

LUDOVIC DENOYER CURRICULUM VITAE

CONTENTS

1 Identity	2
2 Career	2
3 Supervision and Co-supervision of PhD Students	4
4 Main research topics	4
5 Publications (International Journals, Conferences, and Workshops with Program Committee)	6
6 Other Scientific Contributions	12
7 Administrative Research Duties	14
8 Project Management, Participation in Funded Projects	15
9 Relations with industry	16
10 Main Teaching Achievements	17

IDENTITY

- Ludovic Denoyer
- 12th of March 1979 (38 y.o.)
- Full Professor at LIP6 (*Computer Science Laboratory*), University Pierre et Marie Curie, Paris, France
 - *Machine Learning and Information Access* Team
 - *Chargé de mission* reporting to the president of the university on questions concerning *Open Research/Open Data/Open Education*
 - Co-head and co-creator of the Data Science Master (about 100 students)
 - Member of the *France Artificial Intelligence* thinking group reporting to the French Research ministry
- **Website:** <http://www-connex.lip6.fr/~denoyer>
- **Email:** ludovic.Denoyer@lip6.fr

CAREER

2001-2004: PhD Thesis (*Machine Learning*): "*Machine Learning for semi-structured documents classification and mapping*".

- Development of different learning models (Tree-structured HMM, Structured Kernels) for classification of complex tree structures (XML, HTML, Multi-media Documents text+image)
- Definition of a new structured prediction problem (i.e structure matching) aiming at learning how to transform a structured document to a structured document following a different schema (e.g XML to HTML). First propositions of structured learning algorithms.

2004-2005: Assistant Professor (ATER) at UPMC

2005-2014: Associate Professor¹ (*Maître de conférences*) at UPMC in the MALIRE team (*Machine Learning and information Retrieval*)

Décembre 2012: Habilitation à diriger les recherches (HDR²) - "*Learning with Relational Data: Sequential Models and Propagation Models for Structured Classification and Labeling*"

- Representation learning models for graphs and networks. Applications to classification of nodes, link prediction, multi-relational data, heterogeneous graph and information propagation.
- (Budgeted) Reinforcement learning models for data processing. Applications to structured prediction, textual reading, image classification, and features

¹In France, Assistant/Associate/Full professors have to teach 192h each year

²HDR is an additional thesis defended after having co-supervised PhD students, allowing to supervise new students alone, and to apply to full professor positions

selection.

2014-...: Full Professor (*Machine Learning and Information Access* team), LIP6, UPMC

Research Grants

2008-2009: CRCT³, *Exploring Machine Learning Techniques for Social Networks*

2009-2011: Délégation CNRS⁴, *Machine Learning for Social Networks and Sequential Decision Models*

Research Travels

Yahoo Research (Spain) - Information Diffusion and Recommendation - 1 month (2013)

The goal of the visit was to extend the proposed *Learning Social Network Embeddings for Predicting Information Diffusion* model for geographical information spreading i.e how news propagate from places to places in the world.⁵

University of Siena (Italy) - Never-Ending Learning - 1 month (2010): I have been invited by Professor Marco Gori in order to collaborate on the development of *Long-life Learning/Never Ending Learning* models. A post-doc supervised by M. Gori then spent three months under my supervision in Paris on this topic with application to vision systems.

Microsoft Cambridge (UK) - 3 weeks (2005): I have been invited at Microsoft Research in order to collaborate with Hugo Zaragoza on the Machine learning for semi-structured Information Retrieval problem.

³a CRCT is given to a researcher based on a national selection with a small acceptance rate, and removes all teaching duties

⁴a *Délégation CNRS* is given by the CNRS to a researcher based on a national selection with a small acceptance rate, and removes all teaching duties

⁵<http://labs.yahoo.com/event/ludovic-denoyer-speaks-about-learning-graph-embeddings-for-heterogeneous-networks-classification-and-information-diffusion/>

- Francis Maes (2006-2009): Reinforcement Learning for Structured Output Prediction
- Sheng Gao (2009-2013): Latent Factor Models for Link Prediction Problems
- Yann Jacob (2009-2013): Regularized Models for Labelling Heterogeneous Multi-Relational Social Networks
- Gabriel Dulac Arnold (2009-2013): Sequential Methods for Classification
- Maria Laura Maag (CIFRE Alcatel)- (2011-2015): Apprentissage pour l'anonymisation de grands graphes dynamique
- Gabriella Contardo (2013-2017): Budgeted Reinforcement Learning for Features acquisition
- Ali Ziat (CIFRE VEDECOM) (2014-2017): Spatio-Temporal Deep Neural Networks
- Nassim Aklil (ISIR) - (2013-2017): Budgeted Learning for Robots
- Aurélie Léon - (2015-...): Machine Learning for Sequential Interactive Models
- Mickael Chen - (2016-...): Generative models for Multi-views data
- Edouard Delasalles (2016-...) Sequential Models for Time Series
- Perrine Cribier-Delande (CIFRE Renault) (2017-...) Deep Reinforcement Learning in Connected Cars
- Guillaume Lample, (CIFRE Facebook FAIR) (2017-...): Generative models for text

MAIN RESEARCH TOPICS

2001-2009: Machine Learning With Semi-Structured Documents (MLDOC): During my PhD thesis, I mainly worked on machine learning for textual data, and particularly for semi-structured and multimedia documents (XML, HTML, Text+Image). In this context, I proposed a set of learning models for classification and for Information Retrieval based on bayesian networks and Fisher kernels. I also organized multiple international challenges and datasets.

