RESEARCH STATEMENT

Hao Wu, Northeastern University

1 Overview

Machine learning models serve as a data-driven mechanism for reasoning about real-world problems. Recent advances in deep learning have demonstrated that neural networks can achieve remarkable performance when employing large amounts of data and computation. However, there are many problems where we only have small amounts of data or few annotations. Moreover, deep models can produce overconfident predictions, particularly when presented with long-tail inputs or a collection of related tasks.

The challenges are becoming increasingly prominent in domains at the forefront of AI research. One typical application is scene understanding in computer vision, which requires compositional reasoning about data in terms of coherently representing its heterogeneity and the casual relationship between objects. Another example is trajectory prediction in autonomous vehicles, where it is necessary to generalize across distinct-yet-related tasks in a manner that is robust to rare events. Moreover, these challenges emerge from applications in science and healthcare where tasks often reveal themselves as a large amount of imbalanced data or a collection of task datasets, thus discovering the struture of data requires modeling of the underlying domain properties.

My research goal is to develop deep probabilistic models for learning interpretable representations that can uncover the structure of data. This development aids the aforementioned challenges by incorporating structured priors as inductive bias, which regularizes the models and helps to improve generalization, particularly in problems with limited data or few annotations. On the other hand, scaling up variational inference methods to complex domains poses challenges. To address this, my research combines variational inference with importance sampling, as a means of developing methods for accurate inference that can 1) scale to structured models with high-dimensional correlated variables and 2) quantify uncertainty by various amount of data inputs. Next I will illustrate my research contributions and propose future work.

2 Prior Work

2.1 Combining Variational Inference with Importance Sampling

One main challenge in machine learning research is reasoning compositionally about the world as humans do. My research tackles this problem by designing deep generative models with more complex and structured priors. As we start to build up these models, we increasingly see limitations of standard inference methods when scaling them up to high-dimensional, correlated (and discrete) latent variables.

A recent insight is that any importance sampling method can be used to define a variational objective. Initially, this insight has focused on combining importance sampling with traditional variational objectives in the form of a lower bound [1–4]. More recently, it has lead to a resurgence of interest in methods that minimize an inclusive rather than an exclusive KL divergence [5, 6]. This line of developments has inspired me to propose the *amortized population Gibbs* (APG) samplers [7], which frame variational inference as adaptive importance sampling with amortized proposals. This approach exploits structured priors as inductive bias, which allows us to generate high-quality proposals by decomposing a single high-dimensional inference problem into a sequence of lower-dimensional inference problems. This class of methods is applicable to discrete latent variables without reparameterization and score-function that prior methods of variational inference may require.

In my collaboration with Zimmermann et al., we generalized this idea and proposed *nested variational inference* [8], a class of amortized inference methods that construct proposals with nested importance sampling. This research further motivated work on *inference combinators* [9], where Stites and Zimmermann develop a simple grammar to define importance samplers that can be trained using variational methods. With these *inference combinators*, users can concisely tailor sampling strategies to suite particular probabilistic programs, with the guarantee that the resulting sampler always preserves proper weighting with respect to the desired target density. This work sets in motion a precedent for languages of inference, simplifying the development of inference programming and making inference methods accessible to the broader research community.

2.2 Learning Structured Representations with Minimal Supervision

One importance use case of deep generative models is to encode meaningful factors of variation of data in the form of latent representations. In the absence of supervision, the models do not naturally learn representations that are associated with an interpretable data structure. This has motivated the work, on which I collaborated with Esmaeili et al., where we designed a variational objective to induce disentangled representations so that individual dimensions of latent variables represent interpretable factors of variation of data [10].

However, further challenges arise when we move towards problems that have data with complex heterogeneity. Without supervision, uncovering the inherent structure of data requires incorporating inductive bias in the form of structured priors of models. This inspiration led to my design of APG samplers [7], which tackles such problems where the data consists of a corpus of related tasks, within each there is an individual group of observations. By incorporating structured priors, this approach learns representations for not only individual features at the observation level, but also aggregated factors of variation at the corpus and group levels.

For the purpose of unsupervised representation learning, generative models may not achieve meaningful abstractions of data. Since the generators learn to reconstruct data examples given latent variables, those variables must encode factors of variation that give rise to large discrepencies in data space, regardless of whether these factors are semantically meaningful. To provide an alternative to learning a generative process, Esmaeili and I jointly developed the *conjugate energy-based models* [11]. The core idea is to characterize consistency between data and latent variables in the compressed representation space without requiring learning a generator that reconstructs input data. Our models encode image data into more abstract representations that can be used for various downstream tasks such as few-label classification and out-of-domain detection.

