View Reviews

Paper ID

11436

Paper Title

Derivative Manipulation for General Example Weighting

Reviewer #1

Questions

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

This paper introduces a new framework of example weighting to deal with noisy training labels. It relies on a new idea called derivative magnitude that can work with different types of losses under a unified weighting scheme. The authors conduct multiple experiments to verify its robustness on several data sets on which consistent improvements are reported.

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

1. The paper is well-written and generally easy to follow.

2. The use of derivative magnitude for weighting is new and seems to be theoretically sound.

3. The proposed techniques can work with multiple commonly-used losses and with different optimizers (SGD, Adam).

4. Multiple experiments are carried out on several data sets and tasks. Ablation studies are also included to verify the contribution of the proposed weighting function.

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).

1. The sentiment analysis dataset is not a vision dataset. The video retrieval dataset/setting is not very common for robust deep learning.

2. The performances on the CIFAR are not on par with the state-of-the-art published results. One reason for this is the weaker network backbone used in the experiments. The result on Clothing1M is promising but it is still worse than the best-published numbers on this dataset which are around 74.x or above.

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

Borderline

5. [Justification of rating] Please explain how the strengths and weaknesses aforementioned were weighed in for the rating. Please also mention what you expect to see from the rebuttal that may change your rating.

I reviewed this paper before. The current manuscript looks not very different from the previous version. Generally, the method looks sound, new, and is supported by probabilistic explanations. My biggest concern is whether the empirical performance reaches the bar of CVPR as it stands not very strong when compared with state-of-the-art numbers published on these datasets.

Question:

In Clothing1M experiments, did you use the ImageNet pre-trained representation? This setting seems to be different from the CIFAR experiments in which the networks are trained from scratch.

10. [Final rating] Paper rating (post-rebuttal)
Borderline

11. [Justification of final rating] Explanation of final rating. Describe the rationale for your final rating, including notes based on the rebuttal, discussion, and other reviews.
Thanks for the rebuttal. My concern still remain that whether the quality of empirical results meets the bar of CVPR. So I decide to keep my initial rating "borderline". Meanwhile, it makes sense replacing the sentiment with a vision dataset to show the work is tailored to the CV community.

Reviewer #2

Questions

1. [Summary] In 3-5 sentences, describe the key ideas, experiments, and their significance.
The paper proposed a derivative manipulation-based method for addressing the noisy and imbalanced problem of data. On vision and language datasets, the author(s) validated the proposed method.

2. [Strengths] What are the strengths of the paper? Clearly explain why these aspects of the paper are valuable.
The advantages of the proposed method in this paper:

1）By derivative manipulation of the loss function, it can directly handle example weighting;

2）Empirical experiments showed the potential advantages of the proposed method.

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).
Weaknesses of the paper:

1) According to the statement of this paper, a linear classifier is assumed. If a non-linear classifier or non-linear activation function is used in the last layer, can the method in this paper still work?

2) For some existing in-differentiable loss functions, could the method in this paper be extended?

3) Using derivative manipulation to process, will it increase more computational burden or difficulty than that by pure/direct loss function?

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

5. [Justification of rating] Please explain how the strengths and weaknesses aforementioned were weighed in for the rating. Please also mention what you expect to see from the rebuttal that may change your rating.
This paper is generally written well and the work in the paper is interesting. However, there are a few issues/comments with the work:

the author(s) compared several representatives of example weighting algorithms, for improving imbalanced data, it also has some other kinds of methods, such as class weighting. So, does it make sense to compare some representatives of existing class weighting methods?

6. [Detailed comments] Additional comments regarding the paper (e.g. typos, any suggestion to make the submission stronger)
Other minor format issues:

1) Reference [36], the name of journal or conference is missing;

2) For mathematical equations, sometimes with and sometimes without a period, for example, equations (10) and (11).

