

大規模確率場と確率的画像処理の深化と展開

田中和之^{†1}

確率的画像処理はマルコフ確率場とベイズ推定をもとにして定式化される。その具体的なアルゴリズムは確率伝搬法により構成される。マルコフ確率場と確率伝搬法は物理モデルおよび物理的計算技法と深く関係したアプローチである。本稿ではこれらの理論的枠組みとその基礎となる情報統計力学的視点について解説する。

Deeping and Expansion of Large-Scale Random Fields and Probabilistic Image Processing

KAZUYUKI TANAKA^{†1}

The mathematical frameworks of probabilistic image processing are formulated by means of Markov random fields and Bayesian inference. Some practical algorithms are constructed by applying belief propagation methods. Markov random fields and belief propagation methods can be regarded as one of statistical-mechanical approaches to probabilistic information processing. In the present paper, we review some fundamental frameworks and statistical-mechanical approaches of probabilistic image processing.

1. はじめに

確率的画像処理とは画像の生成モデル (Generative Model) を確率をつかって仮定することを基本とする方法である。その用途はノイズ除去、領域分割、エッジ検出などの不完全データを伴う処理にあり、要請される処理を進める上で必要とする情報のすべてがデータの

中に含まれているとは限らない場合に有効に機能することが期待されている。不完全性を伴うデータを扱う情報処理において、その不完全性からくる不確実性を数式を用いて表現する伝統的数学の道具としての確率・統計をもちいるというわけである。通常、あつかわれる自然画像、風景画像は空間的に一様な性質を持っている場合は少なく、従来の空間フィルターだけでは対応しきれない場合がある。信号処理においてはこの問題を解決する手段のひとつとして適応フィルターとよばれるものが提案されている。これに対して確率的画像処理の場合は、生成モデルの性質は空間的に一様な統計的性質をもつものを仮定する。生成される画像はランダムネスを伴うため、その生成モデルの統計的性質が強く反映されている部分と必ずしもそうとはいえない部分が共存する画像が生成される。このランダムネスをともなう統計的性質を系統的に制御する鍵がベイズの公式 (Bayes Formula) である。

確率的画像処理では画像生成モデルとベイズの公式から構成される確率モデルの統計量を計算することが主な手順となる。しかし、この計算には通常は画像の大きさに対して指数関数的に計算量が増加してしまう計算の複雑さ (Computational Complexity) の問題がついてまわる。つい最近まで、あまり有効な方法として認知されることはなかった。しかし、確率伝搬法 (Belief Propagation) とよばれる方法が 1980 年代半ばに人工知能の分野で提案され、その後、実用的な確率的画像処理のアルゴリズムが提案されるようになる。また、画像生成モデル、確率伝搬法の数理構造のなかに物理モデルと物理的計算技法との関係が明らかになり、情報統計力学 (Statistical-Mechanical Informatics) という物理学における新しい研究分野の主要な研究テーマのひとつとしても認知されつつある。¹⁾⁻³⁾

本稿では、もっとも基本的な問題設定であるノイズ除去という用途において確率的画像処理についての説明を進めることにする。ノイズ除去はノイズの加えられてしまった画像からノイズの加えられる前のもとの画像を推定する問題である。これを信号処理の言葉になおせばノイズがデータ、すなわち観測信号であり、もとの画像が原信号ということになる。つまり原信号をデータから推定するという問題である。確率的画像処理の場合ももとの画像の統計的性質とランダムネスを伴うノイズを確率を用いてあらわし、ベイズの公式 (Bayes Formula) を用いて、データが与えられたときに、そのデータを生成した画像はもともとはどのような画像であったかをあらわす確率を導き、推定がおこなわれる。ベイズの公式をつかったこのようなアプローチはベイズ推定 (Bayesian Inference) などとよばれている。ベイズ推定においてももとの画像の統計的性質をあらわす確率を事前確率 (Prior Probability)、データが与えられたときに、そのデータを生成した画像はもともとはどのような画像であったかをあらわす確率を事後確率 (Posterior Probability) とよんでいる。

^{†1} 東北大学大学院情報科学研究科
Graduate School of Information Sciences, Tohoku University

2. 画像生成モデルと大規模確率場

画像処理において確率を導入する場合に、それぞれの画素 (Pixel) での光の強度、すなわち輝度値 (Light Intensity) に対して確率変数 (Random Variable) に対応させることになる。つまり画素の総数と同じ次元をもつ結合確率 (Joint Probability) をつかってあらわされることになる。 $|V|$ 個の画素が正方格子にならべられている場合を想定する。そのひとつひとつの画素に通し番号をつけて、すべての画素からなる集合を $V = \{1, 2, \dots, |V|\}$ とあらわすことにする。 i 番目の画素の輝度値を x_i とするとひとつの画像は $x = (x_1, x_2, \dots, x_{|V|})^T$ という縦ベクトルであらわされることになる。このとき x_i は画素 i における光の強度の状態変数 (State Variable)、縦ベクトル x は画像に対する状態ベクトル (State Vector) とよぶことができる。各画素 i の輝度値に対する確率変数 (Random Variable) を X_i と定義すると、この画像全体には $X = (X_1, X_2, \dots, X_{|V|})^T$ という確率変数のベクトル、すなわち確率ベクトル (Random Vector) がわりあてられることになる。つまり画像 x の確率は $\Pr\{X = x\}$ によりあらわされる。