3 Proposed Work

3.1 Characterizing Model Uncertainty from Sparse Data

In probabilistic modeling, one key question is how to quantify model uncertainty in a manner that appropriately accounts for various amounts of input data. This becomes particularly difficult when we infer task representations over a collection of related tasks, each containing a limited number of examples. Typical problems can be meta-learning the distribution over clusters across a collection of mixture models or the dynamics over trajectories across a collection of state-space models. In these unsupervised few-shot learning cases, we hope that the models can generalize from the training tasks to unseen ones at test time.

To tackle this problem, I propose to reason about the model uncertainty in terms of neural sufficient statistics [7], which is the encoded features of each data example. Specifically this development parameterizes deep generative models in form of (quasi-)conjugate exponential families; This parameterization allows us to quantify the posterior variance by various amount of available data. On the other hand, I will continue to work on inference methods that provide flexible sampling strategies for generating good proposals. Also I will actively collaborate with the researchers in probabilistic programming, where I can contribute to the development of probabilistic programming systems for user-programmable inference methods.

3.2 Likelihood-free Methods for Structured Representation Learning

Despite the success in generation tasks achieved by deep generative models, they still have big challenges in learning semantically meaningful representations. In fact, learning to generate realistic examples may not always lead to more useful representations. When representing scenes, for example, it may be useful to discard information, such as the advertising on the side of a bus. For such use cases, we may want more abstract representations than factors that merely give rise to descrepencies in terms of data reconstruction.

To address this, I propose to learn representations in a likelihood-free manner by combining *contrastive learning* techniques with *structured model priors*. Contrastive learning objectives will support training neural estimators for the ratio between the likelihood and the marginal likelihood, which hopefully will allow us to more easily discard nuisance factors from the data space. On the other hand, structured priors can regularize the latent space with inductive bias, so that the models can encode more meaningful representations that uncover the structure of complex data modalities.

References

[1] Yuri Burda, Roger Grosse, and Ruslan Salakhutdinov. "Importance weighted autoencoders". In: *arXiv preprint arXiv:1509.00519* (2015).

[2] Christian Naesseth, Scott Linderman, Rajesh Ranganath, and David Blei. "Variational sequential monte carlo". In: *International conference on artificial intelligence and statistics*. PMLR. 2018, pp. 968–977.

[3] Tuan Anh Le, Maximilian Igl, Tom Rainforth, Tom Jin, and Frank Wood. "Auto-encoding sequential monte carlo". In: *arXiv preprint arXiv:1705.10306* (2017).

[4] Chris J Maddison, Dieterich Lawson, George Tucker, Nicolas Heess, Mohammad Norouzi, Andriy Mnih, Arnaud Doucet, and Yee Whye Teh. "Filtering variational objectives". In: *arXiv preprint arXiv:1705.09279* (2017).

[5] Jörg Bornschein and Yoshua Bengio. "Reweighted wake-sleep". In: *arXiv preprint arXiv:1406.2751* (2014).

[6] Tuan Anh Le, Adam R Kosiorek, N Siddharth, Yee Whye Teh, and Frank Wood. "Revisiting reweighted wake-sleep". In: (2018).

[7] Hao Wu, Heiko Zimmermann, Eli Sennesh, Tuan Anh Le, and Jan-Willem Van De Meent. "Amortized Population Gibbs Samplers with Neural Sufficient Statistics". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 10421–10431.

[8] Heiko Zimmermann, Hao Wu, Babak Esmaeili, and Jan-Willem van de Meent. "Nested Variational Inference". In: *Conference on Neural Information Processing Systems*. 2021.

[9] Sam Stites, Heiko Zimmermann, Hao Wu, Eli Sennesh, and Jan-Willem van de Meent. "Learning Proposals for Probabilistic Programs with Inference Combinators". In: *Uncertainty in Artificial Intelligence*. 2021.

[10] Babak Esmaeili, Hao Wu, Sarthak Jain, Alican Bozkurt, Narayanaswamy Siddharth, Brooks Paige, Dana H Brooks, Jennifer Dy, and Jan-Willem Meent. "Structured disentangled representations". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2019, pp. 2525–2534.

[11] Hao Wu, Babak Esmaeili, Michael Wick, Jean-Baptiste Tristan, and Jan-Willem van de Meent. "Conjugate Energy-Based Models". In: *International Conference on Machine Learning*. PMLR. 2021.