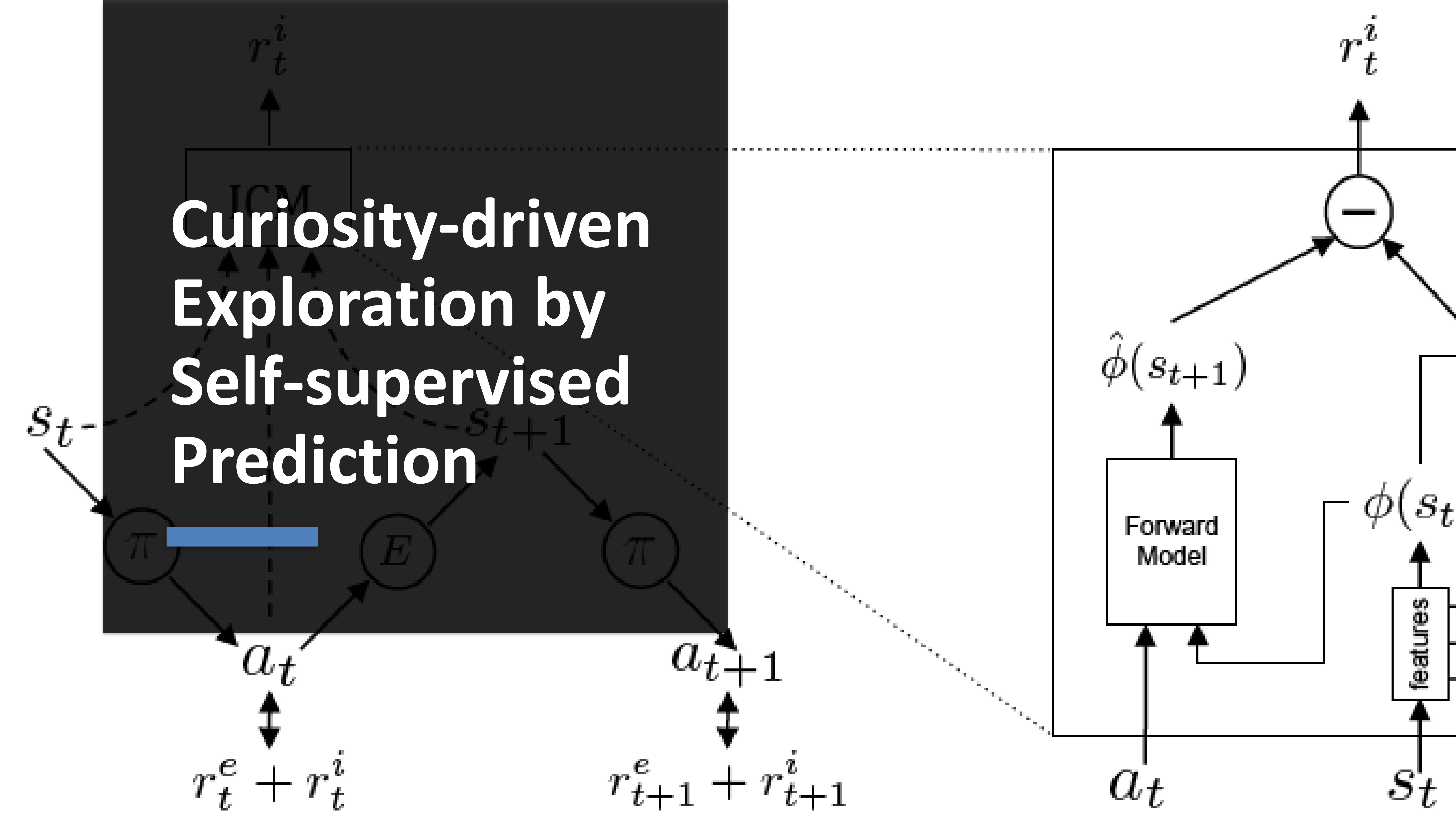
Curiosity-driven Exploration by Self-supervised Prediction



Abstract

In many real-world scenarios, rewards are extremely sparse.
Curiosity can serve as an intrinsic reward.

Curiosity: error in an agent's ability to predict the consequence of its own actions learned by self-supervised inverse dynamic model.

"3-year-old has no trouble entertaining herself using intrinsic motivation or curiosity."

Intrinsic Reward

How hard it is to predict the consequences of actions.

Encourage to explore 'novel' states.
- It requires a statistical model of the distribution of the environment states.

Encourage to perform actions that reduce the error/uncertainty in agent's ability to predict.
-it requires building a model of dynamics.

It is an mechanism for an agent to learn skills that might be helpful in future scenarios.

Intrinsic Reward

Self-supervised prediction for exploration

Given the raw state S_t, encoding it using a DNN into a feature vector $\emptyset(S_t;\theta_E)$
To learn the parameters of this feature encoder, using two sub-modules.

