PhD studentship: Explainable models for detecting anomalies in data streams coupled with textual information

Supervision: LIFO/PRISME (University of Orléans, Orléans, France)

The University of Orléans, France, is offering one fully funded PhD studentship. The position is for 3 years with a monthly net salary of EUR 1,412. **This thesis** is **supported** by the **ANR** (National Research Agency) and the **University of Orléans** in the context of the **project AI.iO** (Artificial Intelligence in Orléans: Learning from heterogeneous data and expert knowledge. Applications in geological and environmental sciences). The thesis will be supervised by researchers from two laboratories of the University of Orléans: LIFO (*Laboratoire d'Informatique Fondamentale d'Orléans*) and PRISME (*Laboratoire Pluridisciplinaire de Recherche en Ingénierie des Systèmes, Mécanique, Énergétique*). These laboratories collaborate via their complementarities in signal/image processing, machine learning and knowledge representation.

We are looking for a candidate who has recently received a M.Sc. degree in a relevant field such as Computer Science, or similar. Particular strength or experience in machine learning/data mining, natural language processing, computer programming or applied mathematics is highly appreciated. French and/or English are the working languages.

The **deadline for submitting the application** is **August 24th 2020**, but the candidates are encouraged to contact us as soon as possible. **Start is expected around October 2020**. Your complete application consists of the documents below, which should be sent as a single PDF file to Marcílio de Souto (marcilio.desouto@univ-orleans.fr) and Adel Hafiane (adel.hafiane@univ-orleans.fr).

- CV with photo
- One-page cover letter (clearly indicating available start date as well as relevant qualifications, experience and motivation)
- University certificates and transcripts/marks (both B.Sc and M.Sc degrees)
- Contact details of up to three referees
- Possibly an English language certificate and a list of publications
- **Attention:** all documents should be in English or French.

Short description of the thesis:

Artificial Intelligence (AI)/Machine Learning algorithms have been shown to be powerful in terms of results and predictions, however the models generated are generally opaque or lack transparency [CPC19]. In other words, most of the models produced are complex black boxes: their internal logic and functioning are hidden from the user and even from experts. This prevents a human, expert or non-expert, from verifying, interpreting and understanding the reasoning of the system and the way in which decisions are taken.

In this thesis, we seek to address heterogeneous data coupling data streams (e.g., sensor data) and textual information (mainly tweets). The idea is to study how the two types of data

can be combined to detect and explain anomalies. Anomaly detection is the problem of identifying observations that do not conform to the expected behavior. They correspond often to critical actionable information in many real-world applications, including condition monitoring, fault prevention and fraud detection.

The coupling can be done, for example, as follows:

1. Detection of an anomaly on the data streams and looking for explanations in both a) the "rules" extracted from the learned model and b) the textual information present in the tweets.
2. Detecting an event/anomaly on tweets (for example, increase in the number of tweets on a given subject, detection of an evolving situation, ...) and its respective link on the data streams.

The detection of anomalies in data streams is a subject widely studied [BZA2020, GRBBG2019, SR2018]. In this context, the challenges posed in this thesis lie in the fact that differently from the works such as [Bertini2020, SSDRB2018] and [BBLC2018] which only approached the problem from the point of view of, respectively, the extraction/generation of rules and textual information from twitter, here both types of information will be used to generate the explanation.

With respect to (2), one of the main tasks is to be carried out in order to relate the target anomalies from an ontological point of view to their "occurrences" in the data streams generated, for example, by sensors as well as in the linguistic productions identifiable via twitter [YDP2019]. Regarding the analysis of tweets, one of the challenges of this thesis is to go beyond the keyword approach and focus on linguistic signals qualified as weak. Several possibilities can be pursued in Natural Language Processing to refine the analysis of weak signals such as the use of word embedding in order to capture any lexical signal close to a cause previously identified, or even adaptation of coreference resolution [DLTLA2018] methods in order to tackle event entities and speech-making through a set of tweets.

With respect to (1), we will first analyze this problem from the point of view of model-agnostic explanation methods (model-independent). These methods are based on the approximation of a black box type of model by an easier to understand white box model (for example, a decision tree model) [Bie18, JLLB18, Lip18, ZWM + 19, LL17 , RSG16]. Although model-agnostic methods are effective in extracting knowledge from different machine learning models (for example, neural networks), it is also important to consider the explanatory methods developed for specific type of models. One of the reasons for this is that, for example, neural networks learn features and concepts in their hidden layers and we need special tools to "extract" them. Finally, with regard to (1) the knowledge extracted from the anomaly detection models generated with the learning techniques will have to be integrated with that extracted from the textual data (e.g., tweets) in order to have a more robust system as well as more comprehensible for the final user.

