

BCSC 9기 | Term 2

Probabilistic Computational Neuroscience

Active Inference: The Free Energy Principle in Mind, Brain, and Behavior

1조 보고서

윤시원	Ch 3, Ch 4, Additional Study
서효정	Ch 7, Ch 8, Additional Study
이서영	Ch 6, Ch 9
이서현	Ch 2, Ch 5

목차

Chapter 2. The Low Road to Active Inference

추론으로서의 지각, 변분 자유 에너지, 행동으로서의 추론, 기대 자유 에너지

Chapter 3. The High Road to Active Inference

생명체의 존재성, Markov blanket, 자기 입증, Hamilton 원리

Chapter 4. Generative Models

로그 증거 유도, Occam의 면도날, POMDP, Jensen 부등식

Chapter 5. Message Passing and Neurobiology

피질 미세회로, 운동 제어, 기저핵, 신경조절과 정밀도

Chapter 6. Active Inference in Practice

4단계 설계 구조, 이산/연속 모델, 시간적 깊이

Chapter 7. Active Inference in Discrete Time

HMM, epistemic value, LFP 신호, 계층적 추론

Chapter 8. Active Inference in Continuous Time

연속 상태 공간, 동역학계, Lorenz 시스템, 혼합 생성 모델

Chapter 9. Model-Based Data Analysis

Meta-Bayesian 추론, 계산적 표현형, 모델 비교 및 검증

Additional Study

현대 AI와 Active Inference 비교 / 논문 리딩

소개

본 보고서는 Thomas Parr, Giovanni Pezzulo, Karl J. Friston의 저서 Active Inference: The Free Energy Principle in Mind, Brain, and Behavior의 핵심 내용을 정리한 것이다.

Active Inference는 자유 에너지 원리(Free Energy Principle)에 기반하여, 생명체가 감각 입력의 놀라움(surprisal)을 최소화하는 방향으로 지각하고 행동한다는 통합적 프레임워크이다. 이 이론은 지각, 행동, 학습, 의사결정을 하나의 수학적 원리—변분 자유 에너지의 최소화—로 설명하며, 신경과학과 인공지능 모두에 깊은 함의를 지닌다.

Chapter 2-5에서는 Active Inference의 이론적 토대를 다룬다. 변분 추론을 통한 지각(Ch. 2), 자유 에너지 원리의 존재론적 근거(Ch. 3), 생성 모델의 수학적 구조(Ch. 4), 그리고 신경생물학적 메시지 전달 메커니즘(Ch. 5)을 순차적으로 살펴본다.

Chapter 6-9에서는 이론의 실제 적용을 탐구한다. Active Inference 모델 설계의 실용적 절차(Ch. 6), 이산 시간(Ch. 7)과 연속 시간(Ch. 8)에서의 구체적 구현, 그리고 모델 기반 데이터 분석과 계산적 표현형 분석(Ch. 9)을 다룬다.

마지막으로 Additional Study에서는 현대 AI(대규모 언어 모델 등)와 Active Inference의 근본적 차이를 세 가지 관점에서 비교 분석하고, Active Inference의 실험적 검증을 다룬 논문을 리뷰한다.

Chapter 2: The Low Road to Active Inference

이서현

Active Inference를 이루는 2가지 축인 High Road와 Low Road 중 Low Road에 대하여 의논하였다. Low Road는 Helmholtz의 ‘무의식적 추론으로서의 지각’이라는 고전적 관점에서 시작하여, 뇌를 통계적 추론 엔진으로 보는 Bayesian brain 가설을 통해 Active Inference를 정당화한다. 우리는 이를 통해 지각뿐만 아니라 행동, 계획, 학습이 어떻게 하나의 통일된 자유 에너지 최소화 원칙으로 통합되는지 탐구하였다.

2.1 지각으로서의 추론 (Perception as Inference)

Active Inference의 관점에서 뇌는 외부 세계의 상태를 추론하고 예측하는 predictive machine이다. 지각은 단순히 감각 입력을 받아들이는 bottom-up 과정이 아니라, 뇌가 가진 내부 model을 통해 감각의 원인을 추측하는 구성적인 top-down 과정이다.

뇌는 관찰 가능한 감각 데이터(y)와 이를 유발하는 숨겨진 상태(x) 사이의 관계를 확률적으로 나타내는 generative model $P(y, x)$ 을 가지고 있다. 새로운 감각 정보가 들어오면 뇌는 Bayes' theorem을 사용하여 prior를 posterior로 업데이트한다.

$$P(x|y) = \frac{P(y|x) P(x)}{P(y)} \quad (2.1)$$

2.2 변분 자유 에너지 (Variational Free Energy)

현실적인 인지 시스템에서 Bayes' theorem의 분모인 model evidence ($P(y)$)를 직접 계산하는 것은 매우 어렵다(intractable). 따라서 뇌는 실제 posterior를 직접 구하는 대신, 다루기 쉬운 variational distribution $Q(x)$ 를 도입하여 이를 실제 분포에 근사시킨다. 이때 최소화하는 지표가 variational free energy (F)이다.

Figure 2.1에서 보듯, free energy (F)는 항상 감각 데이터에 대한 surprisal ($-\ln P(y)$)보다 크거나 같다. 따라서 F 를 최소화하는 것은 간접적으로 surprisal을 최소화하고 model evidence를 최대화하는 것과 동치이다.

$$F[Q, y] = D_{\text{KL}}[Q(x) \| P(x|y)] - \ln P(y) \quad (2.2)$$

여기서 free energy는 complexity와 accuracy의 결합으로 이해될 수 있다. 뇌는 데이터를 정확하게 설명하면서도(accuracy 최대화), 자신의 믿음을 너무 급격하게 수정하지 않는(complexity 최소화) Occam's razor 원칙을 따른다.

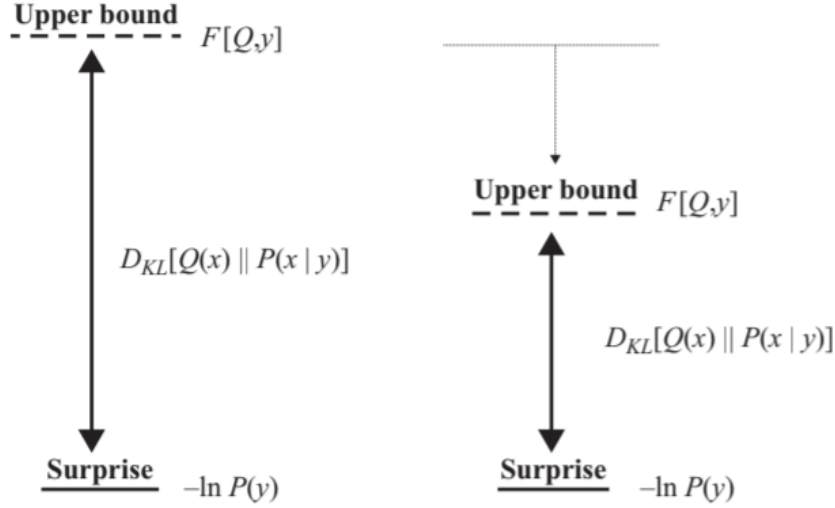


그림 2.1: Variational Free Energy as a Bound on Surprisal

2.3 행동으로서의 추론 (Action as Inference)

지각이 model을 세상에 맞추는 과정이라면, 행동은 세상을 model에 맞추는 과정이다. Active Inference에서 지각과 행동은 동일한 free energy 최소화라는 목적을 공유한다.

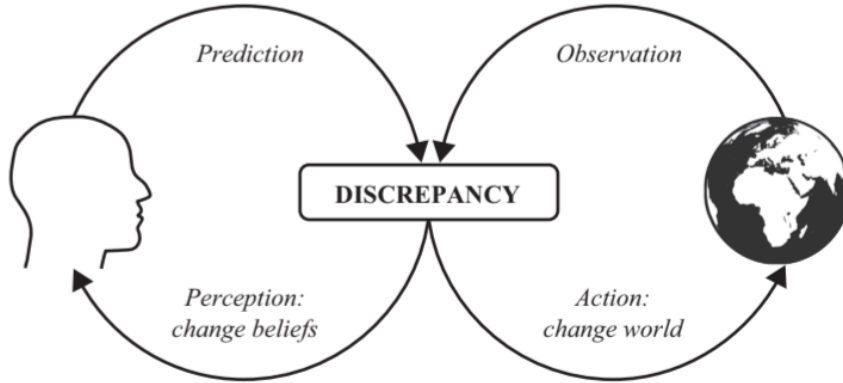


그림 2.2: Perception and Action

Figure 2.2은 free energy를 줄이는 두 가지 경로를 보여준다. 지각(perception)은 믿음(Q)을 수정하여 감각 데이터(y)와의 불일치를 줄인다. 행동(action)은 환경을 변화시켜 자신의 기대에 부합하는 새로운 감각 데이터를 sampling한다.

2.4 기대 자유 에너지와 계획 (Expected Free Energy & Planning)

Free energy가 과거와 현재의 데이터를 다룬다면, 미래를 계획하기 위해서는 미래에 발생할 사건에 대한 expected free energy (G)를 최소화해야 한다. 계획(planning)은 미래에 수행할 policy (π) 중 가장 낮은 expected free energy를 가질 것으로 예측되는 policy를 추론하는 과정이다.

