Stage de fin d'étude école d'ingénieur/master2

Internship subject: ROBUST MULTI-AGENT SYSTEMS BASED AND DISTRIBUTED DATA STREAM LEARNING IN A COLLABORATIVE ENVIRONMENT

Context

Le Commissariat à l'Energie Atomique (CEA) est un acteur européen majeur en matière de recherche, de développement et d'innovation. Cet organisme de recherche technologique intervient dans quatre grands domaines : l'énergie, les technologies pour l'information, la santé et la défense. Situé en île de France sud sur le campus de Saclay, le Laboratoire d'Intégration des Systèmes et des Technologies (LIST) conçoit des systèmes numériques intelligents. Ses 750 collaborateurs accompagnent chaque année 200 entreprises françaises et étrangères sur des projets de recherche appliquée, dans quatre domaines : Manufacturing avancé, Systèmes embarqués, Data Intelligence et Maîtrise des rayonnements pour la santé. Les équipes du CEA LIST sont également partenaires de nombreux laboratoires universitaires, de grandes écoles et d'autres organismes de recherche au travers des projets de recherche collaborative. Elles font partie intégrante de DIGITEO LABS, campus scientifique réunissant plus de 1200 chercheurs spécialistes des technologies de l'information (CEA, INRIA, CNRS, Supélec, Ecole Polytechnique, Université d'Orsay). Au sein cet institut, le SID (Service Intelligence des Données) développe des solutions d'Intelligence Artificielle pour l'aide à la décision, orientées utilisateur et supervision automatique de systèmes complexes.

Subject

Nowadays, data streams are present in more and more applications and domains where dynamism and speed truly matters. In practice those streams represent dynamic data flows, coming from different sources, where their content evolves in time. Research has been done in the subject and many techniques for stream mining have emerged [5]. These algorithms usually sample the data stream in a certain way and deal with them incrementally or online. Despite the results provided by these new techniques, the flow of data is underutilized, which potentially leads to a loss of useful information on the one hand, and to forget what has been previously discovered on the other. Moreover the complexity of current digital applications, and those of the near future, is constantly increasing

due to a combination of aspects such as the large number of sources, the non-linearity of certain processes, the distribution of knowledge and control, the time response, the strong dynamics of its environment or the unpredictability of interactions among others. The aforementioned complexity opens new research challenges about the generation and processing (learning) of these streams, especially in distributed, heterogeneous and collaborative environments. Existing ones lack, in general, the means for collaborating, negotiating, sharing, or validating data streams on such kind of heterogeneous environments. Multi-Agent Systems (MAS) have been demonstrated to be an appropriate technology for dealing with those issues. Their principles enable some of these features but however there is still work to do in order to comply them with the characteristics of data streams.

The aim of this internship is to use agents in data streams to deal with the previously mentioned challenges. Four main blocks are potentially identified: (i) manage non synchronised data streams from different sources, (ii) do research in distributed on-line learning algorithms, (iii) increase the robustness of online learning models that deal with such streams (make them reliable through unexpected environment changes) and (iv) generate new metrics to evaluate the needs mentioned before. This work will rely and improve the STREAMER framework already existing in the lab. STREAMER is a cutting-edge data stream processing (Complex Event Processing) platform devoted to analyzing sequential data for electrical or industrial systems. Such ongoing platform already counts with independent modules which provide their own functionality, such as a graphical interface, machine learning algorithms, communication utilities, etc.

This internship may be potentially extended to a 3 years PhD.

Keywords: Data Stream, Multi-agent Systems, Machine Learning.

Qualifications

We look for a candidate with:
- Engineering/master (Bac + 4) studies, preferable in computer science.
- Strong programming skills (**Java**, R, Python).
- English proficiency speaking.
- French and/or spanish is a plus.
- Knowledge in machine learning is a plus.
- Knowledge in multi-agent systems is a plus.
- Familiar with InfluxDB [1], Kafka [2], Redis [3] is a plus.

Other information

Duration	6 months (plus 2-3 months of administrative delay for contract signature).
Salary	From 700 euros/month (depending on the candidate's CV).
Navettes	Free shuttles CEA from/to the laboratory to many points of île de France.
Transport Card	75% of refund for transport card.

Contact

Send your CV, motivation letter and a recent grades transcript to Sandra Garcia Rodriguez (sandra.garciarodriguez@cea.fr) with email subject [Candidature de stage].

Sandra, GARCIA RODRIGUEZ
Ingénieur - Chercheur

Service Intelligence des Données
Laboratoire Intelligence Artificielle et Apprentissage Automatique

Institut CEA LIST
CEA Saclay DIGITEO LABS **|** Bât. 565-PC192 **|** F-91191 Gif-sur-Yvette Cedex
sandra.garciarodriguez@cea.fr

References:

[1] https://www.influxdata.com/

[2] https://kafka.apache.org/

[3] https://redis.io/

[4] Muthukrishnan, S. (2005). Data streams: Algorithms and applications. *Foundations and Trends® in Theoretical Computer Science*, *1*(2), 117-236.

[5] Palpanas, Themis. "Data series management: The road to big sequence analytics." *ACM SIGMOD Record* 44.2 (2015): 47-52.

[6] Tommasini, Riccardo, Davide Calvaresi, and Jean-Paul Calbimonte. "Stream Reasoning Agents: Blue Sky Ideas Track." *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2019.

list

Laboratoire d'intégration des systèmes et des technologies

[7] Xie, Jing, and Chen-Ching Liu. "Multi-agent systems and their applications." *Journal of International Council on Electrical Engineering* 7.1 (2017): 188-197.

Laboratoire d'intégration des systèmes et des technologies