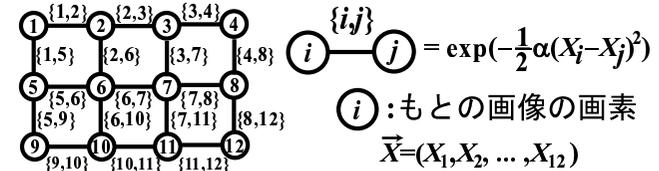
そこでユーザーがノイズ除去において最終的に得たい画像の性質を確率 $\Pr\{X = x\}$ をつかってどのように表現するかについて考えてみよう。たとえば一枚の紙の上に黒い点のひとつのついていたらこれはゴミとみなして手で払ってのけようとするであろう。ノイズ除去はこの操作をアルゴリズム化することが基本となる。このことは「局所的にあるひとつの画素に着目し、その近傍画素と階調値ができるだけ似た値になる画像がノイズが加えられる以前の画像である」と仮定した操作とみなすことができる。ここまでは空間フィルターの考え方と同じである。しかし、確率をつかう場合は上記の操作を選択する場合と選択しない場合の2種類が自動的・適応的に可能であるようにデザインすることができる。確率 $\Pr\{X = x\}$ を用いて、この操作を実現するひとつの選択肢は次の表式である。

$$\Pr\{X = x\} \propto \prod_{\{i,j\} \in E} \exp\left(-\frac{1}{2}\alpha(x_i - x_j)^2\right) \quad (1)$$

E はすべての最近接画素対 $\{i, j\}$ の集合である。画素の集合が $V = 1, 2, \dots, 12$ が縦3行、横4列の正方格子の上に並べられている場合を考えよう。最近接画素対の集合は

$$E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{5, 6\}, \{6, 7\}, \{7, 8\}, \{9, 10\}, \{10, 11\}, \{11, 12\}, \\ \{1, 5\}, \{2, 6\}, \{3, 7\}, \{4, 8\}, \{5, 9\}, \{6, 10\}, \{7, 11\}, \{8, 12\}\} \quad (2)$$

により与えられるものとする。 V と E で定義されるグラフ (V, E) 上で画像 $x = (x_1, x_2, \dots, x_{12})^T$ が与えられる。図1はこのグラフ (V, E) 上で定義される確率 $\Pr\{X = x\}$ のグラフ表現である。



$$V = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

$$E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{5, 6\}, \{6, 7\}, \{7, 8\}, \{9, 10\}, \{10, 11\}, \{10, 11\}, \\ \{11, 12\}, \{1, 5\}, \{2, 6\}, \{3, 7\}, \{4, 8\}, \{5, 9\}, \{6, 10\}, \{7, 11\}, \{8, 12\}\}$$

図1 もとの画像 x に対する式 (1) の確率 $\Pr\{X = x\}$ のグラフ表現である。

この確率は何をあらわすかを2値画像 ($Q = 2$) で説明しよう。式 (1) の確率分布がどのような画像を生成するかについて考えてみる。まずもっとも簡単な $V = \{1, 2\}$, $E = \{\{1, 2\}\}$ によるグラフ (V, E) でまず考える。可能なすべての状態 $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ を並べ上げると $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$,

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ の4つである。この4つの状態の確率の大小関係は} \\ \Pr\{X = \begin{pmatrix} 1 \\ 1 \end{pmatrix}\} = \Pr\{X = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\} > \Pr\{X = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\} = \Pr\{X = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\} \quad (3)$$

と与えられる (図2参照)。つまり隣の画素の状態が決まれば自分の画素の状態が決まることになる。

$$\textcircled{1} - \textcircled{2} = \textcircled{1} - \textcircled{2} > \textcircled{1} \cdots \textcircled{2} = \textcircled{1} \cdots \textcircled{2}$$

図2 $V = \{1, 2\}$, $E = \{\{1, 2\}\}$ で与えられる場合に $x = (0, 0)^T, (0, 1)^T, (1, 0)^T, (1, 1)^T$ の式 (1) の確率 $\Pr\{X = x\}$ の値の大小関係。

次に

$$V = \{1, 2, 3, 4, 5, 6, 7, 8, 9\} \quad (4)$$

$$E = \left\{ \{1, 2\}, \{2, 3\}, \{4, 5\}, \{5, 6\}, \{7, 8\}, \{8, 9\}, \right. \\ \left. \{1, 4\}, \{2, 5\}, \{3, 6\}, \{4, 7\}, \{5, 8\}, \{6, 9\} \right\} \quad (5)$$

与えられるグラフ (V, E) において画素 5 に着目する場合を考える。画素 5 以外の画素 i の輝度値がすべて 1, すなわち $x_i = 1 (\forall i \in V \setminus \{5\})$ であるとき画素 5 の輝度値は $x_5 = 0$ と $x_5 = 1$ で比べれば確率 $\Pr\{X = x\}$ の値が大きくなるのは $x_5 = 1$ のときである (図 3 参照)。

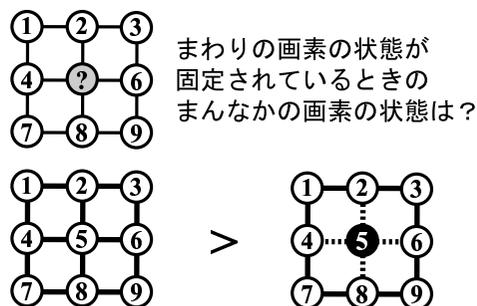
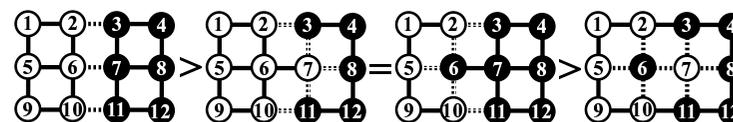
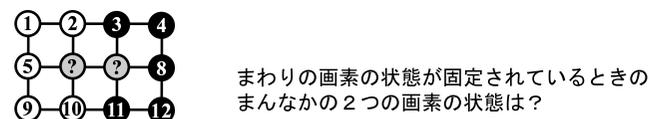