2006–now: Reinforcement Learning for Structured Outputs Prediction (RLSTRUCT): The original idea was to propose new structured outputs prediction models based on reinforcement learning. This work was made at the same time than the apparition of models like LaSo or SEARN, but the proposed methods were based on weaker assumptions, and applied to complex problems like sequences and tree transformations. Part of this work was also dedicated to the idea to directly simulate the "real" inference process during learning resulting in algorithms close to what is presented for example in the recent *Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks* paper.

2009–now: Sequential (Budgeted) Models (BUDGET-RL): I proposed to incorporate

the notion of budget in machine learning models. Typically, the budget can correspond to the cost of the acquired information, or to the computational cost of a particular model. Different models able to handle this external cost have been proposed, particularly using (deep) reinforcement learning models, and with applications to features selection, features acquisition, text reading, or even image classification. The resulting models are close to *Attention* models, but without making assumptions on the nature of the input datapoints. Note that the first RL-based models we proposed on images or text are older than the articles that popularized the notion of *attention* in the Deep Learning domain. I am also working on the Model-free/Model-based paradigm in collaboration with neuro-scientists with the idea to be able to find a good trade-off between these two systems (typically, model-based algorithms are computationally expensive and one has to learn to use them only when it is necessary)

2006-2014: Representation Learning for Graph and Social Networks (GRAPH): We have proposed representation learning models for solving different problems with graphs and social networks: node classification, link prediction, multi-relational link prediction, and prediction about how information spreads in such networks. The basic idea which is now popular was to learn latent representations for the nodes in the graph, and then to perform inference directly on these representations. The most original model was the information diffusion model where the idea was to simulate information diffusion on the real network by a well-known diffusion model (Heat kernel) in the learned latent space.

2015-now: Deep Learning for (Spatio) Temporal series (TIMESERIES): This topic aims at developing new learning models for complex time series forecasting, and particularly for spatio-temporal series. Different models have been proposed with application to car-traffic prediction, meteorology, oceanography and disease spreading prediction.

2015-now: Reinforcement Learning for Human-Machine Interacion (INTERACT): This ongoing work aims at developing models able to interact with a human, like in dialog systems but for more general applications, and particularly for applications where one of the actor (computer or human) aims at helping the other to solve a particular task. The first step in this work is to develop new options discovery models – the second step being to transform these options into natural language.

2016-now: Stochastic/Generative Models (GEN): I have started to work on the (trending) topic of generative models with two structuring ideas: the first one is to incorporate uncertainty in classical predictive models, and particularly in time series forecasting models (see TIMESERIES); the second one is to incorporate uncertainty when the collected information only corresponds to a partial view or a set of partial views about the object on which we want to compute a prediction. Applications here are focused on multi-view prediction and stream/anytime prediction. Additional applications include the use of generative models for reinforcement learning (learn a distribution over the possible futures), natural language generation (Machine translation and Image captioning)

In the following list, bold papers corresponds to the most representative/interesting papers.