1) g: inverse dynamics model

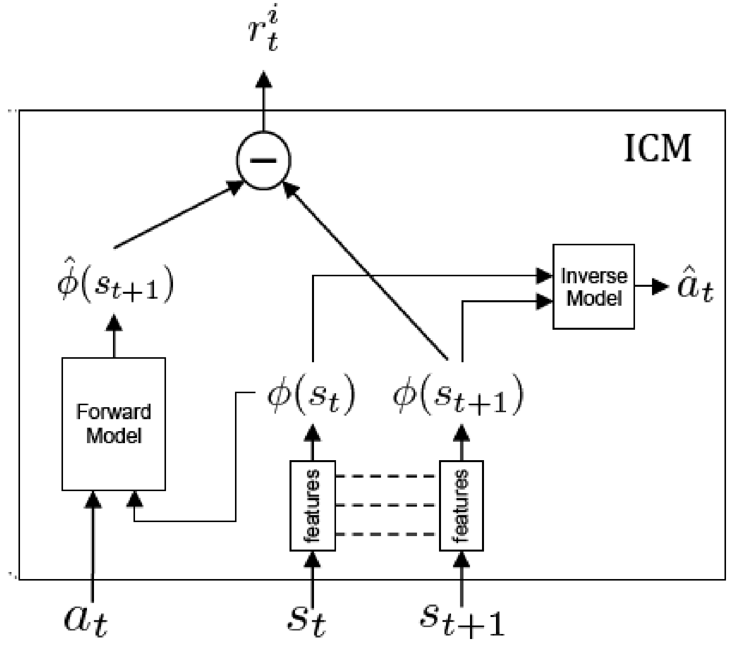
$$\hat{a}_t = g\Big(\phi(s_t), \phi(s_{t+1}); \theta_I\Big)$$

$$\min_{\theta_I,\theta_E} L_I(\hat{a}_t, a_t)$$

2) f: forward dynamics model

$$\hat{\phi}(s_{t+1}) = f\Big(\phi(s_t), a_t; \theta_F\Big)$$

$$\min_{\theta_F,\theta_E} L_F\Big(\hat{\phi}(s_{t+1}), \phi(s_{t+1})\Big)$$

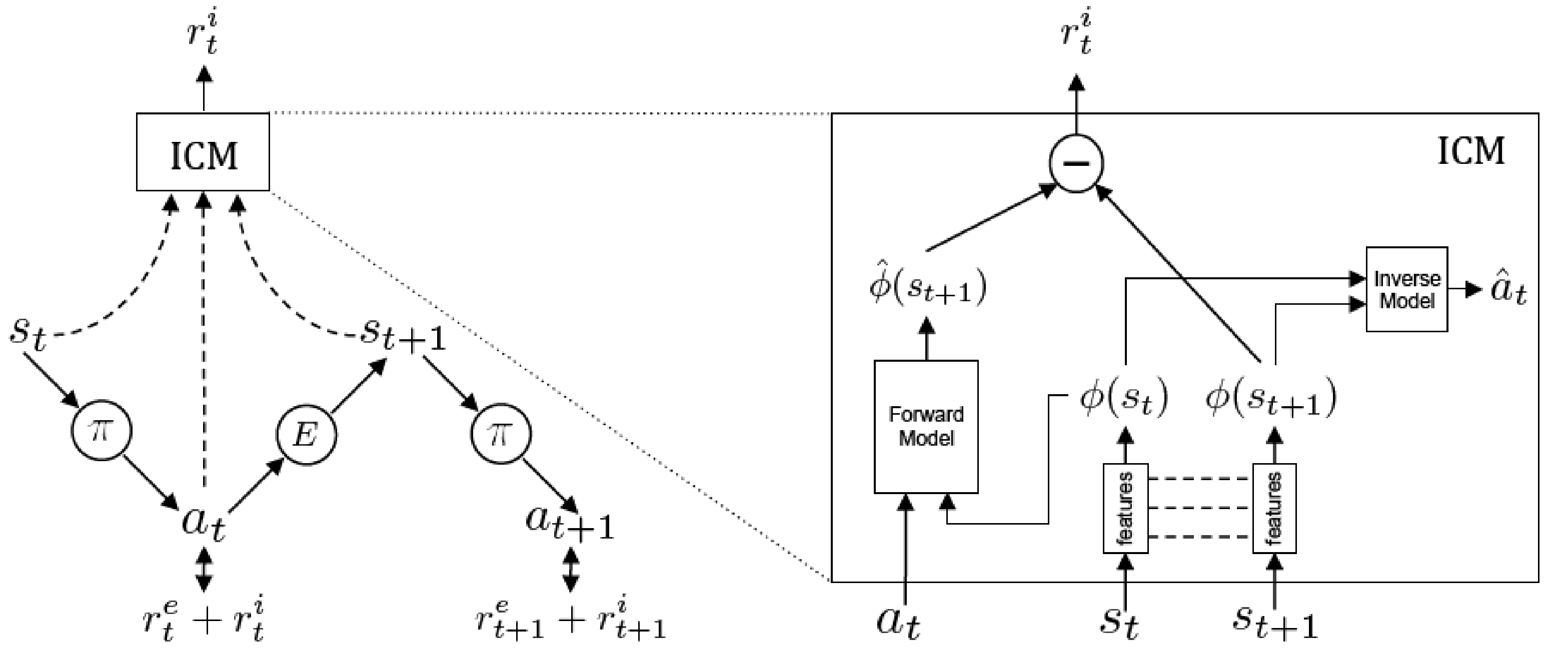


Finally, the intrinsic reward signal r_t^i is computed as,

$$r_t^i = \frac{\eta}{2}\|\hat{\phi}(s_{t+1}) - \phi(s_{t+1})\|_2^2$$

Intrinsic Reward

Self-supervised prediction for exploration



The overall optimization problem can be written as,

$$\min_{\theta_P,\theta_I,\theta_F,\theta_E} \left[-\lambda \mathbb{E}_{\pi(s_t;\theta_P)}[\Sigma_t r_t] + (1-\beta)L_I + \beta L_F \right]$$