解 説

自由エネルギー原理は脳型人工知能の基盤となるか
～アクティブインファレンスの意義について～

大 羽 成 征

京都大学 情報学研究科*

How Does the Free Energy Principle Inspire Artificial Intelligence?

Shigeyuki Oba

Graduate School of Informatics, Kyoto University*

概要

自由エネルギー原理 (FEP) および能動推論 (AIF) は、動物の脳が環境を知覚し認識しこれに基づいて行動を起こす過程の情報処理機構全般を対象とした説明原理であり、脳の神経計算機構のモデルであるのみならず、動物やヒトを取り巻く実環境タスクに向き合う人工知能設計のためにも示唆するところが大きい。本論では、FEP および AIF が依って立つ基盤である変分推論および確率的推論としての制御について、それぞれその基礎を確認することによって FEP および AIF の意義を浮かび上がらせることを試みる。

1. 導 入

ヒトの知能を理解することによって計算機上にヒトの知能と遜色のない機能をつくりたい／ヒトの知能と遜色のない機能を計算機上につくれる程度にまでヒトの知能を理解したい。こうした脳型人工知能を志向する研究動機から、生物の神経細胞やそのネットワーク構造や動態を調べるボトムアップ・アプローチと生物が解いているタスクからそれに必要な計算過程を調べるトップダウン・アプローチの両輪が必要であるが、両者の間のギャップは未だ大きい。

自由エネルギー原理 (free energy principle ; FEP) は、脳における認識・判断・行為の過程の情報処理機構全般を対象とした説明原理であり、予測符号化モデルに端を発して、不確実環境を対象としたベイズ推定を基盤においた拡張・精緻化が加わっている。とくに FEP を用いた能動推論 (active inference ; AIF) のアイデアによって、予測符号化が行動生成過程のモデル¹⁾とリンクしたところに特徴があり、他者の行動を自らの内部モデル中にフィッティングすることによって理解する過程 (ミラーニューロン系) のモデル²⁾や、所要の目的に必要な情報を探索によって獲得する過程における情報獲得の価値 (epistemic value) の導出³⁾などの発展がある。その一方でこれらと平行する形で、

FEP が依って立つ基盤である変分推論および確率的推論としての最適制御の方法論も、機械学習や知的制御の立場から見て急速に研究が進んでいる。

本論では、FEP および AIF が依って立つ方法論の基礎を確認することによって FEP や AIF の意義を浮かび上がらせることを試みる。2 章では変分推論について説明し、自由エネルギーを最適化するという言葉の意味するところを精緻に議論する。3 章では確率的推論としての制御について現代的な視点を解説し、AIF の特色と意義について議論する。4 章では AIF のアルゴリズム導出の詳細な手順を確認し、これの妥当性について議論する。

2. 変 分 推 論

2.1 尤度原理と変分推論

FEP の理論的基盤は、変分推論 (variational inference) もしくは変分ベイズ推定 (Variational Bayesian estimation ; VB) と呼ばれる統計計算手法である。変分推論は尤度原理と事前分布に基づいて未知変量の分布推定を制約付きで行う方法である。最尤推定やベイズ推定と比較するとき、事前分布利用の有無、点推定と分布推定の違い、分布推定に対する制約の有無、を鍵とした対立軸があって、以下のように整理することができる。

- 最尤推定 $x_{\text{ML}}^* = \arg \max_x p(s|x)$

* 〒 606-8501 京都府京都市左京区吉田本町

- MAP 推定 $x_{\text{MAP}}^* = \arg \max_x p(s|x)p(x)$
- 完全ベイズ推定 $q_{\text{Full}}^*(x) = p(x|s) = \frac{p(s|x)p(x)}{p(s)}$
- 変分ベイズ推定 $q_{\text{VB}}^*(x)$ は $q()$ の汎関数 $\mathcal{F}[q; s]$ を $q()$ に関して適当な制約のもとで最大化

ここで $p(s|x)$ は未知変数 x を条件とした観測 s の条件付き確率を表し、これを観測 s を所与としたときの未知変数 x の関数と解釈するとき、未知変数 x の尤度と呼ぶ。尤度関数が未知変数 x の関数であることを強調するために $L(x; s) = p(s|x)$ のように表すことがある。対数尤度の符号反転 $-\ln p(s|x)$ を観測 s のサブライズと呼ぶこともある。

完全ベイズ推定の目的は、天下一的に与えられた未知変数 x の事前確率分布 $p(x)$ の下で、事後確率分布 $q_{\text{Full}}^*(x) = p(x|s)$ を求めることであるが、これを数値計算するうえで必要なエビデンス $p(s) = \int p(s|x)p(x)dx$ の計算が特別な場合を除いて困難となる。変分推論では、求めたい事後確率分布 $q(x)$ に制約を与えることで計算を簡素化する。制約の入れ方は場合によって様々であるが、未知変数 $x = (x_1, \dots, x_N)$ が N 次元ベクトルであるとき、 $q(x) = q(x_1) \cdots q(x_N)$ のように成分毎の確率分布の積で書けるような分布を想定する平均場近似、 $q(x) = \mathcal{N}(x; \mu, \Sigma)$ のようなガウス分布を想定するラプラス近似、 $q(x) = q(x_1, \dots, x_M)q(x_{M+1}, \dots, x_N)$ のように任意の組み合わせ毎に同時確率を考えつつ組み合わせ間の独立性を仮定する独立分解仮定などの制約が広く知られている。変分推論は以下の目的関数を未知確率分布 $q(x)$ に関して上記の制約のもとで最大化する。

$$\begin{aligned} \mathcal{F}[q; s] &\equiv \int q(x) \ln \frac{p(s|x)p(x)}{q(x)} dx \\ &\leq \ln p(s) \end{aligned} \quad (1)$$

