Aggregating Algorithm

Andrew Barraclough

Submitted for the Degree of Master of Science in

Machine Learning

Department of Computer Science
Royal Holloway University of London
Egham, Surrey TW20 0EX, UK

June 19, 2024

Declaration

This report has been prepared on the basis of my own work. Where other published and unpublished source materials have been used, these have been acknowledged.

Word Count:

Student Name: Andrew Barraclough

Date of Submission: 29 August 2024

Signature:

Abstract

Your abstract goes here.

Contents

Acknowledgements

While the contents of this report are on the basis of my own work, none of this would have been possible without the patience and mentorship of my supervisor Dr. Yuri Kalnishkan to whom I am extremely grateful. It was your advice, clear explanations, and expertise that made this project what it is now and something that I am incredibly proud of. I would also like to express my gratitude to the group of friends who made this academic year possible, namely Cougar Tasker, Einstein Ebereonwu, Hayden Amarr, Mohammadreza Yazdian, Niraj Jain, and Ray Mahbub, without whom I would have struggled to maintain my discipline and motivation.

1 Introduction *(1,000)* (728)

1.1 Project Scope and Objectives (205)

The aim of this project is to implement methods of Prediction with Expert Advice, such as the Aggregating Algorithm, and to evaluate their performance in different scenarios, specifically targeting real-world applications.

As an introduction to the concepts explored in the chapters to come, these methods allow for the pooling of different prediction algorithms (known as 'experts') with the goal of improving prediction accuracy—allowing the final prediction to be nearly as accurate as the best-performing expert.

This project will encompass several key areas, including:

- **Explaining the Theory of Prediction with Expert Advice.** To effectively experiment with different methods of Prediction with Expert Advice, the underlying theory must first be understood by conducting a review of the relevant literature.

- **Implementing the Aggregating Algorithm.** This project will primarily investigate the Aggregating Algorithm introduced by Vovk (see [1], [2]).

- **Handling Specialist Experts.** Introduced by Freund [3], *Specialist Experts* may refrain from making predictions at certain points, meaning that the Aggregating Algorithm has to be modified slightly [4].

- **Applying Prediction with Expert Advice to Real-World Data.** The methods described in this report will be applied to real-world datasets in order to evaluate their practicality outside of theoretical models, including an investigation into the perception of randomness by utilising specialist experts.

1.2 Motivation and Interest in the Subject Area (261)

The motivation for selecting a project in this subject area is rooted in both my personal and professional interests, as well as the discussions I had with my now-supervisor, Dr. Yuri Kalnishkan, before finalising my selection.

During this academic year, the module that most piqued my interest was CS5200 – On-line Machine Learning because I was interested in the techniques that allowed machine learning models to gradually improve over time as more data became available to them without the need to retrain the model on the entire newly-updated dataset; something that had not been covered previously by other modules. Due to the module's small size and frequent absentees, I was able to gain a deeper understanding of the module, in large part due to Dr. Kalnishkan's willingness to explain portions of the syllabus in extreme detail. Alongside the lectures, I felt like I was strongly suited to the contents of the module because it has strong ties to the field of statistics – another area that I thoroughly enjoyed throughout my education.

Regarding my professional aspirations, I am set to begin my career later this year and I am of the firm belief that the work that I have done in this subject area is highly relevant, not only to the job I am to start in September, but also for my career plan due to its relevance across a variety of industries – including finance, energy, and insurance.

Ultimately, the combination of all of these factors led me to pursue a project investigating on-line prediction, and prediction with expert advice.

1.3 Structure of the Dissertation (262)

The dissertation is split into distinct chapters, each dedicated to exploring a specific aspect of the work. The following outline guides the reader through the report by providing a brief overview of the contents of each chapter.

Chapters 2 through 5 contain a literature review organised to explain the concepts that the practical portion of the dissertation aims to explore. Chapter 2 defines the problem of On-line Prediction, outlining the scenarios that it applies to, and the protocols that such problems follow. Additionally, it explores how on-line learning differs from traditional batch learning and defines concepts that will be critical to understanding the following sections. Chapter 3 introduces the problem of Prediction with Expert Advice, explaining its significance and applications in the real world, as well as exploring some algorithms that are used to solve such problems – including their theoretical bounds. Chapter 4 introduces the Aggregating Algorithm that this report is centred around, exploring how it differs from other methods of Prediction with Expert Advice. Chapter 5 focuses on Specialist Experts, defining what they are and how the base Aggregating Algorithm must be modified to accommodate them.

