Hiroki Tanabe

Data Scientist / PhD in Computer Science / YouTuber

Experience

2022 Oct –Present **Data Scientist**, *Yahoo Japan Corporation*, Osaka, Japan

As a Data Scientist, I play a critical role in improving our display advertising conversion rate (CVR) prediction models. These models directly impact millions of dollars in daily revenue by optimizing automated bidding processes. I have led several improvements to our data aggregation process, using Hive on a Hadoop cluster. I introduced query and configuration tuning, transitioned from HiveQL to PySpark, and increased system testability by adding unit testing with mock data.

Among my notable contributions, I extended our system's ability to incorporate offline evaluation of training models. Using the workflow management tools, Argo Workflows and Airflow, I designed and implemented a new workflow that includes model evaluation. This addition allows for the routine generation of various metrics and streamlines the evaluation process.

Using the established frameworks of Python, XGBoost, and LightGBM, I develop key machine learning models. While the specifics of these models are confidential, their role in predicting CVRs is critical to increasing our revenue and improving key performance indicators (KPIs) from an advertiser's perspective.

A/B testing is an integral part of my role, providing key insights to validate our models. By analyzing KPIs against a subset of actual traffic, we ensure that only models that are expected to deliver significant improvements in KPIs are deployed.

My work with cloud platforms similar to AWS (such as S3) and GCP (such as Vertex AI) is extensive. I train machine learning models using a system similar to Vertex AI's CustomJob, and store the finished models in an S3-like storage system. I have also developed and deployed an internal-use dashboard built with Streamlit on a Kubernetes cluster. My approach to work is deeply rooted in the principles of DevOps and MLOps. I actively participate in the improvement of our team's assets by raising issues when necessary, initiating discussions with teammates, and submitting Pull Requests when I identify potential improvements. This proactive approach ensures that I am deeply integrated into the team and contribute significantly to our collective goals and continuous improvement.

2017 Feb –Present **YouTuber**

I run three different YouTube channels. My Gaming channel has gained 16,100 subscribers and accumulated 7,063,000 total views, with one of my Speedrun videos surpassing one million views. My music cover channel has gained over 4,500 subscribers and accumulated 1,170,000 total views.

Beyond traditional content creation, I've also embarked on a project where technology meets creativity: an AI virtual YouTuber channel. This channel uses a Python-based AI system to automatically generate aspects of the videos, including talk scripts, audio, motion, and captions. Specifically, the OpenAI API is used to generate the talk scripts. The AI system runs in an environment managed by the Azure Kubernetes service. This project introduced me to the basics of cloud computing platforms and container orchestration technologies.

The hands-on experience gained from managing these YouTube channels has allowed me to develop practical knowledge of digital product growth and community engagement. Combining this experience with my technical skills, I approach each task with a unique blend of creativity and analytical skills.

2020 Apr – 2022 Sept **Research Fellowship for Young Scientists (DC1)**, *Japan Society for the Promotion of Science*, Tokyo, Japan

Selected from the top 20% of Ph.D. applicants, this fellowship allowed me to pursue a research topic in multiobjective optimization algorithms.

My research focused on the development, theoretical analysis, and implementation of first-order multiobjective optimization methods. In contrast to common approaches that optimize a weighted sum of objectives, the algorithms I developed handle the optimization without predefining the weighting parameters. This approach is particularly advantageous when dealing with non-convex objective functions or when simple weights for the objectives are difficult to assign.

I was the primary contributor to all aspects of the research, including problem definition, mathematical proofs, numerical experiments, and paper writing, working primarily for discussion with my advisors. My work resulted in the submission of five papers to peer-reviewed journals, four of which were accepted.

