

View Reviews

Paper ID

3002

Paper Title

Gradual Domain Adaptation via Self-Training of Auxiliary Models

Reviewer #1

Questions

1. [Summary] In 3-5 sentences, describe the key ideas, experiments, and their significance.

This paper proposes a gradual self-training approach for unsupervised domain adaptation. It is mainly inspired by a recent self-training study and tries to introduce multiple intermediate domains to relate the source domain and the target domain. Experiments further validate its effectiveness on several widely-used UDA benchmarks like Office-Home and VisDA-C.

2. [Strengths] What are the strengths of the paper? Clearly explain why these aspects of the paper are valuable.

1. this motivation behind is easy to understand
2. the UDA results along with the SSDA results in the experiments are impressive

3. [Weaknesses] What are the weaknesses of the paper? Clearly explain why these aspects of the paper are weak. Please make the comments very concrete based on facts (e.g. list relevant citations if you feel the ideas are not novel). If applicable, please indicate key issues and questions which, if well addressed during the 1-page rebuttal, might influence you to change your rating.

1. some components within the proposed method sound complicated and can be dropped to make this paper more attractive. For example, section 4.2 tries to enhance the sample selection by incorporating label propagation and clustering-based pseudo labeling, which sounds over-designed. And the experiments in Table 1 indicate that the label enhancement makes a little contribution to the final results.

By the way, it also sounds like that the source selection component is not vital in the proposed method.

2. In Fig.5(b), the results rely on a large number of the intermediate domains (M). Does it mean the proposed method takes a long time to train? The authors need to provide a time complexity analysis and provide some speedup solutions in this paper.

3. Besides, does the proposed method works on a small-sized dataset like Office?

4. The progressive manner is somewhat similar with [a] on partial domain adaptation, the authors need to add the discussion with it.

[a]. Liang, Jian, et al. "A Balanced and Uncertainty-aware Approach for Partial Domain Adaptation." (ECCV 2020).

4. . [Overall rating] Paper rating (pre-rebuttal)

Borderline

5. [Detailed comments] Additional comments regarding the paper (e.g. typos, any suggestions to make the submission stronger).

Overall, the gradual adaptation idea is interesting in this field since most recent UDA works focus on the domain alignment or directly exploit the pseudo labels for self-training.

However, introducing multiple intermediate domains into the training process sounds time-consuming, if this point could be improved, it will definitely be good work in this field.

Besides, dropping some redundant components like the enhanced component and the source data selection will make this paper more attractive.

6. [Reproducibility] Is the method described in this paper reproducible?

I don't know / can't tell

7. [Confidence] Reviewer's confidence in his/her recommendation

Very confident

10. Please provide an "Overall Rating", following the rebuttal and reviewers discussions.

Leaning to reject. I tend to vote for rejecting this submission, but accepting it would not be that bad.

11. Justification of final rating. Describe the rationale for your final rating, including notes based on the rebuttal, discussion, and other reviews.

After reading reviews from other reviewers and the rebuttal, I am learning to reject this paper. An important point raised by R3 (similarity with a TPAMI paper) convinces me that the novelty is somewhat limited. Besides, as written in the initial review, the results in Table 1 do verify that some parts are not vital within the proposed complex method. A simplified one would be much better. The rebuttal does not also convince me that the over-designed component like label enhancement is important.

Reviewer #2

Questions

1. [Summary] In 3-5 sentences, describe the key ideas, experiments, and their significance.

The paper proposes to deal with the domain adaptation problem with a curriculum training strategy, by bridging the gap between source and target domain gradually. The idea is to train a network on intermediary domains by creating hybrid datasets with images from both domains, where the target domain gradually becomes the majority. Furthermore, samples are not chosen randomly but with their proximity to the original domains (source proximity for target samples and vice versa). This approach is novel and yields state-of-the-art results on traditional benchmarks. The relative simplicity of the method and the experiments on real life like situation (such as VisDA) make it significant for the DA field.

