03

Bayesian ensemble learning for better generalization performance

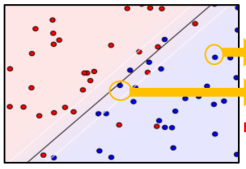
Multiple AIs make better predictions

Abstract

Evaluating the certainty of the prediction is essential for machine learning tasks. For example, certainty is required to assess the predictions' reliability, decision-making, and experimental design problems. We developed a method to efficiently calculate the certainty for a large model such as a neural network using an ensemble of models. Although evaluating the certainty of predictions using an ensemble of models has been widely used in existing work, it was theoretically unclear how to prepare ensembles. Our research theoretically derived an algorithm for preparing ensembles for expressing the certainty of prediction using multiple models. Evaluating the uncertainty is important to make machine learning reliable. We can easily evaluate the certainty using an ensemble of models and expand the range of machine learning applications by proceeding with this research.

The certainty of the prediction

When applying machine learning, the prediction as well as the "**certainty**" of how likely the prediction is essential for some tasks.



High certainty → We adopt the prediction

Low certainty → We reject the prediction

When a reliable prediction is required, such as medical diagnosis, we ask an expert to label it.

Applications using the certainty

- **Experimental design**. Gathering data that is particularly useful for the training.
- **Decision-making problems**. Determining what action to be taken next.

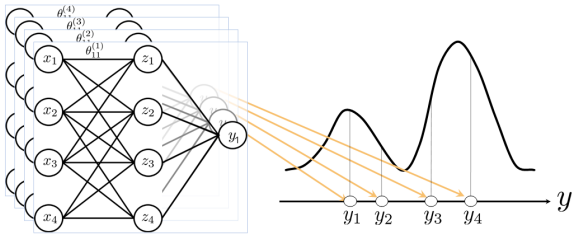
✓A widely used method of obtaining certainty is to use "**Bayesian inference**" to obtain a distribution of predictions.

✓For large models such as neural networks (NN), Bayesian inference requires approximation to perform.

Existing study: The variability of multiple models

We prepare multiple NNs, and approximate the distribution of predictions by the variation of their predictions.

The variation of multiple NNs → The distribution of the prediction → Use this as certainty



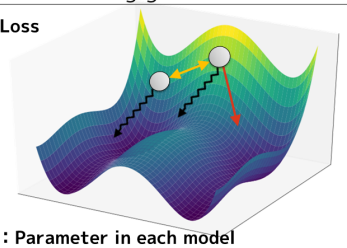
✓When we train NNs in the same way, NNs become similar.

✓How to prepare multiple NNs for the certainty ?

Our study: How to prepare multiple models

Incorporates a "**repulsion term**" that makes models differ from each other into the objective function.

Conceptual diagram of our approach when using gradient descent



Contribution :
The models obtained by our method are guaranteed to approximate a distribution better than without a repulsion term theoretically.

: **Parameter in each model**

Update direction of params = **Gradient of the loss** + **Repulsion term**

Numerical experiments: Decision-making problem

Based on the mushroom features given at each time point, we decide whether to eat or not repeatedly.

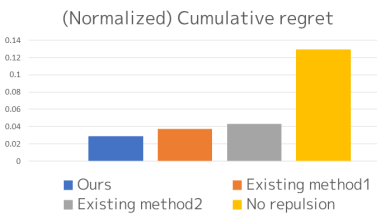
Feature 22dims

Safe → Reward + 5 point

Poison → Reward − 35 point (with 50%)
Reward + 5 point (with 50%)

✓ The goal is to maximize total reward. Only reward information is a learning cue (no labels are given as in classification problems).

✓ Using the certainty, it is necessary to control the trade-offs between exploitation and exploration for gathering information while taking actions that maximize the reward.



✓ The graph shows the cumulative regret when 50000 decisions are made. Thus, the small regret indicates better performance.

✓ We normalized each regret by setting the regret of the completely random decision as one.

✓ We used 20 NNs.

References

[1] F. Futami, T. Iwata, N. Ueda, I. Sato, M. Sugiyama, "Loss function based second-order Jensen inequality and its application to particle variational inference," in *Proc. Neural Information Processing Systems (NeurIPS)*, 2021.

Contact

Futami Futoshi / Ueda Research Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp