

Learning for Companion Robots: Preparation and Adaptation

Xavier Alameda-Pineda

▶ To cite this version:

Xavier Alameda-Pineda. Learning for Companion Robots: Preparation and Adaptation. CAp - RFIAP 2024 - Joint Conférences sur l'Apprentissage automatique and Reconnaissance des Formes, Image, Apprentissage et Perception, Jul 2024, Lille, France. hal-04634985

HAL Id: hal-04634985 https://hal.science/hal-04634985

Submitted on 4 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Learning for Companion Robots: Preparation and Adaptation

Xavier Alameda-Pineda, Inria @ UGA, LJK, CNRS

Joint RFIAP CAP 2024

Where are we?



Why these wandering computers?



Interactive autonomous systems

- ➤ Multi-modal
- Perceive, Act, & React
- Impact on people

Learning-based components

- ➢ High-cost real data
- Limited on-line computation

Need for preparation/adaptation









DVAE

Dynamical Variational AutoEncoders



Successor Feature Representation Learning

SPRING

Social Robotic Experiments in a Day-care Hospital









Apologies in advance for the equations ;)

Dynamical Variational Autoencoders^[1]



Probabilistic models approximate the data distribution with a parametric proxy.

Interest of latent variables: model hidden phenomena with probabilities.





Clean speech // Noisy speech

Ball position // Detection

Latent variables add spice to learning: how to maximise^[2,3,4] the log-likelihood?

$$\log p(\boldsymbol{x}) = \underbrace{\mathbb{E}_{q(\boldsymbol{z})} \left[\log \frac{p(\boldsymbol{z}) p(\boldsymbol{x} | \boldsymbol{z})}{q(\boldsymbol{z})} \right]}_{\text{M-step or VLB}} + \underbrace{D_{\text{KL}} \left(q(\boldsymbol{z}) \left\| p(\boldsymbol{z} | \boldsymbol{x}) \right)}_{\text{E-step}} \right]$$

Step 1: Define the model

$\int p_{oldsymbol{ heta}}(oldsymbol{z}) \& p_{oldsymbol{ heta}}(oldsymbol{x} oldsymbol{z})$	(prior & likelihood)
$\int q_{\phi}(\boldsymbol{z}) pprox p_{\boldsymbol{\theta}}(\boldsymbol{z} \boldsymbol{x})$	(approx. posterior)

[2] A.P. Dempster, et. al., (1977), Journal of the Royal Statistical Society.
[3] M. I. Jordan, et. al., (1999), Machine Learning.
[4] D. Kingma, et. al., (2013), International Conference on Learning Representations.

Step 2: Learning and inference

$$\begin{aligned} \operatorname*{argmax}_{\boldsymbol{\theta}, \boldsymbol{\phi}} \mathcal{L}_{\mathsf{ELBO}}(\boldsymbol{\theta}, \boldsymbol{\phi}) \\ \operatorname*{argmax}_{\boldsymbol{z}} q_{\boldsymbol{\phi}^*}(\boldsymbol{z} | \boldsymbol{x}) \end{aligned}$$

VAEs learn a probabilistic representation with latent non-linear dependencies:



Maximise a lower bound:

$$\boxed{\log p_{\boldsymbol{\theta}}(\boldsymbol{x}) \geq \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z})} \Big[\log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{z})p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})}{q_{\boldsymbol{\phi}}(\boldsymbol{z})}\Big]}$$

From what we've defined:

J	$\int p_{oldsymbol{ heta}}(oldsymbol{z}) \& p_{oldsymbol{ heta}}(oldsymbol{x} oldsymbol{z})$	(prior & likelihood)
	$\left(q_{\boldsymbol{\phi}}(\boldsymbol{z}) pprox p_{\boldsymbol{\theta}}(\boldsymbol{z} \boldsymbol{x}) ight)$	(approx. posterior)

