

Proposition de sujet de thèse 2022

Vers la découverte automatique de zones d'intérêt dans le domaine du transport maritime

1 Contexte

Le transport maritime assure 90% des échanges mondiaux de biens et marchandises. D'un point de vue environnemental, il est de loin le moyen de transport le plus efficient à la tonne de marchandise transportée, bien qu'il doive lui aussi relever le défi de réduire significativement ses émissions. Aujourd'hui l'industrie maritime a initié une profonde transformation qui nécessite un effort important d'innovation et de recherche de la part de l'ensemble des acteurs de l'industrie pour lever les verrous technologiques et scientifiques actuels.

Le projet TNTM (Transformation Numérique du Transport Maritime) s'intègre dans ce contexte d'optimisation de l'industrie maritime et se propose d'agir sur l'ensemble des acteurs aussi bien en mer que sur terre, pour réussir le pari d'un transport moins-carboné voire décarboné. Ainsi, un des axes pour parvenir à ces objectifs est l'excellence opérationnelle qui vise à optimiser l'usage des navires. Cette excellence opérationnelle est basée sur deux innovations majeures, la première innovation est celle de l'IoT et des données, encore très peu présentes sur les navires et les conteneurs maritimes. La deuxième innovation est liée à l'utilisation d'algorithmes d'optimisation pour améliorer la chaîne logistique maritime du transport par conteneur, en intégrant les dernières avancées de la recherche opérationnelle, des méthodes d'optimisation, de la simulation et du traitement de données (machine learning, etc.). C'est dans ce cadre que se situe cette thèse.

Aujourd'hui, de multiples acteurs différents (sites de chargements et déchargements, transporteurs, terminaux, etc.) sont impliqués dans le transport d'un conteneur. Chaque passage de responsabilité entre deux acteurs crée un risque de mauvaise exécution du plan de transport et cette chaîne logistique est désynchronisée. Il n'existe souvent aucune contextualisation dynamique du plan de voyage du conteneur, et ce, même en ayant connaissance des données issues du conteneur intelligent. Par exemple, les zones d'intérêts géographiques (ZOI) du plan de voyage sont créés manuellement par les utilisateurs et les clients. Autre exemple, des surveillances et des extractions manuelles des données sont réalisées quotidiennement pour fournir des alertes et des informations au client face à des événements inattendus.

2 Objectifs

Le contexte scientifique de cette thèse relève du *Trajectory (Data) Mining*, i.e., un ensemble de techniques visant à explorer, analyser, et fouiller des données spatio-temporelles issues des trajectoires d'objets mobiles. Différentes problématiques sont associées à ce domaine, comme par exemple l'analyse exploratoire, le clustering de trajectoires, la classification, la détection d'anomalies, etc.

Parmi les nombreuses problématiques liées à ce domaine, nous nous focaliserons dans cette thèse sur : (1) la reconstruction géographique des trajectoires à partir de données de géopositionnement de conteneurs maritimes, et (2) l'enrichissement sémantique des trajectoires, notamment par l'identification dynamique de ZOI.

Il existe de nombreux modèles mathématiques pour reconstruire des trajectoires à partir du clustering ou de l'identification de patterns dans des données spatio-temporelles [LHW07, LHLG08]. L'algorithme DBscan, et ses variantes, sont souvent les plus utilisés car bien adaptés aux traces de positionnement fournies par les technologies GNSS (Global Navigation Satellite System) [ZNX20, CJW14,

GGCSNM18] ou aux données issues des télécommunications ou du radiopositionnement (Call Detail Records, indoor-positioning, etc.). Le challenge ici réside dans l’application de tels modèles à des données à grosse granularité temporelle comme celles utilisées dans le projet TNTM et qui sont des données AIS (Automatic Identification System) [CLY+20, WCCM21].

Concernant l’enrichissement sémantique des trajectoires [PSR+13], le problème consiste à représenter et exploiter un ensemble d’informations contextuelles telles que la détection du mode de transport, le motif et le contexte de déplacement, la découverte des zones d’intérêts, etc. [CSF+21, dSMB+]. Dans le cadre de cette thèse nous nous limiterons à la détection automatique des ZOI et à leur classification [LHW07]. L’extraction de ZOI à partir de données spatio-temporelles se base généralement sur les algorithmes de reconstruction des trajectoires. On observe toutefois un recours accru aux données issues des connaissances terrain (données sursol, connaissance métier, etc.) pour enrichir et affiner la reconstruction de telles zones. C’est sur ce dernier point que nous concentrerons nos efforts pour développer un modèle de classification des ZOI.

