

Discrete Generative Modeling with Masked Diffusions

Modern generative AI has developed along two distinct paths: autoregressive models for discrete data (such as text) and diffusion models for continuous data (like images). Bridging this divide by adapting diffusion models to handle discrete data represents a compelling avenue for unifying these disparate approaches. However, existing work in this area has been hindered by unnecessarily complex model formulations and unclear relationships between different perspectives, leading to suboptimal parameterization, training objectives, and ad hoc adjustments to counteract these issues. In this talk, I will introduce masked diffusion models, a simple and general framework that unlock the full potential of diffusion models for discrete data. We show that the continuous-time variational objective of such models is a simple weighted integral of cross-entropy losses. Our framework also enables training generalized masked diffusion models with state-dependent masking schedules. When evaluated by perplexity, our models trained on OpenWebText surpass prior diffusion language models at GPT-2 scale and demonstrate superior performance on 4 out of 5 zero-shot language modeling tasks. Furthermore, our models vastly outperform previous discrete diffusion models on pixel-level image modeling, achieving 2.75 (CIFAR-10) and 3.40 (ImageNet 64×64) bits per dimension that are better than autoregressive models of similar sizes.

Author: SHI, Jiaxin (Google DeepMind)

Orateur: SHI, Jiaxin (Google DeepMind)

Classification de Session: ML

Classification de thématique: Machine learning