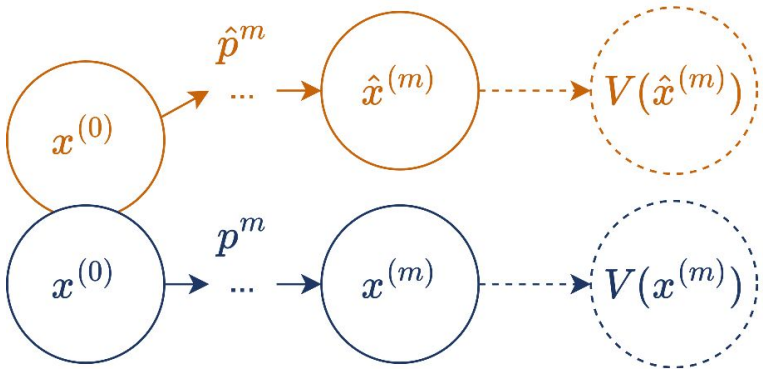
Calibrated Value-Aware Model Learning with Probabilistic Environment Models

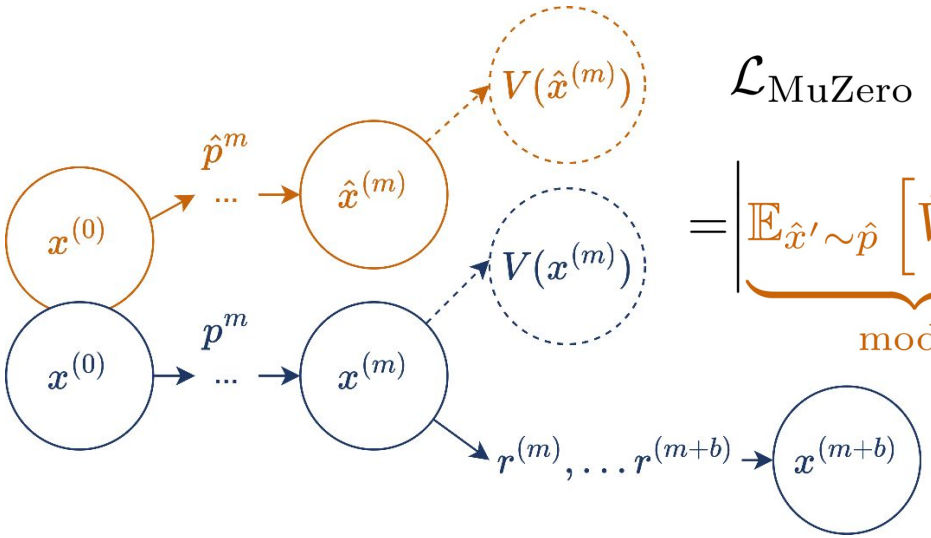
Claas Voelcker, Anastasiia Pedan, Arash Ahmadian,
Romina Abachi, Igor Gilitschenski, Amir-massoud Farahmand

Value Aware Model Learning Losses - VAML



$$\mathcal{L}_{\text{VAML}}\left(\hat{p}|x,\hat{V}\right) = \left|\underbrace{\mathbb{E}_{\hat{x}'\sim\hat{p}}\left[\hat{V}\left(\hat{x}'\right)\right]}_{\text{model}} - \underbrace{\mathbb{E}_{x'\sim\mathcal{P}}\left[\hat{V}\left(x'\right)\right]}_{\text{environment}}\right|^2$$

Value Aware Model Learning Losses - MuZero



$$\mathcal{L}_{\text{MuZero}}\left(\hat{p}, \hat{V} | x\right) = \Big| \underbrace{\mathbb{E}_{\hat{x}' \sim \hat{p}}\left[\hat{V}\left(\hat{x}'\right)\right]}_{\text{model}} - \underbrace{\mathbb{E}_{r,x',x'' \sim \mathcal{P}}\left[r + \gamma \hat{V}\left(x''\right)\right]_{\text{sg}}}_{\text{environment}} \Big|^2$$

Loss approximation

$$\mathcal{L}_{\text{VAML}}\left(\hat{p}|x,\hat{V}\right) = \left|\underbrace{\mathbb{E}_{\hat{x}'\sim\hat{p}}\left[\hat{V}\left(\hat{x}'\right)\right]}_{\text{model}} - \underbrace{\mathbb{E}_{x'\sim\mathcal{P}}\left[\hat{V}\left(x'\right)\right]}_{\text{environment}}\right|^2$$

$$\hat{\mathcal{L}}_{\text{VAML}}\left(\hat{p}|x,\hat{V}\right) = \left|\left(\frac{1}{k}\sum_{i=1}^{k}\hat{V}\left(\hat{x}'_i\right)\right) - \hat{V}\left(x'\right)\right|^2$$