の左辺 $\mathcal{F}$ は確率分布 $q(x)$ の汎関数であり対数エビデンス $\ln p(s)$ の下界となっており、エビデンス変分下界もしくは単に変分下界と呼ばれる。等号は $q(x) \equiv q_{\text{Full}}^*(x)$ のときに成立する。変分下界の符号反転 $-\mathcal{F}$ を変分自由エネルギーもしくは自由エネルギーと呼ぶ^{†1}。尤度 $p(s|x, h)$ や事前確率 $p(x|h)$ にハイパーパラメタを含み、エビデンス基準 $h^* = \arg \max_h p(s|h)$ でハイパーパラメタを決定したい場合も同様に取り扱うことができる。このとき変分下界 $\mathcal{F}[q, h; s]$ は $q(x)$ の汎関数であると同時にハイパーパラメタ h の関数となる。さらに制約付きの事後確率 $q(x)$ の関数形状もハイパーパラ

メタ h' で表すことにするならば、変分下界は事前・事後両確率分布の関数 $\mathcal{F}[h', h; s]$ と解釈することもできる。こうして変分下界の最大化は、関数 $\mathcal{F}[h', h; s]$ の数値最適化問題に帰着するのである。

変分下界の最大化を行う数値最適化によって $q(x)$ を求める逐次近似アルゴリズムは、上界 $\ln p(s)$ が与えられていることで収束が保証されており、 $q(x)$ と h に関する同時最適化アルゴリズムは上界 $\ln p(s|h)$ が h の関数になっているものの、 $\ln p(s|h)$ が h に関して上に発散しない限りは同様に収束保証が得られる。また完全ベイズ推定は $q(x)$ に関して制約のない特別な場合として変分ベイズ推定に含まれると言える。

変分推論は、深層神経網を用いた確率的ニューラルネットワーク (Edward もしくは TensorFlowProbability)⁴⁾ における主要な方法としても用いられている。このフレームワークはおおよそあらゆる確率的推論を変分推論の特別な場合として統一的に考えることで議論を単純化しているのが特徴であり、 $q(x) = \delta(x - x^*)$ を制約として与えた特別な場合として MAP (maximum a posteriori) 推定を、 $q(x)$ を B 個のランダムサンプルの経験分布 $q(x) = \sum_{b=1}^B \delta(x - x^{(b)})$ に制約した特別な場合として MCMC (Markov chain Monte Carlo) 法アルゴリズムを、 $q(x)$ をパラメタ ϕ で規定されたニューラルネットワーク $x = G(x'; \phi)$ にノイズ x' を入力して生成する場合に制約した特別な場合として GAN (generative adversarial network) を、それぞれ変分推論に含めている。最近のチュートリアル論文⁵⁾ が変分推論に関する近年までの進展を近代的な視点から詳しくまとめているので参照されたい。

2.2 FEP は「数値最適化」ではない

多くの機械学習アルゴリズムは目的関数の数値最適化問題の形で定式化されている。数値最適化問題は「陽に定義された目的関数 $f(x)$ が最大値 (もしくは最小値) をとるような $x \in X$ を解空間 X の中に探す問題」を意味する。ここで目的関数が陽に定義されているとは、 x を与えたときに $f(x)$ を数値的に得る方法があるということである。

一方で、脳の計算論における自由エネルギー原理 (脳の FEP) も機械学習同様に「変分自由エネルギーを目的関数とした最適化である」と説明される。実際に特集号中の総説記事中にも最適化という言葉違いが随所で見られる。しかしここでの含意は「数値最適化問題」で言うところの最適化と厳密には異なっている。この違いがしばしば誤解の余地を生むため、以下でその余地を十分に減らすための議論を行っておく。

^{†1} 「変分下界の最大化は、自由エネルギーの最小化」とするのが統計・熱力学の用語法だが、後者の符号を逆とした文献が多いことに注意が必要。

2.2.1 FEP に厳密な最適解は必要でない

深層ニューラルネットなどの複雑なモデル一般におけるパラメタ推定において、局所探索アルゴリズムが広く用いられている。局所探索アルゴリズムは目的関数を評価しながら解空間中で暫定解を逐次的に更新しつつ探索し、局所解（パラメタ空間内の特定の点の近傍に限った最適解）に至って停止する。局所探索アルゴリズムの停止解として得られるのは、解空間における目的関数の局所最適解、鞍点、もしくはこれらの近傍であり、厳密な意味での数値最適化問題の解（大域的最適解）が局所探索の手法で得られることは一般に保証されない。しかし、停止解として得られる解が有用であり高い性能を示すことがニューラルネット学習に関して繰り返し示されている。

自由エネルギーを目的関数とした FEP も典型的な局所探索の手法である勾配法を想定しているが、大域的最適解を必要としているわけではない。

2.2.2 FEP の目的関数は定常でない

変分推論法のもとで観測データに基づいて定常な環境を学習する問題は、バッチ学習・オンライン学習もしくはミニバッチ学習によって等価な定式化が可能である。定常な環境において、観測 $s_1, \dots, s_T$ が尤度関数で表現された生成モデル $p(s|x)$ の i.i.d. 標本であると考えられる。このとき対数尤度関数 $\ln L(x) = \ln p(s_1, \dots, s_T|x) = \sum_t p(s_t|x)$ や自由エネルギー $\mathcal{F} = \int q(x)(\sum_t \ln p(s_t|x) + \ln p(x) - \ln q(x))dx$ の最大化問題を勾配法で解く場合を考えよう。オンライン学習では観測 s_t をひとつ評価するごとに x の一時的な推定分布 $q(x)$ を更新する。ひとつずつでなく、少数の観測をまとめたミニバッチごとに評価する場合をミニバッチ学習と呼ぶ。環境の定常性を仮定できればオンラインでもミニバッチでも通常のバッチ型でも同じ問題を解いていることになり、大域解は同一であることが保証される^{†2}。変分ベイズ学習のオンラインバージョンについては、ATR の佐藤雅昭による 2001 年の論文⁶⁾ がパイオニアワークであり収束条件についても詳しく議論している。

一方でオンライン学習を考えると、必ずしも環境の定常性を仮定しない場合も多い。たとえば Actor-Critic 型強化学習など、複数モジュールから成る学習器がお互いをお互いの環境の一部として扱いながら同時学習する場合には、学習の過程においてお互いの環境が変化し続けていることになる。このときオンライ

ン学習の学習係数は、過去を忘却しつつ現在に適応してゆく忘却プロセスを制御する。FEP が動物や人間の脳のモデルとして想定しているのは、こうした定常でない環境に適応するためのオンライン学習である。

定常性を想定しないオンライン学習において、認識状態 $q(x)$ の評価関数である自由エネルギー（変分下界の符号反転） $-\mathcal{F}[q(x)|s(t)] = -\int q(x)(\ln p(s(t)|x) + \ln p(x) - \ln q(x))dx$ は、時刻 t における認識状態 $q(x)$ に基づく観測 $s(t)$ のサプライズと、認識状態 $q(x)$ の事前確率 $p(x)$ から見た距離とをバランスさせている。これを用いて観測 $s(t)$ 毎（もしくは短期間の観測に対応するミニバッチ毎）に勾配 $\partial \mathcal{F} / \partial x$ を生成し、これを用いて認識状態 $q(x)$ を更新する。自由エネルギーは各時刻の観測 $s(t)$ の関数であるが、その主な機能は認識状態の内部表現更新に用いる勾配ベクトルの生成である。しかるに認識状態や内部モデルを更新し続けてゆくダイナミクスを「自由エネルギーの最適化」と呼ぶのは、自由エネルギーがあたかも定常な目的関数であるかのように誤解させる可能性が高い。

3. 状態空間モデル・モデル予測制御・能動推論

能動推論（active inference ; AIF）は予測コーディングによる世界認識に基づいて運動制御信号を生成する過程のモデルであり、Friston はこれを最適制御の枠組みのオルタナティブであると主張している⁷⁾。この主張は妥当だろうか？ この節では、まず状態空間モデル・最適制御・モデル予測制御・強化学習の枠組みを概観し、次にこれを FEP に基づく能動推論の枠組みと対応づけることで異同を議論する。

3.1 状態空間モデルと推定

状態空間モデル（state space model ; SSM）とは、状態の時間発展規則を記述する力学系モデルと状態に依存した観測の生成過程を記述する観測系モデルの組み合わせであり、典型例として両者について加法的ノイズを想定する以下のモデルが挙げられる。

$$x_{t+1} = f(x_t, u_t) + w_t \quad (2)$$

$$y_t = g(x_t) + z_t \quad (3)$$

ここで x_t, y_t はそれぞれ時刻 t における状態と観測を意味する。 u_t はこの系に対する制御信号である。この節では u_t を既知定数として無視し、次節で制御信号の決定法について議論しよう。 $f()$ をダイナミクスモデル（もしくはシステムモデル）、 $g()$ を観測モデルと呼び、 w_t をシステムノイズ、 z_t を観測ノイズと呼ぶ。ダイナミクスモデル・観測モデルともに一般に非線形関数を許す。

^{†2} ただし、異なる局所解に収束する可能性は排除できない。また、学習係数のスケジューリングに依存して収束が保証されない場合もある。

状態 x_t に関する不確かさを含んだ推定を、事後確率 $q_t = q(x_t) = p(x_t|y_{1:t})$ の形でベイズ推定によって求めることが、状態空間モデリングの基礎的な方法論になっている。とくに過去観測 $y_{1:t}$ に基づく現状態 x_t のベイズ推定を、一つ前の状態 x_{t-1} のベイズ推定を現時点の観測 y_t を用いて逐次的に更新してゆく逐次ベイズ更新則が効率的な計算原理として知られている。

$$q(x_t) \propto \int q(x_{t-1})p(x_t|x_{t-1})p(y_t|x_t)dx_{t-1} \quad (4)$$

ここで $p(x_t|x_{t-1})$ と $p(y_t|x_t)$ はそれぞれダイナミクスモデルと観測モデルの条件付き確率表現である。とくに線形・ガウス状態空間モデル（ダイナミクスモデル $f(x)$ および観測モデル $g(x)$ が入力 x に対して線形であり、システムノイズと観測ノイズが共に既知の共分散行列を持つガウス分布に従う）を想定するとき、ダイナミクスの非線形性や観測ノイズの非ガウス性が想定される場合にも近似法が提案されている。

予測符号化では脳内の認識状態という外部から観測できない隠れ変数 x_t のモデル化に SSM を用いる。すなわち、観測 y_t を予測するのに必要な情報が脳内認識状態 x_t として符号化されており、 x_t を観測 y_t の予測精度が高くなるように更新してゆくと考え。ダイナミクスモデル $f()$ や観測モデル $g()$ がパラメタによって調整可能であるときは、このパラメタも観測 y_t の予測精度が高くなるように更新してゆく。「予測精度が高くなるように更新」の要請は自由エネルギー原理によって満たされる。

SSM を用いる目的は、第一にノイズを含んだ観測の時系列 $y_{1:t}$ から現状態 x_t を推定すること、第二に将来得られるであろう観測 y_{t+1} を予測すること、第三にモデル $f(), g()$ が未知であるときにこれらを推定することである。状態 x_t として対象を表すために十分に高い次元のベクトルを想定したりリッチなシステム同定を目的とする場合もあれば、低い次元の特徴ベクトル状態を想定することで、冗長な観測に対する特徴抽出機能を期待する場合もある。さらに、次節で述べるように、制御信号が力学系の主要項 $f()$ に影響を与える場合を想定して制御目的に沿うような制御信号の決定に用いたり、他エージェントの制御目的を推定するなど、SSM の応用範囲の広さは「予測符号化」の一言で収まらない。

3.2 SSM に基づくモデル予測制御

最適制御は、ダイナミクスモデル $f = f(x_t, u_t)$ が制御信号 u_t によって介入可能であるとき、目的関数 J で定義された制御目的に沿った結果が得られるように

