



CONTRASTIVE LEARNING OF STRUCTURED WORLD MODELS

Вепренцев Иван, 172

Постановка задачи и предлагаемое решение

Задача: построение представлений для состояний в RL

Предлагаемый подход: факторизация каждого состояния (разложение на отдельные объекты, которые находятся попиксельно и затем представляются в виде графа с попарными соотношениями)

Существующие подходы:

- Генеративные модели. Основной недостаток – внимание крупным, но неважным объектам (например фону), но игнорирование мелких, но важных объектов
- Модели с учителем

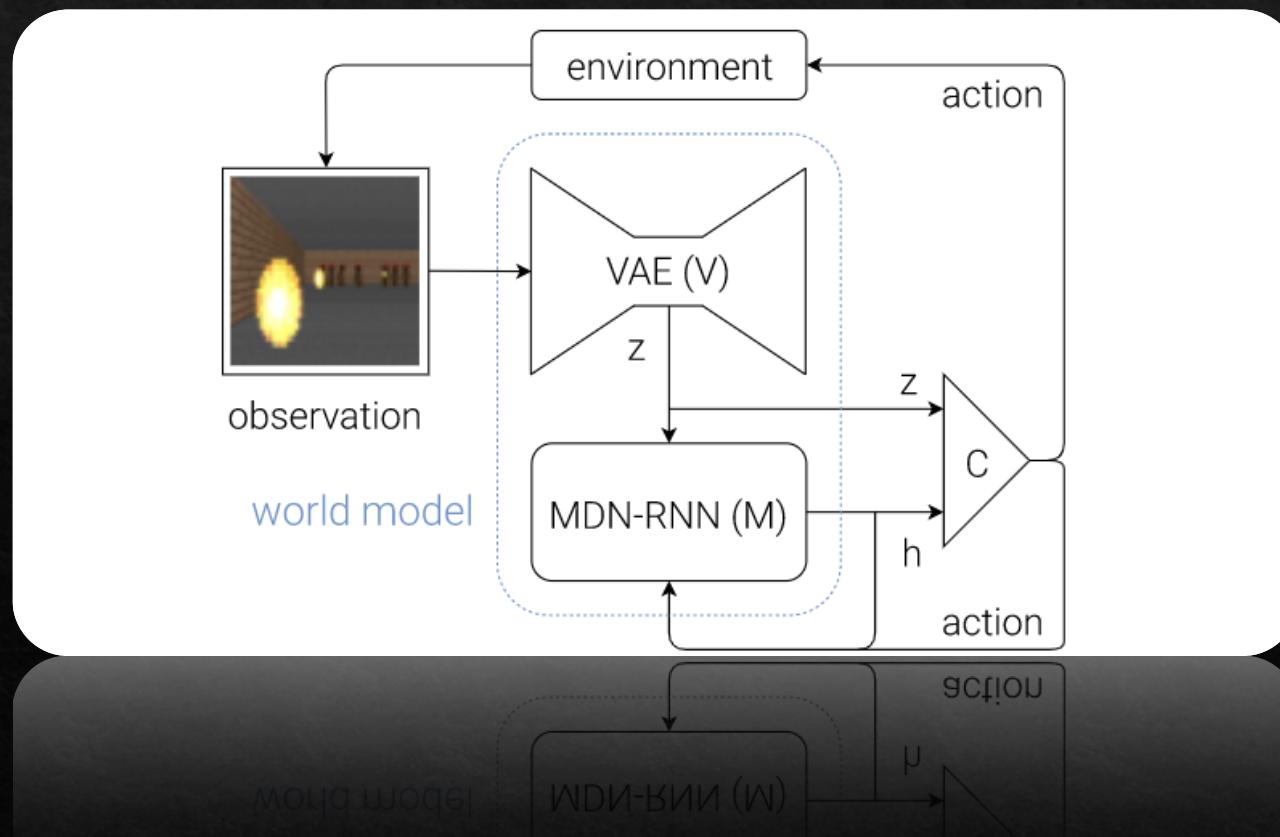
C-SWM опирается на предыдущие работы и опирается на триплексы (состояние-действие-состояние), противопоставляя реальные изменения системы искаженным.

STRUCTURED WORLD MODELS

Основная идея:

“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.”

