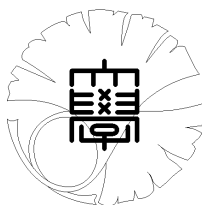


数理科学実践研究レター 2019–8 November 28, 2019

画像の“写真らしさ”に関する
ある確率論的相転移現象について

by

浅尾 泰彦、坂本 龍太郎



UNIVERSITY OF TOKYO
GRADUATE SCHOOL OF MATHEMATICAL SCIENCES
KOMABA, TOKYO, JAPAN

画像の“写真らしさ”に関するある確率論的相転移現象について

浅尾泰彦¹ (東京大学大学院数理科学研究科)

Yasuhiko Asao (Graduate School of Mathematical Sciences, The University of Tokyo)

坂本龍太郎² (東京大学大学院数理科学研究科)

Ryotaro Sakamoto (Graduate School of Mathematical Sciences, The University of Tokyo)

概要

画像認識において、画像全体のうち意味のあるものとそうでないものの境界を見極めることは重要な問題である。本稿では準写真という「写真らしい画像」を定義することで、この問題への数学的アプローチを試みた。「写真らしさとは何か」という問いを数学的に定式化するために、depth という画像に対して実数値をとる関数を導入し、その漸近挙動を調べた。また、例において実際の写真が準写真であることを確かめた。

depth の概念は方体複体の 0 次パーシステントホモロジーの階数から着想を得ており、将来的に高次の階数を解析することでより精密な画像の分類が得られると期待できる。また、画像認識において、近年積極的に活用が進められている深層学習の学習データ選定への応用も期待できると考えている。本稿で画像認識における純粋数学の活用の 1 つのアプローチを提案したい。本稿は、[AS] の内容を証明や手法を含めてより詳しく論じたものである。

1 はじめに

本稿は、[AS] の内容を証明や手法を含めてより詳しく論じたものである。第 7 章と第 8 章で画像の深さに関する確率論的な相転移現象の証明、実際に用いた画像解析の手法について述べた。なお、本序文の以下の部分と第 2 章から第 6 章は [AS] に発表された内容と同じである。

本稿で扱う問題は「コンピュータが“内在的”にどこまで意味のある画像とそうでないものを区別できるか？」というものである。

以下では、図 1 のような意味のある画像のことを「写真」と呼ぶことにする。つまり絵や数字など、我々が目にしてすぐに意味が理解できる、もしくは意味があると判断できるもの全てを「写真」と総称する。一方で「意味のない」画像というのは一見して意味が理解できない、意味があると判断できないものを指す。例えば、図 2 のようなモザイク画像のことを指す。



図 1: りんごの画像。

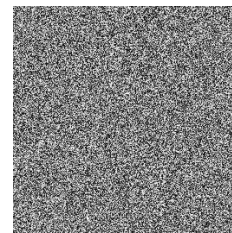


図 2: モザイク画像, depth が非常に大きい。

機械学習では、コンピュータにある性質 A を持つ大量の類似データを学習させることで、新たに読み込ませたデータが性質 A を持つかどうかを判断させることができた。例えば、りんごの写真を学習させた後に図 2 のモザイク画像を読み込ませると、コンピュータはそれがりんごでないと判断できる。我々がここで「内在的」と呼ぶのは、そのような学習の過程を経ないで、ということである。つまり、りんごが何であるかを知らない状況で、コンピュータはりんごの写真とモザイクの画像をきちんと区別できるか？また可能であれば「りんごとモザイク」という極端な分類の他にどの程度分類が可能であるか？

¹asao@ms.u-tokyo.ac.jp

²sakamoto@ms.u-tokyo.ac.jp

本稿では、画像の持つ数理的な性質によって特徴付けられる「準写真」という画像のクラスを導入することで、この問題に取り組んだ。準写真であるという性質は個々の画像に対して数学的に判別できるため、上で述べた意味で内在的である。従って、コンピュータは学習の過程を経ずに画像を準写真とそうでないものに分類することができる。さらに、数学的に定義された準写真であるという性質は、実際の写真にも備わっていることを例で確かめた。

