

統合情報理論の可能性

荻田裕也*

概要

本稿では、2016-1st meeting で私が発表した“我々はどこから来たのか —意識の定式化—”の内容を数理的に詳述するとともに、意識研究にとどまらない統合情報理論の可能性について考察する。統合情報理論は、Giulio Tononi によって意識を定量化するために考案された理論である。この理論が本当に意識を表すかはさておき、複雑に相互作用するネットワークが生み出す機能がある意味で定量化していることは間違いない。統合情報量は、ネットワーク中の機能モジュールを識別する指標として有望である。実際の生命システムに適用することを念頭において、時系列データから統合情報量を近似的に計算する方法を述べ、簡単なシミュレーションによって指標の有用性を確認する。

キーワード: 意識, 情報理論, 統合情報理論, 生物学.

1 意識をどう定式化するか

意識がある状態と無い状態の間には、決定的な違いがあるように直観される。人の意識状態を理解することは、科学者や哲学者の積年の夢であった。一方、近年になってこの問いが実用上も重要であると明らかになった。Locked-in (閉じこめ) 症候群をはじめとする“反応はないが意識はある”状態の発見である。従来、意識の尺度は、医師の指示を理解して、声、手足、言語を通じて反応できるかによって便宜的に定量化されていた。しかし、locked-in 症候群の患者は、脳血栓等により、随意運動を行う能力を失っている(眼球運動のみが随意的に可能な場合も多い)。その結果、感覚や思考は鮮明であるのに、医師の指示に答えられず、植物状態と誤診されてしまう。意識の定量化に成功すれば、臨床医療の場において、意識が無いと勘違いされて非人道的に処置される患者の数を減らせるであろう。

アメリカの神経学者 Giulio Tononi は情報理論における情報量の概念を用いて、意識を定式化しようと試みた。彼のアイデアの中心は、意識は情報が多く統合された場合に表出する、という直観であ

* 東京大学 理学部 物理学科 (2016 年卒業), four_seasons0702 (at) yahoo.co.jp.

る¹. 彼がこのアイデアにたどりついた経緯は一般書 [1] に譲るとして, 主張をおおまかにまとめれば, 次のようになる.

情報が統合される量は統合情報量 Φ で表され, Φ が大きいシステムは意識をもつ.

注意されたいのは, 統合情報理論は提案されて日が浅く, 理論として成熟していない点である. そのため, 定式化の方法にもいくつかのバリエーションがあり, 決定版はない. ただ, 基本的な方針だけはどのバリエーションにおいても一貫している.

- 統合情報量は, 相互作用²し時間発展する変数の組 (相互作用の関係をネットワークと呼ぶ) に対し定義される.
- 統合情報量は, ネットワークを分割した部分ネットワークが生む情報の総和に比べて, もとのネットワーク全体の生む情報がどれだけ大きいかを定量化する.

の二点である. 本稿では, 離散時間, 離散状態のマルコフ遷移における統合情報理論 [4] を紹介する. なお, 本稿を読む際に情報理論の初等的な知識があると読みやすい. 次章で, 記法の導入も兼ねて, 情報理論について多少説明するが, 詳述はしない. 情報理論の入門書としては, 文献 [2] がおすすめである.

2 準備: 情報理論

2.1 情報量とは

前章では, 情報量, 統合情報量 という言葉が登場した. この情報量とはいったい何者であろうか. 情報量は確率事象を扱う文脈で定義される. まず, 定義から述べよう.

確率 p の事象 x が起きた事を知ることで得られる情報量は, $I(x) = -\log_2 p$ bit³ である.

数学的にはこれで終わりであるが, この定義の意味を述べよう. まず, 確率が低い事象が起こった時の方が, 情報量は大きい. ‘昨日の試験でのび太君が満点を取らなかった’⁴ という言説を聞いても, まあそうだろうね, ということになり特に情報量はない. ‘のび太君が満点を取った’ と聞けば, これはすごい事を聞いたぞ, となるわけである. 情報量は, 驚きの程度 *surprisal* と呼ばれることもあるが, その理由も感じていただけたらと思う.

もうひとつの重要な意味は, 情報は分割して伝えることができ, それによって情報量は不変であるという点だ. 例えば, 1 から 100 のどれかの数字がランダムで出てくるビンゴゲームのマシーンを考

¹ 意識のある状態では情報が多く統合されている, というのは直観的に納得できる. しかし, 情報が多く統合されていることこそが意識の本質である, と逆を主張するのが Tononi の仮定である.

² 相互と書いたが, 一方的なものも含む.

³ 対数の底は, 情報量の単位に対応している. 底が 2 の時の単位は bit であるが, 底が 10 の時は dit, 底が自然対数の時は nit もしくは nat が情報量の単位となる. 本稿では, 対数の底が何であるか (単位系が何であるか) にはあまりこだわらない.

⁴ 8 割以上の点数, でもよい.

えよう. 出た数字の情報を伝えるとき, 数字を直接言ってもよいが ($p = 1/100$), まずは, 偶数か奇数かを言ってから ($p = 1/2$), その中で何番目か ($p = 1/50$) を言っても同じ情報量になる. これを式で表すと次のようである. 対数の底は省略している.

$$I(x) = -\log \frac{1}{100} = I(y) + I(z) = -\log \frac{1}{2} - \log \frac{1}{50}.$$

ここで, x は 1 から 100 のどれかを知らせる事象, y は偶奇を知らせる事象, z は 1 から 50 番目までのどれかを知らせる事象を表している.