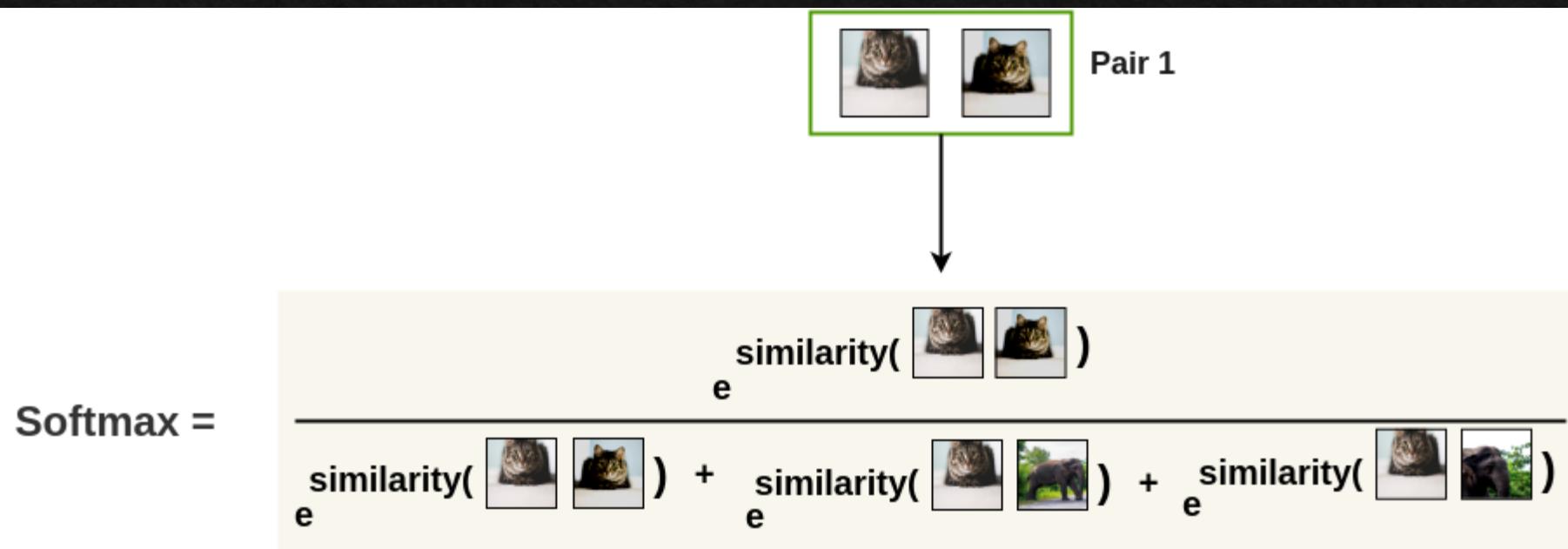
J.W.Forrester



Contrastive Learning

Основная идея:

- Используем косинусную меру, или любую адекватную функцию энергии в качестве меры различия между объектами
- Для того, чтобы убедиться, что изображения похожи, сравниваем текущую разницу в «энергиях» с разницей между непохожими изображениями
- Для того, чтобы можно было минимизировать ошибку составим выражение, например по типу представленного ниже, или же, например, Hinge-Loss



