Machine Consciousness

2020年1月

蒙杰

上海科技大学
ShanghaiTech University
地址：上海市浦东新区华夏中路393号 邮编：201210
Add: 393 Middle Huaxia Road, Pudong, Shanghai 201210, China

Conscious system vs Unconscious system

- *What determines to what extent a system has conscious experience*

The key difference between you and the camera：

information integration

From Neuroscience to Computer Science

A theory of consciousness: information integration theory(IIT)

Φ: *Measuring the capacity to integrate information for a system*

Key: Partition & Measure difference

- *System partition:*

$$S \rightarrow A, B$$

- Measure difference

$$Dist(p_A p_B, p)$$

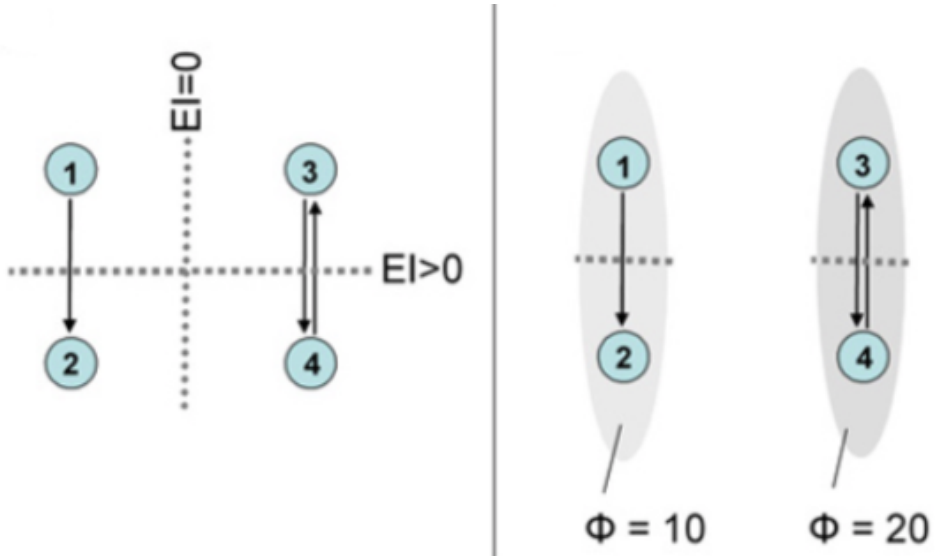
- Calculate phi

$$\phi = \min Dist(p_A p_B, p)$$

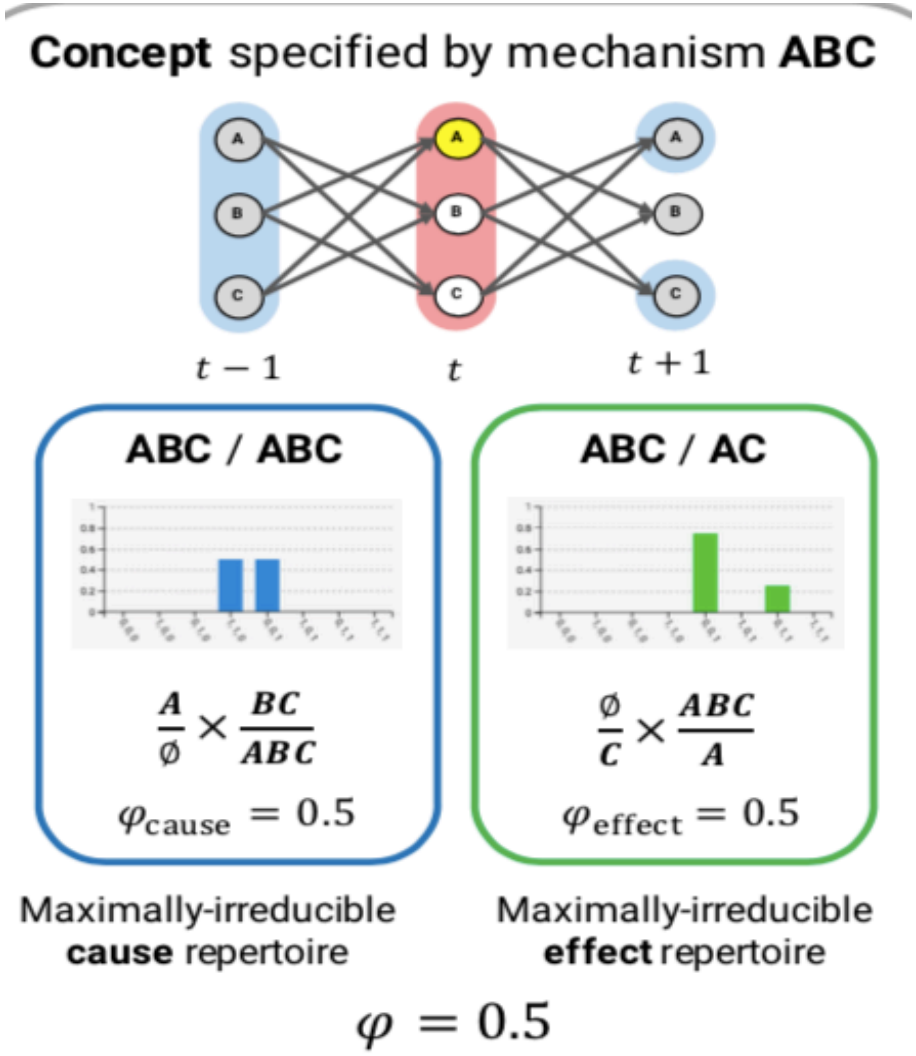
for minimum information partition (A,B)

