Hugo Monzon

Website | 0000-0002-8655-9352 | Google Scholar | hugo91@gmail.com | 070-2626-1265

Summary

Postdoctoral Researcher specializing in Probabilistic Machine Learning and Multi-Objective Optimization. Expert in developing PyTorch-based model merging algorithms and custom optimizers to improve LLM accuracy, safety, and training efficiency. My interest is to combine Multi-Objective and probabilistic Machine Learning to train LLMs that are accurate, robust and explainable. Such approach will lead to efficient training methods that adjust data relevance at each iteration automatically and reduce both training time and conflicts in tasks.

Work Experience

Postdoctoral Researcher at RIKEN AIP, Tokyo Jun 2022 - present

- **Model Merging:** Developed and implemented state-of-the-art model merging algorithms to synthesize multiple fine-tuned checkpoints into a single model. They allow to explore different multitask learning scenarios without retraining each time.
- **Custom Optimizer Engineering:** Mathematically derived and engineered custom PyTorch optimizers using multi-objective techniques to handle conflicting tasks during training.
- **Large-Scale AI Orchestration:** Orchestrated large-scale training and Supervised Fine-Tuning (SFT) jobs on HPC clusters using qsub/SGE. Managed multi-GPU environments to train and evaluate diverse architectures including ViT, BERT, and ResNet18.
- **Probabilistic Machine Learning:** Applied Variational Inference and Bayesian Deep-Learning in model merging and multi-objective multitask learning.

System Solution's Engineer at Technopro IT, Yokohama Apr 2021 - Apr 2022

- **Software Development:** Implemented a label making software (SQL queries, UI design, business logic programming), created installation packages for the product.
- **Japanese skills:** Communicated with clients via email, wrote operation manuals and documentation in Japanese.

Skills

Frameworks	Python: PyTorch, JAX, Scikit-learn, matplotlib, R: Tiddyverse, ggplot
AI/ML	Multi-Task Learning, Supervised FineTunning (SFT), Model Merging, Variational Inference, Bayesian Learning
Tools & Infrastructure	Linux, HPC (qsub/SGE), Git, LaTeX, SQL
Research	Proof of Concepts, Publication and Reviewing for International Conferences
Languages	Spanish: Native, English: Business Level - TOEFL 102/120pts, Japanese: Intermediate Level - JPLT N2.

Education

Oct 2017 - March 2021	**Shinshu University**, Doctor in Engineering.
Oct 2015 - Sep 2017	**Shinshu University**, Master in Engineering
Feb 2009 - Sep 2014	**National University of Asuncion**, Computer Engineer

Publications

Pre-print

- **Hugo Monzon**, Thomas Möllenhoff, Nico Daheim, Iryna Gurevych, Mohammad Emtiyaz Khan, *How to Weight Multitask Finetuning? Fast Previews via Bayesian Model-Merging.* Preprint, December 2024. Pages 21 https://arxiv.org/abs/2412.08147