Expected free energy는 두 가지 가치를 동시에 최적화함으로써 exploration vs exploitation의 딜레마를 해결한다. Pragmatic value는 자신의 preference나 목표에 맞는 결과를 얻으려는 성향(risk 최소화)이다. Epistemic value는 세상에 대한 불확실성을 해소하고 정보를 얻으려는 성향

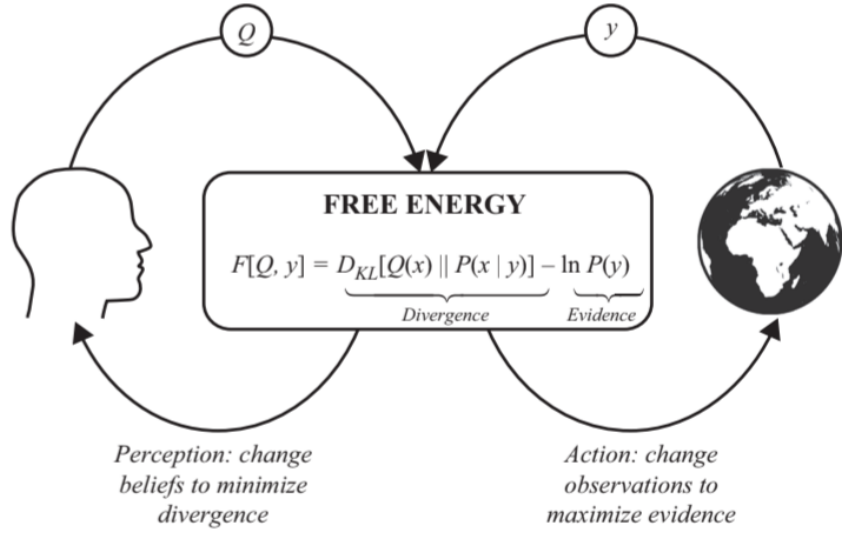


그림 2.3: Components of Expected Free Energy

(ambiguity 최소화)이다.

$$G(\pi) = \underbrace{\text{Ambiguity}}_{\text{epistemic value}} + \underbrace{\text{Risk}}_{\text{pragmatic value}} \quad (2.3)$$

2.5 요약 및 결론

Low Road를 통해 우리는 Active Inference가 단순한 지각 이론을 넘어, 뇌가 어떻게 free energy라는 단일 지표를 통해 지각, 행동, 계획을 통합적으로 제어하는지 확인하였다. 이는 생명체가 자신의 내부 model을 세상과 일치시키려는 끊임없는 self-evidencing의 과정으로 해석될 수 있다.

Chapter 3: The High Road to Active Inference

윤시원

Active Inference를 이루는 2가지 축인 High Road와 Low Road 중 High Road에 대하여 의논하였다. High Road를 통해 우리는 Active Inference를 통해 뇌(더 나아가 생명체)는 어떤 역할을 하며, 왜 하는지에 대한 이유를 탐구하였다.

3.1 생명체의 존재성

생명체에게 있어 존재성 유지는 중요한 목적이다. 생명체는 자신의 존재성을 유지하기 위해 surprise를 최소화하고, 결과적으로 생존한다. 예컨대, 물 밖으로 나온 물고기는 이 상황에서 surprise가 증가하고, 다시 surprise를 최소화하기 위해 물 속으로 들어가 생존한다.

3.2 Markov Blankets

Markov blanket은 생명체의 존재성을 수식적으로 modeling할 수 있는 주요한 도구이다. Blanket b 가 주어졌을 때, 내부 상태 μ 와 외부 상태 x 가 서로 조건부 독립이 성립하는 경계를 Markov blanket이라고 부르며, Active Inference에서는 Figure 3.1와 같이 Markov blanket을 이용하여 modeling한다. 해당 그림에서는 생명체를 나타내는 변수 μ 와 생명체가 살아가는 세상을 나타내는 x 의 dynamic한 상호작용을 modeling하였다. 둘은 blanket인 행동과 감각만을 통해 소통하는 대상으로, 독립성을 통해 생명체의 존재성을 표현할 수 있다.

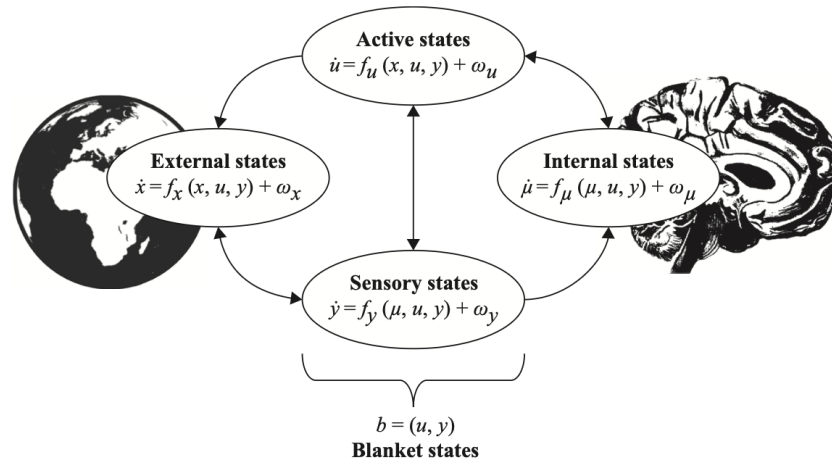


그림 3.1: A dynamic Markov blanket

3.3 Self-Evidencing

Active Inference의 관점에서 뇌는 surprise를 최소화하는 대상이지만, 생명체의 유지를 위해서는 무작정 surprise를 최소화하는 것을 피해야 한다. 예컨대 바닷물 속 물고기가 있는 상황보다 물 만이 존재하는 상황이 자연스러워, 물고기가 무작정 surprise를 최소화하는 소산적 과정을 거치면 존재성을 유지할 수 없다. 때문에 생명체는 생존에 적합한 상태를 가정하는 — 즉, 자신의 존재 자체가 자연스럽다는 — prior 값인 optimism bias (e.g., 체온에 대한 prior 값 $\rightarrow$ 차가운 물 속에서도 자신의 체온을 유지하는 원동력)를 지닌다. 존재를 유지하기 위해서는 감각을 설명할 수 있어야 하는데, 다음을 예로 들 수 있다: 물고기가 바닷물 속에서 차갑게 느껴도 surprise를 줄이지 아니하고, 바닷물에서 차갑게 느껴지는 것은 예측 및 설명 가능하기에 자신의 존재성을 유지한다. 때문에 생명체는 자신의 설명이 옳음을 증명하고자 행동하는데, 이를 self-evidence라고 부른다.

3.4 Hamiltonian & Entropy

Active Inference에서 High Road는 물리 현상을 통해 생명체가 surprise를 최소화하는 것을 정당화함으로써 ‘왜’라는 질문에 답한다. 통계열역학에서 Hamiltonian은 system의 상태가 가지는 energy function을 의미하며, 이는 정보이론의 surprisal ($\mathfrak{S}(y) = -\ln \Pr(y)$)에 대응된다.

Energy function을 의미하는 Hamiltonian은 시간을 매개하여 열역학 제2법칙에 따라 시간이 지남에 따라 증가한다. Active Inference는 surprise를 최소화하기 위해 free energy를 최소화하는 과정이 이와 동치임을 수식으로 보인다. Equation 3.1에서 H 는 (Shannon) entropy를, $\mathfrak{S}(y)$ 는 surprise를 의미한다. 여기서 entropy는 Hamiltonian의 기댓값이다.

$$H[\Pr(y)] = \mathbb{E}_{\Pr(y)}[\mathfrak{S}(y)] = -\mathbb{E}_{\Pr(y)}[\log \Pr(y)] \quad (3.1)$$

Active Inference에서는 생명체를 modeling하기 위해 Markov blanket을 사용하고, 이는 생명체의 존재성을 정당화한다. 생명체와 외부 세상을 분리하는 Markov blanket은 장기적으로 entropy가 증가하는 절대 법칙 속에서 생명체가 self-organization을 통해 단기적으로 entropy를 감소시키는 것을 정당화한다.

3.5 Hierarchical Model

복잡한 인지 활동을 위해서는 다양한 시간 척도에서 추론하는 능력을 요하고, 계층적 model은 이를 만족한다. 계층적 model이 복잡한 인지 활동(여기서는 글 읽기)을 하는 예는 다음과 같다: 상위 계층은 천천히 변하는 글의 맥락을 이해하고, 하위 계층은 빠르게 변하는 단어 및 글자를 이해한다.

3.6 Alternative Futures

Model은 복잡도를 증진하기 위해 temporal depth를 깊게 할 수 있다. 이는 현재와 멀리 떨어진 시점까지 추론할 수 있음을 의미하는데, policies 또는 plans와 직결된다. 각각의 생명체는 서로 다른 model 복잡도 및 temporal depth를 지니고 있다.

Chapter 4: The Generative Models of Active Inference

윤시원

Active Inference에서 우리의 뇌는 generative model이다. 이러한 관점에서 우리의 뇌는 2가지 시간 축에 기반하여 분류 가능하다: 1) 이산 시간 (e.g., 결정을 내리는 상황), 2) 연속 시간 (e.g., 생명체의 몸을 움직이는 상황). 본 chapter에서는 기본적인 generative model의 수학적 탐구부터 2가지 시간 축에 대한 기본적인 모델에 대하여 공부하였다. 다만, 분량을 위하여 본 chapter에서는 이산 시간에 대한 model만을 작성하였다.

4.1 Bayes' Theorem

Generative model은 model evidence $\Pr(y)$ 를 높이는 것과 동치이다. 하지만 이는 실질적으로 직접 구하는 데 어려움이 있어, generative model은 이를 최적화 문제로 치환하여 근사한다.

참고: 확률의 곱과 정보의 합

두 확률의 곱은 정보의 합과 관련 있다:

$$\mathfrak{S}(x, y) = \mathfrak{S}(x) + \mathfrak{S}(y) \quad (4.1)$$

이를 Bayes' theorem에 적용하면 prior 정보와 likelihood 정보를 통해 posterior 정보를 얻을 수 있음을 알 수 있다.

$$\mathfrak{S}(x) + \mathfrak{S}(y|x) = \mathfrak{S}(x|y) + \mathfrak{S}(y) \quad (4.2)$$

$\Pr(y)$ 를 구하는 문제를 최적화 문제로 치환하기 위해 Jensen 부등식을 이용한다.

참고: Jensen's Inequality

오목 함수 ϕ 와 random variable X 에 대하여 다음이 만족한다:

$$\mathbb{E}[\phi(X)] \leq \phi(\mathbb{E}[X]) \quad (4.3)$$

$\log$ 함수의 음값은 오목함수이기에 Jensen 부등식을 적용하면 다음을 만족한다: $\mathbb{E}[\log(X)] \leq \log(\mathbb{E}[X])$.