Les principaux objectifs du projet de thèse sont les suivants :

- Analyser la qualité des données issues des conteneurs intelligents.
- Concevoir un modèle de représentation des trajectoires.
- Identifier les ZOI et leur emprise spatiale à partir des trajectoires spatio-temporelles des conteneurs.
- Classifier ces ZOI en fonction des informations contextuelles et des connaissances métier de notre partenaire industriel dans le projet TNTM.

3 Contributions scientifiques attendues

Nous attendons des contributions pour les problématiques suivantes avec des données spatio-temporelles caractérisées par une grande disparité :

La reconstruction des trajectoires d’objets mobiles à partir de données de positionnement est un domaine toujours d’actualité, dont les ramifications vont du routage dynamique de véhicules à la planification urbaine, en passant par la caractérisation des territoires, l’optimisation de tournées de véhicules, l’optimisation des déplacements, et enfin la reconstruction de trajets avec notamment la détection de ZOI.

Dans cette thèse, nous nous focalisons sur des données issues du transport maritime, et plus principalement sur des données issues de GNSS avec une grosse granularité temporelle. La reconstruction de trajectoires avec ce type de données constitue un véritable défi scientifique.

L’enrichissement sémantique des trajectoires s’appuie sur toute information contextuelle telle que la vitesse ou encore les connaissances issues du sursol (topographie, topologie, bâtiments, emprise spatiale, etc.). Plusieurs modèles de trajectoires sémantiques ont été proposés dans la littérature et le défi scientifique réside dans l’élaboration et la mise en oeuvre de modèles sur des données spatio-temporelles éparses.

L’identification des points d’arrêts Une trajectoire peut être vue comme un ensemble de points ou zones d’intérêts, et ces points correspondent à des activités répertoriées et récurrentes de ”l’objet mobile”. Le challenge de cette thèse est de développer un modèle basé sur la littérature (DBscan, etc.) qui soit adapté aux données TNTM (données AIS à grosse granularité temporelle, données multi-modales, etc.)

La classification des ZOI Les ZOI sont des zones que l’on considère intéressantes selon les différentes caractéristiques d’une trajectoire (répétition sémantique des déplacements, répétition géographique des déplacements, etc.). Le défi de cette thèse est la classification des ZOI en fonction de leur typologie métier (Terminal à conteneur, Dépôt de stockage de conteneurs vides, Terminal ferroviaire, Site de chargements/déchargements etc.) en utilisant notamment les connaissances issues du sursol.

4 Planning du projet

- Mois 1-9 : Familiarisation avec les données issues des conteneurs intelligents, et mise en place de procédures de nettoyage des données. Consolidation de l'état de l'art sur les modèles de représentations des trajectoires, les algorithmes de classification de trajectoires et de détection des zones d'intérêt.
- Mois 10-20 : Spécification d'un algorithme de détection automatique des zones d'intérêt à partir d'un ensemble d'informations contextuelles telles que les points d'arrêt et les informations métier (proximité d'un port, stations ferroviaires, densité des positions de conteneur, etc.)
- Mois 21-30 : Mise en place d'un algorithme de classification des ZOI en fonction de leur typologie (Terminal à conteneur, Dépôt de stockage de conteneurs vides, Terminal ferroviaire, Site de chargements/déchargements etc.) en utilisant les connaissances issues du sursol.
- Mois 12 - 36 : Rédaction d'au moins deux publications scientifiques dans des conférences et/ou revues de renom
- Mois 24 - 36 : Rédaction du rapport de thèse, et préparation de la soutenance

5 Informations complémentaires

Modalités :

Financement : Bourse Projet TNTM

Laboratoire : Laboratoire d'Informatique et Systèmes (LIS)

Équipe d'accueil : [DIAMS](#)

Contacts :

- Etienne Thuillier (etienne.thuillier@univ-amu.fr)
- Sana Sellami (sana.sellami@univ-amu.fr)
- Omar Boucelma (omar.boucelma@univ-amu.fr)

Localisation : Site de St Jérôme

Date de début : Septembre 2022

Salaire net mensuel : 1421.39 euros

Profil du candidat :

- Titulaire d'un master 2 ou diplôme d'ingénieur en informatique
- Solides compétences en apprentissage automatique (machine learning)
- Solides compétences en programmation objet et bases de données (python est un plus)
- Bonne expérience en statistiques
- Connaissances sur les SIG
- Bonne communication orale et écrite en français et anglais

Candidature :

Les dossiers de candidatures doivent comprendre un CV, une lettre de motivation, les deux derniers relevés de notes et des lettres de recommandation.

