



Avis de Soutenance

Monsieur Weizhu QIAN

Présentera ses travaux en soutenance

Soutenance prévue le **lundi 07 décembre 2020** à 10h00
Lieu : UTBM Site de Belfort 12 rue Thierry Mieg 90010 Belfort
Salle : I102

Titre des travaux : Détermination des déplacements humains à partir des données collectées par des smartphones: modèles probabilistes et algorithmes d'apprentissage.

Ecole doctorale : SPIM - Sciences Physiques pour l'Ingénieur et Microtechniques

Section CNU : 27

Unité de recherche : CIAD - Connaissance et Intelligence Artificielle Distribuées

Directeur de thèse : Franck GECHTER

Codirecteur de thèse : HDR NON HDR

Soutenance : Publique A huis clos

Membres du jury :

<u>Nom</u>	<u>Qualité</u>	<u>Etablissement</u>	<u>Rôle</u>
M. Franck GECHTER	Maître de conférences HDR	Université Bourgogne - Franche-Comté	Directeur de thèse
M. Germain FORESTIER	Professeur	Université de Haute-Alsace	Rapporteur
M. Yacine OUZROUT	Professeur	Université Lumière Lyon 2	Rapporteur
M. Sidi-Mohammed SENOUCI	Professeur	Université de Bourgogne	Examineur
M. Ye-Qiong SONG	Professeur	Université de Lorraine	Examineur
M. Fabrice LAURI	Maître de conférences	Université de Technologie de Belfort Montbéliard	Examineur

Mots-clés : Machine Learning, Deep Learning, Probabilistic Models, Variational Inference, Mobile Data,

Résumé de la thèse (en français) :

Les données d'utilisation des smartphones peuvent être utilisées pour étudier la mobilité humaine que ce soit en environnement extérieur ouvert ou à l'intérieur de bâtiments. Dans ce travail, nous étudions ces deux aspects de la mobilité humaine en proposant des algorithmes de machine learning adaptés aux sources d'information disponibles dans chacun des contextes. Pour l'étude de la mobilité en environnement extérieur, nous utilisons les données de coordonnées GPS collectées pour découvrir les schémas de mobilité quotidiens des utilisateurs. Pour cela, nous proposons un algorithme de clustering automatique utilisant le Dirichlet process Gaussian mixture model (DPGMM) afin de regrouper les trajectoires GPS quotidiennes. Cette méthode de segmentation est basée sur l'estimation des densités de probabilité des trajectoires, ce qui atténue les problèmes causés par le bruit des données. Concernant l'étude de la mobilité humaine dans les bâtiments, nous utilisons les données d'empreintes digitales WiFi collectées par les smartphones. Afin de prédire la trajectoire d'un individu à l'intérieur d'un bâtiment, nous avons conçu un modèle hybride d'apprentissage profond, appelé convolutional mixture density recurrent neural network (CMDRNN), qui combine les avantages de différents réseaux de neurones profonds multiples. De plus, en ce qui concerne la localisation précise en intérieur, nous supposons qu'il existe une distribution latente régissant l'entrée et la sortie en même temps. Sur la base de cette hypothèse, nous avons développé un modèle d'apprentissage semi-supervisé basé sur le variational autoencoder (VAE). Dans la procédure d'apprentissage non supervisé, nous utilisons un modèle VAE pour apprendre une distribution latente de l'entrée qui est composée de données d'empreintes digitales WiFi. Dans la procédure d'apprentissage supervisé, nous utilisons un réseau de neurones pour calculer la cible, coordonnées par l'utilisateur. De plus, sur la base de la même hypothèse utilisée dans le modèle d'apprentissage semi-supervisé basé sur le VAE, nous exploitons la théorie des goulots d'étranglement de l'information pour concevoir un modèle basé sur le variational information bottleneck (VIB). Il s'agit d'un modèle d'apprentissage en profondeur de bout en bout plus facile à former et offrant de meilleures performances. Enfin, les méthodes proposées ont été validées sur plusieurs jeux de données publics acquis en situation réelle. Les résultats obtenus ont permis de vérifier l'efficacité de nos méthodes par rapport à l'existant.

Abstract (in English):

Smartphone usage data can be used to study human indoor and outdoor mobility. In our work, we investigate both aspects in proposing machine learning-based algorithms adapted to the different information sources that can be collected. In terms of outdoor mobility, we use the collected GPS coordinate data to discover the daily mobility patterns of the users. To this end, we propose an automatic clustering algorithm using the Dirichlet process Gaussian mixture model (DPGMM) so as to cluster the daily GPS trajectories. This clustering method is based on estimating probability densities of the trajectories, which alleviate the problems caused by the data noise. By contrast, we utilize the collected WiFi fingerprint data to study indoor human mobility. In order to predict the indoor user location at the next time points, we devise a hybrid deep learning model, called the convolutional mixture density recurrent neural network (CMDRNN), which combines the advantages of different multiple deep neural networks. Moreover, as for accurate indoor location recognition, we presume that there exists a latent distribution governing the input and output at the same time. Based on this assumption, we develop a variational auto-encoder (VAE)-based semi-supervised learning model. In the unsupervised learning procedure, we employ a VAE model to learn a latent distribution of the input, the WiFi fingerprint data. In the supervised learning procedure, we use a neural network to compute the target, the user coordinates. Furthermore, based on the same assumption used in the VAE-based semi-supervised learning model, we leverage the information bottleneck theory to devise a variational information bottleneck (VIB)-based model. This is an end-to-end deep learning model which is easier to train and has better performance. Finally, we validate the proposed methods on several public real-world datasets providing thus results that verify the efficiencies of our methods as compared to other existing methods generally used.