2.2 確率分布, 同時確率分布, 独立同時分布

情報量は確率事象を扱う文脈で定義されると書いたが, 本節では事象全体を見たときの性質として, 確率分布についての記号を整理しよう. 起こりうる事象が全部で n 種類あるとき, i 番目の事象の生起確率 p_i をまとめたものが離散確率分布 $\{p_i\}_{i=1}^n$ である. これを簡単に $\{p\}$ と書くこともあり,

$$\{p\} = \{p_i\}_{i=1}^n = \{p_1, p_2, \dots, p_n\}$$

である. 確率分布が事象 $\{x_1, x_2, \dots, x_n\}$ と対応しているとき, 事象のまとまりを大文字の X 一文字で表し, 対応する確率分布を P_X で表すことが多い. つまり, $X = \{x\}$ や, $P_X = \{p(x)\}$ などと書く.

つぎに, ふたつの異なる確率分布を同時に考えることをしよう. 事象 x_i と y_j が同時に起こる確率を同時確率と呼び, $p(x_i, y_j)$ で表す. この時, 同時確率分布は $P_{X,Y} = \{p(x, y)\}$ と書ける. ここで注意したいのは, 一般に, x と y は独立ではないため,

$$p(x_i, y_j) \neq p(x_i)p(y_j)$$

となる点である. よって, ここではさらに, $p_{i,j} = p(x_i)p(y_j)$ についての確率分布も定義したい. この確率分布に特に名前は付いていないが, 便宜上, 独立同時分布とでも呼ぼう. これを $P_X \cdot P_Y$ で表す. 複数の異なる n 種類の事象について独立同時分布を表すときは, 積記号を用いて,

$$\prod_{i=1}^n P_{X_i} = P_{X_1} \cdot P_{X_2} \cdots P_{X_n}$$

と書くこととする. 当然, 考えている全ての事象が独立であるとき, 同時確率分布と独立同時分布は一致する.

2.3 条件つき確率, 条件つき確率分布

それぞれの確率事象は一般には独立ではない. この時, ある事象 y が起こったことがわかっていると, その下で別の事象 x が起こる確率は異なってくる. この時の確率を条件つき確率と呼び, $p(x|y)$ で表す.

この考え方は, 前節で述べた確率分布にも同様に適用できる. つまり, ある事象 y が起こったことがわかっていると, その下で別の事象 $\{x_1, x_2, \dots, x_n\}$ が起こる確率が変わるため, 事象 y の下での

条件つき確率分布が定義できる.

$$P_{X|y} = \{p(x_1|y), p(x_2|y), \dots, p(x_n|y)\}.$$

2.4 情報論的エントロピー, 相互情報量

情報量は, ある特定の事象 x についての性質であった. ここでは, 確率分布 $P_X = \{p(x)\}$ についての性質として, 情報論的エントロピーを導入する. まず, 情報論的エントロピー $H(X)$ は以下で定義される.

$$H(X) = \sum_i p(x_i) I(x_i) = - \sum_i p(x_i) \log p(x_i).$$

情報論的エントロピーは, 事象 x を知ろうとした時に得られるであろう情報量の期待値を表している. この値が大きいほど, 情報を知ること (測定と言ってもよい) が有効となる. つまり, 裏を返せば, 情報論的エントロピーが大きいほど状況が乱雑であり, 前もって予想できない, ということである.

条件つきエントロピーというものも定義できる. 事象 y の下での条件つき確率を用いて,

$$H(X|y) = - \sum_i p(x_i|y) \log p(x_i|y)$$

であり, これをさらに y についても期待値をとると,

$$H(X|Y) = - \sum_j p(y_j) \sum_i p(x_i|y_j) \log p(x_i|y_j)$$

と定義できる. 同様に, 同時確率についてもエントロピーを定義でき,

$$H(X, Y) = - \sum_i \sum_j p(x_i, y_j) \log p(x_i, y_j)$$

と表す.

次に, ふたつの事象の関連度合いを示す相互情報量という量を定義しよう. これは, ふたつの確率分布 P_X, P_Y と, その同時確率分布 $P_{X, Y}$ の性質として求まる. 定義は以下のようになる.

$$I(X; Y) = H(Y) - H(Y|X) = H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y).$$

各等号が成立することは, 各自確認されたい. 意味としては, Y の情報を得たときの X の条件つきエントロピーが, Y の情報を得なかったときのエントロピーに比べてどれだけ小さくなるかを表している. つまり, X と Y の関連度を表していると言える. 式を見てわかるよう, 相互情報量は, X, Y について対称である. 実際, 和の形に書き下して書くと, 以下のようになる.

$$I(X; Y) = - \sum_i \sum_j p(x_i, y_j) \log \frac{p(x_i)p(y_j)}{p(x_i, y_j)}.$$

なお, 相互情報量の記号 $I(X; Y)$ と, 情報量の記号 $I(x)$ が重複していることには注意が必要である. 通常の情報理論では, 情報量に特定の記号を与えることはなく, 相互情報量の方を I で表す. 以降本稿では, 情報量として記号 I を用いることはない.

2.5 Kullback-Leibler divergence

同じ要素数を持つふたつの確率分布がどれだけ異なるかを定量化したい場合がある。例えば、あるイカサマさいころは、全ての目が均等確率で出るような理想的なさいころに比べるとどれだけ異なっているであろうか。この“どれだけ”に答えるのが、情報論的距離 Kullback-Leibler divergence D_{KL} という量である。同じ要素数の事象 X, Y の確率分布 P_X, P_Y を考えよう。定義は以下のようになる。

$$D_{\text{KL}}(P_X||P_Y) = \sum_i p(x_i) \log \frac{p(x_i)}{p(y_i)}.$$

この量は非負であり、0 となるのは P_X と P_Y が厳密に同じ場合のみである⁵。ただし、 D_{KL} は対称性をもたない、すなわち $D_{\text{KL}}(P_X||P_Y) \neq D_{\text{KL}}(P_Y||P_X)$ であるため、実は距離の公理をみたさない。よって正確には、情報論的擬距離と言うべきである。

さらに、 D_{KL} は、各事象 X, Y における label の付け方に依存することにも注意されたい。よって、ただ要素数が同じだけのふたつの確率分布を比べても意味をなさない。先に挙げたさいころの例のように、ふたつの確率分布どうして要素間の対応がつく場合にはじめて意味を持つ。

3 Tononi の統合情報理論

3.1 統合情報量の定義

複数の状態変数 (X でまとめて表す) が相互作用にもとづいて時間発展するとき、その状態変数と相互作用をさしてネットワークと呼ぼう。ネットワーク S が取りうる状態ベクトルのひとつ $\mathbf{x} \in X$ が時刻 t で実現しているとする。 \mathbf{x} の値は離散値をとるものとし、定まった規則によって時間ステップ毎に変化するものとする。この規則は確率で表現されていてもよい。この時、状態ベクトル \mathbf{x} についての統合情報量 $\phi(S; t, \mathbf{x})$ を定義する。そして、ネットワーク S の統合情報量を以下で定義する。

$$\Phi(S; t) = \text{Exp}_{\mathbf{x}}[\phi(S; t, \mathbf{x})] = \sum_{\mathbf{x}} p_t(\mathbf{x}) \phi(S; t, \mathbf{x}). \quad (1)$$

ここで $p_t(\mathbf{x})$ は、状態 \mathbf{x} が時刻 t で観測され得る確率である。以下では、相互作用が定常である特殊な場合を考え、上記の値が t によらない場合を考える⁶。では、 $\phi(S; \mathbf{x})$ の計算手順を述べよう。

$\phi(S; \mathbf{x})$ を計算する上で、ネットワーク S を重なりと余りなく分割する方法を考え、これを $\mathcal{P} = \{M^1, M^2, \dots, M^r\}$ で表す。ここで M^i は S の分割 \mathcal{P} における i 番目の部分ネットワークである。 M^i の状態ベクトルを \mathbf{m}^i で表す。一般に、ひとつのネットワークについて複数の分割が考えられる。この分割の中で、部分ネットワークのみで生成する情報量が最も多い分割を用いて統合情報量

⁵ ここでは証明しないが、 D_{KL} の非負性は Gibbs の不等式として知られている。

⁶ 実際の生物においても、適切な時間スケールを考えればこの仮定は十分成立しうる。相互作用が t に依存する場合にも理論を拡張することは可能であるが、実験で検証することを念頭におくと現実的でないため、ここでは考えない。

を定義する⁷.

さて、上のアイデアを実現するために、ある分割 \mathcal{P} について便宜上の統合情報量として effective information $\tilde{\phi}(S; \mathbf{x}, \mathcal{P})$ を考える。これは、 S 全体から予想される一時刻前の状態の確率分布と、 \mathcal{P} から予想される一時刻前の確率分布の間の Kullback-Leibler divergence D_{KL} を用いて定義される。

$$\tilde{\phi}(S; \mathbf{x}, \mathcal{P}) = D_{\text{KL}} \left(P_{X_0 | \mathbf{x}} \left\| \prod_{i=1}^r P_{M_0^i | m^i} \right. \right). \quad (2)$$

下付添え字の 0 は一時刻前であることを表している。ここで、注目している (部分) ネットワークの外の状態については、そのネットワーク自身は知ることができないとして、ネットワーク外の確率分布としては、最も不確かさの大きいものを仮定する。数式だけではイメージが湧きづらいが、具体的な計算の例を次節で挙げるので、しばらく辛抱いただきたい。

さて、こうして全ての分割について effective information が求まるわけであるが、これらの値を単純に比べることはできない。分割数 r や、部分ネットワークの大きさ $|M^i|$ によって、取り得る値の範囲が大きく異なるからである。そのため、以下の値で規格化した上で比較する。

$$N(\mathcal{P}) = (r - 1) \min_i [H(M_0^i)]. \quad (3)$$

ここで、 $H(M)$ はネットワーク M における状態の確率分布に関する情報論的エントロピーである。

以上で、部分ネットワークのみで生成する情報量が最も多い分割、言い換えれば、もとのネットワーク全体を考えてはじめて生成する情報量が最も小さい分割 (Minimum Information Partition) が求まる。これを、 \mathcal{P}^{MIP} で表す。

$$\mathcal{P}^{\text{MIP}}(\mathbf{x}) = \underset{\mathcal{P}}{\operatorname{argmin}} \left\{ \frac{\tilde{\phi}(S; \mathbf{x}, \mathcal{P})}{N(\mathcal{P})} \right\}. \quad (4)$$

状態 \mathbf{x} に関する統合情報量は以下のように表される。

$$\phi(S; \mathbf{x}) = \tilde{\phi}(S; \mathbf{x}, \mathcal{P}^{\text{MIP}}(\mathbf{x})). \quad (5)$$

ここで、Minimum Information Partition を求めるときのみ規格化を行い、最終的な値では規格化を行っていないことに注意されたい。これは、分割の構造を問題にするか、情報の絶対量を問題にするかの違いである⁸。

