Exercise 2

Hiep Tat Bui (22C01007), Thanh Hoai Le (22C01018)

January 2023

Number of Pages: 5

This exercise includes showing 3 articles that have been published at The 35th Conference on Neural Information Processing Systems (NeurIPS 2021). In this exercise, the writer will try to review those papers quickly, from the author to what makes a difference from the research. Since the writer is not holding ResearcherID or any ORCID of the authors. Therefore, checking via Web of Science to find any personnel researcher information (by name or documents) is not possible. The first article shall start with a group of Vietnamese researchers.

1 Structured Dropout Variational Inference for Bayesian Neural Networks

1.1 Overview the article

The full version can be found at https://doi.org/10.48550/arXiv.2102.07927

Authors:

- First and Corresponding author: Son Nguyen (v.sonnv27@vinai.io);
- Co-author: Duong Nguyen, Khai Nguyen, Khoat Than, Hung Bui, Nhat Ho.

First, we can see this article is contributed by a collaboration between Vietnamese researchers. This article is built on the idea of Son Nguyen (since he is placed in the first position in the list of authors), he is also the corresponding author of this paperwork.

Second, the author Hung Bui and Nhat Ho contributed equally and the affiliation mentioned 3 addresses (2 of them is located in Vietnam, the other from the US).

Third, checking with Google Scholar. This article has been cited by 2 other papers. Son Nguyen have totals 26 times cited and H-index = 3.

1.2 Aiming target

Previous problem: Bayesian Dropout method has simple structures of the mean-field family, its approximation often fails to obtain satisfactory uncertainty estimates. The Variational Dropout based on multiplicative Gaussian noise is also worse when its has improper prior, which leads to making the variational objective undefined.

They proposed a new method called Variational Structured Dropout (VSD) and summarize advantages as follows:

- Maintaining the computational efficiency on low dimensional space;
- Since VSD has a standard Bayesian justification, its avoids critiques from non-Bayesian perspective of the origin Variational Dropout method;
- VSD is more efficient than the previous Bayesian Dropout method;
- VSD converges the local minimal with a smaller spectral norm and stable rank.

New keyword

Variational Structured Dropout (VSD).

1.3 Experiment result

Image classification

Using 3 standard image datasets: MNIST, CIFAR10, SVHN. VSD outperforms almost other methods while its competitive with D.E and SWAG methods.

Integrate VSD into large-scale Convolutional networks

Applied VSD into pre-trained such as AlexNet and ResNet18 then trained with 4 datasets: CIFAR10, SVHN, CIFAR100, STIL10. Overall, the VSD method maintains a good performance with better stability.

Predictive entropy performance

The authors tried to evaluate the predictive uncertainty to see if a model evaluated with an out-of-distribution would it be difference?

To answer it, they train each model with SVHN dataset and then use CIFAR10 and CIFAR100 to test. As it shown in Figure 3, all methods working well with SVHN dataset (especially with MAP and VD are concentrated excessively around 0 more than others given a temporary conclusion that it make overconfident prediction on out-of-distribution data, which is true at the next figure with CIFAR10 and CIFAR100). 3 over 5 models is MCD, BBB, and VSD are well-calibrated, but VSD gains better results overall.

2 TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification

2.1 Overview the article

The full version can be found at https://doi.org/10.48550/arXiv.2106.00908

Authors:

- First author: Zhuchen Shao, Hao Bian, Yang Chen;
- Corresponding author: Yongbing Zhang (ybzhang08@hit.edu.cn);
- Co-author: Yifeng Wang, Jian Zhang, Xiangyang Ji.

Similar to the first section, this article is contributed by a group of Chinese researchers. This article is 3 first author (not 1), because although Zhuchen Shao is staying in first place his work contributed equally with Hao Bian and Yang Chen. For some special cases, some articles may officially write in their own paper about the number of first author. But in this paper, the writer still confirms that there is only 1 first author, who is Zhuchen Shao and he equally shares the same work with others only.

