

# Report on CIMI Thematic Trimester: Machine Learning

## 1 Thematic

The goal of this trimester was to propose a series of scientific events reflecting common interests in Machine Learning of IMT and IRIT. The focus was set on three particular axes: optimization problems in machine learning and image processing, learning with structured data and natural language, and sequential learning. Some of these events were partially dedicated to Master, doctoral and post-doctoral students.

For all events, a more detailed description, the programs, the slides and material of the courses and presentations are available on the website <http://www.irit.fr/cimi-machine-learning/>.

## 2 Organization

**Scientific Comitee:** Francis Bach (INRIA, ENS), Sébastien Bubeck (Microsoft Research), Nicolò Cesa-Bianchi (Università degli Studi di Milano), Rémi Gribonval, (IRISA Rennes), Marc Sebban (Université Jean Monnet St-Etienne), Noah Smith (Carnegie Mellon University), Johan Suykens (KU Leuven), Marc Teboulle (Tel-Aviv University)

**Coordinators:** Stergos Afantenos (IRIT), Aurélien Garivier (IMT), Sandrine Mouysset (IRIT), Mathieu Serrurier (IRIT).

**Summer School - Sept. 14-18** Aurélien Garivier, Sébastien Gerchinovitz (IMT), Josiane Mothe (IRIT), Edouard Pauwels (IRIT)

**Workshop 1 - Oct. 6-7** Nicolas Couellan (IMT), François Malgouyres (IMT) and Sandrine Mouysset (IRIT)

**Workshop 2 - Nov. 9-10** Aurélien Garivier, Sébastien Gerchinovitz, Josiane Mothe

**Big Data days - Nov. 16-17** Nathalie Aussenac-Gilles (IRIT), Philippe Besse (IMT), Wahiba Bahsoun (IRIT), Fabrice Gamboa (IMT), Aurélien Garivier, Jean-Michel Loubes

**Workshop 3 - Dec. 9-11** Stergos Afantenos (IRIT), Jean-Michel Loubes, Mathieu Serrurier, Tim Van de Cruys

## 3 Scientific Events

### 3.1 Summer School

The summer school consisted of four courses, each composed of three lectures, and a hands-on computer session (2h each). In addition, an invited talk by Noah Smith was given on Wednesday morning. In addition to researcher and (post-) doctoral students, all the Master 2 SID students of UPS attended the course.

1. Optimization in Machine Learning, by **Peter Richtárik** ( University of Edinburgh)
2. Information Retrieval and Machine Learning, by **Massimo Melucci** (Università di Padua)
3. Reinforcement Learning, by **Bruno Scherrer** (INRIA, Institut Elie Cartan Lorraine) and **Alessandro Lazaric** (Inria Lille)
4. Dictionary Learning, by **Julien Mairal** (INRIA Grenoble)

Course 3 was supposed to be given by Alessandro Lazzaric, but due to a late injury he was not able to come. One lecture and the hands-on session were given by Aurélien Garivier, the two others lectures were given by Bruno Scherrer, using Alessandro's material.

### 3.2 Workshop 1: Optimization in Machine Learning, Vision and Image Processing

The scientific program included invited talks describing the most recent advances on the theoretical properties, the numerical resolution and the modeling of optimization problems to solve problems in machine learning, vision and image processing. The typical problems optimize large vectors, matrices or tensors according to criteria involving L2, L1 and other sparsity enforcing terms, nuclear norms, Wasserstein distance and/or non-convex criteria.

- Amir Beck, Technion
- Martin Jaggi, ETH Zürich
- Lek Heng Lim, University of Chicago
- Eric Moulines, Telecom Paristech
- Jean Ponce, Ecole Normale Supérieure
- Johan Suykens, KU Leuven
- Yuning Yang, KU Leuven.
- Marco Cuturi, Kyoto University
- Joseph Landsberg, Texas A&M University
- Julien Mairal, INRIA Grenoble
- Gabriel Peyré, University of Paris-Dauphine
- Saverio Salzo, University of Genoa
- Marc Teboulle, Tel-Aviv University.

### 3.3 Workshop 2: Sequential Learning and Applications

Workshop 2 presented some recent advances in learning with sequential and dynamic data. Invited contributions address both theoretical and practical aspects of online learning. Topics included online convex optimization, prediction with expert advice, multi-armed bandits, online combinatorial problems, recurrent neural networks, as well as recent applications for, e.g., recommender systems, information retrieval, and social media analysis. The presentations were recorded and are available at <http://www.cimi.univ-toulouse.fr/en/videos>.

