Mathematical Modeling
(for psychology students…)

Approximate inference (3)—Variational Inference

张洳源

2024/04/11

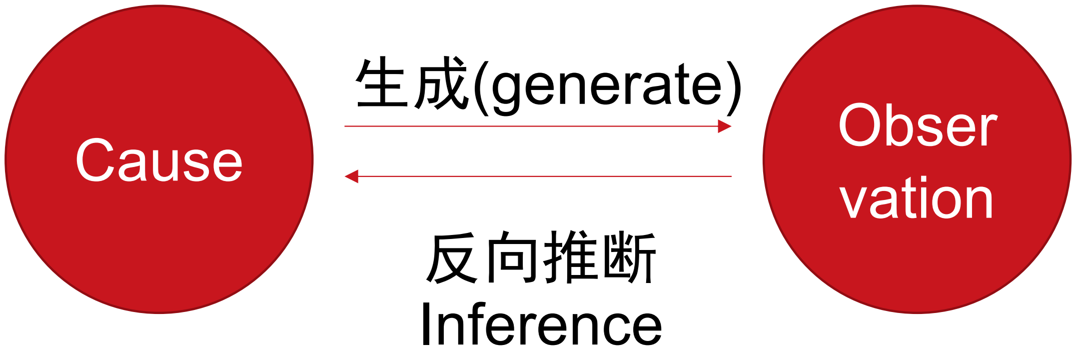
上海交通大学心理与行为科学研究院

上海交通大学医学院附属精神卫生中心

Generative Process and Reverse Inference

- Generative process (由原因生成观察数据)

- Reverse Inference (由观察数据反推原因)



频率学派 vs. 贝叶斯学派

频率学派 ← 最大似然估计

参数点估计

Exact Inference

最大后验估计

贝叶斯学派

马尔科夫链蒙特卡洛采样 (MCMC)

变分推断 (Variational Inference)

贝叶斯参数估计

Approximate Inference

马尔科夫链蒙特卡洛采样 (MCMC)和变分推断(Variational Inference)后面的内容会说到

Outlines

- What is variational Inference
- Variational Inference for parameter estimation
- Variational Inference as a Cognitive Theory

Bayesian Theory

$$\overset{\overset{\text{后验分布}}{\uparrow}}{p(\theta|d)} = \frac{\overset{\overset{\text{似然函数}}{\uparrow}}{p(d|\theta)} * \overset{\overset{\text{先验分布}}{\uparrow}}{p(\theta)}}{p(d) \longrightarrow \text{归一化因子}}$$

有时候写成

$$p(\theta|d) \propto p(d|\theta) * p(\theta)$$

认知心理学中概率模型的本质就是求解后验概率分布$p(\theta|d)$

张洳源 ruyuanzhang@sjtu.edu.cn

MCMC

- MCMC 是通过对后验概率采样的方式来求得后验概率

Variational Inference

目的是求 $p(\phi|s)$

ϕ 是cause, s是data

$$p(\phi|s) = \frac{p(s,\phi)}{p(s)} \tag{1}$$

$$p(s) = \frac{p(s,\phi)}{p(\phi|s)} \tag{2}$$

$$log(p(s)) = log(p(\phi,s)) - log(p(\phi|s)) \tag{3}$$

公式右边引入一个分布 $q(\phi)$, 注意无论$q(\phi)$是什么上式还是成立

$$log(p(s)) = log(\frac{p(\phi,s)}{q(\phi)}) - log(\frac{p(\phi|s)}{q(\phi)}) \tag{4}$$

在公式两边，我们对$q(\phi)$求期望

$$\int log(p(s))q(\phi)d\phi = \int log(\frac{p(\phi,s)}{q(\phi)})q(\phi)d\phi - \int log(\frac{p(\phi|s)}{q(\phi)})q(\phi)d\phi \tag{5}$$

$$log(p(s)) = \int log(\frac{p(\phi,s)}{q(\phi)})q(\phi)d\phi - \int log(\frac{p(\phi|s)}{q(\phi)})q(\phi)d\phi \tag{6}$$

