

Modelling and Inference in Latent Dynamical Systems

Neha Das

neha.das@tum.de

February 9, 2019

1. Introduction

Modelling sequential data and inferring their characteristics is central to a lot of interesting problems, including analysis and deconstruction of auditory sequences [26], prediction in sequential data such as video frames [16] and text [28], and reinforcement learning applications [5]. While the last has been approached through both model-based and model-free techniques, learning a model of the environment has been showed to be more sample efficient [1]. Although a few successful solutions have been proposed for relatively simpler environments and low dimensional observation spaces, scaling them to partially observed dynamical environments with complex transitions (such as interacting particles or switching dynamics) and high-dimensional observations (like images) in a principled manner remains a challenge.

This proposal strives to briefly discuss the characteristics desirable in the learned model of a dynamic system, the hurdles in incorporating those characteristics and possible directions including recently published approaches towards overcoming them.

2. Desirable Characteristics in the model of a dynamical system

- **Incorporation of Latent Representation and Dynamics:** Many interesting dynamic systems have state spaces that are not directly observable or have high dimensional observations such as image sequences which, while easy to obtain, are highly correlated in the pixel (dimensional) space and often lie on much lower dimensional manifold. Incorporating such observations directly in the model's transition dynamics is hard. Many solutions [19, 22] thus assume the underlying graphical model to have a latent dynamical representation (e.g the State Space Models). The two major tasks then, are to learn this generative model, as well as infer the latent states from the indirect observations.
- **Uncertainty Modelling:** Having a quantifiable measure of uncertainty of the model's outputs is essential

to many applications, including data-efficient machine learning and high risk prediction applications such as medical diagnosis and autonomous driving. This uncertainty in outputs may stem from noisy observations (aleatoric uncertainty) and/or the parameters and the structure of the model (epistemic uncertainty) [9]. Representation of both these uncertainty forms is a desirable characteristic in a dynamic system's model.

- **Probabilistic prediction in state-space far into the future:** This is particularly significant for reinforcement learning tasks with long planning trajectories. While a Recurrent Neural Network can store the belief from previous observations and predict observations far into time using deterministic one-step prediction of observation and feedback, it fails to capture the uncertainties in the transition model.
- **Interpretable/Disentangled latent-space representation:** Evaluation of inference models is easier when the latents are interpretable - i.e they can be mapped to some recognizable quantity. In recent experiments, disentanglement in the latent space has been shown to lead to more interpretability[15, 20]. Disentangled latents may also provide added control over generating data with desired properties[29].
- **Incorporation of Partial Knowledge of the Dynamical System:** Many dynamical systems of interest have partially known complexities. Some examples include the presence of interacting particles/agents (for e.g, bouncing balls in box or similar limbs of a robot) and switching dynamics (for e.g, for an agent in a maze environment, the transition dynamics is different when navigating on an obstacle free path from when colliding with walls). A desirable property of the model would be to incorporate these partially known characteristics in a principled manner in order to take advantage of the prior knowledge they present.

3. Related Approaches and Open Problems

This section discusses the previous work and the open challenges in imbibing the desirable characteristics outlined

above in a model of sequential data.