図 3 $V = \{1, 2, \dots, 9\}$, E が式 (5) で与えられる場合に画素 5 に着目し, それ以外の画素の輝度値が $x_i = 1 (\forall i \in V \setminus \{5\})$ であるとき画素 5 の輝度値は $x_5 = 0$ と $x_5 = 1$ で比べれば式 (1) の確率 $\Pr\{X = x\}$ の値が大きくなるのは $x_5 = 1$ のときである。

もう少し大きなグラフとして $V = \{1, 2, \dots, 12\}$ で E が式 (2) で与えられる場合を考えてみよう。このグラフ (V, E) 上で画素 6 と画素 7 に着目し, それ以外の画素の輝度値が $x_i = 1 (i \in \{1, 2, 5, 9, 10\})$, $x_i = 0 (i \in \{3, 4, 8, 11, 12\})$ であるとき画素 6 と 7 の輝度値 $\begin{pmatrix} x_6 \\ x_7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ を比べれば確率 $\Pr\{X = x\}$ の値が最も大きくなるのは $\begin{pmatrix} x_6 \\ x_7 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ のときである。つまり, 確率 (1) に従えば, 画素 6, 7 は周りの画素の状態をみながら自分の画素の状態が決められていることを示している。

もっと画素の数が多い場合を考えてみよう。まず, $2^{|V|}$ 面体のさいころを考えていた



点線の辺の本数が少ないほど確率が高くなる

図 4 $V = \{1, 2, \dots, 12\}$, E が式 (2) で与えられる場合に画素 6 と画素 7 に着目し, それ以外の画素の輝度値が $x_i = 1 (i \in \{1, 2, 5, 9, 10\})$, $x_i = 0 (i \in \{3, 4, 8, 11, 12\})$ であるとき画素 6 と 7 の輝度値は $\begin{pmatrix} x_6 \\ x_7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ で比べれば確率 $\Pr\{X = x\}$ の値が最も大きくなるのは $\begin{pmatrix} x_6 \\ x_7 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ のときである。

だきたい。ただし, $\Pr\{X = x\}$ の値に比例して x がでやすくなっている。もちろん, $x_1 = x_2 = \dots = x_{|V|}$ のときに一番確率が大きくなり, その面が最もでやすくなる。しかし, さいころをふれば常に $x_1 = x_2 = \dots = x_{|V|}$ を満たす画像 x がでるかといえばそうではない。むしろ, まず E に属する $\{i, j\}$ のなかで $x_i = x_j$ を満たす $\{i, j\}$ の本数が多い ($x_i \neq x_j$ を満たす最近接画素対 $\{i, j\}$ の本数が少ない) 画像 x ほどでやすくなると考えていただきたい。式 (1) の確率分布 $\Pr\{X = x\}$ をもとにマルコフ連鎖モンテカルロ (Markov Chain Monte Carlo: MCMC) 法^{4),5)} による画像生成プログラムを構成し, これにより生成された画像 $x = (x_1, x_2, \dots, x_{|V|})^T$ を図 5 に示す。

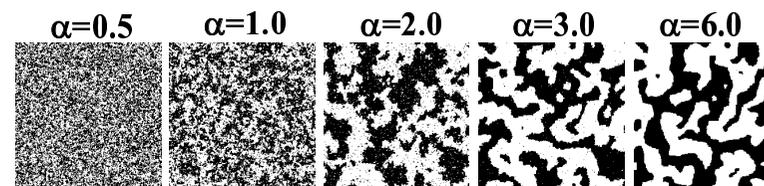


図 5 式 (1) の確率分布 $\Pr\{X = x\}$ をもとに構成されたマルコフ連鎖モンテカルロ法により生成された画像 $x = (x_1, x_2, \dots, x_{|V|})^T$ 。

図 5 で α の小さな時に本来は確率の低いはずの画像が生成されている。このことを説明

するために, $x_i \neq x_j$ を満たす最近接画素対 $\{i, j\}$ の本数

$$K(x) \equiv \sum_{\{i,j\} \in E} (x_i - x_j)^2 = \sum_{\{i,j\} \in E} (1 - \delta_{x_i, x_j}) \quad (6)$$

という量を導入する. 与えられた画像 x は $K(x)$ が小さいほど確率 $\Pr\{X = x\}$ の高い画像とすることができる. ところがこの本数 $K(x)$ が全最近接画素対のなかの 10% 以下の画像 x が出現するという事象 $A = \{x | K(x)/|E| \leq 0.1\}$ の確率 $\Pr\{A\} \equiv \sum_{\{x | K(x)/|E| \leq 0.1\}} \Pr\{X = x\}$

とその余事象 $\bar{A} \equiv \{x | K(x)/|E| > 0.1\}$ の確率 $\Pr\{\bar{A}\} \equiv \sum_{\{x | K(x)/|E| > 0.1\}} \Pr\{X = x\}$ を考えると状況は変わってくる. α がある程度以上に大きいときは不等式

$$\Pr\{A\} > \Pr\{\bar{A}\} \quad (7)$$

が成り立ち, 上で述べてきたことと同じである, しかし, α がある程度より小さくなると不等号の向きが逆転し,

$$\Pr\{A\} < \Pr\{\bar{A}\} \quad (8)$$

という不等式が成り立ってしまう. 後者の場合は事象 A に属するひとつひとつの画像の確率 $\Pr\{X = x\}$ は大きい, その事象の標本点の個数 $|A|$ が少ないため, 総和でみたときに確率 $\Pr\{A\}$ が小さくなってしまふ. α が大きいほど $x_i = x_j (\{i, j\} \in E)$ を満たす画像 x がでやすくなるが, 小さければ $x_i = x_j (\{i, j\} \in E)$ を満たさない画像 x のなかのひとつがでやすくなる. 以上が図 5 で α の大きさによりさまざまな性質をもつ画像 x が生成されることの解釈である.

