Pierre-Edouard PORTIER, PhD

Home Address:
1 rue Marteret
69100 VILLEURBANNE
France

Email: pierre-edouard.portier@insa-lyon.fr ; pierre.edouard.portier@gmail.com
Homepage: https://peportier.me
Cell: +33 (0)6 77 17 54 54
DoB: 30th November 1984

One-page summary

Since 2011, Pierre-Edouard Portier is *associate professor at INSA Lyon*[1], France's leading post-baccalaureate engineering school, within the Computer Science & Information Technologies (CS&IT) department where he also studied and obtained an engineering degree. After his MSc, he obtained in 2010 a PhD degree in computer science from INSA Lyon.

Research topics: His research addresses topics in the fields of *machine learning* and soft computing techniques applied to big data analytics (e.g., anomaly detection, crash prediction, traffic forecasting…), *natural language processing* (e.g., relation extraction…), *semantic web search* and *document engineering*. His particular research interest lies in explainable machine learning and the discovery, modeling and representation of knowledge to be integrated into a data analytics process.

Management & community activities: He is currently *deputy director in charge of corporate relations for INSA Lyon CS&IT*. He is an *active member of IRIXYS*[2], an International Research and Innovation Center in Intelligent Digital Systems. He has also an activity in *academic consulting* (e.g., lead scientist in the context of a long-term collaboration with several companies). He has been a member of various program committees and reviewers for journals, conferences, and workshops such as Elsevier's Expert Systems with Applications, Wiley's Expert Systems, Springer's EPJ Data Science, the European Conference on Information Retrieval, ACM I-KNOW conference, IEEE International Conference on Web Intelligence, etc.

From September 2018 to September 2020, he has been *deputy director in charge of digital strategy at INSA Lyon*. Until September 2018, he managed the organization of seminars with industrial partners invited to share their expertise with INSA Lyon CS&IT students. From 2015 to 2019, he was responsible for the 4th year of studies at INSA Lyon CS&IT. During the academic year 2015-2016, he co-managed its evolution towards a more pronounced project-based learning pedagogy. From 2012 to 2018, he was an elected member of INSA Lyon CS&IT board. From 2012 to 2019, he was an elected representative for the LIRIS[3] laboratory board (2012-2019).

Lecturing: Pierre-Edouard Portier is teaching artificial intelligence, data analytics, agile software development methodologies and formal methods for software quality at INSA Lyon CS&IT. He supervised 7 PhD students in computer science at LIRIS laboratory.

Key publications:

- Pierre-Yves Genest, Pierre-Edouard Portier, Elöd Egyed-Zsigmond, Laurent-Walter Goix: **PromptORE – A Novel Approach Towards Fully Unsupervised Relation Extraction**. CIKM 2022.
- Thomas Véran, Pierre-Edouard Portier, François Fouquet: **Crash prediction for a French highway network with an XAI-informed Bayesian hierarchical model**. IEEE BigData 2020
- Yvan Lucas, Pierre-Edouard Portier, Lea Laporte, Sylvie Calabretto, Olivier Caelen, Liyun He-Guelton, Michael Granitzer: **Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs**. Future Generation Comp. Syst. 2020.
- Johannes Jurgovsky, Michael Granitzer, Konstantin Ziegler, Sylvie Calabretto, Pierre-Edouard Portier, et al.: **Sequence Classification for Credit-Card Fraud Detection**. Expert Systems with Applications, Elsevier, 2018
- Vincent Barrellon, Pierre-Edouard Portier, Sylvie Calabretto, Olivier Ferret: **Linear Extended Annotation Graphs**. ACM Symposium on Document Engineering 2017
- Vincent Barrellon, Pierre-Edouard Portier, Sylvie Calabretto, Olivier Ferret: **Schema-aware Extended Annotation Graphs**. ACM Symposium on Document Engineering 2016
- Mazen Alsarem, Pierre-Edouard Portier, Sylvie Calabretto, Harald Kosch: **Ranking Entities in the Age of Two Webs, an Application to Semantic Snippets**. ESWC 2015: 541-555
- Pierre-Edouard Portier, Sylvie Calabretto: **DINAH, A Philological Platform for the Construction of Multi-structured Documents**. ECDL 2010: 364-375

Patent:

Caelen, O., He-Guelton, L., Portier, P.E., Granitzer, M., Ziegler, K. and Jurgovsky, J., Worldline SA, 2020. **Machine learning system for various computer applications.** U.S. Patent Application 16/632,267.