Variational Inference

$$log(p(s)) = \int log(\frac{p(\phi, s)}{q(\phi)})q(\phi)d\phi \boxed{- \int log(\frac{p(\phi|s)}{q(\phi)})q(\phi)d\phi} \quad (6)$$

$q(\phi)$ 和$p(\phi|s)$ KL divergence: $D_{KL}[q(\phi)||p(\phi|s) = \int log(\frac{q(\phi)}{p(\phi|s)})q(\phi)d\phi$

$$L = \int log(\frac{p(\phi, s)}{q(\phi)})q(\phi)d\phi$$

$$log(p(s)) - L = D_{KL}[q(\phi)||p(\phi|s)$$

我们的要求的是后验概率$p(\phi|s)$，因为这个概率分布很难求，我们引入另外一个关于ϕ的分布$q(\phi)$，那么我们只要让$q(\phi)$ 尽可能的去近似$p(\phi|s)$，那么$q(\phi)$ 就可以代替$p(\phi|s)$成为我们要求的解

Variational Inference

$q(\phi)$ 和$p(\phi|s)$ KL divergence: $D_{KL}[q(\phi)||p(\phi|s)] = \int log(\frac{q(\phi)}{p(\phi|s)})q(\phi)d\phi$

$$L = \int log(\frac{p(\phi,s)}{q(\phi)})q(\phi)d\phi$$ ⟶ Evidence Lower Bound (ELBO)

$$log(p(s)) - L = D_{KL}[q(\phi)||p(\phi|s)$$

我们的要求的是后验概率$p(\phi|s)$，因为这个概率分布很难求，我们引入另外一个关于ϕ的分布$q(\phi)$，那么我们只要让$q(\phi)$ 尽可能的去近似$p(\phi|s)$，那么$q(\phi)$ 就可以代替$p(\phi|s)$成为我们要求的解

让$q(\phi)$ 尽可能的去近似$p(\phi|s)$，即最小化KL divergence $\mathrm{D_{KL}}[q(\phi)||\ p(\phi|s)]$，因为log(p(s))是一个定值，这就转化成L最大。这里的L就叫做Evidence Lower Bound (ELBO)

Variational Inference

$q(\phi)$ 和$p(\phi|s)$ KL divergence: $D_{KL}[q(\phi)||p(\phi|s)] = \int log(\frac{q(\phi)}{p(\phi|s)})q(\phi)d\phi$

$$L = \int log(\frac{p(\phi,s)}{q(\phi)})q(\phi)d\phi$$ ⟶ Evidence Lower Bound (ELBO)

$$log(p(s)) - L = D_{KL}[q(\phi)||p(\phi|s)$$

我们的要求的是后验概率$p(\phi|s)$，因为这个概率分布很难求，我们引入另外一个关于ϕ的分布$q(\phi)$，那么我们只要让$q(\phi)$ 尽可能的去近似$p(\phi|s)$，那么$q(\phi)$ 就可以代替$p(\phi|s)$成为我们要求的解

让$q(\phi)$ 尽可能的去近似$p(\phi|s)$，即最小化KL divergence $D_{KL}[q(\phi)|| p(\phi|s)]$，因为log(p(s))是一个定值，这就转化成L最大。这里的L就叫做Evidence Lower Bound (ELBO)

$$\begin{aligned} q(\hat{\phi}) &= \underset{q(\phi)}{argmax} L \\ &= \underset{q(\phi)}{argmax} \int log(\frac{p(\phi,s)}{q(\phi)})q(\phi)d\phi \\ &= \underset{q(\phi)}{argmax} \int (logp(s|\phi) + logp(\phi) - logq(\phi))q(\phi)d\phi \end{aligned}$$