- **Incorporation of Latent Representation and Dynamics:** Inference in latent dynamical systems is a well studied problem. Classical approaches towards this include Kalman Filters and its variants [18, 17, 27]. However, these approaches assume that the system dynamics are fully known, which limits their applicability. Recent methods for approximate inference through Sequential Monte Carlo (SMC) techniques [6] and Variational Inference [2, 4, 19] have risen as a popular alternative to classical methods. These methods do not make any assumptions regarding the knowledge of the generative model; in fact, the generative model can be learnt alongside inference. These methods, however, are plagued with limitations of their own. SMC methods for instance, trade smaller sample complexity for inference accuracy and are heavily reliant on the proposal distribution, while Variational Inference methods are often limited to factorized Gaussian distributions in the latent space. Recently however, approaches that combine and glean advantages of both methods have been proposed [13, 24]. However, providing a theoretically supported approximate inference model that maintains the delicate balance between inferring from the observations in addition to supporting the learning of transition dynamics remains a challenge for both approaches [19].
- **Uncertainty Modelling:** In 2011, PILCO [5] introduced probabilistic Gaussian Processes in sequence modelling as an alternative to deterministic Recurrent Neural Network based models that are often susceptible to model bias. An added benefit of this approach was that it was data-efficient and could learn from just a few trials. However, PILCO scales poorly to high dimensional observations and tasks that require more number of trials. Deep PILCO [11], therefore, replaces Gaussian Processes with Bayesian Neural Networks (BNN) in order to scale well to high dimensional observations while still avoiding overfitting on small datasets similar to PILCO. Both variants, however, apply transitions in the observational space as opposed to having a latent dynamical system, the advantages of which have been highlighted in the previous section. Recently proposed latent dynamical architectures based on State Space Models on the other hand [19, 22], are explicitly probabilistic (i.e they output distribution parameters) in the models for inference, transition dynamics and emission, thus being able to model both the aleatoric and epistemic uncertainties satisfactorily. Although they do not model weight uncertainty like BNNs do, introduction of weight uncertainty in form of dropout [10] may be straightforward.

It would be interesting to study the effect of this inclusion on the network's performance.

- **Probabilistic prediction in state-space far into the future:** While models like DVBF [19] have the ability to make long term predictions via a series of probabilistic one-step predictions, the prediction quality has been noted to degrade with increasing time. Recently proposed models [12, 14] take a look back at the Partially Observed Markov Decision Process (POMDP) literature and advocate the incorporation of belief states in order to alleviate this problem.
- **Interpretable/Disentangled latent-space representation:** While disentanglement of latent states has been explored in the context of unsupervised modelling of non-temporal datasets [15, 20], there has been limited research towards modelling sequential data with disentangled latents. Recent works by Yinghzen and Mandt [29] and Fraccaro et. al [8] broach upon this subject by disentangling static and dynamic states by explicitly modelling the variables representing the content and the dynamics of the sequence separately. While this model demonstrated disentanglement in the static (applicable over all time steps) latents, it would also be beneficial to model disentanglement in the dynamic latents as well as transition parameters (somewhat accomplished via switching dynamics models - see next section).
- **Incorporation of Partial Knowledge of the Dynamical System:** The problem of structurally incorporating switching dynamics has been explored in many past and recent works [7, 23, 3]. The task of modelling systems with interactive particles and/or agents with multiple co-dependent actuators has also recently been approached using graph neural networks [21, 25] with some success. Both the explored approaches (with the exception of the work by Fox et al. [7]), however, make assumptions regarding the knowledge/or the upper bound on number of transition functions/interacting particles; and in case of the paper by Sanchez et al. [25], explicit knowledge of the structure of the agent. An approach towards dealing with unknown number of interacting components may be found in non-parametric distributions. Similarly dense graph models can be explored in an attempt to model unknown nature connections between components.

4. Conclusion

While research in the field of modelling and inference in Latent Dynamical Systems has come a long way, there still remain a lot of issues that challenge the practical applicability of these models to a complex and real dynamic

environment. This proposal outlines the characteristics of a good general model for complex dynamic environments, the motivations for desiring them and discusses the challenges in imbibing them in learned models.