[1] https://www.insa-lyon.fr/en/insa-lyon
[2] https://irixys.uni-passau.de/
[3] https://liris.cnrs.fr/en

Detailed information

Supervised PhD Thesis and main publications:

2012-2016 — Dr Mazen Alsarem: PhD Thesis co-supervised with Prof. Sylvie Calabretto. In the semantic web context, we introduced the concept of a *semantic snippet*. It enhances a search engine results page with information from the web of data. Links between the traditional web and the web of data are most often inferred from automatic annotation of entities in web pages. Some discovered entities may be wrong, or they may be irrelevant to the user's query. Thus, we developed algorithm LDRANK [1-3] to rank the entities discovered in a web page by their proximity to the user's query. It extends the PageRank algorithm to use external information besides the graph structure, viz. the rank given by a search engine for the webpage containing a given entity, and a textual description of the entity. Taking the latter into account uses an original variant of the singular value decomposition (SVD) algorithm. We obtain two probability vectors which represent the a priori importance of the entities according to two complementary points of view. We then use a belief aggregation operator to optimally combine the two vectors. In the end, we obtain a robust ranking of entities significantly better than the state of the art on a realistic dataset that we collected by a crowdsourcing process.

2014-2017 — Dr Melkamu Beyene: PhD Thesis co-supervised with Prof. Solomon Atnafu and Prof. Sylvie Calabretto. This work is a follow-up of the above thesis applied to dataset linking in a multilingual context [4-5]. We adapted the core ideas of the LDRANK algorithm to higher-order tensors.

2013-2017 — Dr Vincent Barrellon: PhD Thesis co-supervised with Prof. Sylvie Calabretto and Prof. Olivier Ferret. In the context of digital scholarly edition, we developed the annotation formalism eAG (Extended Annotation Graphs), and its schema language SeAG [6-7]. We worked with humanities researchers who build a digital scholarly edition of the *Encyclopédie* of Diderot and D'Alembert[4]. They must annotate documents according to an evolving schema. Moreover, XML documents, with their inherent tree structure, did not meet their needs. Therefore, we proposed an efficient algorithm to validate acyclic graph based on an expressive schema. We also designed a bi-directional algebra to propagate, semi-automatically, the effects on existing annotations of a schema modification. Finally, we defined a markup syntax, reminiscent of XML, to represent eAG documents.

2015-2019 — Dr Yvan Lucas: PhD Thesis co-supervised with Prof. Harald Kosch, Prof. Sylvie Calabretto, and Dr Léa Laporte. As a member of the international research center IRIXYS, I was part of a partnership with one of the major European companies monitoring credit card transactions in Europe. Two PhD thesis were focused on credit card fraud detection. For Yvan Lucas's thesis, we started by modeling the behavior of both fraudsters and normal cardholders. We discovered that the dataset shift phenomenon can be explained by calendar events, sometimes predictable [11]. Then, we conceived an algorithm to integrate into a flexible machine learning model (e.g., random forest), representations of sequential behaviors learned by probabilistic generative models (e.g., hidden Markov model). Our approach outperforms the state of the art due to its ability to model both genuine and fraudulent behaviors [10,12].

2015-2019 — Dr Johannes Jurgovsky: PhD Thesis co-supervised with Prof. Michael Granitzer, and Prof. Sylvie Calabretto. This thesis was also funded by the project mentioned above. First, we proposed to integrate embeddings of knowledge graphs to a machine learning model to improve credit card fraud detection thanks to a richer representation of geographical information linked to cardholders and payment terminals [9]. Then, we are at the origin of a reference study on the use of LSTM neural networks to represent cardholders' behaviors and improve fraud detection [8]. This work resulted in the filing of a patent [15].

