Internship - Master degree or Engineering School
Deep generative models for learning discrete sequence distributions

Duration 5 – 6 months, starting spring 2020

Internship location: Criteo AI Lab, Paris

Keywords: machine learning, deep learning, generative models, sequence distribution models

Contact: Patrick Gallinari (p.gallinari@criteo.com), Matthieu Kirchmeyer (m.kirchmeyer@criteo.com), Amin Mantrach (a.mantrach@criteo.com)

1 Context

Generating and modeling discrete sequences occurs in many domains such as natural language processing, music composition, robotics, speech processing, diagnosis, etc. It also occurs in any sequence of events characterizing user actions or behavior (web and mobiles, medical records, etc). It encompasses several generic machine learning problems like event prediction, modeling sequence distributions, sequence translation... Recent developments have made deep learning the state-of-the-art modeling paradigm for solving several problems in sequence modeling. For example, auto-regressive models exploiting recurrent neural network architectures are ubiquitously used. A more recent development relies on directly learning the distribution of the sequences.

2 Internship Description

Deep Neural Networks offer recent developments for learning complex data distributions, like Variational Autoencoders (VAEs) [Kingma and Welling, 2013], Generative Adversarial Networks (GANs) [Goodfellow et al., 2014] and more recently Normalizing Flow approaches [Dinh et al., 2017, Kingma and Dhariwal, 2018, Grathwohl et al., 2019], a promising approach for learning complex distributions. Up to now this has mainly been used for learning distributions that are continuous and for static data. Current approaches to sequence prediction mainly rely on auto-regressive maximum likelihood approaches. They perform extremely well on some domains but also come with several limitations like: slow sequential generation not adapted to current requirements; and deterministic prediction when there might be several potential futures. Generative models are good candidates for addressing these issues, but are still limited in their development so that their extension to the modeling and generation of text sequences or any sequence of discrete events remains complex. For GAN approaches for example, this does not lead to classical gradient based formulations of learning. Several directions have been explored up to now, based e.g. on reinforcement learning methods [Yu et al., 2016, Xu et al., 2018] or by combining autoregressive and adversarial models [Subramanian et al., 2018]. More recently some initial attempts for learning with normalizing flows have also started to be developed [Ma et al., 2019, Ziegler and Rush, 2019]. These models hold the promise of learning more expressive transformations.

The objective of the internship is to analyze the potential of normalizing flow models and VAE models for modeling sequences of discrete events. The application focus will be on textual and user interaction traces sequences modeling.

3 Requirements

The candidate will have a strong background in Statistical Machine Learning and a first experience in Deep Learning. He/She should be able to develop both theoretical research and practical implementations and experiments.

References

[Dinh et al., 2017] Dinh, L., Sohl-Dickstein, J., and Bengio, S. (2017). Density estimation using real nvp.

[Goodfellow et al., 2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc.

[Grathwohl et al., 2019] Grathwohl, W., Chen, R. T. Q., Bettencourt, J., Sutskever, I., and Duvenaud, D. (2019). Ffjord: Free-form continuous dynamics for scalable reversible generative models. *International Conference on Learning Representations*.

[Kingma and Dhariwal, 2018] Kingma, D. P. and Dhariwal, P. (2018). Glow: Generative flow with invertible 1x1 convolutions. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 10215–10224. Curran Associates, Inc.

[Kingma and Welling, 2013] Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. cite arxiv:1312.6114.

[Ma et al., 2019] Ma, X., Zhou, C., Li, X., Neubig, G., and Hovy, E. (2019). Flowseq: Non-autoregressive conditional sequence generation with generative flow. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Hong Kong.

[Subramanian et al., 2018] Subramanian, S., Mudumba, S. R., Sordoni, A., Trischler, A., Courville, A. C., and Pal, C. (2018). Towards text generation with adversarially learned neural outlines. In *Advances in Neural Information Processing Systems 31*, pages 7551–7563. Curran Associates, Inc.

[Xu et al., 2018] Xu, J., Ren, X., Lin, J., and Sun, X. (2018). Diversity-promoting GAN: A cross-entropy based generative adversarial network for diversified text generation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3940–3949, Brussels, Belgium. Association for Computational Linguistics.

[Yu et al., 2016] Yu, L., Zhang, W., Wang, J., and Yu, Y. (2016). Seqgan: Sequence generative adversarial nets with policy gradient. *CoRR*, abs/1609.05473.

[Ziegler and Rush, 2019] Ziegler, Z. and Rush, A. (2019). Latent normalizing flows for discrete sequences. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 7673–7682, Long Beach, California, USA. PMLR.