[4] D. Kingma, et. al., (2013), International Conference on Learning Representations. Image credit medium.com (https://medium.com/@rushikesh.shende/autoencoders-variational-autoencoders-vae-and-%CE%B2-vae-ceba9998773d)

Fundamental limitation of VAE: frame-wise modeling

From a VAE perspective, these two speech signals are **exactly** the same:



Model sequences of observations & latent variables:

$$p_{\boldsymbol{\theta}}^{\text{DVAE}}(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T}) \neq \prod_{t=1}^{T} p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t}, \boldsymbol{z}_{t}) = p_{\boldsymbol{\theta}}^{\text{VAE}}(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T})$$



As in VAEs, the posterior needs to be approximated. Simplified using **D-separation**.^[5]

$$q_{m{\phi}}(m{z}_{1:T}|m{x}_{1:T}) = \prod_{t=1}^{T} q_{m{\phi}}(m{z}_t|m{z}_{1:t-1},m{x}_{1:T})$$

Learning is performed by maximising the "temporal" ELBO.^[1]

$$\mathcal{L}(\boldsymbol{x}_{1:T}; \boldsymbol{\phi}, \boldsymbol{\theta}) = \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{q}\boldsymbol{\phi}(\boldsymbol{z}_{1:t} | \boldsymbol{x}_{1:T})} \underbrace{\left[\underbrace{\ln p_{\boldsymbol{\theta}_{\boldsymbol{x}}}(\boldsymbol{x}_{t} | \boldsymbol{x}_{1:t-1}, \boldsymbol{z}_{1:t})}_{\text{Reconstruction}} \right] \\ - \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{q}\boldsymbol{\phi}(\boldsymbol{z}_{1:t-1} | \boldsymbol{x}_{1:T})} \underbrace{\left[\underbrace{D_{\text{KL}} \left(q_{\boldsymbol{\phi}}(\boldsymbol{z}_{t} | \boldsymbol{z}_{1:t-1}, \boldsymbol{x}_{1:T}) \, \middle\| \, p_{\boldsymbol{\theta}_{\boldsymbol{z}}}(\boldsymbol{z}_{t} | \boldsymbol{x}_{1:t-1}, \boldsymbol{z}_{1:t-1}) \right)}_{\text{Regularization}} \right]$$

[5] Bishop, C. M., (2006).[1] L. Girin, et. al., (2021), Foundations and Trends on Machine Learning.

Mixtures of DVAEs for unsupervised multi-source tracking^[6]



Motivation: multi-source trajectory modeling



Trajectories @ t-1

Observations @ t

Desired output

MixDVAE Intuition

Multiple tracks: track-detection assignment latent variable. Tracks not observed directly: object position is a latent variable.





GMM (Multi-source No dynamics)

K

KalmanDVAE(1-source(1 sourceLinear dyn.)Non-linear)

MixKalman (Multi-source Linear dyn.) MixDVAE (Multi-source Non-linear)

MixDVAE algorithm

Pre-trained DVAE (synthetic data) + Variational Expectation Maximisation (VEM)



Quick discussion

Let us recall de GMM update and the Kalman update:



$$egin{aligned} m{s}_t = m{P}_t m{o}_t + m{T}_t m{s}_{t-1} \ m{ extstyle extst$$

The MixDVAE update looks like:

$$\boldsymbol{s}_{tn} = \boldsymbol{P}_t \sum_{k} \eta_{kn} \boldsymbol{o}_{tk} + \underbrace{\boldsymbol{T}_t(\boldsymbol{s}_{1:t-1})}_{\text{non-lin. prediction}}$$

Experiments in multiple object tracking, and sound source separation.

Ongoing: adapting non-linear model, complexity, learning from noisy data, ...

Multimodal DVAEs for audio-visual speech modeling^[7]



Can DVAEs model multi-modal data?



Affective audio-visual speech^[8]



What should we model?

