2019-02263 - Progressive Data Series k-NN Similarity Search and Classification with Probabilistic Guarantees

Type de contrat : Convention de stage
Niveau de diplôme exigé : Bac + 5 ou équivalent
Fonction : Stagiaire de la recherche

A propos du centre ou de la direction fonctionnelle

Located at the heart of the main national research and higher education cluster, member of the Université Paris Saclay, a major actor in the French Investments for the Future Programme (Idex, LabEx, IRT, EquipeX) and partner of the main establishments present on the plateau, the centre is particularly active in three major areas: data and knowledge; safety, security and reliability; modelling, simulation and optimisation (with priority given to energy).

The 450 researchers and engineers from Inria and its partners who work in the research centre’s 28 teams, the 60 research support staff members, the high-level equipment at their disposal (image walls, high-performance computing clusters, sensor networks), and the privileged relationships with prestigious industrial partners all make Inria Saclay Ile-de-France a key research centre in the local landscape and one that is oriented towards Europe and the world.

Contexte et atouts du poste

Contacts:
Theo Palmans (Paris Descartes University): themin@mi.parisdescartes.fr
Theophanis Tsandilas (Inria and Université Paris-Saclay): theophanis.tsandilas@inria.fr
Anastasia Bezerianos (Université Paris-Saclay and Inria): Anastasia.Bezerianos@in.fr

Collaboration:
The internship is a collaboration between Inria (EXISTU and ILDA teams) with an expertise in Visualization and Interaction and the University of Paris Descartes with an expertize in data mining.

Context:

Data series are ordered sequences of values measured and recorded from a wide range of human activities and natural processes, such as seismic activity, human brain oxygen saturation, or electroencephalography (EEG) signal recordings. The analysis of such sequence is becoming increasingly challenging given their sizes that are often times measured in the order of multiple terabytes [1].

Data series analysis involves pattern matching, anomaly detection, frequent pattern mining, clustering, and classification. These tasks rely on the notion of data series similarity. Data series similarity is often domain- and visualization-dependent, and in many situations, analysts depend on time-consuming manual analysis processes [2].

For example, neuroscientists manually inspect the EEG data of their patients, using visual analysis tools, so as to identify patterns of interest. In such cases, it is important to have techniques that operate within interactive response times in order to enable analysts to complete their tasks easily and quickly. To allow for interactive response times when users analyze large data series collections, we need to consider progressive and iterative visual analytics approaches [3,4]. Such approaches provide progressive answers to users’ requests, sometimes based on algorithms that return quick approximate answers. The goal is to support exploration and decision making by providing progressive (i.e., intermediate) results, before the final and exact ones have been computed.

References:


Mission confiée

We have developed a probabilistic learning-based method that provides probabilistic time and distance guarantees for progressive k-Nearest Neighbor (1-NN) query answering. Our approach learns from a small set of queries and builds statistical models that assess how far we are from exact answers. We can then communicate this information to users, and users can decide to terminate a progressive search to reduce waiting times.

The goal of the internship is to develop new techniques that extend this approach to more complex data-series analysis tasks, for example, to the problems of k-Nearest Neighbor similarity search, and k-Nearest Neighbor classification. In all cases, we need to develop techniques that will produce answers much earlier than running the full algorithm, and at the same time provide probabilistic guarantees on the quality of the produced results.

Informations générales

- Thème/Domaine : Interaction et visualisation
- Ville : Orsay and Paris
- Centre Inria : CRI Saclay - Ile-de-France
- Date de prise de fonction souhaitée : 2020-03-01
- Durée de contrat : 6 mois
- Date limite pour postuler : 2020-02-28

Contacts

- Equipe Inria : EX-SITU
- Recruteur : Theophanis.Tsandilas@inria.fr

A propos d'Inria

Inria, est l’institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 chercheurs et ingénieurs pour relever les défis du numérique, souvent à l’interface d’autres disciplines. L’institut fait appel à de nombreux talents dans plus d’une quarantaine de métiers différents. 900 personnels d’appui à la recherche contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde.

Consignes pour postuler

- CV
- Cover letter

Sécurité défense :

Ce poste est susceptible d’être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L’autorisation d’accès à une zone est délivrée par le chef d’établissement, après avis ministériel favorable, tel que défini dans l’arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l’annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

Attention : Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d’autres canaux n’est pas garanti.
Compétences
The candidate is expected to have a Computer Science degree (M1/M2 level), knowledge of statistical analysis methods, and experience with C/C++ and R programming.

Avantages
- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage