

# World Models-Paper Review

Ayush Aaryan

18-07-2019

## 1 Problem

- 1.The paper aims at training the agent to get better at a 2D racing game whose environment is available on OpenAI Gym with the help of Deep Reinforcement Learning.The agent can control the car by deciding the steering angle  $[-1, 1] \rightarrow [\text{Left}, \text{Right}]$ , acceleration and brake.
- 2.The paper also aims to train an agent to play the ViZDoom (DOOM with tuning for RL methods benchmarking) environment.

## 2 What makes the model different

The model learns using the states it dreams about ,which technically means using the latent state.The model only focusses on the important parts of the environment(like roads,fireballs coming rather than the sky and background color) like we humans do in our dreams or when we visualise things mentally.This hugely cuts down on the amount of data necessary which is very useful when we use iterative training procedure.

## 3 Model

The model basically consists of 4 components-:

- 1.Random Policy Search
- 2.Variational Autoencoder
- 3.RNN,specifically an LSTM network(Memory Model)
- 4.Controller Model

We will discuss all of them briefly-:

### 3.1 Random Policy

The model collects a bunch of samples(video frames) of environment through Random Policy search and gives the observation(frame) as output.

## 3.2 Variational Autoencoder

We will begin with what these autoencoders do.

Traditional autoencoders

They consist of two parts:-

1.Encoder-They collect data from an input and compress this data into lower dimensional data(like a bottleneck).

2.Decoder-They generate a new output using this lower dimensional data.

What these Variational Encoders do is add some randomness in the data encoded,so that we can get slightly different versions of the input.

Now,talking about the model,the observation frame is passed on to the Variational Autoencoder and the model starts focussing on the dynamic part(turns etc),formally called the latent entries of the frame rather than the static part(like ceiling,forest around etc) and finally we get an output different from the given input.These outputs are the different situations the model may think of.

## 3.3 RNN(Memory Model)

The RNN helps in decision making of the model,so that each observation is not a surprise to the model.It helps to predict what the next state of the environment be based on your current state and action.It helps to track how your states evolve over your actions.This model finally gives a new latent state as the output taking input the previous latent state.The model consists of a 256 hidden units LSTM network.

### 3.3.1 Temperature parameter

Temperature parameter is a measure of randomness we want,the more the parameter,more is the randomness.So,the model can imagine a whole lot of futures to remove wrong decisions based on exploitations.

## 3.4 Controller Model

The Controller is simply a densely connected neural network which helps in learning the weights and bias using  $Z_t$ (latent state) and  $h_t$ (hidden state) as input.The 3 output neurons correspond to the three actions(acceleration,brakes,steering direction).

## 4 Hype around the "Dream" word

The model actually performed a surprisingly good task on VizDoom, where the RNN model learnt to generate monsters that shoot fireballs and the controller discovered policies to actually dodge these fireballs, the VAE finally decoded this setup and helped the agent to transfer this style of playing to the real environment (real game).

## 5 Experimental Results

The experiment found out that the scores greatly varied with the temperature parameter. An optimum temperature parameter ensured greater scores as well less variance when the model was applied to the real environment.