

제한적 데이터를 이용하는 정현파 함수 회귀에 대한 다양한 전이 학습 시나리오에서의 MAML 의 적합성에 관한 연구

윤지석, 김홍찬, 박세웅

서울대학교 전기·정보공학부 및 뉴미디어통신공동연구소

jsyoun@netlab.snu.ac.kr, hckim@netlab.snu.ac.kr, sbahk@snu.ac.kr

A Study on the Suitability of MAML for Sinusoidal Regression Using Limited Data in Various Transfer Learning Scenarios

Jiseok Youn, Hongchan Kim, and Saewoong Bahk

Department of Electrical and Computer Engineering and INMC,
Seoul National University, Korea

Abstract

In this paper, we empirically show that MAML (Model-Agnostic Meta Learning) can generate an adequate regression model for an unfamiliar sinusoid function with limited data. We compare the base model generated by MAML and by baseline, in terms of regression loss for the function after model update with few data examples.

I. Introduction

One of the hallmarks of human learning is that a person can rapidly learn an unseen thing when having past experience of learning other relevant things. For instance, a skilled inline-skater can get used to ice skating after far less trials than ordinary people. The performance will be better if the experience is more similar. Bringing this intuition to machine learning, a methodology called 'transfer learning' have emerged. It aims to use base model, which is a pre-trained model for one or more tasks (i.e., training tasks), to conduct inference on an unseen task (i.e., target task). The base model needs to rapidly adapt to the target task. In specific, base model parameters need to be updated using only few data from the target task to make an adequate model. This is important especially for some practical applications such as mobile sensing [1]. For example, It is undesirable to have a user speak every word many times each to get enough data for voice recognition.

Finn et al. [2] proposed MAML (Model-Agnostic Meta Learning) as a promising technique for this data-constrained transfer learning, showing its adaptability on sinusoidal regression, image classification, and reinforcement learning applications. However, there is no further illustration of how robust MAML is against the discrepancy between training tasks and target task. In this paper, we perform a detailed experiment on sinusoidal regression with various data-constrained transfer learning scenarios to show the robustness of

MAML that enables to make an adequate regression model for an unfamiliar sinusoid function.

II. Background

Sinusoidal regression. A sinusoid function can be expressed as $y = a \cdot \sin(wx - \theta) + b$ where x, y, a, w, θ, b stand for the input, output, amplitude, frequency, phase, and offset respectively. A regression model is trained with data examples $\{(x, y)\}$ so it can predict the corresponding output for any given input value. In sinusoidal regression, a task is defined by a fixed tuple (a, w, θ, b) .

MAML operation. MAML operation consists of two processes; base model generation and target task adaptation.

There are three steps for base model generation. 1) Firstly, MAML samples multiple training tasks. 2) For each training task, MAML picks $2K$ data examples $\{(x_1, y_1), (x_2, y_2), \dots, (x_{2K}, y_{2K})\}$, products a temporary model which is a model updated from base using K examples via gradient-descent, and computes a post-update loss for the task using remaining K examples and the temporary model. We note that this temporary model is not for practical use, but only for obtaining the post-update loss. 3) Then, MAML updates the base model via gradient-descent to minimize the sum of post-update losses over all training tasks. Step 1) ~ 3) are repeated several times. α, β are learning rates for gradient-descent in step 2) and 3), respectively.

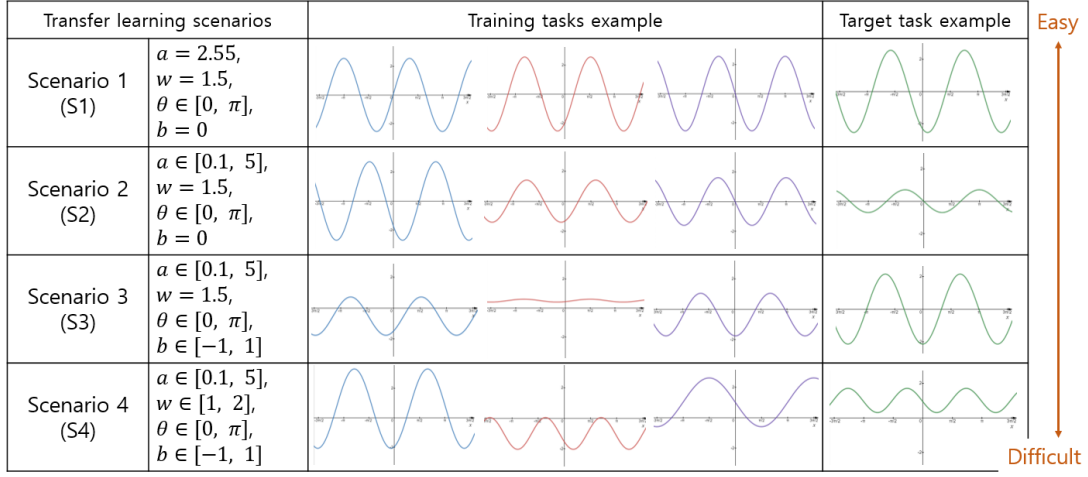


Figure 1. Four transfer learning scenarios.