制御信号 u_t を設計する。

$$J = J(x_{t+1:T}, u_{t+1:T}) = \sum_{t'=t+1}^T J_{t'}(x_{t'}, u_{t'}) \quad (5)$$

ここで右辺の $J_{t'}(x_{t'}, u_{t'})$ は時刻 t' における状態 $x_{t'}$ および制御信号 $u_{t'}$ の好ましさを表す。 T は目的関数の定義に関わる未来時区間の限界を表し、 T を無限遠にとることもある。目的関数 J や、その局時的表現 $J_{t'}$ は設計者が任意に与えるが、具体例として時刻 t' における目標状態 $x_{t'}^*$ を与えつつ、

$$J_{t'}(x_{t'}, u_{t'}) = -(x_{t'} - x_{t'}^*)^T Q (x_{t'} - x_{t'}^*) - u_{t'}^T R u_{t'} \quad (6)$$

のような二次形式の目的関数がよく研究されている。目的関数 J が x, u に関する二次形式であり、なおかつダイナミクスモデル $f = f(x, u)$ が x, u に関して線形であるときの最適制御器 $u = u_{\text{LQR}}(x)$ を線形二次制御器 (linear quadratic regulator; LQR) と呼ぶ。ダイナミクスモデルが x, u に関して非線形であるときも、 $f(x, u)$ の局所的線形化によって $u_{\text{LQR}}(x)$ を求めて用いることで将来の制御信号 $u_{t+1:T}$ を最適化することができる (開ループ制御)。

モデル予測制御によれば、開ループ制御の計算を随時 (理想的には $t' = t+1, t+2, \dots$ の各時刻において) 行つて、その時刻 t' から見た将来制御信号 $u_{t'+1:T}$ とこれによる将来軌道計画 $x_{t'+1:T}$ をその都度同時に決定することで、目標に到達するまでに軌道計画を随時修正しながら目的を達することができる。ダイナミクスモデル $f(x, u)$ が真のダイナミクスと異なることによって生じるモデル化誤差や、計画外の擾乱に対して頑健になりやすい。

モデル予測制御は、これ自体がモデル化誤差や擾乱を想定しているが、さらに確率モデルを入れ込むことによる改善法が考慮されている。線形ダイナミクスモデルにガウス雑音の入った状態空間モデルを想定して、現状態の事後確率をカルマンフィルターで求めながら開ループ制御を行う線形二次ガウス制御器 (LQG; linear quadratic Gaussian regulator) を基礎として、非線形ダイナミクスモデルが局所線形化できる場合への拡張がよく研究されている⁸⁾。観測ノイズやシステムノイズのモデルがガウス分布的でない場合のモデルを想定したシステム同定法としてパーティクルフィルタが広く使われているが、将来軌道計画までをパーティクル表現しながらモデル予測制御を行う方法 (PF-MPC)⁹⁾ も提案されている。計測時雑音の多い場合のシステム同定一般に適用可能であり、生体メタボリックシステ

ムの制御¹⁰⁾などの事例で有効であった。

3.3 確率的推論としての制御

過去までの観測に基づいて、現時刻 $t = 1$ における状態 x_1 の事後確率 $q(x_1)$ が与えられているものとする。またダイナミクスモデル $p(x_{t+1}|x_t, u_t)$ も既知であるものとする。以上の前提のもとで目的関数 J を最大にする最適制御 $u_t, t = 1, \dots$ を求める問題を、確率的推論の枠組みに乗せるのが確率的推論としての制御である。

制御信号の生成を確率的推論の一貫として行うというアイデアは、これまでにカルマンフィルタと線形最適レギュレーターの双対性 (Todorov, 2008)¹¹⁾、KL 制御 (Kappen, 2011)¹²⁾ (Kappen et al., 2012)¹³⁾、確率的最適制御 (Toussaint, 2009)¹⁴⁾などで提案されてきた。

近年これらを統一的に理解する試み (Levin, 2018)¹⁵⁾として「各時刻の制御信号が最適であるか否かを指定する隠れ変数」を導入する説明方法が考案され示唆に富んでいる。

Levin (2018)¹⁵⁾ は将来時刻 t における制御の最適性を表す隠れ変数 $o_t \in [0, 1]$ を導入し、これの生成モデルを以下で表すことで最適制御を確率モデルに変換した。

$$p(o_t = 1|x_t, u_t) = \exp(J_t(x_t, u_t)) \quad (7)$$

ここで定式化を簡単化するために $J_t(x_t, u_t) < 0$ を仮定するが、適当な変換によってこの仮定を外すことは容易であるため一般性を失わない。すると時刻 $t = 1, \dots, T$ における制御信号の全てが最適 $o_t = 1$ であるという条件 $\mathbf{o}$ のもとで、全ての未知変数の同時確率を以下で表すことができる。

$$\begin{aligned} p(x_{1:T}, u_{1:T}|\mathbf{o}) &\propto p(x_{1:T}, u_{1:T}, \mathbf{o}) \\ &= p(x_1) \prod_{t=1}^T p(o_t = 1|x_t, u_t) p(x_{t+1}|x_t, u_t) p(u_t) \\ &= p(x_1) \prod_{t=1}^T p(x_{t+1}|x_t, u_t) p(u_t) \exp\left(\sum_{t=1}^T J_t\right) \end{aligned} \quad (8)$$

ここで $p(u_t)$ は制御信号の事前確率であり最も簡単には一様分布を考えてよい。このモデルは全体として単純な隠れマルコフモデル (図 1) になっている。そこで、制御信号が最適であった場合の状態・行動系列 $(x_t, u_t), t = 1, \dots, T$ の事後確率を推定する問題は、具体的には、確率伝搬法の要領で $\beta(u_t, x_t) \equiv p(\mathbf{o}|u_t, x_t)$ の値を $t = T, \dots, 1$ の順に後ろ向きに決めてゆき、これ