- INTERACT** A Meta-Learning Approach to One-Step Active Learning – Gabriella Contardo, Ludovic Denoyer, Thierry Artieres, Arxiv
- GEN** Fader Networks: Manipulating Images by Sliding Attributes – Guillaume Lample, Neil Zeghidour, Nicolas Usunier, Antoine Bordes, Ludovic Denoyer, Marc'Aurelio Ranzato – Arxiv
- BUDGET-RL** Learning Time-Efficient Deep Architectures with Budgeted Super Networks – Tom Veniat, Ludovic Denoyer – Arxiv
- BUDGET-RL** **Options Discovery with Budgeted Reinforcement Learning, A Léon, L Denoyer, Deep RL workshop at NIPS 2016**
Comment: This is not yet a published paper, but it presents a simple and elegant idea that options naturally emerge when an agent aims at solving a task but also at minimizing its cognitive effort. Here, the cognitive effort is modeled through a cost over the acquired information (i.e observations). The policy is modeled to optimize the trade-off between efficiency and cognitive effort by using policy gradient methods over a deep recurrent architecture.
- GEN** Multi-view Generative Adversarial Networks M Chen, L Denoyer, Generative Adversarial Learning workshop at NIPS 2016 and **ECML/PKDD 2017**
- GEN** Modelling Relational Time Series using Gaussian Embeddings pdf Ludovic Dos Santos, Ali Ziat, Ludovic Denoyer, Benjamin Piwowarski, Patrick Gallinari, NIPS Time Series workshop 2016
- BUDGET-RL** Recurrent neural networks for adaptive feature acquisition, G Contardo, L Denoyer, T Artières, International Conference on Neural Information Processing (Rank A), **Best student paper award**
- BUDGET-RL** Sequential Cost-Sensitive Feature Acquisition, G Contardo, L Denoyer, T Artières, Intelligent Data Analysis (Rank A) IDA 2016
- TIMESERIES** Learning Embeddings for Completion and Prediction of Relational Multivariate Time-Series, A Ziat, G Contardo, N Baskiotis, L Denoyer European Symposium on Artificial Neural Networks, Computational Intelligence (Rank B) ESANN 2016
- BUDGET-RL** Policy-gradient methods for Decision Trees, A Léon, L Denoyer, European Symposium on Artificial Neural Networks, Computational Intelligence (Rank B) ESANN 2016
- TIMESERIES** Joint prediction of road-traffic and parking occupancy over a city with representation learning, A Ziat, B Leroy, N Baskiotis, L Denoyer Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International on transportation systems , 2016
- (Bandit) Whichstreams: A dynamic approach for focused data capture from large social media, T Gisselbrecht, L Denoyer, P Gallinari, S Lamprier Ninth International AAAI Conference on Web and Social Media 2015 (Acceptance rate = 20%)
- BUDGET-RL** Reinforced Decision Trees, A Léon, L Denoyer, European workshop on Reinforce-

ment Learning EWRL 2015

TIMESERIES Car-traffic forecasting: A representation learning approach, A Ziat, G Contardo, N Baskiotis, L Denoyer, Workshop MUD, Mining Urban Data 2015 à ICML 2015

INTERACT Representation Learning for cold-start recommendation, G Contardo, L Denoyer, T Artieres, ICLR (Short Paper) 2014

GRAPH Learning Information Spread in Content Networks, C Lagnier, S Bourigault, S Lamprier, L Denoyer, P Gallinari, International Conference on Learning Representations (short paper) ICLR 2014

GRAPH Graph Anonymization using Machine Learning, ML Maag, L Denoyer, P Gallinari, IEEE 28th International Conference on Advanced Information Networking (AINA) 2014 (Rank B)

BUDGET-RL **Deep Sequential Neural Networks, L Denoyer, P Gallinari, EWRL 2015 (invited submission) and Workshop Deep Learning NIPS 2014 (Peer reviewed) 2014**

Comment: This paper has never been submitted to a conference or journal for different timing reasons. But it was one of the first paper about ‘*conditionnal computation*’. The idea is to model the fact that different datapoints do not require the same computations. The DSNN model is a graph of neural networks layers which are learned together with a policy which objective is to guide any datapoint into the computation structure: two datapoints will follow two different paths in the graph, and will thus be processed by different layers. This paper has been at the source of different new models like the *Reinforced Decision Trees* published later, and a budgeted version is still under investigation. The same idea has been used for recent papers in the community like *Neural Fabrics* or even *Outrageously large neural networks*.

GRAPH Learning latent representations of nodes for classifying in heterogeneous social networks Y Jacob, L Denoyer, P Gallinari, Proceedings of the 7th ACM international conference on Web search and data mining (Rank A+) WSDM 2014

GRAPH **Learning social network embeddings for predicting information diffusion, S Bourigault, C Lagnier, S Lamprier, L Denoyer, P Gallinari, Proceedings of the 7th ACM international conference on Web search and data mining (Rank A+) WSDM 2014**

Comment: This paper is based on the idea that based on there is no satisfactory model about how information propagates into social network. So, instead of modeling information diffusion in the network, we have proposed to learn a mapping from the network to a latent space such that information spreading can be easily modeled in the latent space by a known continuous diffusion process (i.e heat kernel). This model is efficient and allows to model information spreading on graphs with millions of nodes when existing models (e.g Cascade models) are limited to hundred of nodes.

(Pure RL)) Learning States Representations in POMDP, G Contardo, L Denoyer, T Artieres, P Gallinari, International Conference on Learning Representations (short paper) ICLR 2014

BUDGET-RL **Sequentially Generated Instance-Dependent Image Representations for Classification G Dulac-Arnold, L Denoyer, N Thome, M Cord, P Gallinari, International Conference on Learning Representations - ICLR 2014 Long Paper**

Comment: This paper proposes an original Reinforcement Learning model (and learning algorithm) which model the problem of image classification by a sequential model that acquire, at each timestep, a patch over the incoming image. As far as I know it was one of the first paper to learn an *attention model* over images (note that, when published, the term *attention model* was not in use in the machine learning community and appears later).

