View Reviews

Paper ID
1809

Paper Title
Gradual Domain Adaptation via Self-Training of Auxiliary Models

Track Name
Main Track

Reviewer #1

Questions

1. {Summary} Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).
This work proposes a new method of domain adaptation. It repeats generation of dataset of intermediate domain and training of auxiliary model multiple times to gradually switch from source domain to target domain. Experiments indicate that the proposed method likely improves the test performance from existing methods on domain-adaptation benchmarks.

2. {Novelty} How novel are the concepts, problems addressed, or methods introduced in the paper?
Good: The paper makes non-trivial advances over the current state-of-the-art.

3. {Soundness} Is the paper technically sound?
Fair: The paper has minor, easily fixable, technical flaws that do not impact the validity of the main results.

4. {Impact} How do you rate the likely impact of the paper on the AI research community?
Fair: The paper is likely to have moderate impact within a subfield of AI.

5. {Clarity} Is the paper well-organized and clearly written?
Fair: The paper is somewhat clear, but some important details are missing or unclear.

6. {Evaluation} If applicable, are the main claims well supported by experiments?
Good: The experimental evaluation is adequate, and the results convincingly support the main claims.

7. {Resources} If applicable, how would you rate the new resources (code, data sets) the paper contributes? (It might help to consult the paper's reproducibility checklist)
Not applicable: For instance, the primary contributions of the paper are theoretical.

8. {Reproducibility} Are the results (e.g., theorems, experimental results) in the paper easily reproducible? (It may help to consult the paper's reproducibility checklist.)
Fair: key resources (e.g., proofs, code, data) are unavailable but key details (e.g., proof sketches, experimental setup) are sufficiently well-described for an expert to confidently reproduce the main results.

9. {Ethical Considerations} Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias), etc.?
Not Applicable: The paper does not have any ethical considerations to address.

10. {Reasons to Accept} Please list the key strengths of the paper (explain and summarize your rationale for your evaluations with respect to questions 1-9 above).
The concept explanation is understandable since domain adaptation with intermediate domains has been known to succeed in most cases. The strategy of selecting target samples with high class-prediction scores and dropping source samples with low class-prediction scores is simple, though it requires many repetitions.

Experimental results are convincing; the proposed method likely yields higher test performance compared with existing domain-adaptation methods. The volume of experiments seems adequate. Ablation experiments well clarify the major improvement factors.

11. {Reasons to Reject} Please list the key weaknesses of the paper (explain and summarize your rationale for your evaluations with respect to questions 1-9 above).
c_s defined in Eq.(6) and f-hat are explained as the prediction probability, but I am not sure why it is so. Explanation is needed.

Eq.(9) is hard to understand. It seems that authors want to express that there are two types of f_m such that one on the lhs is the 'ensemble' f_m and one on the rhs is the 'independent' f_m. If so, Eq.(9) is not a mathematically valid expression. If not so, f_m = 1/2(f-hat + f-tilde), which does not make sense.

Parametrization of f-tilde is unclear. It says that f-tilde is obtained by minimizing Eq.(11), but it is not possible without knowing the way of parametrization.

12. {Questions for the Authors} Please provide questions that you would like the authors to answer during the author feedback period. Please number them.
Authors call f_m as 'indicator', but I am not sure why it is an indicator. Do authors mean f_m as an indicator function?

Is J(W) in Fig. 5(a) defined?

13. {Detailed Feedback for the Authors} Please provide other detailed, constructive, feedback to the authors.
Figure numbers / table numbers should be referred in a sequential fashion.

Ticks in figures are too small to see.

14. (OVERALL EVALUATION) Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper. Ideally, we should have: - No more than 25% of the submitted papers in (Accept + Strong Accept + Very Strong Accept + Award Quality) categories; - No more than 20% of the submitted papers in (Strong Accept + Very Strong Accept + Award Quality) categories; - No more than 10% of the submitted papers in (Very Strong Accept + Award Quality) categories - No more than 1% of the submitted papers in the Award Quality category
Borderline accept: Technically solid paper where reasons to accept, e.g., novelty, outweigh reasons to reject, e.g., limited evaluation. Please use sparingly.

