Efficient Exploration in Deep Reinforcement Learning

A Gaussian-Processes-Tensor-Decomposition Approach

Baddar, Mohamed

July 16, 2020

Table of contents

Table of Contents

Exploration vs Exploitation

- Main RL objective : find best sequence of action in an uncertain environment. [1]
- Exploitation: Apply highest reward action based on model at hand
- Exploration: Explore *possibly* higher reward action
- Usually have limitation over number of interaction with the environment (time, cost)
- Finding the right balance is usually challenging
- One of the main stream research direction:
  - Quantify uncertainty in the reward model, for example the Q-NN.[2][3]
  - Utilized quantified uncertainty for efficient exploration[4]

Bayesian Reinforcement Learning

- Stochastic model for reward function
  $r_t = E_{\theta \sim p(\theta|D)}(R_\theta(a_t, s_t))$ [5]
- D is the training data, tuples of records of actions, states and corresponding rewards (s_t, a_t, r_t)
- Quantify uncertainty in reward distribution and use it for efficient exploration

Table of Contents

Thompson Sampling

- Model the uncertainty in the model parameters via approximate posterior[6]
- The uncertainty propagates to the target (reward) via sampling
- Proven to converge asymptotically to optimal reward [7][8]

for *t in 1:T* do
    Sample $\hat{\theta}$ from (approximate) posterior q
    $a_t \leftarrow \text{argmax}_{a_t} E_{\theta \sim q}(R_\theta(a_t, s_t))$
    Apply a_t
    Observe s_t and r_t
    Update approximate posterior q
end

Table of Contents

Cast Deep Neural Network as GP-LVM

- Core idea: Create as stochastic generative process that "emulates" deep neural networks[9][2][3]
- Variational Methods to approximate the latent posterior distribution and find the posterior predictive distribution $P(y^*|y) = \int p(y^*, x, z|y)$ [2][10][11][12]

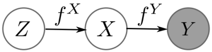

Figure: Two-Layer Gaussian Process

$$
\begin{aligned}
&y_{nd} = f_d^Y(\mathrm{x}_n) + \epsilon_{nd}^Y, \quad d = 1, \ldots, D, \quad \mathrm{x}_n \in \mathcal{R}^Q \\
&x_{nq} = f_q^X(\mathrm{z}_n) + \epsilon_{nq}^X, \quad q = 1, \ldots, Q, \quad \mathrm{z}_n \in \mathcal{R}^Q z \\
&f^Y \sim \mathcal{GP}\left(0, k^Y(\mathrm{X}, \mathrm{X})\right) \\
&f^X \sim \mathcal{GP}\left(0, k^X(\mathrm{Z}, \mathrm{Z})\right) \\
&k(\mathrm{x}_i, \mathrm{x}_j) = (\sigma_{se})^2 \exp\left(-\frac{(\mathrm{x}_i - \mathrm{x}_j)^2}{2l^2}\right)
\end{aligned}
\tag{1}
$$

Bayesian Training I

- Optimize the log evidence[2]

$$\log p(\mathrm{Y}) = \log \int_{\mathrm{X},\mathrm{Z}} p(\mathrm{Y} \mid \mathrm{X}) p(\mathrm{X} \mid \mathrm{Z}) p(\mathrm{Z}) \tag{2}$$

- maximize the ELBO instead

$$\mathcal{F}_v = \int_{\mathrm{X},\mathrm{Z},\mathrm{F}^Y,\mathrm{F}^X} \mathcal{Q} \log \frac{p\left(\mathrm{Y}, \mathrm{F}^Y, \mathrm{F}^X, \mathrm{X}, \mathrm{Z}\right)}{\mathcal{Q}} \tag{3}$$