La date limite de candidature est fixée au **05/06/2022**.

References

- [CJW14] Wen Chen, Minhe Ji, and Jianmei Wang. T-dbscan: A spatiotemporal density clustering for gps trajectory segmentation. *International Journal of Online and Biomedical Engineering (iJOE)*, 10(6):pp. 19–24, Oct. 2014.
- [CLY⁺20] Xinqiang Chen, Jun Ling, Yongsheng Yang, Hailin Zheng, Pengwen Xiong, Octavian Postolache, and Yong Xiong. Ship trajectory reconstruction from ais sensory data via data quality control and prediction. *Mathematical Problems in Engineering*, 2020:1–9, 08 2020.
- [CSF⁺21] Cécile Cayère, Christian Sallaberry, Cyril Faucher, Marie-Noelle Bessagnet, and Philippe Roose. Proposition d’un modèle de trajectoires multi-aspects et multi-niveaux appliqué au tourisme. In Maxime Lefrançois, editor, *Journées Francophones d’Ingénierie des Connaissances (IC) Plate-Forme Intelligence Artificielle (PFIA’21)*, pages pp 56–64, Bordeaux, France, 2021.
- [dSMBA⁺] Ronaldo dos Santos Mello, Vania Bogorny, Luis Otávio Alvares, Luiz Henrique Zambom Santana, Carlos Andres Ferrero, Angelo Augusto Frozza, Geomar Andre Schreiner, and Chiara Renso. MASTER: A multiple aspects view on trajectories. In Maristella Agosti, Maurizio Atzori, Paolo Ciaccia, and Letizia Tanca, editors, *Proceedings of the 28th Italian Symposium on Advanced Database Systems, Villasimius, Sud Sardegna, Italy (virtual due to Covid-19 pandemic), June 21-24, 2020*, volume 2646 of *CEUR Workshop Proceedings*, pages 136–143. CEUR-WS.org.
- [GGCSNM18] Christian Eduardo Verdonk Gallego, Victor Fernando Gómez Comendador, Francisco Javier Saez Nieto, and Miguel Garcia Martinez. Discussion on density-based clustering methods applied for automated identification of airspace flows. In *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, pages 1–10, 2018.
- [LHLG08] Jae-Gil Lee, Jiawei Han, Xiaolei Li, and Hector Gonzalez. *TraClass*: trajectory classification using hierarchical region-based and trajectory-based clustering. *Proc. VLDB Endow.*, 1(1):1081–1094, 2008.
- [LHW07] Jae-Gil Lee, Jiawei Han, and Kyu-Young Whang. Trajectory clustering: A partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, SIGMOD ’07, page 593–604, New York, NY, USA, 2007. Association for Computing Machinery.
- [PSR⁺13] Christine Parent, Stefano Spaccapietra, Chiara Renso, Gennady L. Andrienko, Natalia V. Andrienko, Vania Bogorny, Maria Luisa Damiani, Aris Gkoulalas-Divanis, José Antônio Fernandes de Macêdo, Nikos Pelekis, Yannis Theodoridis, and Zhixian Yan. Semantic trajectories modeling and analysis. *ACM Comput. Surv.*, 45(4):42:1–42:32, 2013.
- [WCCM21] Lianhui Wang, Pengfei Chen, Linying Chen, and Junmin Mou. Ship ais trajectory clustering: An hdbscan-based approach. *Journal of Marine Science and Engineering*, 9(6), 2021.
- [ZNX20] Zhiyuan Zhang, Guoxin Ni, and Yanguo Xu. Comparison of trajectory clustering methods based on k-means and dbscan. In *2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, volume 1, pages 557–561, 2020.