3.2 統合情報量の計算例

ここでは前節で述べた $\tilde{\phi}$ の計算法の具体例を挙げる⁹。ここで考えるのは図 1 のようなネットワークである。簡単のため、各ノードが 0 か 1 のみをとる 2 値の場合を考える。時間発展は決定論的に次の規則で決まっているものとする。すなわち、どのノードも自分に 1 の入力がある時は次

⁷ これは逆に言えば、 ϕ の値が小さくなる分割である。Minimum Information Partition と呼ばれ、 \mathcal{P}^{MIP} と表される。ネットワーク全体としてはじめて生成される情報量を考えるために、このような分割に着目する。

⁸ 規格化の方法については議論があり、この方法に異を唱える人も多い。

⁹ 2016-1st の発表で説明した例とは異なるものを挙げた。より簡単な例については、2016-1st の発表 slide を見られたい。

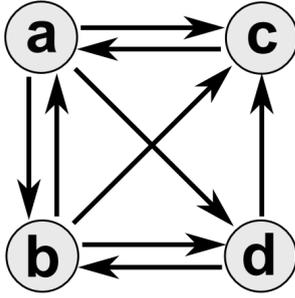


図1 具体例のネットワーク.

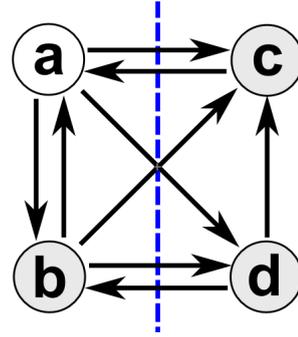


図2 初期状態と分割. 白色が1, 灰色が0を表す.

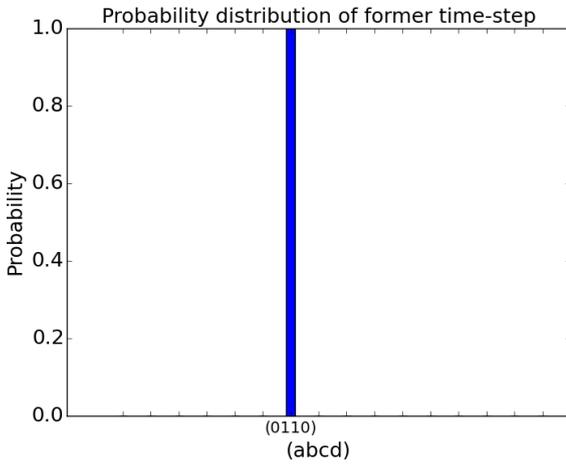


図3 全体から推定した確率分布.

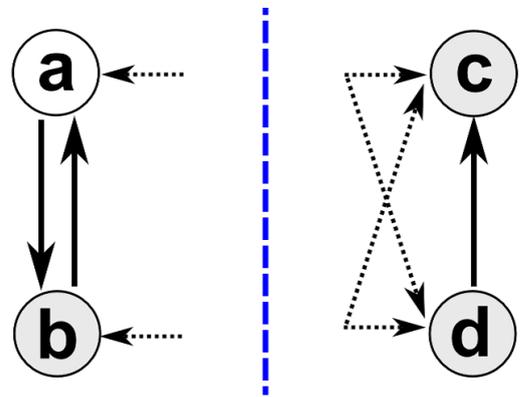


図4 部分ネットワークにおける考え方の概念図.

の時間ステップで0になり, 1の入力がふたつ以上の時は次の時間ステップで1にする. 状態ベクトルは $\mathbf{x} = (abcd)$ で表せる. ここでは具体例として状態が $\mathbf{x} = (1000)$ であるとき, 図2のような分割に関する $\tilde{\phi}$ を計算する.

まずはじめに, ネットワーク全体の情報から一時刻前の状態の確率分布を考えよう. $a = 1$ のためには, $c_0 = b_0 = 1$ でなければならない. $d = 0$ のためには, $b_0 = 1$ より, 確実に $a_0 = 0$ である. 最後に, $c = 0$ のためには, $d_0 = 0$ が必要とわかるから, 一時刻前の状態が確実に推定できて, $\mathbf{x}_0 = (0110)$ とわかる (図3).

次に, 分割後の部分ネットワークについて, 一時刻前の状態分布を推定する (図4). 図の点線の矢印は着目系の外からの入力であり, この部分の値はランダムに定まるノイズのように考える.

左部の部分ネットワーク M^1 について, 現在の状態ベクトル \mathbf{m}^1 は, $\mathbf{m}^1 = (ab) = (10)$ である. $a = 1$ より, $b_0 = 1$ は確実である. $\mathbf{m}_0^1 = (a_0b_0) = (01)$ の場合, $\mathbf{m}^1 = (10)$ を与えるノイズ入力は, $(c_0d_0) = (10), (11)$ の2通りである. 同様に, $\mathbf{m}_0^1 = (11)$ の場合, $\mathbf{m}^1 = (10)$ を与えるノイズ入力は, $(c_0d_0) = (10)$ の1通りである. ノイズ入力は最も不確かであると仮定しているから, 各ノイズ入力が起こる確率は均等で, 完全にランダムである. よって, $\mathbf{m}_0^1 = (01)$ であった確率は $2/3$, $\mathbf{m}_0^1 = (11)$

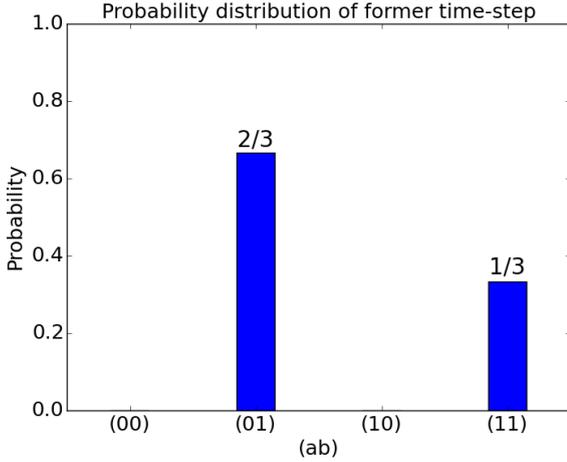


図5 M^1 から推定される確率分布.

