Перевод статьи: Tutorial on Active Inference

Active inference is the Free Energy principle of the brain applied to action. It is relatively supported by experimental neuroscience studies and is a popular model of 'how the brain works'. In this tutorial, we will consider the latest version, which is formulated as planning in discrete state-space and time. The initial motivation of active inference is that the agent wants to remain alive, by maintaining its homeostasis. To this end, the agent must ensure that important parameters, like body temperature or blood oxygenation, don't deviate too much from the norm, i.e. are not surprising. But since it's only possible to infer these parameters from sensory measurements, the agent minimizes surprise of sensory observations instead. Interestingly, this is equivalent to continuously improving agent's model of the world, as we will see shortly. So in short: remain alive -> maintain homeostasis -> avoid surprising states -> avoid surprising observations -> by minimizing approximation to surprise(free energy). While the previous post aimed to give a mere intuition on Free Energy, here we will get our hands dirty. No technical background is necessary except the probability theory and Bayes Theorem. The idea goes as follows. The brain avoids surprise by having a good model of the environment. On the one hand, the environment has a true probabilistic hidden variable (called state) s, which generates probabilistic observations o. It's important to emphasize that the state s is hidden, meaning that we can only observe o. For example, it rained at night (s), so the grass is wet (o). This is the generative process R(s,o).

On the other hand, the brain tries to infer the probability of different hidden states given the observation, p(s|o), from the prior belief p(s) and likelihood p(o|s). Thus, the brain constructs a generative model defined as a joint p(s,o).

Let's see a quick example. Imagine the following generative process — there is an apple and an orange tree in the garden. The identity of a fruit (apple or orange) is the hidden variable s. Suppose the apple tree is slightly to the left from the orange one, so when the fruits fall on the ground, we observe the following:

It seems like there are about 70 % of oranges and 30 % of apples. This is the true distribution of the hidden state s. When the fruits fall, they don't hit the same location but somewhat randomly disperse around. This is the probability of observations conditioned on s. Inference in the generative model lies in finding posterior p(s|o) — the probability that the fruit is an apple (or orange) if it lies at a specific location. Learning the generative model consists in estimating the distributions (like parameters mean and variance for a bell-shaped Normal) of the hidden state p(s), of the state-observation mapping p(o|s). Inferring p(s|o) via Bayes formula requires us to compute p(o), which is also interesting by itself, since the better our model, the higher will be the probability of the observed data p(o). It is also called 1) 'model evidence', since it quantifies how well is our model predicting the real data, 2) 'marginal likelihood', because we marginalize, or sum out, the hidden state s. Compare 2 models below, in which we estimated p(o,s): the one on the left is obviously much better — p(s) correctly shows that there are more oranges than apples, and p(o|s) is well centered on each cluster. We can quantify the quality of each model with model evidence p(o), marginalizing the hidden variable s:

Note: technically this is a likelihood of model parameters p(o|parameters, model) and to get the actual model evidence p(o|model) we would need to marginalize over parameters as well. We'll come back to it in the end of the post. So ideally, we want to choose model parameters that will lead to as high model evidence p(o), as possible. How does it connect to active inference and an agent that avoids surprising observations? In fact, maximizing model evidence is equivalent to minimizing surprise, which is just a negative log of p(o). If probability is 1 — surprise is 0, probability is 0 — surprise is infinite. Here is surprise as a function of probability:

And here is surprise -log p(o), overlaid with model evidence p(o):

As we discussed above for model evidence, to evaluate surprise -log p(o), we need to sum out the hidden variable s from the joint distribution p(o,s):

While this is just a summation over all possible values of s, it can be overly hard if there are many possible values of s (you'd have to sum up gazillion values). Fortunately, there is a trick to approach this impossible summation. Instead of computing surprise directly, we can approximate it by something that is close enough, but much easier to work with. First, we introduce a dummy distribution q over the space of s (which will turn out really useful later). We could safely bring it inside the summation by multiplying and dividing at the same time:

This gives us a weighted sum, where for each s, the ratio p(o,s)/q(s) is weighted by q(s). It is still the same surprise -log p(o), but now, we are in a good shape to replace it by an approximation. Now let's put surprise aside for a second, and reflect on a possible approximation. We know that the function of surprise -log looks like a valley or a bowl (i.e. it is convex). Colloquially, the definition says: the function is convex if, when you drop a stick on it, the stick gets trapped in the function like in a bowl. We express this idea formally as follows. Let's take 2 points in the function's domain (x and y), and pick a number between 0 and 1 (w). As pictured below, we can move between x and y by computing the weighted sum and changing w from 0 to 1. So you could consider w and (1–w) as parameters of a simple distribution. Actually, there is a short name for 'weighted sums, in which the weights are defined by a distribution' — expectation. Now back to the stick in a bowl: if you evaluate the function of this expectation, it will always be lower or equal to the expectation of the function evaluated at x and y.

Since we earlier expanded surprise as a function (-log) of the expectation (weighted sum of the ratio p(o,s)/q(s)), it looks suitable for our approximation, following this inequality of convexity! Formally, it's called 'an upper bound', a quantity that approaches from above. Lo and behold, this upper bound is... the Free Energy:

We could go a step further and remove the minus in front of the log. This brings us to the definition used in Active Inference papers.

The cool thing about Free Energy is that weights in the summation are defined by q(s), and we have a full control over q(s). So basically we can wiggle q(s) in a way that minimizes Free Energy. Using a couple of standard identities, Free Energy (usually appended with 'Variational' in the literature) can be decomposed in 2 equivalent ways.