- Decompose the Joint distribution

$$\begin{gathered} p\left(\mathrm{Y}, \mathrm{F}^Y, \mathrm{F}^X, \mathrm{X}, \mathrm{Z}\right) = \\ p\left(\mathrm{Y} \mid \mathrm{F}^Y\right) p\left(\mathrm{F}^Y \mid \mathrm{X}\right) p\left(\mathrm{X} \mid \mathrm{F}^X\right) p\left(\mathrm{F}^X \mid \mathrm{Z}\right) p(\mathrm{Z}) \end{gathered} \tag{4}$$

- the terms X, Z appear in highly non linear manner$P(F|X), p(F|Z)$ respectively

Bayesian Training II

- The term appears in double exponential function (first exp from Gaussian definition , second from kernel definition). Also the second level is an inverse function, which makes integration intractable

Table of Contents

Challenges

- Mathematical tractability (illustrated above)[2][12]
- Scalability [10][13]
- Generality [14][15][16]

Scalability

- Gaussian Process $f(x)$ calculation

$$
\begin{gathered}
f(x) \sim \mathcal{N}(m(x), k_{\theta}(x, x^{'}) \\
m_{\mathrm{y}}(\mathrm{x}) = K_{\mathrm{x}n}\left(\sigma^2 I + K_{nn}\right)^{-1} \mathrm{y} \\
k_{\mathrm{y}}\left(\mathrm{x}, \mathrm{x}'\right) = k\left(\mathrm{x}, \mathrm{x}'\right) - K_{\mathrm{x}n}\left(\sigma^2 I + K_{nn}\right)^{-1} K_{n\mathrm{x}'}
\end{gathered} \tag{5}
$$

- Inversion of a NxN matrix is of $O(N^3)$

Inducing points

- Select set of pseudo-inputs (or latents) X_m where $M << N$ with corresponding Gaussian process value of f_m[10][12]
- The input (or latent) space will be (X, X_m) with corresponding Gaussian process distribution $P(f, f_m)$
- Assume that X_m selection is independent from X selection (Key assumption in mathematical derivation for lower bounds.
- Variational-EM approach by maximizing the ELBO for marginal likelihood $log(p(y))$ to simultaneously select X_m and optimize its variational posterior parameters $\phi(f_m) \sim N(\mu, A)$
- Inducing points solve two problems 1. Mathematical tractability (find a closed from lower bound for Marginal likelihood for the GP-LVM model, and reduce computation cost from $O(N^3) to O(NM^2)$

Tensor + GP + Inducing points

- Tensors can learn high-order correlation from data efficiently
- Recent work has been published trying to mathematically connect GP with Tensor Regression[13]
- Furthermore, another direction is to apply hybrid inducing points + Tensors to scale GP to billions of inducing points [17]

Current Research direction(s)

- Deep understanding of inducing points methods [10][12]
- understand the connection between Tensor Regression and inducing points[13]
- Explore applying Tensor + GP methods for scalable uncertainty quantification in the context of RL exploration scalable˙gp˙tensor˙train˙dec
- Explore tackling the Generality problem (no determined yet, Generalized Gaussian Process or Normalizing flows) [14][15]

[1] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018, ISBN: 0262039249.

[2] A. Damianou and N. Lawrence, "Deep gaussian processes," in *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, C. M. Carvalho and P. Ravikumar, Eds., ser. Proceedings of Machine Learning Research, vol. 31, Scottsdale, Arizona, USA: PMLR, 29 Apr–01 May 2013, pp. 207–215. [Online]. Available: http://proceedings.mlr.press/v31/damianou13a.html.

[3] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *Proceedings of The 33rd International Conference on Machine Learning*, M. F. Balcan and K. Q. Weinberger, Eds., ser. Proceedings of Machine Learning Research, vol. 48, New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 1050–1059. [Online]. Available: http://proceedings.mlr.press/v48/gal16.html.

[4] C. Riquelme, G. Tucker, and J. Snoek, *Deep bayesian bandits showdown: An empirical comparison of bayesian deep networks for thompson sampling*, 2018. arXiv: 1802.09127 [stat.ML].

[5] M. Ghavamzadeh, S. Mannor, J. Pineau, and A. Tamar, "Bayesian reinforcement learning: A survey," *CoRR*, vol. abs/1609.04436, 2016. arXiv: 1609.04436. [Online]. Available: http://arxiv.org/abs/1609.04436.

[6] D. Russo, B. V. Roy, A. Kazerouni, and I. Osband, "A tutorial on thompson sampling," *CoRR*, vol. abs/1707.02038, 2017. arXiv: 1707.02038. [Online]. Available: http://arxiv.org/abs/1707.02038.

[7] E. Kaufmann, N. Korda, and R. Munos, “Thompson sampling: An asymptotically optimal finite-time analysis,” in *Proceedings of the 23rd International Conference on Algorithmic Learning Theory*, ser. ALT'12, Lyon, France: Springer-Verlag, 2012, pp. 199–213, ISBN: 9783642341052. DOI: 10.1007/978-3-642-34106-9_18. [Online]. Available: https://doi.org/10.1007/978-3-642-34106-9_18.

[8] S. Agrawal and N. Goyal, “Analysis of thompson sampling for the multi-armed bandit problem,” in *Proceedings of the 25th Annual Conference on Learning Theory*, S. Mannor, N. Srebro, and R. C. Williamson, Eds., ser. Proceedings of Machine Learning Research, vol. 23, Edinburgh, Scotland: PMLR, 25–27 Jun 2012, pp. 39.1–39.26. [Online]. Available: http://proceedings.mlr.press/v23/agrawal12.html.

[9] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, ser. Adaptive Computation and Machine Learning. Cambridge, MA, USA: MIT Press, Jan. 2006, p. 248.

[10] M. Titsias, "Variational learning of inducing variables in sparse gaussian processes," in *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, D. van Dyk and M. Welling, Eds., ser. Proceedings of Machine Learning Research, vol. 5, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA: PMLR, 16–18 Apr 2009, pp. 567–574. [Online]. Available: http://proceedings.mlr.press/v5/titsias09a.html.

[11] M. K. Titsias, "Variational model selection for sparse gaussian process regression,", 2008.

[12] M. Titsias and N. D. Lawrence, "Bayesian gaussian process latent variable model," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Y. W. Teh and M. Titterington, Eds., ser. Proceedings of Machine Learning Research, vol. 9, Chia Laguna Resort, Sardinia, Italy: PMLR, 13–15 May 2010, pp. 844–851. [Online]. Available: http://proceedings.mlr.press/v9/titsias10a.html.

[13] R. Yu, G. Li, and Y. Liu, *Tensor regression meets gaussian processes*, 2017. arXiv: 1710.11345 [cs.LG].

[14] B. Wang and J. Q. Shi, "Generalized gaussian process regression model for non-gaussian functional data," *Journal of the American Statistical Association*, vol. 109, no. 507, pp. 1123–1133, 2014, ISSN: 01621459. [Online]. Available: http://www.jstor.org/stable/24247440.

[15] I. Kobyzev, S. Prince, and M. A. Brubaker, *Normalizing flows: An introduction and review of current methods*, 2019. arXiv: 1908.09257 [stat.ML].

[16] D. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach and D. Blei, Eds., ser. Proceedings of Machine Learning Research, vol. 37, Lille, France: PMLR, Jul. 2015, pp. 1530–1538. [Online]. Available: http://proceedings.mlr.press/v37/rezende15.html.

[17] P. Izmailov, A. Novikov, and D. Kropotov, "Scalable gaussian processes with billions of inducing inputs via tensor train decomposition," in *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, A. Storkey and F. Perez-Cruz, Eds., ser. Proceedings of Machine Learning Research, vol. 84, Playa Blanca, Lanzarote, Canary Islands: PMLR, Sep. 2018, pp. 726–735. [Online]. Available: http://proceedings.mlr.press/v84/izmailov18a.html.