図1 確率的推論としての最適制御を説明するグラフィカルモデル¹⁵⁾。将来時刻 t における制御信号 u_t が状態 x_t のもとで最適であるか否かを示す変数 $o_t \in \{0, 1\}$ を導入し、確率モデルと即時制御コスト関数 J_t を $p(o_t = 1|x_t, u_t) = \exp(J_t(x_t, u_t))$ によって繋ぐ。すると将来の制御と状態の系列を $o_t = 1, t = 1, \dots, T$ を所与としてベイズ推定することで、最適制御と等価な結果が得られる。

の周辺化 $\beta(x_t) = \int p(\mathbf{o}|u_t, x_t) p(u_t) du_t$ を求めておけば、最適な制御方を $\pi(u_t|x_t, \mathbf{o}) = \beta(u_t, x_t)/\beta(x_t)$ のようにして求めることができる。

以上の定式化はシンプルであるがゆえに汎用性が高い。たとえば、ダイナミクスのモデル

$p(x_1) \prod_{t=1}^T p(x_{t+1}|x_t, u_t)$ が全体として未知であるとき、これを推定する過程は変分推論で定番の方法を使って導出できる。まずモデルを表す因子の各々をパラメトリックな未知確率分布 $q(x_1; \phi_{x_1}), q(x_{t+1}|x_t, u_t, \phi_x)$ としてモデリングすれば、エビデンスの変分下界はパラメタ ϕ_{x_1}, ϕ_x の関数で表すことができ、問題は第2章で述べたような最適化問題に帰着する。

3.4 AIF と最適制御の計算モデル

Friston (2010) は制御信号 u を計算するうえで、AIF と最適制御の方法論を比較して、両者に大きな違いがあると主張している¹⁾。しかし前節で概観したように当該論文発表当時と比べて、「確率的推論としての最適制御」に関する知見は進んでいる。近年の水準と改めて比べた場合に本質的な違いはあるだろうか？ また、そこに本質的なアドバンテージはあるだろうか。以下、計算モデル、運動学習モデル、実装モデル、3つの論点から比較しながら論じたい。

まず FEP に基づく AIF で行う計算過程は「確率的推論としての最適制御」の特別な場合として含まれている。Friston の AIF は確率微分方程式系で定式化さ

を紹介しているので、合わせて参照いただきたい。

脳内の認識状態 x は各時刻の感覚状態 s を観測しながら更新されてゆく。このさい、時々刻々と認識状態 x に基づく感覚状態 s の予測が行われて、予測誤差 ξ_s が計算されている。認識状態 x は、予測誤差 ξ_s をゼロに近づけるべく誤差フィードバックによって更新されてゆくものとする。予測符号化とはここまでのプロセスのことを言う。FEP では認識状態を事後確率分布 $q(x)$ で表し、予測を期待サプライズ $-\int q(x) \ln p(s|x) dx$ で表すという違いはあるが、AIF の説明上ではこの違いは本質的でない。

簡単のためシステム全体の制御目的を、 x の空間内で定義された目標状態 x^* の形で表わすものとする。これは二次形式の目的関数 $J_t(x) = -(x - x^*)^T Q (x - x^*)$ で表すこともできるし、これに対応した正規分布形の事前確率分布 $p(x)$ で表すこともできる。これは、内部誤差 $\xi_x = \Pi_x(x - x^*)$ をゼロに近づけることと言い換えることができる。ここで行列 Π_x は認識状態事後確率 $q(x)$ の分散や二次形式目的関数の重み行列 Q に依存して定まる。ここまでは、AIF も通常の最適制御も同様である。

AIF と最適制御の間の本質的な違いは制御信号 a の計算を行う逆モデル関数においてみられる。最適制御において逆モデルは認識状態 x と目標状態 x^* の関数となっているのに対して、AIF における逆モデルは感覚誤差 ξ_s のみの関数となっている。Friston らはこの点を AIF の特長として主張している¹⁾。つまり、AIF ならば感覚誤差情報のみを用いて制御信号を生成できれば神経実装が容易であり、脊髄における実装も可能である。最適制御的にこれを計算したいならば、認識状態と目標状態の情報を集約しつつさらに制御信号空間に変換しなければならない。

AIF の方式でもう一つ着目すべき点は、認識状態 x を更新するさいに感覚誤差 ξ_s と内部誤差 ξ_x からのフィードバック信号を両方受け取る点である。この結果 AIF では、認識状態 x が内部誤差 ξ_x の影響を受けて目標状態 x^* 方向に偏って推定される。AIF では、目標状態 x^* 方向に偏った認識状態 x による感覚状態の予測 $\hat{s} = g(\hat{x})$ と実際の感覚 s との差分 $\xi_s = \Pi_s(s - \hat{s})$ をゼロに近づけるような制御信号 a を出力することで、状態を目標状態 x^* に近づけてゆくのである。最適制御では感覚誤差信号 ξ_s のみを受け取ることで認識状態 x を偏りのない「客観的」な推定値に保持しつつ、これを目標状態に近づけるべく最適な制御信号を生成する。

4. AIF の実装と微分可能ダイナミクス

AIF による行動の実装は環境ダイナミクスモデルの微分に依存する。とくに環境と認識状態の確率微分方程式のもとで AIF を定式化した Action and Behaviour 論文¹⁾では、自由エネルギー $\mathcal{F}$ が知覚 $s = s(a)$ を通してアクション a の関数として解釈できることを用いて、アクション a を勾配方向 $\mathcal{F}_a \equiv \partial \mathcal{F} / \partial a$ に向けて更新する。(ここで右下添字が変数であるとき、変数による偏微分を意味するものとする表記法を用いている。) この更新式は当該論文の Appendix の (A3.3 式) において以下のように示されている。

$$\dot{a} = -\mathcal{F}_a = -\tilde{\epsilon}_a^T \tilde{\xi} \quad (9)$$

$$\tilde{\epsilon}_a = \mathbf{g}_{\mathbf{x}} \sum_i D^{-i} \mathbf{f}_{\mathbf{x}}^{i-1} \mathbf{f}_a \quad (10)$$

