INTERNSHIP PROPOSAL

Master 2 or Engineer (COMPUTER SCIENCE or BIONFORMATICS)

Title: Extraction of rules from a neural network applied to the healthcare
Keywords: artificial neural networks, deep learning, rule extraction, electronic patient records.
Supervisor: Farida Zehraoui, Lecturer, AROB@S Team, IBISC laboratory, Paris-Saclay University, Univ. Evry

Contact: farida.zehraoui@univ-evry.fr
Duration of internship: 6 months
Location: IBISC Laboratory, IBGBI, University of Evry, 23 Boulevard de France, 91000 Evry.

Description:
Deep learning is a class of machine learning methods that is able to model data with different levels of abstraction. These methods are mainly based on artificial neural networks. They have enabled significant progress in several areas such as object recognition, signal analysis, automated natural language processing, etc.
Despite their predictive power, deep neural networks are considered black boxes, which makes their interpretation difficult, especially in sensitive areas such as healthcare.
Recently, there has been a growing interest in explaining these models [1], particularly in extracting rules for interpretation. Researchers have proposed various methods aimed at extracting symbolic decision rules from neural networks [2][3][4][5].
The objective of the project is to study the different existing methods and to propose a new method for extracting rules from a deep neural network. The proposed approach will be applied to real data extracted from electronic records of patients admitted to intensive care units to predict "Sepsis" [6].

Project steps:
- Study of different approaches to extracting rules from neural networks.
- Proposition and implementation of a new rule extraction method from a neural network that satisfy some desirable properties.
- Application of the implemented method to real data from patients admitted to intensive care units to early prediction of "Sepsis".

Bibliography

1. R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM Comput. Surv.*, 51(5):93:1–93:42, Aug. 2018. ISSN 0360-0300.
2. R. Andrews, J. Diederich, and A. B. Tickle. Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge-Based Systems*, 8(6):373 – 389, 1995. ISSN 0950-7051. Knowledge- based neural networks.
3. G. Bologna and Y. Hayashi. A rule extraction study on a neural network trained by deep learning. In *2016 International Joint Conference on Neural Networks, IJCNN 2016, Vancouver, BC, Canada, July 24-29, 2016*, pages 668–675. IEEE, 2016.

4. A. Bondarenko and L. Aleksejeva. Methodology for knowledge extraction from trained artificial neural net- works. *Information Technology & Management Science (RTU Publishing House)*, 21, 2018.
5. M. T. Ribeiro, S. Singh, and C. Guestrin. Anchors: High-precision model-agnostic explanations. In *Proceed- ings of the AAAI conference on artificial intelligence*, volume 32, 2018.
6. https://physionet.org/content/challenge-2019/1.0.0/