2. [Strengths] What are the strengths of the paper? Clearly explain why these aspects of the paper are valuable.

The paper is very well written, easy to read and to understand. All the different figures are concise and efficient.

The experiments are thorough, especially for the control ones (on the rotated MNIST dataset) and the ablation studies.

The base idea is interesting and novel (only a few recent papers published about gradual domain adaptation) while being simple enough for the method to be used in the wild.

3. [Weaknesses] What are the weaknesses of the paper? Clearly explain why these aspects of the paper are weak. Please make the comments very concrete based on facts (e.g. list relevant citations if you feel the ideas are not novel). If applicable, please indicate key issues and questions which, if well addressed during the 1-page rebuttal, might influence you to change your rating.

The motivation section just shows that the larger the gap between the source domain and the target domain, the worse a model trained on the source domain performs. Provided that a model can be adapted if the domain gap is small, it is sufficient to ground the idea of a curriculum training strategy. However, it could be interesting to show that the adapted performance (by using a baseline domain adaptation method) is worse the larger the gap, and not the source performance. This would give credit to the idea that gradually adapting the model is key when the domain gap is large. It could also be interesting to apply the proposed method on the rotated MNIST dataset (which is done) and show the results alongside the source training in figure 1.

An explanation or some clues or ideas as to why there is a correlation between the domain shift and the maximum softmax probability could be interesting, especially when (as stated in the paper) deep networks

sometimes yield very high probability for out-of-distribution data. If a theoretical clue cannot be found, it could be interesting to add more empirical analyses on other synthetic / toy datasets, as MNIST based datasets may behave differently from the others because of their simplicity.

4. . [Overall rating] Paper rating (pre-rebuttal)

Strong Accept

5. [Detailed comments] Additional comments regarding the paper (e.g. typos, any suggestions to make the submission stronger).

Introduction : clear, motivation is explained, contributions are listed.

Related Works : complete.

Motivation : weaknesses are detailed in the Weaknesses section of the review.

Proposed Method : the advanced data augmentation A() used is not described, even briefly, in the paper and lead the reader to a reference directly. The paper could benefit from a single sentence explaining the principle of this data augmentation, and its complexity, as it seems that a correct choice of data augmentation is important to reach peak performance (as explained in the ablation studies).

Experiments : control studies are interesting and not always included, they make sure that the method work as intended.

6. [Reproducibility] Is the method described in this paper reproducible?

The paper includes information that would make it possible to reproduce the methods and experiments

7. [Confidence] Reviewer's confidence in his/her recommendation

Confident

10. Please provide an "Overall Rating", following the rebuttal and reviewers discussions.

Leaning to Accept. I tend to vote for accepting this submission, but rejecting it would not be that bad.

11. Justification of final rating. Describe the rationale for your final rating, including notes based on the rebuttal, discussion, and other reviews.

After reading the other reviews and the authors rebuttal, I am downgrading my grade and only lean to accept this paper. The matter of the correlation between domain shift and maximum prediction probability has been addressed, but the novelty of the paper is lesser than I thought initially and the results are not discussed enough.

Reviewer #3

Questions

1. [Summary] In 3-5 sentences, describe the key ideas, experiments, and their significance.

This paper proposes a domain adaptation approach by gradually adapting the source model to the target domain by self-training of auxiliary models on evolving intermediate domains. The experimental results on four datasets have proved the effectiveness of the proposed method.

2. [Strengths] What are the strengths of the paper? Clearly explain why these aspects of the paper are valuable.

1. The Enhanced Indicator for Sample Selection is a little interesting.
2. The experiments on rotating MNIST dataset are well-designed.

3. [Weaknesses] What are the weaknesses of the paper? Clearly explain why these aspects of the paper are weak. Please make the comments very concrete based on facts (e.g. list relevant citations if you feel the ideas are not novel). If applicable, please indicate key issues and questions which, if well addressed during the 1-page rebuttal, might influence you to change your rating.

