

② Evolving loss

method: multimodal learning.

①

Evolving Loss

$$L = \sum_m \sum_t \lambda_{m,t} L_{m,t} + \sum_d \lambda_d L_d$$

task

modality

distillation

weighted [0,1]

constraint

Distillation: $L_d(L_i, m_i) = \|L_i - m_i\|_2$

layer in main network

layer in another network.

Evolving an unsupervised loss function?

① constraint

Zipf's distribution matching.

feature $x_{P(A,B)} \in \mathbb{R}^D$ $= E_{\text{regh}}(\mathbb{I}) \rightarrow$ cluster into k .

$$P(x|c_i) = \frac{1}{\sqrt{2\sigma^2\pi}} \exp\left(-\frac{(x-c_i)^2}{2\sigma^2}\right) \quad // \text{centroid } c_i \in \mathbb{R}^D$$

... $\{c_1 \dots c_k\}$

$$P(c_i|x) = \frac{P(c_i) P(x|c_i)}{\sum_j P(x|c_j) P(c_j)} = \frac{\exp(-(x-c_i))^2}{\sum_{j=1}^k \exp(-(x-c_j))^2}$$

prior of $q(c_i) = \frac{1/c_i^s \rightarrow \text{real constant} \quad !! ?}{H_{k,s}}$

$\hookrightarrow k$ th harmonic number !! ?

law of total prob.

$$KL(P||Q) = \sum_i P(c_i) \log \frac{P(c_i)}{q(c_i)} \quad 1/P(c_i) = \frac{1}{N} \sum_{x \in V} P(c_i|x)$$