3 QUESTIONS À KARINE ZEITOUNI

Enseignante-chercheuse au DAVID*, Karine Zeitouni fait partie de l'équipe ADAM - Ambient Data Access and Mining, et est actuellement impliquée principalement dans les projets Polluscope, Master H2020, Astroide, et Dataia.

Publié dans La lettre de la recherche n°45, octobre 2018

**Données et Algorithmes pour une ville intelligente et durable*

1/ En quoi consiste votre recherche ?

« Je travaille sur la gestion de données et l'analyse par des techniques de fouilles de données. La donnée est au centre de tout et s'apparente à un matériau finalement. De la même façon que l'on fait une recette avec plusieurs ingrédients, l'on saisit l'opportunité

de créer de la connaissance nouvelle à condition de combiner et d'intégrer ces données d'une certaine manière.
Ces informations peuvent être d'une grande variété dans leur nature et leur provenance : données du Web, données ouvertes, données d'observations (univers, capteurs, localisation), transactions, clients, tweets, blogs, etc.
Chacune de ces informations comporte des caractéristiques qu'il est souvent nécessaire d'exhiber et de prendre en compte à la fois dans le traitement par requêtes et dans les algorithmes de fouille de données. Ainsi, les spécificités des données impactent la manière d'optimiser les requêtes. Leur analyse implique le développement de méthodes de fouille de données spécifiques aux caractéristiques des données, ainsi qu'à leur volumétrie. Les domaines d'application sont multiples : ville intelligente, mobilité urbaine (notamment avec VEDECOM, les collectivités territoriales, Versailles Grand Parc), environnement, santé, etc. »

2/ Vous êtes la porteuse scientifique du projet Polluscope. Pouvez-vous nous en parler ?
« L'objectif de ce projet ANR, démarré en 2016, est de mieux comprendre l'exposition individuelle aux risques de pollution et l'effet sur la santé individuelle, en prenant en compte différents lieux tels que celui du travail, du domicile, des transports, et ainsi de générer des profils d'exposition. Il s'agit d'une infrastructure de mesurage de cette exposition en personnalisé puisque les personnes portent des capteurs fournissant ces informations. Les données fournies sont donc très précises et permettent des recommandations selon les variations fines, contrairement aux cartes de pollution qui se basent sur des stations de mesure éparses.
Projet pluridisciplinaire, notre laboratoire collabore entre autres avec le Laboratoire des sciences du climat et de l'environnement (LSCE), Airparif en charge de l'observation de la qualité de l'air en Ile-de-France, l'Inserm avec une équipe mixte de Paris Sorbonne, université reconnue dans le domaine de l'expologie. Ce sont deux cents personnes qui seront observées lors de la campagne de mesures qui durera deux ans sur les cinq années du projet.
Outre la partie expérimentale consistant en la gestion des données via des plateformes dans le cloud, le projet pose des problèmes de recherche directement liés à la nature spécifique des mesures par capteurs mobiles. En effet, les données reçues massivement en continu n'ont de sens que si elles sont contextualisées, pour analyser qui est exposé à quoi, où et quand. La prise en compte du contexte spatial et temporel est un des principaux verrous scientifiques ».

3/ À quels autres projets participez-vous en parallèle ?

« Parlons tout d'abord du projet MASTER H2020 (Multiple Aspects Trajectory management and analysis) dans lequel je suis impliquée. L'objectif est de faire collaborer des équipes européennes et internationales autour d'une thématique abordant des problèmes scientifiques d'analyse de données de trajectoires, lesquelles peuvent être très variées : trajectoires de vie, sémantiques, ou liées à des trajectoires de positions géographiques, de points d'intérêts, de Tweets.
Dans un tout autre domaine, je participe au projet ASTROIDE, co-financé par l'UVSQ et le CNES. Il se base sur l'observation spatiale de positions et de paramètres physiques (températures, éloignement en années-lumière, magnitude) des astres. Ces observations sont de plus en plus nombreuses et précises grâce aux caméras de très haute résolution et aux capteurs puissants associés. Les données massives générées nécessitent un système de gestion dédié et le développement d'une méthodologie pour en permettre la manipulation.
Enfin, je travaille aussi pour DATAIA, l'institut de convergence interdisciplinaire de l' université Paris-Saclay pour la recherche sur les données, dont le consortium rassemble des universités, des instituts de recherche nationaux et des grandes écoles. Plus particulièrement, le projet StreamOps vise une nouvelle plateforme logicielle qui analysera des données temporelles générées par l'avancée technologique en matière de capteurs et de développement de l'IoT (Internet of Things). Concernant les données médicales par exemple, il s'agirait de définir un outil informatique qui gèrerait les données en temps réel des patients en post-opératoire afin de réagir par anticipation à un risque potentiel ».

INFORMATIONS COMPLÉMENTAIRES

En savoir plus
Crédit photo : UVSQ
> DAVID
> Polluscope
> Master H2020
> ASTROIDE
> DATAIA