不等式 (7) から不等式 (8) に切り替わる α の値がどこかにあるはずである. しかもその切り替わりの点でどのようなことがおこっているのかも興味深い. 図 5 で $\alpha = 2$ のときの画像が α の小さいときの性質と α の大きいときの性質をあわせもつ画像であることに気がつくであろう. このときになにかのゆらぎが大きくなっていることが予想される. どのようなゆらぎが大きくなっているのであろうか. 最近接画素対の輝度値の相関に対するゆらぎと考えるのが自然であろう. そのゆらぎをあらわす量のひとつとして共分散

$$\text{Cov}[X_i, X_j] \equiv \sum_{x_1=0}^1 \sum_{x_2=0}^1 \cdots \sum_{x_{|V|=0}}^1 (x_i - \mu_i)(x_j - \mu_j) \Pr\{X = x\} \quad (\forall \{i, j\} \in E) \quad (9)$$

があげられる. μ_i は X_i の期待値である.

$$\mu_i \equiv \sum_{x_1=0}^1 \sum_{x_2=0}^1 \cdots \sum_{x_{|V|=0}}^1 x_i \Pr\{X = x\} \quad (10)$$

$|V| \rightarrow +\infty$ における式 (9) の共分散 $\text{Cov}[X_i, X_j]$ の α 依存性は図 6 の通りであり, $\alpha = 1.7627\dots$ において最大値をいることがわかる. つまりゆらぎが最大となるということである. 図 5 で $\alpha = 2$ のときの画像はこのゆらぎが最大となる点の付近で生成される画像と見

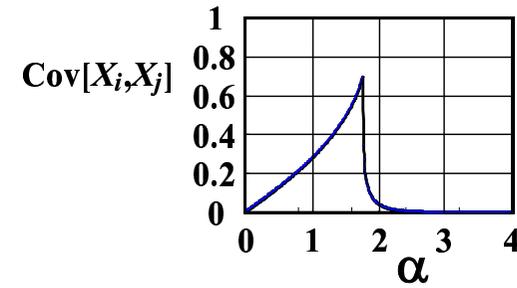


図 6 $|V| \rightarrow +\infty$ における式 (9) の共分散 $\text{Cov}[X_i, X_j]$.

なすことができる. 通常自然画像を 2 値化した画像と図 5 の $\alpha = 2$ のときの生成画像を比較すると, 全体としてみれば似ているとは言い難いが部分的なパターンを取り出してみるとそこに類似性が見いだされる (図 7 参照). この類似性を画像処理に利用するというのがベイズ推定による画像処理の発想の基本である. $Q = 256$ の場合の確率分布 (1) におけるマルコフ連鎖モンテカルロ法による生成画像 x は図 8 の通りである.

ちょっと違う視点から図 3 と図 4 の例を説明しよう. 図 3 の例で $V \setminus \{5\}$ の画素の状態が固定されている時の画素 5 の状態に対する確率は条件付き確率をつかって次のようにあらわされる.

$$\begin{aligned} \Pr\{X_5 = x_5 | X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, X_6 = 1, X_7 = 1, X_8 = 1, X_9 = 1\} \\ = \frac{\Pr\{X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = x_5, X_6 = 1, X_7 = 1, X_8 = 1, X_9 = 1\}}{\Pr\{X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, X_6 = 1, X_7 = 1, X_8 = 1, X_9 = 1\}} \end{aligned} \quad (11)$$

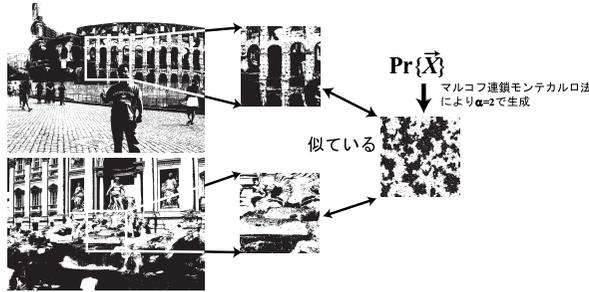


図 7 通常の自然画像を 2 値化した画像と図 5 の $\alpha = 2$ のときの生成画像の類似性.

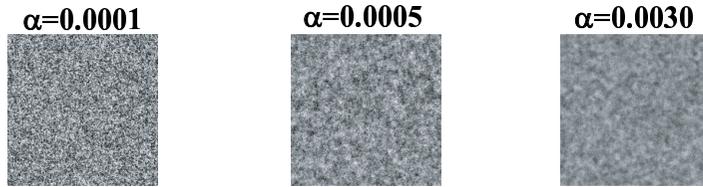


図 8 式 (1) の確率分布におけるマルコフ連鎖モンテカルロ (Markov Chain Monte Carlo) 法による生成画像 x .

$$\begin{aligned} & \Pr\{X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, X_6 = 1, X_7 = 1, X_8 = 1, X_9 = 1\} \\ &= \sum_{x_5=0}^1 \Pr\{X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = x_5, X_6 = 1, X_7 = 1, X_8 = 1, X_9 = 1\} \end{aligned} \quad (12)$$