Variational Inference

$q(\phi)$ 和$p(\phi|s)$ KL divergence: $D_{KL}[q(\phi)||p(\phi|s)] = \int log(\frac{q(\phi)}{p(\phi|s)})q(\phi)d\phi$

$$L = \int log(\frac{p(\phi, s)}{q(\phi)})q(\phi)d\phi \longrightarrow \text{Evidence Lower Bound (ELBO)}$$

$$log(p(s)) - L = D_{KL}[q(\phi)||p(\phi|s)$$

让$q(\phi)$ 尽可能的去近似$p(\phi|s)$，即最小化KL divergence $D_{KL}[q(\phi)|| p(\phi|s)]$最小，因为log(p(s))是一个定值，这就转化成L最大。这里的L就叫做Evidence Lower Bound (ELBO)

$$\begin{aligned} q(\hat{\phi}) &= \underset{q(\phi)}{argmax} L \\ &= \underset{q(\phi)}{argmax} \int log(\frac{p(\phi, s)}{q(\phi)})q(\phi)d\phi \\ &= \underset{q(\phi)}{argmax} \int (\underbrace{logp(s|\phi)}_{\text{likelihood}} + \overbrace{logp(\phi)}^{\text{prior}} - \underbrace{logq(\phi)}_{\text{Variational distribution}})q(\phi)d\phi \end{aligned}$$

Variational Inference

最后的最后总结

目的是求 $p(\phi|s)$

最大化 $= \underset{q(\phi)}{argmax} \int (logp(s|\phi) + logp(\phi) - logq(\phi))q(\phi)d\phi$

这里有个积分很难求怎么办？？ 蒙特卡洛法暴力解决问题！！！ 从$q(\phi)$里面采样然后做近似

$$L = \int (logp(s|\phi) + logp(\phi) - logq(\phi))q(\phi)d\phi$$

$$\approx \frac{1}{N}\sum_{l=1}^{N} logp(s|\phi)^{(l)} + logp(\phi)^{(l)} - logq(\phi)^{(l)}$$

$$q(\hat{\phi}) = \underset{q(\phi)}{argmax} L$$

$$\approx \underset{q(\phi)}{argmax} \frac{1}{N}\sum_{l=1}^{N} logp(s|\phi)^{(l)} + logp(\phi)^{(l)} - logq(\phi)^{(l)}$$

$$= \underset{q(\phi)}{argmin} \frac{1}{N}\sum_{l=1}^{N} logq(\phi)^{(l)} - logp(s|\phi)^{(l)} - logp(\phi)^{(l)}$$

利用变分推断估计一个难估计的概率VI1.ipynb

Variational Inference

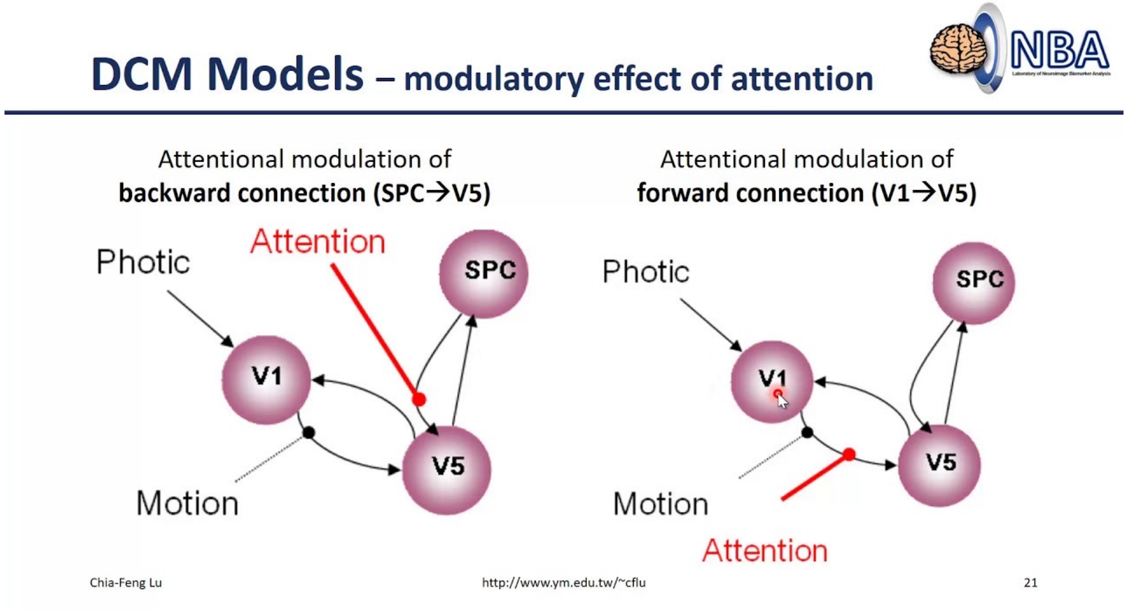
为什么我们需要Variational Inference?

