RESEARCH STATEMENT

Hao Wu, Northeastern University

Overview. Recent advances in deep learning have proven that models can achieve remarkable performance when employing deep neural networks and a large amount of data and computation. Nevertheless, overparameterized deep models can produce overconfident predictions. Moreover, there are many problems where we only have small amount of data or little annotation, in which cases the models can fail to generalize to long-tail cases or unseen tasks.

My research goal is to advance deep probabilistic programming as a means of developing interpretable models that uncover structure of problems, reason about uncertainty, and robustly generalize across tasks. This development combines strengths of deep learning and probabilistic programming. Deep probabilistic programs employ neural networks to parameterize models, thus the models inherit the scalability of deep architectures and benefit from optimization strategies by stochastic gradient descent. Probabilistic programming, on the other hand, allows us to design deep generative models that incorporate necessary domain knowledge as inductive bias if need be. Inductive bias regularizes the models in form of structured priors, which helps to characterize uncertainty and improve generalization, particularly in problems where limited data and little supervision is available. Next I will illustrate my research contributions and propose work in future.

Amortized Inference with Probabilistic Programs. Amortized inference methods learn neural proposals that approximate the model distribution of interest by optimizing a variational objective. As we start to design deep generative models with more complex and structured priors, we increasingly see limitations of standard inference methods when scaling them up to models with a large number of correlated latent variables. To address this, there has been a body of work on developing lower-variance gradient estimation methods for more stable training, and optimizing network architectures to achieve better scalability.

A recent insight is that any importance sampling method can be used to define a variational objective. This has inspires me to propose the amortized population Gibbs (APG) samplers [1], a general framework that frames variational inference as adaptive importance sampling with amortized proposals. This sampling strategy exploits the structure of models, specifically the conditional independence, to generate high-quality proposals in an incremental manner; This approach allows us to a single high-dimensional inference into a sequence of lower-dimensional inference problems. This class of methods have a broad range of applications including models with discrete variables, since its optimization does not require reparameterization. I collaborated with Zimmermann et al. to make a further stride on this idea. We developed the nested varitional inference [2], a class of amortized inference methods that construct proposals with the nested importance sampling [3].

In collaboration with Stites and Zimmermann et al., I contributed to the development of a language for inference programming in probabilistic programs, referred to as inference combinators [4]. We demonstrate how variational methods can be used to learn proposals and thereby define a language for user-programmable amortized inference. We implemented APG samplers as probabilistic programs using inference combinators, which is built upon the Probablistic Torch library [5].

Learning Structured Representation of Complex Domains. Deep generative models represent data in low-dimensional latent variables by associating a prior over latent variables with a generator. With no supervision, latent variables do not naturally encode meaningful factors of variation of data. This has motivated the work that I collaborated with Esmaeili et al., where we designed a variational objective to induce disentangled representations, so that individual dimensions of latent variables can uncover interpretable factors of variation of data [6].

However, further challenges arise when we move towards real-world domains, because those domains typically have data with complex heterogeneity. In the absence of supervision, I think it requires incorporating inductive bias that can provide the inherent structure of data but also retain enough flexibility. This inspiration led to my design of APG samplers [1], which tackle such problems where the data consists of a corpus of related tasks, within each there is an individual group of observations. By encoding structured priors as inductive bias, the APG generative models learn representations for not only individual features at observation level, but also aggregated factors of variation at corpus level and group level.

Furthurmore, complex data modalities may exhibit a large number of underlying generative factors that give rise to a combinatorial explosion of possible inputs. This poses challenges when training deep generative models for unsupervised representation learning; Since the generators need to accurately reconstruct data examples given latent variables, those latent variables must encode factors of variation that give rise to large discrepencies in data space, regardless of whether these factors are semantically meaningful. Thus, as an alternative, Esmaeili and I jointly developed the conjugate energy-based models [7], a class of latent-variable models that combines a conjugate prior with an energy-based likelihood. The core idea is to characterize consistency between data and latent variables at some compressed repsentation, 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.

Proposed Work. Real-World problems tyically have data that is imbalanced-distributed and thus only provides a limited feature subspace. This poses challenges for models to generalize across tasks and adapt to unseen domains. I believe that deep probabilistic programming is a way to aid robustness and generalization, because deep probabilistic models have presented potential to meta-learn the distribution over unsupervised domains by inductive bias. For this reason I propose to develop new models to tackle unsupervised few-shot learning tasks by learning structured representations of complex data modalities. To train these models, I will continue to work on amortized inference methods that can generate high-quality proposals using importance sampling schemes. In addition, I will actively collaborate with the researchers in probabilistic programming, where I can contribute to the integration of deep learning with user-programmable inference methods.

References

[1] 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.

[2] Heiko Zimmermann, Hao Wu, Babak Esmaeili, and Jan-Willem van de Meent. "Nested Variational Inference". In: *arXiv preprint arXiv:2106.11302* (2021).

[3] Christian Naesseth, Fredrik Lindsten, and Thomas Schon. "Nested sequential monte carlo methods". In: *International Conference on Machine Learning*. PMLR. 2015, pp. 1292–1301.

[4] Sam Stites, Heiko Zimmermann, Hao Wu, Eli Sennesh, and Jan-Willem van de Meent. "Learning Proposals for Probabilistic Programs with Inference Combinators". In: *arXiv preprint arXiv:2103.00668* (2021).

[5] N. Siddharth, Brooks Paige, Jan-Willem van de Meent, Alban Desmaison, Noah D. Goodman, Pushmeet Kohli, Frank Wood, and Philip Torr. "Learning Disentangled Representations with Semi-Supervised Deep Generative Models". In: *Advances in Neural Information Processing Systems 30*. 2017, pp. 5927–5937.

[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: *The 22nd 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: 2021. arXiv: `2106.13798 [cs.LG]`.