GRAPH Latent factor blockmodel for modelling relational data, S Gao, L Denoyer, P Gallinari, J Guo, European Conference on Information Retrieval (Rank B), 2013

GRAPH Predicting information diffusion in social networks using content and user's profiles, C Lagnier, L Denoyer, E Gaussier, P Gallinari European conference on information retrieval, 74-85 2013

GRAPH Latent factor blockmodel for modelling relational data, S Gao, L Denoyer, P Gallinari, J Guo, European Conference on Information Retrieval, 447-458 2013

BUDGET-RL Choosing which message to publish on social networks: A Contextual bandit approach, R Lage, L Denoyer, P Gallinari, P Dolog, ASONAM 2013

BUDGET-RL **Sequential approaches for learning datum-wise sparse representations, G Dulac-Arnold, L Denoyer, P Preux, P Gallinari, Machine learning 89 (1-2), 87-122, 2012**

Comment: We have proposed a reinforcement learning-based family of models able to learn to acquire features under various budgeted constraints. The main interest of such an approach is first that the features acquisition process is adaptative and depends on the input datapoint, second that this type of model can handle various types of budget (e.g acquisition per block of features, single features, cost-sensitive problems, etc...)

GRAPH Probabilistic latent tensor factorization model for link pattern prediction in multi-relational networks, GAO Sheng, L Denoyer, P Gallinari, GUO Jun, The Journal of China Universities of Posts and Telecommunications 19, 172-181, 2012

GRAPH Probabilistic Latent Tensor Factorization model for link pattern prediction, S Gao, L Denoyer, P Gallinari, IEEE Network Infrastructure and Digital Content (IC-NIDC), 2012

GRAPH Predicting information diffusion on social networks with partial knowledge, A Najjar, L Denoyer, P Gallinari, Proceedings of the 21st International Conference on World Wide Web, 1197-1204

GRAPH Link prediction via latent factor blockmodel, S Gao, L Denoyer, P Gallinari, Proceedings of the 21st International Conference on World Wide Web, 507-508 (small paper)

GRAPH Iterative multi-label multi-relational classification algorithm for complex social networks, S Peters, Y Jacob, L Denoyer, P Gallinari Social Network Analysis and Mining 2 (1), 17-29, 2012

(Pure RL) Fast reinforcement learning with large action sets using error-correcting output codes for MDP factorization, G Dulac-Arnold, L Denoyer, P Preux, P Gallinari, Machine Learning and Knowledge Discovery in Databases (ECML), 180-194, 2012

RLSTRUCT Corpus-Based Structure Mapping of XML Document Corpora: A Reinforcement Learning Based Model, F Maes, L Denoyer, P Gallinari (Book Chapter) Modeling, Learning, and Processing of Text Technological Data, 2012

GRAPH Classification and annotation in social corpora using multiple relations, Y Jacob, L Denoyer, P Gallinari, Proceedings of the 20th ACM international conference on Information and Knowledge Management CIKM, 2011

GRAPH Temporal link prediction by integrating content and structure information S Gao, L Denoyer, P Gallinari, Proceedings of the 20th ACM international conference on Information and Knowledge Management CIKM, 2011

GRAPH Link pattern prediction with tensor decomposition in multi-relational networks, S Gao, L Denoyer, P Gallinari, Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on ... 2011

BUDGET-RL Datum-wise classification: A sequential approach to sparsity, G Dulac-Arnold, L Denoyer, P Preux, P Gallinari, Machine Learning and Knowledge Discovery in Databases (ECML), 375-390

BUDGET-RL Text classification: a sequential reading approach, G Dulac-Arnold, L Denoyer, P Gallinari, Advances in Information Retrieval (ECIR), 2011

GRAPH Tensor decomposition model for link prediction in multi-relational networks, S Gao, L Denoyer, P Gallinari, Network Infrastructure and Digital Content, 2010 2nd IEEE International ...2010

GRAPH Madspam consortium at the ecml/pkdd discovery challenge 2010, LDA Sokolov, T Urvoy, O Ricard, Proceedings of the ECML/PKDD 2010

MLDOC Report on INEX 2009, T Beckers, P Bellot, G Demartini, L Denoyer, CM De Vries, A Doucet, ..., ACM SIGIR Forum 44 (1), 38-57 2010

RLSTRUCT,GRAPH Iterative annotation of multi-relational social networks, S Peters, L Denoyer, P Gallinari, Advances in Social Networks Analysis and Mining (ASONAM), 2010

GRAPH A ranking based model for automatic image annotation in a social network, L Denoyer, P Gallinari, Proceedings of the 4th international AAAI conference on weblogs social media, 2010