準写真は画像の「depth」という概念を数学的に定義することで得られ、depthは近年データサイエンスの分野で広く認知されているパーシステントホモロジーから着想を得ている。本稿においてパーシステントホモロジーなど純粋数学で成熟した道具を、画像認識に活用する1つのアプローチを提案したい。

2 画像の depth

2.1 節において画像の定義について説明し、2.2 節で画像の depth を定義する。

2.1 画像の定式化

0 以上 1 未満の実数の集合を $[0, 1)$ で表す。自然数 N に対して集合 \square_N を

$$\square_N := \left\{ \frac{0}{2^N}, \frac{1}{2^N}, \dots, \frac{2^N - 1}{2^N} \right\} \times \left\{ \frac{0}{2^N}, \frac{1}{2^N}, \dots, \frac{2^N - 1}{2^N} \right\} \subseteq [0, 1) \times [0, 1).$$

で定義する。また

$$C := \{0, 1, \dots, n - 1\}$$

を白黒の濃淡を表す色の集合とする。 \square_N と $[0, 1) \times [0, 1)$ を $2^N \times 2^N$ 等分割したマスの集合との間に自然な一対一対応が存在するため、写像 $\square_N \rightarrow C$ は $2^N \times 2^N$ ピクセルのモノクロ画像と考えられる。

次に、非負整数 $d \leq N$ に対して \square_N を $2^d \times 2^d$ 分割することを考える。 \square_N の部分集合族 \square_N^d を

$$\square_N^d := \left\{ \frac{1}{2^d} \square_{N-d} + x \subseteq \square_N \mid x \in \square_d \right\}$$

で定める。例えば $\square_N^0 = \{\square_N\}$, $\square_N^N = \{\{x\} \mid x \in \square_N\}$ である。このとき \square_N^d は \square_N の $2^d \times 2^d$ 分割

$$\square_N = \bigsqcup_{\square \in \square_N^d} \square$$

を与える。

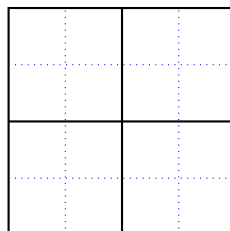


図 3: 実線は \square_N^1 , 点線は \square_N^2 .

2.2 深さの定義

非負整数 $d \leq N$ と 2 色モノクロ画像 $f: \square_N \rightarrow C = \{0, 1\}$ に対して、色の偏りを測る指標として $\varphi_d(f) \in \mathbb{Z}_{\geq 0}$ を

$$\varphi_d(f) := \min_{\square \in \square_N^d} \left(\sum_{i, j \in \square} |f(i) - f(j)| \right)$$

で定義する。即ち $\varphi_d(f)$ は $2^d \times 2^d$ 分割したそれぞれのマスのうち、最も色が偏ったマスの偏り具合を数値として表す関数である。ある $\square \in \square_N^d$ で $\#f(\square) = 1$ ならば $\varphi_d(f) = 0$ である。つまり、1 つでも全てが白または黒であるマス $\square \in \square_N^d$ が存在すれば $\varphi_d(f)$ の値は 0 となる。特に、 $\varphi_N(f) = 0$ が常に成り立つ。

2 色と限らない画像 $f: \square_N \rightarrow C = \{0, \dots, n-1\}$ の depth を「どれくらい分割を細かくした時に (d を N に近づけたときに) 初めて $\varphi_d(f) = 0$ となるか」を測ることで定義する。

定義 2.1. $2^N \times 2^N$ ピクセル n 色画像 $f: \square_N \rightarrow C$ に対して

$$\text{depth}(f) := \frac{1}{N} \min \left\{ d \in \{0, \dots, N\} \mid 0 = \min_{p: \text{Im}(f) \rightarrow \{0,1\}} \varphi_d(p \circ f) \right\}$$