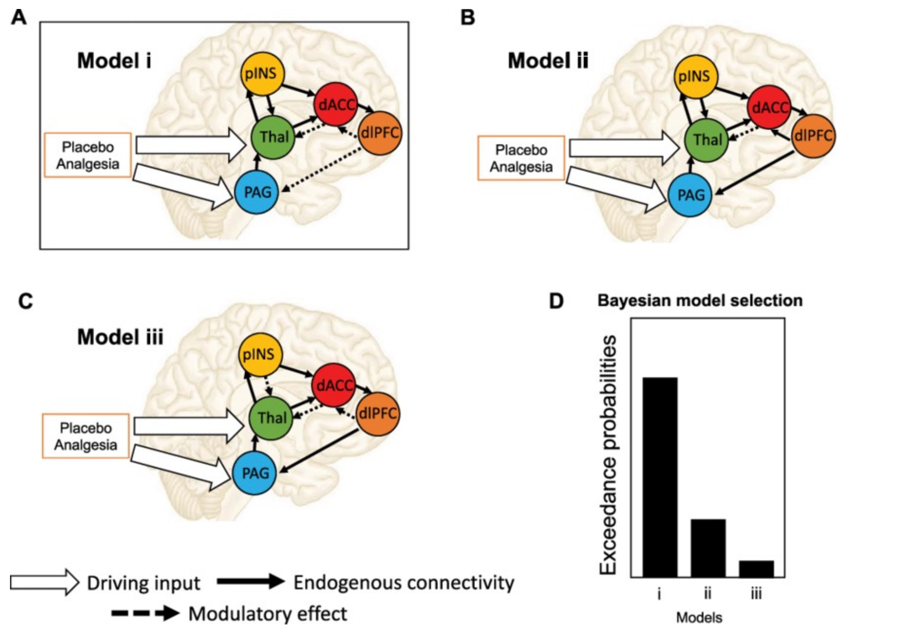
- 从效率上来讲，Variational Inference要比MCMC快得多
  - 在目前真实的machine learning研究中，几乎没有人用MCMC，实在太慢了。。

- 从理论上讲，很多时候我们永远无法知道真实的generative model是什么。。

Variational Inference for parameter estimation

- e.g., Dynamic causal modeling in neuroimaging





Variational Inference as a cognitive theory

- Free-energy principle

The “Bayesian brain” Hypothesis

- 如果人脑是按照贝叶斯推断的方式来完成各项认知功能，那么我们脑子里需要表征prior distribution，likelihood function和posterior distribution

- 人脑如何表征这些成分是当前计算神经科学最前沿的研究之一

$$p(\theta|d) = \frac{p(d|\theta) * p(\theta)}{p(d)}$$

The Free-Energy Theory

The Genius Neuroscientist Who Might Hold the Key to True AI

Karl Friston's free energy principle might be the most all-encompassing idea since the theory of natural selection. But to understand it, you need to peer inside the mind of Friston himself.

The Free-Energy Theory

Karl John Friston FRS FMedSci FRSB (born 12 July 1959) is a British neuroscientist and theoretician at University College London. He is an authority on brain imaging and theoretical neuroscience, especially the use of physics-inspired statistical methods to model neuroimaging data and other random dynamical systems.[2][4][5][6][7][8][9][10] Friston is a key architect of the free energy principle and active inference. In imaging neuroscience he is best known for statistical parametric mapping and dynamic causal modelling. In October 2022, he joined VERSES Inc, a California-based cognitive computing company focusing on artificial intelligence designed using the principles of active inference, as Chief Scientist.

Friston is one of the most highly cited living scientists[11] and in 2016 was ranked No. 1 by Semantic Scholar in the list of top 10 most influential neuroscientists.[12]

The Free-Energy Theory

nature reviews neuroscience

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature reviews neuroscience > review articles > article

Review Article | Published: 13 January 2010

The free-energy principle: a unified brain theory?

