Contrat Doctoral — ED Galilée

Titre du sujet : Diffusion for Structure-aware Modern Natural Language Processing

- ➢ Unité de recherche : LIPN
- ➢ Discipline : Informatique
- ➢ Direction de thèse : Joseph Le Roux (PU), Antoine Rozenknop (MCF)
- ➢ Contact : rozenknop@lipn.fr, leroux@lipn.fr
- ➢ Domaine de recherche : Informatique, traitement automatique des langues, apprentissage automatique
- ➢ Mots-clés : Modèles de langues, modèles génératifs, modèles de diffusion

Context

Many recent advances in generative modelling for complex data such as language, image or video, are based on a notion of transport between two probability distributions, from noise to target. Such models include diffusion models [10] and the many variants of flows [9]. These architectures parametrize a denoiser by reversing a time-dependent noise-adding mechanism (close the diffusion phenomenon from Physics [7]). This has become a new paradigm for the probabilistic approach to Machine Learning with recent successes for image and video generation [14]. Still, these models' performance are often hindered by their simplistic denoising probability distributions.

This is the case for instance in Language Modelling (generating follow-up texts). While being at the heart of an active research community, with interesting properties such as theoretically parallelizable generation, flow-based and diffusion-based language models (DLMs) [1] struggle to reach the accuracy of auto-regressive (AR) language models that are the current standard for large language models [12]. In practice, generating with AR LMs makes use of coarse-to-fine cascading such as *speculative decoding* [8] where a small model is used to generate candidates which are simply validated by the large model. Since the bottleneck is the generation and that validation can be parallelized, this helps improving generation speed. This kind of method is not available for DLMs (although some early works exist such as [6]). Moreover, the so-called *reasonning* capabilities of AR LMs are often implemented as chain-of-thought generation [17], where the model is asked not only to respond to a query, but to generate an step-by-step explanation of the rationale behind the response. This paradigm is difficult to implement in the DLM training procedure. Moreover in *retrieval-augmented generation* (RAG) the generated depends not-only on a user prompt but also on a set of documents (or simply excerpts) that are retrieved from in accordance with the prompt but also on the generated text itself. Again this change of input size is difficult to reconciliate in DLMs. In summary, the DLM paradigm still lags behind the ARLM state-of-the-art.

Recent works, at LIPN [3, 15] and more globally [5, 13] showed that diffusion is compatible with structured prediction, typically Markov Random Fields (MRFs). MRFs have been used extensively in the pre-neural era of Natural Language Processing (NLP) and Computer Vision (CV). They assign probabilities to correlated random variables allowing to take linguistic/graphical structure into account (via sequences, trees, or more general graphs). In the case of flows and diffusion, this can help the denoising process by enforcing well-structuredness constraints, giving a priori information so that denoising is faster (require less intermediate steps).

Project

In this project we want to explore the structure of the probability field defined by Diffusion Models, more precisely the interaction between variational methods for MRFs [16] and generative models.

We plan to work on two axes:

1. **Acelerating Diffusion Inference** especially in the case of DLMs. We plan to start from methods developed for latent-variables MRFs such as *coarse-to-fine* (see for instance [11] for an example in NLP), that use marginal probabilities to iteratively prune the search space. Also we want to incorporate the method of [4] to diffuse only parts of the information to improve speed while preserving quality in our MRF models. In order cope with the size of the MRF, and the complexity of inference, we will use the Mean-Field approximation of [2].

2. **Coping with Structural Constraints in Multimodal Corpora**, with a focus on image-text tasks such as captionning and guided image generation. Following [5, 13] a MRF can be used to define probabilities inside in the image generation process by modelling interactions between pixels, or a latent reprentation of a subset of pixels. Recently [3] building on DLMs of [1] proposed a CRF model for text labelling. We want to explore how the definition of single MRF/CRF encompassing pixels and words can improve the image generation, with possible application to text generation of image/video by using the MRF *in the other direction*. This would allow a tighter coupling of texts and images, for more accurate generation.

