DOCTORAL PRIZES 2020

Abstacts

HEC Lausanne/November 2020

An Information-Theoretic Perspective on Trustworthy Machine Learning

Prix de la Fondation Nicolas et Hélène Porphyrogenis

Dr. **Natasa Tagasovska**, PhD in Information System. Thesis supervised by Prof. Valérie Chavez and Prof. Periklis Andritsos.

We live in an age where we witness how machines learned to talk, beat us at video games, dream, paint and advance making scientific discoveries. However, having little or no underlying principle that explains the working of these learning systems, leads to a lack of confidence in their efficiency. An inevitable consequence of the lack of trust and certainty, is limited productionalization of machine learning. This notably holds for sensitive domain applications, such as precision medicine, engineering or self-driving vehicles to name a few.

This thesis is an attempt to make a contribution towards more understandable and trustworthy machine learning models, humans' pivotal asset for achieving (general) artificial intelligence. A recurring theme throughout the text will be the information-theoretic perspective of learning systems. In particular, we are interested in compression techniques, which have their interpretation as a "measure of intelligence". Such interpretation is motivated by the fact that "being able to compress well, is closely related to - acting intelligently". The intuition is that shedding redundancy from the data leads to meaningful summarisations and pattern discovery i.e generalization, one way of recognising intelligent behaviour. This thesis is a result of combining four papers which all relate to information compression. We start by presenting an efficient approach for extracting relevant information from large datasets which enables human to consume (parse) big data. In particular, the focus in chapter 2 is on distributed clustering recast as information compression. We then continue in chapter 3, towards information theoretic causal discovery, where an optimal code length description is leveraged in distinguishing causes from effects in observational data. Although not yet entirely established, an information-theoretic interpretation is been posed as theoretical grounding and explanation of deep learning models. This brings us to chapter 4, where we propose a generative hybrid model that exhibits explorative and flexible properties at the same time. Our generative model benefits from the nonparametric modelling of copulas on top of a compressed version of the training data (i.e. embedded features of an autoencoder), that makes up for an efficient, easy add-on technique, which altogether paves its way to myriad of applications. In the last chapter of the thesis, we present simple, scalable, single-model uncertainty estimates constructed by benefiting from the most informative feature representations (outputs of the last layer) of a deep model.

In conclusion, by tackling causal, generative and uncertainty-aware machine learning methods, we hope to increase the integrity and aid the adaptation of such models in any domain of application.

Three Essays in Dynamic Corporate Finance

Prix de la Banque Cantonale Vaudoise

Dr. **Jakub Hajda**, PhD in Economics, subject area Finance. Thesis supervised by Prof. Boris Nikolov.

Recent advances in computational power and the availability of vast volumes of high quality data allow us to venture into novel and previously unexplored areas of corporate finance. In my thesis, I combine a data-driven approach with theoretical models of firms to better understand how firms make financing and investment decisions and what the economy-wide implications of these decisions are.

I argue that the composition of cash flow variability is a vital factor influencing firms' capital structure decisions. I demonstrate that distinguishing between cash flow variability affecting long-run prospects of firms (e.g. reflecting changes to technology) and that subsiding over time (e.g. driven by a natural disaster) is crucial. As the effects of persistent shocks are long-lasting, they incentivize firms to be financially conservative.

Firms' revenues consist of product revenue. One source of cash flow variability is thus related to product life cycle, which implies a negative relationship between product revenue and age. I document that product life cycle is an important force shaping the evolution of firms' cash flows and that accounting for this mechanism helps understand corporate valuation as well as investment and financing decisions.

Debt financing by firms has implications for innovation as well. In particular, I argue that debt fosters innovation and growth in the economy. This is the result of two opposing forces. First, debt hampers innovation by incumbent firms, as indebted firms invest less than they would in absence of debt due to agency frictions. Second, debt incentivizes entry as it increases the value of the industry, which intensifies competition and stimulates innovation by entering firms.

Statistics of Extremes, Matrix Distributions and Applications in Non-Life Insurance Modeling

Prix de la Banque Cantonale Vaudoise

Dr. **Martin Bladt**, PhD in Actuarial Science. Thesis supervised by Prof. Hansjoerg Albrecher.

The thesis deals with the mathematical modeling and statistical estimation of risks. The thesis first develops new variants of probabilistic models for the surplus of an insurance company. Through exact solutions as well as simulation, important aspects of the process are studied, such as the ruin probability, or the expected discounted dividends until ruin in a modified setting with a ratcheting strategy. The second theme of the thesis concerns the estimation of heavy-tailed risks. The main ideas involve combining expert information with statistical practice, and removing low-importance data-points from a sample in order to gain statistical stability. The methods are also extended to incorporate right-censored observations, as for some insurance claim data-points only a lower bound is available. The third and final theme concerns the specification of new distributions arising from the generalization of already established ones when replacing the parameter with a matrix. The resulting classes inherit many of the properties from the underlying building blocks, while at the same time gaining large flexibility to model any kind of positive data.

Managing Advanced Synchronization Aspects in Logistics Systems

Prix de Faculté

Dr. **Marc-Antoine Coindreau**, PhD in Business Analytics. Thesis supervised by Prof. Olivier Gallay.

Logistics (or supply chain management) refer to the process of moving goods, materials, and people within a company. The optimization of such processes enables firms to gain competitiveness, and constant innovative methodological approaches must be unveiled to sustain this potential for improvement. Currently, to be optimized, the supply chain is divided into several echelons (or levels) (e.g., production, distribution, storage, etc.), each of them being optimized individually and independently. As a result, the obtained solutions for one echelon become then the input data for the next echelon, which leads globally to sub-optimality. In that regard, synchronization (i.e., aiming at a harmonious collaboration of various actors at different levels of the supply chain) appears to be a promising avenue for improvement. In this thesis, three practical cases, inspired by different industrial partners, are analyzed. The considered situations range from the transportation of raw materials from suppliers to production plants to home parcel delivery. In each of these cases, we quantify the potential offered by adopting a methodological approach based on synchronization and we compare the obtained solutions to current industrial practices. Results show that aspects related to costs, labor conditions, or even environmental impacts of the supply chain can be significantly improved by obtaining and introducing synchronized solutions in the field of logistics. In this respect, this thesis focuses also on building the cutting-edge algorithms that are necessary to solve such resulting problems and to obtain such synchronized solutions.

Three Essays in Political Economics

Prix de Faculté

Dr. **Ulrich J. Eberle**, PhD in Economics, subject area Political Economy. Thesis supervised by Prof. Dominic Rohner.

The ambition of this thesis is to quantitatively assess research questions of high policy-relevance relating to conflict. The results presented in all three chapters stem from analyses making use of cutting-edge econometric and spatial methods which - in conjunction - offer a powerful tool set for causal inference.
The first thesis chapter investigates the impact of newly constructed dams on local conflict around the world. Large-scale infrastructure projects are a frequent source of local protest, which can escalate into violent combats. The analysis reveals that once new dams are built, they disrupt the local economy and increase the likelihood of conflict in the regions surrounding dams. Long-standing ethnic grievances and low political competition appear to be closely related to the causes of conflict.
The second chapter, joint with Dominic Rohner and Mathias Thoenig studies conflict between settlers and nomads in Africa. The paper finds that in years of drought, resource scarcity forces nomadic herders to migrate towards the fringes of deserts where they compete with settled farmers over fertile land, resulting in violent conflict. Institutional features such as land dispute resolution mechanisms and secure property rights appear to be effective towards preventing these issues.
The thematic sphere of the third chapter of my thesis focuses on the link between ethnic diversity, social tensions and urbanization. This work is joint with Vernon Henderson, Dominic Rohner and

Kurt Schmidheiny. The global analysis at the province level finds that increased ethno-linguistic fractionalization and polarization are associated with lower urbanization and an increased role for secondary cities relative to the primate city of a province. Policy makers should spend special attention on intermediately democratized countries, as those nations appear to be most prone to experience the adverse effects of ethnic diversity on urbanization.

More details: http://ulricheberle.com/

Contributions to Establishing Causality in Management Studies: Three Essays on Experimental Methods, Latent Moderated Models and the Origins of Organizational Leadership

Prix de Faculté

Dr. **Sirio Lonati**, PhD in Economics, subject area Management. Thesis supervised by Prof. John Antonakis and Prof. Christian Zehnder.

Over the past years, the management field has been challenged to raise its "methodological game", especially concerning the estimation of causal relationships. My doctoral dissertation tackles some of the empirical and theoretical difficulties linked to establishing causality credibly in our discipline. The first chapter studies how to ensure causally interpretable experimental effects, with a specific attention to organizational research. The second essay focuses on a workhorse for management studies, namely, moderated models. Specifically, it derives a novel test for identifying misspecifications due to violations of distributional assumptions in latent interaction models, further probing this new statistical procedure in a large simulation study. In the last chapter of the dissertation, I turn to field data and explore empirically the ultimate causes of participative leadership style in organizations, employing an unusual theoretical prism inspired by evolutionary-informed social sciences.