View Reviews

Paper ID 3254

Paper Title Variational Inference with Non-invertible Flow

Reviewer #1

Questions

1. Contributions: Please list three things this paper contributes (e.g., theoretical, methodological, algorithmic, empirical contributions; bridging fields; or providing an important critical analysis). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions.

1- This Paper introduces Non-invertible flow method which extends normalizing flow (NF) technique using non-invertible transformation to infer more accurate posteriori. Significance: it is a good idea depending on its computation complexity

2- the paper suggested a non-invertible transformation. The pdf of new random variable obtained from transformation can be calculated analytically based on equation 7 of the paper. Significance: good result but not very practical

3- To solve the computational complexity issue, authors proposed an auxiliary distribution and formulation to Monte-carlo sample from new bound. Significance: the method showed certain capability in inference in experiments section

2. Detailed comments: Please provide a thorough review of the submission, including its originality, quality, clarity, and significance. Hover over the "?" next to this prompt to see a brief description of these metrics.

This Paper introduces Non-invertible flow method which extends normalizing flow (NF) technique using non-invertible transformation to infer more accurate posteriori. The paper identifies issues for non-mean field techniques offered so far and use the proposed approach to address those problems. The problem is well motivated by trying to find a transformation which is computationally efficient in high dimension (unlike normalizing flow in which the posteriori computational complexity can be quadratic to number of dimensions). Also the proposed transformation is not restrictive and slow to generate samples from such as Inverse Autoregressive Flow (IAF) or some other method which needs many Monte-Carlo samples to approximate the posteriori.

To address above problem the paper suggested a non-invertible transformation. The pdf of new random variable obtained from transformation can be calculated analyticity based on equation 7 of the paper. This approach will need to find root and local inverse of transformation which are computationally non-trivial. To solve that issue, authors proposed an auxiliary distribution (in this case a normal distribution) and formulation to Monte-carlo sample from new bound. Authors shows how after adding new auxiliary distribution, the ELBO changes and made a good discussion about loss function and its components' meaning.

In experimental results paper has used different set of problems to address. Namely VAE and BNN regression. For the VAE, MNIST and Fashion-Mnist has been used and for BNN regression four different dataset have been used and compared the existing method performance against the baselines. Paper have showed good improvement, specifically on MNIST data.

Quality:
Motivation, claims and and supporting material in main paper and supplementary material are clearly explained and I could not find any significant technical issues with the details of claims made in this paper; although, there are some derivation details could be explained better that will be discussed in clarity part of this review.

In the paper, it is claimed that many other distribution can be used as the auxiliary distribution beside normal distribution , but no example of them given neither on synthetic nor on real data experiments. That would be interesting to see how this new proposed model behave given data are generated from latent variables which have multi-modal and/or overdispresed distributions. Also it is interesting to discuss under what condition we can use multiple non-invertible transformation (like NF) and how expensive they are computationally since it will be auxiliary distribution inference at each transformation step.

Clarity:
I think the paper objectives and explanation are pretty clear and flow of material is pretty smooth up to Experiments section. Here are a few clarification points that can make the paper easier to read and comprehend.
- it makes the paper easier to read if derivation for equation 6 and 7 in the paper be added to appendix

-Should K be used in equation 3 instead of M?

- In Experiments section, many details of experiments set ups have been eliminated these details needs to be discussed specially on how auxiliary distribution parameters have been set up and learned, also it will be great to explain the competing algorithms parameters set ups.

Originality:
As mentioned in summary the main contribution of this paper could be summarized as bellow
- How to use the non-invertible transformation instead of invertible transformation to address the main issues of Normalizing flow and IAF

- Then in order to avoid finding roots of the non-invertible function , paper proposed to use Gaussian approximation of delta function and an auxiliary distribution to infer the posteriori.

Authors had done a detailed literature review and most of related works have been mentioned and contribution of this paper have been compared and highlighted clearly.

Significance:
The method is new and original. Main shortcoming I would see with in this paper is the extent of the experiments and clarity of performed experiments.

-As suggested in quality section of this review, it is desired to see how this new proposed transformation will infer the posteriori if data are generated from latent variables with multi-modal and/or overdispresed distributions.

-it should be clear under which condition the comparision on VAE experiments being made between algorithms (e.g. set ups for the SIVI-VAE number of samples and network architecture)

-It is interesting to see what would be advantage and limitation of using other auxiliary variables and a discussion about the outcome on certain example

3. Please provide an "overall score" for this submission.
6: Marginally above the acceptance threshold. I tend to vote for accepting this submission, but rejecting it would not be that bad.

4. Please provide a "confidence score" for your assessment of this submission.
4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.

5. Improvements: What would the authors have to do for you to increase your score?
It is explained in significance part of the review

Reviewer #2

Questions

1. Contributions: Please list three things this paper contributes (e.g., theoretical, methodological, algorithmic, empirical contributions; bridging fields; or providing an important critical analysis). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions.

The main contribution of the paper is an algorithm to use non invertible flows for variational inference. Up to my knowledge, flows so far have always been based on invertible transformations with a tractable jacobian, and this paper explores the possibility of using a non invertible transformation.

