Nicolas Goix

date of birth: 07-30-1989
nationality: French

Télécom ParisTech, CNRS-LTCI (Office: E305)
46 rue Barrault, 75013 Paris, France
☎ *+33 (0) 1 45 81 73 24*
✉ *goix [dot] nicolas [at] gmail [dot] com*
http://ngoix.github.io

Areas of Expertise

Statistical Learning, Stochastic Modelling, Anomaly Detection.
I have a strong mathematical background and solid programming skills, with expertise in Machine Learning. It provides me with well-founded problem solving capabilities and analytical insight.

Education

Oct 2013 – Dec 2016 **Ph.D. in applied mathematics**, *Télécom ParisTech*, Paris, France.
- title: Machine Learning Methods to Anomaly Detection.
- main contributions:
  - theoretical advances in the dependence estimation of multivariate extreme events [4].
  - an anomaly detection algorithm based on multivariate extreme value theory, with theoretical and empirical guarantees [3, 7, 6].
  - an efficient way to evaluate unsupervised anomaly detection algorithms [5, 1].
  - a one-class random forest algorithm, which structurally extend random forests to one-class classification [2].
- supervisors: Stéphan Clémençon and Anne Sabourin.
- in parallel: contributor to scikit-learn (main contributions: implementation of Isolation Forest and Local Outlier Factor algorithms).
- in parallel: member of the academic research team of the chair 'machine learning for big data' funded by Safran, PSA Peugeot Citroën, Criteo and BNP Parisbas.

2011 – 2013 **Normalien in mathematics**, *Ecole Normale Supérieure*, Cachan, France.
- ranked 5th to the national examination (3th year entry).
- Normalien status : effectively a trainee civil servant provided with a government salary until graduated.
- 3th year: MSc université Paris VI (see below).
- 4th year: research internship (see section work experiences).

2011 – 2012 **MSc (Master 2) in probability and stochastic models**, *université Paris VI*.
- brownian motion and stochastic calculus, limit theorems and large deviations, Markov processes, Levy processes, stochastic flows methods.
- master thesis: concentration inequality and applications to random graphs (supervisor: Rama Cont).
- graduated with 77.5/100.

2009 – 2011 **Magistère in fundamental and applied mathematics**, *université Paris-Sud XI*, Orsay, France.

2007 – 2009 **Classes préparatoires MPSI-MP* (mathematics, physics, computer science)**, *Lycée Hoche*, Versailles.

Work Experiences

Fev 2017 – present **Independent Consultant**, *BPCE Group*, Paris, France.
Building machine learning based tools to detect credit card fraud, money laundering, terrorism financing in a "Big Data" setting.

May 2016 – Aug 2016 **Research scholar visitor**, *NYU center for data science*.
- modelling sensor data from black holes, and work on scikit-learn.
- supervisor: Andreas Müller.

Sep 2014 – Aug 2015 **Part-time programmer for scikit-learn**, *Paris-Saclay center for data science*.
- implementation of Isolation Forest algorithm (merged on scikit-learn), and work on issues and pull requests.
- supervisor: Alexandre Gramfort.

Sep 2012 – Aug 2013 **One-year research internship**, *LPMA, Paris VI*.
- systemic risk in interbank networks : stochastic modeling and asymptotic analysis.
- supervisor: Rama Cont.

Participation in Research Projects

2014 – present **Machine learning for big data**, a chair led by Stéphan Clémençon and funded by major companies such as Safran, PSA Peugeot Citroën, Criteo and BNP Parisbas.

Computer Skills

OS UNIX/Linux, Windows.

Programming Python

Invited Talks

Mars 2016 Damex: detecting anomalies in high dimension.
- description: my work was selected to be presented at the french ministry of industry in front of industral companies within the 'bourse aux technologies–industrie du futur–smart manufacturing'.
- venue: ministère de l'économie, de l'industrie et du numérique, Paris, France.

Dec 2015 Anomaly detection in scikit-learn and new tools from multivariate extreme value theory.
- description: introduction to anomaly detection through scikit-learn and presentation of my PhD work on EVT.
- venue: Télécom ParisTech, TSI department seminar, Paris, France.

Oct 2015 Anomaly detection algorithms in scikit-learn.
- description: presentation of my contribution to scikit-learn.
- venue: OSI day (Open Software Initiative), Paris-Saclay center for data science, Orsay, France.

Oct 2015 Anomaly detection with multivariate extremes.
- description: presentation of my work, DAMEX (Detecting Anomalies using Multivariate EXtreme) algorithm, to Safran, PSA Peugeot Citroën, and BNP Parisbas.
- venue: machine learning for big data chair - GT predictive maintenance, Paris.

May 2015 Scoring anomalies among multivariate extreme observations.
- description: presentation of my PhD work on extremes value theory and anomaly detection.
- venue: séminaire de statistique AgroParisTech, Paris.

Jan 2015 Approximating hierarchical MV-sets for hierarchical clustering.
- description: presentation of an article from A. Glazer, O. Weissbrod, M. Lindenbaum, S. Markovitch.
- venue: SMILE seminar, 'NIPS defriefing', Paris.

Languages

French (mother tongue), **English**(fluent), **German**(knowledge).

Publications

[1] N. Goix. How to Evaluate the Quality of Unsupervised Anomaly Detection Algorithms? In *ICML Workshop on Anomaly Detection, co-winner of the best paper award sponsored by Google*, 2016.

[2] N. Goix, R. Brault, N. Drougard, and M. Chiapino. One Class Splitting Criteria for Random Forests with Application to Anomaly Detection. 2016. Submitted paper.

[3] N. Goix, A. Sabourin, and S. Clémençon. Sparse Representation of Multivariate Extremes with Applications to Anomaly Detection. 2017. To appear in Journal of Multivariate Analysis.

[4] N. Goix, A. Sabourin, and S. Clémençon. Learning the dependence structure of rare events: a non-asymptotic study. In *COLT*, 2015.

[5] N. Goix, A. Sabourin, and S. Clémençon. On Anomaly Ranking and Excess-Mass Curves. In *AISTATS*, 2015.

[6] N. Goix, A. Sabourin, and S. Clémençon. Sparse Representation of Multivariate Extremes. In *NIPS Workshop on Nonparametric Methods for Large Scale Representation Learning*, 2015.

[7] N. Goix, A. Sabourin, and S. Clémençon. Sparse Representation of Multivariate Extremes with Applications to Anomaly Ranking. In *AISTATS*, 2016.