自由エネルギー原理
～知覚と行動と学習の統一原理～

理論

1. イントロダクション

次に示す2つの画像1, 2を見てください．図1の突起は左から順に凸凹凸の形状に見えます．一方，図2の突起は左から順に凹凸凹の形状に見えます．

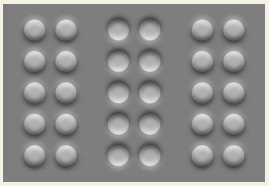

図1

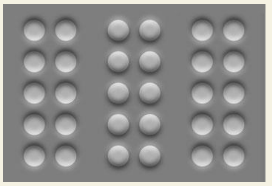

図2

出典：http://www.ritsumei.ac.jp/~akitaoka/

実は右の画像は左の画像を180°回転させたものです．これは人の脳が「光源は上にある」，「光は直進する」といった**先見的な知識に基づいて認識**した結果として生じる錯視と呼ばれる現象の一つです．

それでは，人はどのように先見的な知識と呼ばれるものを獲得しているのでしょうか．現在は，これらの知識獲得を教師なし学習によって行っているという説が有力です．先の錯視の例に倣うと，その大まかな流れは次のようになります．

1. 人は視覚によって得られた網膜画像から外部の三次元形状を**推定**します．
2. 今度は逆に，1で得られた三次元形状をもとに，外部環境のシミュレーションのようなことを行うことで，網膜に映る画像を**予測**します．
3. 2で予測した網膜画像と実際の網膜画像を比較し，その誤差をもとに脳内のネットワークを更新します．この行為が**学習**に該当します．

これらの手順を繰り返すことで，人は幼少期に凹凸当てゲームのようなものをしなくても，自然とこれらの技術を身につけることができます．

自由エネルギー原理（FEP）とは，これら一連の複雑な流れを統一的に扱うための原理です．

2. ベイズの定理

それでは早速FEPの説明に入りたいのですが，そのためにはベイズの定理について説明をする必要があります．

ある事象Yのもとで事象Xが起きる条件付き確率を

$$p(X \mid Y) := \frac{p(X,Y)}{p(Y)}$$

で定義します．ただし，$p(Y) \neq 0$とします．これより

$$p(X,Y) = p(X \mid Y)p(Y)$$

を得ます．また，同様の手順により

$$p(X,Y) = p(Y \mid X)p(X)$$

を得ます．したがって，初めの式は

$$p(X \mid Y) := \frac{p(X,Y)}{p(Y)} = \frac{p(Y \mid X)p(X)}{p(Y)}$$

と表すことができます．最後の式を「ベイズの定理」といいます．ここでXを原因，Yを結果とすると，左辺は「結果から原因」への過程であるのに対し，右辺は「原因から結果」へと過程を反転できます．

3. FEP

2章の最後に原因と結果の話をしました．ここで1章の錯視の話と絡めるためにXを外部環境，Yを網膜画像とします．網膜画像から外部環境を推定するには，2章で導出したベイズの定理を用いて

$$p(X \mid Y) = \frac{p(Y \mid X)p(X)}{p(Y)}$$

を計算すれば良いです．しかしながら，右辺の分母にある$p(Y)$を厳密に計算することは非常に難しいことが一般に知られています．人の脳も当然ながら有限の演算能力しか持たないため，必ずしもこの計算が実行できるとは限りません．

そこで人は自分の演算力で処理できる範囲で，左辺の確率分布$p(X \mid Y)$を近似していると考えられます．以降，この近似した関数を$q(X)$とします．関数を近似するとなると，それがどの程度良い近似になっているかを知る必要があります．確率分布間の"近さ"のようなものを評価するものの一つにKullback-Leibler divergence（KLD）と呼ばれる量が存在します．

2つの確率分布P, Qに対して

$$\mathrm{D_{KL}}(P \parallel Q) = \sum_x p(x) \ln \frac{p(x)}{q(x)}$$

で定義される量をPとQのKLDといいます．KLDは次の2つの特徴を持ちます．

1. 常に非負値を取る．
2. $\mathrm{D_{KL}}(P \parallel Q) = 0 \Leftrightarrow P = Q$

よって，KLDが小さければ小さいほど二つの確率分布は似ていると解釈することができます．

それでは元の話に戻って近似の良さを評価しましょう．

$$\mathrm{D_{KL}}\big(q(x) \parallel p(x \mid y)\big) = \sum_x q(x) \ln \frac{q(x)}{p(x \mid y)}$$

この右辺を変形すると次の形にまで変形できます．

$$\mathrm{D_{KL}}\big(q(x) \parallel p(x \mid y)\big) = \mathrm{D_{KL}}(q(x) \parallel p(x, y)) + \ln p(y)$$

この右辺第一項をFとおくと，KLDの性質から

$$\begin{aligned} F &= \mathrm{D_{KL}}\big(q(x) \parallel p(x \mid y)\big) - \ln p(y) \\ &\geq - \ln p(y) \end{aligned}$$

となるので，Fを最小化すると$p(X \mid Y)$と$q(X)$が完全に一致することがわかります．このFを自由エネルギーと呼びます．

最後にFEPの主張内容をまとめます．FEPは「人は1章で述べたような知覚・行動・学習を自由エネルギーFという単一変数を最小化することで実現している」という事を主張する原理です．

※簡単のため，ここでは「人は～」としましたがFEP自体は人以外のより一般の対象についても扱うことができます．

4. POMDP

それでは，実際にモデルを設計して，人が行動を選択する過程を見てみましょう．ここでは部分観測マルコフ決定過程（POMDP）と呼ばれる過程を考えます．
説明のため，いくつかの記号を導入します．