以下では、これの導出とともに、この更新式が意味するところを解説したい。

4.1 AIF 行動更新式の導出

自由エネルギー $\mathcal{F}$ の主要項は $\frac{1}{2} \tilde{\epsilon}^T \tilde{\Pi} \tilde{\epsilon}$ という二次形式で書くことができる。ここで $\tilde{\epsilon} \equiv \tilde{s} - g(\tilde{\mu})$ は、現時点の一般化感覚状態 $\tilde{s}$ と、これに関する予測 $g(\tilde{\mu})$ の間に生じる誤差ベクトルを表す。現時点の認識状態 $q(\tilde{x})$ を平均 $\tilde{\mu}$ と分散 $\tilde{\Sigma}$ を持つガウス分布で表している。二次形式の重み $\tilde{\Pi}$ は認識状態の不確定性の影響を受けているが、この影響を $\tilde{\xi} \equiv \tilde{\Pi} \tilde{\epsilon}$ に押し付けている。知覚 $\tilde{s}$ は真の観測モデル $\tilde{s} = \mathbf{g}(\tilde{\mathbf{x}}) + \tilde{\mathbf{z}}$ から得られ、 $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}(a)$ を通してアクションとつながる。

以上の定式化において気をつけるべき点が2つある。第一に、物理環境の真の状態を表す $\mathbf{x}$ と脳内の認識状態 x は一般に異なるモデルに従う別の変数であって対応関係はなく、値だけでなくベクトルの次元数まで異なってもよい。太文字と細文字を注意深く区別する必要がある。図3に真のSSMと認識状態のSSMを並列して示す。アクション a は太文字で書かれた真の物

	外部世界のモデル	脳内の認識状態モデル
状態	$\tilde{\mathbf{x}}$	$\tilde{x}$
時間発展	$\dot{\tilde{\mathbf{x}}} = \mathbf{f}(\tilde{\mathbf{x}}, a) + \tilde{\mathbf{z}}$	$\dot{\tilde{x}} = f(\tilde{x}) + \tilde{z}$
感覚	$\tilde{s} = \mathbf{g}(\tilde{\mathbf{x}}) + \tilde{\mathbf{w}}$	$\tilde{s} = g(\tilde{x}) + \tilde{w}$

図3 世界と認識の状態空間モデル。太字で表現されている変数や関数 $\mathbf{x}, \mathbf{f}(), \mathbf{g}()$, etc. は外部世界モデルを記述しており、認識状態やそのダイナミクスモデルからは直接アクセスできないことに注意。

理状態の側を制御しているということにも注意が要る。誤差 ϵ の定義式（サプライズを定義付けていると言ってもよい）を以下のように書き直しておく誤解の余地がすくなくなるだろう。

$$\begin{aligned}\tilde{\epsilon} &\equiv \tilde{s} - g(\tilde{\mu}) = \mathbf{g}(\tilde{\mathbf{x}}) - g(\tilde{\mu}) \\ &= \mathbf{g}(\tilde{\mathbf{x}}(a)) - g(\tilde{\mu})\end{aligned}\quad (11)$$

したがって $\tilde{\epsilon}_a \equiv \partial \tilde{\epsilon} / \partial a$ の主要項は $\partial \mathbf{g}() / \partial a$ であり、ここに $g()$ は直接関与しない。

第二に、状態は一般化座標系で表わされていることにも注意が必要である。一般化知覚状態 $\tilde{s} = (s, \dot{s}, \ddot{s}, \dots)$ 、一般化状態 $\tilde{\mathbf{x}} = (\mathbf{x}, \dot{\mathbf{x}}, \ddot{\mathbf{x}}, \dots)$ 、一般化認識状態 $\tilde{x} = (x, \dot{x}, \ddot{x}, \dots)$ の三種ともに一般化座標系利用の有無は、数式上ではチルダの有無によって注意深く区別する必要がある。

ブロック行列作用素 D を用いると、この形で定義された一般化状態（たとえば特に $\tilde{\mathbf{x}}$ ）の時間微分を

$$\dot{\tilde{\mathbf{x}}} = D\tilde{\mathbf{x}} \quad (12)$$

で計算することができる。例えば 2 次元平面上の位置座標に対応するような状態変数 $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^T$ を考えるとき、これに対応する一般化座標は $\tilde{\mathbf{x}} = (\mathbf{x}_1, \mathbf{x}_2, \dot{\mathbf{x}}_1, \dot{\mathbf{x}}_2, \dots)^T$ と書くことができ、これに作用素 D をかけることによって $\dot{\tilde{\mathbf{x}}} = D\tilde{\mathbf{x}} = (\dot{\mathbf{x}}_1, \dot{\mathbf{x}}_2, \ddot{\mathbf{x}}_1, \ddot{\mathbf{x}}_2, \dots)^T$ となる。 $D\tilde{\mathbf{x}}$ は $\tilde{\mathbf{x}}$ に対してベクトルの要素を 2 つぶんにずらしたことを意味する。同様に $D^i, i = 0, \pm 1, \pm 2, \dots$ という表現によってベクトル要素を 2, 4, 6 ずつ上や下にずらす操作を表現することができる。ここで存在しない成分にはゼロを埋める。一般化座標系を考えるさいに、回数無制限の高次時間微分に対応してベクトル $\tilde{\mathbf{x}}$ の次元数も無制限になっているように見えるが、数値計算上は有限で止めることに問題は無い。現実世界の物理的対象の運動方程式は高々 2 次や 3 次微分までしか含まず、真の状態空間モデル・脳内認識状態の状態空間モデルともに高階微分まで必要としない。作用素 D を扱う負荷は数式処理・数値演算のうえでも軽く、神経回路上の実装も容易であることが、一般化状態表現の妥当性の根拠のひとつとなっている。