While the right branch is usually used in practice, let's focus on the left one, as it gives us a nice theoretical insight [note: we can remove expectation (the sum over s of q(s)) in front of -log p(o), because p(o) does not depend on s, and q(s), as a probability distribution, sums up to 1]. It says that Free Energy is equal to KL (Kullback — Leibler) divergence between q(s) and p(s|o), and surprise -log p(o).

KL divergence quantifies how different are 2 distributions. For example, if a distribution p over coin landing head or tail is [0.5, 0.5] and another distribution q is also [0.5, 0.5], KL divergence is 0. So by minimizing Free Energy, the arbitrary distribution q(s) gets closer to the posterior p(s|o), and if they match, KL term becomes 0 and Free Energy is exactly equal to surprise. So the more we minimize Free Energy by wiggling q(s), the more it becomes closer to surprise (since it's an upper bound), and by minimizing Free Energy by wiggling parameters of p(o,s) we can minimize surprise even further. This is illustrated in detailed in the post on predictive coding, which is a corollary of the Free Energy principle applied to perception. Here is the summary so far: minimize approximation to surprise (free energy) -> avoid surprising observations -> avoid surprising states -> maintain homeostasis -> remain alive So far, we dealt with a static situation, with one hidden state and one set of observations, but the real world is dynamic. So we would have a hidden state at each point in time, and since things tend to depend on what has just happened before, we'll assume that s at a

certain time t depends on s at the previous time point. For example, a probability of seeing a rainbow depends directly on whether it rained before.

As before, we try to model the true generative process by learning a generative model p(o,s) and obtaining an approximation to the posterior q(s) at each time step. As in a simple static situation, we can find in which direction to change the parameters of p(o,s) and q(s) to decrease the Free Energy, and then make many little steps in that direction.

This would work, but we would be just passively observing the environment. What if we also act on it? In this case, s at a certain time t would depend on the s at the previous time point and our action u. In other words, action can directly affect the state of the world, so a different action would lead to a different future (e.g. we can physically move things by our actions).

Now the inference becomes active! We just have to find a way to choose a good action at each time step. In reality, it seems that we don't consider consequences of action at only next time step, but we plan the whole policy, a series of action, oriented towards goals that are distant in time. So if there are many possible actions and many future time points -> there are many potential policies we can undertake. Active inference says: just consider all of them. So we do the inference, approximating p(s|o) with q(s), simultaneously (in parallel) for every possible policy pi.

And since our goal is still the same - to minimize surprise by minimizing the Free Energy - we can compute the Free Energy (and the direction to change q(s) that will minimize FE) under each possible policy and time step. Policies that minimize Free Energy in the future are preferred. Let's say we plan for 10 time steps ahead. If we are at time t=1, for each policy, we sum up the Free Energies for time steps 1 to 10, and pick a policy that has a minimum cumulative Free Energy in the future. Let's see again the big picture of how Free Energy can be calculated:

The left branch shows us important theoretical properties of Free Energy minimization (e.g. that q(s) approaches p(s|o)), but it is impractical, since we use Free Energy to approximate -log p(o) in the first place. So let's look on the right branch, which is a standard way to compute Free Energy:

These 2 terms are usually called Complexity and Accuracy. Complexity indicates how much the approximation of the posterior q(s) deviates from the prior p(s), and quantifies how many extra bits of information that are not in the prior we want to encode in the approximate posterior q(s). Accuracy (expected p(o|s)) scores how likely are states s given a specific outcome o. While we can easily compute the Complexity term, there is a problem in estimating Accuracy for the future time steps — simply because we haven't observed the outcomes yet. Thus, we need another way to compute the Free Energy, so we'll go by the left branch of FE decomposition. But here is a catch: since we don't have access to future observations — we'd have to guess what they could look like, and take the weighted sum of the Free Energy over the guesses p(o|s). Note, we will condition probabilities on pi, since Free Energy is estimated separately for every policy. Only state-observation mapping p(o|s) is assumed to be the same for all policies.

0.1 Liza's part

Итак, нам нужно предсказывать как будущие состояния $q(s|\pi)$, так и наблюдения. А также оценивать свободную энергию на их основе. Давайте пройдем шаг за шагом, сначала сосредоточившись на знаменателе $p(o, s|\pi)$. Рассматривая левую ветвь разложения свободной энергии, показанную выше, мы можем преобразовать ее так же, как $p(s|o) \cdot p(o)$:

$$= \sum_{o,s} p(o_t|s_t)\ q(s_t|\pi) \log \frac{q(s_t|\pi)}{p(o_t, s_t|\pi)} \qquad p(a,b) = p(b|a)\ p(a)$$

$$= \sum_{o,s} p(o_t|s_t)\ q(s_t|\pi) \log \frac{q(s_t|\pi)}{p(s_t|o_t, \pi)\ p(\boldsymbol{o}_t)} \leftarrow \textit{prior preferences on future outcomes}$$

$p(o)$ - это априорные предпочтения относительно будущих наблюдений (the prior preference on the future outcomes) (которые пропорциональны вознаграждению в классическом обучении с подкреплением). Разделив логарифм, мы получаем следующие две величины: отрицательное эпистемическое (т.е. знание) значение (negative epistemic value) и ожидаемое предшествующее предпочтение (expected prior preference):

$$\textit{- Epistemic value}$$
$$= \sum_{o,s} p(o_t|s_t)\ q(s_t|\pi) \log \frac{q(s_t|\pi)}{p(s_t|o_t, \pi)} - \sum_{o,s} p(o_t|s_t)\ q(s_t|\pi) \log\ p(\boldsymbol{o}_t)$$

