Chapter 1

Active Inference(AIF)

1.1 Active Inference

Recall the definition of POMDP, we will define a 7-tuple for the whole decision making process: $\{\mathbb{S}, \mathbb{A}, \mathbb{O}, \mathbf{T}, \mathbf{R}, \mathbf{O}, \gamma\}$, where at time step t, the state $s_t \in \mathbb{S}$ are hidden in the real world, its transition mechanism is illustrated by the transition model $T(s_{t+1}|s_t, a_t)$, action will be taken according to a policy $\pi(a_t|s_t)$, the observation comes from the observation model $o_t = O(o_t|s_t)$. To optimize the policy, we define a reward model $r_t = R(s_t, a_t)$ and the weighted accumulated total return $\mathbb{E}[U_t^\pi] = \mathbb{E}\left[\sum_{i=t}^\infty \gamma^{i-t} r_i\right]$.
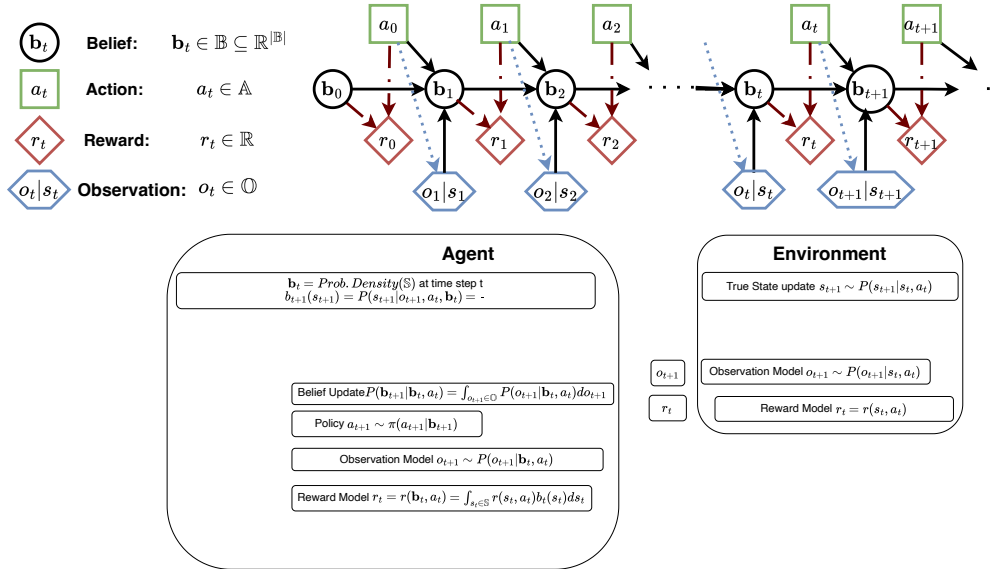

Figure 1.1: Original POMDP Model

There are only one model for the real world

After understanding the basics of Partially Observable Markov Decision Process (POMDP), we can To have an basic understanding of Active Inference shown in the Figure 1.2, we should clarify that there are two models defined, i.e. the real world model described by the generative process $P^\star(s^\star, o)$ and the generative model $P(s, o) = P(s|o)P(o) = P(o|s)P(s)$.

The true process is always unknown, the agent has a generative model about how the outside world works. The only true connection between the interior and external world is the outcome (or observation).

The internal generative model would be updated every time observe some outcome. The prior belief is defined as the probability density of the state $P(s)$, the posterior is defined as

$$P(s|o) = \frac{P(s, o)}{P(o)} = \frac{P(o|s)P(s)}{P(o)} \tag{1.1}$$

Due to the difficulty of calculating the evidence $P(o) = \int_{S \in \mathbb{S}} P(o|S)P(S)$, we may use $Q(s)$ to approximate the posterior distribution $P(s|o)$. The information loss caused when using $Q(s)$ to approximate $P(s|o)$ is given by the relative Kullback-Divergence:

$$D_{KL}[Q(s)\|P(s|o)] = \mathbb{E}_{Q(s)}[\ln(Q(s)) - \ln(P(s|o))] \tag{1.2}$$

When substitute the Eq. 1.1 to the Eq. 1.2 we could get:

$$D_{KL}[Q(s)\|P(s|o)] = \mathbb{E}_{Q(s)}[\ln Q(s) - \ln P(s|o)] \tag{1.3a}$$

$$= \mathbb{E}_{Q(s)}\left[\ln Q(s) - \ln\left(\frac{P(s, o)}{P(o)}\right)\right] \tag{1.3b}$$

$$= \mathbb{E}_{Q(s)}[\ln Q(s) - \ln(P(s, o))] + \mathbb{E}_{Q(s)}[\ln P(o)] \tag{1.3c}$$

The active inference framework is based on the premise that perception and learning can be understood as minimizing a quantity known as **variational free energy** (VFE), and that action selection, planning and decision-makoing can be understood as minimizing the **expected free energy** (EFE).

- **Perception**: posterior state inference after each new observation

- **Learning**: slowly updating the priors and likelihood distributions in the model over many observations (which facilitates more accurate state inference in the long run).

- **Action selection, planning and deccision-making**: to select policy that will bring about future observations that minimize VFE.

Table 1.1: The physical meaning of **variational free energy** $\mathcal{F}$

Variational	$=$	$D_{KL}\left[Q(s)\|P(s|o)\right]$	$+$	$-\mathbb{E}_{Q(s)}[\ln P(o)]$ Eq. 1.4b
Free	$=$	Divergence	$+$	Surprisal
Energy	$=$	$D_{KL}\left[Q(s)\|P(s)\right]$	$-$	$\mathbb{E}_{Q(s)}[\ln P(o|s)]$ Eq. 1.4c
$\mathcal{F}$	$=$	Complexity	$-$	Accuracy (Goodness of fit)

Variational Free Energy (VFE) $\mathcal{F}$

The first term in Eq. 1.3c is defined as the **variational free energy** $\mathcal{F} \overset{\text{def}}{=\!=} \mathbb{E}_{Q(s)}\left[\ln(Q(s)) - \ln(P(s,o))\right] = D_{KL}[Q(s)\|P(s,o)]$ Then we can write that

$$\mathcal{F} = D_{KL}\left[Q(s)\|P(s,o)\right] \tag{1.4a}$$
$$= D_{KL}\left[Q(s)\|P(s|o)\right] - \mathbb{E}_{Q(s)}[\ln P(o)] \tag{1.4b}$$
$$= D_{KL}\left[Q(s)\|P(s)\right] - \mathbb{E}_{Q(s)}[\ln P(o|s)] \tag{1.4c}$$

The physical meaning of the **variational free energy** (VFE) is shown in Table 1.1:

From the Eq. 1.4b, the term Divergence is the information loss when using the approximated density to approximate the posterior, which is no less than zero. Minimizing the **variational free energy** is equivalent to minimize the **surprisal** (defined as the negative log evidence) and also equivalent to maximize the evidence.

From the Eq. 1.4c, the Complexity is a measure of the difference between the approximation model and the prior density. And the accuracy is a meassure of the likelihood, to show how well the model fit the observation. Minimizing VFE means only increase the accuracy of the model, also reducing the change effort when observing new observation.

In summary, by minimizing the **variational free energy**, we can perform the task of both perception and learning to find (approximatedly) optimal posterior beliefs after each new observation.

Expected Free Energy (EFE) $\mathcal{G}$

The task of action selection and planning is to select policies that will bring about future observations that minimize VFE. However, the future outocmes have not yet been observed. Actions must therefore be selected such that they minimize expected free energy (EFE).

