PhD position with CIFRE contract: Fraud detection on credit card payments

KEY INFORMATIONS

CIFRE contract for a PhD position within the context of the Lusis - CentraleSupelec AI chair partnership
University: Université Paris-Saclay
Doctoral School: Information and Communication Sciences and Technologies
Domain: Computer Science
Research Unit: LRI - Laboratoire de Recherche en Informatique
Team: AO/LAHDAK
PhD Thesis advisor: Bich-Liên Doan
Co-advisor: Fabrice Popineau
Contacts for application: Fabrice Popineau (Fabrice.Popineau@lri.fr) - Bich-Liên Doan (Bich-Lien.Doan@lri.fr) - Arpad Rimmel (Arpad.Rimmel@lri.fr)
Starting date: september 2020
Deadline for application: <no deadline>

Keywords

Fraud detection, machine learning, explainability

Profile and skills required

The applicant should demonstrate good programming skills and knowledge of toolkits dedicated to machine learning such as scikit-learn and tensorflow.

Context

Lusis is the publisher of TANGO, a high-performance transactional platform for payment systems and market finance. Based on this platform, Lusis builds complete payment systems, including fraud detection, as well as extremely rich and complex front-to-back trading platforms on which more than 5 billion dollars a day are currently traded, half of which are in Forex, Indices and Commodities. Lusis and CentraleSupélec have joined forces by creating a research chair to strengthen their collaboration in the field of artificial intelligence applied to banking.

Detailed presentation of the doctoral project

Today, electronic payments by credit card have become widespread on the planet thanks to the Internet. These payments represent an ever-increasing volume of more than US$ 500 billion annually, and fraud accounts for less than 0.5% of the number of these transactions. Globally, this represents an estimated US$ 25 billion in losses per year.

Irrespective of the amount of this fraud, all possible detection measures must be implemented to limit its spread. This is why a great deal of research has been carried out on this issue over the last 20 years.

This is of course of strategic interest to individuals, banks and companies. As the publisher of a high-performance transactional platform for payment systems, LUSIS is therefore primarily interested in fraud countermeasures. Within the framework of the LUSIS chair, we want to study fraud detection both from the point of view of algorithm performance, but also with a constraint of realism of implementation and this on real data.

Credit card fraud detection has been studied for many years, but the lack of labelled real data makes the task difficult, both for learning and for evaluation (Parthasarathy et al., 2019).

Operationally, false positives should be avoided as much as possible: customers do not like to have their payments wrongly refused. For reasons of process control, systems deployed today are rules-based. As a result, they are cumbersome to process and especially to update. Systems based on machine learning are essentially under study at the moment.

In recent years, machine learning and deep learning algorithms have proven their effectiveness in many areas. The main machine learning methods have all been tested on this problem: decision trees, random forests, Bayesian networks, Hidden Markov Models, neural networks and deep neural networks (Ryman-Tubb et al., 2018). The specific difficulty with fraud lies in a very large imbalance between the "fraud" and "non-fraud" classes (less than 0.5% fraud) and the essentially online nature of the process, which is not taken into account in offline studies (Carcillo et al., 2018).

(Zojaji et al., 2016) classifies techniques into two main approaches to fraud detection, namely abuse (supervised) and anomaly detection (unsupervised): a classification of techniques is proposed based on the ability to handle both numerical and categorical datasets. (Kamaruddin & Ravi, 2016) used a single-class classification approach by developing a hybrid architecture composed of a swarm optimization algorithm and a self-associative neural network. Graph approaches as described by (Zhang et al., 2019) or (Belle et al., 2019) are very recent and promising.

As new fraud strategies emerge, the paradigm of anomaly detection in consumer payment habits is becoming increasingly important. In this context, it is necessary to take into account the

fact that these habits evolve over time: this is what we call *concept drift* for a consumer. The FraudMemory model developed by (Yang & Xu, 2019) captures the sequential patterns associated with each transaction and exploits memory networks to improve both performance and interpretability while taking into account conceptual drift.

Finally, one last point remains relatively unexplored in the work on payment fraud detection: explicability. This point becomes crucial with the GDPR law, since any consumer for whow a valid payment has been refused will be able to ask the provider for an explanation.

In conclusion, despite the very good results obtained by techniques such as random forests on collected data sets, the problem of fraud detection is not solved and many avenues remain to be explored to achieve a system that is robust over time and explainable.

Thanks to our partnership with the company LUSIS, we have access to real data sets and a quality field of experimentation with business experts. The PhD thesis project that we propose will focus on four research orientations:

1. attributed networks,
2. anomaly detection,
3. conceptual drift,
4. explainability.

In a first phase of bibliographical study, the student will have to synthesize the state of the art on the existing methods and their advantages and disadvantages from a practical point of view. Then, it will be a question of proposing an architecture, possibly hybrid, meeting the criteria of performance, robustness and explainability.

Bibliography

Belle, R. V., Mitrovic, S., & Weerdt, J. D. (2019). *Graph Representation Learning for Fraud Prediction : A Nearest Neighbour Approach*. 6.

Carcillo, F., Le Borgne, Y.-A., Caelen, O., & Bontempi, G. (2018). Streaming active learning strategies for real-life credit card fraud detection : Assessment and visualization. *International Journal of Data Science and Analytics*, *5*(4), 285‑300. https://doi.org/10.1007/s41060-018-0116-z

Kamaruddin, Sk., & Ravi, V. (2016). Credit Card Fraud Detection using Big Data Analytics : Use of PSOAANN based One-Class Classification. *Proceedings of the International Conference on Informatics and Analytics - ICIA-16*, 1‑8. https://doi.org/10.1145/2980258.2980319

Parthasarathy, G., Ramanathan, L., JustinDhas, Y., Saravanakumar, J., & Darwin, J. (2019). Comparative Case Study of Machine Learning Classification Techniques Using Imbalanced Credit Card Fraud Datasets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3351584

Ryman-Tubb, N. F., Krause, P., & Garn, W. (2018). How Artificial Intelligence and machine learning research impacts payment card fraud detection : A survey and industry

benchmark. *Engineering Applications of Artificial Intelligence*, *76*, 130‑157. https://doi.org/10.1016/j.engappai.2018.07.008

Yang, K., & Xu, W. (2019, janvier 8). *FraudMemory : Explainable Memory-Enhanced Sequential Neural Networks for Financial Fraud Detection*. https://doi.org/10.24251/HICSS.2019.126

Zhang, Z., Cui, P., & Zhu, W. (2019). Deep Learning on Graphs : A Survey. *arXiv:1812.04202 [cs, stat]*. http://arxiv.org/abs/1812.04202

Zojaji, Z., Atani, R. E., & Monadjemi, A. H. (2016). *A Survey of Credit Card Fraud Detection Techniques : Data and Technique Oriented Perspective*. 26.