Generative Multi-Agent Behavioral Cloning*

Zhan et al.

2018

1 What

Generative models are hard. Hierarchical modelling is super hard. Multi-agent behaviour modelling is hard. The paper combines all of this and presents a method capable of modelling complex interaction between basketball players and generate realistic trajectories of multiple players over long time periods. The research impact is not limited to sports only, it can also be applied in modelling animal behaviour, video-game and music generation.

Contributions:

- novel Behaviour Cloning (BC) problem setting, where an optimal policy is probabilistic, multi-modal and closed-loop (an agent must take other guys around into consideration)
- hierarchical generative policy
- an approach that uses amortised variational inference
- empirical demonstration of the superiority of the approach to existing baselines

2 Why

The paper is working in a multi-agent scenario when the behaviour is multi-modal. The agents are contiually interacting with each other and, knowing each others' macro goals should, for instance, prevent two players from doing the same (if my teammate is in this region of the pitch, I shouldn't go there).

*Notes by Vitaly Kurin `https://yobibyte.github.io/`

3 How (kudos to authors for providing the tex sources)

3.1 Background

A variational autoencoder (VAE) [Kingma and Welling, 2013] is a generative model for non-sequential data that injects latent variables $\mathbf{z}$ into the joint distribution $p_\theta(\mathbf{x}, \mathbf{z})$ and introduces an inference network parametrised by ϕ to approximate the posterior $q_\phi(\mathbf{z} \mid \mathbf{x})$. The learning objective is to maximise the evidence lower-bound (ELBO) of the log-likelihood with respect to the model parameters θ and ϕ:

$$\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] + D_{KL}(q_\phi(\mathbf{z} \mid \mathbf{x})||p_\theta(\mathbf{z})) \tag{1}$$

VRNN [Chung et al., 2015] = VAEs + RNNs, which can be trained by maximizing ELBO:

$$p_\theta(\mathbf{z}_t|\mathbf{x}_{<t}, \mathbf{z}_{<t}) = \varphi_{\text{prior}}(\mathbf{h}_{t-1}) \qquad \text{(prior)} \tag{2}$$

$$q_\phi(\mathbf{z}_t|\mathbf{x}_{\leq T}, \mathbf{z}_{<t}) = \varphi_{\text{enc}}(\mathbf{x}_t, \mathbf{h}_{t-1}) \qquad \text{(inference)} \tag{3}$$

$$p_\theta(\mathbf{x}_t|\mathbf{z}_{\leq t}, \mathbf{x}_{<t}) = \varphi_{\text{dec}}(\mathbf{z}_t, \mathbf{h}_{t-1}) \qquad \text{(generation)} \tag{4}$$

$$\mathbf{h}_t; = f(\mathbf{x}_t, \mathbf{z}_t, \mathbf{h}_{t-1}). \qquad \text{(recurrence)} \tag{5}$$

$$\mathbb{E}_{q_\phi(\mathbf{z}_{\leq T}|\mathbf{x}_{\leq T})}\Bigg[\sum_{t=1}^{T} \log p_\theta(\mathbf{x}_t \mid \mathbf{z}_{\leq T}, \mathbf{x}_{<t}) - D_{KL}\Big(q_\phi(\mathbf{z}_t \mid \mathbf{x}_{\leq T}, \mathbf{z}_{<t})||p_\theta(\mathbf{z}_t \mid \mathbf{x}_{<t}, \mathbf{z}_{<t})\Big)\Bigg] \tag{6}$$

3.2 Notation

- Let $\mathcal{X}, \mathcal{A}$ denote the state, action space.
- Let $\mathbf{s}_{\leq T} = \{\mathbf{s}_t\}_{1 \leq t \leq T}$ denote a demonstration, where $\mathbf{s}_t = (\mathbf{x}_t, \mathbf{a}_t) = (\{\mathbf{x}_t^k\}_{\text{agents } k}, \{\mathbf{a}_t^k\}_{\text{agents } k})$. $\mathbf{x}_t^k \in \mathcal{X}$, $\mathbf{a}_t^k \in \mathcal{A}$ are the state, action of agent k at time t.
- Let $\tau_t = \{(\mathbf{x}_u, \mathbf{a}_u)\}_{1 \leq u \leq t}$ denote the history of state-action pairs.
- Let $\pi_\theta(\mathbf{x}_t, \tau_{t-1})$ denote a (multi-agent) stochastic policy parametrised by θ that samples actions from the probability distribution $p_\theta(\mathbf{a}_t|\mathbf{x}_t, \tau_{t-1})$.
- Let π_E denote the (multi-agent) expert stochastic policy that generated the data $\mathcal{D}$, and $\mathbf{s}_{\leq T} \sim \pi_E$ to denote that $\mathbf{s}_{\leq T}$ was generated from policy π_E.
- Let $\mathcal{M}(\mathbf{x}_t, \mathbf{a}_t)$ denote a (possibly probabilistic) transition function on states: $\mathbf{x}_{t+1} \sim p_{\mathcal{M}}(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{a}_t)$.