以上の準備のもとで式 (10) の導出は容易である。まず真の状態空間モデルを作用素 D を使って以下のように書き、

$$D\tilde{\mathbf{x}} = \mathbf{f}(\tilde{\mathbf{x}}, a) + \tilde{\mathbf{w}} \quad (13)$$

この演算子の逆 D^{-1} を両辺に適用、

$$\tilde{\mathbf{x}} = D^{-1}\mathbf{f}(\tilde{\mathbf{x}}, a) \quad (14)$$

両辺でアクション a に関する全微分をとる。

$$\dot{\tilde{\mathbf{x}}}_a = D^{-1}\mathbf{f}_{\tilde{\mathbf{x}}}\dot{\tilde{\mathbf{x}}}_a + D^{-1}\mathbf{f}_a \quad (15)$$

右辺の第一項に含まれる $\tilde{\mathbf{x}}_a$ に右辺全体を代入する、

$$\dot{\tilde{\mathbf{x}}}_a = D^{-2}(\mathbf{f}_{\tilde{\mathbf{x}}})^2\tilde{\mathbf{x}}_a + D^{-2}\mathbf{f}_{\tilde{\mathbf{x}}}\mathbf{f}_a + D^{-1}\mathbf{f}_a \quad (16)$$

この操作を繰り返して

$$\dot{\tilde{\mathbf{x}}}_a = \sum_{i=1}^{\infty} D^{-i}(\mathbf{f}_{\tilde{\mathbf{x}}})^{i-1}\mathbf{f}_a \quad (17)$$

以上を用いると、真の観測モデル $\mathbf{g}(\tilde{\mathbf{x}}(a))$ の a による微分が得られる。

$$\tilde{\epsilon}_a = \mathbf{g}_{\tilde{\mathbf{x}}}\dot{\tilde{\mathbf{x}}}_a = \mathbf{g}_{\tilde{\mathbf{x}}} \sum_{i=1}^{\infty} D^{-i}(\mathbf{f}_{\tilde{\mathbf{x}}})^{i-1}\mathbf{f}_a \quad (18)$$

数値的には環境状態変数 $\mathbf{x}$ を M 次元縦ベクトルとし、一般化座標系への拡張 $\tilde{\mathbf{x}}$ を MK 次元縦ベクトルとすると、 $\mathbf{x}_a$ や $\mathbf{f}_a$ は MK 次元縦ベクトルとなる。また $\mathbf{g}_{\tilde{\mathbf{x}}}$ や $\mathbf{f}_{\tilde{\mathbf{x}}}$ は $MK \times MK$ 行列となり、演算子 D は同じ大きさのスパースな $MK \times MK$ 行列で表わされるものになる。なおこのとき、いかなる行列 A に対しても $D^{-i}A$ は $i > K$ においてゼロ行列になるため一般に無限和で書かれている箇所は事実上は有限和として計算されることにも注意しておく。

4.2 AIF 計算過程の妥当性

前節でみたように AIF の行動更新計算の過程は

$$\dot{a} = -\mathcal{F}_a = -\tilde{\epsilon}_a^T \tilde{\xi} \quad (19)$$

で表すことができる。これは神経計算として妥当なモデルと言えるだろうか。

$\tilde{\epsilon}$ は環境モデル（太字）に基づく観測と認識状態モデル（細字）に基づく予測の差分であり、Action and Behaviour 論文¹⁾の定式化（および対応する MATLAB 実装）において、アクション a は認識状態モデル $g()$ ではなく、真の環境状態のモデル $\mathbf{g}()$ のなかに含まれている。つまり AIF は真の環境状態 $\mathbf{x}$ や真の状態空間モデル $\mathbf{f}()$ 、 $\mathbf{g}()$ および、また式 10 のように真のモデルに関する微分 $\mathbf{g}_{\tilde{\mathbf{x}}}$ 、 $\mathbf{f}_{\tilde{\mathbf{x}}}$ 、 $\mathbf{f}_a$ が天下り的に与えられていることを想定し、これらを用いてアクション a を計算する機構となっている。

しかしここで「真の環境」なるものを、認識モデルの部品として使うのは筋が通らない。例えば Friston らの AIF モデルのシミュレーション^{1,2)} アクション生成に用いる「予測誤差のアクション微分」 $\tilde{\epsilon}_a$ は、シミュレーション実験環境として設定された状態空間モデルを真値として直接用いて導出していたが、これは「真

の環境」を認識モデルの部品として使っていることになる。

では「予測誤差のアクション微分」 $\tilde{\epsilon}_a$ をどのように与えれば、この無理筋を解消できるだろうか。

第一に、真の環境状態モデル $\mathbf{f}()$, $\mathbf{g}()$ を天下り的に与えるのではなく、近似モデルを用いる方法を考えよう。近似モデルを学習によって得ることができれば、これを用いて $\tilde{\epsilon}_a$ を導くことができる。具体的には、例えば細字で書かれる認識モデル $f()$, $g()$ を真の環境モデル $\mathbf{f}()$, $\mathbf{g}()$ の代わりに用いることとし、事前確率の影響なしに適用する。通常の認識状態 $q(\tilde{x})$ が無情報でない階層的な事前確率のもとでバイアス入りになっているのに対して、事前確率の影響なしの推定状態 $\mathbf{x}$ を別途与えれば、これらの差分から $\tilde{\epsilon}$ および $\tilde{\epsilon}_a$ の近似値をそれぞれ得ることができて AIF が成立する。しかし、 $\mathbf{f}()$, $\mathbf{g}()$ が階層を持ったり高次元入出力を持つなど複雑なものである場合に、 $\tilde{\epsilon}_a$ を式 10 から求める過程は複雑であり、神経計算機構にとって妥当とは考え難い。