を f の **depth(深さ)** という。ここで p は f の像 $\text{Im}(f)$ から集合 $\{0, 1\}$ への全射全体を動く。

例えば、 $\text{depth}(f) = 0$ と f が定値写像 (つまり単色画像) は同値である。逆に図 2 のような複雑な画像の depth は 1 となる。即ち、depth は画像の複雑さを測る指標である。意味のある画像はある程度色に偏りがあると考えられる。そのため depth はある程度低いことが期待される。図 4 は d を次第に大きくした時の $\left(\sum_{i, j \in \square} |f(i) - f(j)| \right)_{\square \in \square_N^d}$ を色の濃さとして表したものである。 d が増加するとマスは細かくなり、初めて真っ黒なマスが現れた時の d が depth に対応する。



図 4: $N = 10$ のキリンの画像。左上から右下にかけて $0 \leq d \leq 10$ が大きくなっていく。 $d = 5$ で初めて黒いマスが現れるので $\text{depth} = 0.5$ である。

3 深さの漸近挙動と準写真の定義

この章では前章で定義した depth を用いて、どれくらいの割合の画像が「とても複雑」(つまり写真らしくない)かを計算する。それに基づいて準写真という数学的に「写真らしい」画像のクラスを定義する。

\square_N から C への写像全体 ($2^N \times 2^N$ ピクセルのモノクロ画像全体) を $\text{Map}(\square_N, C)$ で表す。実数 $\alpha \in [0, 1]$ に対して

$$\mathcal{P}_{N,C}(\alpha) := \{f \in \text{Map}(\square_N, C) \mid \text{depth}(f) \leq \alpha\}$$

で depth が α 以下の画像全体の集合を表す. 有限集合 X の要素の個数を $\#X$ で表すことにすれば, $\#\text{Map}(\square_N, C) = n^{4N}$ より depth が α 以下の画像全体の割合は $\#\mathcal{P}_{N,C}(\alpha)/n^{4N}$ である. このとき次の計算結果を得る.

補題 3.1 (系 7.3). 実数 $\alpha \in [0, 1]$ に対して

$$\lim_{N \rightarrow +\infty} \frac{\#\mathcal{P}_{N,C}(\alpha)}{n^{4N}} = \begin{cases} 0 & \text{if } 0 \leq \alpha < 1, \\ 1 & \text{if } \alpha = 1. \end{cases}$$

即ち「ほとんど全ての画像がとても複雑」であることがわかる. この結果は, ランダムに生成した画像が写真であることは稀有である, という我々の認識に関する直感と矛盾しない. 上の結果を $\#\mathcal{P}_{N,C}(\alpha)$ の主要項の漸近挙動だと考え, 第 2 主要項を計算すると以下の結果を得る.

補題 3.2 (系 7.4). 実数 α に対して

$$\lim_{N \rightarrow +\infty} \frac{\#\mathcal{P}_{N,C}(1 - \alpha \frac{\log N}{N})}{n^{4N}} = \begin{cases} 0 & \text{if } 1/\log 4 < \alpha, \\ 1 & \text{if } 0 \leq \alpha \leq 1/\log 4. \end{cases}$$

従って, 画像全体のうち depth が $1 - \log N/(N \log 4)$ 未満であるような「複雑すぎないもの」は非常に少ないことがわかる. 前章でも述べたように意味のある画像は depth が低いと期待されるため, これらを準写真と定義する.

定義 3.3. $2^N \times 2^N$ ピクセルのモノクロ画像 $f: \square_N \rightarrow C$ が

$$\text{depth}(f) < 1 - \frac{\log N}{N \log 4}$$

を満たすとき, f は**準写真**であるという.

例えば, $N = 10$ のとき $1 - \log 10/(10 \log 4) = 0.83390359525 \dots$ より図 4 のキリンの画像 (depth = 0.5) は準写真である.

