Sony R&D Brussels Laboratory Internship Proposals

About Sony R&D Center Brussels Laboratory

Sony R&D BRL is a software and Artificial Intelligence research & development group, focusing its activities on AI, security and new technologies such as human sensing and blockchain. It is wholly owned by Sony Corporation and exclusively researches and develops technology for Sony products and services.

Sony R&D BRL's offices are located in Zaventem, near Brussels, Belgium. It takes 45 minutes to reach by train from Brussels Midi (Eurostar / Thalys terminal) and 5 minutes by shuttle bus from Brussels Airport.

The group is international and the working language is English.

In order to build up collaboration with academia, we would like to offer some internship proposals to Master students from local universities. Due to the current Covid19 restrictions most of the internship will be done online, however the candidate should be located in Belgium.

Anomaly Detection Platform

Annona is a flexible machine learning (ML) platform for anomaly detection (AD) that supports the full ML cycle from data exploration to model building and execution. Historically the main area of interest has been mostly fraud detection but is being extended to new ML problems.

Internship Proposal: Auto Feature Engineering Compiler Improvement

At the heart of Annona is the automated feature engineering component capable of generating a huge number of features from which the most effective are selected automatically. We are always looking to improve the performance of this compiler. The goal is implementing and evaluating a number of potential improvements. Optimization of DSL generation by identification of invariant functions for a model type is part of this.

- Identify nodes in the feature graph that end up with same results (feature values are the same) by taking different paths.
- Improve pre-feature selection. For example in the current pre-feature selection log(x) is preferred to x, although both features should give the same model result in the end.

- Analyze compiled features to identify patterns. Maybe there are features that are always discarded or kept in the pre-selection step.
- How many times should time window function be applied until the result becomes meaningless? In other words, determine max hierarchy level for a windowed functions
- Update compiler to limit file read. Read a file and compile all dependent features.
- What features are the best for a use with XGBoost algorithm? In other words, maybe there is a particular feature distribution giving better results. This could also be used to limit generated features.

For the Annona AD platform we are constantly evaluating new algorithms and technology to improve performance and functionality.

Internship Proposal: Evaluation of TabNet on fraud dataset

TabNet is a DNN architecture well-suited for tabular data. It enables interpretability like tree-based models, and efficient learning for the large datasets like DNN. It alleviates the need for feature engineering. TabNet is frequently used in tabular data competitions in Kaggle recently, and getting the higher position in the ranking. It's already deployed in AI platform on GCP and according to Google it is suited for financial applications such as fraud detection. The goal is to evaluate TabNet on a fraud dataset and compare it to the traditional models in use.

Internship Proposal: Evaluation of experiment management solutions

One new feature on the roadmap for the Annona AD platform is experiment management. Requirements for properly managing experiments are amongst others reproducibility, versioning of datasets, models, parameters and environment, convenient performance comparison, and availability to access and collaborate on within the organization. The goal would be to explore the current market of experiment management solutions both commercial and open source, and develop a PoC with the selected solution.

Internship Proposal: Evaluation of H20 for AutoML

The Annona AutoML component is responsible for building and optimizing models for production based on the features selected by Annona's Auto Feature Engineering. The goal would be to evaluate whether H2O would be a suitable underlying technology. It's performance should be compared to other of the shelf solutions e.g. GCP built models.

Cognisance Platform

Cognisance is a system that provides users with capabilities to gather sensor data from mobile and embedded devices, upload it to the cloud, perform ML training and deploy resulting models back to the devices. The system can be deployed component wise in an automated way to Google Cloud and separate components for mobile and embedded devices can ingest the resulting models.

Internship Proposal: Operational dashboard for end-to-end ML pipeline

The task for the internship is to enhance the deployment scripts of the system to instantiate a monitoring dashboard with Grafana on Kubernetes that would aggregate all real time monitoring of the system. This will include extracting metrics from already available sources as well as adding new emissions to sources that don't emit needed metrics in real time yet (for example: adding real time monitoring of accuracy during training).

This component should be encapsulated and separated from other components, so that it could be instantiated and destroyed any time (possibly persiting all metrics data as an option).

Human Sensing

The human sensing team is investigating how to automatically interpret human behavior with a current focus on mental health detection. By integrating changes in physiology with behavioral context like motion, the goal is to build algorithms which are applicable to be used in real-life conditions.

Internship Proposal: Predict heart rate changes due to physical activity

The task of the internship is to model changes in heart rate which are related to physical activity. The candidate will thereby learn how to plan and execute data recordings, process different types of physiological data (pre-processing, filtering, interpolation, fusion, ...) and create a structured dataset he/she can use for the creation of the algorithm. The candidate will explore different machine learning methods including deep learning and will learn how to validate the results. The optimal candidate should be interested in programming (Python), working with time-series data and applying statistical methods. The candidate should be able to do moderate physical activity.

Security of AI at Edge

Sony leverages Artificial Intelligence in several consumer and B2B products. The deployment of AI at the Edge of distributed systems can reduce latency of tasks, save battery power and guarantee user privacy. But it also makes these models vulnerable to attacks.

Our team thus evaluates the security strength of AI deployments using state-of-the-art side-channel attacks in order to design countermeasures against such attacks.

In particular, electromagnetic (EM) traces captured close to a processor executing an inference or power traces, can leak information about the model that is used. This might seem difficult due to several sources of EM/power noise (e.g. other operations being performed on the same processor, other electronic components, environment), unknown timing of the operations, and the large volume of data.

