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 causal 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 in life science and healthcare, where robustly modeling imbalanced data requires incorporating domain knowledge.

My research goal is to develop deep probabilistic models for learning interpretable representations that uncover structure from data. This development aids the aforementioned challenges by incorporating inductive bias in form of structured priors, which regularize deep generative models and improve generalization and robustness to sparse data. 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 amounts 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 to compositionally reason about the world as humans do. My research tackles this problem by designing structured priors as inductive bias in deep generative models. Our hope is that these inductive biases can provide necessary domain knowledge while retaining model flexibility. As we start to build up these models, we increasingly see limitations of variational inference methods when scaling them up to high-dimensional, correlated (and discrete) latent variables.

A recent insight is combining importance sampling techniques methods with variational methods [1, 2]. This line of developments has inspired me to propose the *amortized population Gibbs samplers* [3], which performs amortized inference on deep generative models with neural Gibbs proposals. This approach iterates between approximate Gibbs updates to blocks of latent variables, which means we decompose a high-dimensional inference problem into a sequence of lower-dimensional inference problems. Also, we define a variational objective based on importance sampling, so that this approach is applicable to models with hybrid discrete-continuous latent variables, where prior work may require reparameterization or score-function technique for gradient computation.

In my collaboration with Zimmermann et al., we generalized this idea and proposed *nested variational inference* [4], a class of amortized inference methods that construct proposals with nested importance sampling. This research further motivated work on *inference combinators* [5], where Stites and Zimmermann develop a simple grammar to define importance samplers that can be trained using variational methods. 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 [6].

However, further challenges arise when we move toward data with more complex heterogeneity. In absence of supervision, inference tasks require incorporating inductive bias as domain knowledge. Amortized population Gibbs samplers [3] exploit structured priors as inductive biases for learning representations on a collection of individual tasks such as clustering and unsupervised tracking. Our inference methods learn representations that are coherent in hierarchical structure inferring global features by aggregating local features.

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 discrepancies 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* [7]. 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 downstream tasks such as few-label classification and out-of-domain detection.

3 Proposed Work

3.1 Inferring Task Representations From A Small Number of Examples

As we build more complicated deep generative models, we also hope to develop fast and accurate inference methods. Many applications have a typical scenario where we hope to solve a collection of related inference tasks. In the domains of autonomous vehicles and healthcare analysis, a task is typically defined by inferring time-invariant features such as time dynamics or object representations over a small number of time series; In the domains of large-scale data analysis, common tasks, such as clustering and classification, are defined by certain factors of variation (e.g. colors or shapes of objects) that divide data into subgroups.

Thus one key question is how to infer *task representations* from a small number of training examples, where the task representations are characteristics that are jointly presented by certain groups of examples. Modeling the distributions over those characteristics allows us to efficiently make inference on unseen-yet-related tasks and reason about prediction uncertainty.

To address this, I propose to design inference algorithms that incorporate the structure of deep generative models, which is analogous to variational messaging passing in conjugate probabilistic models. We have shown the potential of this idea by proposing neural sufficient statistics [3], which define approximate conjugate models that allow us to infer high-level characteristics by aggregating lower-level features. Also I plan to exploit different neural architectures such as Transformers, which can provide good scalability and flexibility. On the other hand, I will continue to work on inference methods that provide flexible sampling strategies for generating good proposals. I will continue to collaborate with researchers in probabilistic programming, where I can contribute to developing 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 discrepancies 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 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] Jörg Bornschein and Yoshua Bengio. "Reweighted wake-sleep". In: *arXiv preprint arXiv:1406.2751* (2014).

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

[3] 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.

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

[5] 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.

[6] 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.

[7] 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.