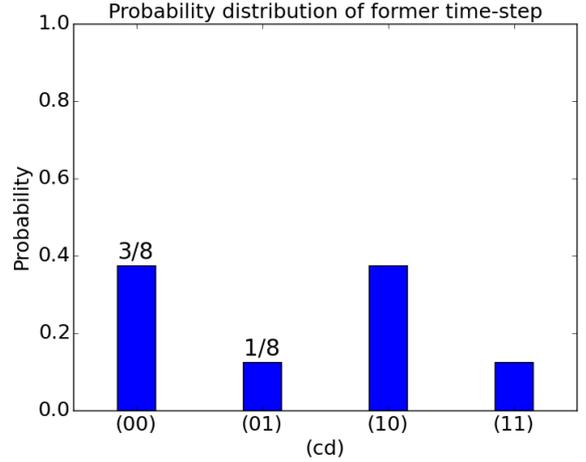


図6 M^2 から推定される確率分布.

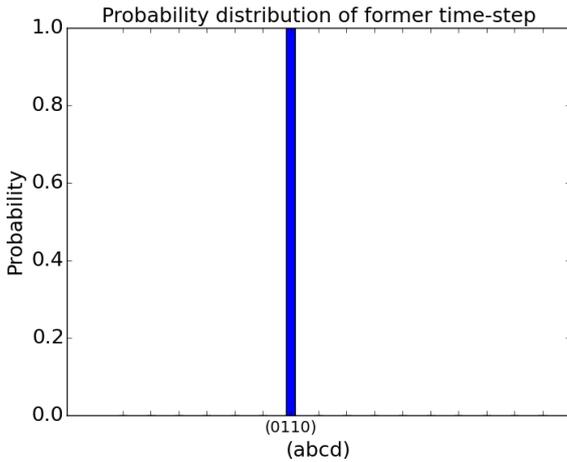


図7 全体から推定した確率分布 (再掲).

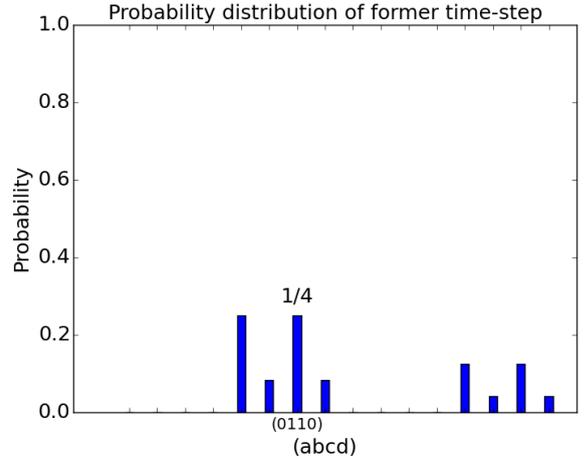


図8 分割から推定した確率分布.

であった確率は $1/3$ と推測できる. すなわち, 部分ネットワークの現在の状態 m^1 を用いて一時刻前の部分ネットワーク M_0^1 の状態について推定した確率分布 $P_{M_0^1|m^1}$ は図5のようになる.

右側の部分ネットワーク M^2 については, 同様に, 各 m_0^2 に対し, $m^2 = (00)$ を与えるノイズ入力 (a_0b_0) が何通りあるかを数え上げればよい. 結果として, 図6のような確率分布 $P_{M_0^2|m^2}$ が推定される.

以上を合わせると, 分割から推定される一時刻前の確率分布 $\prod_{i=1}^2 P_{M_0^i|m^i}$ は図8のようになる. あとは D_{KL} を計算するだけで, 次のように計算できる.

$$\tilde{\phi}(S; \mathbf{x}, \mathcal{P}) = D_{KL} \left(P_{X_0|\mathbf{x}} \left\| \prod_{i=1}^2 P_{M_0^i|m^i} \right. \right) = -\log \frac{1}{4} = 2 \text{ bit (単位は log の底を 2 とした)}.$$

これで, この分割についての effective information が求まった. あとは, 同様の計算を全ての分割のパターンについて行えば, このネットワーク S が状態 \mathbf{x} について生成する統合情報量を求めるこ

とができる.

4 機能モジュールの識別

さて, 前章で統合情報理論のバリエーションのひとつを紹介したが, 意識というものをかなり抽象化して議論していることが実感いただけたと思う. 実際, 統合情報理論それ自体は, 意識の問題とは独立した理論である. Tononi は, この統合情報量と意識を対応付けたのであるが, 別の見方もできる. つまり, 統合情報量は意識に限らず, 機能を持つモジュールを一般に識別できるのではないか, という予想である.

以降では, この立場にたつて議論を進める. 実際, 私自身は統合情報理論が意識を表すかどうかに関して懐疑的である. 一方で, 情報の統合に着目して定量化した理論は歴史的にも新しく, その意味で統合情報理論は, 生物や経済などの複雑系を考える上で有望な武器になると考える.