1. The novelty is limited. The proposed method is not neat. It seems to put up lots of existing modules together. Interpolating source domain and target domain is not new. Gradually adapting from source domain to target domain is not new. The strategies of selecting source domain samples and target domain samples are also trivial.
2. Is there theoretical analysis or guarantee for the correlation between the domain shift and maximum prediction probability?
3. The experimental results on some categories (Table 1) or some settings (Table 3) are not SOTA. The authors should provide some in-depth analyses for the categories and settings in which the proposed method is not very effective.
4. The effectiveness of enhanced indicator is not fully justified. The authors only compare (9) with vanilla f_m. Since there are in total three components, the combination of each two components should also be studied. There is no analysis for the hyper-parameter λ in (11).

4. . [Overall rating] Paper rating (pre-rebuttal)

Weak Reject

5. [Detailed comments] Additional comments regarding the paper (e.g. typos, any suggestions to make the submission stronger).

1. The paper writing needs to be improved. For example, the sentences in Line 98-103 is very long and hard to understand.
2. The authors should provide some in-depth analyses for the categories and settings in which the proposed method is not very effective.
3. The authors should study the combination of each two components in (9) and the hyper-parameter λ in (11).

6. [Reproducibility] Is the method described in this paper reproducible?

I don't know / can't tell

7. [Confidence] Reviewer's confidence in his/her recommendation

Very confident

10. Please provide an "Overall Rating", following the rebuttal and reviewers discussions.

Reject. I vote and argue for rejecting this submission.

11. Justification of final rating. Describe the rationale for your final rating, including notes based on the rebuttal, discussion, and other reviews.

The authors do not address my concerns well, so I recommend rejecting this paper.

- 1) The authors did not provide theoretical analysis or guarantee for the correlation between domain shift and predictive uncertainty. The correlation between domain shift and predictive uncertainty has been explored as early as in 2009, which also gradually adapts from source domain to target domain based on prediction confidence. The differences are only some technical details and the proposed method is a combination of existing techniques.

[1] Bruzzone, Lorenzo, and Mattia Marconcini. "Domain adaptation problems: A DASVM classification technique and a circular validation strategy." IEEE transactions on pattern analysis and machine intelligence 32.5 (2009): 770-787.

- 2) Lots of important experiments are missing and adding them in the rebuttal is a little post-hoc.
- 3) I agree with Reviewer#5 that this paper is not so well-written and needs to be polished.

Reviewer #5

Questions

1. [Summary] In 3-5 sentences, describe the key ideas, experiments, and their significance.

This is a mostly theoretical paper about Domain Adaptation with some mostly toy in depth experiments and some experiments for SoTa comparisons.

The authors are motivated by analysis that they did on source reliability for domain adaptation(Sec. 3), in which

they have observed that

larger divergence leads to worse results when naively adapting the source domain model to the target domain(Fig. 1).

Along those lines, they propose a self-training of auxiliary models (AuxSelfTrain) that learns models for intermediate domains and gradually

combats the distancing shifts across domains(Fig. 2).

They introduce:

- evolving intermediate domains as combinations of decreasing proportion of source data and increasing proportion of

target data, which are sampled to minimize the domain distance between consecutive domains.

This helps the source model to steadily adapt for the use in the target domain

by self-training of auxiliary models on evolving intermediate domains.

- They also introduce an enhanced indicator for sample selection via implicit ensemble and extend the proposed method to semi-supervised domain adaptation.

- They perform experiments on benchmark datasets of unsupervised and semi-supervised domain adaptation verify its efficacy(Tables 3,4,5).

Overall, their model (AuxSelfTrain) is based on self-training of auxiliary models on auxiliary domains which are created

using a sample selection strategy to select decreasing proportion of source data and increasing proportion of target data to construct evolving intermediate domains under the assumption of minimal domain shift i.e. the intermediate domains have small shifts and are gradually transformed to T(target domain).