Короче говоря, эпистемическое значение (epistemic value) говорит нам, насколько будущие наблюдения могут уменьшить нашу неопределенность относительно аппроксимации апостериорного $q(s)$ (the approximate posterior q(s)). Давайте сначала закончим вывод, а затем обсудим его подробно. У нас есть истинное апостериорное $p(s|o, \pi)$ в знаменателе левого слагаемого, которое, как мы знаем, трудно вычислить, особенно в будущем (имейте в виду, что мы предсказываем будущие состояния и наблюдения). Таким образом, мы могли бы применить формулу Байеса, чтобы повторно выразить это соотношение в более вычислимых выражениях:

$$\textit{- Epistemic value}$$
$$= \sum_{o,s} p(o_t|s_t)\ q(s_t|\pi) \log \frac{q(s_t|\pi)}{p(s_t|o_t, \pi)} - \sum_{o,s} p(o_t|s_t)\ q(s_t|\pi) \log\ p(\boldsymbol{o}_t)$$

Bayes rule $\quad \frac{q(s_t|\pi)}{p(s_t|o_t, \pi)} = \frac{q(s_t|\pi)\ q(o_t|\pi)}{p(o_t|s_t, \pi)\ q(s_t|\pi)} = \frac{q(o_t|\pi)}{p(o_t|s_t, \pi)} \qquad p(a|b) = \frac{p(b|a)\ p(a)}{p(b)}$

$$= \sum_{o,s} p(o_t|s_t)\ q(s_t|\pi) \log \frac{q(o_t|\pi)}{p(o_t|s_t, \pi)} - \sum_{o,s} p(o_t|s_t)\ q(s_t|\pi) \log\ p(\boldsymbol{o}_t)$$
$$\textit{- Epistemic value}$$

Это было бы легче сделать, если бы вы пренебрегли зависимостью от π. Тогда $q(o)$ это вроде маргинального распределения (marginal likelihood), $p(o|s)$ – вероятность (likelihood) и $q(s_t|\pi)$ –

априорное распределение (?). Вы также можете заметить, что для некоторых распределений «p» заменяется на «q». Это результат аппроксимации, так как, например, $p(s|o,\pi)$ является истинным апостериорным распределением, которое мы можем аппроксимировать с помощью $q(s|o,\pi)$, что при разложении приведет к тому, что все компоненты формулы Байеса будут формой q. Также обратите внимание, что $q(o|s,\pi)$ это то же самое, что и $p(o|s,\pi)$, поскольку это все еще относится к той же вероятности (likelihood).

Мы также могли бы объединить логарифмы и снова разделить их по-другому, что приводит нас к окончательной форме ожидаемой свободной энергии:

$$= \sum_{o,s} p(o_t|s_t)\, q(s_t|\pi) \log \frac{q(o_t|\pi)}{p(\boldsymbol{o_t})\, p(o_t|s_t,\pi)} \quad \longleftarrow \textit{predicted outcomes}$$

$$= \sum_{o,s} q(s_t|\pi)\; p(o_t|s_t)\; \log \frac{q(o_t|\pi)}{p(\boldsymbol{o_t})} - \sum_{s} q(s_t|\pi) \sum_{o} p(o_t|s_t) \log p(o_t|s_t)$$

$$KL(p(a)||q(a)) = \sum_{a} p(a) \log \frac{p(a)}{q(a)} \qquad H[p(a)] = -\sum_{a} p(a) \log p(a)$$

$$= KL(q(o_t|\pi)||p(\boldsymbol{o_t})) + \sum_{s} q(s_t|\pi) H[p(o_t|s_t)]$$

$$\textit{Expected cost} \qquad \textit{Expected Ambiguity}$$

Примечание: на левой стороне мы можем повторно выразить $q(s|\pi)p(o|s)$ как совместное распределение $q(s,o)$ и просуммировать по всем s, чтобы получить $q(o)$. Это даст нам ожидаемую дивергенцию Кульбака-Лейблера. А на правой стороне мы можем удалить π из $p(o|s,\pi)$, потому что вероятность (likelihood) одинакова для каждой политики.

Левое слагаемое (называемое "Издержки"или "Риск") – это дивергенция Кульбака-Лейблера между двумя распределениями: ожидаемыми в рамках политики π наблюдениями $q(o|\pi)$ и априорными предпочтениями (prior preferences). Таким образом, минимизация ожидаемой свободной энергии будет способствовать политике, которая приведет к наблюдениям, которые нам нравятся. И правое слагаемое, неоднозначность (Ambiguity), количественно определяет, насколько неопределенным является отображение между состоянием и наблюдениями $p(o|s)$. И это окончательная формула свободной энергии в будущем. Давайте теперь посмотрим на эпистемическое значение (слагаемое, которое появилось раньше, окрашено в желтый цвет).

Эпистемическое значение говорит нам, как много мы могли бы извлечь из окружающей среды, если бы следовали этой политике. Так происходит потому, что эпистемическое значение представляет собой взаимную информацию (mutual information) между скрытыми состояниями s и ожидаемыми наблюдениями o. Взаимная информация количественно определяет, насколько неопределенность (H) по одной переменной уменьшается, если мы знаем другую.

$$MI(a,b) \begin{matrix} = H[p(a)] - H[p(a|b)] \\ = H[p(b)] - H[p(b|a)] \end{matrix} \qquad H[p(a)] = -\sum_{a} p(a) \log p(a)$$

Аналогично, взаимная информация может быть повторно выражена как дивергенция Кульбака-Лейблера между совместным распределением двух переменных (если взаимная информация велика, то знание одной переменной говорит нам много о распределении другой) и произведением их маргинальных плотностей (marginals) (как если бы они были полностью независимы).