Similarly to the previous introduced posterior approximation $P(s|o) \approx Q(s)$, the approximation of observations are written as $Q(o)$, the joint distribution of state and observation is $Q(s,o) = Q(o|s)Q(s)$
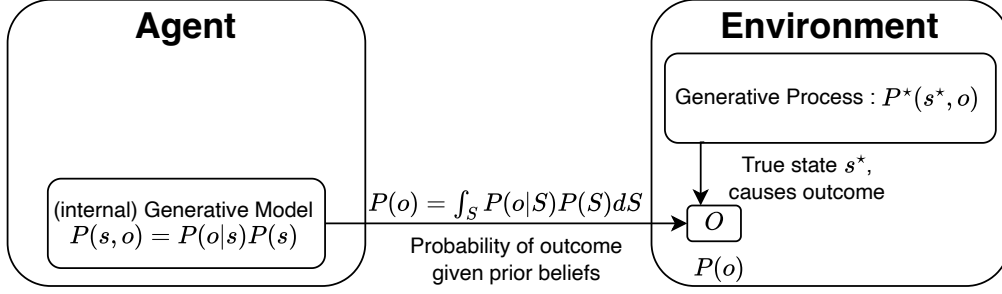
Figure 1.2: Graphical Representation of the Generative Process(based on the true states $\mathbf{s}^\star$) in the world and the corresponding (internal) generative model (based on probabilistic beliefs random variables $\mathbf{s}$) that best explain the outcomes $\mathbf{o}$. The outcomes are shared between the generative process and model. [**sajid2021active**]

The expected free energy can be seen as the expectation of the variational free energy over observations

$$\mathcal{G} \overset{\text{def}}{=} \mathbb{E}_{Q(o|s)}\left[\mathcal{F}\right] = \mathbb{E}_{Q(s,o)}\left[\ln Q(s) - \ln P(s,o)\right] \tag{1.5a}$$

$$= \mathbb{E}_{Q(o,s)}\left[\ln Q(s) - \ln P(s|o)\right] - \mathbb{E}_{Q(o,s)}[\ln P(o)] \tag{1.5b}$$

$$= \mathbb{E}_{Q(o,s)}\left[\ln Q(s) - \ln P(s|o)\right] - \mathbb{E}_{Q(o)}[\ln P(o)] \tag{1.5c}$$

$$\approx \mathbb{E}_{Q(o,s)}\left[\ln Q(s) - \ln Q(s|o)\right] - \mathbb{E}_{Q(o)}[\ln P(o)] \tag{1.5d}$$

$$\approx \mathbb{E}_{Q(o,s)}\left[\ln Q(s) - \ln Q(s|o)\right] - \mathbb{E}_{Q(o)}[\ln P(o|C)] \tag{1.5e}$$

$$= \mathbb{E}_{Q(o,s)}\left[\ln Q(o) - \ln Q(o|s)\right] - \mathbb{E}_{Q(o)}[\ln P(o|C)] \tag{1.5f}$$

$$= \mathbb{E}_{Q(o,s)}\left[\ln Q(o) - \ln P(o|s)\right] - \mathbb{E}_{Q(o)}[\ln P(o|C)] \tag{1.5g}$$

$$= D_{KL}[Q(o)\|P(o|C)] + \mathbb{E}_{Q(s)}\left[H[P(o|s)]\right] \tag{1.5h}$$

where the second line Eq. 1.5b uses the product rule of probability $P(s,o) = P(s|o)P(o)$ to rearrange EFE into two terms that can be associated with information-seeking (epistemic value) and reward-seeking (pragmatic value). In active inference, we did not define the reward specifically. But instead, we have observed that the organism has an instinction to observe some thing in their expectation. If the observation aligns with their expectation, even the observation itself is a bad phenomena like a war happening or a disaster, if it was predicted by themselves, it is a signal of reward. Which means that the internal model of the world is quite effective. ???. With this explanation, we could treat the expectation of observation as the reward. And the difference between the true posterior model and the approximation could be treated as the information-seeking. Minimizing the expected free energy means minimizing the discrepancy between the approximation and the true posterior and at the same time maximizing the possibility of observing desired results.

The third line Eq. 1.5d uses a new introduced approximation of true posterior $P(s|o) \approx Q(s|o)$.

The fourth line Eq. 1.5e uses a new introduced terminology prior expectation over observations $p(o|C)$, that plays the role of preference.

The fifth line Eq. 1.5f uses the equality $\frac{Q(s)}{Q(s|o)} = \frac{Q(s)Q(o)}{Q(o|s)Q(s)} = \frac{Q(o)}{Q(o|s)}$

The sixth line Eq. 1.5g assumes that $P(o|s) = Q(o|s)$.