References

- [1] C. G. Atkeson and J. C. Santamaria. A comparison of direct and model-based reinforcement learning. In *Robotics and Automation, 1997. Proceedings., 1997 IEEE International Conference on*, volume 4, pages 3557–3564. IEEE, 1997. 1
- [2] J. Bayer and C. Osendorfer. Learning stochastic recurrent networks. *arXiv preprint arXiv:1411.7610*, 2014. 2
- [3] P. Becker-Ehmck, J. Peters, and P. van der Smagt. Switching linear dynamics for variational bayes filtering, 2019. 2
- [4] J. Chung, K. Kastner, L. Dinh, K. Goel, A. C. Courville, and Y. Bengio. A recurrent latent variable model for sequential data. In *Advances in neural information processing systems*, pages 2980–2988, 2015. 2
- [5] M. Deisenroth and C. E. Rasmussen. Pilco: A model-based and data-efficient approach to policy search. In *Proceedings of the 28th International Conference on machine learning (ICML-11)*, pages 465–472, 2011. 1, 2
- [6] A. Doucet, N. De Freitas, and N. Gordon. An introduction to sequential monte carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer, 2001. 2
- [7] E. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky. Nonparametric bayesian learning of switching linear dynamical systems. In *Advances in Neural Information Processing Systems*, pages 457–464, 2009. 2
- [8] M. Fraccaro, S. Kamronn, U. Paquet, and O. Winther. A disentangled recognition and nonlinear dynamics model for unsupervised learning. In *Advances in Neural Information Processing Systems*, pages 3601–3610, 2017. 2
- [9] Y. Gal. Uncertainty in deep learning. *University of Cambridge*, 2016. 1
- [10] Y. Gal and Z. Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059, 2016. 2
- [11] Y. Gal, R. McAllister, and C. E. Rasmussen. Improving pilco with bayesian neural network dynamics models. In *Data-Efficient Machine Learning workshop, ICML*, 2016. 2
- [12] K. Gregor and F. Besse. Temporal difference variational auto-encoder. *CoRR*, abs/1806.03107, 2018. 2
- [13] S. S. Gu, Z. Ghahramani, and R. E. Turner. Neural adaptive sequential monte carlo. In *Advances in Neural Information Processing Systems*, pages 2629–2637, 2015. 2
- [14] Z. D. Guo, M. G. Azar, B. Piot, B. A. Pires, T. Pohlen, and R. Munos. Neural predictive belief representations. *CoRR*, abs/1811.06407, 2018. 2
- [15] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2017. 1, 2
- [16] D. Jayaraman, F. Ebert, A. A. Efros, and S. Levine. Time-agnostic prediction: Predicting predictable video frames. *arXiv preprint arXiv:1808.07784*, 2018. 1
- [17] R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *Journal of basic engineering*, 83(1):95–108, 1961. 2
- [18] R. E. Kalman et al. Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana*, 5(2):102–119, 1960. 2
- [19] M. Karl, M. Soelch, J. Bayer, and P. van der Smagt. Deep variational bayes filters: Unsupervised learning of state space models from raw data. *arXiv preprint arXiv:1605.06432*, 2016. 1, 2
- [20] H. Kim and A. Mnih. Disentangling by factorising. *arXiv preprint arXiv:1802.05983*, 2018. 1, 2
- [21] T. Kipf, E. Fetaya, K.-C. Wang, M. Welling, and R. Zemel. Neural relational inference for interacting systems. *arXiv preprint arXiv:1802.04687*, 2018. 2
- [22] R. G. Krishnan, U. Shalit, and D. Sontag. Deep kalman filters. *arXiv preprint arXiv:1511.05121*, 2015. 1, 2
- [23] S. W. Linderman, A. C. Miller, R. P. Adams, D. M. Blei, L. Paninski, and M. J. Johnson. Recurrent switching linear dynamical systems. *arXiv preprint arXiv:1610.08466*, 2016. 2
- [24] C. A. Naesseth, S. W. Linderman, R. Ranganath, and D. M. Blei. Variational sequential monte carlo. *arXiv preprint arXiv:1705.11140*, 2017. 2
- [25] A. Sanchez-Gonzalez, N. Heess, J. T. Springenberg, J. Merel, M. Riedmiller, R. Hadsell, and P. Battaglia. Graph networks as learnable physics engines for inference and control. *arXiv preprint arXiv:1806.01242*, 2018. 2
- [26] R. E. Turner. *Statistical models for natural sounds*. PhD thesis, UCL (University College London), 2010. 1

- [27] E. A. Wan and R. Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158. Ieee, 2000. [2](#)
- [28] Z. Xie. Neural text generation: A practical guide. *arXiv preprint arXiv:1711.09534*, 2017. [1](#)
- [29] L. Yingzhen and S. Mandt. Disentangled sequential autoencoder. In *International Conference on Machine Learning*, pages 5656–5665, 2018. [1](#), [2](#)