Journal

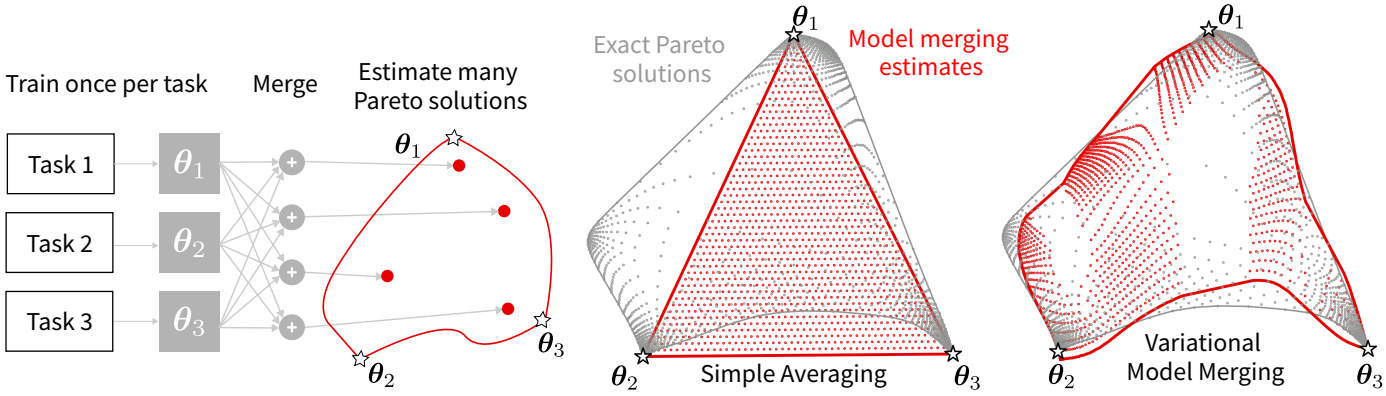
- **Hugo Monzon**, Hernan Aguirre, Sebastien Verel, Arnaud Liefooghe, Bilel Derbel, Kiyoshi Tanaka. *Understanding Population Dynamics in Multi- and Many-objective Evolutionary Algorithms for High-Resolution Approximations.* In Advances in Operation Research, Hindawi. December, 2021. Pages 16. https://doi.org/10.1155/2021/6699277.
- **Hugo Monzon**, Hernan Aguirre, Sebastien Verel, Arnaud Liefooghe, Bilel Derbel, Kiyoshi Tanaka. *Estimating Hypervolume using Population Features from Dynamic Compartmental Models.* In Transactions of the Japanese Society for Evolutionary Computation. December, 2020. Pages 14. https://doi.org/10.11394/tjpnsec.12.12

Conferences

- **Hugo Monzon**, Saul Zapotecas-Martinez. *A Dynamic Penalty Function within MOEA/D for Constrained Multi-objective Optimization Problems.* In Proceedings of IEEE Congress on Evolutionary Computation (CEC), Krakow, July, 2021. Pages 8 (1470-1477). https://doi.org/10.1109/CEC45853.2021.9504940
- **Hugo Monzon**, Hernan Aguirre, Sebastien Verel, Arnaud Liefooghe, Bilel Derbel, Kiyoshi Tanaka. *Dynamic Compartmental Models for Large Multi-objective Landscapes and Performance Estimation.* In Proceedings of the European Conference on Evolutionary Computation in Combinatorial Optimization (EvoCOP '20), Seville, April, 2020. Pages 15 (99-113). https://doi.org/10.1007/978-3-030-43680-3_7. **Best Paper Nomination**
- **Hugo Monzon**, Hernan Aguirre, Sebastien Verel, Arnaud Liefooghe, Bilel Derbel, Kiyoshi Tanaka. *Dynamic compartmental models for algorithm analysis and population size estimation.* In Proceedings of the Companion Publication of the Genetic and Evolutionary Computation Conference (GECCO '19), Prague, July, 2019. Pages 4 (2044-2047). https://dl.acm.org/doi/10.1145/3319619.3326912 **Best Student Paper Nomination**
- **Hugo Monzon**, Hernan Aguirre, Sebastien Verel, Arnaud Liefooghe, Bilel Derbel, Kiyoshi Tanaka. *Studying compartmental models interpolation to estimate MOEAs population size.* In Proceedings of the Companion Publication of the Genetic and Evolutionary Computation Conference (GECCO '19), Prague, July, 2019. Pages 2 (227-228). https://dl.acm.org/doi/10.1145/3319619.3321985
- **Hugo Monzon**, Hernan Aguirre, Sebastien Verel, Arnaud Liefooghe, Bilel Derbel, Kiyoshi Tanaka. *Studying MOEAs Dynamics and their Performance using a Three Compartmental Model.* In Companion Publication of the Genetic and Evolutionary Computation Conference (GECCO '18), Kyoto, July, 2018. Pages 2 (191-192). https://dl.acm.org/doi/10.1145/3205651.3205739
- **Hugo Monzon**, Hernan Aguirre, Sebastien Verel, Arnaud Liefooghe, Bilel Derbel, Kiyoshi Tanaka. *Closed State Model for Understanding the Dynamics of MOEAs.* In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'17), Berlin, July, 2017. Pages 8 (606-616). https://dl.acm.org/doi/10.1145/3071178.3071259
- Christian von Lucken, **Hugo Monzon**, Carlos Brizuela, Benjamin Baran. *Dimensionality Reduction in Many-objective Problems Combining PCA and Spectral Clustering.* In Companion Publication of the Genetic and Evolutionary Computation Conference (GECCO '15), Madrid, July, 2015. Pages 2 (1511-1512). https://dl.acm.org/doi/10.1145/2739482.2764636