References

[1] Jacob Austin et al. "Structured Denoising Diffusion Models in Discrete State-Spaces". In: *Advances in Neural Information Processing Systems*. Ed. by M. Ranzato et al. Vol. 34. Curran Associates, Inc., 2021, pp. 17981–17993. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/958c530554f78bcd8e97125b70e6973d-Paper.pdf.

[2] Justin Domke. "Learning Graphical Model Parameters With Approximate Marginal Inference". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.10 (2013), pp. 2454–2467. DOI: 10.1109/tpami.2013.31. URL: http://dx.doi.org/10.1109/TPAMI.2013.31.

[3] Nicolas Floquet, Joseph Le Roux, and Nadi Tomeh. "Approximate Structured Diffusion for Sequence Labelling". In: *ARR Jan 26*. under review. 2026. URL: https://openreview.net/forum?id=2yQIrlLSDP.

[4] Daniel Mingyi Israel, Guy Van den Broeck, and Aditya Grover. "Accelerating Diffusion LLMs via Adaptive Parallel Decoding". In: *The Thirty-ninth Annual Conference on Neural Information Processing Systems*. 2026. URL: https://openreview.net/forum?id=xwqTt26NJf.

[5] Sadeep Jayasumana et al. "MarkovGen: Structured Prediction for Efficient Text-to-Image Generation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2024, pp. 9316–9325.

[6] Tero Karras et al. "Guiding a Diffusion Model With a Bad Version of Itself". In: *CoRR* (2024). arXiv: 2406.02507 [cs.CV]. URL: http://arxiv.org/abs/2406.02507v3.

[7] Paul Langevin. "Sur la théorie du mouvement brownien". In: *Comptes-Rendus de l'Académie des Sciences* 146 (1908), pp. 530–532.

[8] Yaniv Leviathan, Matan Kalman, and Yossi Matias. "Fast Inference From Transformers Via Speculative Decoding". In: *CoRR* (2022). arXiv: 2211.17192 [cs.LG]. URL: http://arxiv.org/abs/2211.17192v2.

[9] Yaron Lipman et al. *Flow Matching Guide and Code*. 2024. arXiv: 2412.06264 [cs.LG]. URL: https://arxiv.org/abs/2412.06264.

[10] Calvin Luo. "Understanding Diffusion Models: a Unified Perspective". In: *CoRR* (2022). arXiv: 2208.11970 [cs.LG]. URL: http://arxiv.org/abs/2208.11970v1.

[11] Slav Petrov and Dan Klein. "Improved Inference for Unlexicalized Parsing". In: *Proceedings of the conference on Human Language Technologies and the conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL'07)*. 2007.

[12] Alec Radford et al. "Language Models are Unsupervised Multitask Learners". In: (2019).

[13] Kanchana Ranasinghe et al. *LatentCRF: Continuous CRF for Efficient Latent Diffusion*. 2024. arXiv: 2412.18596 [cs.CV]. URL: https://arxiv.org/abs/2412.18596.

[14] Robin Rombach et al. *High-Resolution Image Synthesis with Latent Diffusion Models*. 2022. arXiv: 2112.10752 [cs.CV]. URL: https://arxiv.org/abs/2112.10752.

[15] Alexandre Schulz. "Improved Combinatorial Optimization Learning with Graphical Models". PhD thesis. Université Sorbonne Paris Nord, 2026.

[16] Martin J Wainwright and Michael I Jordan. "Graphical models, exponential families, and variational inference". In: *Foundations and Trends® in Machine Learning* 1.1–2 (2008), pp. 1–305.

[17] Jason Wei et al. "Chain-Of-Thought Prompting Elicits Reasoning in Large Language Models". In: *CoRR* (2022). arXiv: 2201.11903 [cs.CL]. URL: http://arxiv.org/abs/2201.11903v6.