The algorithm is based on a known result regarding the pdf of a transformed random variable after a non-invertible transformation. However, this result is not applicable right away, since the resulting pdf after the non invertible transformation is hard to compute (requires finding the roots of a complex system). In order to deal with this the authors propose two things: the use of a (known) bound with auxiliary variables, together with a new approximation for the intractable pdf, which consists on replacing a delta function with a sharp (low variance) Gaussian. The paper combines known results/algorithms with a new approximation in a novel way.

2. Detailed comments: Please provide a thorough review of the submission, including its originality, quality, clarity, and significance. Hover over the "?" next to this prompt to see a brief description of these metrics.

Summary and originality: The performance of VI depends on the flexibility of the variational distribution proposed. Flows are used to transform random variables and create more flexible distributions. These typically use invertible transformations with tractable jacobians. This work proposes the use of non-invertible transformations. A closed for expression for the resulting pdf exists in some cases (required to compute the ELBO). However, it depends on the roots of a complex non-linear system. To avoid this, the authors propose an approximation for the intractable pdf and the use of an auxiliary bound for the ELBO. Up to my knowledge, this is the first work that proposes the use of non-invertible transformations this way.

Clarity: Most of the paper is clearly written. The one part that I think requires a more detailed description is the experimental section. In the "Variaitonal Autoencoder" part (line 223), it is not clear to me if the results for the other models (Yin and Zhou 2018, Titsias and Ruiz 2019) were obtained by running the models or if they were extracted from the papers. It is also not clear to me if the same architecture (decoder network, dimension of the latent space) was used for all models or not. Also, in order for a reader to be able to reproduce the results, the training parameters should be included: learning rate, minibatches, number of epochs, etc. Many of these things could be added in the supplementary material of the paper.

Quality and significance: The work is technically sound. While the algorithm proposed is new and original, I feel that some aspects of the results are not completely compelling.
1- In the "imputation with VAE" part of the results (line 258), the method is compared only against mean field variational inference. Looking at the figure, the proposed method seems to work better. However, a mean field approximation with a Gaussian distribution is one of the simplest possible variational distributions one could use, and it is known to have issues when latent variables are correlated. I think that comparing against other flow-based approaches in this experiment should be done.
2- In the "BNN regression" part of the results (line 273) it can be observed that the proposed method does not perform considerably better than mean field VI (the intervals [mean +- std] for each method have a considerable overlap). Again, in this part it would be interesting to see how the method compares agains other widely used flow-based approaches (ie, Normalizing flows, IAF, Real-NVP).

Despite the fact that results could be compared against more related methods, I think that the use of non-invertible transformations is an interesting idea that is worth exploring.

A few typos and a missing reference:
- lines 22-23: "A popular approach is to address the problem of the integral is to approximate..." The middle part of that sentence should be removed.
- Equation 1: Wrong sign, the right hand side of that equation should be multiplied by -1.
- line 71: The ELBO is maximized, not minimized.
- The following paper should be cited: "A family of non-parametric density estimation algorithms", by Tabak and Turner. This introduced the idea of flows, before the "Variational Inference with Normalizing Flows" paper.

3. Please provide an "overall score" for this submission.
6: Marginally above the acceptance threshold. I tend to vote for accepting this submission, but rejecting it would not be that bad.

4. Please provide a "confidence score" for your assessment of this submission.
3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

5. Improvements: What would the authors have to do for you to increase your score?
Since the paper presents a new flow-based approach to build flexible variational distributions, I think a comparison against other flow-based methods should be included. For example, Normalizing flows, Inverse Autorregressive flows, and/or Real NVP.

Reviewer #3

Questions

1. Contributions: Please list three things this paper contributes (e.g., theoretical, methodological, algorithmic, empirical contributions; bridging fields; or providing an important critical analysis). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions.
This paper proposes an extension to normalizing flow that allows for non-invertible transformations on simple distributions to construct complex distributions.
The authors use Dirac delta function to represent the conditional PDF of the target variable and propose to approximate it with a Gaussian distribution.
The authors introduced an auxiliary distribution accompanied with a DNN, and derived a variational lower bound for inferring the parameters of DNN.

2. Detailed comments: Please provide a thorough review of the submission, including its originality, quality, clarity, and significance. Hover over the "?" next to this prompt to see a brief description of these metrics.
In this paper, the authors extend the normalizing flow to the realm of non-invertible transformations. The approximation to the conditional distribution of target variable and the auxiliary distribution enable easy sampling target variable and variational inference. The paper is well-written and readable.

My main concerns are
1. In Eq.9, the Gaussian distribution approximates the delta function with a small alpha. Was alpha optimized in the experiments since the authors used the word "initialize"? If so, how was it optimized? If not, how was it chosen?
The work would be complete if the authors can discuss the effect of the magnitude of alpha theoretically or empirically.
2. The similar question to the beta in \tilde{q}(z|x).
3. What is the size of the Monte Carlo samples mentioned in Line 157. The authors claim their approach does not require a large sample size. Please clarify if it indicates the data or Monte Carlo. It would be more convincing if the authors provide quantitative comparisons regarding to this claim.
4. The work would be more complete if the authors show how fast the learning is.

3. Please provide an "overall score" for this submission.

6: Marginally above the acceptance threshold. I tend to vote for accepting this submission, but rejecting it would not be that bad.

4. Please provide a "confidence score" for your assessment of this submission.

4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.

5. Improvements: What would the authors have to do for you to increase your score?

I would increase the score if the authors answer my concerns in the comments.