Important assumption for motion tracking set up! The transitions are deterministic: $x_{t+1} = x_t + a_t$.

There are more simplifying assumptions in the paper, but I'll not copy them here since there is already a lot of copypaste =).

Each agent is modelled using its own latent $\mathbf{z}_t^k$:

$$\pi_\theta^k(\tau_{t-1}) \sim p_\theta^k(\mathbf{x}_t^k|\mathbf{x}_{<t}) = \varphi^k(\mathbf{z}_t^k, \mathbf{h}_{t-1}^k), \tag{7}$$

$$\mathbf{h}_t^k = f^k(\mathbf{x}_t^k, \mathbf{z}_t^k, \mathbf{h}_{t-1}^k). \tag{8}$$

Macro-goal variables $\mathbf{g}_t$:

$$\pi_\theta^k(\tau_{t-1}) \sim p_\theta(\mathbf{x}_t^k|\mathbf{x}_{<t}) = \varphi^k(\mathbf{z}_t^k, \mathbf{h}_{t-1}^k, \mathbf{g}_t) \tag{9}$$

Macro-goals sampling:

$$p(\mathbf{g}_t|\mathbf{g}_{<t}) = \varphi_g(\mathbf{h}_{g,t-1}, \mathbf{x}_{t-1}), \tag{10}$$

$$\mathbf{h}_{g,t} = f_g(\mathbf{g}_t, \mathbf{h}_{g,t-1}). \tag{11}$$

We can train an agent independently from macro-goal policies, maximising log-likelihood of macro-goals $\mathbf{g}_{\leq T}$ and maximizing VRNN's ELBO for each of the agents:

$$\mathbb{E}_{q^k(\mathbf{z}^k_{\leq T}|\mathbf{x}^k_{\leq T}, \mathbf{g}_{\leq T})} \Bigg[\sum_{t=1}^{T} \log p_\theta^k(\mathbf{x}_t^k|\mathbf{z}^k_{\leq T}, \mathbf{x}^k_{<t}, \mathbf{g}_{\leq T}) \\ - D_{KL}\Big(q_\phi^k(\mathbf{z}_t^k|\mathbf{x}^k_{\leq T}, \mathbf{z}^k_{<t}, \mathbf{g}_{\leq T}) \Big\| p_\theta^k(\mathbf{z}_t^k|\mathbf{x}^k_{<t}, \mathbf{z}^k_{<t}, \mathbf{g}_{<t})\Big) \Bigg]. \tag{12}$$

4 Evaluation

The authors provide an extensive description of the whole experimental set up and code as well [1]

The baselines are RNN and two variants of VRNN.

4.1 Quantitative evaluation

This one is interesting. Log-likelihood of the proposed approach is only slightly better than the VRNN-indep version from the baselines. The authors mention [Theis et al., 2015] saying that log-likelihood does not necessarily indicate higher quality of generated samples. They recruit 14 experts to compare their

[1] `https://github.com/ezhan94/gen-MA-BC`

model with VRNN, and the proposed model is superior to the baselines (VRNN-indep is preferred in only 6 out of 25 comparisons and 4 times it's a tie).

It's super hard to compare generative models, and it's great that the authors have made some steps beyond eye-balling. However, it would be interesting to think about possible flaws in this evaluation apart from having a larger amount of humans to test (and asking non-professionals for their judgement).

4.2 Qualitative evaluation

An interesting thought here is that hierarchical approach makes models more interpretable. We can reason about the intent of the agents by looking at the sampled macro-goals.

Qualitative analysis shows that the proposed BC might still suffer from compounding error and have difficulties with generalisation.

The cool thing is that having these macro-goals enables us to have more control over the agents during inference. We can ground macro-goals manually instead of sampling them from the model.

4.3 Models analysis

Another great addition to the paper. The authors change parameters and observe the analyse the model behaviour under these different conditions. The result of this reasoning influenced the future decisions on model architectures. More details in the original paper.

5 Comments

- Demo is nice `http://basketball-ai.com/`
- one possible research direction is to use something more sophisticated than BC in the same set up
- In related work section the authors seem to confuse learning the cost/reward and learning the value when they're talking about IRL
- I really like the paper, though it's not that easy to grasp because of notation/assumption heaviness, though it's really great that the author explicitly point out all the assumptions and we do not need to guess
- the authors ignore ball modelling, which makes a problem somehow harder, somehow easier. It would be interesting to see it in a follow up
- macro-goals here are location-specific due to the nature of the task, it would be cool to see the development of this to more abstract goals

References

[Chung et al., 2015] Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A. C., and Bengio, Y. (2015). A recurrent latent variable model for sequential data. In *Advances in neural information processing systems*, pages 2980–2988.

[Kingma and Welling, 2013] Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

[Theis et al., 2015] Theis, L., Oord, A. v. d., and Bethge, M. (2015). A note on the evaluation of generative models. *arXiv preprint arXiv:1511.01844*.