An essential part of my research was the translation of theory into practice. To this end, I implemented the proposed methods as a Python module, published it on GitHub and PyPI under the MIT license, and enabled others in the field to use the results of my research. This ensures not only the applicability of my results, but also their potential for further development and iteration in the broader research community.

| 2019 – 2020
Feb — Mar | **Data Scientist Internship**, *HACARUS Inc.*, Kyoto, Japan |

2019 – 2020
Feb   Mar

Data Scientist Internship, *HACARUS Inc.*, Kyoto, Japan

I have designed and developed several machine learning models for the manufacturing and agricultural industries, mainly by analyzing a small number of image datasets, typically using Python. In particular, I developed models for surface defect detection using various techniques, including dictionary learning and fused lasso, and hyperspectral image segmentation using online dictionary learning and typical classification algorithms (e.g., support vector machine, random forest). Reviewed scientific papers and implemented algorithms from them using some machine learning libraries (e.g., scikit-learn), image processing libraries (e.g., open-cv, scikit-image), data visualization libraries (e.g., matplotlib, seaborn), and notebook environments (e.g., Jupyter).

More minor tasks included: Added some new algorithms (e.g., alternating direction methods of multipliers for the generalized lasso, matching pursuit for sparse coding) to the open source library "spm-image", a scikit-learn compatible library for sparse modeling and compressive sensing. Implemented batch-OMP, an improved version of the orthogonal matching pursuit algorithm, to contribute to the research project of NEDO (the New Energy and Development Organization), Japan's largest government R&D organization. I developed the front-end of an in-house web application that generates an interactive visualization of the quality of image datasets using flasks, pandas, and bokeh.

Education [*]

2019 – 2022 Oct — Sept	**PhD, Computer Science**, *Kyoto University*, Kyoto, Japan
2018 – 2019 Apr — Sept	**MSc, Computer Science**, *Kyoto University*, Kyoto, Japan, Valedictorian, early degree completion
2014 – 2018 Apr — Mar	**BSc, Computer Science**, *Kyoto University*, Kyoto, Japan, Valedictorian

Skills Summary

Technical

| Industry Knowledge | Optimization, Machine Learning, Data Engineering, DevOps, MLOps, Generative AI | Tools & Technologies | Python, TypeScript, Next.js, SQL, Hive, Hadoop, Spark, GitHub Actions, Git, Docker, Kubernetes, OpenAI API, LangChain, OpenCV |

Language

| English | Professional working proficiency | Japanese | Native or bilingual proficiency |

Publications

Papers

- Hiroki Tanabe, Ellen H. Fukuda, and Nobuo Yamashita, New merit functions for multiobjective optimization and their properties. *To appear in Optimization*, 2023. [pdf]
- Hiroki Tanabe, Ellen H. Fukuda, and Nobuo Yamashita, An accelerated proximal gradient method for multiobjective optimization. *To appear in Computational Optimization and Applications*, 2023. [doi, pdf]
- Hiroki Tanabe, Ellen H. Fukuda, and Nobuo Yamashita, Convergence rates analysis of a multiobjective proximal gradient method. *Optimization Letters*, 17, pp. 333–350, 2023. [doi, pdf]
- Hiroki Tanabe, Ellen H. Fukuda, and Nobuo Yamashita, Proximal gradient methods for multiobjective optimization and their applications, *Computational Optimization and Applications*, 72(2), pp. 339–361, 2019. [doi, pdf]
- Hiroki Tanabe, Ellen H. Fukuda, and Nobuo Yamashita, A globally convergent fast iterative shrinkage-thresholding algorithm with a new momentum factor for single and multi-objective convex optimization. *Submitted*, 2022. [pdf]

Books

- Takashi Someda, Naoki Kitora, Ippei Usami, Ryuji Masui, and Hiroki Tanabe, An Introduction to Sparse Modeling for IT Engineers (in Japanese), Shoeisha, 2021.

References

Available upon request.

---

[*]In Japan, Doctorate Degree = Doctor of Philosophy (PhD) = Doctor of Science (D.Sc.), Master's Degree = Master of Science (MSc), Bachelor's Degree = Bachelor of Science (BSc), and Informatics = Information Science = Computer Science are often used interchangeably. For this reason, PhD, MSc, BSc, and Computer Science are used here as universal translations to reduce confusion.