The completed base model is updated (or adapted) once or more to a given target task using K_{target} data examples. Adaptation is similar to the step 2) in base model generation, but this adapted model is for practical use. Post-update (or post-adaptation) loss derived using another K_{target} examples shows the performance of the adapted model.

Baseline operation. We compare the performance of MAML with the baseline in [2]. The latter also consists of aforementioned two processes and three steps, but in step 2) there is no model update and only pre-update loss derivation using K examples and the base model itself. Accordingly, in step 3) baseline conducts gradient-descent to minimize sum of pre-update losses over all training tasks. In short, base model generation process of baseline is unaware of target task adaptation.

III. Experiment Setting & Result

To show the robustness of MAML against discrepancy between training tasks and target task, we set four transfer learning scenarios (S1 ~ S4) illustrated by Figure 1. In scenario 4 (S4), training tasks are determined by sampling a, w, θ and b from corresponding range and used to make a base model. For evaluation, the base model adapts to multiple target tasks sampled from same range and outputs post-adaptation losses over all target tasks. So, the discrepancy between tasks (or difficulty) decreases from S4 to S1 because all tasks in S3 share common w , those in S2 share w and b , and so on.

We set $x \in [-5, 5]$, $K = K_{target} = 10$ and $\alpha = \beta = 10^{-3}$. Mean Squared Error is used for loss function. We use two fully-connected hidden layers with ReLU activation function and 40 nodes each. Step 1) ~ 3) are repeated 70,000 times. 600 target tasks are sampled.

Figure 2 shows the performance of adapted model from base model generated by MAML or baseline. the x-axis stands for the number of adaptation and y-axis stands for mean (marked line) and 95% confidence interval (shaded area) of post-adaptation losses. When using a same scheme, the performance degrades from S1 to S4, due to increased difficulty. MAML outperforms

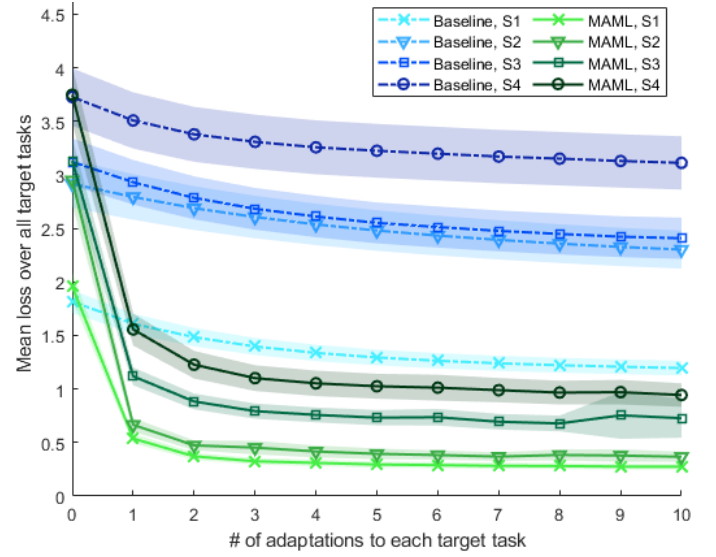


Figure 2. Performance of adapted models.

ms baseline in all scenarios since it generates base model in awareness of target task adaptation. Moreover, unlike baseline, MAML can make a good target model even with a single adaptation that employs few data.

IV. Conclusion

We demonstrate that MAML is suitable for sinusoidal regression in various data-constrained transfer learning scenarios through a detailed experiment.

ACKNOWLEDGMENT

This research was supported by a grant to Bio-Mimetic Robot Research Center Funded by Defense Acquisition Program Administration, and by Agency for Defense Development (UD190018ID).

References

- [1] T. Gong, Y. Kim, J. Shin, and S. Lee, "MetaSense: Few-Shot Adaptation to Untrained Conditions in Deep Mobile Sensing", *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, 2019.
- [2] C. Finn, P. Abbeel, and S. Levine, "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks", *Proceedings of the 34th International Conference on Machine Learning*, 2017.