MLDOC Overview of the INEX 2009 XML mining track: Clustering and classification of XML documents, R Nayak, C De Vries, S Kutty, S Geva, L Denoyer, P Gallinari, Focused Retrieval and Evaluation, 366-378, 2010

MLDOC Report on the xml mining classification track at inex 2009, L Denoyer, P Gallinari, INitiative for the Evaluation of XML Retrieval 2009 Workshop Preproceedings, 2009

RLSTRUCT **Structured prediction with reinforcement learning, F Maes, L Denoyer, P Gallinari, Machine learning 77 (2-3), 271, 2009**

Comment : This paper adresses the generic problem of structured output prediction with a new class of models coming from the reinforcement learning domain. It shows how structured output prediction can be solved with sequential models, with applications to sequence labelling and to tree-transformations. In this last application, other sequential-based methods (Like LaSo or SEARN) cannot be used since they rely on stronger assumptions. Our technique is able to still solve these complex problems. As far as we know, it was one of the first paper to propose to use reinforcement learning for dealing with classical machine learning problems.

MLDOC Report on INEX 2008, G Demartin, L Denoye, A Douce, KN Fachry, P Gallinar,

S Gev, ACM SIGIR Forum 43 (1), 17-36, 2009

(Medecine) Comparative study of aspheric intraocular lenses with negative spherical aberration or no aberration, A Denoyer, L Denoyer, J Halfon, S Majzoub, PJ Pisella, Journal of Cataract & Refractive Surgery 35 (3), 496-503, 2009

- Automatic Identification of Paedophile Keywords, C Belbeze, D Chavalarias, L Denoyer, R Fournier, JL Guillaume, ... Measurements and Analysis of P2P Activity Against Paedophile Content Project, 2009
- UPMC/LIP6 at ImageCLEFannotation 2009: Large Scale Visual Concept Detection and Annotation, A Fakeri-Tabrizi, S Tollari, L Denoyer, P Gallinari, CLEF working notes, 2009

RLSTRUCT,**GRAPH** Simulated iterative classification a new learning procedure for graph labeling, F Maes, S Peters, L Denoyer, P Gallinari, Machine Learning and Knowledge Discovery in Databases (ECML), 47-62, 2009

MLDOC Overview of the INEX 2008 XML mining track, L Denoyer, P Gallinari, Advances in Focused Retrieval, 401-411, 2009

(Medecine) Intraindividual comparative study of corneal and ocular wavefront aberrations after biaxial microincision versus coaxial small-incision cataract surgery, A Denoyer, L Denoyer, D Marotte, M Georget, PJ Pisella, British Journal of Ophthalmology 92 (12), 1679-1684, 2008

MLDOC Report on the XML mining track at INEX 2007 categorization and clustering of XML documents, L Denoyer, P Gallinari, ACM SIGIR Forum 42 (1), 22-28, 2008

(Medecine) Influence of the Corneal Incision Size on Corneal and Ocular Higher-Order Aberrations After Cataract Surgery: Microincision versus Small-Incision, A Denoyer, L Denoyer, J Halfon, PJ Pisella, Investigative Ophthalmology & Visual Science 49 (13), 5662-5662, 2008

GRAPH Web spam challenge 2008, C Castillo, K Chellapilla, L Denoyer, Proceedings of the 4th International Workshop on Adversarial Information 2008

RLSTRUCT Applications of Reinforcement Learning to Structured Prediction, F Maes, L Denoyer, P Gallinari, Recent Advances in Reinforcement Learning, 2008

MLDOC Machine Learning for Semi-structured Multimedia Documents: Application to Pornographic Filtering and Thematic Categorization, L Denoyer, P Gallinari, (Book Chapter) Machine Learning Techniques for Multimedia, 227-247, 2008

MLDOC Probabilistic model for structured document mapping, G Wisniewski, F Maes, L Denoyer, P Gallinari, International Workshop on Machine Learning and Data Mining in Pattern Recognition, 2007

- The shared corpora working group report, A Meyers, N Ide, L Denoyer, Y Shinyama, Proceedings of the linguistic annotation workshop, 184-190, 2007

MLDOC Document Track-XML Structure Mapping, F Maes, L Denoyer, P Gallinari, Lecture Notes in Computer Science 4518, 540-552, 2007

GRAPH Web spam challenge track ii, C Castillo, BD Davison, L Denoyer, P Gallinari, ECML/PKDD Graph Labelling Workshop, Warsaw, Poland, 2007

- GRAPH** PROCEEDINGS OF THE GRAPH LABELLING WORKSHOP AND WEB SPAM CHALLENGE, C Castillo, B Davison, L Denoyer, ECML PKDD 2007
- MLDOC** XML Structure Mapping, F Maes, L Denoyer, P Gallinari, Comparative Evaluation of XML Information Retrieval Systems, 540-551, 2007
- MLDOC** Mining XML documents, L Candillier, L Denoyer, P Gallinari, MC Rousset, A Termier, ..., (Book Chapter) Data mining patterns: new methods and applications, 2007
- RLSTRUCT** Sequence labeling with reinforcement learning and ranking algorithms, F Maes, L Denoyer, P Gallinari, Machine Learning: ECML 2007, 648-657, 2007
- MLDOC** XML structure mapping application to the PASCAL/INEX 2006 XML document mining track, F Maes, L Denoyer, P Gallinari, Advances in XML Information Retrieval and Evaluation: Fifth Workshop of the ..., 2006
- MLDOC** Report on the XML Mining Track at INEX 2005 and INEX 2006, L Denoyer, P Gallinari, AM Vercoustre, Comparative Evaluation of XML Information Retrieval Systems, 432-443, 2007
- MLDOC** **The XML Wikipedia Corpus, L Denoyer, P Gallinari, Sigir Forum 40 (1), 64-69, 2006**
Comment: Not a complex paper, but still a complex work at that time due to the size of the dataset. This paper is a structring paper in the information retrieval community and the corresponding dataset has been used in many hundreds of research projects and papers.
- MLDOC** Xml mining challenge at inex 2005, L Denoyer, P Gallinari, AM Vercoustren Technical report, University of Paris VI, INRIA, 2006
- MLDOC** Semi-Structured Document Classification, L Denoyer, P Gallinari, (Book Chapter) Encyclopedia of Data Warehousing and Mining, 2005
- MLDOC** Stochastic models for document restructuration, P Gallinari, F Maes, L Denoyer, ECML Workshop On Relational Machine Learning, 2005
- MLDOC** XML document mining challenge, L Denoyer, P Gallinari, Advances in XML Information Retrieval and Evaluation, 5th International ..., 2005
- MLDOC** Bayesian network model for semi-structured document classification, L Denoyer, P Gallinari, Information processing & management 40 (5), 807-827, 2004
- MLDOC** Document structure matching for heterogeneous corpora, L Denoyer, G Wisniewski, P Gallinari, SIGIR, Workshop on IR and XML, 2004
- MLDOC** Structured multimedia document classification, L Denoyer, JN Vittaut, P Gallinari, S Brunessaux, S Brunessaux, Proceedings of the 2003 ACM symposium on Document engineering, 153-160, 2003
- MLDOC** Using Belief Networks and Fisher Kernels for structured document classification, L Denoyer, P Gallinari, Knowledge Discovery in Databases: PKDD 2003, 120-131, 2003
- MLDOC** A belief networks-based generative model for structured documents. An application to the XML categorization, L Denoyer, P Gallinari, Machine Learning and Data Mining in Pattern Recognition, 277-302

(Text classification) HMM-based passage models for document classification and ranking, L Denoyer, H Zaragoza, P Gallinari, Proceedings of ECIR-01, 23rd European Colloquium on Information Retrieval ECIR 2001

OTHER SCIENTIFIC CONTRIBUTIONS

Open Softwares and Datasets

- 2017 - RL PyTorch: A RL⁶ simple package for PyTorch and OpenAI Gym (\approx 150 stars on GitHub)
- 2016 - RL Torch: A RL⁷ simple package for Torch and OpenAI Gym (\approx 35 stars on GitHub)
- Multiple source codes on GitHub
- Wikipedia XML Corpus (2006): First large XML datasets used in many international competitions (INEX, CLEF, ...). About 500 publications are directly citing this dataset, and many other are using it without citations...

Collaborations with external researchers

- Carlos Castillo, Brian Davison - (2006 – 2008). Creation of the international WebSpam challenge⁸
- Richi Nayak et Chris De Vries - Queensland University of Technology (2005-2009). Creation/Sueprvision of the XML Mining Challenge during 5 years
- Philippe Preux - INRIA LNE (Equipe Sequel) - (2008 – 2014) Budgeted Reinforcement Learning
- Benoit Girard et Medhi Khamassi (Neuroscience) - ISIR - Paris - (2013–...) Model-based versus Model-free models.
- Marc Aurelio Ranzato - Facebook FAIR - Paris/New York - (2016–...) Generative models for machine translation and image captionning

Participation to program committees

- Journal of Information Processing and Management, Pattern Recognition Letters, Journal Data and Knowledge Engineering, International Journal of Approximate Reasoning on Information Retrieval and Graphical Models, Journal of Systems and Software, Information Retrieval, Journal of Social Networks Analysis and Mining, International Journal of Intelligent Information and Database Systems (IJIIDS), RAIRO ITA
- CIKM (2007,2011,2012,2013), CORIA (2008, 2009, 2010, 2011,2012,2013,2014), INEX (2005, 2006, 2007, 2008, 2009, 2010), SIGIR (2012,2013), ECAI (2012), AN-NPR (2012,2013), Cap(2014,2015,2016,2017), NIPS (2014,2015,2016,2017), ICML