Нам просто нужно перевернуть номинатор и знаменатель (потому что эпистемическое значение отрицательно в уравнениях). Обратите внимание, что мы можем удалить π из $p(o|s,\pi)$, потому что вероятность одинакова для каждой политики.

- Epistemic value

$$\sum_{o,s} p(o_t|s_t)\, q(s_t|\pi) \log \frac{q(o_t|\pi)}{p(o_t|s_t,\pi)}$$

Mutual information

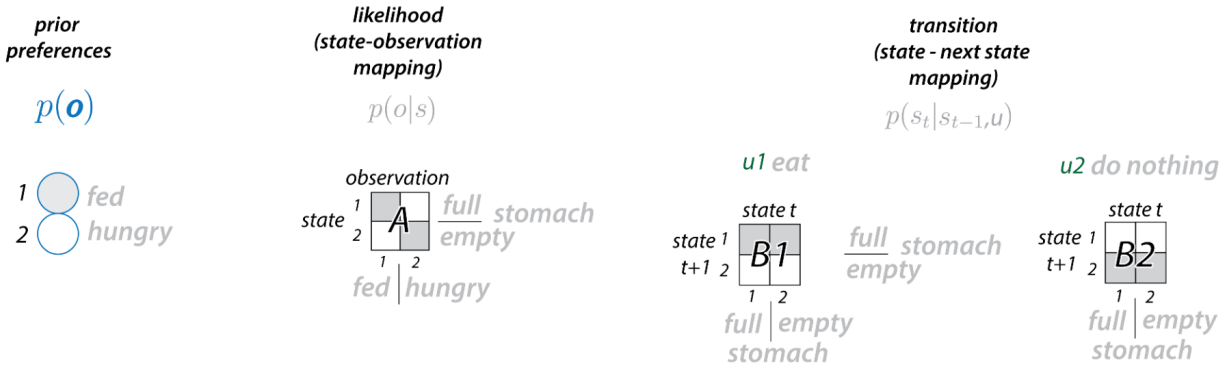
$$MI(a,b) = \sum_{a,b} p(a,b) \log \frac{p(a,b)}{p(a)\, p(b)}$$

$$-\log(a) = \log(\frac{1}{a})$$

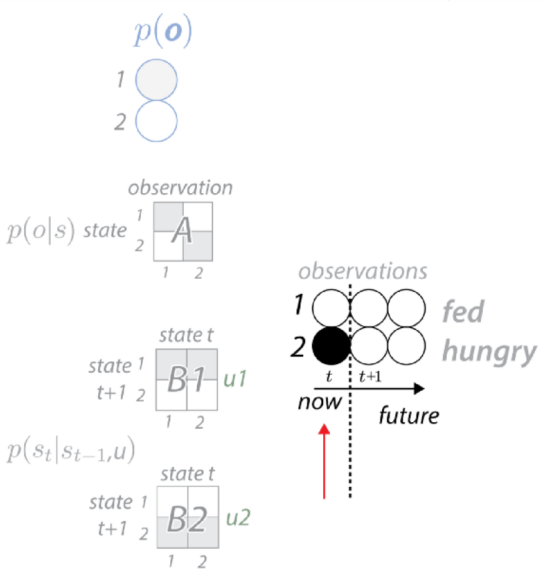
$$= \sum_{a,b} p(a|b)\, p(b) \log \frac{p(a|b)\, p(b)}{p(a)\, p(b)} = \sum_{a,b} p(a|b)\, p(b) \log \frac{p(a|b)}{p(a)}$$

На практике эпистемическое значение зависит от неопределенности относительно будущих состояний $q(s|\pi)$. Если вы абсолютно уверены, тогда $H[q(s|\pi)]$ невелико, вам больше нечему учиться, поэтому эпистемическая ценность будет низкой. Но если вы не уверены, $H[q(s|\pi)]$ высоко, и существует сильная зависимость между состояниями и наблюдениями (поэтому $H[q(s|o)]$ низок), то взаимная информация будет высокой (см. формулы взаимной информации в терминах энтропий выше). Надеюсь, что прочитав приведенным описаниям несколько раз и пройдя их самостоятельно, оно станет простым.