4 人間の認識に関する予想

図 4 では左上の真っ白な画像から始まって次第に細度が上がっていきキリンの姿が浮かび上がり, 最後には真っ黒な画像になる. 真っ黒になる 1 つ手前の画像はもうほぼ写真と変わらないが, もう 1 つ手前までいくと写真と認識できるものやや画質が悪いという印象を持つ. これら 11 枚の画像に真っ白なものから順番に $0/10, 1/10, \dots, 10/10$ と数を振ると, $8/10$ と $9/10$ の間がちょうど画質の良し悪しを判断する境目ということになる. 我々の予想はこの境目の値がおおよそ $1 - \log N/(N \log 4)$ に対応するのではないかというものである. 実際, 前章でみたように $1 - \log 10/(10 \log 4) \sim 8/10$ である.

予想 1. 各 $x \in \square_{N-1}$ に対して x を含む \square_N^{N-1} の唯一の元を \square_x と表す. 画像 $f_N = f: \square_N \rightarrow C$ に対して画像 $f_{N-1}: \square_{N-1} \rightarrow C$ を

$$f_{N-1}(x) := \left\lfloor \frac{1}{4} \sum_{y \in \square_x} f(y) \right\rfloor$$

で定義する. 帰納的に $f_{N-i} := (f_{N-i+1})_{N-i}$ と定義する. 画像の列 $f_N, f_{N-1}, f_{N-2}, \dots, f_0$ は次第に画質が荒くなっていくが, 画質の良し悪しの変化を認識する境目は

$$\frac{k}{N} \leq 1 - \frac{\log N}{N \log 4} \leq \frac{k+1}{N}$$

を満たすような f_k と f_{k+1} の間である.

5 パーシステントホモロジーによる高次化

パーシステントホモロジーの一般論については [H] が詳しい. depth と 0 次パーシステントホモロジーとの関連を見るために, 画像 $f: \square_N \rightarrow C$ に対して次のようなフィルター付き方体複体 $C_d(f)$ を考える. 頂点集合は \square_N であり, 2 つの相異なる頂点 a, b はそれらがある $\square \in \square_N^d$ に隣り合って含まれていてかつ $f(a) = f(b)$ であるときに 1 方体で結ばれているとする. 2 方体についても同様に定義する. このとき複体 $C_d(f)$ の 0 次ホモロジーの階数は d について広義単調増加関数である. depth の概念はこの関数から着想を得ている. 本研究では 0 次パーシステントホモロジーしか考えていないが, 同様に高次の階数から画像の内在的な情報を取り出せると期待できる.

6 画像認識への応用の展望

機械学習や深層学習において, 例えばコンピュータに犬と猫の写真を分類させようとするとき, 必要な学習データはそれぞれの写真 10000 枚程度とされている. 10000 枚のデータを人の手で収集し, それをコンピュータに読み込ませることはかなりのコストを費やすため近年では収集・読み込みの自動化が試みられている. 一方で, depth を用いた方法は画像の内在的な情報を数理的に引き出すことで画像の分類をしているため学習データを用意する必要がない. 未だ精度が荒く実用化への障害はあるものの, パーシステントホモロジーを始めとする様々な数学を用いた画像認識への新たなアプローチとして期待できると考えている.

7 補題の証明

本章では画像の深さに関する確率論的相転移現象である補題 3.1 と補題 3.2 の証明をする.

非負整数 $2 \leq n$ と $d < N$ を固定する. $C := \{0, \dots, n-1\}$ とする. depth が d/N 以上の画像全体の集合を

$$\mathcal{Q}_N(d) := \{f: \square_N \rightarrow C \mid \text{depth}(f) \geq d/N\}$$

で表す. $q_N(d) := \#\mathcal{Q}_N(d)$ とおく. このとき

$$1 - \frac{q_N(d+1)}{n^{4N}} = \frac{\#\mathcal{P}_{N,C}(d/N)}{n^{4N}} \quad (1)$$

に注意する.