従来の生物学, 経済学は, 要素間の因果や相関に注目してきたわけであるが, 一般に因果, 相関と, その結果として生まれる機能は対応しない, という問題がある. 例えば, 全く同じように発火する神経細胞群は, 相関こそ高いが, 単純すぎてそれだけでは高次の機能を持ち得ない. つまり, 機能モジュールを発見するためには, 因果, 相関のみに頼った解析では不十分であり, 新たな方法論が必要とされている. 統合情報理論がその方法論となるのではないか, ということだ.

もちろん, 統合情報理論の適用にあたっては実用上の問題がある. 最も大きな障害は計算量である. 一般にネットワークを分割する場合の数は, Bell 数と呼ばれる数であり, ネットワークの大きさに対して爆発的に増大する. さらに, 前章で紹介した理論は, 離散時間, 離散状態におけるものである. 実際のデータのほとんどは連続状態であるから¹⁰, この点についても拡張の必要がある. 以上の点を解消して, 実験データに適用できる指標としたのが, 次に述べる Barrett の経験的統合情報量である.

5 Barrett の経験的統合情報量

5.1 Tononi の統合情報量についての重要な表現

Barrett の経験的統合情報量は Tononi の統合情報量を実験データから近似的に求めることを目標に考案された. その導入の前に, Tononi の統合情報量を変形しておく, 見通しが良い. さて, 統合情報量は,

$$\Phi(S) = \text{Exp}_{\mathbf{x}}[\phi(S; \mathbf{x})] = \sum_{\mathbf{x}} P(\mathbf{x})\phi(S; \mathbf{x})$$

であった. ここで, 状態 \mathbf{x} に関する統合情報量の表式は,

$$\phi(S; \mathbf{x}) = \tilde{\phi}(S; \mathbf{x}, \mathcal{P}^{\text{MIP}}) = D_{\text{KL}} \left(P_{X_0|\mathbf{x}} \left\| \prod_{i=1}^r P_{M^{\text{MIP}}_0 | m^i} \right. \right)$$

¹⁰ 時間については, サンプルング時間を考えると離散のままでも問題ない.

である. ここで, $M^{\text{MIP}^i}_0$ は Minimum Information Partition における i 番目の部分ネットワークの一時刻前の状態を表している. 注意しなければいけないことは, \mathcal{P}^{MIP} が \mathbf{x} に依存する点である. このままでは, $\phi(S; \mathbf{x})$ の具体形を代入しても, 計算を進めることができない. そこで, 分割を固定した上で effective information の期待値 $\tilde{\Phi}(S; \mathcal{P})$ を考えることとする. すると今度はうまく計算できる. D_{KL} の定義,

$$D_{\text{KL}}(P_X || P_Y) = \sum_i p(x_i) \log \frac{p(x_i)}{p(y_i)}.$$

を思い出そう. すると, 以下のように計算できる.

$$\begin{aligned} \tilde{\Phi}(S; \mathcal{P}) &= \text{Exp}_{\mathbf{x}}[\tilde{\phi}(S; \mathbf{x}, \mathcal{P})] = \sum_{\mathbf{x}} p(\mathbf{x}) D_{\text{KL}} \left(P_{X_0|\mathbf{x}} \left\| \prod_{i=1}^r P_{M_0^i|\mathbf{m}^i} \right. \right) \\ &= \sum_{\mathbf{x}} p(\mathbf{x}) \left[\left(\sum_j p(x_{0j}|\mathbf{x}) \log p(x_{0j}|\mathbf{x}) \right) - \left(\sum_{\{j_k\}_{k=1}^r} p(x_{0\{j_k\}_{k=1}^r}|\mathbf{x}) \log \prod_i p(m_{0j_i}^i|\mathbf{m}^i) \right) \right] \\ &= \left(\sum_{\mathbf{x}} p(\mathbf{x}) \sum_j p(x_{0j}|\mathbf{x}) \log p(x_{0j}|\mathbf{x}) \right) \\ &\quad - \left(\sum_{\mathbf{x}} p(\mathbf{x}) \sum_{j_1} \sum_{j_2} \cdots \sum_{j_r} p(m_{0j_1}^1, m_{0j_2}^2, \dots, m_{0j_r}^r|\mathbf{x}) \sum_i \log p(m_{0j_i}^i|\mathbf{m}^i) \right) \\ &= -H(X_0|X) - \sum_{\mathbf{x}} p(\mathbf{x}) \left(\sum_{j_1} \sum_{j_2} \cdots \sum_{j_r} p(m_{0j_1}^1, m_{0j_2}^2, \dots, m_{0j_r}^r|\mathbf{x}) \sum_i \log p(m_{0j_i}^i|\mathbf{m}^i) \right). \end{aligned}$$

式変形の二列目において, 独立同時分布をあらわに表す際に index が複雑になっていることに注意されたい. ここで, 第二項の括弧の中身について, i の和の部分分解し, $i = 1$ のみを考えてみよう. すると, j_1 以外の和が実行できて,

$$\sum_{j_1} \cdots \sum_{j_r} p(m_{0j_1}^1, \dots, m_{0j_r}^r|\mathbf{x}) \log p(m_{0j_1}^1|\mathbf{m}^1) = \sum_{j_1} p(m_{0j_1}^1|\mathbf{x}) \log p(m_{0j_1}^1|\mathbf{m}^1).$$