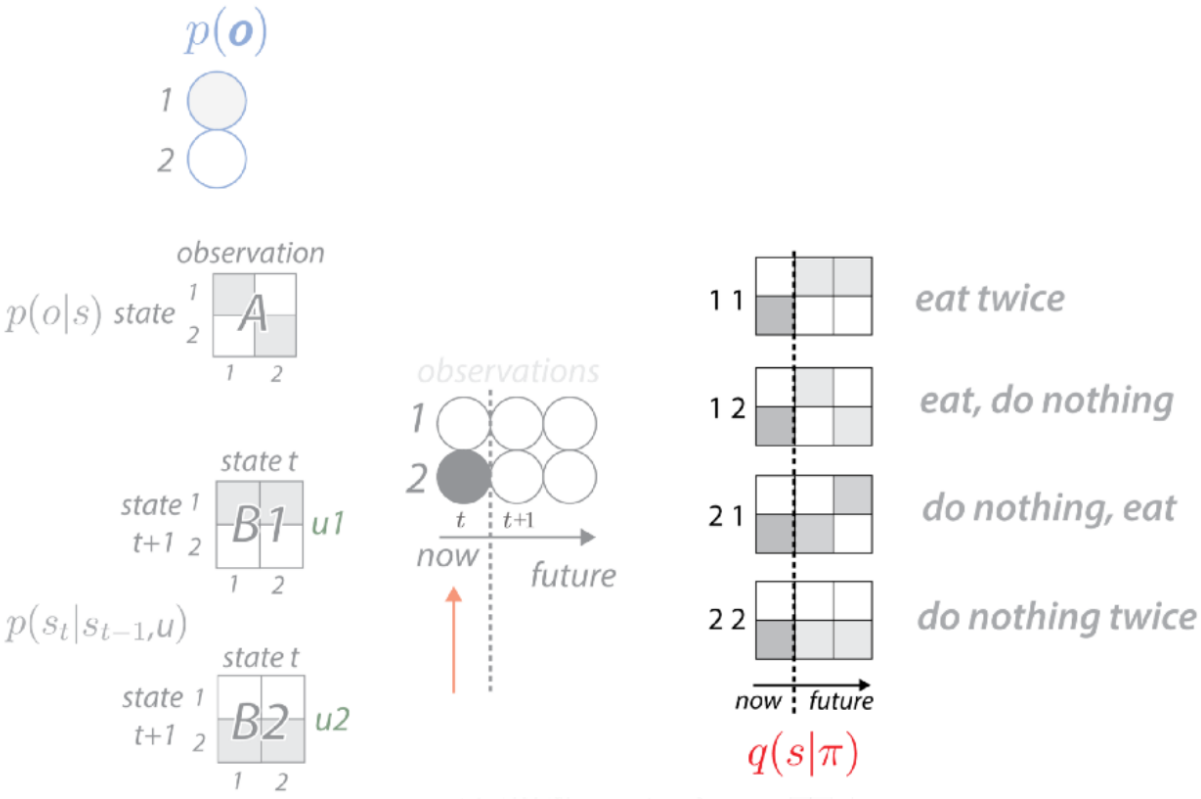
Вот небольшой пример того, что на самом деле произошло бы в мозгу агента, если бы он использовал активное умозаключение (active inference). Для простоты предположим, что окружающая среда имеет только два состояния s: «1» и «2», например, есть пища в вашем желудке (1) или нет (2). Аналогично, есть только два возможных наблюдения : «1» и «2», вы чувствуете себя сытым (1) или голодным (2). Предположим, что мы уже знаем параметры генеративной модели (generative model) $p(o,s)$. Вероятность (называемая матрицей A) $p(o|s)$ сопоставляет состояния с наблюдениями – если у вас есть еда – вас кормят и наоборот. Вероятность перехода $p(s_t|s_{t-1},u)$ отображает предыдущее состояние в следующее. Но поскольку переход также зависит от действия u, мы можем выразить его в виде отдельной матрицы переходов (B) для каждого действия. Предположим, что мы можем либо пойти за едой (u_1), либо ничего не делать (u_2). Так что если мы выберем u_1 – у нас будет еда в следующем состоянии, независимо от того, есть ли она у нас сейчас, и наоборот. Наконец, у нас также есть предварительные предпочтения $p(o)$ – мы любим, чтобы нас кормили и не голодали, поэтому мы приписываем более высокую вероятность наблюдению «1» – кормили. Другими словами, Мы выражаем предпочтения по отношению к наблюдениям как вероятность $p(o)$.



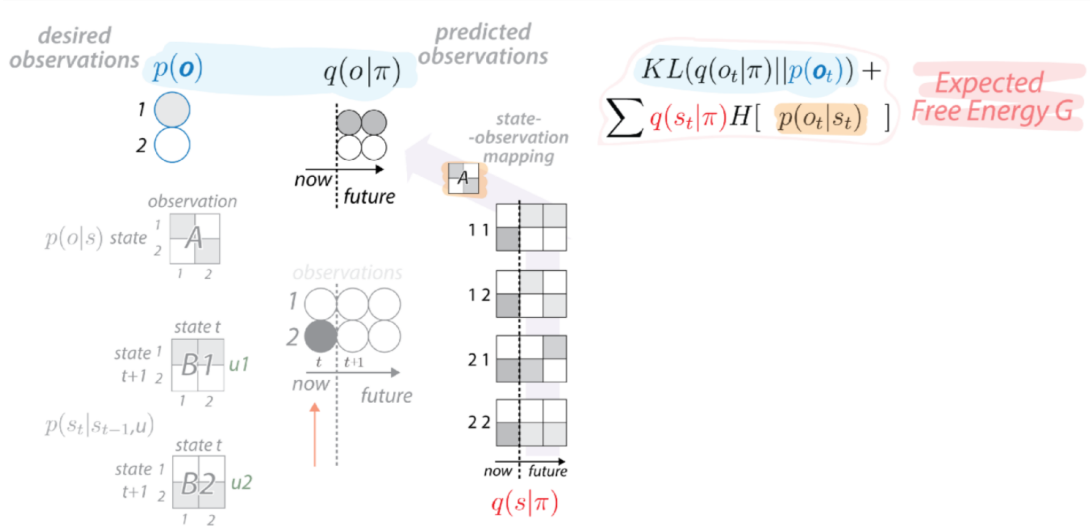
Теперь представьте, что мы наблюдаем, что мы голодны, и должны планировать свои действия на два шага вперед.