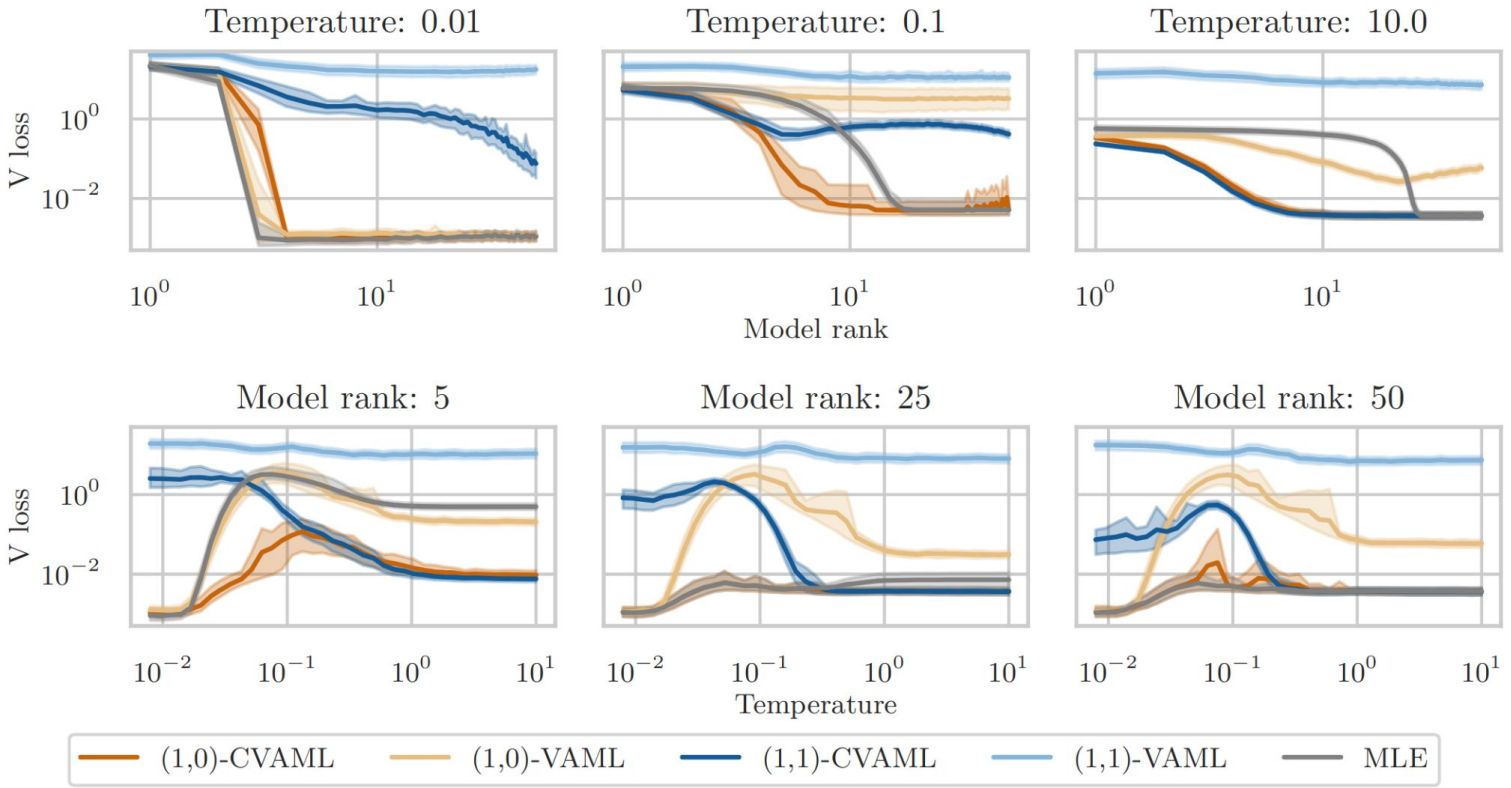
$$\mathcal{L}_{\text{MuZero}}\left(\hat{p},\hat{V}|x\right) = \left|\underbrace{\mathbb{E}_{\hat{x}'\sim\hat{p}}\left[\hat{V}\left(\hat{x}'\right)\right]}_{\text{model}} - \underbrace{\mathbb{E}_{r,x',x''\sim\mathcal{P}}\left[r+\gamma\hat{V}\left(x''\right)\right]_{\text{sg}}}_{\text{environment}}\right|^2$$

$$\hat{\mathcal{L}}_{\text{MuZero}}\left(\hat{p},\hat{V}|x\right) = \left|\left(\frac{1}{k}\sum_{i=1}^{k}\hat{V}\left(\hat{x}'_i\right)\right) - \left[r+\gamma\hat{V}\left(x''\right)\right]_{\text{sg}}\right|^2$$

Variance Error Term

$$\mathbb{E}_{\hat{p}}\left[\hat{\mathcal{L}}_{\mathrm{VAML}}\left(\hat{p}|x,\hat{V}\right)\right]$$
$$=\underbrace{\mathcal{L}_{\mathrm{VAML}}\left(\hat{p}|x,\hat{V}\right)}_{\text{desired}}+$$

Empirical results - Gridworld



Claas Voelcker

Anastasiia Pedan

Arash Ahmadian

Romina Abachi

Igor Gilitschenski

Amir-massoud Farahmand