補題 7.1.

$$q_N(d+1) = \left(\sum_{i=1}^n (-1)^{n-i} \binom{n}{i} i^{4N-d} \right)^{4^d}. \quad (2)$$

証明. 画像 $f: \square_N \rightarrow C$ の depth が $d+1$ 以上であるための必要十分条件は, 全ての全射 $p: \text{Im}(f) \rightarrow \{0, 1\}$ に対して $\varphi_d(p \circ f) \neq 0$ となることである. この条件は任意の $\square \in \square_N^d$ に対して $f(\square) = C$ と同値である. \square_N^d は \square_N の分割を与えるので

$$\begin{aligned} q_N(d+1) &= \prod_{\square \in \square_N^d} \#\{f \in \text{Map}(\square, C) \mid f \text{ は全射}\} \\ &= \left(\sum_{i=1}^n (-1)^{n-i} \binom{n}{i} i^{4N-d} \right)^{4^d} \end{aligned}$$

を得る. □

等式 (2) を用いて $q_N(d+1)$ を閉区間 $[0, N-1]$ 上の連続写像に拡張する. 次に, d と N を動かしたときの連続関数 $q_N(d+1)/n^{4^N}$ の漸近挙動をより詳しく見る.

$$f_{N,d} := \sum_{i=1}^{n-1} (-1)^{n-i} \binom{n}{i} \left(\frac{i}{n}\right)^{4^{N-d}}$$

とおく. 等式 (2) より

$$\frac{q_N(d+1)}{n^{4^N}} = (1 + f_{N,d})^{4^d} \quad (3)$$

である. 定義より $0 < q_N(d+1)/n^{4^N} < 1$ なので不等式 $-1 < f_{N,d} < 0$ を得る. また

$$f_{N,d} \times \left(-n^{-1} \left(1 - \frac{1}{n}\right)^{-4^{N-d}}\right) = 1 + \sum_{i=1}^{n-2} \frac{(-1)^{n-i}}{n} \binom{n}{i} \left(\frac{i}{n-1}\right)^{4^{N-d}} \rightarrow 1 \quad \text{as } N \rightarrow +\infty \quad (4)$$

に注意する.

補題 7.2. $\beta: \mathbb{R}_{>0} \rightarrow \mathbb{R}$ を写像とし

$$\lim_{x \rightarrow +\infty} \beta(x) = \lim_{x \rightarrow +\infty} (x - \beta(x)) = +\infty$$

を仮定する. このとき

$$\log \left(\frac{q_N(N - \beta(N) + 1)}{n^{4^N}} \right) \sim -n4^{N-\beta(N)} \left(1 - \frac{1}{n}\right)^{4^{\beta(N)}} \quad \text{as } N \rightarrow +\infty$$

である.

証明. 等式 (3) より

$$\begin{aligned} g(N) &:= \log \left(\frac{q_N(N - \beta(N) + 1)}{n^{4^N}} \right) \\ &= 4^{N-\beta(N)} \log(1 + f_{N,N-\beta(N)}) \end{aligned}$$

である. 実数 $-1 < x < 0$ に対して $x \leq \log(1+x) \leq x/(1+x)$ が成り立つことに注意すると, $-1 < f_{N,N-\beta(N)} < 0$ より

$$4^{N-\beta(N)} f_{N,N-\beta(N)} \leq g(N) \leq \frac{4^{N-\beta(N)} f_{N,N-\beta(N)}}{1 + f_{N,N-\beta(N)}}$$

が従う. 特に (4) より

$$g(N) \sim -n4^{N-\beta(N)} \left(1 - \frac{1}{n}\right)^{4^{\beta(N)}} \quad \text{as } N \rightarrow +\infty$$