References

[Bie18] Przemyslaw Biecek. 2018. Dalex: Explainers for complex predictive models in r. J. Mach. Learn. Res., 19(1) :3245–3249, January 2018**.**

[BBLC2018] B. Belabbess, M. Bairat, J. Lhez , O. Curé O. 2018. Combining Machine Learning and Semantics for Anomaly Detection. In: Faron Zucker C., Ghidini C., Napoli A., Toussaint Y. (eds) Knowledge Engineering and Knowledge Management. EKAW 2018. Lecture Notes in Computer Science, vol 11313. Springer.

[Bertini2020] João Roberto Bertini. 2020. Graph embedded rules for explainable predictions in data streams, Neural Networks, Volume 129, 2020, Pages 174-192, ISSN 0893-6080.

[BZA2020] Azzedine Boukerche, Lining Zheng, and Omar Alfandi. 2020. Outlier Detection: Methods, Models, and Classification. ACM Computing Surveys 53, 3, Article 55 (June 2020), 37 pages. DOI:https://doi.org/10.1145/3381028

[CPC19] D.V. Carvalho, E.M. Pereira, and J.S. Cardoso. 2019. Machine learning interpretability: A survey on methods and metrics. Electronics, 8(8) :832.

[DLTLA2018] A. Désoyer, F. Landragin, I. Tellier, A. Lefeuvre-Halftermeyer, J.-Y. Antoine, et al. 2018. Coreference Resolution for French Oral Data: Machine Learning Experiments with ANCOR. *Computational Linguistics and Intelligent Text Processing.*, n° 9623-9624, Springer, 2018, Lecture Notes in Computer Science. ⟨hal-01889593⟩

[GRBBG2019] Heitor Murilo Gomes, Jesse Read, Albert Bifet, Jean Paul Barddal, and João Gama. 2019. Machine learning for streaming data: state of the art, challenges, and opportunities. ACM SIGKDD Explorations Newsletter 21, 2 (December 2019), 6–22. DOI:https://doi.org/10.1145/3373464.3373470

[LdSdO09] Teresa B. Ludermir, Marcílio C. P. de Souto, and Wilson R. de Oliveira. 2009. On a hybrid weightless neural system. International Journal of Bio-Inspired Computation, 1(1/2) :93–104.

[Lip18] Zachary C. Lipton. The mythos of model interpretability. 2018. Commun. ACM, 61(10) :36–43.

[LL17] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. 2017. In Annual Conference on Neural Information Processing Systems, Long Beach, CA, USA, pages 4765–4774.

[RSG16] Marco Túlio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "why should I trust you ?" : Explaining the predictions of any classifier. CoRR, abs/1602.04938.

[SR2018] Mahsa Salehi and Lida Rashidi. 2018. A Survey on Anomaly detection in Evolving Data: [with Application to Forest Fire Risk Prediction]. ACM SIGKDD Explorations Newsletter 20, 1 (June 2018), 13–23. DOI:https://doi.org/10.1145/3229329.3229332

[SSDRB2018] Fei Song, Arnaud Stiegler, Yanlei Diao, Jesse Read, Albert Bifet. 2018. EXAD: A System for Explainable Anomaly Detection on Big Data Traces. ICDMW 2018 - IEEE International Conference on Data Mining Workshops, Singapore, Singapore

[YDP2019] Yousefinaghani, S., Dara, R., Poljak, Z. et al. 2019. The Assessment of Twitter's Potential for Outbreak Detection: Avian Influenza Case Study. Sci Rep 9, 18147. https://doi.org/10.1038/s41598-019-54388-4

Contact details

For further information about the studentship, please, contact Marcílio de Souto (marcilio.desouto@univ-orleans.fr) and Adel Hafiane (adel.hafiane@univ-orleans.fr).

For more information about the research teams at the LIFO and of PRISME concerned by the thesis, see:

http://www.univ-orleans.fr/lifo/equipe.php?lang=en&id=1
https://www.univ-orleans.fr/fr/prisme/la-recherche/axe-image-vision

For more information on the University see http://www.univ-orleans.fr/. Orléans, France, is one hour by train from Paris and not far from the famous castles of the Loire River valley.