さらに, 条件つき確率の定義式,

$$P(x|y) = \frac{P(x, y)}{P(y)}$$

を用いれば,

$$\sum_{\mathbf{x}} p(\mathbf{x}) \sum_{j_1} p(m_{0j_1}^1|\mathbf{x}) \log p(m_{0j_1}^1|\mathbf{m}^1) = \sum_{\mathbf{x}} p(\mathbf{m}^1) \sum_{j_1} p(m_{0j_1}^1, \mathbf{m}^2, \dots, \mathbf{m}^r|\mathbf{m}^1) \log p(m_{0j_1}^1|\mathbf{m}^1).$$

\mathbf{x} の和について, 各 $\mathbf{m}^1, \mathbf{m}^2, \dots, \mathbf{m}^r$ の和に分解して実行すると, \mathbf{m}^1 以外は実行できて,

$$\sum_{\mathbf{x}} p(\mathbf{m}^1) \sum_{j_1} p(m_{0j_1}^1, \mathbf{m}^2, \dots, \mathbf{m}^r|\mathbf{m}^1) \log p(m_{0j_1}^1|\mathbf{m}^1) = \sum_{\mathbf{m}^1} p(\mathbf{m}^1) \sum_{j_1} p(m_{0j_1}^1|\mathbf{m}^1) \log p(m_{0j_1}^1|\mathbf{m}^1).$$

この最後の表式は, $-H(M_0^1|M^1)$ に他ならない. よって結局, 以下が成立する.

$$\tilde{\Phi}(S; \mathcal{P}) = \sum_{i=1}^r H(M_0^i|M^i) - H(X_0|X). \quad (6)$$

この表式は、分割が非常にうまくいったと仮定すると、さらに簡単にすることができる。Minimum Information Partition では、ネットワークを情報の観点での繋がりが最も弱い部分で分割したことを思い出そう。分割が最もうまくいったときは、部分ネットワーク同士の相関は 0 である。この時、全体のエントロピーは、単純に各部分ネットワークのエントロピーの和となる。

$$H(X_0) = \sum_i H(M_0^i)$$

現実の Minimum Information Partition でもこれに近い関係が成立すると考えれば、

$$H(X_0) \sim \sum_i H(M_0^i) \quad (7)$$

と近似できる。この近似のもとでは $\tilde{\Phi}$ はさらに使いやすい形となり、

$$\tilde{\Phi}(S; \mathcal{P}) \sim I(X_0; X) - \sum_{i=1}^r I(M_0^i; M^i) \quad (8)$$

となる。

5.2 経験的統合情報量

Tononi の統合情報量の定義では、離散状態のみを扱うため実際のデータに適用することが困難であった。そこで Barrett は式 (8) からの類推で、実験データから統合情報量の近似値を計算する指標を考案し、計算機実験でこの指標が統合情報量の良い近似となることを確かめた [3]。この指標は経験的な (empirical) 統合情報量と名付けられ、 $\Phi_E(S; \tau)$ で表される。計算の手順を述べよう。まず empirical な effective information の期待値として、次を考える。

$$\tilde{\Phi}_E(S; \tau, \mathcal{B}) = I(X_{t-\tau}; X_t) - \sum_{i=1}^2 I(M_{t-\tau}^i; M_t^i). \quad (9)$$

ここで、 \mathcal{B} は二分割 (Bipartition) を表す¹¹。あとは、Tononi の場合と同じように、Minimum Information Bipartition を探せばよい。つまり、経験的統合情報量の定義は、

$$\Phi_E(S; \tau) = \tilde{\Phi}_E(S; \tau, \mathcal{B}^{\text{MIB}}) \quad (10)$$

であり、ここで、

$$\mathcal{B}^{\text{MIB}} = \underset{\mathcal{B}}{\operatorname{argmin}} \left\{ \frac{\tilde{\Phi}_E(S; \tau, \mathcal{B})}{N_e(\mathcal{B})} \right\} \quad (11)$$

$$N_e(\mathcal{B}) = \min[H(M_t^1), H(M_t^2)] \quad (12)$$

である¹²。ここで、Tononi の場合とは、分割を選ぶ手順と期待値をとる手順の順番が逆になっていることに注意。以上の表式はデータ間の相互情報量のみで表現できているため、実験データから計算できる。

¹¹ 分割数を一般の形に拡張することは容易であるが、計算量の問題で、実用上は分割数を一定の値で区切る必要がある。Barret は最も単純な場合として、二分割で定義した。

¹² 定常状態を仮定しているため、一般にこれらの値は t によらず、 τ のみによる。 τ は時間のスケーリングと関連する。

ただしこの場合、状態の確率分布や条件つき確率分布をデータから構築する必要があるため、膨大なサンプルが必要となる。そこで、データが多変量正規分布に従う場合に限り、共分散を用いたより簡便な計算法を用いることができる。今回は計算法の詳細には触れないが、興味のある方は文献 [3] を参照されたい。なお、次章のシミュレーションでは共分散を用いた計算法を用いている。

6 シミュレーション

統合情報量が時系列データから機能モジュール¹³を特定しうるか、簡単なモデルについてシミュレーションを行ってみた。用いたのは、ガウス雑音のもとでの、線形かつマルコフ的な系である。この系においては、状態の時間発展は、

$$X_t = A \cdot X_{t-1} + E_t$$

で記述される。ここで、 A は重み付きの隣接行列であり、ネットワークの構造を表している。 E_t はガウス雑音を表すベクトルであり、各成分は独立に、平均 0、分散 1 のガウシアンノイズとする。