- Sébastien Bubeck, Microsoft Research
- Richard Combes, Supélec
- Ronen Eldan, Weizmann Institute, Israel
- Jean-Patrice Lasserre, LAAS
- Joseph Lehec, Université Paris Dauphine
- Stephano Mizzaro, University of Udine
- Jian-Yun Nie, Université de Montréal
- Vianney Perchet, Université Paris Diderot
- Eric San Juan, Université d'Avignon
- Nicolo Cesa-Bianchi, Università degli Studi di Milano
- Moez Draief, Huawei Research
- Nicolas Grislain, AlephD
- Alessandro Lazzaric, INRIA Lille
- Odalric-Ambrym Maillard, INRIA Saclay
- Fionn Murtagh, University of Derby and University of London
- Yann Ollivier, CNRS and Université Paris Saclay
- Benjamin Piwowarski, CNRS and Université Paris 6
- Gilles Stoltz, CNRS and HEC Paris

### 3.4 Big Data Days

For the third consecutive year, the CIMI organized (together with Labex AMIES, GDR MADICS and MascotNUM and CMI SID) an event dedicated to Big Data. The first Day was more specifically dedicated to research on Big Data and Computer experiments, with two working sessions on 'Big Data and Computer Experiments' and 'Big Data and HPC'. The second day had a broader, industrial perspective, and was ended by a 'Job and Internship Forum'. During both days, students from several teaching programs of Toulouse participated (in particular CMI SID, INSA Toulouse, TSE). More details are available on the website: <http://www.math.univ-toulouse.fr/~agarivie/bigdata/>.

- Didier Auroux, Université de Nice
- Balázs Kegl, Université Paris 11
- Alexandra Carpentiers, Universität Potsdam
- Clément Chevalier, Université Neuchâtel
- Pierre Jolivet, CNRS and IRIT
- Tim Van de Cruys, IRIT
- Jérôme Lacaille, SNECMA-SAFRAN.
- Guillaume Allain, Schibsted
- Philippe Goudal, Mediamobile
- Daniel Bouche, CEA Bruyères le Châtel (canceled)
- Michel Rochette, ANSYS Lyon
- Alfredo Buttari, IRIT
- Clément Calauzènes, Criteo
- Marc Spigai, IRT Saint-Exupéry
- Carole Bernon, IRIT
- Olivier Teste, IRIT

### 3.5 Workshop 3: Learning with Structured Data and applications on Natural Language and Biology

The scientific program included invited talks describing the most recent advances on the area of structured predictions with applications to natural language processing and biology. The workshop ended with an introductory tutorial to Deep Learning (3h lectures + 2h hands-on).

- **Alexandre Allauzen**, Université Paris XI
- **Xavier Carreras**, Xerox Research Centre Europe
- **François Coste**, IRISA and INRIA Rennes
- **Pascal Denis**, INRIA Lille
- **Hachem Kadri**, Université Aix-Marseille
- **Alessandro Moshitti**, Qatar Computing Research Institute and Università da Trento
- **Jian-Yun Nie**, University of Montreal
- **Christine Sinoquet**, Université de Nantes
- **Eustasio del Barrio**, Universidad de Valladolid
- **James Cussens**, University of York
- **Christophe Gonzalès**, Université Paris VI
- **André Martins**, Priberam Labs Lisbon
- **Ariadna Quattoni**, Xerox Research Centre Europe
- **Andreas Vlachos**, University of Sheffield

### 3.6 Participants

Origin	Summer School	Workshop 1	Workshop 2	Workshop 3
IMT	16	12	28	6
IRIT	43	18	32	34
Other labs in Toulouse	28	19	16	10
Other labs in France	7	7	23	12
Other labs outside France	9	1	7	6
from industry	6	0	8	6
Total	109	57	114	74

These numbers do not include the students that participated to these events as part of their Master training.

The Big Data days were attended by approximately 200 participants in total, including students from Licence and Master SID, INSA, TSE, and other masters. There were more than 40 industrial participants.

## 4 Outcomes - Interactions between Mathematics and Computer Science

It is obviously too early to measure the impact of the trimester on research in CIMI. As always with this kind of event, it permitted to foster collaboration networks, to diffuse knowledge, extend the expertise of the participants, and to be a basis for future joint publications (works in progress, to be submitted in 2016). Some of the invitees (Ronen Eldan, Sébastien Bubeck, Emilie Kaufmann, etc.) have spent up to two weeks in Toulouse, allowing for the preparation of future publications but also for discussions with several members of the laboratories even outside the field of Machine Learning.

Besides, the organization of a joint project between IMT and IRIT has strongly accelerated the connections between mathematicians and computer scientists working in data science. In particular, it permitted to imply Edouard Pauwels, just hired at IRIT, into these joint activities. It was also an opportunity to strengthen the links with other Toulouse laboratories (as testified, for instance, by the presence of speakers like Jean-Bernard Lasserre from LAAS, and by the number of participants given above).

Let us just mention four modest but immediate outcomes. Workshop 2 has led to a proposition on a European project by Josiane Mothe and Fionn Murtagh, to be submitted in January 2016. Furthermore, a collaborative work is starting between Josiane Mothe and Jian-Yun Nie from workshop 2 the results of which will be submitted at the end of 2016. Aurélien Garivier, Jean-Michel Loubes, and Mathieu Serrurier have started the supervision of a PhD on machine learning for network security in December. Workshop 3 has led to the common supervision of a Masters 2 thesis (starting March 2016) between Edouard Pauwels, Mathieu Serrurier and Stergos Afantenos.