Answering from Sure to Uncertain: Uncertainty-Aware Curriculum Learning for Video Question Answering

Haopeng Li[1], Mohammed Bennamoun[2], Jun Liu[3], Hossein Rahmani[3], Qiuhong Ke[4]
[1]School of Computing and Information Systems, The University of Melbourne
[2]Computer Science & Software Engineering, University of Western Australia
[3]School of Computing and Communications, Lancaster University
[4]Department of Data Science & AI, Monash University

BMVC 2025

Motivations

- Previous methods for VideoQA do not consider the inherent **uncertainty** in data, which could have a negative impact on video, especially on the **fine-grained** motion analysis.
- Existing **self-paced curriculum learning** (SPL) exploit the loss as difficulty measurement, which is problematic because 1) the loss cannot precisely reflect the **data difficulty**; and 2) the **loss varies greatly** over training process.

Contributions

- We propose probabilistic modeling for VideoQA by considering VideoQA as a **stochastic computation graph**, where the hidden representations are modeled **as random variables** to capture the data uncertainty.
- We develop a novel **self-paced curriculum learning** framework, where the difficulty is measured by the **uncertainty** defined based on the probabilistic modeling.
- We integrate VideoQA into our **uncertainty-aware curriculum learning** framework and conduct extensive experiments. The results show that our method achieves **state-of-the-art performance** on several benchmarks.

Probabilistic Modeling for VideoQA

Variational Inference

$$p_{\phi,\psi,\theta}(y|V,Q) = \int_{m,n} p_{\phi,\psi,\theta}(y,m,n|V,Q)\mathrm{d}m\mathrm{d}n = \int_{m,n} p_\theta(y|m,n)q_\phi(m|V)q_\psi(n|Q)\mathrm{d}m\mathrm{d}n.$$

Monte Carlo Integration →

$$p_{\phi,\psi,\theta}(y|V,Q) \approx \frac{1}{K}\sum_{k=1}^{K} p_\theta(y|m_k,n_k)$$

Optimization (ELBO)

$$\begin{aligned}&\log p_{\theta,\phi,\psi}(y|V,Q)\\ &\geq \mathbb{E}_{(m,n)\sim q_{\phi,\psi}(m,n|V,Q)}\left[\log \frac{p_{\theta,\phi,\psi}(y,m,n|V,Q)}{q_{\phi,\psi}(m,n|V,Q)}\right]\\ &= \mathbb{E}_{(m,n)\sim q_{\phi,\psi}(m,n|V,Q)}\left[\log p_\theta(y|m,n)\right] - \log p(V)p(Q)\\ &- D_{\mathrm{KL}}\left(q_\phi(m|V)||p(m)\right) - D_{\mathrm{KL}}\left(q_\psi(n|Q)||p(n)\right),\end{aligned}$$

Monte Carlo Integration →

$$\begin{aligned}\mathcal{L}(\theta,\phi,\psi) = &-\frac{1}{K}\sum_{k=1}^{K}\log p_\theta(y|m_k,n_k)\\ &+\frac{\alpha}{2}\sum_d(1+\log(\sigma_m)_d^2-(\mu_m)_d^2-(\sigma_m)_d^2)\\ &+\frac{\alpha}{2}\sum_d(1+\log(\sigma_n)_d^2-(\mu_n)_d^2-(\sigma_n)_d^2),\end{aligned}$$

Uncertainty-Based Self-paced Curriculum Learning (SPL)

Loss Function

$$\mathcal{L}(\xi;e) = \frac{1}{D}\sum_{i=1}^{D} w_i l\left(f_\xi(x_i),y_i\right)$$

Uncertainty-Based Weight

$$w_i = 1-\sigma\left(\frac{U_i}{\lambda(e)}\right)$$

Detached Weight

$$\frac{\partial w_i}{\partial \xi} := 0, \forall i = 1,\cdots,D$$

Feature Uncertainty

$$cls:\ U_F = \frac{1}{C}\left\|\frac{1}{K}\sum_{k=1}^{K}(g_k-\bar{g})^2\right\|_1$$

$$reg:\ U_F = \frac{1}{K}\sum_{k=1}^{K}(\mu_k-\bar{\mu})^2,$$

Predictive Uncertainty

$$cls:\ U_P = -\sum_{c=1}^{C} p_c \log p_c,$$

$$reg:\ U_P = \frac{1}{K}\sum_{k=1}^{K}\sigma_k^2,$$

Normalization

$$\bar{U} = \frac{U-\mathrm{E}[U]}{\sqrt{\mathrm{Var}[U]}}$$

Experiments

Results on TGIF-QA

Models	*Count*↓	*FrameQA*	*Action*	*Transition*
B2A	3.71	57.5	75.9	82.6
HAIR	3.88	60.2	77.8	82.3
HOSTER	4.13	58.2	75.6	82.1
MASN	3.75	59.5	<u>82.3</u>	86.2
HQGA	3.97	**61.3**	76.9	85.6
IGV	<u>3.67</u>	53.8	78.5	85.7
UCLQA$_F$	**3.22**	<u>60.5</u>	**84.0**	<u>87.7</u>
UCLQA$_P$	**3.22**	60.4	**84.0**	**87.8**

Results on NExT-QA

Models	Testing			
	Causal	Temp.	Descrip.	All
CoMem	45.9	50.0	54.4	48.5
HCRN	47.1	49.3	54.0	48.8
HME	46.8	48.9	57.4	49.2
HGA	48.1	49.1	57.8	50.0
IGV	48.6	<u>51.7</u>	59.6	51.3
HQGA	49.0	52.3	59.4	<u>51.8</u>
UCLQA$_F$	<u>49.4</u>	50.6	**61.9**	<u>51.8</u>
UCLQA$_P$	**50.3**	50.5	<u>61.8</u>	52.2