Lately, the effectiveness of attacks has increased manifolds because attackers can automate several steps in the attack chain. Following this evolution, BRL wants to investigate the advantages that attackers can achieve by applying techniques such as machine learning and signal processing to the attacking task.

If you have enthusiasm, and want to experience working in an industrial R&D environment, here are three possible challenges that we would like to take on together with you:

1. Analyzing performance of different types of filters on reducing the noise in raw traces
2. Developing tools for the profiling of EM/power traces
3. Using machine learning to extract neural network topology from EM/power traces

Internship Proposal

You will analyze the state of the art and use the results to

- implement and apply filters to acquired EM signal for eliminating noise; evaluate and compare the performance of filters,
- efficiently profile device behavior while performing known operations,
- use ML techniques to find profile patterns in new traces.

You will use industry best practices to develop, manage and deliver software.

Requirements

- Enthusiasm
- Basic understanding of computing hardware and software execution on embedded devices.
- Understanding of neural network, machine leaning
- Good knowledge of statistical analysis
- Development experience in C/C++, Python

Additionally, the following might prove very useful

- Familiarity with signal processing, signal alignment
- Static and dynamic assembly analysis (on ARM platforms)
- SCM (github), CI/CD (github action)
- Knowledge of Rust or Golang

LDP with heterogeneous noise distribution

Differential Privacy (Dwork, 2006) and its extension Local Differential Privacy (LDP) (Kasiviswanathan, et al., 2011) are promising statistical approaches for the anonymization of user generated databases. However, significant challenges remain for its industrial application. In LDP, a noise with known distribution is added to the individual data before it is collected. This noise allows to make individual data as useless as possible while allowing to use the collected data to compute the posterior distribution for a statistic of interest. One challenge of LDP is to handle heterogeneous noise distribution in the collected data. The objective of this task is to identify and characterize opportunities in that field.

Research in LDP often concerns either continuous or discrete variables, with significantly different approaches. The canonical method for dealing with continuous variables is to add a Laplacian noise to the original data, with a standard deviation relating to both variable range and privacy parameters (Dwork, et al., 2014). Dealing with categorical data most often relies on the Randomized Response method (Warner, 1965) combined with different information encoding strategies.

Internship Proposal

The proposed activities revolve around the most well-known randomization processes, setting aside any more complex and case dependent ideas. We consider two cases:

1. Mean estimation of a single continuous variable defined in the extended real numbers domain [L, U] with an added Gaussian noise. What if a noise is used instead? (see section 3.3 of (Dwork, 2006))
2. Probability of success of a binary variable using randomized response (see section 3.2 of (Dwork, 2006)). Pertinent reference (Ye, et al., 2019)

In all cases the goal is to derive the distribution of the considered metric when the privacy parameters are known but heterogeneous. The most important is to first obtain an unbiased estimator for the mean, and second for the standard deviation. An analytical solution might exist, but the solution can also be an algorithm to obtain a numerical solution. Special cases with simpler solutions can also be of interest.

Finally, the case when the privacy parameters are unknown is also of interest but might be more complicated to develop.

Required Background

Mathematics, Physics, or Engineering student with specialty in applied mathematics and/or statistics.

Secure AI

Internship Proposal: Adversarial Evasion Attacks improvements

Adversarial Evasion Attacks have been demonstrated on “side models” trained on tabular data, assuming that train set of production model is (at least partially) available. Moreover, attack transferability to production systems was also successfully performed for a few cases.

The goal of this project is to deepen the study of Adversarial Evasion Attacks and improve their performances, with an emphasis to the production setting. The candidate should start with a literature research of Evasion Attacks, taking over from what it has been already done in BRL Lab. After this, the candidate will learn how to extend the created attack algorithms to produce effective attacks that overcome some of the restrictions and assumptions made in our previous work. These activities include the study of black box attacks setting, where the train set is not available, the study of novel norms and metrics to ensure attacks imperceptibility and the study and evaluation of attack transferability, with the ultimate goal of improving the attack success rate towards a real production system.

The candidate will explore different machine learning algorithms with relative attack techniques. The ideal candidate should be interested in programming, Artificial Intelligence/Machine Learning and security topics.

Internship Proposal: Deploy Adversarial Training with a few Adversarial Attacks in train data

Adversarial Training has proven to be a simple but effective defense technique, used against adversarial attacks that are performed on Artificial Intelligence systems. By simply including in the training set the Adversarial Examples received in the past at execution time, the training procedures generates a model that is more robust against future adversarial attacks. However, given the novelty of Adversarial AI techniques, not so many historical Adversarial Examples are usually available in practice, during the training procedure. On the other hand, given the increasing popularity of publications related to

Adversarial AI, it is expected that these fraudulent practices will be more and more adopted in the near feature by fraudsters all over the world.

The goal of this project is to study a reliable procedure to generate a robust adversarial trained model for tabular transactional data, also in the cases where a few historical Adversarial Examples are available. The candidate should start with a literature research of Evasion Attacks and Defenses, taking over from what it has been already done in BRL Lab. After this, the candidate will learn how to perform and improve Adversarial Training and will come up with techniques to estimate and simulate realistic Adversarial Examples to be used in the training procedure. Finally, the output of the project will consist of a comparative study of the performances obtained, where the pros and cons of the proposed solutions will be detailed.

The candidate will explore different machine learning algorithms with relative attack and defenses techniques. The ideal candidate should be interested in programming, Artificial Intelligence/Machine Learning and security topics.