Log evidence $\log \Pr(y)$ 에 대하여 approximate posterior $Q(x)$ 를 가정하고 Jensen 부등식을 적용하면 다음과 같은 결과를 얻을 수 있다:

$$\begin{aligned} \log \Pr(y) &= \log \sum_x \Pr(y, x) = \log \sum_x Q(x) \frac{\Pr(y, x)}{Q(x)} \\ &= \log \mathbb{E}_{Q(x)} \left[\frac{\Pr(y, x)}{Q(x)} \right] \geq \mathbb{E}_{Q(x)} \left[\log \frac{\Pr(y, x)}{Q(x)} \right] =: -F[Q, y] \end{aligned} \quad (4.4)$$

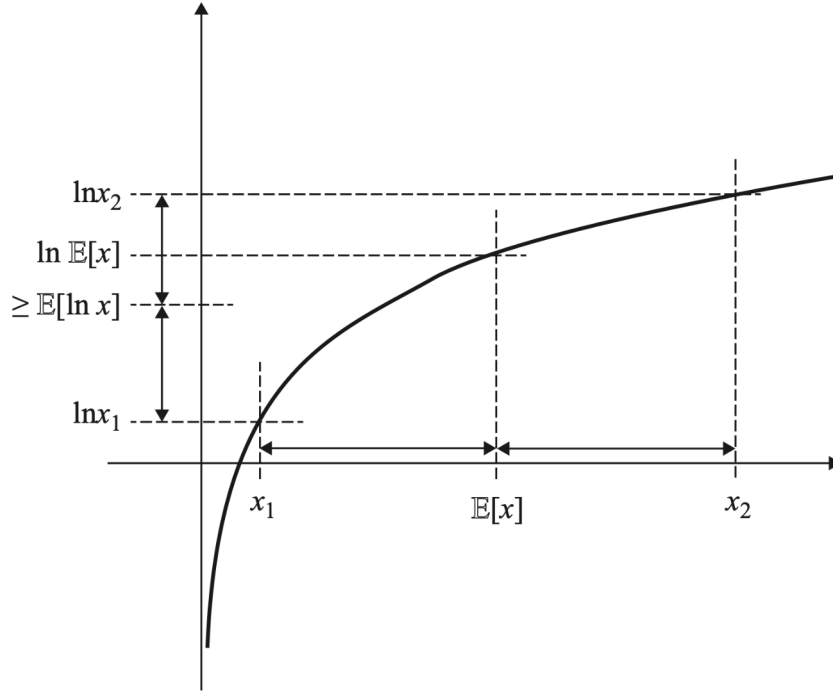


그림 4.1: Logarithmic function providing intuition for Jensen's inequality

이를 정리하면 다음과 같다:

$$F[Q, y] = D_{\text{KL}}[Q(x) \| P(x|y)] - \log \Pr(y) \quad (4.5)$$

Equation ??에서 $F[Q, y]$ 는 variational free energy이며, F 를 최소화함으로써 log evidence 값을 높일 수 있음을 알 수 있다. 앞서 언급한 generative model의 목표와 이는 일치한다. Variational free energy는 complexity ($D_{\text{KL}}[Q \| P]$)와 accuracy ($\mathbb{E}_Q[\log \Pr(y | x)]$)로 분류 가능하다. 이는 데이터를 정확하게 설명하되(accuracy 최대화), 불필요하게 복잡한 가정을 지양하고 기존 믿음을 유지하려는(complexity 최소화) Occam's razor 원칙을 수학적으로 구현한 것이다.

4.2 Generative Models

Free energy를 계산하기 위해 다음 3가지의 요소가 필요하다: data y , 변분 분포 $Q(x)$ 의 집합, 그리고 generative model $\Pr(x, y)$. 이때 각 생명체는 고유한 generative model을 가지기에 서로 다른 출력을 생성한다.

Active Inference에서는 다음 두 형식의 generative model이 사용된다: discrete model, continuous model. Discrete model의 경우 discrete한 시간에서 작동하며 선택과 같은 인지 기능을 해당 model로 modeling할 수 있다. 반면, continuous model의 경우 continuous한 시간에서 작동하며 움직임 등 연속성이 필요한 요소를 modeling한다.

Figure ??의 (a)는 discrete model의 대표적 예인 Partially Observable Markov Decision Process (POMDP)이며, (b)는 continuous model의 대표적인 예이다. 본 chapter에서는 POMDP를 중점으로 소개한다. POMDP는 시간에 따라 진화하는 상태 열(s)로 설명된다. 각 discrete한 시간 step에서 상태는 이전 상태와 agent가 목표한 policy(π)에 의존한다. 최종적으로 상태는 observation(o)을 생성한다.

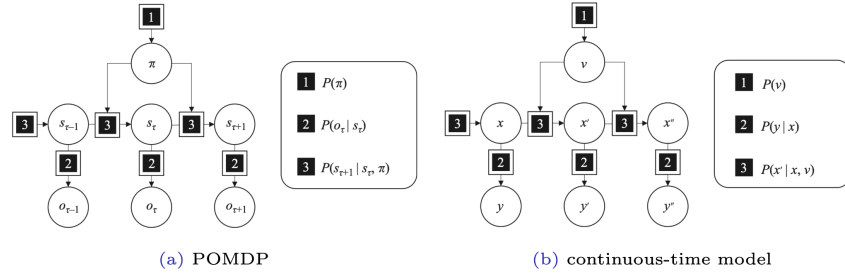


그림 4.2: Two dynamic generative models

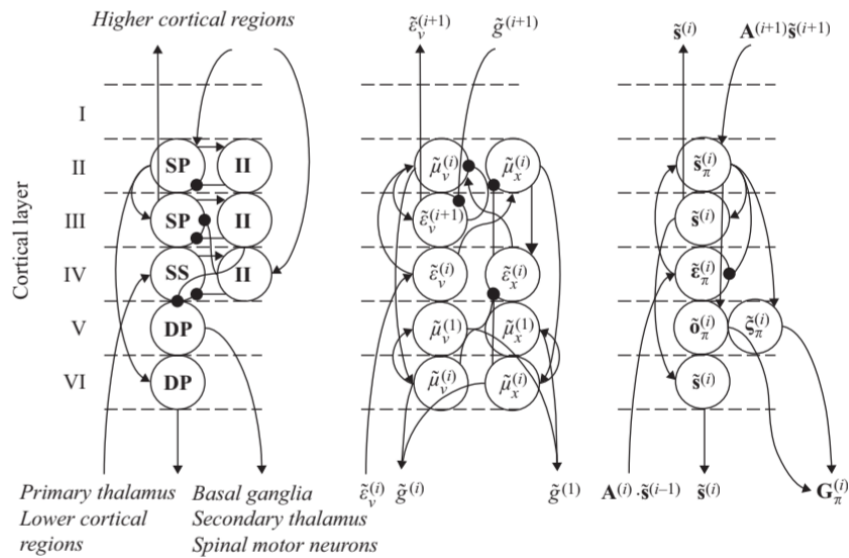
Chapter 5: Message Passing and Neurobiology

이서현

Active Inference의 수학적 원리가 실제 뇌라는 생물학적 기질에서 어떻게 구현되는지 탐구하는 process theory에 대하여 의논하였다. 뇌의 신경 구조와 신경 전달 물질의 역동을 variational message passing 알고리즘과 매핑함으로써, 추상적인 free energy 최소화 원리가 어떻게 실제 신경 세포의 활동으로 변환되는지 분석하였다.

5.1 피질 미세회로와 메시지 전달

뇌의 대뇌피질은 정형화된 6층 구조의 미세회로를 가지며, 이는 계층적 generative model의 message passing 구조를 수학적으로 구현하기에 최적화되어 있다. 신경세포의 활동 전위는 상태에 대한 믿음을 업데이트하는 gradient descent 과정으로, 시냅스의 연결 강도는 model의 매개변수로 해석된다.



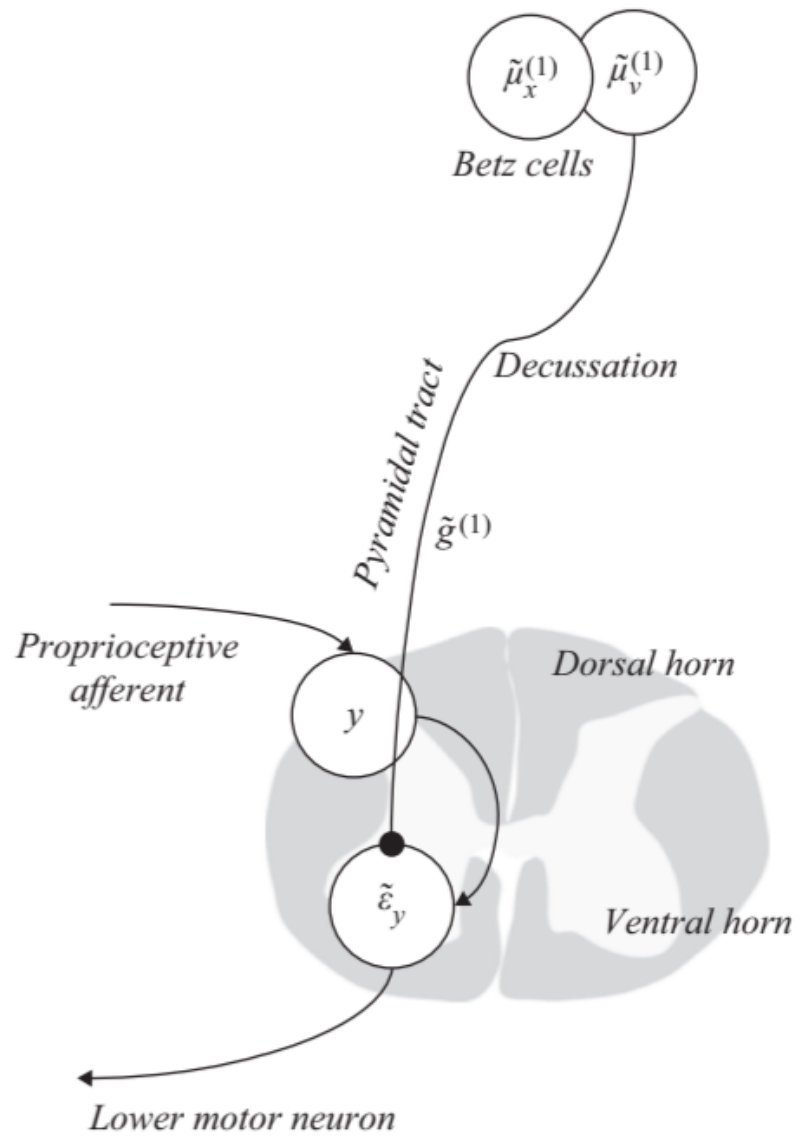


그림 5.2: Neuroanatomy of spinal motor reflexes

움직임으로 발생하는 감각 정보를 일시적으로 무시하는 sensory attenuation 과정이 필수적으로 수반된다.