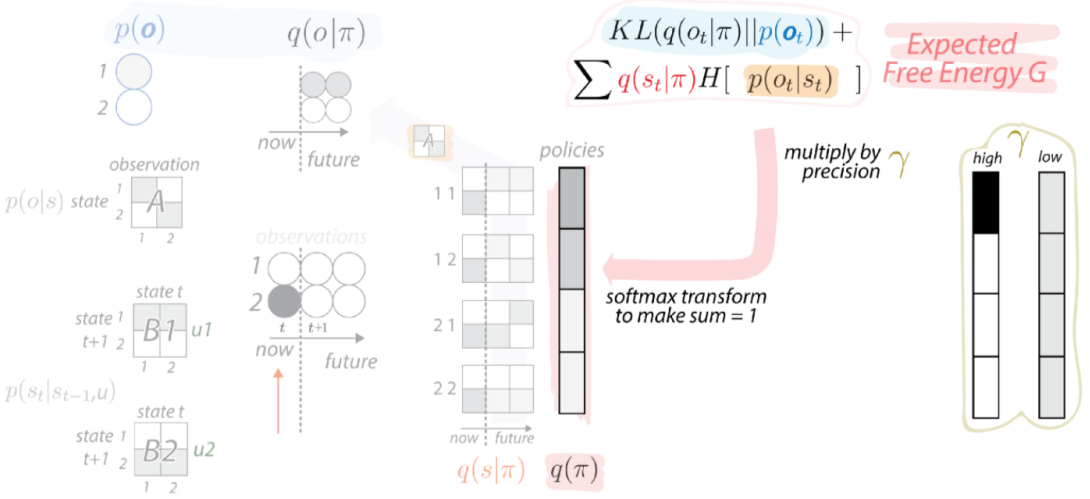
Поскольку есть только два возможных действия и два шага, мы можем оценить все возможные политики: $1-1$ (идти за едой оба раза), $1-2$, $2-1$, $2-2$. На самом деле мы будем оценивать апостериорное распределение по будущим состояниям (будет ли пища находиться в нашем желудке), при каждой из этих политик:



Поскольку мы знаем, как состояния связаны с наблюдениями ($p(o|s)$, матрица A), мы можем оценить прогнозируемое наблюдение для каждой политики $q(o|\pi)$ и вычислить дивергенцию Кульбака-Лейблера (заштрихована синим цветом) ожидаемой свободной энергии (??). Аналогично, мы также можем оценить неоднозначность (Ambiguity) (заштрихована оранжевым цветом), который зависит от $p(o|s)$:

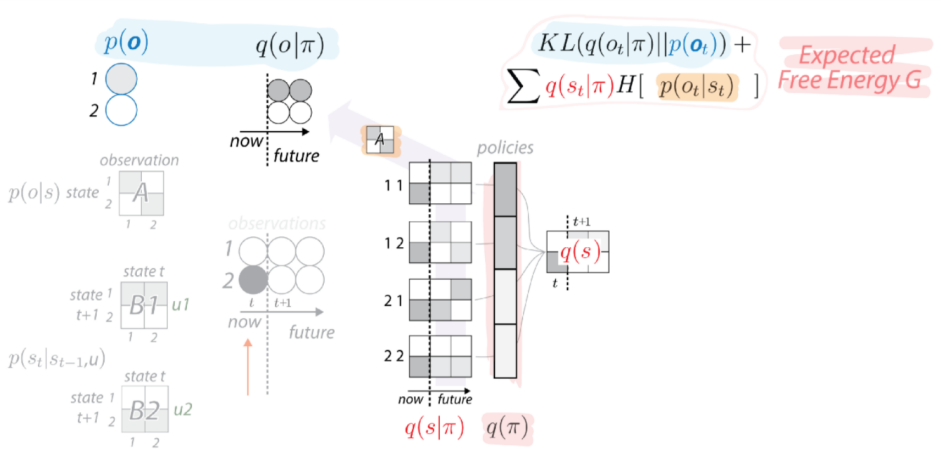


Затем мы суммируем ожидаемую свободную энергию по будущим действиям и преобразуем ее в распределение вероятностей для политики (probability distribution over policy) $q(\pi)$. Так что чем меньше свободная энергия, тем выше вероятность проведения политики. Интересно, что при преобразовании свободная энергия взвешивается (умножается) на точность (precision), которая определяет, насколько мы уверены в своих убеждениях по отношению к политике (то есть, изменяя точность до ее крайних значений, наши убеждения могут разрушаться на одной политике или распространяться равномерно (our beliefs can collapse on a single policy or spread uniformly)). Это важно для определения исследования/эксплуатации (exploration/exploitation), так как чем больше вы уверены в том, что у вас есть хорошая политика (т. е. высокая точность), тем меньше вы исследуете и наоборот.