Karl Friston

Nature Reviews Neuroscience **11**, 127–138 (2010) | Cite this article

71k Accesses | **3680** Citations | **654** Altmetric | Metrics

Home > Biological Cybernetics > Article

Action and behavior: a free-energy formulation

Original Paper | Open access | Published: 11 February 2010

Volume 102, pages 227–260, (2010) Cite this article

Download PDF

You have full access to this open access article

Karl J. Friston, Jean Daunizeau, James Kilner & Stefan J. Kiebel

Cell Press Trends in Cognitive Sciences *Supports open access*

This journal Journals Publish News & events About

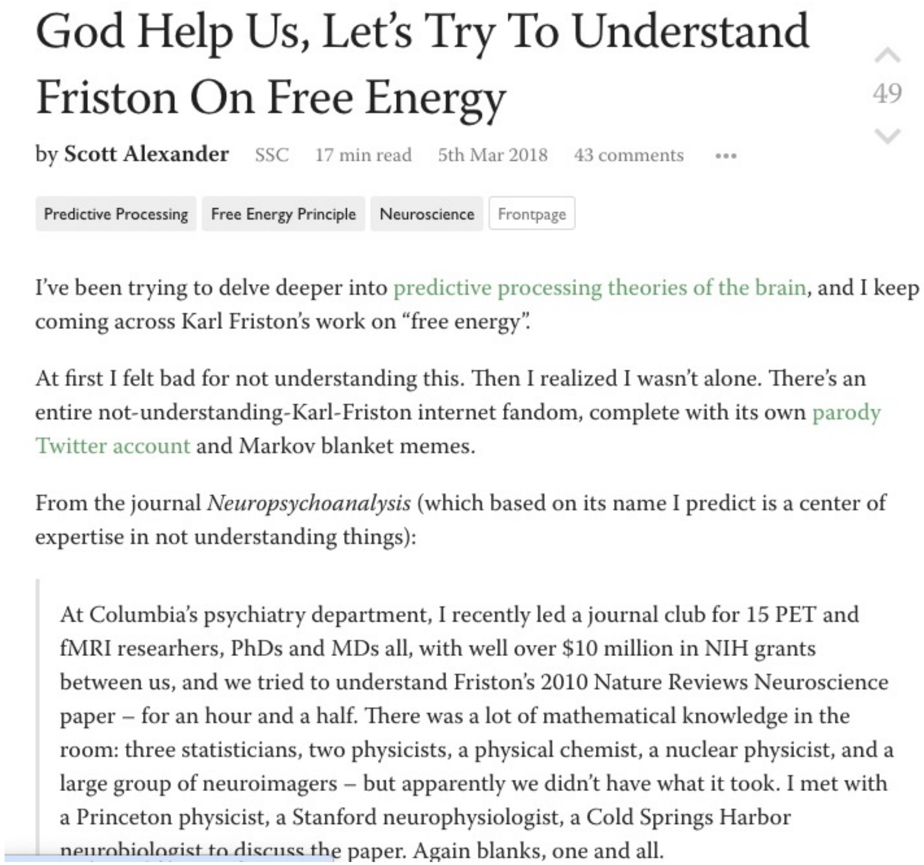
OPINION | VOLUME 13, ISSUE 7, P293-301, JULY 2009 Download Full Issue

The free-energy principle: a rough guide to the brain?

Karl Friston

Published: June 25, 2009 • DOI: https://doi.org/10.1016/j.tics.2009.04.005

The Free-Energy Theory



God Help Us, Let's Try To Understand Friston On Free Energy

49

by **Scott Alexander** SSC 17 min read 5th Mar 2018 43 comments •••

Predictive Processing | Free Energy Principle | Neuroscience | Frontpage

I've been trying to delve deeper into predictive processing theories of the brain, and I keep coming across Karl Friston's work on "free energy".