5.3 기저핵과 정책 선택

의사결정과 미래 계획 수립은 basal ganglia를 통한 policy 선택 과정으로 설명된다. 이는 이산적 시간 model(POMDP)에서의 추론 과정과 대응된다.

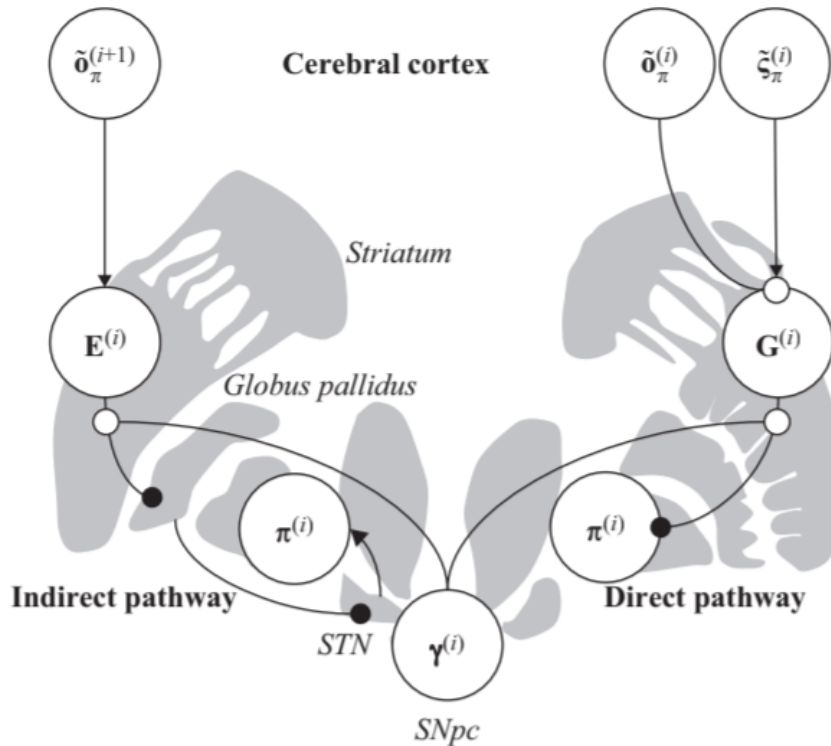


그림 5.3: Basal ganglia and policy selection

Basal ganglia의 직접 및 간접 경로는 다양한 policy들의 expected free energy (G)를 평가하여 특정 행동을 촉진하거나 억제한다. 이때 dopamine은 보상 자체가 아니라, 선택된 policy가 미래의 free energy를 얼마나 잘 최소화할 수 있는지에 대한 precision 혹은 신뢰도를 나타내는 신호로 작용한다.

5.4 신경변조와 정밀도

Generative model에서 각 정보의 신뢰도를 결정하는 precision parameter는 뇌의 주요 신경전달물질 시스템과 매핑된다. 각 물질은 서로 다른 유형의 precision을 조절함으로써 인지 기능을 수행한다.

Acetylcholine (ACh)은 감각 데이터(likelihood)의 precision을 높여 외부 자극에 집중하게 한다. Noradrenaline (NE)은 상태 전이(transitions)의 precision을 조절하여 환경 변화에 대응한다. Dopamine (DA)은 policy의 precision을 조절하여 행동 개시와 자신감을 제어한다.

5.5 주요 수식 정리 (Message Passing)

이산적 시간 model에서 상태에 대한 업데이트 역동은 다음과 같이 표현된다.

$$\dot{\mathbf{s}}_\tau = \ln \mathbf{A} \cdot \mathbf{o}_\tau + \ln \mathbf{B}_\pi \mathbf{s}_{\tau-1} + \ln \mathbf{B}_\pi^\top \mathbf{s}_{\tau+1} - \ln \mathbf{s}_\tau \quad (5.1)$$

연속적 시간 model에서 신경세포의 활동 역동은 variational free energy의 경사를 따라가는 과정이다.

$$\dot{\tilde{\mu}} = D\tilde{\mu} - \frac{\partial F}{\partial \tilde{\mu}} \quad (5.2)$$

Policy 선택에 대한 posterior는 expected free energy (G)와 free energy (F)의 softmax 함수로 결정된다.

$$\pi = \sigma(-G_\pi - F_\pi) \quad (5.3)$$

Chapter 6: Active Inference Design

이서영

6.1 개요

Active Inference는 생물학적 혹은 인공지능 시스템이 환경을 인식하고 행동을 선택하는 과정을 설명하는 이론적 framework이다. 이 접근법에서는 뇌의 실제 계산 과정을 그대로 모사하기보다는, 해결하려는 문제를 설명할 수 있는 generative model을 먼저 설계하고 그 model을 통해 인지와 행동을 설명하는 방식을 사용한다. 이러한 방식은 bottom-up 방식이 아니라 top-down 방식의 model 설계 접근에 해당한다.

즉 Active Inference에서 중요한 것은 특정 algorithm을 적용하는 것이 아니라 어떤 generative model을 설계하느냐이다. Model의 구조에 따라 Active Inference의 형태와 동작 방식이 달라질 수 있다. 이러한 model 설계를 위해 일반적으로 4단계 설계 구조가 사용된다.

6.2 Active Inference 모델 설계 구조

Active Inference model은 다음 네 단계의 설계를 통해 구성된다. 1) System modeling: model링할 시스템의 경계를 정의한다. 2) Form of generative model: model의 수학적 구조를 결정한다. 3) Setting up generative model: model의 변수와 확률 구조를 설정한다. 4) Setting up generative process: 실제 환경 과정을 정의한다. 이 네 단계가 정의되면 perception과 action은 free energy 최소화 원리에 따라 자동적으로 도출된다.

6.3 System Modeling

6.3.1 시스템 경계 정의

Model 설계의 첫 단계는 model링할 시스템의 경계를 정의하는 것이다. 이 과정에서는 무엇을 환경으로 볼 것인지, 무엇을 agent의 내부 상태로 볼 것인지, 환경과 agent가 어떤 방식으로 상호작용하는지의 관계를 정의하기 위해 Markov blanket 개념이 사용된다.

6.3.2 Markov Blanket

Markov blanket은 내부 상태와 외부 상태 사이의 직접적인 연결을 차단하고, 두 영역의 상호작용을 특정 상태들을 통해 매개하는 구조이다. 이 구조는 sensory states(외부 세계가 내부 상태에 영향을 주는 경로)와 active states(내부 상태가 외부 세계에 영향을 주는 경로)의 두 가지 상태를 포함한다.

따라서 시스템은 다음과 같은 구조를 가진다: 환경 $\rightarrow$ sensory states $\rightarrow$ 내부 상태, 그리고 내부 상태 $\rightarrow$ active states $\rightarrow$ 환경. 이 구조를 통해 내부 상태와 외부 상태의 경계가 명확해지며, free energy 최소화가 어떤 시스템에 대해 수행되는지 정의된다. Markov blanket은 반드시 뇌 전체에만 적용되는 것이 아니라 다양한 scale에서 정의될 수 있다.

6.4 Generative Model의 구조 결정

6.4.1 Discrete Model과 Continuous Model

Generative model은 사용되는 변수의 형태에 따라 discrete model 또는 continuous model로 구성될 수 있다.

Discrete model은 상태 범주나 행동 선택과 같이 명확하게 구분되는 변수들을 사용하며, 시간 단계마다 상태 전이 확률로 표현된다. 이러한 방식은 의사결정 문제나 policy 선택을 model링할 때 자주 사용된다.

반면 continuous model은 위치, 속도, 온도, 밝기와 같은 연속적으로 변화하는 물리량을 표현하는 데 사용된다. 이러한 경우 시스템의 변화는 미분 방정식과 같은 형태로 표현된다. 특히 continuous generative model은 predictive coding 구조와 잘 맞는 특징을 가진다.

6.4.2 Shallow Model과 Deep Model

Generative model은 단일 수준 구조를 가질 수도 있고, 여러 수준의 계층 구조를 가질 수도 있다. 예를 들어 언어 처리 과정에서는 음소 → 단어 → 문장과 같은 계층 구조가 존재한다. Model을 설계할 때는 관심 있는 수준에 따라 하위 계산을 model에 포함할 수도 있고, 하위 수준의 처리를 단순히 관측 값으로 처리할 수도 있다. 일반적으로 복잡한 인지 기능을 설명할 때는 계층적 구조를 갖는 deep generative model이 더 적합하다.

6.4.3 Temporal Depth

Generative model 설계에서 중요한 또 하나의 요소는 temporal depth이다. Temporal depth는 agent가 행동을 선택할 때 얼마나 먼 미래까지 고려하는지를 의미한다. Temporal depth가 없는 경우 agent는 현재 관측만을 기반으로 행동을 선택한다. 반면 temporal depth가 존재하면 미래 결과를 고려하여 행동을 계획할 수 있다. Active Inference에서 planning은 특정 행동 policy에 대해 expected free energy를 계산하고 그 값이 가장 작은 policy를 선택하는 방식으로 이루어진다. 따라서 미래 결과를 고려한 행동 선택을 위해서는 temporal depth가 필요하다.

6.5 Generative Model 구성

6.5.1 모델 변수 설정

세 번째 단계에서는 generative model의 구체적인 구성 요소를 정의한다. 여기서는 hidden states, observations, actions, priors, likelihood의 요소를 정의한다. 즉 환경이 어떤 구조로 데이터를 생성하는지에 대한 확률적 model을 정의하는 과정이다.

6.5.2 Sensory Modalities

Active Inference model에서는 다양한 감각 정보를 동일한 방식으로 처리한다. 감각 정보는 다음 세 가지 유형으로 구분된다: exteroceptive sensations (외부 세계에 대한 감각 정보, e.g., 시각, 청각), proprioceptive sensations (신체의 위치와 움직임에 대한 감각), interoceptive sensations (신체 내부 상태에 대한 감각, e.g., 심박수). 이러한 감각 정보들은 모두 generative model을 통해 해석된다.

6.5.3 Prior와 Likelihood

Generative model을 구성할 때 핵심적인 요소는 prior와 likelihood이다. Prior는 hidden state에 대한 사전 믿음을 의미하며, likelihood는 특정 hidden state에서 특정 observation이 발생할 확률을 의미한다. Agent는 이러한 확률 구조를 기반으로 Bayesian inference를 수행하여 환경 상태에 대한 믿음을 업데이트한다.