Chapter 6 contains the practical portion of the dissertation, explaining how the research problem was handled based on the concepts explored in the literature review, the findings from conducting the requirements analysis and design processes, and the results found when comparing an individual's idea of "random" to that of a random number generator.

Finally, Chapter 7 contains a conclusion which discusses the findings of the investigation as well as a self-evaluation of the project.

2 On-line Prediction *(1,250)*

2.1 Preliminaries

2.2 Games

3 Prediction with Expert Advice *(1,250)*

- Outcomes $(\omega_1, \omega_2, ...)$ happen in sequence. - All outcomes come from an outcome space, $\omega \in \Omega$. - In the Simple Prediction Game, $\Omega = \{0, 1\}$, i.e. bits.

- Before we see any outcomes, we make a prediction γ_t. - All predictions come from a predction space $\gamma \in \Gamma$. - In the Simple Prediction Game, $\Gamma = \{0, 1\}$, i.e. bits.

Algorithm 1 Prediction with Expert Advice

FOR $t = 1, 2, ...$
    the experts $E_1, ..., E_N$ output predictions $\gamma_t^1, ..., \gamma_t^N \in \Gamma$
    the learner L outputs $\gamma_t \in \Gamma$
    the nature outputs $\omega_t \in \Omega$
    the experts $E_1, ..., E_N$ suffer losses $\lambda(\gamma_t^1, \omega_t), ..., \lambda(\gamma_t^N, \omega_t)$
    the learner L suffers loss $\lambda(\gamma_t, \omega_t)$
END FOR

4 Aggregating Algorithm *(2,500)*

Describe the concept of the aggregating algorithm, its purpose, and how it synthesizes predictions from multiple experts to improve overall accuracy.

Algorithm 2 Aggregating Algorithm

initialise weights $w_0^i = q_i, i = 1, 2, \ldots, N$
FOR $t = 1, 2, \ldots$
  read experts' predictions $\gamma_t^i, i = 1, 2, \ldots, N$
  normalise the weights $p_{t-1}^i = w_{t-1}^i / \sum_{j=1}^N w_{t-1}^j$
  output $\gamma_t \in \Gamma$ satisfying for all $\omega \in \Omega$ the inequality
    $\lambda(\gamma_t, \omega) \leq -\frac{C}{\eta} \ln \sum_{i=1}^N p_{t-1}^i e^{-\eta\lambda(\gamma_t^i, \omega)}$
  observe the outcome ω_t
  update the experts' weights $w_t^i = w_{t-1}^i e^{-\eta\lambda(\gamma_t^i, \omega_t)}, i = 1, 2, \ldots, N$
END FOR

4.1 Weak Aggregating Algorithm

Explain the weak aggregating algorithm, its methodology, and its advantages. Discuss how it differs from stronger aggregating methods and its specific use cases.

4.2 Fixed Share Algorithm

Discuss the fixed share algorithm, its mechanics, and how it balances the use of different experts over time. Explain its relevance and application in dynamic environments.

4.3 Switching Experts

Analyze the strategy of switching between experts based on performance. Discuss the criteria for switching and its impact on prediction accuracy.

4.4 Specialist Experts & Sleeping Experts

Describe the role of specialist experts who focus on specific types of data or conditions. Discuss how their specialized knowledge enhances overall predictive performance.

Algorithm 3 Aggregating Algorithm for Specialist Experts (AASE)

initialise weights $w_0^i = q_i, i = 1, 2, \dots, N$

FOR $t = 1, 2, \dots$

  read the predictions, γ_t^n, of awake experts

  normalise the weights of awake experts

   $p_{t-1}^i = w_{t-1}^i / \sum_{i: E_i \text{ is awake}} w_{t-1}^i$

  solve the system $(\omega \in \Omega)$:

   $\lambda(\gamma, \omega) \leq -\frac{C}{\eta} \ln \sum_{n: E_n \text{ is awake}} p_t^n e^{-\eta\lambda(\gamma_t^n, \omega)}$

   w.r.t. γ and output a solution γ_t

  observe the outcome ω_t

  update the awake experts' weights $w_t^n = w_{t-1}^n e^{-\eta\lambda(\gamma_t^n, \omega)}$

  update the sleeping experts' weights $w_t^n = w_{t-1}^n e^{-\eta\lambda(\gamma_t, \omega)/C(\eta)}$

END FOR

4.5 Comparison with Model Selection

Compare the approach of prediction with expert advice to traditional model selection methods. Highlight the advantages and limitations of each approach.