である. □

系 7.3 (補題 3.1). 実数 $\alpha \in [0, 1]$ に対して

$$\lim_{N \rightarrow +\infty} \frac{\#\mathcal{P}_{N,C}(\alpha)}{n^{4^N}} = \begin{cases} 0 & \text{if } \alpha < 1, \\ 1 & \text{if } \alpha = 1. \end{cases}$$

証明. $\alpha = 1$ のとき主張は明らか, よって $0 \leq \alpha < 1$ としてよい. 等式 (1) より

$$\lim_{N \rightarrow +\infty} \frac{q_N(\alpha N + 1)}{n^{4^N}} = 1$$

を示せば十分である. これは $\beta(x) := (1 - \alpha)x$ として補題 7.2 を適用すれば従う. □

系 7.4 (補題 3.2). 実数 $\alpha \geq 0$ に対して

$$\lim_{N \rightarrow +\infty} \frac{\#\mathcal{P}_{N,C}(1 - \alpha \frac{\log N}{N})}{n^{4N}} = \begin{cases} 0 & \text{if } 1/\log 4 < \alpha, \\ 1 & \text{if } 0 \leq \alpha \leq 1/\log 4 \end{cases}$$

が成立する.

証明. $\beta(x) := \alpha \log x$ とおく. このとき $4^{\beta(N)} = N^{\alpha \log 4}$ である. $4(1 - 1/n) > 1$ に注意すれば補題 7.2 より

$$\lim_{N \rightarrow +\infty} \frac{q_N(N - \alpha \log N + 1)}{n^{4N}} = \begin{cases} 1 & \text{if } 1/\log 4 < \alpha, \\ 0 & \text{if } 0 \leq \alpha \leq 1/\log 4 \end{cases}$$

となる. よって, この補題は (1) より従う. □

8 図 4 のプログラムソースコード

本章では 'kirin.jpg' という画像データから図 4 を描くプログラムのソースコードを紹介する.

```

1 import numpy as np
2 import cv2
3
4 imori = cv2.imread('kirin.jpg', cv2.IMREAD_GRAYSCALE)
5 t = []
6 for N in range(11):
7     l = 2**N
8     m = 2**(10-N)
9     a = []
10    am = imori.astype('int64')
11    for p in range(m):
12        c = []
13        for q in range(m):
14            b = 0
15            for i in range(l-1):
16                for j in range(l-1):
17                    b = b + abs(am[p*l+i,q*l+j] - am[p*l+i,q*l+j+1]) + abs(am[p*
18                        l+i,q*l+j] - am[p*l+i+1,q*l+j])
19            c.append(b)
20        a.append(c)
21    for p in range(m):
22        for q in range(m):
23            for i in range(1):
24                for j in range(1):
25                    am[l*p+i,l*q+j] = 50*a[p][q]/(2*(1**2-2*1+2))
26    t.append(am)
27 im_h1 = cv2.hconcat([im,t[10],t[9],t[8],t[7],t[6]])
28 im_h2 = cv2.hconcat([t[5],t[4],t[3],t[2],t[1],t[0]])
29 im_v = cv2.vconcat([im_h1,im_h2])
30 cv2.imwrite('result.jpg', im_v)

```

9 謝辞

本研究において, 画像認識についての解説などで尽力して下さった株式会社ニコン研究開発本部数理技術研究所の皆様, 特に深層学習との関連や論文に対する貴重なコメントを下された高山侑也さ

ん, 中村ちからさんに心から感謝申し上げます. またセミナーの時間調整や全般に関わるコメントをして頂いた東京大学数理科学研究科特任助教(当時)の土岡俊介さんにも御礼申し上げます. 最後に注意深く論文を読み, 有益なアドバイスを下さった査読者の方に感謝いたします.

参考文献

- [H] 平岡裕章, タンパク質構造とトポロジー: パーシステントホモロジー群入門, 共立出版 (2013).
- [AS] Y. Asao, R. Sakamoto, 画像の“写真らしさ”に関する数学的アプローチについて, 人工知能学会全国大会論文集 JSAI2019 (2019), 3K3J205–3K3J205.