式が長いので

$$\mathbf{X}_V = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \\ X_9 \end{pmatrix}, \quad \mathbf{X}_{V \setminus \{5\}} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_6 \\ X_7 \\ X_8 \\ X_9 \end{pmatrix}, \quad \mathbf{x}_V = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{pmatrix}, \quad \mathbf{x}_{V \setminus \{5\}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{pmatrix} \quad (13)$$

という記号を導入する. この記号をつかうと式 (11)-(12) は画素 5 以外の画素の輝度値が $x_i (\forall i \in V \setminus \{5\})$ であるときの画素 5 の輝度値に対する確率は次のように与えられる.

$$\Pr\{X_5 = x_5 | \mathbf{X}_{V \setminus \{5\}} = \mathbf{x}_{V \setminus \{5\}}\} = \frac{\Pr\{\mathbf{X}_V = \mathbf{x}_V\}}{\Pr\{\mathbf{X}_{V \setminus \{5\}} = \mathbf{x}_{V \setminus \{5\}}\}} \quad (14)$$

$$\Pr\{\mathbf{X}_{V \setminus \{5\}} = \mathbf{x}_{V \setminus \{5\}}\} = \sum_{x_5=0}^1 \Pr\{\mathbf{X}_V = \mathbf{x}_V\} \quad (15)$$

式 (14)-(15) に式 (1) を代入すると次の式が得られる.

$$\Pr\{X_5 = x_5 | \mathbf{X}_{V \setminus \{5\}} = \mathbf{x}_{V \setminus \{5\}}\} \propto \prod_{j \in \partial_5} \exp\left(-\frac{1}{2}\alpha(x_5 - x_j)^2\right) \quad (16)$$

$$\partial_5 \equiv \{2, 4, 6, 8\} \quad (17)$$

∂_5 は画素 5 のすべての最近接画素の集合をあらわしている. 式 (16)

$$\Pr\{X_5 = x_5 | \mathbf{X}_{V \setminus \{5\}} = \mathbf{x}_{V \setminus \{5\}}\} = \Pr\{X_5 = x_5 | X_k = x_k, \forall k \in \partial_5\} \quad (18)$$

が成り立つことを示している.

同様に、図 4 の例で $V \setminus \{6, 7\}$ の画素の状態が固定されている場合に

$$\mathbf{X}_V = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \\ X_9 \\ X_{10} \\ X_{11} \\ X_{12} \end{pmatrix}, \mathbf{X}_{V \setminus \{6,7\}} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_8 \\ X_9 \\ X_{10} \\ X_{11} \\ X_{12} \end{pmatrix}, \mathbf{x}_V = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11} \\ x_{12} \end{pmatrix}, \mathbf{x}_{V \setminus \{6,7\}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11} \\ x_{12} \end{pmatrix} \quad (19)$$

という記号を導入することで、画素 6 と 7 の状態に対する確率は条件付き確率をつかって次のようにあらわされる。

$$\Pr\{X_6 = x_6, X_7 = x_7 | \mathbf{X}_{V \setminus \{6,7\}} = \mathbf{x}_{V \setminus \{6,7\}}\} = \frac{\Pr\{\mathbf{X}_V = \mathbf{x}_V\}}{\Pr\{\mathbf{X}_{V \setminus \{6,7\}} = \mathbf{x}_{V \setminus \{6,7\}}\}} \quad (20)$$

$$\Pr\{\mathbf{X}_{V \setminus \{6,7\}} = \mathbf{x}_{V \setminus \{6,7\}}\} = \sum_{x_6=0}^1 \sum_{x_7=0}^1 \Pr\{\mathbf{X}_V = \mathbf{x}_V\} \quad (21)$$

式 (20)-(21) に式 (1) を代入すると次の式が得られる。

$$\begin{aligned} & \Pr\{X_6 = x_6, X_7 = x_7 | \mathbf{X}_{V \setminus \{6,7\}} = \mathbf{x}_{V \setminus \{6,7\}}\} \\ & \propto \left(\prod_{k \in \partial 6 \setminus \{7\}} \exp\left(-\frac{1}{2}\alpha(x_6 - x_k)^2\right) \right) \\ & \quad \times \exp\left(-\frac{1}{2}\alpha(x_6 - x_7)^2\right) \left(\prod_{k \in \partial 7 \setminus \{6\}} \exp\left(-\frac{1}{2}\alpha(x_7 - x_k)^2\right) \right) \end{aligned} \quad (22)$$

$$\partial\{6\} \equiv \{2, 5, 7, 10\}, \quad \partial\{7\} \equiv \{3, 6, 8, 11\} \quad (23)$$

やはり

$$\begin{aligned} & \Pr\{X_6 = x_6, X_7 = x_7 | \mathbf{X}_{V \setminus \{6,7\}} = \mathbf{x}_{V \setminus \{6,7\}}\} \\ & = \Pr\{X_6 = x_6, X_7 = x_7 | X_k = x_k, \forall k \in \partial\{6, 7\}\} \end{aligned} \quad (24)$$

が成り立つこととなる。

式 (22) と式 (23) はいずれも着目画素以外の画素の状態を固定したとき、着目画素の状態はその最近接画素の状態のみによって決定されることをあらわしている。この性質をもつ確率場 \mathbf{X}_V をマルコフ確率場 (Markov Random Field: MRF) とよばれ、次の等式としてまとめられる。

$$\Pr\{X_i = x_i | \mathbf{X}_{V \setminus \{i\}} = \mathbf{x}_{V \setminus \{i\}}\} = \Pr\{X_i = x_i | X_k = x_k, \forall k \in \partial\{i\}\} \quad (\forall i \in V) \quad (25)$$