⁶<https://github.com/ludc/rl>

⁷<https://github.com/ludc/rltorch>

⁸<http://webspam.lip6.fr>

(2014,2015,2016,2017), ICLR (2014,2015,2016,2017), IJCAI (2015, 2016,2017), SIGIR 2005, WSOM 2005, RJCRI 2006

- Workshops : AirWeb (2007, 2008, 2009), FLAC (2007, 2008), Reiso (2010,2011) Infosird (2011), ISMIS 2011, Web Social (EGC 2011), WWW-MSND (2012,2013), Web Social - EGC (2012), ISMIS Social Networks (2011), Multiples workshops at NIPS, ICML

Participation to PhD/HDR juries

PhD thesis

- Robin Allesiardo, 2016, Multi-armed bandits for non-stationary data streams
- Tzirizo RABENORO, 2015, Statistical tools for preventive maintenance over aircraft engines
- Mohamed EL MAHRSI, 2013, Machine learning with trajectories
- Cedric Lagnier, 2013, Information diffusion in social networks
- Laurent Boyer, 2011, Edit distances learning
- Matthieu Durut, 2011, Decentralized learning models
- Alfonso Romero, 2009, Document Classification Models Based On Bayesian Networks
- As a reviewer:
 - Alberto GARCÍA-DURÁN, 2016, Learning representations in multi-relational graphs : algorithms and applications
 - Timothé Collet, 2016, Optimism in Active Learning
 - Maria-Irina Nicolae, 2016, Learning Similarities for Linear Classification: Theoretical Foundations and Algorithms
 - Djalel BENBOUZID, 2014, Sequential prediction for budgeted learning : Application to trigger design
 - Bilal Piot, 2014, Off-line reinforcement learning based on expert traces
 - Julien Becker, 2014, Protein Structural Annotation: Multi-task learning and feature selection
 - Olivier Nicol, 2014, Data Driven Evaluation of Contextual Bandits

Habilitation à diriger les recherches

- Albert Bifet, 2016, Big Data Stream Mining, Large Scale Real-Time Analytics

Scientific Popularization

- Interview for *La Recherche* on AlphaGo

- November 2016: TEDX La Rochelle, Machine Learning
- Reinforcement Learning for Data Processing, Machine Learning Meetup, Paris 2015
- Deep Reinforcement Learning, Deep Learning Meetup, Paris 2015
- Le controle Parental Sur Internet, Qu'en savez-vous vraiment ? Conférence pour les curieux, 2006

ADMINISTRATIVE RESEARCH DUTIES

2016: Member of the program committee at *National Research Agency (ANR) – CE23 "Big Data, Knowledge, and Parallelized Computation"*

2014 – 2016: Co-organizer of the Machine learning and information retrieval *Hackdays* (4 events) (CORIA 2014, CORIA 2016, CAP 2014 and CAP 2016)⁹ – The idea of the hackdays is to propose a full research day dedicated to an original domain, and to provide a complete datasets to participants. PhD students and researchers are grouped into teams and develop original algorithms/software for a particular problem they have imagined. The proposed topics were *Computational Cooking*, *Budgeted Learning* and *TV Series/Movies*

2015: Industry Chair, Conference IEEE DSAA 2015

2015: Co-president of the scientific committee of the French machine learning conference CAP 2015

2013: Co-organizer of the *Sequential Prediction workshop* at ICML.

2005-2010: Organizer of the *XML Mining Challenge* (PASCAL and DELOS European Networks of excellence) regrouping more than 40 research teams on the problem of machine learning with complex graphs.

2007-2008: Co-Organizer of the *Web Spam/Graph Labeling Challenge* in collaboration with Yahoo! Research + workshop at ECML/PKDD.

Other Administrative Duties

2017: Member of the *France Artificial Intelligence* thinking group reporting to the French Research ministry.

2016-2017: *Chargé de mission* reporting to the president of the university on questions concerning *Open Research/Open Data/Open Education*

2011-2014: Member of the expert committee¹⁰ in Computer Science at UPMC

2009-2010: Member of the laboratory council

Member of selection committees: (In charge of recruiting new professors)

- University Pierre et Marie Curie - 2013, 2014, 2015

⁹<http://hackday.lip6.fr>

¹⁰In charge of recruiting new professors

- University Lille 3 - 2009, 2010, 2011, 2012, 2013
- University of Saint Etienne - 2010, 2012, 2013, 2014
- University of Grenoble - 2010

PROJECT MANAGEMENT, PARTICIPATION IN FUNDED PROJECTS

In France, all research funds (except permanent researchers salaries) for travels, computers and master/phd/post-doc students are obtained through collaborative projects that are launched based on a selection procedure with a low acceptance rate $\approx 10\%$. ANR corresponds to the French Research agency, while DIM and FUI projects are regional projects. Each project corresponds to 150 K€ to 300K€ for each partner