- ➤ □Static AV (ID, emotion)
- Dynamic AV(lip-audio corr.)
- Dynamic Audio (pitch, tone)
- Dynamic Visual (eye AUs)

Multimodal DVAE (MDVAE)



What should we model?

- ➤ □Static AV (ID, emotion)
- Dynamic AV(lip-audio corr.)
- Dynamic Audio (pitch, tone)
- Dynamic Visual (eye AUs)

Quantization helps!

MDVAE works better coupled with quantisation^[9]



Qualitative results (1)



Same visual dynamics

Same audio-visual dyn.

Qualitative results (2)



Emotion interpolation

Appearance interpolation

23

Successor Feature Representations^[10]



Motivation

Teach a humanoid robot to join a group of people

- Combine audio, visual and optimal motion?
- Deep Reinforcement Learning?



Basics of value-based RL

 $(\mathcal{S}, \mathcal{A}, T, R, \gamma)$

Markov decision process:

- state space,
- action space,
- transition probability,
- reward function,

- discount factor.



$$Q^{\pi}(s_t, a_t) = R(s_t, a_t) + \gamma \mathbb{E}_{\pi, T \mid s_t, a_t} \{ R(s_{t+1}, a_{t+1}) + \gamma^2 R(s_{t+2}, a_{t+2}) + \ldots \}$$

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(R(s_t, a_t) + \gamma Q(s_{t+1}, a^*) - Q(s_t, a_t))$$

Let's use *features*

At least two issues: data-greedy and reward design.



Dependency of the reward function on these features.

$$R(\phi) = R(\phi_{\text{object}}, \phi_{\text{people}}, \phi_{\text{speed}})$$

Long training for every new task!



Problem: even if the case of feature-based rewards, we need lots of training data.

Can we re-use what we learn thanks to these feature-based rewards?

Transferring knowledge?



First paradigm was linear^[11] dependencies:

$$R_k(s_t, a_t) = w_k^\top \phi(s_t, a_t)$$

How does the action-state value function write?

$$Q_k^{\pi}(s_t, a_t) = \mathbb{E}_{\pi} \left\{ \sum_{\ell \ge 0} \gamma^{\ell} R_k(s_{t+\ell}, a_{t+\ell} | s_t, a_t) \right\}$$
$$= \mathbb{E}_{\pi} \left\{ \sum_{\ell \ge 0} \gamma^{\ell} w_k^{\top} \phi(s_{t+\ell}, a_{t+\ell} | s_t, a_t) \right\}$$
$$= w_k^{\top} \sum_{\ell \ge 0} \gamma^{\ell} \mathbb{E}_{\pi} \left\{ \phi(s_{t+\ell}, a_{t+\ell} | s_t, a_t) \right\}$$
$$= w_k^{\top} \psi^{\pi}(s_t, a_t)$$

Definition of "Q"

$$R_k(s_t, a_t) = w_k^\top \phi(s_t, a_t)$$

Task-policy separation

Re-using knowledge?



Generalised policy improvement: A cumulative greedy policy...

...cannot be too far away.

$$\tilde{\pi}(s_t) = \arg\max_{a} \max_{j \in [1, K-1]} w_K^\top \psi^{\pi_j}(s_t, a)$$

$$\|\pi_K - \tilde{\pi}\| \le \rho(\min_j \|R_K - R_j\|)$$

31

Reward does not always depend linearly on the features: we need an extension.

$$\begin{split} Q_k^{\pi}(s_t, a_t) &= \mathbb{E}_{\pi} \left\{ \sum_{\ell \ge 0} \gamma^{\ell} R_k(s_{t+\ell}, a_{t+\ell} | s_t, a_t) \right\} & R_k(s_t, a_t) = R_k(\phi(s_t, a_t)) \\ &= \mathbb{E}_{\pi} \left\{ \sum_{\ell \ge 0} \gamma^{\ell} R_k(\phi(s_{t+\ell}, a_{t+\ell} | s_t, a_t)) \right\} \\ &= \sum_{\ell \ge 0} \gamma^{\ell} \int_{\Phi} R_k(\phi) p(\phi(s_{t+\ell}, a_{t+\ell}) | \pi, s_t, a_t) \mathrm{d}\phi \\ &= \int_{\Phi} R_k(\phi) \xi^{\pi}(\phi, s_t, a_t) \mathrm{d}\phi \end{split}$$
 Task-policy separation 32