- x：人が知りたい情報（**状態**）
- s：人が直接観測できる情報（**感覚入力**）
- a：人が取る**行動**

これらの記号を用いて，POMDPの流れを図3に示します．

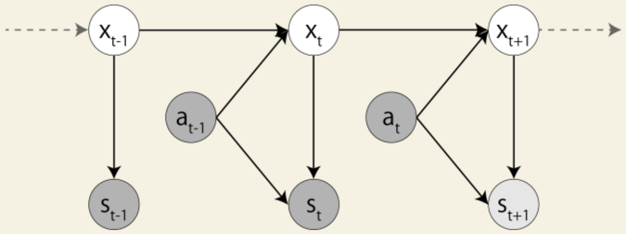

図3

各記号の右下についている添字は時刻を表しています．各時刻での流れは次のようになります．

1. 観測値 s_t と前の時刻の行動 a_{t-1} から状態 x_t を推定します． （この推定した状態を**信念**と呼びます）
2. その信念から何らかの**方策**に基づいて行動 a_t を決定します．
3. a_t により状態は状態遷移確率 $p_T(x_{t+1}\,|\,x_t,\,a_t)$にしたがって状態遷移します．
4. その結果，時刻 t+1 での観測値 s_{t+1} は観測確率関数 $p_s(s_{t+1}\,|\,x_{t+1},\,a_t)$ により観測されます．

以上の4ステップを繰り返していきます．

ここで，人が直接知ることができる情報としては， s の感覚入力だけでなく， a の行動も自分が選択したことなので観測可能であることに注意します．

そこで，次のような観測量の履歴

$$h_t = \{a_0, s_1, a_1, \ldots, a_{t-1}, s_t\}$$

を考えることができて，今後は時刻 t までの観測量 h_t を基に状態 x_tを推定することにします．すなわち，信念を次のように定義します．

$$b(x\,|\,h_t) = \Pr[X_t = x\,|\,H_t = h_t]$$

表記の煩雑さを減らすため，以降これを $b_t(x)$ と略記することにします．

こうして定めた $b_t(x)$ について，少々複雑な計算をすることで次のような関係式が導かれます．

$$b_{t+1}(x') = \frac{p_s(s_{t+1}\,|\,a_t,\ x')\sum_x p_T(x'\,|\,x, a_t)\,b_t(x)}{\sum_{x'} p_s(s_{t+1}\,|\,a_t,\ x')\sum_x p_T(x'\,|\,x, a_t)\,b_t(x)}$$

この式に現れる変数をよく観察すると，信念 $b_t(x)$は時刻 t での行動 a_t と時刻 t+1 での感覚入力 s_{t+1} があれば更新できることがわかります．

※統計学用語では，これを「 $\{b_t, a_t, s_{t+1}\}$ は h_{t+1} の十分統計量である」と言うことがあります．

5. Active Inference

さて，4章で設計したモデルを使用して自由エネルギー Fを計算するとどうなるでしょうか．

$$F = \mathrm{D_{KL}}\big(q(x) \parallel p(x,h)\big) = \sum_x q(x) \ln \frac{q(x)}{p(x,h)}$$

$$= \sum_x q(x) \{\ln q(x) - \ln p(x,h)\}$$

$$= \sum_x q(x) \left\{ \ln q(x) - \ln \left[p_s(s_{t+1}|x,a_t) \sum_{\tilde{x}} p_T(x|\tilde{x},a_t) b_t(x) \right] \right\}$$

最後の式を見ると，時刻 t での自由エネルギー Fは行動 a_t と感覚入力 s_{t+1} があれば計算できることがわかります．

しかし，これには大きな問題があります．時刻 t の段階では，当然ながら未来の時刻 t+1 の感覚入力 s_{t+1} は分かりません．よって，このままではFを計算することはできません．

そこで， s_{t+1} について期待値をとることを考え，これを期待自由エネルギーと呼び， G で表記します．

$$G(a_t) = \mathbb{E}_{s_{t+1}}[F(a_t, s_{t+1})]$$

4章の途中で人は信念から何らかの**方策**に基づいて行動を決定すると述べましたが，FEPの枠組みにおいては自由エネルギー Fの最小化が目標であるので，ここでの方策とは期待自由エネルギー Gを最小化するような行動を選択することになります．このような行動選択を行う枠組みを**能動推論**（Active Inference ）と呼びます．

6. まとめ

- 知覚：現在の外界の状態の推定
- 行動：未来の外界の状態の予測
- 学習：予測と実際のズレによる推定モデルの更新

FEPは上に示した「知覚」と「行動」と「学習」の統一原理であって，人は自由エネルギーという単一変数の最小化によってこれを実現していると主張しています．また，実際にPOMDPのようなモデルを考えると直接自由エネルギーを計算することはできず，代わりに期待自由エネルギーを計算することで，これを最小化するように行動を決定するという枠組みを能動推論といいます．

参考文献

[1] 山たー. “掛け算のニューラルネットワークとベイズ推定”, 知識のサラダボウル. https://omedstu.jimdofree.com/2018/03/21/掛け算のニューラルネットワークとベイズ推定/, (参照2020-9-12).

[2] Friston, K. J. A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 815–836 (2005)

[3] 森村哲郎. 強化学習. 東京, 講談社, 2019, 320p.

[4] 吉田正俊. “自由エネルギー原理入門” http://pooneil.sakura.ne.jp/archives/permalink/001663.php,（参照2020-9-12).