They also introduce an enhanced selection indicator by improving the quality of predictive uncertainty to domain shift via implicit ensemble(Sec. 4.2). They average the basic model (f_m) with one classifier based on label propagation[55](\tilde{f}) and a clustering based classifier(\hat{f}).

Their experiment thoroughly show their train of thought on a toy dataset(Rotating MNIST) and compare themselves to previous methods(Tables 3,4,5), briefly on public benchmarks.

2. [Strengths] What are the strengths of the paper? Clearly explain why these aspects of the paper are valuable.

The authors clearly motivate their idea and position themselves compared to previous work in the domain.

Their motivation experiments are clear and clean and their figure(2) is indicative of their idea. The Algorithm 1 flow chart is clear and explains how their method works in a high-level. Figure 3 clearly shows how intermediate domains evolve(SupMat) also does so.

Moreover, their goal and problem setup is clear, intuitive and their approach seems to achieve their goal as the divergence between is minimal between different steps(Figure 4.) Finally, the feature discriminability study indicates again a proof of their hypothesis with a nice ablation(Figure 5.)

Overall, the idea is well-motivated and supported with solid arguments. Not in all of its points, but at least in its high-level form. It is nice to see various ablation studies even in a toy dataset.

Introduction, Related Work, Abstract and Motivation are very well-written and clear. This assists the reader to immediately perceive the main idea of the paper and how this is different from what previous researchers have done. It also beneficial for the reader the thoughts for future work which you are providing in end of the Conclusion.

3. [Weaknesses] What are the weaknesses of the paper? Clearly explain why these aspects of the paper are weak. Please make the comments very concrete based on facts (e.g. list relevant citations if you feel the ideas are not novel). If applicable, please indicate key issues and questions which, if well addressed during the 1-page rebuttal, might influence you to change your rating.

The method of the paper is not so well-written and clear for the reader. I understand that your paper is mostly theoretical but in many point that you write an equation, you omit to give the intuitive explanation of it which is

very important for the reader. L302 you could have written what is the k th entry. L306 you could also intuitively explain the equation which is the maximum probability of x_t that comes from the source model. Moreover, in L414 you do not explain what is the prototype classifier and you do not explain why one needs a target classifier. There are a couple of points that I had to read multiple times. Your equations are clean. However, I would like to see more "pauses" in the text to explain the intuition between the steps and what different symbols indicate in equations semantically. Although, notation-wise you are explaining every quantity you omit to explain it semantically.

Moreover, Sec 4.2 is written in a bit compressed way and I think it needs some more explanations except citing relevant papers on why this exact classifier ensembling is a good choice. The explanation of (11)-somewhat explained in SupMat-, (10) are again lacking a semantic, high-level view.

Sometimes, it might be easy for the reader to infer those, but it is a bit tiring when reading the paper to miss the intuition behind important concepts of the paper.

In Analyses section, L737 Selection-S column does not behave exactly the same as Selection-T. The plane category, for example, has the same performance while bus person, skateboard perform better when your approach is not followed. In Table 1, your method(full) does not outperform the other methods in all categories and this is merely explained. The experiments on the SoTA Tables are not described in detail and you do not mention how the methods you compare to are behaving worse or better. Overall, you have spend a substantial amount of space on analyzing and motivating your idea(some of these could be part of SupMat) leaving only 2 pages for experiments and discussion which do not solve reader's questions about your failures and successes. In conclusion, although you propose an interesting approach, the discussion of the experimental results is not sufficient and analyzed enough to convey the message that you want. You could include a more compressed version of Rotating MNIST experiments and explain more the closer to the "real-world" benchmarks and your results on them(Currently 5 lines in 5.2.2).

Last but not least, it would be great for the community and quality of the paper if you could provide an explanatory code with your experiments and experimental setup in a form of a notebook or code repo. This would vastly help the reader understand your different steps and gain the intuition that may be missing from your manuscript.

I am really on the border right now but after the author's reply and discussion with other reviewers I will provide a clear sign for the AC. I am happy to hear your thoughts and responses to my suggestions.

