Proposal of Final Project
Zhengxian Lin
ONID: linzhe
Student ID: 933317086

The project is about reducing space of the enormous states by variational auto-encoder for a game called 2048(https://play2048.co/), and then train a model-based agent on the representation states, comparing to several model-free agent like Q-learning agent, SARSA agent.

Step 1:

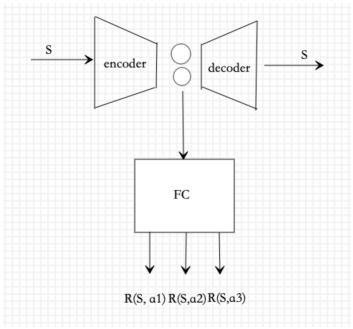
Creating a demo 2048 game (without UI)

Step 2:

Creating model. Build a VAE structure, training it by dataset from states of the 2048 game. Use the encoder to encode input state of the game to get the distribution of state.

Step 3:

However, the distribution of the states is not good enough to train the agent, because the distribution is about state itself. It is not guided by rewards. So, I will plug in a DQN model to connect with the VAE, taking the distribution as input. The output is the q-value of each action, comparing with the q-value from the game and do the backpropagation to DQN and the encoder of VAE. Thus, there will be two backpropagation way, first one update the decoder and encoder, another one update the DQN model and encoder of VAE.



Step 4:

Tune hyper-parameter, record the performance.

Step 5:

Create one or several model free agents like UCT or regular DQN to learn the best policy of the game and compare the model-based agent and model free agent.

The state space of 2048: 16^{16}

The representation state space : 2^8(maybe)

The shape of probabilistic model: $2^8 * 2^8$

Performance measure: the score of game

For the check point:

Finish step 1 and step 2.

However, for the step 2, training the VAE to get a good result is hard to tell until test the

performance after step 4. So, after checkpoint, redo the step 2, 3 is necessary.

Plan B:
Because this mode is designed by myself (I don’ t know if there is someone who have done the things like this before.), so it is possible to be not working.
Also, I have read a paper called DARLA: Improving Zero-Shot Transfer in Reinforcement Learning, I am interesting on that, and it also provide codes. However, the reason why I don’ t want to do something on the paper is that the model of this paper is not guided by rewards. What it exactly does is extracting the high-level features from raw observation of video game, but it may can extract something from images that are relative, but may not for my case, since the raw observation in my case is not image, and the similarity visual state can have opposite strategy. However, if I have to try something one this paper, I will change the neutral network structure of this model. The first step is removing the convolution layers, because I don’ t need that anymore. And see that it may also be working on my case.

The model of DARLA:

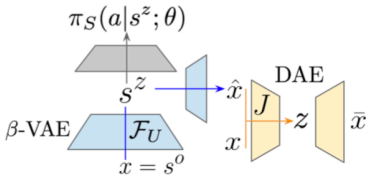


Figure 1. Schematic representation of DARLA. Yellow represents the denoising autoencoder part of the model, blue represents the β-VAE part of the model, and grey represents the policy learning part of the model.