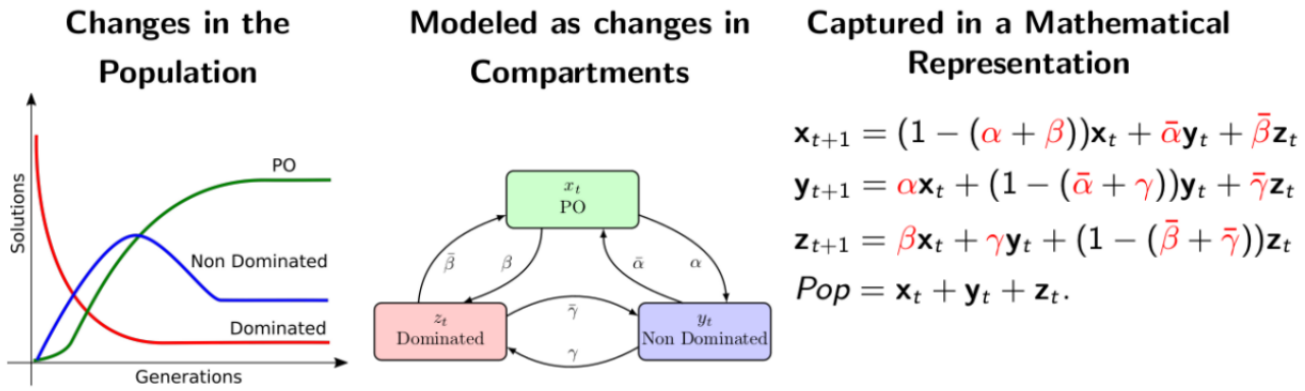
Recent Projects

Variational Model Merging for Pareto Front Estimation in Multitask Finetuning



I worked on variational Bayesian learning, my project connects the weighting used to merge models to Pareto trade-offs on an equivalent multi-task finetuned single model on all the tasks. Using probabilistic machine learning and variational inference, we show an unified framework for Simple Averaging, Task Arithmetic and other merging methods, by connecting it different complexity classes of Gaussian posteriors, allowing the development of new and more accurate merging methods like one based on mixture of Gaussians. Our framework allowed to explore multiple ways of weighting each task to achieve a desired trade-off between them, avoiding the costly retraining and using model merging as a proxy for the underlying multitask finetuning problem. Currently I am developing a new multi-objective optimizer that extends the classic Multi-Objective Gradient Descent Algorithm and Multi-Objective Newton to optimize instead of a single set of parameters a distribution over them, and allow a more robust descent on conflicting tasks, for example during the training of LLMs on instruction following and safety.

Dynamic Compartmental and Performance Models for Analysis and Configuration of MOEAs



During my doctoral course I studied Multi-Objective Evolutionary Algorithms (MOEAs), a method that simulates natural evolution by a population of solutions, and uses operators such as recombination (takes parameters from two solutions and merges them) and mutation (changes parameters at random) iteratively improving them and reaching the Pareto Set of optimal and non-dominated solutions. I proposed a model that captures changes in the optimality of solutions present in the population and correlates it to performance of the algorithm. Dynamic Compartmental Models (DCM) simulate how individuals in different stages of evolution (optimality) in the population interact and affect each other. Compartments are determined based on Pareto dominance status and presence or not at certain iteration of the solutions. The proportion in each compartment changes as the algorithm progresses in the search of the Pareto Set and this can be used to predict its performance.