

Apprentissage automatique et fusion de données pour la surveillance environnementale. Application à la prévention des risques d'exposition aux polluants dans les environnements de travail.

MOTS CLES : données spatio-temporelles, apprentissage automatique, tenseurs, fusion de données

CONTEXTE ET OBJECTIFS

La thèse se déroulera au sein des deux organisations :

- L'**INRS** (Institut National de Recherche et de Sécurité), 1 rue du Morvan, CS60027, 54501 Vandœuvre-les-Nancy Cedex. Personne à contacter : Dr Ph. DUQUENNE ; tel : 03 83 50 98 75, E-mail : philippe.duquenne@inrs.fr
- Le **CRAN** (Centre de Recherche en Automatique de Nancy), Campus sciences - Bd des Aiguillettes B.P. 70239, Vandœuvre-lès-Nancy Cedex. Personne à contacter : Dr S. MIRON ; tel : 03 72 74 53 37 ; E-mail : sebastian.miron@univ-lorraine.fr

Les récents développements technologiques (réseau de capteurs, géolocalisation, scanner 3D etc.) ouvrent des nouveaux champs d'application en surveillance environnementale et prévention des risques d'exposition aux polluants. Toutefois, l'analyse et la représentation de données acquises grâce à ces nouvelles technologies soulèvent des questions scientifiques qui doivent être solutionnées pour pouvoir bénéficier pleinement des progrès techniques. Ces questions concernent principalement la création et l'interprétation des cartes spatio-temporelles de pollution et la fusion efficace d'informations fournies par des capteurs de natures différentes (gaz / particules), voir par d'autres systèmes d'acquisition (Scanner 3D, système de localisation etc.).

L'objectif de la thèse est de développer une méthode automatique pour la cartographie spatio-temporelle des concentrations des polluants ainsi qu'une méthodologie pour l'évaluation des risques d'exposition des travailleurs, en utilisant les données acquises par différents types d'instruments de mesure déployés dans l'environnement de travail. Ce projet s'inscrit naturellement dans le cadre de la transformation numérique actuelle de la fabrication/production et des industries connexes, communément appelé « Industrie 4.0 ».

DESCRIPTION

La thèse débutera par un travail bibliographique et de recherche d'information, visant à mettre à jour les informations publiées dans la littérature sur l'analyse spatio-temporelle, l'apprentissage automatique et la fusion de données pour la cartographie. Le travail de recherche sera ensuite organisé en 3 étapes :

L'étape 1 vise à construire un outil d'analyse et de représentation des données spatio-temporelles collectées par des réseaux de capteurs spécifiques. Cet outil doit permettre d'utiliser conjointement la cartographie dynamique des concentrations, la représentation numérique des locaux et la localisation des équipements et celle des opérateurs dans ces locaux. Il sera fondé principalement sur des méthodes d'apprentissage automatique à partir des données multi-capteur, sur des algorithmes d'interpolation 2D-3D (e.g., Krigeage [1]) et des approches tensorielles pour la fusion des données [2,3,4].

A partir des cartes de pollution ainsi construites et des données issues des capteurs de géolocalisation qui équiperont les travailleurs, une méthodologie pour l'estimation du risque d'exposition aux polluants sera ensuite proposée.

L'étape 2 concerne la validation de la méthode développée et prévoit des essais expérimentaux. L'INRS dispose d'une chambre expérimentale avec des paramètres environnementaux contrôlés (température, humidité, etc.). Pour de telles conditions, il existe des modèles mathématiques qui permettent de prédire la dispersion d'un gaz dans l'espace et dans le temps. Nous utiliserons ces prédictions mathématiques comme une forme de « vérité terrain » pour valider la nouvelle méthode proposée.

L'étape 3 vise à effectuer des applications en entreprises (conditions réelles d'exposition) de l'outil développé.

La thèse s'inscrit dans une étude pluridisciplinaire qui mobilisera des personnes, des compétences et des connaissances multiples (électronique, informatique, instrumentation, stratégie de mesures, exploitation des données etc.). Le doctorant devra veiller à la programmation et à la coordination de ses activités avec les différents acteurs et participera activement à la rédaction des publications scientifiques issues des travaux. Il pilotera et participera aux essais expérimentaux au laboratoire et en entreprises et assurera le traitement des données collectées.

REFERENCES

- [1] Stein, Michael L. Interpolation of spatial data: some theory for kriging. Springer Science & Business Media, 2012.
- [2] Acar E., Bro R., and Smilde A.K. Data fusion in metabolomics using coupled matrix and tensor factorizations. Proceedings of the IEEE, 103(9):1602–1620, 2015.
- [3] Farias R.C., Cohen J.E., and Comon P. Exploring multimodal data fusion through joint decompositions with flexible couplings. IEEE Transactions on Signal Processing, 64(18):4830–4844, 2016.
- [4] Kanatsoulis CI, Fu X, Sidiropoulos ND, Ma WK. Hyperspectral super-resolution: A coupled tensor factorization approach. IEEE Transactions on Signal Processing. 2018 Oct 17;66(24):6503-17.

PROFIL ET APTITUDES SOUHAITÉES:

- Étudiant.e en MASTER deuxième année (niveau bac + 5)
- Connaissances en analyse de données et/ou traitement du signal
- Connaissances en programmation et maîtrise des logiciels de calcul numérique (e.g., Matlab)
- Connaissances de base sur les réseaux de capteurs
- Lecture de l'anglais, autonomie

CONDITIONS DE LA THESE :

- Thèse de 36 mois à partir de janvier 2022
- Thèse rémunéré (environ 2500 € bruts/mois)
- Le doctorant sera basé à Vandœuvre-lès-Nancy (54)