Linear vs. non-linear

	Linear	Non-linear
$R_k(s_t, a_t)$	$w_k^{ op}\phi(s_t,a_t)$	$R_k(\phi(s_t, a_t))$
$Q_k^{\pi}(s_t, a_t)$	$w_k^{ op}\psi^{\pi}(s_t, a_t)$	$\int_{\Phi} R_k(\phi) \xi^{\pi}(\phi, s_t, a_t) \mathrm{d}\phi$

The generalised policy improvement (greedy cumulative policy) holds for both.

What are we missing so far? The learning operators!

The learning operators

$$\begin{aligned} & \text{For a given transition } (s_t, a_t, s_{t+1}) & a^* = \arg \max_a Q_k^{\pi}(s_t, a) \\ & Q_k^{\pi}(s_t, a_t) & Q_k^{\pi}(s_t, a_t) + = \alpha (R_k(s_t, a_t) + \gamma Q_k^{\pi}(s_{t+1}, a^*) - Q_k^{\pi}(s_t, a_t)) \\ & w_k^{\top} \psi^{\pi}(s_t, a_t) & \psi^{\pi}(s_t, a_t) + = \alpha (\phi(s_t, a_t) + \gamma \psi^{\pi}(s_{t+1}, a^*) - \psi^{\pi}(s_t, a_t)) \\ & \int_{\Phi} R_k(\phi) \xi^{\pi}(\phi, s_t, a_t) d\phi & \xi^{\pi}(s_t, a_t) + = \alpha (p(\phi|s_t, a_t) + \gamma \xi^{\pi}(s_{t+1}, a^*) - \xi^{\pi}(s_t, a_t)) \end{aligned}$$

Where is the task/environment in the feature-based learning operators?

A few results



Socially Pertinent Robots in Gerontological Healthcare^[12]

(a lot of people)

[12] X. Alameda-Pineda, et al., (2024), International Journal of Social Robotics, Under review.

Goal: to develop socially assistive robots with the capacity of performing multi-person interactions and open-domain dialogue.



Experiments at gerontological day-care Broca hospital (AP-HP)

Data Protection & Ethics



Ethics compliance

Phase 1 (CER Paris Université):

- Ethnographic study
- Living lab simulated study

Phase 2 (CPP):

- Day-care hospital waiting room

Software architecture

More than 50 software modules interconnected using ROS



Acceptability and usability were measure with standard questionnaires:

Wave (#pers)	First (20)	Second (49)
Acceptability ^[14] (max 10)	3.96 ± 0.05	6.17 ± 0.33
Usability ^[15] (max 10)	4.79 ± 2.40	5.70 ± 2.30

Comments:

- Scales for consumer products (so proxy).
- Outstanding results for an experimental platform.
- Improvement after taking feedback into account... and including LLMs.

- Deep is nice, useful, performance, etc.
 But it's great to ground things with mathematical models.
- > Fusing multiple modalities helps **representing each other**.
- > Very challenging to deploy models in robotic platforms.
- Beyond limited computation, there scientific challenges linked to data distribution shifts.
- > Complete software architectures required **combined specific expertises**.
- When experimenting with social robots (and hence with humans) one must be very careful: what is the **impact on the user**?

Thanks

All RobotLearn-ers and collaborators!



The funding programs supporting our research:







And you for listening. Question (and answer?) time!

I am available for longer discussions: I would like to hear about your topic(s)!!!