4. . [Overall rating] Paper rating (pre-rebuttal)

Borderline

5. [Detailed comments] Additional comments regarding the paper (e.g. typos, any suggestions to make the submission stronger).

L 358-359 please rephrase your point is not clear

L 362 typo

L 415 What is a target prototype classifier? Please write this more clear.

L 302 k th entry means k th- class associated entry ?i.e. the probability for the k th class?

6. [Reproducibility] Is the method described in this paper reproducible?

I don't know / can't tell

7. [Confidence] Reviewer's confidence in his/her recommendation

Somewhat confident

10. Please provide an "Overall Rating", following the rebuttal and reviewers discussions.

Leaning to reject. I tend to vote for rejecting this submission, but accepting it would not be that bad.

11. Justification of final rating. Describe the rationale for your final rating, including notes based on the rebuttal, discussion, and other reviews.

First, I would like to thank the authors for their submission. Their paper contains a very interesting idea, backed up with some experimental results which help the reader better understand the value of the idea. The overall approach is interesting evolves intuitively, although some steps towards the practical implementation of it are a bit over complex and described in brevity.

Based on these final thoughts, I was led to my rating.

1. Labelling enhanced is indeed free of cost but really does not give an important improvement based on Table 1. Something that you can get free of cost is not something that you should include absolutely. It increases complexity which is already inherent and repels the reader/user of the paper/method. In that case, I do not think it is essential part and it might include more complexity which could be added in the Sup.Mat to clarify the benefits of your method using the main components.

2. The selection enhancement method indeed improves your results. However, I do not see its vital role. As mentioned in my review:

"In Analyses section, L737 Selection-S column does not behave exactly the same as Selection-T. The plane category, for example, has the same performance while bus person, skateboard perform better when your approach is not followed. In Table 1, your method(full) does not outperform the other methods in all categories and this is merely explained. The experiments on the SoTA Tables are not described in detail and you do not mention how the methods you compare to are behaving worse or better...". In general, the explanation of the results is a bit quick and dirty and lacks intuition. In my opinion, this makes the papers imbalanced and overcomplex in terms of theoretical content which is not absolutely necessary about better performance.

3. The experiments provided in the rebuttal are essential parts of the manuscript and, if not all, most of them should be included in the final submission.

Overall, you spend substantial amount of space on analyzing and motivating your idea(some of these could be part of SupMat) leaving only 2 pages for experiments and discussion which do not solve reader's questions about your failures and successes. In conclusion, although an interesting approach, the discussion of the experimental results is not sufficient and analyzed enough to convey the message that you want.

4. Overall, you do not provide any theoretical(not experimental motivation from a previous work) guarantees for the correlation between domain shift and maximum prediction probability. It is true that you mention it in the rebuttal, promising to investigate it as a next step and that is why I consider this a minor point towards improving your paper.

5. Finally, about novelty, the authors' approach is indeed interesting. However, the comparison, discussion and inclusion of results from [1] is essential and critical for the reader and future user. The authors in [1] have slightly different objective(cross domain class match vs evolving domains), but the authors in this paper do not include their results(which are much better in OfficeHome for example) and not mention this approach. Moreover, in [2] the authors propose a method of evolving domains which is along the path that you are following. Their method[2] is not discussed or mentioned in your paper, although idea-wise it contains similar intuition to yours. I also consider this a minor point, but necessary to add.

The changes and additions would vastly transform the paper. I encourage the authors -in case of rejection- to take advantage of the reviews and transform their work. Their idea, motivation and line of work is valuable and based on the comments can get accepted to a future venue.

Given the aforementioned, I believe that the paper is not ready for publication yet. Although, not a terrible idea if someone accepts this.

[1]. A Balanced and Uncertainty-aware Approach for Partial Domain Adaptation, ECCV 2020

[2]. Domain Adaptation Problems: A DASVM Classification Technique and a Circular Validation Strategy, PAMI 2010