The final line Eq. 1.5h introduces a new terminology: Entropy of a distribution $P(x)$ is defined as $H[P(x)] \stackrel{\text{def}}{=\!=} -\mathbb{E}_{P(x)}[\ln P(x)]$. Mutual information of x and y is defined as

$$I(x,y) \stackrel{\text{def}}{=\!=} H[p(x)] - H[p(x|y)] = H[p(y)] - H[p(y|x)] \tag{1.6}$$

$I(x,y)$ is symmetric and scores the reduction in uncertainty (entropy) about the value of a variable x afoorded by knowledge of another variable y. If these two variables are independent, mutual information is 0. To derive Eq. 1.5h, we start from Eq. 1.5g

$$\mathcal{G} = \mathbb{E}_{Q(o,s)}\left[\ln Q(o) - \ln P(o|s)\right] - \mathbb{E}_{Q(o)}[\ln P(o|C)] \tag{1.7a}$$

$$= \mathbb{E}_{Q(o,s)}\left[\ln \frac{Q(o)}{P(o|s)}\right] - \mathbb{E}_{Q(o,s)}[\ln P(o|C)] \tag{1.7b}$$

$$= \mathbb{E}_{Q(o,s)}\left[\ln \frac{Q(o)}{P(o|s)} - \ln P(o|C)\right] \tag{1.7c}$$

$$= \mathbb{E}_{Q(o,s)}\left[\ln \frac{Q(o)}{P(o|s)P(o|C)}\right] \tag{1.7d}$$

$$= \mathbb{E}_{Q(o,s)}\left[\ln \frac{Q(o)}{P(o|C)} - \ln P(o|s)\right] \tag{1.7e}$$

$$= \mathbb{E}_{Q(o,s)}\left[\ln \frac{Q(o)}{P(o|C)}\right] - \mathbb{E}_{Q(o,s)}\left[\ln P(o|s)\right] \tag{1.7f}$$

$$= \mathbb{E}_{Q(o)}\left[\ln \frac{Q(o)}{P(o|C)}\right] - \mathbb{E}_{Q(o|s)Q(s)}\left[\ln P(o|s)\right] \tag{1.7g}$$

$$= \mathbb{E}_{Q(o)}\left[\ln \frac{Q(o)}{P(o|C)}\right] - \mathbb{E}_{Q(s)}\left[\mathbb{E}_{Q(o|s)}[\ln P(o|s)]\right] \tag{1.7h}$$

$$= \mathbb{E}_{Q(o)}\left[\ln \frac{Q(o)}{P(o|C)}\right] - \mathbb{E}_{Q(s)}\left[\mathbb{E}_{P(o|s)}[\ln P(o|s)]\right] \tag{1.7i}$$

$$= D_{KL}[Q(o)\|P(o|C)] - \mathbb{E}_{Q(s)}\left[H(P(o|s))\right] \tag{1.7j}$$

1.1.1 Data Collection

The research employs multiple datasets to ensure comprehensive evaluation:

- **Dataset A**: Contains 10,000 samples with 50 features...

- **Dataset B**: Comprises time-series data spanning 5 years...

- **Dataset C**: Includes multimodal data from various sources...

1.2 Algorithm Design

The proposed algorithm builds upon the foundation established by [**lee2019optimization**]. The pseudocode is presented in Algorithm 1.

Input: X, θ
Output: Best solution in P
1 Initialize population $P \leftarrow \emptyset$;
2 **for** $i \leftarrow 1$ **to** N **do**
3     $x_i \leftarrow$ **InitializeSolution()**;
4     $P \leftarrow P \cup \{x_i\}$;
5 **end**
6 **while** *not converged* **do**
7   Evaluate fitness for all $x \in P$;
8   Select parents for reproduction;
9   Apply crossover and mutation;
10   Update population P;
11 **end**
12 **return** best solution in P;

Algorithm 1: Proposed Optimization Algorithm

1.3 Experimental Setup

All experiments were conducted on a computing cluster with the following specifications:

- CPU: Intel Xeon Gold 6248R (3.0 GHz)

- GPU: NVIDIA A100 (40GB)

- Memory: 256 GB DDR4

- Storage: 2 TB NVMe SSD