2019-2022 — Dr Thomas Véran: PhD Thesis co-supervised with Prof. Jean-Marc Petit. This thesis was funded in collaboration with an industrial partner: Data New Road[5], a joint venture between IT company Amiltone and APRR[6], a subsidiary of Eiffage, one of the largest European motorway managers. We developed a Bayesian hierarchical model to discover risk factors linked to highway accidents [13]. This work is a contribution to the domain of eXplainable Machine Learning (XAI). Its originality lies in two aspects. First, the automatic discovery of a hierarchical structure by leveraging the features' importance of a black box model (e.g., XGBM) given by their Shapley values. Second, the automatic discovery of significant features' interactions by an auto-adaptive polynomial neural network. Moreover, in a recent work, currently under review, we achieve an even better balance between predictive performance and interpretability by using an innovative symbolic regression model.

2021-… — Pierre-Yves Genest: PhD Thesis co-supervised with Dr Elöd Egyed-Zsigmond. This thesis is funded in the context of a collaboration with Alteca[7], a French IT company. We are working on unsupervised relation extraction for natural language processing (NLP). We developed a "Prompt-based Open Relation Extraction" model (PromptORE) [14] by adapting the prompt-tuning paradigm to work in an unsupervised setting and using it to create embeddings for sentences expressing a relation. We then cluster these embeddings to discover candidate relations. PromptORE is the first unsupervised relation extraction model that doesn't need any hyperparameter tuning and consistently outperforms state-of-the-art models.

[4] http://enccre.academie-sciences.fr/encyclopedie/
[5] https://www.data-newroad.com/?lang=en
[6] https://aprr.com/en
[7] https://www.alteca.fr/

Academic consulting and scientific expertise:

Reviewer: for international journals, conferences, and workshops, such as: Elsevier's Expert Systems with Applications, Wiley's Expert Systems, Wiley's Concurrency and Computation, Springer's EPJ Data Science, MDPI's Sensor Networks, MDPI's Applied Sciences, IEEE Transactions on Neural Networks and Learning Systems, the European Conference on Information Retrieval, ACM I-KNOW conference, IEEE International Conference on Web Intelligence, …

Academic consulting: I actively participate in DataValor[8], a structure that we have set up to facilitate the transfer of academic innovations to industrial partners. For example, I took part in setting up a data science training curriculum for a large French bank. With this same partner, I did several consulting missions to help them with machine learning and data science challenges. I also assisted them in recruiting data scientists. Moreover, I was lead scientist in a multi-year contract for one of the largest European motorway managers through the Data New Road joint venture. We developed long-term and short-term traffic prediction models by adapting state-of-the-art algorithms.

Teaching:

Advanced Algorithm for AI: A module for M1 students at INSA Lyon CS&IT department, in which I introduce (1) algorithms for solving combinatorial problems formulated as a graph search (A*, IDA*, etc.), (2) algorithms to find configurations in a state space graph (e.g., simulated annealing, tabu search), (3) the PageRank algorithm, (4) Gödel incompleteness theorems, (5) some aspects of combinatorial game theory, …

Formal approach for building software: A module for M2 students at INSA Lyon CS&IT department, in which I introduce how to write programs correct by construction by transforming a formal specification into a program. We follow the approach promoted by Dijkstra and based on weakest precondition predicates.

Object oriented conception and Agile development: A module for M2 students at INSA Lyon CS&IT department, in which, with a colleague researcher, we have students develop in groups of 6 an application for building optimal city-wide delivery rounds.

Introduction to Machine Learning: A module for M2 students at INSA Lyon CS&IT department, in which I introduce the fundamental concepts of supervised machine learning (e.g., regularization, bias, variance…).

Indexing large volume of textual data: A module for M2 students at INSA Lyon CS&IT department, where I introduce the main algorithms for indexing and querying large amount of textual data.

Collective and general interest responsibilities:

2020-… — Deputy Director in charge of Corporate Relation at INSA Lyon CS&IT department: I initiated the creation of a club of industrial partners to consolidate fruitful relations with a restricted group of companies recognized for their excellence and their values (e.g., Avanade, BioMérieux, CGI, Crédit Agricole, Deloitte, L'Oréal, Société Générale, Sopra Steria, VMware, Onepoint…).

2020-… — Deputy Director CISR (Network Services Inter-Universities Center): This center is a central platform for all university projects linked to network and infrastructures for data storage and computation in the Lyon (France) region. In this context, I am associated with the management of a large data center construction project.

2016-… — Active member of IRIXYS: International Research & Innovation Center on Digital Intelligent Systems.