∂_i は画素 i のすべての最近接画素の集合をあらわしている。また、式 (22) と式 (23) を一般の正方形格子からなるグラフ (V, E) に対して書き下すと次のような表式としてまとめられる。

$$\Pr\{X_i = x_i | \mathbf{X}_{V \setminus \{i\}} = \mathbf{x}_{V \setminus \{i\}}\} \propto \prod_{k \in \partial i} \exp\left(-\frac{1}{2}\alpha(x_i - x_k)^2\right) \quad (26)$$

$$\begin{aligned} & \Pr\{X_i = x_i, X_j = x_j | \mathbf{X}_{V \setminus \{i,j\}} = \mathbf{x}_{V \setminus \{i,j\}}\} \\ & \propto \left(\prod_{k \in \partial i \setminus \{j\}} \exp\left(-\frac{1}{2}\alpha(x_i - x_k)^2\right) \right) \\ & \quad \times \exp\left(-\frac{1}{2}\alpha(x_i - x_j)^2\right) \left(\prod_{k \in \partial j \setminus \{i\}} \exp\left(-\frac{1}{2}\alpha(x_j - x_k)^2\right) \right) \end{aligned} \quad (27)$$

3. ノイズ生成モデルとベイズ推定

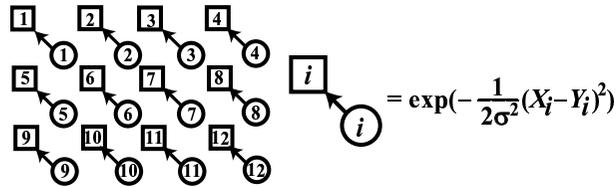
マルコフ確率場の性質をもつ確率分布 (1) は画像のさまざまな性質をあらわすことが可能であるが、それだけでは個々のデータに含まれる情報を取りこむことができない。このデータを取りこむメカニズムが必要である。ノイズ除去の場合にはデータはもともとの画像 x にノイズが加えられることで生成されたと考える。最も基本的にはノイズは加法的白色ガウスノイズ (Additive White Gaussian Noise) である。各画素ごとに独立に発生した $n_i (i \in V)$ という値による構成されるノイズ $n = (n_1, n_2, \dots, n_{|V|})$ が画像に加算されて生成されたのがデータ $y = (y_1, y_2, \dots, y_{|V|})$ であると仮定する。

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \quad (28)$$

データ $\mathbf{y} = (y_1, y_2, \dots, y_{|V|})$ の確率ベクトルを $\mathbf{Y} = (Y_1, Y_2, \dots, Y_{|V|})$ とあらわすことにすると加法的白色ノイズの確率は次の式で与えられる (図 9 参照)。

$$\Pr\{\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}\} \propto \prod_{i \in V} \exp\left(-\frac{1}{2\sigma^2}(y_i - x_i)^2\right) \quad (29)$$

$\Pr\{\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}\}$ はもとの画像 x が与えられたときのデータ y が生成される条件付き確率であるが今ほしいのは逆で、データ y が与えられたときのもとの画像 x に対する条件付き確率 $\Pr\{\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}\}$ である。これは式 (1) と式 (29) からベイズの公式 (Bayes



○ : もとの画像の画素 $\vec{X}=(X_1, X_2, \dots, X_{12})$
□ : データの画素 $\vec{Y}=(Y_1, Y_2, \dots, Y_{12})$

図 9 式 (29) の条件付き確率分布のグラフ表現.

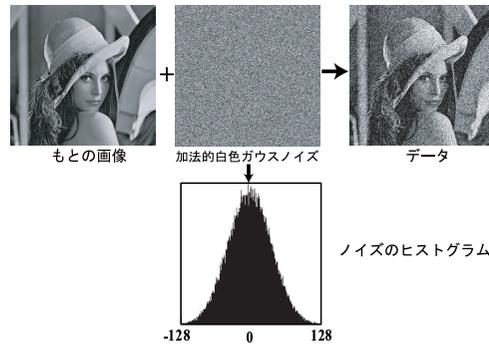


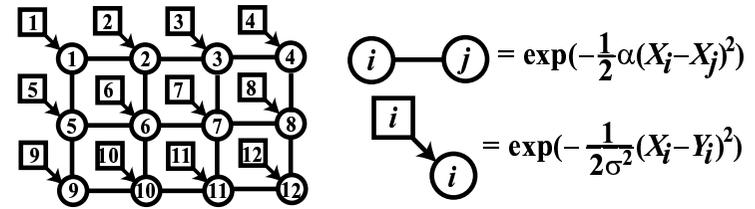
図 10 式 (29) の条件付き確率分布によるデータ y の生成過程.

Formula) をつかって次のように導かれる.

$$\Pr\{X = x|Y = y\} = \frac{\Pr\{Y = y|X = x\}\Pr\{X = x\}}{\Pr\{Y = y\}} \propto \left(\prod_{i \in V} \exp\left(-\frac{1}{2\sigma^2}(y_i - x_i)^2\right) \right) \left(\prod_{\{i,j\} \in E} \exp\left(-\frac{1}{2}\alpha(x_i - x_j)^2\right) \right) \quad (30)$$

ベイズの公式により導出された式 (30) の条件付き確率 $\Pr\{X = x|Y = y\}$ は事後確率 (Posterior Probability) とよばれる. また, 確率 $\Pr\{X = x\}$ は事前確率 (Prior Probability) とよばれる.

