

[← Back to **Author Console** \(/group?id=ICML.cc/2023/Conference/Authors#your-submissions\)](/group?id=ICML.cc/2023/Conference/Authors#your-submissions)



Domain Generalization with Small Data



 (/pdf?)

id=HIYL2Zn35W)

Kecheng Chen (/profile?id=~Kecheng_Chen1), Elena Gal (/profile?id=~Elena_Gal1), Hong Yan (/profile?id=~Hong_Yan2), Haoliang Li (/profile?id=~Haoliang_Li2) 👁

 26 Jan 2023, 13:19 (modified: 11 Feb 2023, 06:45)
  ICML 2023 Conference Submission
  Conference, Area Chairs, Authors, Reviewers, Senior Area Chairs
  Revisions (/revisions?id=HfYL2Zn35W)

Keywords: domain generalization, healthcare, small data, medical imaging

TLDR: A novel domain generalization method in the context of insufficient data is proposed in this work

Abstract:

In this work, we propose to tackle the problem of domain generalization in the context of insufficient samples. Instead of extracting latent feature embeddings based on deterministic models, we propose to learn a domain-invariant representation based on the probabilistic framework by mapping each data point into probabilistic embeddings. Specifically, we first extend empirical maximum mean discrepancy (MMD) to a novel probabilistic MMD that can measure the discrepancy between mixture distributions (i.e., source domains) consisted of a serial of latent distributions rather than latent points. Moreover, instead of imposing the contrastive semantic alignment (CSA) loss based on pairs of latent points, a novel probabilistic CSA loss encourages positive probabilistic embedding pairs to be closer while pulling other negative ones apart. Benefiting from the learned representation captured by probabilistic models, our proposed method can marriage the measurement on the distribution over distributions (i.e., the global perspective alignment) and the distribution-based contrastive semantic alignment (i.e., the local perspective alignment). Extensive experimental results on three challenging medical datasets show the effectiveness of our proposed method in the context of insufficient data compared with state-of-the-art baseline methods.

Paper Checklist Guidelines: I certify that all co-authors of this work have read and commit to adhering to the Paper Checklist Guidelines, Call for Papers and Publication Ethics.

Submission Number: 3386

Filter by reply type...

Filter by author...

Search keywords...

Sort: Newest First

-

=

≡

Everyone

Program Chairs

Area Chairs

Authors

Reviewer gTaf

Submitted

3 / 3 replies shown

Senior Area Chairs

Reviewer s5G1

Reviewer 14zj

Add: **Withdrawal** **Rebuttal**

Official Review by Reviewer gTaf

Official Review  Reviewer qTaf  08 Mar 2023, 11:15 (modified: 14 Mar 2023, 21:41)

Program Chairs, Area Chairs, Authors, Reviewer qTaf, Submitted, Senior Area Chairs

Summary:

The paper proposes a probabilistic method of domain generalization (DG) with small data with three medical datasets as examples. The method repurposes maximum mean discrepancy (MMD) and contrastive semantic alignment (CSA) in a probabilistic way for global and local perspectives, respectively. The empirical results show that the proposed method outperforms various other known DG methods.

Strengths And Weaknesses:

The paper proposes a novel and effective method to handle domain generalization. The empirical results look good. The writing of the theory was a bit hard to follow. Figure 2 is great; it helps a reader to understand the concept better. Overall, the paper looks fine though. The proposed method seems to be useful for medical image related problems given how often the medical datasets are from various different domains in small chunks.

Questions:

See the limitations section below.

Limitations:

What are some limitations of using this probabilistic MMD and CSA vs the original version? What are some tradeoffs between the proposed method vs other methods compared? When would one prefer to use another method instead of the proposed method?

Ethics Flag: No

Soundness: 3 good

Presentation: 3 good

Contribution: 3 good

Rating: 6: Weak Accept: Technically solid, moderate-to-high impact paper, with no major concerns with respect to evaluation, resources, reproducibility, ethical considerations.

Confidence: 2: You are willing to defend your assessment, but it is quite likely that you did not understand the central parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

Code Of Conduct: Yes

Add: **Rebuttal**

Official Review by Reviewer s5G1

Official Review  Reviewer s5G1  07 Mar 2023, 00:13 (modified: 14 Mar 2023, 21:41)

 Program Chairs, Area Chairs, Authors, Reviewer s5G1, Submitted, Senior Area Chairs

Summary:

The paper proposes a method for domain generalization employing probabilistic MMD over the features samples from a Bayesian network using a probabilistic contrastive semantic alignment. The main contributions are extending MMD with deterministically extracted features to a the bayesian network (which uses $p(z|x,W)$) using RKHS. To enable non-linear learning on distributions they further propose using a second kernel. They present results on medical image datasets, where the domain shifts occurs while using data from different hospitals. The proposed method gives consistent gains over the compared images. The paper also presents results on PACS, Officehome and VLCS datasets and achieve comparable performance to the SOTA methods.

Strengths And Weaknesses:

Strengths

- The paper motivates use Bayesian networks and empirical observations seem to support the case
- The observation of using P-MMD seems to preserve high level statistics, when compared to MMD over the first moment appears interesting.
- The results on the medical datasets suggest the efficacy of the proposed approach

Weakness

- An ERM baseline is warranted, given the recent findings in the field (Gulrajani et al. In search of lost domain generalization). DomainBed ensures standardisation across experiments, a major issue with DG experiments.

- MIRO + SWAD comparisons are missing and would be worth an addition to the paper (Cha et al. Domain Generalization by Mutual-Information Regularization with Pre-trained Models)
- The paper does not talk about the source of the dataset, is it collected by authors or taken from public forums. If it is collected on their own and some more experiments on public datasets are necessary. For example, using WILDS dataset (e.g. Camelyon17 is for Tissue slides across hospital, <https://wilds.stanford.edu/datasets/> (<https://wilds.stanford.edu/datasets/>)). Not sure if the authors were unaware of this dataset or they choose not to perform experiments on it.
- The method seems to work well on small dataset. Would it work on larger datasets as well? Results on Domain-net would be useful.

Questions:

Please address the concerns raised in the weaknesses section. Specially on not doing experiments on Camelyon17 and lacking any observations when the method is utilized on larger datasets like Domain-net. Standardised experimental protocol is almost of utmost importance, some clarifying comments on that would be useful. Also, I would suggest the authors to put up an ERM baseline. My final rating will heavily depend on the rebuttal.

Limitations:

I did not find any discussion on the limitations of the proposed approach and it is concerning from my personal perspective. DG is a tricky area and entire community has been beating around the bush for a long time (rarely algorithms perform better to ERM). Any new tailored approach needs transparency and adding limitation section would help in that.

Ethics Flag: No

Soundness: 3 good

Presentation: 3 good

Contribution: 3 good

Rating: 5: Borderline accept: Technically solid paper where reasons to accept outweigh reasons to reject, e.g., limited evaluation. Please use sparingly.

Confidence: 5: You are absolutely certain about your assessment. You are very familiar with the related work and checked the math/other details carefully.

Code Of Conduct: Yes

Add: **Rebuttal**

Official Review by Reviewer 14zj

Official Review  Reviewer 14zj  05 Mar 2023, 20:14 (modified: 14 Mar 2023, 21:41)

 Program Chairs, Area Chairs, Authors, Reviewer 14zj, Submitted, Senior Area Chairs

Summary:

The paper tackles the domain generalisation problem. The learner is given sets of samples from diverse domains to train on and is required to generalise to a novel domain at test time. A typical approach is to learn domain-invariant features - for example via distribution matching (e.g. MMD). This paper is interested in the scenario when there are not so many samples per domain. To address the lack of training data, the paper proposes to use a Bayesian neural network with a distribution over the network parameters to obtain probabilistic feature embeddings for each sample. The paper then applies a distribution matching objective between two distributions (samples from each domain) of distributions (probabilistic embeddings). On top of the distribution matching, the paper applies contrastive objective with the class labels (probabilistic contrastive semantic alignment). The paper shows empirical results on three datasets from the healthcare domain.

Strengths And Weaknesses:**Strengths**

Technically, the concept of meta-distribution is quite interesting. I'd be interested in its applications beyond the domain generalisation setup.

The paper writing is generally okay.

Weaknesses

The paper does not make a convincing case for why probabilistic embeddings are really necessary. The paper argues, for example, that "the probabilistic neural network is imposed to enable high-quality and powerful feature representation in the context of insufficient samples" (l.085 right). I'm not sure if it's generally understood that BNNs improve generalisation for few-shot learning scenarios. To me, this is not very obvious. I do agree that BNNs may help identify the samples with higher epistemic uncertainty (samples without so many neighbouring training data points), but not so sure about BNNs actually improving the generalisation performance. It would be great if the authors could provide a few prior works discussing this matter. Please point me to the part of the submission, if I have happened to miss this discussion.

The experiments section also doesn't convincingly show how turning the embeddings into distributions help the model cope with the lack of training data. For example, one could show that the overall variance for the domains with less training samples is generally lower. Or, in the ablative studies with changing fractions of training data (l.396 right), one could also show the change in sample variance for the probabilistic embeddings as the number of training samples decreases. The experimental results are a bit too focused on numbers and do not show much insight.

The proposed method is not very practical. It is both conceptually complicated and computationally expensive. When a method is conceptually complicated, it is difficult for practitioners to know what to do to apply it to their own problems. Example:

- which sigma value for the RBF kernels should I use?
- which number of MC samples would be needed?
- how to set the loss weights in Equation 9?

Likewise, the method seems computationally expensive, as it relies on MC sampling during training and inference.

There's a branch of research on probabilistic embeddings -- see <https://arxiv.org/abs/2202.06768> (<https://arxiv.org/abs/2202.06768>) for an overview. Not saying one should read and cite this arXiv paper, just saying that the references in this arXiv paper are at least remotely relevant to the submission, as they also introduce distributions over the feature space for each input sample. The key difference is that the source of randomness for the probabilistic embedding for the current submission is the distribution over the parameters (BNN), while the mentioned branch of research is more interested in the prediction of distributional parameters (e.g. Gaussian mean and cov) for the probabilistic embedding. Perhaps the authors can find a good way to argue that the randomness coming from BNN is more suitable in their case as it copes better with the lack of training data. (Not too sure about this though) Or perhaps the authors find this line of work interesting and seek ways to incorporate those probabilistic embeddings into their future work.

Minor

- Title: "Small" data means each data sample is small in size - e.g. low resolution image. One should say "few" data instead, for example.
- The first part of the introduction (up to l.090 left) is too verbose and should be shortened to around 50-70% of the current length.
- l.135 left: "is" --> "be"
- Equation (5): brackets not balanced
- l.178 left: "6" --> "Equation 6"
- l.278 left: remove "denotes"

Questions:

Please address the weaknesses. It's not very clear to me yet how the probabilistic embeddings help the model overcome the lack of training data.

Limitations:

Not applicable.

Ethics Flag: No

Soundness: 2 fair

Presentation: 3 good

Contribution: 2 fair

Rating: 4: Borderline reject: Technically solid paper where reasons to reject, e.g., limited evaluation, outweigh reasons to accept, e.g., good evaluation. Please use sparingly.

Confidence: 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.

Code Of Conduct: Yes

Add: **Rebuttal**

[About OpenReview \(/about\)](/about)

[Hosting a Venue \(/group?id=OpenReview.net/Support\)](/group?id=OpenReview.net/Support)

[All Venues \(/venues\)](/venues)

[Sponsors \(/sponsors\)](/sponsors)

[Frequently Asked Questions](#)

[\(https://docs.openreview.net/getting-started/frequently-asked-questions\)](https://docs.openreview.net/getting-started/frequently-asked-questions)

[Contact \(/contact\)](/contact)

[Feedback](#)

[Terms of Service \(/legal/terms\)](/legal/terms)

[Privacy Policy \(/legal/privacy\)](/legal/privacy)

[OpenReview \(/about\)](/about) is a long-term project to advance science through improved peer review, with legal nonprofit status through [Code for Science & Society \(https://codeforscience.org/\)](https://codeforscience.org/). We gratefully acknowledge the support of the [OpenReview Sponsors \(/sponsors\)](/sponsors).