Information integration theory

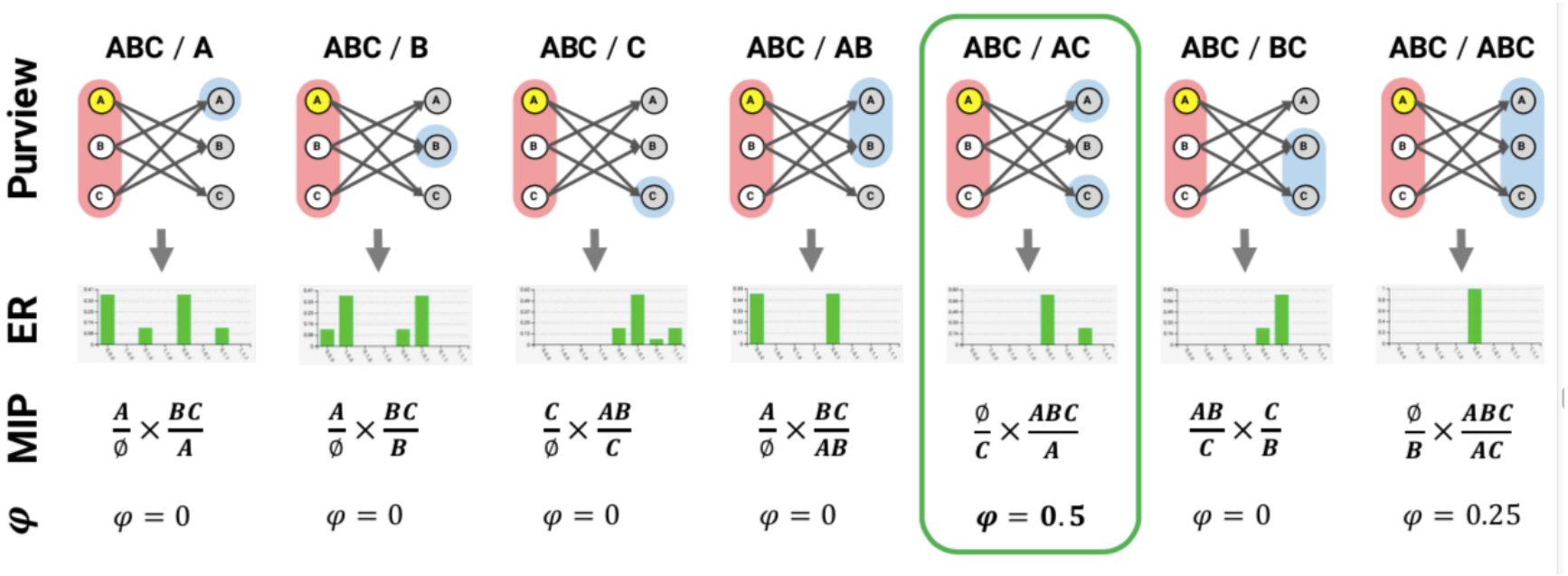
Key: Partition & Measure difference



Information integration theory 3.0



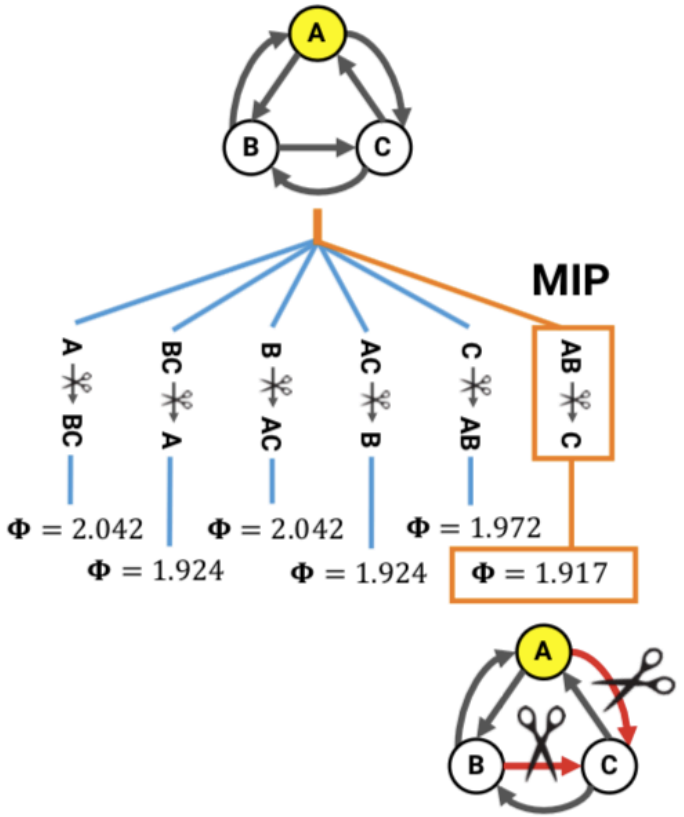
Information integration theory 3.0



Information integration theory 3.0

- The minimal Φ-value, Φ^{MIP}, is the Φ of the whole candidate system

- As with mechanisms, the cut that makes the *least difference* to the candidate system captures how intrinsically irreducible it is



Our Work

- Generalize to continuous system

- Calculate φ for Neural Network

From Discrete to Continuous

- Transition Probability Matrix → Probability Distribution