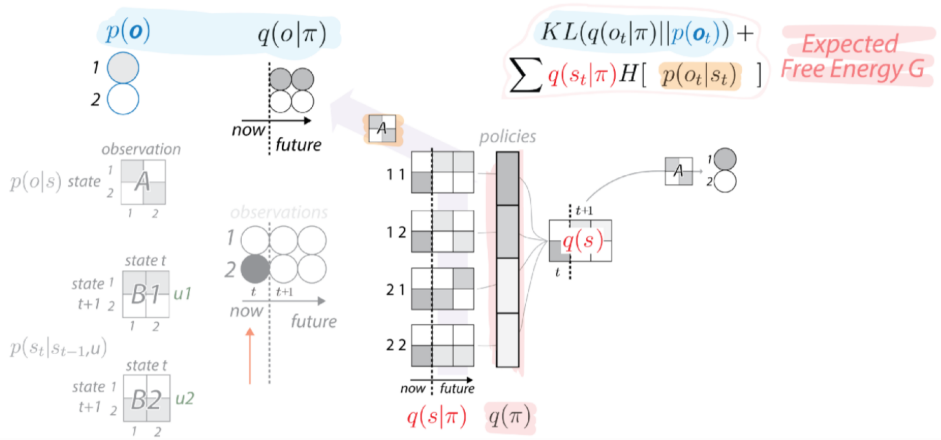


Теперь вы можете просто выбрать политику, которая максимально минимизирует (?) свободную энергию, но есть более аккуратный способ: вместо выбора 1 политики мы будем делать усреднение по ним. Подумайте об этом так, как если бы в определенный момент времени существовала одна политика, которая давала бы самую низкую свободную энергию, а другие политики были бы просто немного хуже. По мере того как вы будете получать новые наблюдения с течением времени, может оказаться, что лучшая в мире политика была среди тех, которые были "немного хуже"раньше. Но что делать, если вы уже выбрали действие, которое не позволяет вам следовать этой недавно обнаруженной политике? Таким образом, мы позволяем каждой политике голосовать во время выбора действий, причем политики, которые максимально ми-

нимизируют свободную энергию, имеют наибольший вес. Мы берем математическое ожидание $q(s|\pi)$ при известном $q(\pi)$ (expectation of $q(s|\pi)$ under $q(\pi)$) – взвешенная сумма, где веса определяются вероятностью каждой политики. Это приводит к маргинальному распределению $q(s)$, которое неявно включает политику (и, следовательно, ожидаемую свободную энергию).

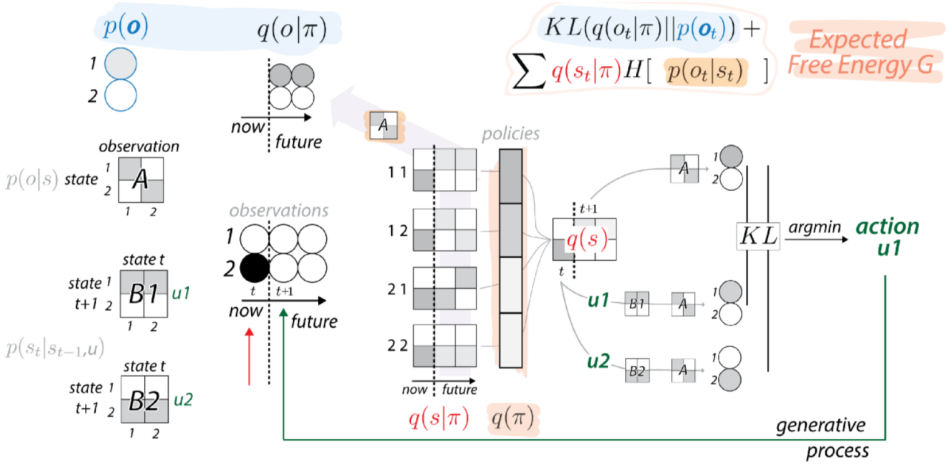


И вот теперь мы готовы действовать! Мы выбираем действие, которое минимизирует разницу (дивергенцию Кульбака-Лейблера) между тем, что мы ожидаем увидеть на следующем временном шаге, и тем, что мы должны были бы увидеть, если бы выбрали конкретное действие. Сперва чтобы получить (вероятность) ожидаемых наблюдений, мы умножаем наше убеждение о скрытом состоянии на следующем временном шаге $q(s_{t+1})$ на $p(o|s)$, сопоставляющее состояния и наблюдения (матрица A) (by the state-observation mapping p(o|s) ('A' matrix)) :



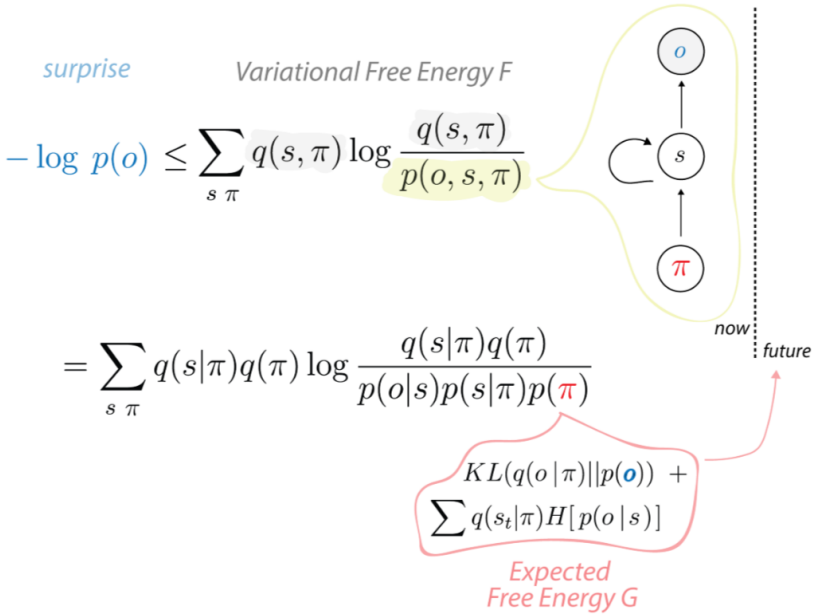
И чтобы получить ожидаемые наблюдения, если мы должны были предпринять определенные действия, мы умножаем нашу веру в скрытое состояние в текущий момент времени $q(s_t)$ на переходную матрицу B для данного действия, чтобы получить гипотетическое следующее состояние, а затем на матрицу A, сопоставляющую состояния и наблюдения, чтобы получить гипотетическое наблюдение (And to get the expected observations if we were to take a certain action, we take our belief on the hidden state at the current time step q(s_t), multiply it by the action-specific state-transition matrix B (to get the hypothetical next state), and then by state-observation matrix A (to get the hypothetical observation)). Выполняется действие, минимизирующее дивергенцию Кульбака-Лейблера, генеративный процесс возвращает нам следующее наблюдение в момент времени $t + 1$, и процесс начинается снова. И это все.