式 (1) の事前確率 $\Pr\{X = x\}$ と式 (30) の事後確率 $\Pr\{X = x|Y = y\}$ はそれぞれ図 1 と図 9 の正方形子によるグラフ表現であたえられる. 具体的な計算は結合確率



○ : もとの画像の画素 $\vec{X}=(X_1, X_2, \dots, X_{12})^T$
□ : データの画素 $\vec{Y}=(Y_1, Y_2, \dots, Y_{12})^T$

図 11 式 (30) の事後確率のグラフ表現.

$\Pr\{X = x|Y = y\}$ について, 各確率変数ごとのモーメントや共分散などの統計量を計算する問題に帰着される. 例えば確率変数 X_i の期待値 (1 次のモーメント) を求めようとする場合, 定義に従えば

$$E[X_i] = \sum_{x_1=0}^{Q-1} \sum_{x_2=0}^{Q-1} \dots \sum_{x_{|V|=0}^{Q-1} x_i \Pr\{X = x|Y = y\} \quad (31)$$

という操作をおこなうことになる. これは一般には $\mathcal{O}(\exp(|V|))$ の計算量を必要としてしまう. この統計量を (数値計算の誤差を除いたとしても) 厳密なものではないが問題によっては厳密計算の値に近い値が得られるように設計することができる近似計算技法として確率伝搬法 (Belief Propagation) が最近注目をあつめている.

確率伝搬法は各最近接格子点対 $\{i, j\}$ に対して $M_{i \rightarrow j}(x_j)$ と $M_{j \rightarrow i}(x_i)$ の 2 種類のメッセージ (Message) とよばれる量が導入される. 式 (30) で与えられた事後確率に対してはこのメッセージは次のメッセージ伝搬規則 (Message Propagation Rule) によるくりかえし計算を収束するまで実行することで計算される (図 12 参照).

$$M_{j \rightarrow i}(x_i) \leftarrow \text{Constant} \times \sum_{z=0}^{Q-1} \exp\left(-\frac{1}{2}\alpha(x_i - z)^2 - \frac{1}{2\sigma^2}(z - y_j)^2\right) \times \prod_{k \in \partial_j \setminus \{i\}} M_{k \rightarrow j}(z) \quad (\forall \{i, j\} \in E) \quad (32)$$

$$M_{i \rightarrow j}(x_j) \leftarrow \text{Constant} \times \sum_{z=0}^{Q-1} \exp\left(-\frac{1}{2}\alpha(x_j - z)^2 - \frac{1}{2\sigma^2}(z - y_i)^2\right) \times \prod_{k \in \partial_i \setminus \{j\}} M_{k \rightarrow j}(z) \quad (\forall \{i, j\} \in E) \quad (33)$$

∂_j は画素 j の最近接画素の集合、 $\partial_j \setminus \{i\}$ は画素 j の最近接画素の集合から画素 i をのぞいた集合を意味する。メッセージ伝搬規則 (32)-(33) にあられる $\prod_{k \in \partial_j \setminus \{i\}}$ は、高々、各画素の最近接画素数なので計算量は $O(1)$ となる。つまり、くりかえし計算によるアルゴリズムの 1 回の更新ごとの計算量は $O(|V|)$ というわけである。モーメントや共分散はメッセージ $\{M_{i \rightarrow j}(x), M_{j \rightarrow i}(x) | \forall \{i, j\} \in E\}$ をもちいてあらわすことができるので、得られたメッセージを代入することで求められる。図 13 に示した出力結果はこの確率伝搬法により得られた

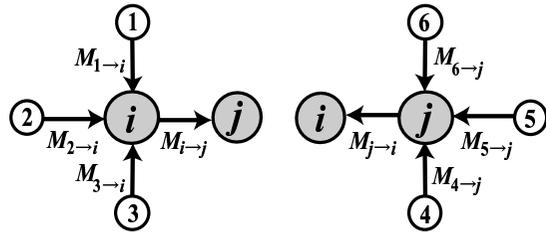


図 12 確率伝搬法におけるメッセージ伝搬規則のグラフ表現.

ものである。

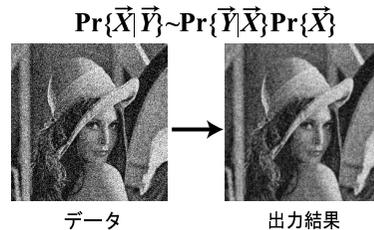


図 13 確率伝搬法によるノイズ除去例.

確率伝搬法を用いた画像処理の基本的なアルゴリズム更正法の詳細は文献²⁾を参照して

いただきたい。確率伝搬法は人工知能の確率推論におけるアルゴリズムとして 1980 年代に考案され、その後、誤り訂正符号における低密度パリティ検査 (Low Density Parity Check: LDPC) 符号、移動体通信の符号分割多重接続 (Code Division Multiple Access: CDMA) 方式、組み合わせ最適化における充足可能性 (Satisfiability: SAT) 問題に対するキーアプリケーションとしても新しい計算の概念を生み出しつつある。その最近の展開については文献^{3),6)-8)}を参照していただきたい。さらに確率伝搬法は物理学における平均場理論との類似性が指摘され、情報統計力学の研究テーマとしても研究が進められている^{3),6),9)}。