- Assumption:
  - Current state: Gaussian distribution
  - Transition probability: Gaussian distribution
- Key: Partition & Measure difference

$$p_t = Mp_{t-1}; p_t = \tilde{M}p_{t+1}$$

  - Cause distance

$$D_c = Dist\left(p_t, M^A p_{t-1}^A \otimes M^B p_{t-1}^B\right)$$

  - Effect distance

$$D_e = Dist\left(p_t, \tilde{M}^A p_{t+1}^A \otimes \tilde{M}^B p_{t+1}^B\right)$$

  - Calculate phi

$$\phi = \min_{(A,B)}\left(D_c + D_e\right)$$

From Discrete to Continuous

- Transition Probability Distribution:

$$M \sim Norm(\mu, \Sigma)$$

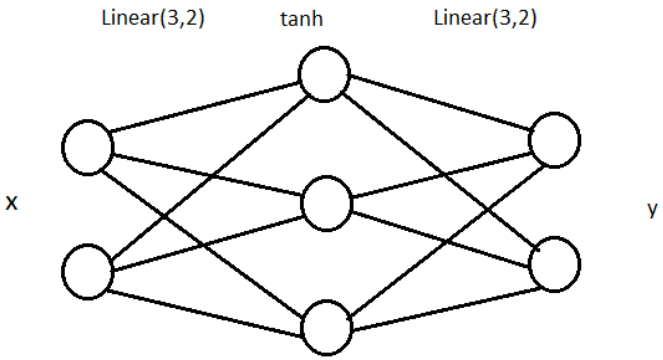
- Ma, Mb are normal distribution, their parameters are:

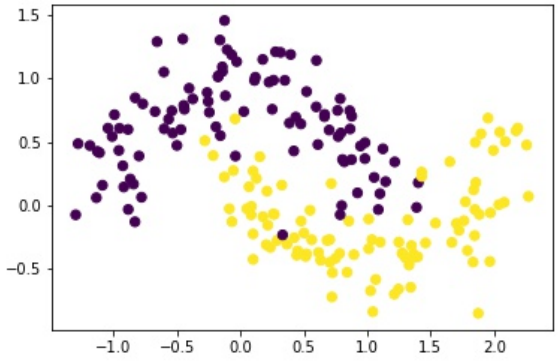
$$(\mu_A, \Sigma_A) = f(\mu, \Sigma)$$

$$(\mu_B, \Sigma_B) = g(\mu, \Sigma)$$

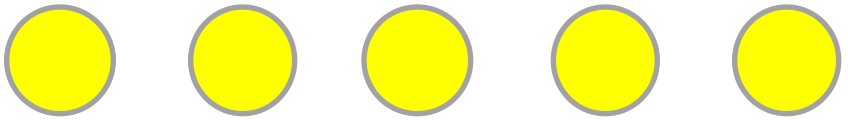
Experiment on simple neural network

MLP for classification: A feed-forward system

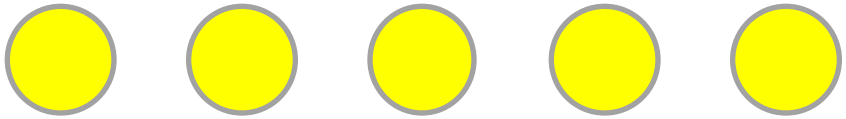




Sampling for Sigma; mu



Epoch t-1



Epoch t+1

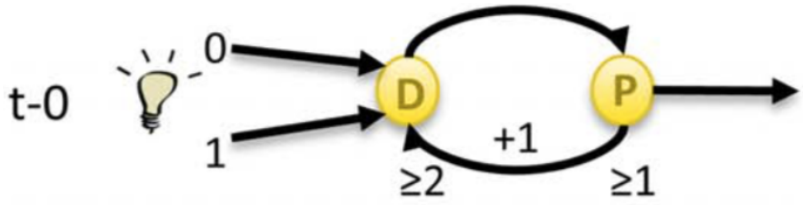
Result

$$\Phi = 36.8 > 0$$

Influence factors:

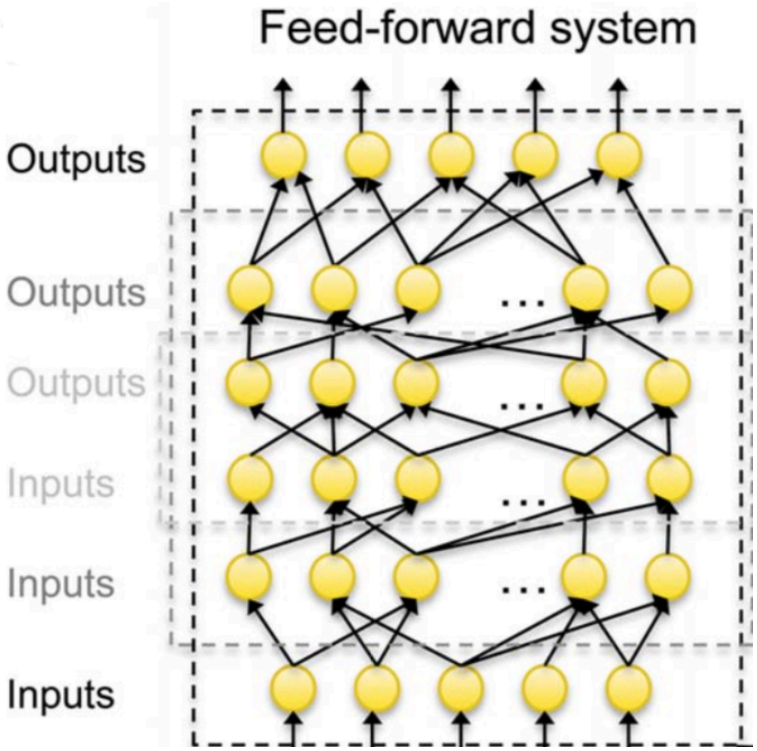
1. Linear vs Non-linear
2. Definition of distance metric
3. Input distribution

Intuition



Φ>0

A system with feedback circuit is more conscious?



Φ=0


谢谢！
上海雾计算实验室SHIFT