Paper Review

DiffAutoML: Differentiable Joint Optimization for Efficient End-to-End Automated Machine Learning

Review authors: Aymeric Côme & Omar Iken

January 27, 2022

Summary

The paper [1] is concerned with the realization of an automated "data to model" process including different modeling components, namely data augmentation, neural architecture search and hyperparameter optimization. This approach first uses data augmentation to select the argument transformation of the data. It attempts to select the examples that result in higher training loss for the model in order to handle the difficult examples. Then, it uses the DAG for the search of a neural architecture. Given the data and the architecture, it alternately updates the model parameter and the hyper parameter. The overall proposed framework is end-to-end. The experiment on ImageNet shows a slight performance improvement over existing approaches. The authors also perform an ablation study to show the effectiveness of joint modeling of the three components (data augmentation, neural architecture search, hyperparameter optimization).

Strengths of the paper

The paper is overall clearly written and easy to follow. The purpose is logical and the results support the claims. Moreover, the problem addressed by the article was of obvious interest: full end-to-end auto-ML encompassing DA, NAS and HPO. Also work considers three important tasks in the modeling process, including data argumentation, HPO, and NAS. The different components may interact with each other and impact performance. However, the main contribution of the paper is that it's the first work showing that end-to-end AutoML is possible in a fully differential manner.

Weaknesses of the paper

Although the article seems very important, it raises many questions and remarks about the work done. First, is DiffAutoML generalizable to other cases or is it task specific? Also, why did the authors group NAS and DA together and not HPO and NAS? One may also wonder about the contributions of this paper as it may seem that the contributions are minor? Another question is about the hyperparameters, in fact they only use one hyperparameter for the main experiment, so does it work for more than one? A final question is about the comparison between DiffAutoML and DSNAS, is it fair to compare DiffAutoML to DSNAS since the latter does not use a validation set?

Response to weaknesses

To verify the generalizability of DiffAutoML, the authors applied their approach to two different datasets, the first being a very large and rich dataset (ImageNet). Moreover, the proposed approach can be applied to different tasks, such as object detection, semantic segmentation, or even NLP tasks. Regarding the grouping of NAS and DA, the reason is that research on neural architecture and data augmentation uses training loss for optimization, in fact, the optimization of neural architecture search aims to reduce training loss to find better structures, while the optimization of data augmentation aims to increase training loss so that the training process pays more attention to complicated samples. Regarding the comment on contributions, DiffAutoML is the first work showing that end-to-end AutoML is possible in a fully differential manner. Concerning the remark about the limited number of hyperparameters, the authors claim that their work can be applied to a larger number of hyperparameters. Finally, it is fair to compare DiffAutoML to DSNAS since the DiffAutoML validation set is sampled from the training set.

[1] Zhou, K., Lanqing, H. O. N. G., Zhou, F., Ru, B., Li, Z., Niki, T., & Feng, J. (2020). DiffAutoML: Differentiable Joint Optimization for Efficient End-to-End Automated Machine Learning.