World Models

Manan Tomar : ED14B023

Abstract

This paper ((Ha & Schmidhuber, 2018)) proposes a model based technique for learning a latent space representation, and then using it to predict in the latent space, thus allowing for control in Reinforcement Learning tasks. The proposed method is tested on the Car Racing task and the on the Vizdoom shooting task.

1. Approach

The work presents learning through three modules refered as **V**, **M**, and **C** for VAE, Model and Control. The Variational Autoencoder (VAE) (Kingma & Welling, 2013) module learns a latent representation z_t of the input image at time t. The Model module uses this representation to learn to predict in the latent space. This is done with a recurrent connection using a history of z. The output here is z_{t+1} given z_t, the hidden representation h_t, and the action a_t. The control module then uses z_t and h_t to produce the action a_t using only a linear layer. The **M** module therefore serves two purposes, that of providing a hidden representation h_t to the control module and of prediction which allows learning using synthetic samples. Note that the prediction is not deterministic but a probability density function based on Gaussian Mixture Models. Also, since the controller design and tasks considered are fairly simple, the authors choose the Covariance-Matrix Adaptation Evolution Strategy (CMA-ES) (Hansen, 2006) as the control algorithm.

The training is carried in three steps, first learning the VAE module using samples from the environment, followed by learning to predict based on the learnt latent representation z and then finally using both z and prediction module to learn to control either on the real environment or on a virtual one, using artificially generated samples. In the latter case, the learnt policy is then deployed to the real world simulator.

Since the environment is not guaranteed to be explored fully using a random policy, making the learnt model inaccurate for such cases, the authors propose learning in an iterative manner by toggling between learning in a virtual environment and generating more samples from the real world to train the prediction module further if the task remains incomplete.

2. Experiments

2.1. Car Racing

In this case, the control module is trained for both with and without the prediction module **M**. The authors observe that adding a history component to **C** allows learning a smoother behavior.

2.2. Vizdoom

The authors consider the shooting task and learn to control in a ”hallucinated” environment, i.e. a virtual environment created using artificial samples of future states through the learned prediction module. The authors show how learning only in such a virtual environment can allow transferring the learnt policy successfully to the real world.

3. Critique and Future Work

- Although Vizdoom is a 3D domain, both the tasks considered require only two dimensional actions which limits the generality of the method to harder 3D tasks such as maze navigation.
- Although the authors suggest an iterative approach to tackling divergence in model training in cases where tasks are fairly difficult, no clear method is described and it is unsure what policy needs to be used to train the prediction module further. Therefore, this part is loosely concluded.
- The authors also do not mention a method for fine-tuning after deployment to the real world. This is essential to overcome the inaccuracies present in the learnt environment model.

References

Ha, David and Schmidhuber, Jürgen. World models. *arXiv preprint arXiv:1803.10122*, 2018.

Hansen, Nikolaus. The cma evolution strategy: a comparing review. In *Towards a new evolutionary computation*, pp. 75–102. Springer, 2006.

Kingma, Diederik P and Welling, Max. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.