6.6 Inference와 Learning

Active Inference에서는 inference와 learning이 구분된다. Inference는 현재 관측을 기반으로 hidden state를 업데이트하는 과정이며 비교적 빠르게 이루어진다. 반면 learning은 generative model의 parameter를 업데이트하는 과정으로 여러 번의 경험을 통해 점진적으로 이루어진다. 예를 들어 Gaussian 분포를 사용하는 경우 평균과 분산과 같은 parameter에 대한 prior belief가 존재하며, 새로운 데이터를 관찰할 때마다 posterior로 업데이트된다.

6.7 Generative Model과 Generative Process

6.7.1 두 개념의 차이

Active Inference에서 중요한 점은 generative model과 generative process가 서로 다른 개념이라는 것이다. Generative model은 agent가 내부적으로 가지고 있는 환경에 대한 믿음이며, generative process는 실제 환경에서 발생하는 물리적 과정이다. Agent는 환경을 완전히 알 수 없기 때문에 generative model은 실제 generative process와 일치하지 않을 수 있다.

6.7.2 Sensorimotor Model

어떤 시스템에서는 환경 전체를 정확히 model링하지 않고도 행동 수행이 가능하다. 이러한 경우 행동 수행에 필요한 정보만 포함하는 sensorimotor model 또는 action-oriented model을 사용할 수 있다. 예를 들어 passive dynamic walker와 같은 시스템에서는 환경의 물리 법칙을 완전히 model링하지 않아도 환경의 물리적 특성을 활용하여 안정적인 행동이 가능하다.

6.8 Active Inference의 실제 활용

Active Inference model을 실제 문제에 적용할 때는 generative model과 generative process를 정의하는 위의 절차를 기반으로 model을 설계한다. 이러한 구조가 설정되면 내부 상태와 행동 상태는 free energy 함수를 최소화하는 방향으로 업데이트된다. 이 과정에서 perception과 action은 동시에 설명되며, 실제 model링에서는 상황에 따라 temporal depth를 조정하거나 expected free energy의 일부 항을 제거하는 ablation 분석 등을 수행하여 model을 단순화하거나 비교할 수 있다.

Chapter 7: Active Inference in Discrete Time

서효정

〈Active Inference〉 7장은 이 이론을 실제 시스템(생물학적 혹은 인공적)에 구현할 때 가장 핵심이 되는 이산 시간(discrete time) model을 다룬다. 앞서 개념적으로 설명된 Active Inference라는 이론을 실제로 어떻게 쓰이는지, 구체적인 예를 통해 보여주겠다는 것이 이 장의 목적이다. 이는 뇌가 어떻게 불연속적인 행동(선택)을 결정하고, 정보를 탐색하며, 계층적인 구조를 통해 복잡한 인지 기능을 수행하는지 이해하는 데 필수적이다.

7.1 개요: 이산 시간(Discrete Time) 모델의 필요성

Active Inference 이론을 실제 시스템에 구현하기 위한 핵심 요소로서 이산 시간 model을 제시한다. 뇌의 인지 과정은 연속적인 환경 속에서도 saccade와 같은 불연속적 정보 수집 단계를 기반으로 작동하며, 각 시간 단계마다 감각 결과(o)를 통해 숨겨진 상태(s)를 역추론하고 다음 행동을 결정한다. 이를 설명하기 위해 본문은 POMDP(Partially Observable Markov Decision Process)를 model의 틀로 사용한다. 이러한 이산 시간 추론 구조는 theta rhythm과 같은 신경생물학적 리듬 및 hierarchical inference 개념과 정합적으로 연결되며, free energy principle이 실제 신경 활동과 행동으로 구현되는 방식을 설명한다.

7.2 Hidden Markov Model

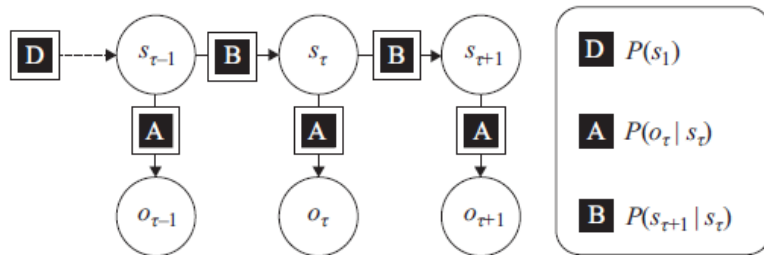


그림 7.1: Hidden Markov Model

Hidden Markov Model (HMM)은 4장에서 도입된 표기법을 사용하여 시간에 따라 변화하는 상태열(s)을 표현한다. 각 시점의 상태는 관측 가능한 결과(o)를 생성하며, 현재 상태는 직전 상태에만 의존하는 Markov 성질을 가진다. 이러한 상태 전이는 전이 행렬(B)로 표현되며, 초기 상태는 사전 분포(D)에 의해 결정된다. 7장에서는 이러한 이산 시간 문제를 다루는 Active Inference model을 HMM 또는 POMDP의 형태로 정식화한다. 제시된 예시는 지각 처리 model과 이산적 포식 선택 model로, 이는 보상을 획득하기 위해 의사결정 지점에서 좌·우 중 하나를 선택하는 문제로 구성된다. 또한 정보 탐색, 학습, 새로움 추구와 같은 과정 역시 이산 시간 Active Inference의 틀 안에서 다루어진다.

7.3 Epistemic Value

정보 탐색(information seeking)의 수학적 원리를 epistemic value 개념을 통해 설명한다. Epistemic value는 agent가 단순한 보상 획득(exploitation)을 넘어 탐색(exploration)을 수행하는 생물학적 동기를 정식화한 지표이다. 제시된 수식 $I(\pi)$ 는 특정 행동 policy(π)이 기대되는 정보 획득량을 나타내며, 이는 여러 동등한 방식으로 해석될 수 있다.

$$\begin{aligned}
 \underbrace{I(\pi)}_{\text{Epistemic value}} &= \underbrace{H[Q(\tilde{o}|\pi)]}_{\text{Post. pred. entropy}} - \underbrace{E_{Q(\tilde{s}|\pi)}[H[P(\tilde{o}|\tilde{s})]]}_{\text{Expected ambiguity}} \\
 &= \underbrace{D_{KL}[P(\tilde{o}|\tilde{s})Q(\tilde{s}|\pi) \parallel Q(\tilde{o}|\pi)Q(\tilde{s}|\pi)]}_{\text{Mutual information}} \\
 &= \underbrace{E_{Q(\tilde{o}|\pi)}[D_{KL}[Q(\tilde{s}|\pi, \tilde{o}) \parallel Q(\tilde{s}|\pi)]]}_{\text{Information gain, salience, Bayesian surprise}} ; \quad Q(\tilde{s}|\pi, \tilde{o}) \triangleq \frac{P(\tilde{o}|\tilde{s})Q(\tilde{s}|\pi)}{Q(\tilde{o}|\pi)} \quad (7.1)
 \end{aligned}$$

Epistemic value는 특정 행동이 내부 신념 또는 상태 추정에 얼마나 큰 변화를 유발하는지를 정량화한다. 이에 따라 agent는 정보 이득이 최대화되는 행동을 선택하며, 이러한 메커니즘을 통해 정보 탐색 이후 보상 획득으로 이어지는 효율적인 생존 전략이 자연스럽게 도출된다.

7.4 실제 신경 신호(LFP)와 계층적 추론

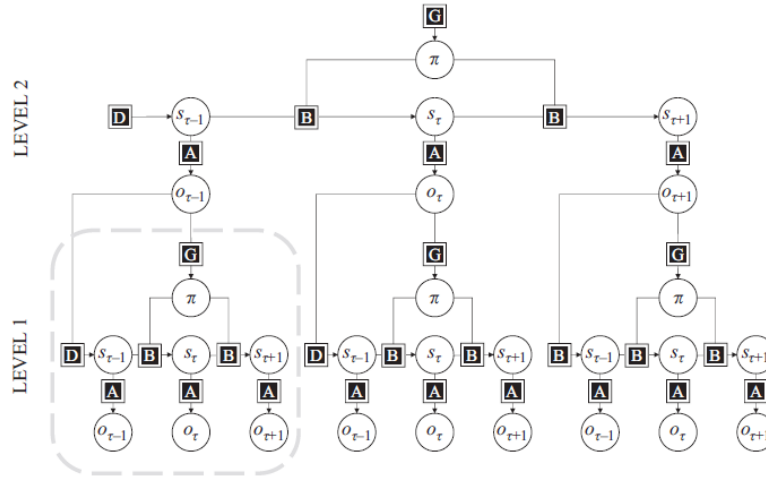


그림 7.2: 서로 다른 시간 척도에서의 계층적 추론과 LFP 신호

서로 다른 시간 척도(timescale)를 갖는 정보가 뇌에서 어떻게 계층적으로 처리되는지를 보여 준다. 상위 계층(Level 2)은 하위 계층(Level 1)에 비해 더 긴 시간 동안 유지되며, 하위 계층에서 반복적인 상태 전이가 발생하는 동안 하나의 맥락(context)을 안정적으로 유지한다. 이와 같은 시간적 분리는 단기적인 감각 변화와 장기적인 맥락 추론을 동시에 가능하게 한다.

상위 계층의 상태는 직접 관찰되지 않으며, 하위 계층의 prior로 작용하여 하위 계층의 추론을 제약한다. 이러한 구조를 통해 상위 계층은 하위 계층의 해석 공간을 제한하고, 하위 계층은 상위

계층이 제공한 맥락 내에서 빠른 추론과 상태 전이를 수행한다. 상·하위 계층은 동일한 POMDP 구조를 공유하며, 이는 동일한 추론 algorithm이 계층적으로 반복 적용됨을 의미한다.

본 model은 문장(sentence)과 단어(word)와 같이 서로 다른 시간 척도를 갖는 인지 요소 간의 상호작용을 모사한다. 또한 model 하단에 제시된 LFP 신호는 Active Inference에서의 prediction error에 대응하며, 새로운 단어의 인식이나 문맥 변화에 따른 신념 갱신 과정이 신경 신호 수준에서 나타남을 보여준다.

