Master Thesis Proposals

Examiner: Prof. Vladimir Vlassov (vladv@kth.se)
Department of Computer Science, SCS division, KTH Kista
Academic supervisors: Zainab Abbas and Sana Imtiaz (PhD students, {zainabab,sanaim@kth.se})
External academic supervisor: M. Arsalan (PhD student, Infineon Technologies AG, Germany)

1. Time-series forecasting using spiking neural networks

Spiking neural networks (SNNs) mimic the behaviour of biological brains by transferring information over neurons via a sequence of spikes. They are counterparts to the deep neural networks (DNNs) and possess great potential in future applications to replace the current DNNs in terms of resource efficiency and robustness because they have very less compute requirements compared to other DNNs.

In this thesis, we aim to explore the application of SNNs for the task of time-series prediction compared to the current state-of-the-art time series prediction model based on deep learning. In particular, we will assess the SNNs with respect to 1) robustness towards learning on noisy data for user privacy, and 2) resource efficiency in the context of a smart health application.

Key Areas: Spiking neural networks, Deep learning, distributed systems, user privacy, smart health applications

2. Online machine learning on data streams

Traditional machine learning techniques rely on training an ML-based model over a batch of available historic data. Once the model is trained, its parameters do not change and the model is deployed for inference. This approach is not efficient for handling streaming data, which arrives in real-time and changes its behaviour over time as well.
In order to accurately perform learning tasks on streams, such as classification or prediction, the model should be updated based on recent information in the streams. This thesis work focuses on implementing an online ML framework, where the model is updated based on the new incoming data and deployed to serve the stream of new data. For example, in a data prediction task, the model not only predicts in real-time but also learns in real-time.

Key Areas: streaming systems, online machine learning, smart health applications

3. Robust learning using Graph Neural Networks

Graphs are used to represent many real-world relational data such as social networks. Recently there is an increasing interest in learning graph-structured data using neural networks. The neural networks that are used for learning graph data are called Graph Neural Networks (GNNs). GNNs can be used in graph processing tasks such as link prediction and node classification etc. Most of the existing GNNs work on static graphs and little work is done on exploring GNNs for streaming graphs. Moreover, the applications designed on streaming data often process private or sensitive information, and user privacy concerns should be respected in the learning process. In this thesis, we aim to explore the potential of using GNNs for streaming graphs, where the graph structure can evolve dynamically. Moreover, we aim to create a robust GNN model which can handle noisy data for differential user privacy and while giving an acceptable performance for sensitive data.

Key Areas: streaming systems, graph neural networks, differential privacy

4. Synthetic dataset generation using for user privacy

Collaborative research in the social and health sciences domain requires sharing and processing of sensitive and private user data. GDPR requires respecting user privacy by designing privacy-preserving systems. A possible approach is to use synthetic but accurate representative datasets based on real, sensitive data. Generative adversarial networks (GANs) are highly popular machine learning based data augmentation and generation solutions. However, training GANs is a hard task due to long training times. On the other hand, powerful tools exist in the natural language processing domain that use deep learning and are capable of creating human-like text. This thesis aims to explore the possibility of using powerful NLP prediction tools like Generative Pre-trained Transformer (gpt) or BigBird, for data generation for respecting user privacy. Moreover, we will compare this approach to the generally popular synthetic data generation approaches like GANs in context of training time and data usability.

Key Areas: deep learning, generative adversarial networks, smart health care, user privacy