PRINCETON UNIVERSITY

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Inference via Interpolation: Contrastive Representations Provably Enable Planning and Inference

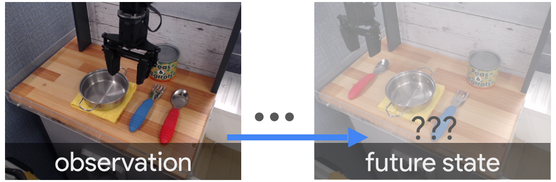
Benjamin Eysenbach*, Vivek Myers*, Ruslan Salakhutdinov, Sergey Levine

Code and paper!

https://github.com/vivekmyers/contrastive_planning

Q: Inference in high-dimensional time series?

Prediction/forecasting:



Planning/inpainting:

References

[1] Sermanet, P., et al. (2018). Time-contrastive networks: Self-supervised learning from video. ICRA.
[2] Oord, A. v. d., Li, Y., and Vinyals, O. (2018). Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748.
[3] Eysenbach, B, et al. (2022). Contrastive learning as goal-conditioned reinforcement learning. NeurIPS.
[4] Ma, Z., Collins, M. (2018). Noise Contrastive Estimation and Negative Sampling for Conditional Models: Consistency and Statistical Efficiency. EMNLP.
[5] Wang, T., Isola, P. (2020). Understanding contrastive representation learning through alignment and uniformity on the hypersphere. ICML.
[6] Zheng, Chongyi, et al. (2024). Stabilizing Contrastive RL: Techniques for Robotic Goal Reaching from Offline Data. ICLR.
[7] Ma, Yecheng Jason, et al. (2023). VIP: Towards Universal Visual Reward and Representation via Value-Implicit Pre-Training." ICLR.
[8] Zhao, Shengjia, et al. (2017). Towards deeper understanding of variational autoencoding models." arXiv preprint arXiv:1702.08658.

Open Questions

1. Applications to various time series datasets?
2. Can the assumptions always be satisfied?
3. How to extend this to large video/audio/text datasets?
4. Contrastive learning across multiple datasets (e.g., text ←→ videos ←→ audio)
5. How does this relate to dynamic programming?
6. Might it inherit some TD-like properties (e.g., combinatorial generalization, shorter paths)?

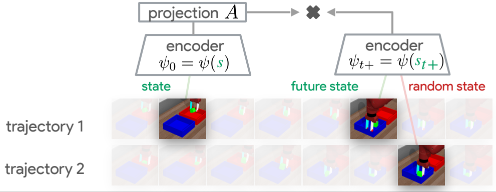
Key Idea

Key idea: predict representations instead of observations.

- How should these representations be learned so they retain bits relevant to prediction and planning?
- Avoid reconstruction methods (e.g., seq. VAE [Zhao '17]), which are computationally expensive.

There's already exists a method in the literature that does this!

Temporal Contrastive Learning [1,2,3]



Setup and assumptions

Objective: (Symmetric) InfoNCE, with a quadratic energy function:

$$\min_{A,\psi(\cdot)} \mathbb{E}_{\substack{(x_0, x_T^{(1)}) \sim p(x_0, x_T) \\ x_T^{(i)} \sim p(x_T)}} \left[\log \frac{e^{-\left\| A\psi(x_0) - \psi(x_T^{(1)}) \right\|_2^2}}{\sum_{i=1}^B e^{-\left\| A\psi(x_0) - \psi(x_T^{(i)}) \right\|_2^2}} \right]$$

such that *expected* representation norm is small: $\frac{1}{k}\mathbb{E}_{p(x)} \left[\|\psi(x)\|_2^2 \right] \le c$

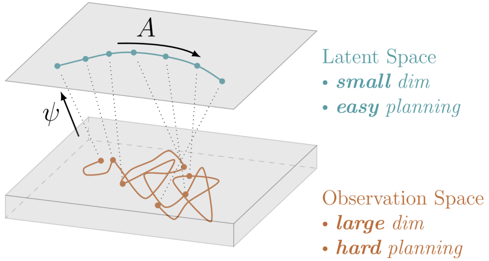
Assumption 1: (Bayes-optimal classifier) [4]

$$e^{-\frac{1}{2}\|A\psi(x_0) - \psi(x_T)\|_2^2} = \frac{p(x_T \mid x_0)}{p(x)C}$$

Assumption 2: (Pushforward measure is Gaussian) [5]

$$p(\psi) = \mathcal{N}(\psi; \mu = 0, \sigma = c \cdot I).$$

Theoretical Results



Theorem 1 (Prediction): Representation of a future state is Normally distributed, with mean that is a linear function of the current state representation.

$$p(\psi_{t+} = \psi \mid \psi_0) = \mathcal{N}\left(\mu = \frac{c}{c+1} A\psi_0, \Sigma = \frac{c}{c+1} I\right)$$

Theorem 2 (Planning): Representation of a intermediate state is Normally distributed, with mean that is a linear function of the current and final state representations.

$$p(\psi_w \mid \psi_0, \psi_{t+}) = \mathcal{N}\left(\psi_w; \mu = \Sigma^{-1}(A^T\psi_{t+} + A\psi_0), \Sigma = \tfrac{c}{c+1} A^T A + \tfrac{c+1}{c} I\right)$$

Theorem 3 (Multi-Step Planning): Joint distribution of representations is a Gaussian Markov chain, so intermediate representations are linear functions of initial and final representations.

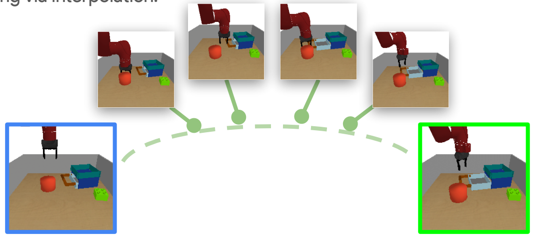
$$p(\psi_i \mid \psi_0, \psi_{t+}) = \mathcal{N}\left(\psi_i; \mu_i = (\Sigma\eta)^{(i)}, \Sigma_i = (\Lambda^{-1})^{(i,i)}\right)$$

(Σ and Λ are block matrices constructed from A.)

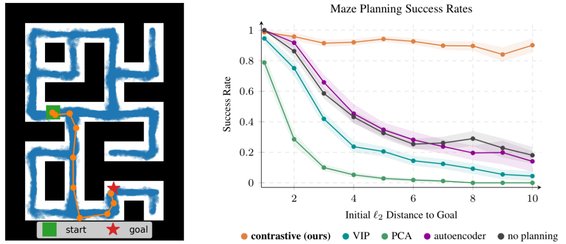
Corollary 3.1 (Special Case of Multi-Step Planning): If c is large and A is a rotation matrix, intermediate representations are a simple convex combination of initial and final representations.

Explains observation from prior work [6]

Planning via interpolation:



Application: Planning via (warped) interpolation



Visualizing the plans (TSNE)