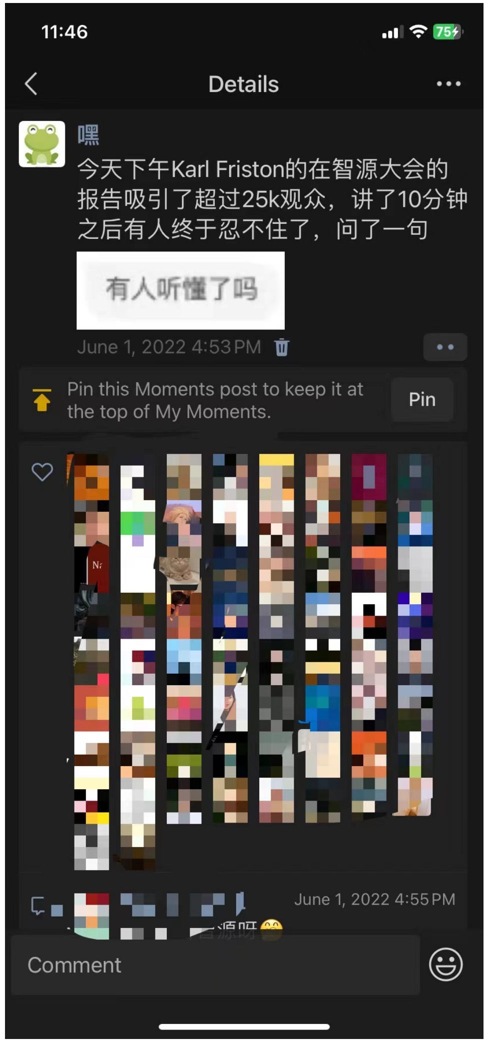
At first I felt bad for not understanding this. Then I realized I wasn't alone. There's an entire not-understanding-Karl-Friston internet fandom, complete with its own parody Twitter account and Markov blanket memes.

From the journal *Neuropsychoanalysis* (which based on its name I predict is a center of expertise in not understanding things):

> At Columbia's psychiatry department, I recently led a journal club for 15 PET and fMRI researhers, PhDs and MDs all, with well over $10 million in NIH grants between us, and we tried to understand Friston's 2010 Nature Reviews Neuroscience paper – for an hour and a half. There was a lot of mathematical knowledge in the room: three statisticians, two physicists, a physical chemist, a nuclear physicist, and a large group of neuroimagers – but apparently we didn't have what it took. I met with a Princeton physicist, a Stanford neurophysiologist, a Cold Springs Harbor neurobiologist to discuss the paper. Again blanks, one and all.

https://www.lesswrong.com/posts/wpZJvgQ4HvJE2bysy/god-help-us-let-s-try-to-understand-friston-on-free-energy

The Free-Energy Theory



The Free-Energy Theory

Trevor Robbins
Cambridge

The Free-Energy Theory

$$log(p(s)) \boxed{-L} = D_{KL}[q(\phi)||p(\phi|s)]$$

$$L = \int log(\frac{p(\phi, s)}{q(\phi)})q(\phi)d\phi$$ Evidence Lower Bound (ELBO)

Free energy !!!

- Free energy就是ELBO的负数
- 最小化Free energy就是最大化ELBO

$$\approx \underset{q(\phi)}{argmax} \frac{1}{N} \sum_{l=1}^{N} logp(s|\phi)^{(l)} + logp(\phi)^{(l)} - logq(\phi)^{(l)}$$ → 最大化ELBO

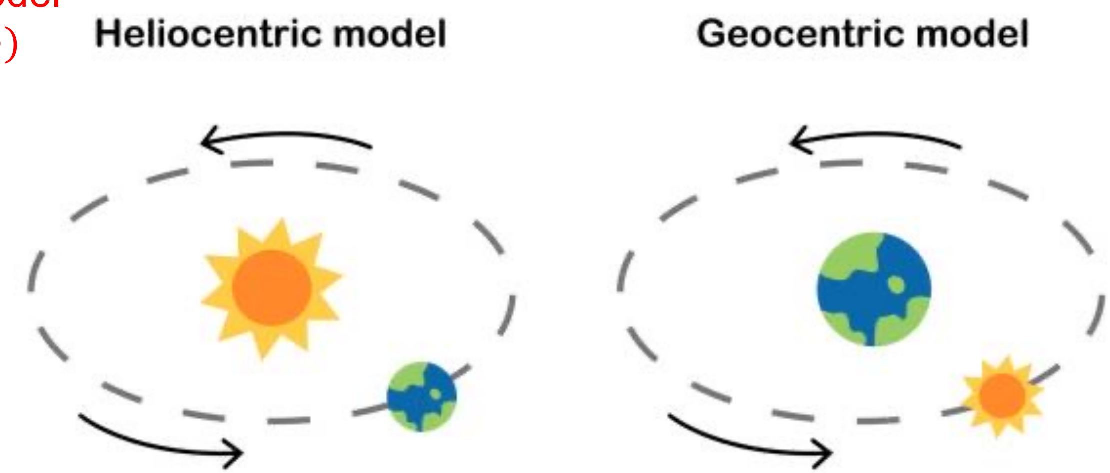
$$= \underset{q(\phi)}{argmin} \frac{1}{N} \sum_{l=1}^{N} logq(\phi)^{(l)} - logp(s|\phi)^{(l)} - logp(\phi)^{(l)}$$ → 最小化free energy