そこで第二案として、単純なアクション a 毎に単純な状態空間モデル $\mathbf{f}(\mathbf{x}, a)$, $\mathbf{g}(\mathbf{x})$ を天下り的に与えることを考える。ここで単純なアクションとはたとえば特定の筋肉ひとつの出力に対応するスカラー値を想定する。この程度にまで単純なアクションだけを考える場合、これに対応する環境状態 $\mathbf{x}$ やこれに基づく感覚 $\mathbf{g}(\mathbf{x})$ の内容として、 a に対応するごく限られた触覚や体性感覚のみを含むように設定することが可能である。そうすれば $\tilde{\epsilon}_a$ は単純な関数となるし、この近似が悪くても大きな問題は生じない。なぜならば、AIF において軌道計画や状態空間モデリングはアクションを含まない認識状態を対象としており、軌道のズレ補正のみが $\tilde{\epsilon}_a$ の役割となるからである。誤差のアクション微分 $\tilde{\epsilon}_a$ に含まれ得るバイアス・バリエーションともに、認識状態の状態空間モデル学習に容易に吸収され得る。したがって、状態空間モデルからの導出も考えずに $\tilde{\epsilon}_a$ が天下り的に与えられていてもよい。この条件下でならば、脊髄における小さな神経回路による実装も難しくない。Friston らは第二案の状況を想定しているものと考えられるが、この点に限っては前提条件のもとでの最適性を求めるあまり現実性を欠いた解釈が許される定式化になっている。第二案の状況を明に想定したより適切な単純化近似が必要であろう。

5. ま と め

FEP および AIF の理論的基盤であるところの、変分推論 (2 章) と確率的推論としての制御 (3 章) について概説した。変分推論では事後確率 $q()$ に対する

恣意的な制約、確率的推論としての制御では目的関数 $J()$ の恣意的な設定のもとで、原理に基づいたオンライン最適化過程の形で認識状態ダイナミクスモデルが導出できることを示した。これに対し FEP および AIF 特有の差分は、 $q()$ および $J()$ の恣意的な設定に対して、動物の脳神経計算特有の制約とその妥当性を議論の俎上に載せている点にあることを指摘した。とくに経験に基づく階層的知覚モデルの学習が、AIF において制御目的関数 $J()$ の学習による獲得と同一視される点は、脳型人工知能の設計において示唆するところが大きい。4 章で示したように AIF 計算過程の実装を詳細に見てゆくと無理筋を含む箇所も残るが、致命的な困難とまでは思われない。動物における AIF の設計原理は計算知能設計の基盤としておおいに参考となつてゆくだろう。

謝辞

本研究は独立行政法人新エネルギー・産業技術総合発機構 (NEDO 技術開発機構) および文部科学省ポスト「京」萌芽的課題アプリケーション開発の研究助成を受けて行われた。

参 考 文 献

- 1) Friston, K., Daunizeau, J., Kilner, J., Kiebel, S. (2010): Action and behavior: a free-energy formulation, *Biol. Cybern.*, Vol.102, No.3, pp.227–260.
- 2) Friston, K., Mattout, J., Kilner, J. (2011): Action understanding and active inference, *Biol. Cybern.*, Vol.104, pp.137–160.
- 3) Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., Pezzulo, G. (2015): Active inference and epistemic value, *Cognitive Neuroscience*, Vol.6, No.4, pp.187–214.
- 4) Tran, D., Kucukelbir, A., Dieng, A.B., Rudolph, M., Liang, D., Blei, D.M. (2016): Edward: A library for probabilistic modeling, inference, and criticism, *arXiv preprint arXiv*, 1610.09787.
- 5) Zhang, C., Butepage, J., Kjellstrom, H., Mandt, S. (2017): Advances in variational inference, *arXiv preprint arXiv*, 1711.05597.
- 6) Sato, M. (2001): On-line model selection based on the variational bayes, *Neural Computation*, Vol.13, No.7, pp.1649–1681.
- 7) Friston, K., Daunizeau, J., Kiebel, S. (2009): Reinforcement learning or active inference?, *PLoS One*, Vol.4, No.7, e6421.
- 8) Todorov, E., Li, W. (2005): A generalized iterative lqg method for locally-optimal feed-

- back control of constrained nonlinear stochastic systems. In IEEE Proceedings of the American Control Conference, pp.300–306.
- 9) Stahl, D., Hauth, J. (2011): Pf-mpc: Particle filtermodel predictive control, Systems and Control Letters, Vol.60, No.8, pp.632–643.
 - 10) Murakami, Y., Koyama, M., Oba, S., Kuroda, S., Ishii, S. (2017): Model-based control of the temporal patterns of intracellular signaling in silico, Biophys. Physicobiol., Vol.14, pp.29–40.
 - 11) Todorov, E. (2008): General duality between optimal control and estimation. In Conference on Decision and Control (CDC).
 - 12) Kappen, H. (2011): Optimal control theory and the linear bellman equation, Inference and Learning in Dynamic Models, pp.363–387.
 - 13) Kappen, H.J., Gomez, V., Opper, M. (2012): Optimal control as a graphical model inference problem, Machine Learning, Vol.87, No.2, pp.159–182.
 - 14) Toussaint, M. (2009): Robot trajectory optimization using approximate inference. In International Conference on Machine Learning (ICML).
 - 15) Levin, S. (2018): Reinforcement learning and control as probabilistic inference: Tutorial and review. arXiv preprint arXiv, 1805.00909.
 - 16) Friston, K., Trujillo-Barreto, N., Daunizeau, J. D. (2008): A variational treatment of dynamic systems, Neuroimage, Vol.41, No.3, pp.849–885.
 - 17) 乾 敏郎 (2018) : 知覚・認知・運動・感情・意思決定をつなぐ自由エネルギー原理, 日本神経回路学会誌, Vol.25, No.3, pp.123–134.
 - 18) 磯村拓哉 (2018) : 自由エネルギー原理の解説 : 知覚・行動・他者の思考の推論, 日本神経回路学会誌, Vol.25, No.3, pp.71–85.