Academic Projects

- ANR Marmota - 2005...2008 - MACHine learning pRObabilistic MOdels Tree LANguages - Participants : Mostrare (INRIA LNE), BDAA (LIF - Marseille), Laboratoire Huber Curien (Saint Etienne), LIP6 (Paris)
- ANR Lampada - 2009...2014 - Learning Algorithms, Models an sPARSE representations for structured DATA - Participants : Mostrare (INRIA LNE), SequeL (INRIA LNE), BDAA (LIF - Marseille), Laboratoire Huber Curien (Saint Etienne), LIP6 (Paris)
- DIM/DIGITEO REMI - 2011..2014 - RELational data based MACHines for Image annotation - Participants : Telecom ParisTech (Paris), LIP6 (Paris)
- ANR MLVIS - 2012-2015 - Machine Learning for visual annotations in large social networks- Participants : Telecom ParisTech (Paris), LIP6 (Paris)
- LABEX SMART OnBul - 2013 - 2016 - Online Budgeted Learning (**Manager**) - Participants : ISIR (Paris), LIP6 (Paris)
- (IUIS - UPMC) APTITUDE - 2016 - Deep LEarning pour l'aide à la chirurgie de la cataracte (**Manager**) - Participants: LIP6 (Paris) et Institut de la Vision (Paris)
- ANR Deep In France - 2016 - 2020: regrouping Deep Learning researchers in France

Academic/Industrial Projects

- ANR ATASH - 2005...2008 - Machine Learning for Documents Transformations - Participants : Xerox RD (Grenoble), Mostrare (INRIA LNE), LIP6
- ANR MADSPAM - 2008...2011 - Automatic SPAMdexing detection in large Information Networks - Participants : BlogSpirit (Malako), Orange RD (Issy les Moulineaux), LIP6 (Paris)
- ANR/DGCIS ExDeus/Cèdres 2009...2012 - Social Networks Analysis - Participants : KXEN, LIPN (University Paris 13), Skyrock, AF83, Heaven, Mondomix, LIP6

- Systematic Vigies 2008...2011 - Visualization and Electronic Interceptions – Participants : Alcatel-Lucent, France Telecom, Onera, Vecsys, LIP6
- FUI DIFAC 2010...2013 - Information Spreading and Opinion Leader Detection in Social Networks – Participants : BlogSpirit (Malako), Orange RD (Issy les Moulineaux), LIP6 (Paris)
- ANR LIVES 2016 - 2020 - Multi View Learning - Participants: LIF, Laboratoire Huber Curien, Hopital de la Timone
- ANR PAMELA 2016 - 2020 - Decentralized Privacy-safe Machine Learning - Participants: LIP6, INRIA Lille, INRIA Rennes, Mediego, SNIPS

New incoming Projects

- Projet *Deep Learning for Autonomous Sailing* in collaboration with XXXX¹¹ for the development of automatic sailing pilots for race boats (Vendée Globe and America Cup)

RELATIONS WITH INDUSTRY

Invited talks in companies

- Overview of Deep Learning, Math/Industry meeting 2017
- Budgeted Learning - Google (SF) - 2017
- Representation learning for graphs and social networks , Criteo, 2017
- Sequential Budgeted Learning - Facebook FAIR (Paris) - 2016
- Deep Learning, Roundtable, Conference CESA 2016
- Machine Learning, Deep Learning - Artificial Intelligence symposium - Renault 2016
- Budgeted Sequential Learning for Data Processing, Criteo , 2015
- Big Data, Data Science, Enseignement et Recherche, Séminaire DEP 2014 (Data Quality) – Working group regrouping many big french companies
- Learning Graph Embeddings for Heterogeneous networks classification and Information Diffusion, Yahoo! Barcelona, 2013

Other activities

- Multiple participations to industrial projects
- Co-supervision of PhD student with companies (*Thèse CIFRE*) – Alcatel/Nokia, Renault and Facebook
- Participation in the joint Thalès/LIP6 research team
- Many connections with the *French tech* (AI startups group in Paris)

¹¹confidential

MAIN TEACHING ACHIEVEMENTS

- 2014- now:** Co-head of the *Data Science* Master at UPMC – <http://dac.lip6.fr/master> which is a very successful master program with about 100 students accepted over many hundred of candidates.
- 2014-2016:** Head and creator of the *Business Intelligence Course*
- 2011-2016:** Head and creator of the *Web Technologies Course*
- 2012-now:** Head and creator of the *Introduction to Data Science Course*
- 2012-2014:** Co-head of one of the speciality of the computer science master (Machine Learning and Data Mining)
- (approx...) 2006-2011:** Head of the Oriented Object Programming course
- 2006-2014:** Head of the Machine Learning Course at ENSIIE (Grandes ecoles)