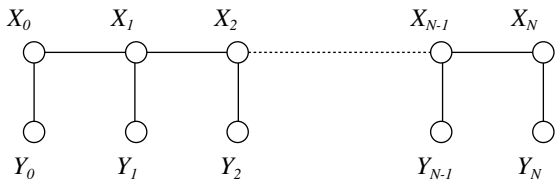

図 7: 確率変数の依存性：1 次元の隠れマルコフモデル

$y_0, y_1, \cdots, y_T$ が与えられているとする．求めたいのは

$$\begin{aligned} \mathrm{Prob}(x_0, \cdots, x_T \mid y_0, \cdots, y_T) &= \frac{\mathrm{Prob}(x_0, \cdots, x_T,\ y_0, \cdots, y_T)}{\mathrm{Prob}(y_0, y_1, \cdots, y_T)} && (30) \\ &= c\ \mathrm{Prob}(x_0, \cdots, x_T,\ y_0, \cdots, y_T) && (31) \\ &= c \prod_{n=0}^{T} \mathrm{Prob}(y_n|x_n)\mathrm{Prob}(x_0) \prod_{n=1}^{T} \mathrm{Prob}(x_n|x_{n-1}) && (32) \end{aligned}$$

を最大にする $x_0, x_1, \cdots, x_T$ であった．最大化には関係のない定数になる部分を c とした．

1. 初期値として $x_0, x_1, \cdots, x_T$ に 0,1 をランダムに割り当てる．
2. 0 から T までの数から一つをランダムに選び，それを i とし，x_i に着目する．

3. 式 (32) を最大にすることを考えた場合，x_i に関係するのは，$\mathrm{Prob}(x_i|x_{i-1}), \mathrm{Prob}(x_{i+1}|x_i), \mathrm{Prob}(y_i|x_i)$ の 3 項である．図で考えると 3 本のリンクに対応している．x_i の値が 0 の場合と，1 の場合について，これらの 3 項の値を計算し掛け算した値をそれぞれ h_0, h_1 とする．[0:1] の一様乱数を出力し，それが $h_0/(h_0+h_1)$ より小さければ，x_i を 0 に設定し，そうでなければ 1 に設定する（$h_0 > h_1$ なら $x_i = 0$ そうでなければ 1 というのでもおそらくうまくいく）．

4. 2. にもどり繰り返す.

（この方法で，どうして近似解が得られるのかという議論が必要．ここでは省略）

2 次元格子型のマルコフ確率場の場合，動的計画法が実質使えなくなるので，この方法が有効となる．

8 正確な事後確率の計算

事後確率を最大にする MAP 推定値 $\hat{x}_0, \hat{x}_1, \cdots, \hat{x}_T$ を計算することはできたが，その正確な事後確率

$$\max_{x_0, x_1, \cdots, x_T} \mathrm{Prob}(x_0, x_1, \cdots, x_T \mid y_0, y_1, \cdots, y_T) \tag{33}$$

の値は非常に小さく，通常は計算できない．実は，線型グラフの場合は，これを計算機で正確に効率よく計算する方法がある．

【ヒント】分子と分母を入れ替えて 3 節の計算と似た作業をおこなう [2].

9 マルコフ確率場とギブス分布

確率変数の依存関係が一般的にグラフ（ノードとリンク）で書ける.「マルコフ確率場とギブス分布の同値性」について説明できなかった．互いの定義だけ以下に書いておく．以下では，$\boldsymbol{X}$ を確率変数のベクトル，$\boldsymbol{x}$ を具体的な値のベクトル，${}_i\boldsymbol{x}$ を $\boldsymbol{x}$ から x_i を除いたベクトル，$\mathrm{NB}(i)$ を i 番目のノードと近傍関係にあるノードの集合とする．

定義： マルコフ確率場（MRF）

確率分布が

$$P(X_i = x_i|{}_i\boldsymbol{X} =_i \boldsymbol{x}) = P(X_i = x_i|\boldsymbol{X}_{\mathrm{NB}(i)} = \boldsymbol{x}_{\mathrm{NB}(i)}) \tag{34}$$

を満たすとき，その分布はグラフ G に関するマルコフ確率場であるという．上の式は長たらしい．通常は $P(x_i|{}_i\boldsymbol{x}) = P(x_i|x_{\mathrm{NB}(i)})$ と略して書く．

定義： Gibbs 分布

どの $\boldsymbol{x}$ に対しても $P(\boldsymbol{x}) > 0$ であり，$\mathcal{C}$ をクリークの集合としたとき

$$P(\boldsymbol{x}) = \prod_{c \in \mathcal{C}} f_c(x_c) \text{ と書ける確率分布を Gibbs 分布という．} \tag{35}$$

ほかにも面白い性質がたくさんある...

参考文献

[1] H. Künsch, S. Geman, and A. Kehagias, “Hidden Markov random fields,” Annals of Applied Probability, vol.5, no.3, pp.557–602, 1995.

[2] S. Geman and K. Kochanek. “Dynamic programming and the graphical representation of error-correcting codes,” IEEE Trans. Information Theory, vol.47, no.2, pp.549–567, Feb. 2001.