シミュレーションの手順は以下の通りである。

1. $X_t = A \cdot X_{t-1} + E_t$ に従って時系列データを生成する。
2. データの成分の中から特定の成分の組を選ぶ (部分ネットワーク)。
3. 選んだ成分の組を 2 つの群に分ける (Bipartition)。
4. データから共分散行列を計算し、 $\tilde{\Phi}_E$ を計算する¹⁴。
5. 手順の 3–4 を繰り返すことで、Minimum Information Bipartition を見つけ、 Φ_E を求める。
6. 手順の 2–5 を繰り返すことで、 Φ_E の値が大きい部分ネットワークを見つける。

今回シミュレーションしたのは、図 9 のようなネットワークである。ノード間の結合重みは 0.25 と小さくしたため、生データをみればわかるよう (図 10)、激しいノイズののったデータとなっている。このデータのみから構造を推定することを試みた。生データとしては、初期条件から十分ランダムに変化したと思われる時間 (3000 step) を $t = 0$ として、そこから 3000 step 分のデータを用いた。 $\tau = 1$ とし、多変数ガウス分布の仮定の下で計算を行った。なお、今回は時間とプログラム技術の関係で¹⁵、ノードが 3 つから構成されるネットワークに絞って網羅的に調べた。計算の結果は表 6 のようになった。

この結果をみると、(123), (456) の部分ネットワークだけ桁違いに統合情報量が大きい。今回のように雑音が多いデータについても、モジュール識別のために統合情報量が有用である可能性が示唆される¹⁶。

¹³ このモデルは単純であるため、機能と相関との区別がつかない。統合情報量の有用性を検証するためには、相関と機能が一致しない系を調べる必要がある。生物や論理回路を実験的に調べるのが近道と考える。

¹⁴ 共分散を用いた計算方法については、文献 [3]。

¹⁵ 自分は、網羅的な分割を簡潔に記述できるアルゴリズムを考えつかなかった。

¹⁶ 今回のモデルは非常に簡単であるため、この結果が統合情報量の威力を保証するわけではないが、最低限、識別できてほしいものは識別できた。

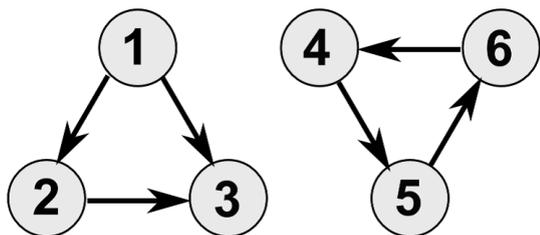


図9 シミュレーションしたネットワーク. フィードフォワードとフィードバックのネットワークが分離して存在している. 各エッジの重みは全て 0.25 とした.

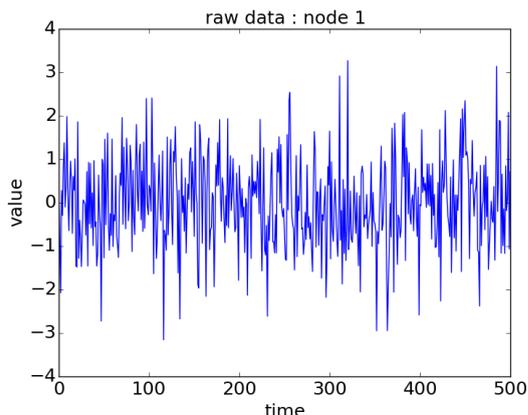


図10 ノード1 についての生データ.

表1 ノード3つの部分ネットワークについての統合情報量の計算. 単位は bit.

部分ネットワーク	Φ_E	部分ネットワーク	Φ_E	部分ネットワーク	Φ_E
(123)	0.072	(456)	0.095	(124)	0.001
(356)	0.00076	(125)	0.00086	(346)	0.00089
(126)	0.0009	(345)	0.00038	(134)	0.00069
(256)	0.00062	(135)	0.00088	(246)	0.00077
(136)	0.00087	(245)	0.00075	(145)	0.0012
(236)	0.0011	(146)	0.00065	(235)	0.00017
(156)	0.001	(234)	0.00092		

7 まとめ

統合情報理論は意識を記述するために考案されたが,それを適用できる範囲は広く,複雑系の機能の程度を一般的に表現すると期待される. 本稿では, Tononi による統合情報理論を紹介したのち, 実験に適用する際の有用な近似として, Barrett の機能的統合情報量を導入した. これを用いて, 時系列データからシステムの機能モジュールを推定できることを, 簡単なモデル上で実際に確かめた.

一般に因果関係の構造と, 機能モジュールとは一致しないはずである. 近年の因果推論に関する研究の発展はめざましいが, それだけでは高次の機能を解き明かすには足りない. 統合情報理論が複雑系を解明する突破口となることを夢想して, 本稿を終えたい.

謝辞

Acta Epsilonica や Group Epsilon の活動を通して, 本稿を書くきっかけを与えてくださった山下弘一郎先生に感謝いたします. また, 本稿の Review を通して, 多くの有用なコメントをくださった

Referee の方に感謝いたします.

参考文献

- [1] MARCELLO MASSIMINI AND GIULIO TONONI, 花本知子 (訳): 意識はいつ生まれるのか, 亜紀書房, 2015.
- [2] 甘利俊一: 情報理論, ちくま学芸文庫, 2011.
- [3] ADAM B. BARRETT AND ANIL K. SETH: Practical Measures of Integrated Information for Time-Series Data, *PLoS Comput Biol* **7** (2011) 1.
- [4] DAVID BALDUZZI AND GIULIO TONONI: Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework, *PLoS Comput Biol* **4** (2008) 6.