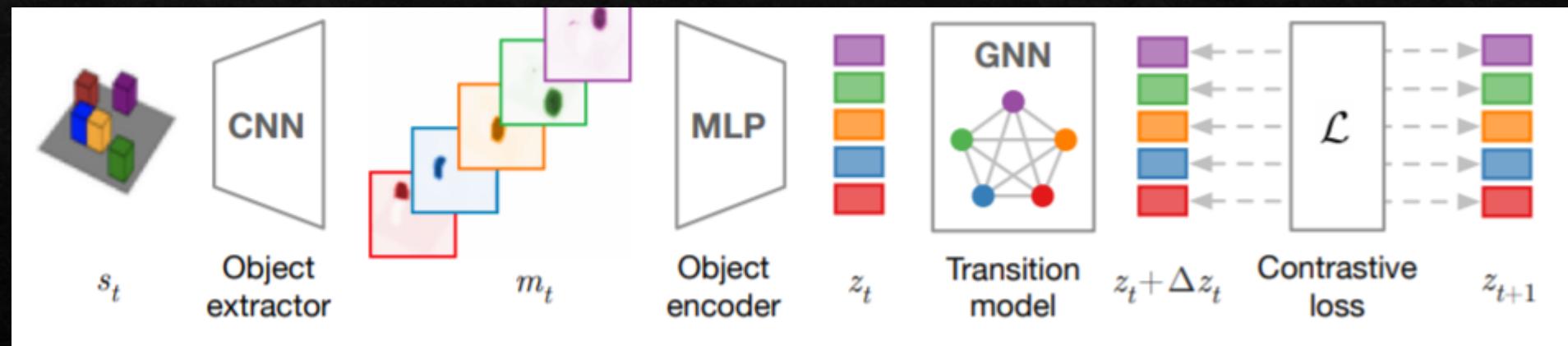
Реализация концепта целиком

1) Будем оперировать буфером памяти $B = (s_t, a_t, s_{t+1})$, где есть информация о состояниях, действии и новом состоянии после действия. Хотим получить обобщённое z_t для каждого s_t для реализации *STRUCTURED WORLD MODELS*

- ❖ $E : S \rightarrow Z$
- ❖ $T : Z \times A \rightarrow Z$

2) Используем *CONTRASTIVE LEARNING* для функции энергии $H = d(z_t + T(z_t, a_t), z_{t+1})$ и получим ошибку $L = d(z_t + T(z_t, a_t), z_{t+1}) + \max(0, \gamma - d(z_t, z_{t+1}))$,
 \mathbf{z}_t – дискриминативное изображение

3) Получим следующую архитектуру:



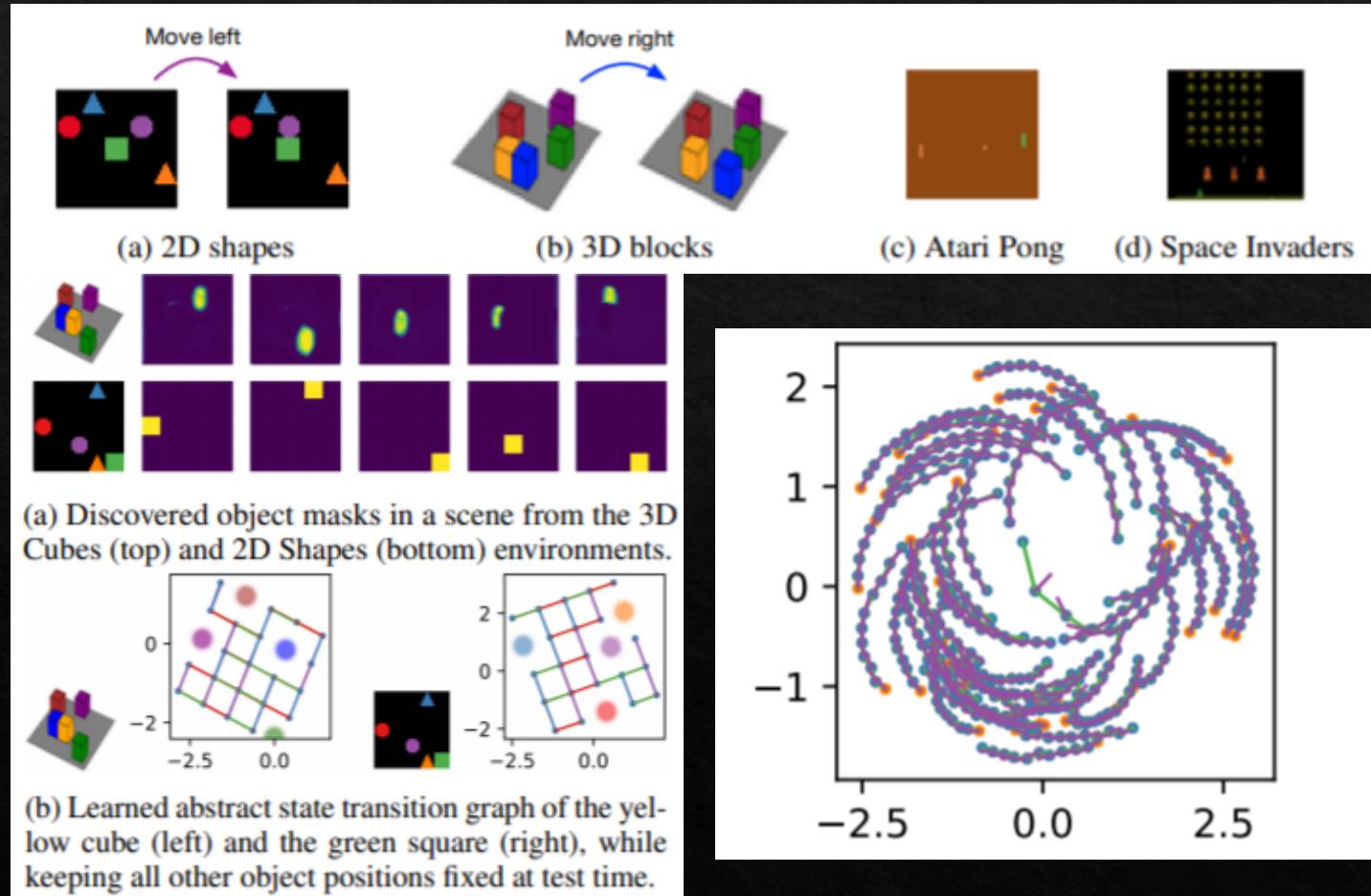
Немного деталей

- ❖ Авторы «доставали» объекты, используя две модели (CNN+MLP). Работали именно с объектами по отдельности (факторизовали состояние), чтобы получить лучшую обобщающую способность. Объекты сначала извлекались, а затем для каждого состояния им присваивались веса.
- ❖ В качестве Transition Model использовался полносвязный GNN, который получал текущее состояние на вход и предсказывал изменение состояния Δz , после чего следующее состояние получалось сложением текущего вектора и Δz . Тогда функция потерь выходит такая:

$$H = \frac{1}{K} \sum_{k=1}^K d(z_t^k + T^k(z_t, a_t), z_{t+1}^k), \quad \tilde{H} = \frac{1}{K} \sum_{k=1}^K d(\tilde{z}_t^k, z_{t+1}^k)$$
$$\mathcal{L} = H + \max(0, \gamma - \tilde{H})$$

Эксперименты авторов

❖ 2D и 3D эксперименты. Несколько тысяч эпизодов примерно по 100 шагов в зависимости от среды. В качестве метрики оценивания – ранкинговая система



