Tanguy Lefort

Montpellier, France tanguv.lefort@inria.fr

GitHub: tanglef Website: https://tanglef.github.io

PostDoc in machine learning: expertise in deep-learning and crowdsourcing

Experience

PostDoc

INRIA Lille, France

- Supervisor: Odalric-Ambrym Maillard (INRIA)
- Research Focus: User expertise profiling with reinforcement learning and data collection via recommender systems
- Model user expertise for data annotation in a large-scale botanical setting
- Create a recommendation system to better exploit user interactions in the Pl@ntNet system
- Obtain theoretical guarantees and implemented algorithms for uncertainty quantification regarding predictions and user profiles in a crowdsourcing setting

Ph.D Student

IMAG, University of Montpellier and INRIA Montpellier, France

- Supervisors: Benjamin Charlier (CNRS), Alexis Joly (INRIA) and, Joseph Salmon (CNRS)
- Research Focus: Noisy labels in crowdsourced classification datasets with expert feedback
- Identify data ambiguity in cooperative annoation datasets using neural networks margins
- Standardized framework in a Python library: peerannot (https://peerannot.github.io/)
- Compare crowdsourcing aggregation strategies in a high class and large number of workers real dataset with Pl@ntNet

Intern

IMAG, Montpellier, France

- Master's thesis on High dimensional optimization for penalized linear models with interactions
- Supervisors: Benjamin Charlier and Joseph Salmon
- Benchmark descent methods for linear models with L1 and L2 penalties and first-order interactions
- Applied strategies for GPU acceleration and applied them to genomics datasets

IMAG, Montpellier, France

- Contributed to the PyKeOps library under the supervision of Benjamin Charlier
- Rewrote Scipy's Fortran conjugate gradient routine for symbolic matrices in PyKeOps
- Benchmarked Ridge-Tikhonov regularization and worked on optimal transport problems

10/2024 - 03/2026

03/2021 - 08/2021

07/2020 - 08/2020

10/2021 - 09/2024

Education

Master Biostatistics, Montpellier, France				
• Theoretical and applied statistics and probabilities				
• Classification algorithms, survival analysis, and modelization of populations				
Bachelor in mathematics, Dijon, France	2016-2019			
• Bachelor with honors, specialized in applied mathematics				
• Final project on skeletonization algorithm for gamma-ray surgery				

Publications

Journal

- Cooperative learning of Pl@ntNet's Artificial Intelligence algorithm: how does it work and how can we improve it? in Methods in Ecology and Evolution (under review) by T. Lefort , A. Affouard , B. Charlier , J-C. Lombardo , M. Chouet, H. Goëau , J. Salmon , P. Bonnet and A. Joly
- Peerannot: classification for crowdsourced image datasets with Python in Computo by T. Lefort, B. Charlier, A. Joly and J. Salmon in 2024
- Identify ambiguous tasks combining crowdsourced labels by weighting Areas Under the Margin in TMLR by T. Lefort, B. Charlier, A. Joly and J. Salmon 2024

In Proceedings of Conferences

- Weighted majority vote using Shapley values in crowdsourcing in CAp by T. Lefort, B. Charlier, A. Joly and J. Salmon 2024
- Cooperative learning of Pl@ntNet's Artificial Intelligence algorithm using label aggregation in Journées des statistiques de France by T. Lefort, A. Affouard, , B. Charlier, J. Salmon, P. Bonnet and A. Joly 2024
- peerannot: A framework for label aggregation in crowdsourced datasets in Journées des statistiques de France by A. Dubar, T. Lefort and J. Salmon 2024
- Weighting areas under the margin in crowdsourced datasets in Journées des statistiques de France by T. Lefort, B. Charlier, A. Joly and J. Salmon 2023
- Crowdsourcing label noise simulation on image classification tasks in Journées des statistiques de France by T. Lefort, B. Charlier, A. Joly and J. Salmon 2022
- Benchopt: Reproducible, efficient and collaborative optimization benchmarks in NeurIPS 2022 by
 T. Moreau, M. Massias et. al
 2022

Other involvements

Open source library contributions

- Lead developer of the Peerannot library for handling crowdsourced datasets in image classification
- Developer on the BenchOpt library for reproducible benchmarks in optimization problems

Community services

•	Reviewer for TMLR international journal	2024 – Current
•	Reviewer for Computo journal (SFDS)	2023 – Current
•	Co-organizer of the Ph.D seminar at IMAG, University of Montpellier	2022 - 2023

Teaching

- Co-advisor or Guillaume Demoor (Mines Alès) with F-D. Collin and G. Durif, Master's internship of 4 months on *Converting math formula images to LaTeX encoding with machine learning* 2024
- Advisor of 4 master students in the creation of a data visualization workshop on genomics data at the Montpellier Omics Days conference 2023
- TA for Convex Optimization to undergraduate mathematics students 2021–2023
- TA for a first-year biology course covering mathematical concepts 2021-2023
- TA for mathematical undergraduates covering logic and proof techniques 2023
- TA for second year undergraduates in chemistry: linear algebra, derivation and integration 2023

Talks (a selection)

•	Keeping humans	in the loop	via citizen .	science Univ.	Lille 11	/2024
	r					/

- Crowdsourcing and citizen science: how to handle our data? INRIA, Lille 07/2024
- Cooperative data enrichment algorithms at Journées des Statistiques de France (JDS) Univ. Bordeaux
 05/2024
- Apprentissage collaboratif d'espèces de plantes et agrégation de labels dans Pl@ntNet, IA-ECO seminar, UMR MARBEC. 03/2024
- ChatGPT & co, Myths and Reality. Everything you wanted to ask about Deep Learning but did not dare to ask, General knowledge Seminar with Francois David Collin, IMAG. 10/2023
- Data collection from a crowd: where is the noise coming from? at Ph.D students seminar, IMAG. 09/2023
- Weighting areas under the margin in crowdsourced datasets at Journées des Statistiques de France (JDS) Univ. Bruxelles 07/2023
- Learning from crowds: going beyond aggregation schemes at ML-MTP seminar, Univ. Montpellier 02/2023
- Improve learning combining crowdsourced labels by weighting Areas Under the Margin at ML-MTP seminar, Univ. Montpellier 10/2022
- Crowdsourcing label noise simulation on image classification tasks at Journées des Statistiques de France (JDS) Univ. Lyon.
 06/2022
- High dimensional optimization for penalized linear models with interactions using graphics card computational power at Probability and Statistics (EPS) team seminar 11/2021
- Introduction to neural networks with Joseph Salmon at ML-MTP seminar, Univ. Montpellier 10/2021
- Paper club Ridge Regularization: an Essential Concept in Data Science by Trevor Hastie with Florent Bascou at ML-MTP seminar 04/2021

Skills

- Tools and Languages: Python, R,Git, ${\rm I\!AT}_{\!E\!} \! X,$ JavaScript, HTML, CSS
- **Machine learning algorithms:** logistic regression, trees, SVM, dimension reduction methods, KNN, neural networks (from CNN to ViT and more), bandits, Recommender systems
- Languages: French, English, Spanish, Italian (in decreasing order)