Итак, мы никогда не моделируем действие непосредственно, а вместо этого: оцениваем скрытые состояния при каждой политике → оцениваем ожидаемую свободную энергию для каждой



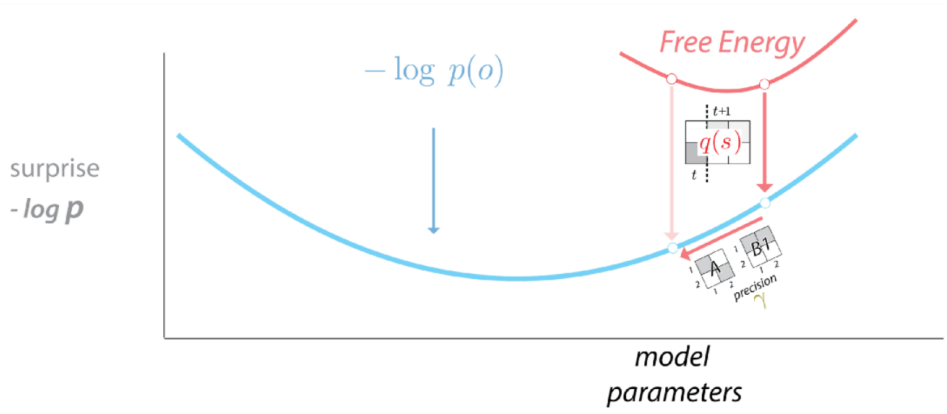
политики → преобразуем ее в вероятность политики → усредняем состояния для всех политик и, наконец, → действуем так, как будто мы наблюдаем то, что мы ожидаем.

Формально связывание вероятности каждой политики с ожидаемой свободной энергией G осуществляется следующим образом. Модель $p(o, s)$ фактически также включает в себя политику, больше похожую на $p(o, s, \pi)$. Аналогично, аппроксимация апостериорного q также включает политику и выглядит как $q(s, \pi)$. Решение для оптимальной политики показано аналитически, и в дополнение к ожидаемой свободной энергии G, включает в себя предыдущие предпочтения (prior preferences) по политике $p(\pi)$ и свободную энергию с прошлых временных шагов (где мы фактически можем оценить точность, так как наблюдения уже известны). Однако, если мы предположим, что все политики в прошлом одинаковы – состоят из уже выполненных действий – аппроксимация апостериорного распределения политики $q(\pi)$ действительно зависит только от ожидаемой свободной энергии G в будущем. Минимизация свободной энергии в будущем – это «сильно закодированный» априор («hard-coded» prior), потому что вы связываете вероятность выбора определенной политики с ее ожидаемой свободной энергией, и это дополнительное предположение, которое вы должны сделать. Это приводит к минимизации одной свободной энергии (G) внутри другой (F). Вот вариационная свободная энергия с совместным распределением, развернутым во второй строке (Here is the variational FE with the joint expanded in the second line):



Наконец, в дополнение к выводу, мы также изучаем параметры модели, такие, как сопостав-

ление состояний и наблюдений (матрица A), переход от состояния к состоянию (матрица) и точность (γ). Таким образом, минимизируя свободную энергию относительно параметров модели, мы не только делаем свободную энергию лучшим приближением неожиданности (surprise), но и минимизируем саму неожиданность.



Мы могли бы либо найти точечную оценку параметров, либо искать все распределение. В последнем случае мы включаем неопределенность параметров в нашу модель p, а также в аппроксимацию апостериорного q. Чтобы получить обоснованность модели/предельное правдоподобие/неожиданность (model evidence/marginal likelihood/surprise), мы затем суммируем не только латентные переменные, но и параметры. Таким образом, технически, поскольку мы не убрали параметры в этом посте, мы работали с вероятностью параметров $p(o|parameters, model)$ (likelihood of the parameters), но не с обоснованностью модели $p(o|model)$ (model evidence).

Конечно, есть и некоторые ограничения. Например, мы предполагаем, что аппроксимации апостериорного распределения $(q_t|\pi)$ независимы в каждый момент времени, и что они раскладываются на латентные переменные и параметры. Но что еще более важно, масштабируемость этой схемы ограничена, и до сих пор моделирование включает простые ситуации с небольшими пространствами состояний и наблюдений, так что вычисления (например, определение априорных предпочтений (aprior preferences) $p(o)$, что трудно, если существует много возможных наблюдений) остаются отслеживаемыми. Однако эта схема скорее предназначена для доказательства принципа, и ее вполне достаточно, чтобы быть полезной в вычислительной психиатрии. Например, считается, что точность (precision) отражает функцию дофамина, а это означает, что неспособность определить оптимальную точность будет иметь неблагоприятные психопатологические последствия. Аналогично, в недавнем исследовании использовался активный вывод (active inference) с акцентом на мотивацию, предполагая, что люди имеют либо целенаправленные, либо однородные априорные предпочтения (prior preferences) $p(o)$. Хотя эта схема была применена к игре DOOM в OpenAI Gym, она по существу сводится к тому же самому принципу, обсуждаемому здесь, поскольку пространство состояний было дискретизировано до максимума из 10 возможных состояний.