

Postdoctoral fellowship on “Improving security for Paris 2024 Olympic Games: discovering anomalous urban situations via real-time analysis of mobile phone data”

- **Level of qualifications required:** PhD degree in Computer Science, Machine Learning, Applied Math or strictly-related field;
- **Starting date:** August-October 2020;
- **Duration of contract:** 12 months
- **Deadline to apply:** applications will be reviewed until the position is filled;
- **Main Location:** Lyon, France (LICIT, UNIVERSITÉ GUSTAVE EIFFEL/ENTPE): <https://goo.gl/maps/K19HBR4ETZ92>
- **Project Team:**
 - Nour-Eddin EL FAOUZI (Lab Director), LICIT, University of Lyon, ENTPE, UNIVERSITÉ GUSTAVE EIFFEL-COSYS
 - Angelo FURNO (Researcher), LICIT, University of Lyon, ENTPE, UNIVERSITÉ GUSTAVE EIFFEL-COSYS
 - Marco FIORE (Researcher), CNR - IEIT, Turin, Italy
 - Zbigniew SMOREDA (Researcher), Orange, France
 - Eric GAUME (Lab Director), UNIVERSITÉ GUSTAVE EIFFEL-GERS

Context

We are looking for an enthusiastic Postdoctoral candidate to carry out research in the context of the ANR-funded DISCRET projet (French title: *Démonstrateur d' Identification de Situations Critiques via la Remontée de données multisources pour l'aERte en Temps-réel*; English translation: *Prototype for the identification of Critical Situations via multi-source data for real-time alert*).

The goal of the project is to demonstrate the possibility to detect and locate, in real-time, unusual or critical situations in urban areas (e.g., attacks, fires, sudden weather-related events, etc.), based on the analysis of mobile phone probe network data. This detection will be complemented with information extracted from social networks (i.e., Twitter in the context of the project) and other sources of contextual data.

Several recent research works have shown that major events induce locally significant modifications of the amount and nature of cellular network communications [1]. These anomalies, typically concomitant with the unusual event, may be detected and located based on the network of cell phone antennas and the associated user-generated traffic information. The early detection and localization of the events allow for a more effective retrieval of information from the social networks. That permits to provide elements of description and context for the detected event and, therefore, to increase the confidence and the amount of information conveyed by the population via channels that are not explicitly conceived for alerting purposes.

A prototype of a warning platform for security and emergency operators will be implemented, tested and demonstrated as part of the whole project. The final prototype is expected to offer TRL 6 solutions by the end of the project that could be subsequently industrialized and operated by 2023, in the context of the Olympic Games that will be held in Paris in 2024.

The subject is at the interface between machine learning, big data processing, networking and transportation.

Main activities

The Postdoc will have the unique opportunity to work on large-scale, already available mobile phone datasets, collected by the Orange French network provider, consisting in 2G, 3G and 4G network probe data, as well as more traditional Call Detail Records (CDR).

Additionally, novel highly-detailed datasets on the usages of Internet mobile phone apps from mobile phone users will be specifically collected in the framework of the project, as well as detailed information on the nature, occurrence and location of possible incidents during the observed events.

In a first phase, the activity of the postdoctoral candidate will consist in analyzing the collected data and extracting, via machine learning techniques and previous work from the team [2, 3], spatio-temporal fine-grained signatures of the typical network activity (aggregated at the antenna level) with different temporal resolution (5, 10, 60 minutes).

In a second phase, the postdoctoral fellow is expected to explore and define efficient classification techniques [4, 5] for the inference of atypical situations (increase in the volume of the communication and consumption activity of certain services, sudden growth of mobility-related events, change of signal shape, etc.) compared to prototypical mobile phone signatures as identified from phase 1.

This second phase will also focus on the creation of a process for periodic updating of signatures in order to adapt them to changes in the actual communication activity at certain places of interest.

From a methodological point of view, the main challenge is to develop a classification method that can work in real time. To achieve this goal, a method combining artificial intelligence (AI) and statistical learning will be implemented (neural networks-based approaches, curve classification method, kernel method or generative method). It will be completed by sequential statistics (Wald test) to detect in real time the deviations from prototypes (profiles or prototypical signatures).

As the volume of data to be analyzed is expected to be significant, the detection of anomalies and the periodic updating of the standard signatures will have to be carried out as close as possible to the source of the data flows so as to minimize network load and latency (Mobile Edge Computing). To update the standard signatures, in cooperation with our Orange partner, it is expected to implement the proposed methods according to a distributed learning process based on Federated Learning [6] implemented via Edge Computing [7].

From a technology perspective, the computational complexity of data mining requires the use of appropriate solutions for real-time and scalable implementation. As part of the project we will explore the use of Apache Kafka, TensorFlow and Apache Flink open source platforms to meet this criterion.

It is expected that the successful candidate will contribute to top-tier computer networks, self-adaptive distributed systems, big data and machine-learning conferences and journals (IEEE INFOCOM, IEEE ICDM, ACM SIGKDD, IEEE Big Data, IEEE Transactions on Autonomous and Adaptive Systems,

Transportation Research Board, IEEE Intelligent Transportation Systems, Transportation Research, etc.).