Model	1 Step		5 Steps		10 Steps	
	H@1	MRR	H@1	MRR	H@1	MRR
2D SHAPES						
C-SWM	100 \pm 0.0	100 \pm 0.0	100 \pm 0.0	100 \pm 0.0	99.9 \pm 0.0	100 \pm 0.0
– latent GNN	99.9 \pm 0.0	100 \pm 0.0	97.4 \pm 0.1	98.4 \pm 0.0	89.7 \pm 0.3	93.1 \pm 0.2
– factored states	54.5 \pm 18.1	65.0 \pm 15.9	34.4 \pm 16.0	47.4 \pm 16.0	24.1 \pm 11.2	37.0 \pm 12.1
– contrastive loss	49.9 \pm 0.9	55.2 \pm 0.9	6.5 \pm 0.5	9.3 \pm 0.7	1.4 \pm 0.1	2.6 \pm 0.2
World Model (AE)	98.7 \pm 0.5	99.2 \pm 0.3	36.1 \pm 8.1	44.1 \pm 8.1	6.5 \pm 2.6	10.5 \pm 3.6
World Model (VAE)	94.2 \pm 1.0	96.4 \pm 0.6	14.1 \pm 1.1	21.4 \pm 1.4	1.4 \pm 0.2	3.5 \pm 0.4
3D BLOCKS						
C-SWM	99.9 \pm 0.0	100 \pm 0.0	99.9 \pm 0.0	100 \pm 0.0	99.9 \pm 0.0	99.9 \pm 0.0
– latent GNN	99.9 \pm 0.0	99.9 \pm 0.0	96.3 \pm 0.4	97.7 \pm 0.3	86.0 \pm 1.8	90.2 \pm 1.5
– factored states	74.2 \pm 9.3	82.5 \pm 8.3	48.7 \pm 12.9	62.6 \pm 13.0	65.8 \pm 14.0	49.6 \pm 11.0
– contrastive loss	48.9 \pm 16.8	52.5 \pm 17.8	12.2 \pm 5.8	16.3 \pm 7.1	3.1 \pm 1.9	5.3 \pm 2.8
World Model (AE)	93.5 \pm 0.8	95.6 \pm 0.6	26.7 \pm 0.7	35.6 \pm 0.8	4.0 \pm 0.2	7.6 \pm 0.3
World Model (VAE)	90.9 \pm 0.7	94.2 \pm 0.6	31.3 \pm 2.3	41.8 \pm 2.3	7.2 \pm 0.9	12.9 \pm 1.3
ATARI PONG						
C-SWM ($K = 5$)	20.5 \pm 3.5	41.8 \pm 2.9	9.5 \pm 2.2	22.2 \pm 3.3	5.3 \pm 1.6	15.8 \pm 2.8
C-SWM ($K = 3$)	34.8 \pm 5.3	54.3 \pm 5.2	12.8 \pm 3.4	28.1 \pm 4.2	9.5 \pm 1.7	21.1 \pm 2.8
C-SWM ($K = 1$)	36.5 \pm 5.6	56.2 \pm 6.2	18.3 \pm 1.9	35.7 \pm 2.3	11.5 \pm 1.0	26.0 \pm 1.2
World Model (AE)	23.8 \pm 3.3	44.7 \pm 2.4	1.7 \pm 0.5	8.0 \pm 0.5	1.2 \pm 0.8	5.3 \pm 0.8
World Model (VAE)	1.0 \pm 0.0	5.1 \pm 0.1	1.0 \pm 0.0	5.2 \pm 0.0	1.0 \pm 0.0	5.2 \pm 0.0
SPACE INVADERS						
C-SWM ($K = 5$)	48.5 \pm 7.0	66.1 \pm 6.6	16.8 \pm 2.7	35.7 \pm 3.7	11.8 \pm 3.0	26.0 \pm 4.1
C-SWM ($K = 3$)	46.2 \pm 13.0	62.3 \pm 11.5	10.8 \pm 3.7	28.5 \pm 5.8	6.0 \pm 0.4	20.9 \pm 0.9
C-SWM ($K = 1$)	31.5 \pm 13.1	48.6 \pm 11.8	10.0 \pm 2.3	23.9 \pm 3.6	6.0 \pm 1.7	19.8 \pm 3.3
World Model (AE)	40.2 \pm 3.6	59.6 \pm 3.5	5.2 \pm 1.1	14.1 \pm 2.0	3.8 \pm 0.8	10.4 \pm 1.3
World Model (VAE)	1.0 \pm 0.0	5.3 \pm 0.1	0.8 \pm 0.2	5.2 \pm 0.0	1.0 \pm 0.0	5.2 \pm 0.0
3-BODY PHYSICS						
C-SWM	100 \pm 0.0	100 \pm 0.0	97.2 \pm 0.9	98.5 \pm 0.5	75.5 \pm 4.7	85.2 \pm 3.1
World Model (AE)	100 \pm 0.0	100 \pm 0.0	97.7 \pm 0.3	98.8 \pm 0.2	67.9 \pm 2.4	78.4 \pm 1.8
World Model (VAE)	100 \pm 0.0	100 \pm 0.0	83.1 \pm 2.5	90.3 \pm 1.6	23.6 \pm 4.2	37.5 \pm 4.8
Physics WM (PAIG)	89.2 \pm 3.5	90.7 \pm 3.4	57.7 \pm 12.0	63.1 \pm 11.1	25.1 \pm 13.0	33.1 \pm 13.4

Использованные источники

- ◆ Contrastive Learning of Structured World Models *Thomas Kipf, Elise van der Pol, Max Welling*
- ◆ A Tutorial on Energy-Based Learning *Yann LeCun, Sumit Chopra, Raia Hadsell, Marc'Aurelio Ranzato, and Fu Jie Huang*
- ◆ What Is Contrastive Learning? *Ram Sagar*
- ◆ World Models *David Ha, Jürgen Schmidhuber*