10. [Final rating] Paper rating (post-rebuttal)
Weak accept

11. [Justification of final rating] Explanation of final rating. Describe the rationale for your final rating, including notes based on the rebuttal, discussion, and other reviews.
Thanks for the responses. The author(s) basically cleared up my doubts except "the extensions of in-differentiable loss functions" and "the difficulty increase" in my questions 2 and 3. I understood that the computational burden will be reduced, but my concern is that, in order to use the proposed method, as the author(s) said, a derivative function needs to be known in advance, so for specified loss functions, every time we need to know their derivative functions first, we cannot directly use the standard framework. Considering this, I would not increase but keep my initial rating.

Reviewer #3

Questions

1. [Summary] In 3-5 sentences, describe the key ideas, experiments, and their significance.
This paper explores the connection of the loss function and the weighting scheme in noisy label problem. It studies some common loss (CCE, MAE, MSE, and GCE) and suggests that a loss actually defines an implicit weighting scheme by its derivatives.
Motivated by this finding, the authors propose Derivative Manipulation (DM), a more flexible method that modifies the derivative directly. Extensive experiments across datasets verify its effectiveness on DNN training under label noise.

2. [Strengths] What are the strengths of the paper? Clearly explain why these aspects of the paper are valuable.
The perspective of loss function as an implicit reweighting scheme is interesting.
It has taken a long while to improve the robustness against noisy labels in machine learning community. There exist quantities of different methods, including robust losses (GCE, MAE, SL, etc.) and reweighting schemes (curriculum learning, self-paced, etc.). Therefore, it is delighted to see a unified perspective, which perhaps provides a better understanding on all these proposed methods in the further. Specifically, the beta distribution weighting is an interesting extension since it includes other loss derivative forms. Besides, the authors conducted extensive experiments across datasets (CIFAR-10/100, Clothing1M, MARS, and so on), and the results seem convincing.

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).

Major:

1.It is not convincing that the proposed approach unifies the design of example weighting and loss function in Line 152-154, since authors only consider four losses (CCE, MAE, MSE, and GCE). Can authors provide more derivate forms of more robust losses proposed in previous research (e.g., SL[56], Forward loss[13], PENCIL, and so on)?
PENCIL: Kun Yi, Jianxin Wu. Probabilistic End-to-end Noise Correction for Learning with Noisy Labels. In CVPR, 2019.

2.There are two hyper-parameters in DM methods, and the best selection varies in noise types, which brings much burden in practice. Can authors provide a better guideline for selection instead of comparison on val set?

3. It is difficult to understand some concepts, especially "Emphasis Mode" in Line 229-232 and "Emphasis Variance" in Line 233-236. Can the author provide more intuition and elaboration on these introduced concepts?

4.Actually, it is not novel to regard the loss as an implicit reweighting since the paper of GCE has already provided the reweighting formulas for CCE, MAE, and GCE. Besides, with the "unified" framework, can the author provides a unified understanding on the performance of different robust losses and the reweighting methods? The description in Line 616-619 seems too rough.

Minor:

5.Is there any evidence to support "A noisy example should have smaller derivative magnitude" in Line 115? A robust model can handle the impacts from a noisy example, which does not mean the derivative magnitude from the noisy example is small.

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

Weak reject

5. [Justification of rating] Please explain how the strengths and weaknesses aforementioned were weighed in for the rating. Please also mention what you expect to see from the rebuttal that may change your rating.

The perspective in the paper, regarding loss functions as a reweighting scheme, is interesting. However, some claims still lack enough confidence and the writting can be improve further. In conclusion, the current version is not ready for being published.

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

1.To convince readers that this is actually a unified framework, can authors provide more derivate forms of other more robust losses in previous research?
2.With two hyper-parameters in DM, is there a guideline for hyper-parameter selection in practical scenarios?
3.To improve the writing, can author provide more intuition and elaboration on the introduced concepts, especially "Emphasis Mode" and "Emphasis Variance"?

10. [Final rating] Paper rating (post-rebuttal)

Weak reject

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

After reading the rebuttal and other comments, my concerns are not addressed. Regarding the loss as an implicit reweighting is not novel, and the proposed method did not beat the SOTA, i.e., the authors did not provide new insights and new SOTA results. Besides, the choosing of hyper parameter is decided on eval set. There is no insights or guidelines for it. Thus, I keep my weak rejection rating.