Reviewer #2

Questions

1. {Summary} Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).
This paper proposes a self-training method that learns models for intermediate domains by gradually decreasing the proportion of source data and increasing the proportion of target data.

2. {Novelty} How novel are the concepts, problems addressed, or methods introduced in the paper?
Fair: The paper contributes some new ideas.

3. {Soundness} Is the paper technically sound?
Good: The paper appears to be technically sound, but I have not carefully checked the details.

4. {Impact} How do you rate the likely impact of the paper on the AI research community?
Fair: The paper is likely to have moderate impact within a subfield of AI.

5. {Clarity} Is the paper well-organized and clearly written?
Good: The paper is well organized but the presentation could be improved.

6. {Evaluation} If applicable, are the main claims well supported by experiments?
Fair: The experimental evaluation is weak: important baselines are missing, or the results do not adequately support the main claims.

7. {Resources} If applicable, how would you rate the new resources (code, data sets) the paper contributes? (It might help to consult the paper's reproducibility checklist)
Not applicable: For instance, the primary contributions of the paper are theoretical.

8. {Reproducibility} Are the results (e.g., theorems, experimental results) in the paper easily reproducible? (It may help to consult the paper's reproducibility checklist.)
Fair: key resources (e.g., proofs, code, data) are unavailable but key details (e.g., proof sketches, experimental setup) are sufficiently well-described for an expert to confidently reproduce the main results.

9. {Ethical Considerations} Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias), etc.?
Not Applicable: The paper does not have any ethical considerations to address.

10. {Reasons to Accept} Please list the key strengths of the paper (explain and summarize your rationale for your evaluations with respect to questions 1-9 above).
1. The idea is reasonable to address the gradual domain adaptation problem.
2. This paper is well-written and easy to follow.
3. The experiment results demonstrate effectiveness of each component of the proposed method.

11. {Reasons to Reject} Please list the key weaknesses of the paper (explain and summarize your rationale for your evaluations with respect to questions 1-9 above).
1. Novelty concern:
(a) The main problem that this paper is addressing has a strong assumption: small distribution divergence between source and target, and the intermediate domains are available. This assumption is not practical. Moreover, the experimental results are not impressive, which indicates that the proposed method may not be effective for real-world or general domain adaptation problem.
(b) The difference from Kumar et al. [1] is not clear. Kumar et al. also deals with gradual domain adaptation problems via self-training. The contribution would not be clear if the difference is not clarified.

2. Claim is not very clear:
(a) In the Discussions, the paper mentions that "Moreover, the mix of source and target data could be conducted in other ways, such as pixel-level mixup (Zhang et al. 2018)". However, the "pixel-level mixup" method is not mentioned in the paper.

3. Some technical details are not clear:
(a) This paper adopts ensemble models for further improvement. However, the reasons for selecting accessory models are not clear. This paper only mentions that clustering methods and label propagation are widely adopted in domain adaptation. This makes people feel that the ensemble model is obtained by trial-and-error without cleat motivation (e.g., DomainNet is much larger than Office-Home). Obtaining M only by the "Rw --> Cl" task of Office-Home could not be appropriate for DomainNet or VisDA. It is better to have more analyses in the paper.

4. Experiments:
(a) The number of the intermediate domains M is critical to gradual domain adaptation, especially when the intermediate domains are not defined. Therefore, M could be very different in different datasets since the scales of these datasets are different.

12. {Questions for the Authors} Please provide questions that you would like the authors to answer during the author feedback period. Please number them.
See 11.

13. {Detailed Feedback for the Authors} Please provide other detailed, constructive, feedback to the authors.
See 11.

14. (OVERALL EVALUATION) Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper. Ideally, we should have: - No more than 25% of the submitted papers in (Accept + Strong Accept + Very Strong Accept + Award Quality) categories; - No

more than 20% of the submitted papers in (Strong Accept + Very Strong Accept + Award Quality) categories; - No more than 10% of the submitted papers in (Very Strong Accept + Award Quality) categories - No more than 1% of the submitted papers in the Award Quality category

Borderline reject: Technically solid paper where reasons to reject, e.g., lack of novelty, outweigh reasons to accept, e.g., good evaluation. Please use sparingly.

20. I acknowledge that I have read the author's rebuttal and made whatever changes to my review where necessary.

Agreement accepted