Variational inference的思想是不去直接求复杂的后验分布，而是另外假设一个简单的分布去近似后延分布，当两者距离足够小的时候，简单分布就是所求

Free-energy principle认为人脑也是这么做的！！！

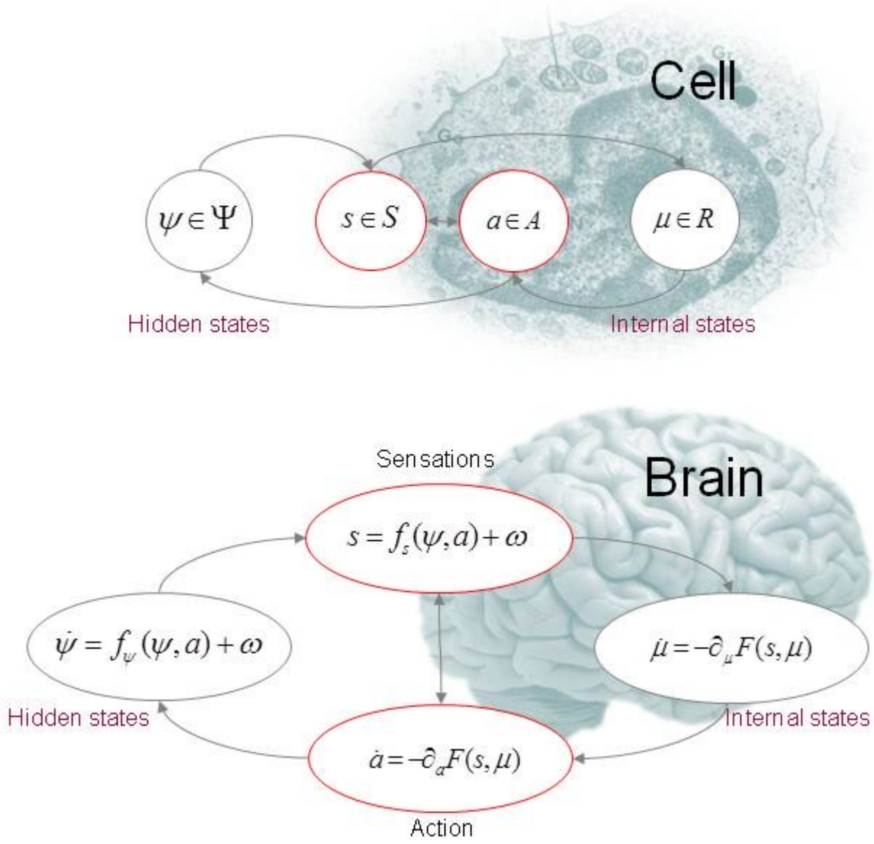
日心说 vs. 地心说

Generative model
$p(\phi|s) * p(\phi)$



Inference model
$q(\phi)$

The Free-Energy Theory



The Free-Energy Theory

- Problems???
- Unfalsifiable
  - 要么证明人不是在做贝叶斯推断?
  - 要么证明人做贝叶斯推断不是用的变分推断的形式?

ORIGINAL ARTICLE

Theory

What does the free energy principle tell us about the brain?

Samuel J. Gershman[1*]

说说Karl Friston的自由能原理

Ruyuan Zhang
心理学话题下的优秀答主

Feitong Yang、滤波插值 等 121 人赞同了该文章