The postdoc will closely interact with researchers of Orange Labs, in relation to both data collection/analysis (Sense Lab in Paris) as well as in prototyping the distributed real-time platform for anomaly detection (Nice). The postdoc will also collaborate with researchers from the UNIVERSITÉ GUSTAVE EIFFEL of the Géotechnique, Environnement, Risques naturels et Sciences de la Terre (GERS) department, in relation to methods and models for risk prediction and prevention, as well as researchers of the Université de technologie de Troyes (UTT) to integrate the anomaly detection solutions with a Twitter engine providing contextual information on the detected events from tweets concurrently collected in the observed areas.

Required Skills

We look for strongly motivated candidates with a solid background in computer science, mathematics and probability/statistics. Candidates with proven skills in the fields of machine learning, artificial intelligence, big data, data mining. Knowledge of distributed programming will be strongly appreciated. Programming skills with Python, Scala or Java are desired.

Proven written and verbal communication skills with fluency in written and spoken English.

Benefits package

- Fully reimbursed visiting periods to Paris and Nice to work at Orange facilities;
- Abroad visiting periods in Italy (CNR-IEIIT);
- Subsidized catering service
- Partially reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Access to sports facilities

Salary

- 2,600-2,900 euros (gross monthly salary, depending on years of experience),
- 2,000-2,300 euros (net salary, depending on years of experience)

Recruiters' Contacts

- Dr. Angelo FURNO / angelo.furno@Universit  Gustave Eiffel.fr
- Prof. Nour-Eddin EL FAOUZI / nour-eddin.elfaouzi@Universit  Gustave Eiffel.fr
- Dr. Marco FIORE / marco.fiore@ieiit.cnr.it
- Dr. Zbigniew SMOREDA / zbigniew.smoreda@orange.com

Required documents

- A curriculum vitae;
- The complete record of master grades (relevé de notes M1 and M2 for French candidates)
- The Phd Thesis manuscript;
- A two-page motivation letter discussing how the candidate's background and research interests relate to the proposed subject and bibliographic references [1-7].

About the Transport and Traffic Engineering Lab, UNIVERSITÉ GUSTAVE EIFFEL - ENTPE

The Transport and Traffic Engineering Laboratory (LICIT) is a Joint Research Unit under the dual administrative supervision of the [French Institute of Science and Technology for Transport, Development and Networks](#) (UNIVERSITÉ GUSTAVE EIFFEL) and the [National Post-Graduate School of Public Civil Engineering](#) (ENTPE). It is recognized for its work in traffic modelling and engineering. The laboratory has already developed many successful applications for both traffic information and simulation tools

The UNIVERSITÉ GUSTAVE EIFFEL is a state-financed scientific and technological institute under the supervision of the Ministry of Research and the Ministry of Transport. The Institute's activities involve diverse fields as acoustics, mechanics, mathematics, computer science, electronics and electro-technical sciences. The diversity of the approach used to carry out the different research programmes gives a multidisciplinary characteristic to the UNIVERSITÉ GUSTAVE EIFFEL research teams.

The UNIVERSITÉ GUSTAVE EIFFEL research programme covers many aspects of work involved within this project including driving aids, information, assistance and automation; transport networks and services; sustainability, environment and road safety. UNIVERSITÉ GUSTAVE EIFFEL has been and still is heavily involved in various projects, especially European Commission-supported research like the INTRO project. UNIVERSITÉ GUSTAVE EIFFEL's experience will be used to assist in achieving the aims of this project.

Relevant References

- [1] D. Naboulsi, **M. Fiore**, S. Ribot, and R. Stanica, "Large-scale mobile traffic analysis: a survey," IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 124–161, 2015.
- [2] **A. Furno**, **M. Fiore** and R. Stanica, "Joint spatial and temporal classification of mobile traffic demands," in INFOCOM 2017-IEEE Conference on Computer Communications, IEEE, 2017.

- [3] **A. Furno, M. Fiore**, R. Stanica, C. Ziemlicki and **Z. Smoreda**, "A tale of ten cities: Characterizing signatures of mobile traffic in urban areas," IEEE Transactions on Mobile Computing, vol. 16, pp. 2682-2696, 2017.
- [4] L. Fahrmeir, T. Kneib, S. Lang, B. Marx. "Regression : Models, Methods and Applications". Berlin: Springer. p. 663, 2013.
- [5] A. Ben-Aissa, **N.-E. El Faouzi**, and E. Lefevre, "Classification multisource via la théorie des fonctions de croyance: application à l'estimation du temps de parcours," Revue de Statistique Appliquée, p. 17p, 2009.
- [6] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, "In-edge ai: Intelligentizing mobile edge computing, caching and communication by federated learning," arXiv preprint arXiv:1809.07857, 2018.
- [7] N. di Pietro, M. Merluzzi, E. C. Strinati, and S. Barbarossa, "Resilient design of 5g mobile-edge computing over intermittent mmwave links," arXiv preprint arXiv:1901.01894, 2019.