Next, the affiliation mentioned 4 addresses (2 of them is located in the same institution - Tsinghua University). The writer temporarily concludes that this paper is highly good. Because none of those institutes in the affiliation is in low rank. Rechecking by ARWU- Universities Ranking (since all those universities is from China, checking THE Ranking or QS Ranking may not be given an accurate ranking). Tsinghua University (Top 26th Universities Ranking); Peking University (Top 34th Universities Ranking); Harbin Institute of Technology (Top 151-200 Universities Ranking).

Checking with Google Scholar. This article has been cited 112 times.

The coding is available at: https://github.com/szc19990412/TransMIL

2.2 Aiming target

Previous problem: As the article mentioned, Multiple instance learning (MIL) is being applied to using for digital pathology, which use to deal with the huge size and the lack of pixel-level annotations. However, MIL methods are under the assumption as all the instances are independent and identically distributed (i.i.d). But too good to be true, in reality, a pathologist needs to put those instances under the right context or consider the correlation information between different areas when making diagnostic decisions. Then, it might have some correlation between different instances in MIL.

They proposed a new framework called correlated MIL, which provided proof of convergence. Then, the authors devised a Transformer based on MIL, called TransMIL and summarize its advantages as follows:

- Maintaining the computational efficiency on unbalanced/balanced and binary/multiple classifications;
- Improving the performance, faster convergence than the state-of-the-art deep MIL models.

New keyword

correlated MIL, TransMIL.

2.3 Experiment result

- CAMELYON16, a public dataset for metastasis detection in breast cancer;
- TCGA-NSCLC, a dataset including Lung Squamous Cell Carcinoma (TGCA-LUSC) and Lung Adenocarcinoma (TCGA-LUAD);
- TGCA-RCC, a dataset with a total of 884 diagnostic including Kidney Chromophobe Renal Cell Carcinoma (TGCA-KICH), Kidney Renal Clear Cell Carcinoma (TCGA-KIRC) and Kidney Renal Papillary Cell Carcinoma (TCGA-KIRP).

The test AUC for the binary tumor classification can be up to 93.09% over CAMELYON16 dataset. And the AUC over the cancer subtypes classification can be up to 96.03% and 98.82% over TCGA-NSCLC dataset and TCGA-RCC dataset.

3 Metropolis-Hastings Data Augmentation for Graph Neural Networks

3.1 Overview the article

The full version can be found at https://doi.org/10.48550/arXiv.2203.14082

Authors:

- First author: Hyeonjin Park, Seunghun Lee;
- Corresponding author: Hyunwoo J. Kim;
- Co-author: Sihyeon Kim, Jinyoung Park, Jisu Jeong, Kyung-Min Kim, , Jung-Woo Ha.

The paper print denoted that the first two authors have equal contributions. Therefore, there are 2 first authors.

Checking with Google Scholar. This article has been cited 13 times by others researched.

3.2 Aiming target

Graph Neural Networks (GNNs) have been widely used for representation learning on graph-structured data, it has impressive performance for diverse datasets such as social networks, physics, etc. However, because of the non-Euclidean nature of data space and the dependencies between samples, applying an effective augmentation on graphs is challenging. The authors proposed a new framework called Metropolis-Hastings Data Augmentation (MH-Aug), which MH algorithm is a Markov chain Monte Carlo method to draw random samples from a target distribution when direct sampling is difficult. Therefore, MH-Aug will produce a sequence of the augmented graph. Then improving the performance of GNNs. Their mainly work in this paper is:

- Proposed Metropolis-Hastings Data Augmentation (MH-Aug);
- Prove MH-Aug works in theoretically or empirically;
- Propose a target distribution that flexibly controls the strength and diversity of augmentation;
- Introducing a simple and effective semi-supervised learning strategy leveraging sequentially generated samples from the method.

New keyword

Metropolis-Hastings Data Augmentation (MH-Aug).

3.3 Experiment result

- Citation networks with 2 datasets: CORA and CITESEER;
- Amazon product networks with 2 datasets: Computers (Compu.) and Photo;
- Coauthor Networks with CS dataset.

The framework MH-Aug, which is trained in the semi-supervised setting, consistently achieves the best performance in all datasets and the improvement against the vanilla models (without augmentation) is 3.16% on average.