2018-2020 — Deputy Director in charge of Digital at INSA Lyon: I have initiated several projects of digital transformation (e.g., integration of an electronic document management and collaborative work platform, positioning of INSA Lyon as a pilot site for a new national school management software, support for start-ups resulting from student initiatives…). I had also a role in the management of the COVID crisis to ensure continuity of service and help students and staff who were in difficulty due to poor access to digital resources.

2012-2018 — Elected member of INSA Lyon CS&IT department board

2012-2019 — Elected member of LIRIS laboratory board

[8] https://www.insavalor.fr/datavalor

References:

[1] Alsarem, M., Portier, P.E., Calabretto, S. and Kosch, H., 2015, March. Ordonnancement d'entités appliqué à la construction de snippets sémantiques. In *CORIA 2015: COnférence en Recherche d'Information et Applications* (pp. 253-268).

[2] Alsarem, M., Portier, P.E., Calabretto, S. and Kosch, H., 2015. Ordonnancement d'entités pour la rencontre du web des documents et du web des données. *Document numerique*, *18*(2), pp.123-154.

[3] Alsarem, M., Portier, P.E., Calabretto, S. and Kosch, H., 2015, May. Ranking entities in the age of two webs, an application to semantic snippets. In *European Semantic Web Conference* (pp. 541-555). Springer, Cham.

[4] Beyene, M., Portier, P.E., Atnafu, S. and Calabretto, S., 2016, November. Dataset linking in a multilingual linked open data context. In *Proceedings of the 8th International Conference on Management of Digital EcoSystems* (pp. 149-157).

[5] Beyene, M., Portier, P.E., Atnafu, S. and Calabretto, S., 2015, October. Tensor factorization for cross lingual entity coreference resolution in the linked open data. In *Proceedings of the 7th International Conference on Management of computational and collective intElligence in Digital EcoSystems* (pp. 17-23).

[6] Barrellon, V., Portier, P.E., Calabretto, S. and Ferret, O., 2016, September. Schema-aware extended annotation graphs. In *Proceedings of the 2016 ACM Symposium on Document Engineering* (pp. 45-54).

[7] Barrellon, V., Portier, P.E., Calabretto, S. and Ferret, O., 2017, August. Linear extended annotation graphs. In *Proceedings of the 2017 ACM Symposium on Document Engineering* (pp. 9-18).

[8] Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.E., He-Guelton, L. and Caelen, O., 2018. Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, *100*, pp.234-245.

[9] Ziegler, K., Caelen, O., Garchery, M., Granitzer, M., He-Guelton, L., Jurgovsky, J., Portier, P.E. and Zwicklbauer, S., 2017, June. Injecting semantic background knowledge into neural networks using graph embeddings. In *2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)* (pp. 200-205). IEEE.

[10] Lucas, Y., Portier, P.E., Laporte, L., Calabretto, S., Caelen, O., He-Guelton, L. and Granitzer, M., 2019, April. Multiple perspectives HMM-based feature engineering for credit card fraud detection. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing* (pp. 1359-1361).

[11] Lucas, Y., Portier, P.E., Laporte, L., Calabretto, S., He-Guelton, L., Oblé, F. and Granitzer, M., 2019, June. Dataset shift quantification for credit card fraud detection. In *2019 IEEE second international conference on artificial intelligence and knowledge engineering (AIKE)* (pp. 97-100). IEEE.

[12] Lucas, Y., Portier, P.E., Laporte, L., He-Guelton, L., Caelen, O., Granitzer, M. and Calabretto, S., 2020. Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs. *Future Generation Computer Systems*, *102*, pp.393-402.

[13] Veran, T., Portier, P.E. and Fouquet, F., 2020, December. Crash prediction for a French highway network with an XAI-informed Bayesian hierarchical model. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 1256-1265). IEEE.

[14] Genest, P.Y., Portier, P.E., Egyed-Zsigmond, E and Goix, L.W., 2022, October. PromptORE — A novel approach towards fully unsupervised relation extraction. *CIKM 2022*. To be published.

[15] Caelen, O., He-Guelton, L., Portier, P.E., Granitzer, M., Ziegler, K. and Jurgovsky, J., Worldline SA, 2020. Machine learning system for various computer applications. U.S. Patent Application 16/632,267.