

# Influence of hyperparameters on online aggregation with countable experts

S. M. Kunin-Bogoiavlenskii<sup>1\*</sup>, A. V. Zukhba<sup>1</sup>, R. D. Zukhba<sup>1</sup>

<sup>1</sup>*Moscow Insitute of Physics and Technology*

*kunin-bogoiavlenskii.sm@phystech.edu; a\_\_l@mail.ru; zukhba.rd@phystech.edu*

*\*Presenting author*

---

**Abstract.** Aggregating forecasts from multiple experts is a valuable method to improve prediction accuracy. Our work examines the influence of hyperparameters on the accuracy of the aggregation algorithm for a countable number of experts. We implement a time series generator with specified properties and an aggregating forecasting model. We conduct a series of experiments with various hyperparameters of the algorithm. The experiments confirm that these hyperparameters have a significant influence on the algorithm's performance.

**Keywords.** *online learning; aggregating algorithm; prediction with experts' advice; Fixed Share, Mixing Past Posteriors (MPP)*

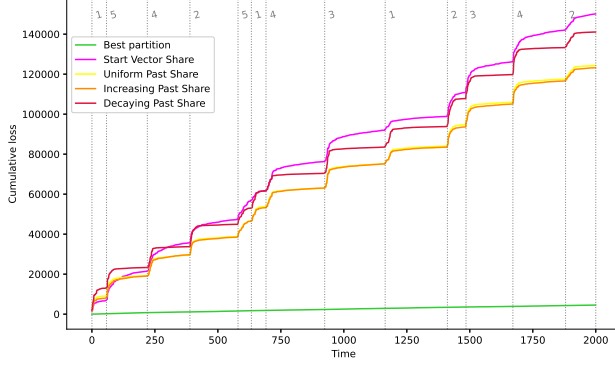
UDC: 519.246 + 519.837

---

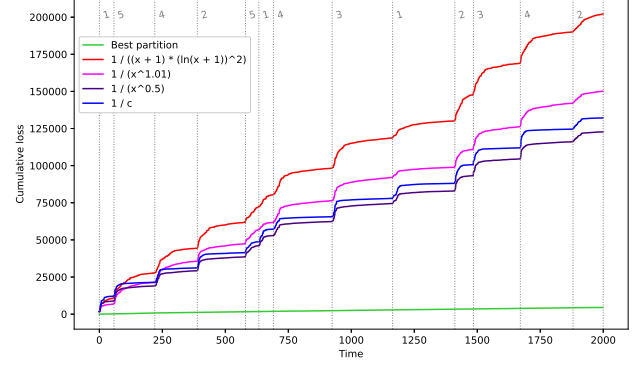
Our work focuses on the problem of online time series forecasting using expert advice, particularly when dealing with a countable number of experts. This means that the pool of potential experts is not fixed beforehand, but rather new experts can be introduced dynamically over time. Inspired by the algorithm developed in [1], which handles online prediction with countable experts, we investigate the role of hyperparameter optimization in achieving optimal performance. Our approach makes no assumptions about the underlying nature of the data, allowing it to handle deterministic, stochastic, or other types of time series.

The standard online learning framework for prediction with expert advice involves a master algorithm aggregating predictions from a set of expert models. At each time step, the master algorithm combines expert predictions, makes its own prediction, and then receives feedback in the form of a loss function. The goal is to minimize the difference between the master's cumulative loss and the best expert's cumulative loss, a metric known as regret. Traditional algorithms like Fixed Share [2] and Mixing Past Posteriors (MPP) [3] assume a fixed set of experts. However, in many real-world scenarios, the set of experts can expand dynamically. This necessitates algorithms capable of handling countable expert sets, such as the GMPP algorithm proposed in [1], where a new expert is introduced at each time step.

We explore how key hyperparameters affect the GMPP algorithm’s performance. We use a synthetic time series generator with specific properties to conduct an experimental study, which helps us control the underlying data dynamics and evaluate the impact of different hyperparameter configurations.



(a) Total losses for different mixing schemes



(b) Total losses for different weight functions

Figure 1: Total losses of the best partition and the master algorithms

Figures 1a and 1b illustrate the algorithm’s performance under different hyperparameter settings. As the metric of our experiments, we use regret - the difference between the master algorithm’s cumulative loss and the cumulative loss of the best partition. Figure 1a compares the cumulative loss of the master algorithm using different mixing update schemes, including "Start Vector Share", the default scheme in [1]. Figure 1b depicts the cumulative loss using different weight initialization functions, including the default function  $1/((x + 1) \ln^2(x + 1))$ . These examples demonstrate the significant influence of mixing schemes and weight functions on the GMPP algorithm’s performance.

Our findings highlight the importance of considering hyperparameter choices when deploying the GMPP algorithm for online time series forecasting with countable experts. Different mixing schemes and weight initialization functions can significantly impact the algorithm’s cumulative loss. In our paper we explore the effects of other hyperparameters, such as the training window size, noise variance and weights update parameters. A comprehensive understanding of these factors may guide the development of more effective and robust online aggregation strategies for countable expert sets.

## References

- [1] V. V’yugin and V. Trunov. Prognozirovaniye lokal’no statsionarnykh dannykh s ispol’zovaniem predskazanii ekspertnykh strategiy [prediction of locally stationary data

- using prediction with expert advice]. 2023.
- [2] M. Herbster and M. Warmuth. Tracking the best expert. 1998.
- [3] O. Bousquet and M. Warmuth. Tracking a small set of experts by mixing past posteriors. 2002.