World Models

D. Ha and J. Schmidhuber, ArXiv, 2018.

R244 Large-scale data processing and optimisation
A (short) presentation by Anna Talas on 24/11/2022

Motivation

- Reinforcement learning (RL) problems are difficult
- Vast amount of information to process
  - Bottlenecked by credit assignment problem
- Training is slow

The inspiration

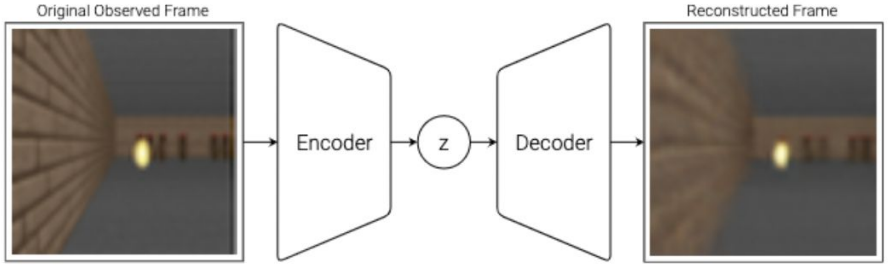
- Can we build an agent that thinks like a human?
- **“The image of the world around us, which we carry in our head, is just a model. Nobody in his head imagines all the world, government or country. He has only selected concepts, and relationships between them, and uses those to represent the real system.” (Forrester, 1971)**
- Having an abstract representation of the world is sufficient
- Can an agent learn inside a dream?

The idea

A world model that can be trained **quickly** in an **unsupervised** manner to learn a **compressed representation** of the environment.

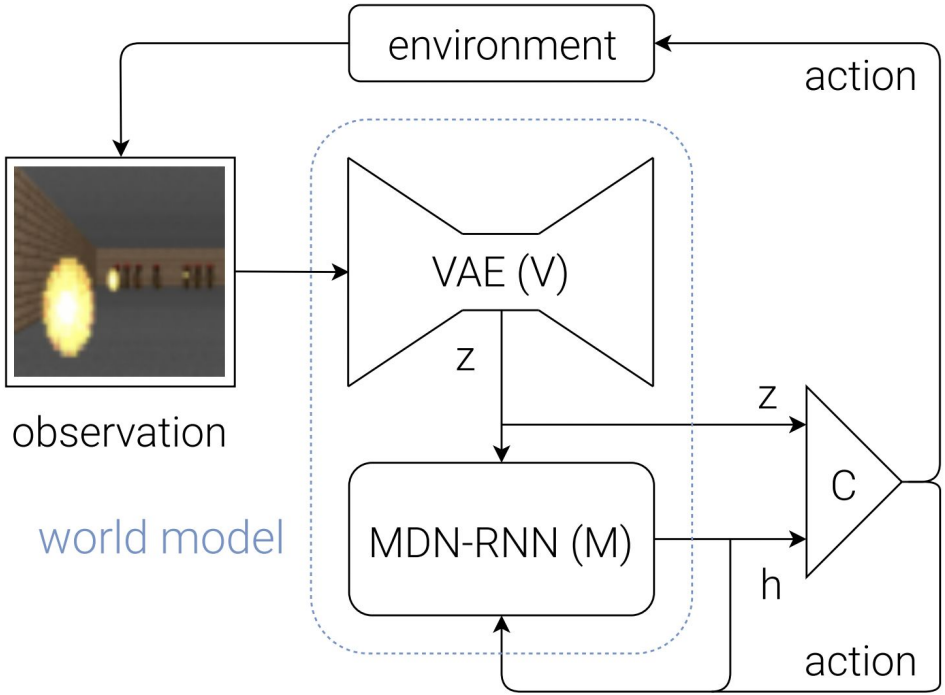
The framework

- *World model* consisting of V and M
- Vision model (V)
  - Variational Autoencoder
  - Encodes input to latent vector
- Memory model (M)
  - Predict future states
  - Probability distribution
- Controller model (C)
  - Make decisions
  - As simple and small as possible
  - Trained separately
  - Only part with access to rewards information



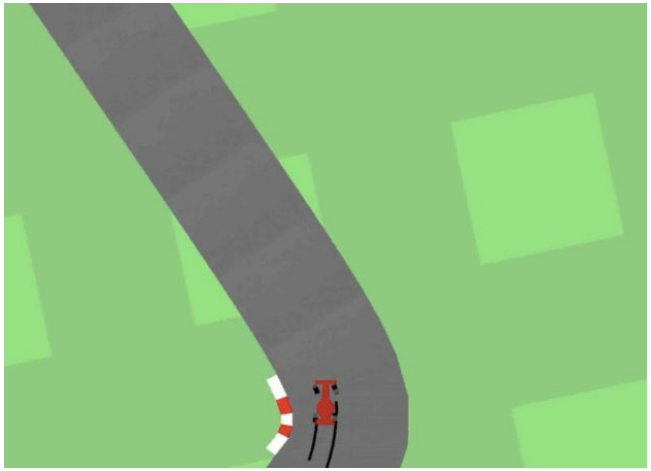
Model	Parameter Count
VAE	4,348,547
MDN-RNN	422,368
Controller	867

How it all fits together



Evaluation (1/2)

- CarRacing-v0 (OpenAI Gym)
- Procedure:
  - Collect 10,000 random rollouts
  - Train V model to encode frames
  - M model to predict next state
  - Define controller
  - Solve for max expected cumulative reward
- First to solve the problem
  - Average 900 steps
- Both V and M model needed



METHOD	AVG. SCORE
DQN (PRIEUR, 2017)	343 ± 18
A3C (CONTINUOUS) (JANG ET AL., 2017)	591 ± 45
A3C (DISCRETE) (KHAN & ELIBOL, 2016)	652 ± 10
CEOBILLIONAIRE (GYM LEADERBOARD)	838 ± 11
V MODEL	632 ± 251
V MODEL WITH HIDDEN LAYER	788 ± 141
FULL WORLD MODEL	$\mathbf{906 \pm 21}$

Table 1. `CarRacing-v0` scores achieved using various methods.

Evaluation (2/2)

- VizDoom
- 750 time steps to solve
- Learning inside a hallucinated dream
  - Using M model to predict the next state
  - No connection to actual environment
  - 900 steps on average
- Policy transferred to real world
  - 1100 steps on average
- “Dream” world can be made more difficult
  - uncertainty

Critique

- **Pros:**
  - An innovative way to think
  - State-of-the-art results
  - Could minimise computational effort needed
- **Cons:**
  - The *world model* can be cheated
  - Tested on a limited set of problems with relatively little noise
  - Is unsupervised learning a viable option for more complex problems?

References

- Ha, David R and Jürgen Schmidhuber. “World Models.” ArXiv abs/1803.10122 (2018): n. Pag.
- https://worldmodels.github.io/ (interactive version of the paper)
- Forrester, Jay W.. “Counterintuitive behavior of social systems.” Theory and Decision 2 (1971): 109-140.

Thank you!