The DEformer: An Order-Agnostic Distribution Estimating Transformer

PDF **(/pdf?id=H1qfnDmta6)**

Michael A. Alcorn (/profile?id=~Michael_A._Alcorn1), Anh Nguyen (/profile?id=~Anh_Nguyen1)

Published: 15 Jun 2021, Last Modified: 21 Oct 2023 INNF+ 2021 poster Readers: Everyone Show Bibtex Show Revisions (/revisions?id=H1qfnDmta6)

Revision

Abstract: Order-agnostic autoregressive distribution (density) estimation (OADE), i.e., autoregressive distribution estimation where the features can occur in an arbitrary order, is a challenging problem in generative machine learning. Prior work on OADE has encoded feature identity by assigning each feature to a distinct fixed position in an input vector. As a result, architectures built for these inputs must strategically mask either the input or model weights to learn the various conditional distributions necessary for inferring the full joint distribution of the dataset in an order-agnostic way. In this paper, we propose an alternative approach for encoding feature identities, where each feature's identity is included alongside its value in the input. This feature identity encoding strategy allows neural architectures designed for sequential data to be applied to the OADE task without modification. As a proof of concept, we show that a Transformer trained on this input (which we refer to as "the DEformer", i.e., the distribution estimating Transformer) can effectively model binarized-MNIST, approaching the performance of fixed-order autoregressive distribution estimating algorithms while still being entirely order-agnostic. Additionally, we find that the DEformer surpasses the performance of recent flow-based architectures when modeling a tabular dataset.

Community Implementations: [x] 1 code implementation (https://www.catalyzex.com/paper/arxiv:2106.06989/code)

Add **Withdraw**

Reply Type: all Author: everybody Visible To: all readers **3 Replies**

Hidden From: nobody

[–] Paper Decision

ICML 2021 Workshop INNF Program Chairs

14 Jun 2021 ICML 2021 Workshop INNF Paper4 Decision Readers: Everyone Show Revisions (/revisions?id=lVyk-70FFt)

Decision: Accept (poster)

[–] Official Review of Paper4 by Reviewer LFQV

ICML 2021 Workshop INNF Paper4 Reviewer LFQV

11 Jun 2021 ICML 2021 Workshop INNF Paper4 Official Review Readers: Everyone Show Revisions (/revisions?id=Er-cRJmmCs)

Summary:
The paper proposes a transformer-based density model that can represent arbitrary factorization of a joint distribution of data. More specifically, the paper aims at modeling autoregressive models for arbitrary index orderings, but with a single transformer. For a single factorization, the transformer conditions on the order of indices of random variables; for image data, the author uses pixel's xy-coordinate to represent the index. The author emphasizes that the proposed conditioning is computationally more efficient than previous MADE-based models.

During training, the transformer learns arbitrary factorizations by randomizing the orderings; however, one may note that the densities for different index orders won't be consistent with each other. For the test, the learned model computes the likelihood by conditioning on an ordering depending on tasks. As an example, for a given inference task, one can choose an order that is the best fit for the task. Moreover, the model can average an ensemble of the likelihoods of different orderings.

In the experiments, the paper demonstrates that the ensemble estimate of the proposed method achieves as good likelihood as strong autoregressive model baselines. Furthermore, the author shows a potential usage of the proposed method for out-of-distribution detection.

Rating: Borderline Accept

Justification For Rating:
Overall I like the general direction of the paper. I found that experiments can be further improved so that the paper can clearly motivate the benefits of learning arbitrary orderings.

Confidence: 4: The reviewer is confident but not absolutely certain that the evaluation is correct

[−] **Official Review of Paper4 by Reviewer 9p4k**

ICML 2021 Workshop INNF Paper4 Reviewer 9p4k

11 Jun 2021 ICML 2021 Workshop INNF Paper4 Official Review Readers: Everyone Show Revisions (/revisions?id=W6HMTmfqYCM)

Summary:
This paper explores autoregressive transformer models for estimation of arbitrary marginal distributions. The model consists of a general transformer architecture that takes pixels and their identifiers (row and column indices in their experiment), applies masking of random ordering, and outputs conditional distributions over the output.

Rating: Borderline Accept

Justification For Rating:
The paper is well written, and the topic is a good fit for this workshop.

Calling this "order agnostic" is a bit confusing. The authors write "we shuffled the order of the pixels in each training image to encourage the DEformer to learn a joint distribution of the dataset that is approximately permutation invariant with respect to the ordering of the pixels." While it is true that the model will hopefully learn a density that is approximately invariant to the order in which the pixels are presented, it will definitely not be invariant to permutations of pixels that change their spatial location, since this location is used as input to the network. This is in contrast to the NADE and MADE models that the authors compare against: these models are actually permutation invariant and do not use the spatial structure of the image at all.

The novelty of the presented method is somewhat limited, as models of this type have previously been explored: e.g. in XLNet, which the authors do cite. The experimental validation is also somewhat preliminary (binarized MNIST).

Confidence: 4: The reviewer is confident but not absolutely certain that the evaluation is correct