4. 情報統計力学と統計的性能評価

前節で説明したベイズ推定による画像処理システムの性能の評価をするひとつの方法としてたくさんのサンプルを生成して標本平均 (Sample Average) を計算するという方法がまず考えられる。これは、まず事前確率 (1) に従ってもとの画像 x を 100 個生成し、それぞれに対して各画像 x ごとに条件付き確率 (29) に従って 100 個ずつのデータ y を生成する。そして、得られた 10000 個のサンプルデータ y から $\Pr\{\bar{X}|\bar{Y} = \bar{y}\}$ に従ってもとの画像 x の推定値 $\hat{x} = \hat{x}(y)$ をそれぞれに求めた上で x と \hat{x} との平均二乗誤差 $d(x, \hat{x}) = \frac{1}{|V|} \sum_{i \in V} (x_i - \hat{x}_i)^2$ についての標本平均を計算すると、この事後確率に対する画素あたりの誤差を実験的に評価できる。しかし、これではいずれの場合も 10000 回のテストを行わなければならないことになる。また、より信頼性の高い結果を得るためには 1000000 回、100000000 回とテスト回数を増やす必要があるかもしれない。ところが、このプロセスを式で表すと

$$\sum_{x_1=0}^{Q-1} \sum_{x_2=0}^{Q-1} \cdots \sum_{x_{|V|=0}^{Q-1} \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} d(x, \hat{x}(y)) \times \Pr\{Y = y | X = x\} dy_1 dy_2 \cdots dy_{|V|} \right) \Pr\{X = x\} \quad (34)$$

という量を計算することに相当し、これを計算できてしまえばテストを 10000 回行わなくてもよいことになる。実は式 (34) であらわされる量は 20 世紀の後半に多くの物理学者がスピングラス理論 (Spin Glass Theory) とよばれる分野のなかで挑戦してきた量、すなわち配位平均 (Configuration Average) の初等的な場合に相当する。このことは、20 世紀後半に行われたスピングラス理論に関する多くの理論的成果が確率の情報処理の問題の理解に役立てられる可能性を示しており、実際に画像処理はもとより、誤り訂正符号、移動体通信における CDMA 復調方式、組み合わせ最適化などの様々な問題で情報統計力学によるアプローチが新しい研究成果をあげている^{3),10)-14)}。

5. ま と め

本稿ではノイズ除去を例にとって確率的画像処理について解説した。事前確率 (1) を領域分割, エッジ検出などの用途にあわせて適切に変更することにより, さまざまな確率的画像処理フィルターを設計することができる。そのさまざまな用途に対する最近の展開については文献¹⁵⁾ で概観することができる。確率的画像処理の情報統計力学的展開についての詳細は文献^{1),11)} を参照していただきたい。

謝辞

本研究の一部は文部科学省科学研究費補助金 (No.18079002, No.18079008) および東北大学電気情報系グローバル COE プログラム「情報エレクトロニクスシステム教育研究拠点」の補助のもとでおこなわれたものである。

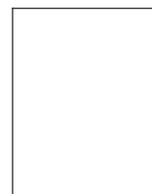
参 考 文 献

- 1) K. Tanaka: Statistical-Mechanical Approach to Image Processing (Topical Review), Journal of Physics A: Mathematical and General, Vol.35, No.37, pp.R81-R150, 2002.
- 2) 田中和之: 確率モデルによる画像処理技術入門, 森北出版, 2006.
- 3) 田中和之 (編著): 臨時別冊・数理科学 SGC ライブラリ「確率的情報処理と統計力学—様々なアプローチとそのチュートリアル」, サイエンス社, 2006.
- 4) 伊庭幸人: ベイズ統計と統計物理, 岩波書店 (2003).
- 5) 伊庭幸人, 種村正美: 統計科学のフロンティア 計算統計 II —マルコフ連鎖モンテカルロ法とその周辺, 岩波書店, 2005.
- 6) 汪金芳, 田栗正章, 手塚集, 樺島祥介, 上田修功: 統計科学のフロンティア 計算統計 I —確率計算の新しい手法, 岩波書店, 2003.

- 7) C. M. Bishop : Pattern Recognition and Machine Learning, Springer, 2006.
- 8) M. J. Wainwright and M. I. Jordan: Graphical Models, Exponential Families, and Variational Inference, now Publishing Inc, 2008.
- 9) M. Opper and D. Saad (eds): Advanced Mean Field Methods —Theory and Practice, MIT Press, 2001.
- 10) 西森秀稔: スピングラス理論と情報統計力学, 岩波書店, 1999.
- 11) H. Nishimori: Statistical Physics of Spin Glasses and Information Processing, —An Introduction, Oxford University Press, 2001.
- 12) 樺島祥介: 物理の世界/学習と情報の平均場理論, 岩波書店, 2002.
- 13) Y. Kabashima and D. Saad: Statistial Mechanics of Low-Density Parity-Check Codes (Topical Review), Journal of Physics A: Mathematical and General, Vol. 37, No.6, pp.R1-R43, 2004.
- 14) M. Mézard and A. Montanari: Information, Physics, and Computation, Oxford University Press, 2009.
- 15) A. S. Willsky: Multiresolution Markov models for signal and image processing, Proceedings of IEEE, vol.90, no.8, pp.1396-1458, 2002.

(平成 21 年 5 月 9 日受付)

(平成 年 月 日採録)



田中 和之

昭和 36 年生。平成元年東北大学大学院工学研究科電子工学専攻博士課程修了。同年東北大学工学部助手。平成 6 年室蘭工業大学工学部助教授。平成 11 年東北大学大学院情報科学研究科助教授。平成 19 年より東北大学大学院情報科学研究科教授。確率的情報処理と情報統計力学に関する研究に従事。工学博士。日本物理学会, 電子情報通信学会, 人工知能学会, 情報理論とその応用学会, IEEE-CIS 各会員。