5 Specialist Experts *(2,500)*

First introduced in the work of Freund et al. [3], 'specialist experts' can be thought of as an extension to the traditional on-line prediction framework that allows 'experts' to abstain from making predictions.

These experts are referred to as 'specialists' because they can be thought of as only making predictions "when the instance to be predicted falls within their area of expertise." In such cases where the expert is actively making a prediction, the expert is deemed to be 'awake', and is 'asleep' otherwise.

A prediction algorihtm may see that its internal confidence is low and decide to skip a turn in order to re-train. Alternatively, an algorithm may simply break down.

A natural idea for handling sleeping experts is to assume that a sleep expert "joins the crowd". Imagine that the sleeping expert sides with the learner and outputs the learner's prediction for that turn.

This modifies the Aggregating Algorithm, but that's to be discussed later!

In order to accommodate these 'specialist experts', we must modify the on-line prediction framework slightly. Similarly to the traditional framework, on-line learning with specialist experts can be thought of as a game that is played between a prediction algorithm, hereafter referred to as the 'learner', and an adversary, hereafter referred to as the 'nature'. The game is played in discrete iterations $t = 1, \ldots, T$, consisting of the same five steps:

1. The nature chooses a set of specialists that are 'awake' at iteration t, $E_t \subseteq \{1, \ldots, N\}$.
2. For each 'awake' specialist in the set chosen by the nature, $i \in E_t$, a prediction for that discrete iteration is output, $\hat{y}_i^t$.
3. The learner makes its own prediction based on the predictions of each awake specialist, $\hat{y}_t$.
4. The nature chooses an outcome y_t.
5. The learner suffers loss $\ell_L^t = L(\hat{y}_t, y_t)$.
6. The 'awake' specialists suffer loss $\ell_i^t = L(\hat{y}_t^i, y_t)$ while specialists that are asleep suffer no loss.

- We still assume that there are N 'experts', some of which may be 'specialist', indexed from $\{1, \ldots, N\}$.
- Predictions and Outcomes are real-valued numbers from a bounded range $[0, 1]$.
- Loss: $L : [0, 1] \times [0, 1] \to [0, \infty)$ associates a non-negative loss to each (prediction, outcome) pair.

To-Do

- How do we evaluate the performance of an algorithm?
  - In order to give the algortihm a meaningful bound, we compare the difference between the total loss of the algorithm and the total loss of the experts.
  - The total loss of the insomniac algorithm is compared against the loss of the best expert, but this doesn't make sense in this scenario because not all the experts are awake all the time, and may not make predictions.
  - Prove that the algorithm doesn't suffer large losses regardless of the adversary's strategy.
- How does this improve computational efficiency?
  - Discuss Problem Decomposition.
  - Naive algorithms make use of extremely large sets of experts which can make the calculation of predictions computationally expensive and infeasible.
  - By allowing for specialist experts, only a select handful of those experts make a prediction at a given time which can significantly reduce the computational load required.
- Discuss the applications of 'Specialist Experts'.
  - Markov Models
    * Talk about your theory and what your implemented code is about!
  - Switching Experts

5.1 Applications of Specialist Experts

6 Practical *(5,000)*

7 Conclusion *(1,500)*

References

[1] V. Vovk, "Aggregating strategies," in *Colt Proceedings 1990*, pp. 371–383, San Francisco: Morgan Kaufmann, 1990.

[2] V. Vovk, "A game of prediction with expert advice," *Journal of Computer and System Sciences*, vol. 56, no. 2, pp. 153–173, 1998.

[3] Y.Freund, R. Schapire, Y. Singer, and M. Warmuth, "Using and combining predictors that specialize," *Conference Proceedings of the Annual ACM Symposium on Theory of Computing*, 01 1997.

[4] Y. Kalnishkan, D. Adamskiy, A. Chernov, and T. Scarfe, "Specialist experts for prediction with side information," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pp. 1470–1477, 2015.