Chapter 8: Active Inference in Continuous Time

서효정

〈Active Inference〉 8장은 7장에서 다룬 이산 시간(discrete time) model을 보완하여, 연속 상태 공간(continuous state-space) model을 중심으로 Active Inference 이론의 구현을 확장한다. 본 장은 감각 수용기에 작용하는 물리적 변동과 근육과 같은 효과기의 연속적인 운동을 기술하는 데 적합한 generative model을 다루며, 실제 생물학적 시스템과 물리적 환경을 보다 정밀하게 묘사하는 것을 목표로 한다.

8.1 Continuous State-space Models

연속적 상태 공간 model은 감각 수용기에 영향을 미치는 물리적 변동과 신체 효과기의 연속적인 움직임을 기술하기 위해 사용되는 generative model이다. 이러한 model은 환경과 신체 상태가 시간에 따라 연속적으로 변화하는 실제 생물학적 시스템을 정밀하게 표현하는 데 적합하며, 감각 입력과 운동 출력이 불연속적인 사건이 아니라 연속적인 과정으로 이루어진다는 점을 전제로 한다.

연속적 상태 공간 model을 통해 운동 제어 model의 기본 원리와 동역학계(dynamical systems)가 수행하는 역할을 설명한다. 또한 개별 상태들의 상호작용이 시간에 따라 안정적으로 결합되는 과정을 일반화된 동조(generalized synchrony) 개념을 통해 다룬다. 궁극적으로 연속적 상태 공간 model은 이산 시간(discrete) model과의 통합을 통해, Active Inference 이론이 감각·운동·환경 상호작용을 단일한 generative model framework 안에서 설명할 수 있도록 한다.

8.2 Dynamical Systems

동역학계(dynamical systems)는 Active Inference에서 신체의 연속적인 움직임이 어떻게 생성되고 조절되는지를 설명하는 핵심적인 수학적 틀이다. 본 장에서는 특히 척수 반사(spinal reflexes) 수준에서의 운동 제어를 예로 들어, 인지적 추론과 물리적 움직임이 어떻게 연결되는지를 설명한다.

Agent는 자신의 신체 상태에 대해 특정한 신념(belief)을 가지며, 이는 generative model 내에서 감각 결과가 어떠해야 하는지에 대한 예측으로 표현된다. 실제 감각 입력과 이러한 예측 간의 차이는 prediction error로 나타나며, 이는 척수 수준에서 즉각적인 행동(action)을 유발한다. 이때 행동은 외부 세계를 직접 조작하기 위한 것이 아니라, 감각 입력이 예측과 일치하도록 신체 상태를 조정하는 역할을 수행한다.

이러한 과정에서 행동은 감각 prediction error를 최소화하는 방향으로 지속적으로 조정되며, 예측과 실제 감각이 일치하는 상태에 도달하면 움직임은 자연스럽게 종료된다. 이를 통해 동역학계 기반 Active Inference model은 고차원적 계획 없이도 연속적인 운동이 생성될 수 있음을 보여준다.

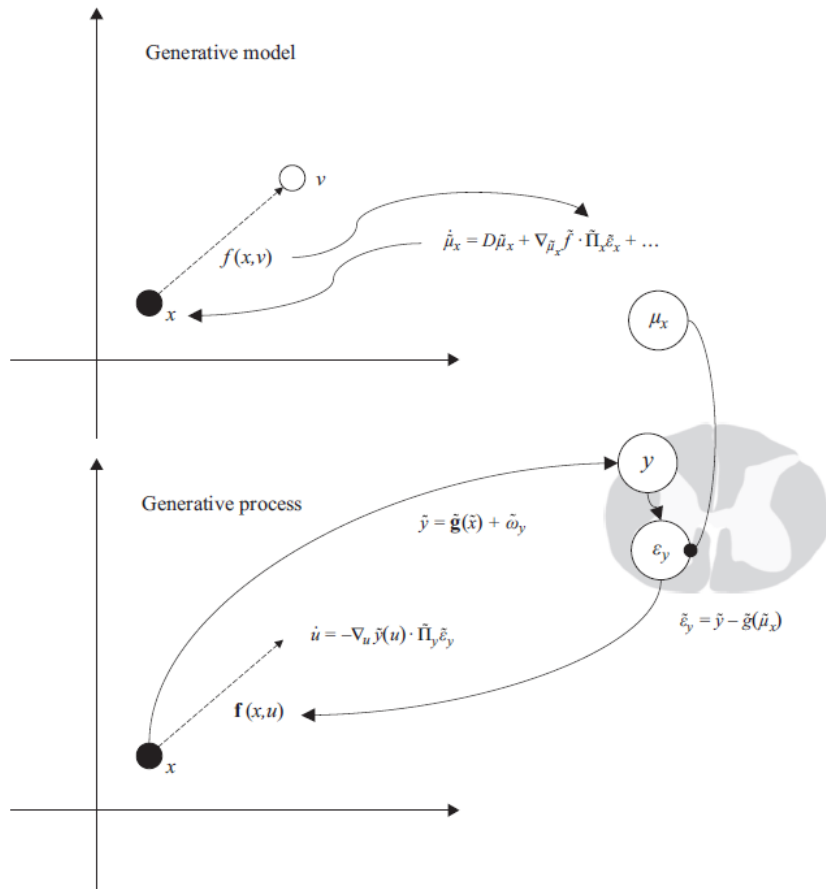


그림 8.1: Dynamical systems에 기반한 운동 제어

$$\dot{x} = \begin{bmatrix} \sigma(x_2 - x_1) \\ x_1(\rho - x_3) - x_2 \\ x_1x_2 - \beta x_3 \end{bmatrix}$$

그림 8.2: Lorenz system의 혼돈적 궤적

8.3 Lorenz System

Lorenz system은 특정한 평형점에 수렴하지 않고 지속적으로 변화하는 비선형 동역학계의 대표적인 예로, 3차원 공간에서 혼돈(chaos)과 순환성을 동시에 지닌다. 이 system은 초기 조건에 매우 민감하게 반응하며, 시간이 지남에 따라 복잡하고 예측 불가능한 궤적을 생성한다는 특징을 가진다.

Active Inference 맥락에서 Lorenz system은 고정된 상태에 머무르지 않고 지속적으로 변화하는 system의 동역학을 modeling하는 데 활용된다. 이러한 특성으로 인해 Lorenz system은 매우 역동적이고 변화무쌍한 감각 데이터를 생성할 수 있으며, 단순한 선형 model로는 설명하기 어려운 복잡한 행동 패턴을 표현하는 데 적합하다.

특히 system이 끊임없이 상태를 갱신하며 진화한다는 점에서, Lorenz system은 의도성(intentionality), 즉 agent가 단순한 반응을 넘어 목적 지향적으로 행동하는 양상을 설명하는 데 유용한 예로 사용된다. 이를 통해 비선형 동역학계는 Active Inference에서 복잡한 행동과 상태 전이를 기술하기 위한 중요한 수학적 기반을 제공한다.

8.4 Hybrid Generative Model

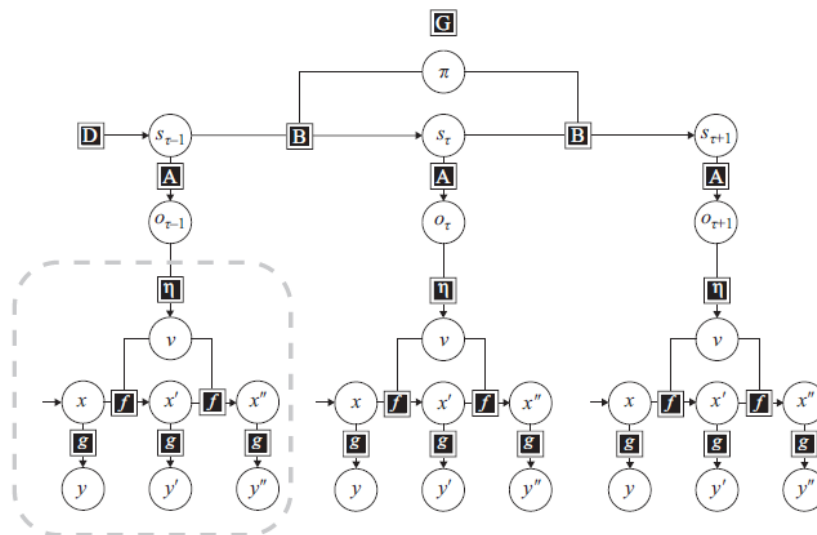


그림 8.3: Hybrid generative model의 계층 구조

Hybrid generative model은 Active Inference에서 이산적 추론과 연속적 동역학을 하나의 framework로 통합하기 위해 제안된다. 상위 계층은 POMDP 구조를 통해 이산적인 상태와 시간 척도에서 계획 수립과 전략적 선택을 수행하며, 하위 계층은 연속적 상태 공간에서 신체의 실제 운동을 생성하고 조절한다.

이러한 구조에서 상위 계층의 이산적 상태는 하위 계층에서 실행될 연속적인 운동 궤적의 구간을 선택하거나 전환하는 역할을 한다. 즉, 추상적인 의사결정은 구체적인 연속 운동으로 변환되며, 하위 계층의 감각 피드백은 다시 상위 계층의 추론을 갱신하는 데 사용된다. 이를 통해 서로 다른 시간 척도와 표현 방식을 갖는 두 수준의 model이 유기적으로 연결된다.

Hybrid generative model의 궁극적인 목적은 이산적 추론 구조와 연속적 운동 제어 간의 불일치를 해소하는 데 있다. 이를 통해 Active Inference는 고수준의 계획과 저수준의 물리적 실행을 단일한 generative model 안에서 설명할 수 있으며, 실제 생물학적 행동과 인공 시스템의 작동 방식을 보다

일관되게 기술할 수 있다.

Chapter 9: Model-based Data Analysis

이서영

9.1 Active Inference를 이용한 행동 데이터 분석의 개요

Active Inference는 인지 과정을 설명하는 이론일 뿐만 아니라, 실제 실험에서 얻어진 행동 데이터를 분석하는 방법으로도 사용될 수 있다. 연구자는 피험자의 행동 데이터를 관찰한 뒤, 그 행동이 어떤 generative model에서 생성되었는지를 추론함으로써 피험자의 내부 인지 구조를 추정할 수 있다.

이 과정에서는 두 가지 model을 구분하여 생각해야 한다. 먼저 subjective model은 피험자가 실제로 가지고 있다고 가정되는 내부 generative model을 의미한다. 반면 objective model은 연구자가 피험자의 행동 데이터를 설명하기 위해 설정한 분석용 model을 의미한다.

연구자는 objective model을 데이터에 맞추어 행동을 설명하고, 그 결과를 이용하여 피험자의 subjective model에 포함된 parameter를 추정한다. 이러한 과정은 연구자가 피험자의 추론 과정을 다시 추론한다는 의미에서 meta-Bayesian inference라고 설명된다.

9.2 행동 데이터로부터 Generative Model Parameter를 추정하는 과정

행동 데이터를 분석하는 중요한 목적은 피험자의 generative model parameter를 추정하는 것이다. 이를 통해 피험자가 어떤 prior belief를 가지고 행동했는지, 혹은 어떤 정보 처리 방식을 사용했는지를 파악할 수 있다.

9.2.1 Computational Phenotyping

행동 데이터를 가장 잘 설명하는 model parameter를 추정하면 해당 parameter는 피험자의 인지적 특성이나 정보 처리 방식을 나타내는 지표로 해석될 수 있다. 이러한 접근을 computational phenotyping이라고 부른다.

즉 개인의 행동 특성이나 의사결정 성향을 단순한 행동 패턴이 아니라 model parameter 값으로 표현하는 방식이다. 예를 들어 어떤 사람은 탐색 행동을 더 선호할 수 있으며, 또 다른 사람은 위험을 회피하는 경향이 더 강할 수 있다. 이러한 차이는 generative model의 parameter 값 차이로 표현될 수 있다.

9.2.2 여러 Generative Model 사이의 비교

행동 데이터를 설명할 수 있는 model은 하나만 존재하는 것이 아니라 여러 가지가 가능하다. 따라서 연구자는 여러 generative model을 설정한 뒤, 어떤 model이 데이터를 가장 잘 설명하는지를 비교해야 한다.

이 과정에서 중요한 기준이 되는 것은 model evidence이다. Model evidence는 특정 model이 관찰된 행동 데이터를 얼마나 잘 설명하는지를 나타내는 지표이다. 연구자는 여러 model의 model evidence를 비교하여 가장 설명력이 높은 model을 선택하게 된다.

9.3 Bayesian 구조에서의 행동 데이터 분석

행동 데이터 분석은 기본적으로 Bayesian inference 구조를 따른다. 연구자는 행동 데이터를 이용하여 model parameter와 model 구조를 동시에 추론하게 된다.

이 과정에서는 다음과 같은 관계가 사용된다. 특정 model과 parameter가 주어졌을 때 행동 데이터가 나타날 확률을 likelihood라고 부른다. Parameter에 대해 연구자가 사전에 설정한 믿음을 prior라고 부른다. 행동 데이터를 관찰한 이후 업데이트된 parameter 분포를 posterior라고 부른다.

또한 특정 model이 전체 데이터를 설명하는 정도를 나타내는 값이 model evidence이다.

이러한 구조를 통해 연구자는 두 가지 중요한 문제를 해결할 수 있다. 첫 번째 문제는 특정 model이 주어졌을 때 어떤 parameter 값이 데이터를 가장 잘 설명하는지를 찾는 parameter estimation 문제이다. 두 번째 문제는 여러 model 중 어떤 model이 데이터를 가장 잘 설명하는지를 비교하는 model comparison 문제이다.

9.4 Model-based Analysis의 실제 분석 절차

실제 연구에서 Active Inference model을 이용하여 행동 데이터를 분석할 때는 여러 단계의 절차가 필요하다.

9.4.1 행동 데이터 수집 단계

먼저 피험자가 특정 과제를 수행하면서 나타난 행동 데이터를 수집해야 한다. 예를 들어 피험자가 어떤 자극을 보았을 때 어떤 선택을 했는지, 혹은 어떤 행동을 수행했는지를 기록한다. 이러한 데이터는 이후 model 기반 분석의 입력 자료로 사용된다. 또한 여러 피험자로부터 데이터를 수집하면 개인 간 차이를 분석할 수 있으며, 이를 통해 집단 수준 분석을 수행할 수 있다.

9.4.2 POMDP 기반 Generative Model 설정 단계

다음 단계에서는 해당 행동 과제를 설명할 수 있는 generative model을 설정한다. Active Inference에서는 이러한 generative model을 보통 POMDP 구조로 표현한다. 이 model에는 hidden state, observation, action, state transition 구조, 그리고 prior preference와 같은 요소들이 포함된다. 이러한 구조는 피험자가 환경을 어떻게 인식하고 행동을 선택하는지를 설명하는 내부 model을 나타낸다.

9.4.3 Likelihood 함수 설정 단계

Model이 설정되면, 해당 model이 특정 행동을 생성할 확률을 계산할 수 있다. 이러한 확률을 likelihood 함수라고 부른다. 연구자는 likelihood 함수를 이용하여 실제로 관찰된 행동 데이터가 해당 model에서 얼마나 잘 설명되는지를 계산하게 된다.

9.4.4 Parameter에 대한 Prior Belief 설정 단계

다음 단계에서는 model parameter에 대한 prior belief를 설정한다. 이 단계에서는 parameter의 기대값이나 분산과 같은 값이 정의된다. 이러한 prior는 피험자의 행동 성향이 어떤 범위 안에 있을 것이라는 연구자의 가정을 수학적으로 표현한 것이다.

9.4.5 Posterior와 Model Evidence 계산 단계

행동 데이터와 prior belief가 설정되면, 연구자는 parameter의 posterior 분포를 계산할 수 있다. 또한 각 model이 데이터를 얼마나 잘 설명하는지를 나타내는 model evidence도 계산할 수 있다. 이 과정에서는 Variational Laplace와 같은 표준적인 Bayesian inference 방법이 사용된다. 이를 통해 연구자는 각 피험자의 generative model parameter를 추정할 수 있다.

9.4.6 집단 수준 분석 단계

여러 피험자의 parameter를 추정한 뒤에는 집단 수준 분석을 수행할 수 있다. 이 단계에서는 Parametric Empirical Bayes(PEB)와 같은 방법이 사용된다.

PEB 분석을 통해 연구자는 개인 간 행동 차이가 어떤 parameter 차이로 설명되는지를 분석할 수 있다. 또한 환자 집단과 정상 집단 사이의 인지적 차이를 model parameter 수준에서 설명할 수 있다.

9.5 Model 검증 과정

Model 기반 분석에서는 model이 실제 데이터를 제대로 설명하는지 확인하는 과정이 필요하다.

9.5.1 Face Validity 검증

먼저 model이 생성한 가짜 데이터가 실제 행동 데이터와 유사한지를 확인해야 한다. 이를 통해 model이 실제 행동 패턴을 현실적으로 재현할 수 있는지를 평가할 수 있다.

9.5.2 Parameter Recovery 검증

또한 model이 parameter 값을 정확하게 추정할 수 있는지도 검증해야 한다. 이를 위해 연구자는 알려진 parameter 값을 사용하여 가짜 데이터를 생성한 뒤, model 역추정을 수행하여 동일한 parameter 값을 다시 복원할 수 있는지를 확인한다. 이 과정을 통해 특정 parameter가 실제로 식별 가능한지 평가할 수 있다.

9.6 Generative Model 연구 사례

Active Inference 기반 generative model은 다양한 행동 연구에서 사용되어 왔다.

9.6.1 Adams 연구

Adams의 연구에서는 smooth pursuit eye movement를 설명하기 위한 continuous generative model이 사용되었다. 이 연구에서는 눈의 위치와 속도와 같은 연속적인 변수를 이용하여 generative model을 구성하였다.

이 model을 이용하여 감각 신호의 신뢰도와 관련된 parameter를 추정할 수 있었으며, 행동 데이터와 함께 neuroimaging 데이터를 이용하여 subjective model과 뇌 활동 사이의 관계도 분석하였다.

9.6.2 Mirza 연구

Mirza의 연구에서는 시각 탐색 과제를 수행하는 동안 나타나는 saccadic eye movement를 분석하였다. 연구자는 POMDP 기반 generative model을 이용하여 피험자가 어떤 위치로 눈을 이동시키는지 설명하였다.

이 연구에서는 특히 epistemic value를 포함한 model이 더 높은 model evidence를 보였다는 결과가 보고되었다. 이는 인간의 행동이 단순히 보상을 얻기 위한 것이 아니라 정보를 얻기 위한 탐색 행동도 포함한다는 점을 보여준다.

9.7 False Inference와 정신 질환 Modeling

Active Inference는 정신 질환을 설명하는 이론적 model로도 사용될 수 있다. 정신 질환은 종종 잘못된 generative model이나 잘못된 추론 과정으로 해석될 수 있다.

예를 들어 잘못된 prior belief가 존재하거나 감각 정보의 precision이 비정상적으로 설정되면 환경에 대한 추론이 왜곡될 수 있다. 이러한 왜곡된 추론 과정은 환각이나 망상과 같은 증상으로 나타날 수 있다.

9.8 Precision의 역할

Active Inference에서는 불확실성의 정도를 나타내는 precision이 중요한 역할을 한다. Precision은 특정 정보가 얼마나 신뢰할 수 있는지를 나타내는 개념이다. 예를 들어 likelihood precision은 감각 정보가 얼마나 신뢰할 수 있는지를 의미한다. 반면 prior precision은 기존의 믿음이 얼마나 강하게 유지되는지를 의미한다.

이 두 값의 균형이 깨지면 잘못된 추론이 발생할 수 있다. 예를 들어 likelihood precision이 지나치게 높으면 noise가 포함된 감각 정보를 과도하게 신뢰하게 될 수 있다. 반대로 prior precision이 지나치게 높으면 새로운 정보를 받아들이지 못하고 기존 믿음을 계속 유지하게 될 수 있다.

이러한 precision의 이상은 환각이나 망상과 같은 정신 질환 현상을 설명하는 데 사용될 수 있다.

Additional Study

윤시원, 서효정

현대 AI Model과 Active Inference 비교

작성자: 윤시원

현대 AI model(e.g., LLM, ANN)의 문제점을 살펴보고 Active Inference와 비교하였다.

A.1 No Real-Time Update

현대 LLM model은 data를 가지고 학습을 통해 weight 값을 변화시킨다. 하지만 사용자와 대화하는 경우 frozen weight를 사용하여 실시간 학습이 불가능하다. 반면 Active Inference는 연속 시간에서 세상과 생명체가 관찰과 행동을 통해 서로 상호작용함을 가정한다.

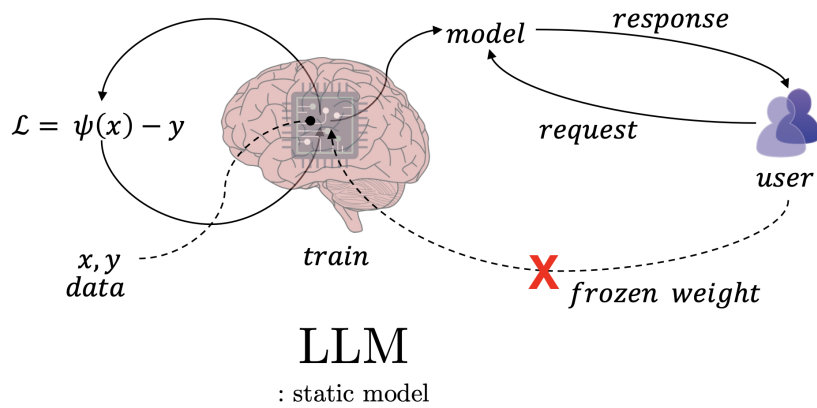
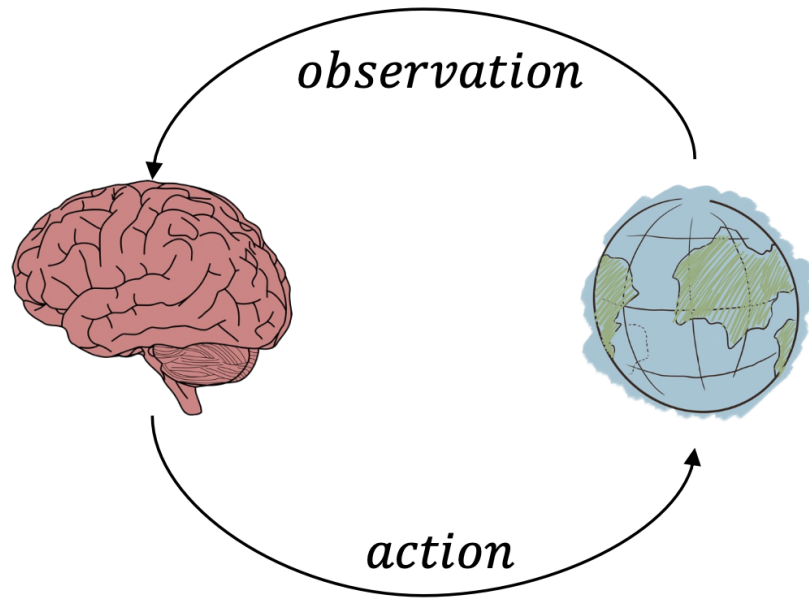


그림 A.1: 현대 LLM의 frozen weight 구조와 Active Inference의 실시간 상호작용 비교

A.2 Collapsed Markov Blanket

현대 LLM은 학습한 data를 학습하여 앞으로 다음 단어를 예측한다. 이러한 model은 공감을 나타낸 문헌들을 학습하여 인간의 공감을 흉내낼 수는 있으나 상대의 상황을 자신의 상황으로 simulation(Theory of Mind, ToM)하여 나타나는 진정한 의미의 공감이 불가능하다. 비단 ToM의 어려움뿐 아니라 계층 model의 부재로 깊은 temporal depth를 가질 수 없으며 이는 높은 temporal depth를 필요로 하는 신뢰, 용서, 허상, 종교, 더 나아가 공동체를 형성하는 데 어려울 수 있다.



Active Inference

: action-perception loop

그림 A.2: Active Inference 에이전트의 실시간 업데이트 메커니즘

A.3 Lacks First Principles

ANN은 해결하고자 하는 문제를 최적화 문제로 바꾸어 해결하는 기법이다. 하지만, 이 방법이 현실에서 작동해야 한다는 원리가 없다. Active Inference는 Hamiltonian을 생명체의 first principles을 만족함을 주장한다.

Experimental Validation of the Free-Energy Principle with In Vitro Neural Networks

작성자: 서효정

교재 내용을 기반으로 자유 주제를 선정하여 후속 스터디를 진행하였다. 본 파트에서는 논문 리딩 결과를 정리한다.

A.1 논문 핵심 요약

주제는 실제 생물학적 신경 network(in vitro)가 free energy 최소화를 통해 학습하고 환경에 적응하는지를 실험적으로 검증하는 것이다.

배경으로, Karl Friston이 제안한 free energy principle은 지각이 있는 존재가 자신의 내부 model과 외부 환경 사이의 불일치(surprise 또는 free energy)를 최소화하려는 방향으로 행동하고 학습한다고 설명한다. 하지만 그동안은 주로 수학적 modeling에 그쳤고, 실제 neuron 수준에서의 증거는 부족하였다.

실험 방법으로는, 쥐의 뇌 세포를 배양한 신경 network(MEA, Multi-Electrode Array)에 특정 전기 자극 pattern을 주어 환경을 simulation하고, neuron들이 이 자극에 어떻게 반응하고 학습하는지를 관찰하였다.

결과적으로, 실제 neuron 집단이 free energy를 최소화하는 방향으로 연결성(synaptic plasticity)을 수정하며, 외부 자극을 예측하는 내부 model을 형성한다는 것을 확인하였다.

A.2 역공학 접근의 중요성

이 논문의 또 하나의 중요한 특징은 접근 방식 자체에 있다. 기존 연구들은 보통 뇌가 Bayesian inference를 한다고 가정한 뒤 그 가정에 맞는 update 규칙을 설계하는 방식이었다. 그러나 이 논문은 반대로 접근한다. 실제 신경망의 반응을 먼저 관찰한 다음, 그 반응을 가장 잘 설명하는 generative model이 무엇인지 역으로 추론한다.

이 접근이 중요한 이유는 두 가지이다. 첫째, 생물학적 신경망이 실제로 어떤 세계 model을 사용하고 있는지를 추정할 수 있다. 둘째, 그 generative model을 알게 되면 단순한 사후 설명을 넘어 앞으로의 학습과 자기 조직화를 정량적으로 예측할 수 있다. 즉, 이 연구는 해석 가능성뿐만 아니라 예측 가능성을 동시에 확보하려는 시도라고 볼 수 있다.

A.3 실험적 검증: 시험관 내 신경망

이 접근이 실제로 검증되었는지를 살펴보면, 연구진은 쥐 배아의 피질 신경세포로 구성된 시험관 내(in vitro) 신경망을 사용하였다. 그리고 이 신경망의 자극-반응 pattern을 가장 잘 설명하는 POMDP 기반 generative model을 역설계하였다.

그다음, 이 generative model로부터 가소성 규칙을 도출하고 실제 신경망이 시간에 따라 어떻게 자기 조직화되는지를 예측하였다. 그 결과, 실제 신경망의 변화는 free energy 최소화의 경사 하강 법과 매우 잘 일치하였다. 즉, 이 연구는 이론이 data를 설명하는 데 그치지 않고 실제 생물학적 변화를 예측한다는 점을 보여준다.

A.4 신경 신호와 Bayesian 변수의 대응 관계

이 논문에서 특히 주목할 점은 신경학적 변수와 Bayesian 변수 사이의 구체적인 대응 관계를 제시하였다는 것이다. 실험 결과에 따르면, 신경 활동은 hidden state의 posterior를 encoding하고, 기저 흥분성은 상태에 대한 prior를, 시냅스 강도는 model parameter의 posterior를 encoding하는 것으로 해석된다.

즉, 이 신경망은 단순한 값이 아니라 실제로 확률 분포를 표현하고 조정하는 system으로 이해할 수 있다. 이러한 점은 이 연구를 단순한 비유 수준이 아니라 수학적으로 엄밀한 이론과 연결시키는 핵심 요소이다.

A.5 단순한 Hebbian Learning의 한계

이 논문은 단순한 Hebbian learning이 왜 충분하지 않은지도 설명한다. 단순 Hebbian learning은 사실상 잘못된 prior를 가진 비최적화된 generative model에 해당한다.

특히 blind source separation 문제를 제대로 해결하기 위해서는 Hebbian plasticity와 항상성 plasticity의 균형이 필요하다고 주장한다. 그리고 이러한 균형이 필요한 이유는 free energy principle을 통해서만 일관되게 설명될 수 있다. 즉, 이 연구는 특정 학습 규칙이 작동한다는 사실을 넘어서 왜 그러한 규칙이어야 하는지를 이론적으로 정당화한다.

A.6 정신 질환, 약물, 지각 왜곡과의 연결

이 논문은 이론적 논의를 넘어 정신 질환과 약물 효과까지 연결한다. 연구에서 기저 흥분성은 prior에 해당한다. 따라서 prior가 왜곡되면 지각 자체도 왜곡된다.

이 관점은 환각, 망상, 그리고 항불안제나 환각제가 지각에 미치는 영향을 하나의 통합된 framework로 설명할 수 있게 한다. 즉, 정신 질환을 단순한 이상 현상이 아니라 확률 추론 system의 오류로 이해할 수 있다.

A.7 AI 및 뉴로모픽 컴퓨팅으로의 확장

이 논문은 AI로의 확장 가능성도 제시한다. Free energy 기반 학습은 역전파가 필요 없고 국소 정보만으로 작동하며 생물학적으로 타당하다.

이러한 특성은 저전력, 고효율, 해석 가능한 AI 설계로 이어질 수 있다. 즉, 이 연구는 뇌 이론인 동시에 차세대 AI algorithm의 후보로도 볼 수 있다.

A.8 결론

정리하면 이 연구는 다음과 같은 점을 보여준다. 첫째, free energy principle은 실제 신경망을 정량적으로 예측할 수 있다. 둘째, 추론과 학습을 하나의 원리로 통합한다. 셋째, 생물학적 data와 이론을 직접적으로 연결한다. 넷째, 정신 질환, AI, 뇌과학 전반에 응용 가능하다.

요약하면, 이 연구는 free energy principle이 단순한 설명 이론이 아니라 실제 생물학적 신경망의 추론, 학습, 자기 조직화를 정량적으로 예측할 수 있는 법칙임을 실험적으로 입증하였다.