

Proposition de Sujet de Thèse pour Contrat Doctoral UCA

Adresse e-mail à utiliser pour toute correspondance :

paola.goatin@inria.fr

Titre de la thèse

Analyse de données et modélisation du trafic routier

Thesis Title

Big Data Analysis and Modeling of Road Traffic

Directeur de Thèse (HDR ou assimilé)

Nom : Goatin

Prénom : Paola

Téléphone : +33492387834

Courriel : paola.goatin@inria.fr

Laboratoire d'accueil

INRIA-ACUMES

Co-directeur

Nom : Binois

Prénom : Mickael

HDR : Non

Unité de recherche : INRIA-ACUMES

Téléphone :

Courriel : mickael.binois@inria.fr

Domaine Scientifique

DS1 - Mathématiques et leurs Interactions

Description du sujet

À l'échelle macroscopique, le trafic est généralement représenté comme un fluide s'écoulant sur le réseau routier. Afin de caractériser l'évolution de ce système, nous définissons des quantités agrégées telles que le débit ou la densité de véhicules. L'enjeu est de pouvoir comprendre, reproduire et anticiper l'évolution de la densité et du débit dans l'espace et dans le temps, en s'appuyant à la fois sur cette modélisation mathématique et sur des données de trafic.

Il existe actuellement diverses sources de données de trafic routier (boucles magnétiques, caméras, GPS, Bluetooth, etc.), qui peuvent être utilisées pour la calibration du modèle mathématique. C'est-à-dire, identifier des valeurs plausibles pour les paramètres (éventuellement non physiques) du modèle de trafic. En pratique, l'hétérogénéité du trafic routier dans les régimes congestionnés rend difficile une simulation satisfaisante de la réalité, empêchant ainsi d'obtenir des prévisions fiables de l'état du trafic au-delà d'horizons courts (15-30 min). En effet, les modèles de trafic reposent généralement sur un choix unique du diagramme fondamental (relation débit-densité), indépendamment de l'heure et du jour de la semaine. Cependant, les paramètres du modèle peuvent dépendre du temps, car certaines conditions de circulation et le comportement des conducteurs peuvent être corrélés à des plages horaires spécifiques. En outre, les techniques d'analyse de données peuvent aider à ouvrir de nouvelles perspectives en termes d'interprétation de ces données mais aussi de modélisation du trafic routier.

Dans le cadre d'une collaboration entre l'équipe-projet ACUMES d'Inria Sophia Antipolis et le laboratoire LJAD de Nice, nous souhaitons analyser les informations fournies par les données de trafic à l'aide de méthodes novatrices d'apprentissage automatique et les exploiter à l'aide de modèles mathématiques déterministes à base d'équations aux dérivées partielles. Les méthodes statistiques ciblées reposent notamment sur le co-clustering et la modélisation par processus gaussiens. Les données de trafic seront considérées dans ce contexte comme des fonctions temporelles et les méthodes statistiques identifieront des groupes de boucles homogènes (types de réseau routier) ainsi que des groupes d'intervalles temporels (jours de la semaine). Cette analyse pourrait nous permettre d'affiner les modèles mathématiques développés par ACUMES en identifiant différentes journées standard ou une typologie spatiale de réseau. L'analyse devrait nous permettre également de calibrer plus efficacement certains paramètres fondamentaux de nos modèles.

Dans cette perspective, notre objectif est de calibrer et de valider des modèles macroscopiques de premier et second ordre sur des ensembles de données appropriés. L'étude sera basée sur les étapes suivantes:

* Préparation des données. Avant l'analyse, les données doivent être triées par capteur et par direction, puis agrégées sur des périodes de temps (généralement 1 à 6 minutes) pour obtenir des quantités moyennes, mesurant la densité, la vitesse moyenne et le débit.

* Calibration du modèle. Les paramètres du modèle, liés en particulier au choix de la fonction de vitesse moyenne, seront calibrés sur un sous-ensemble de l'ensemble de données, en distinguant éventuellement les jours de la semaine et les créneaux horaires. Cela nécessitera de développer de nouvelles méthodes de calibration pilotées par les

données, faisant appel à des algorithmes d'apprentissage et capables de gérer la nature fonctionnelle de ces données.

* Validation du modèle. Les paramètres résultants seront utilisés pour exécuter des simulations qui seront comparées à l'évolution du trafic décrite par des données réelles appartenant à un sous-ensemble de test, différent de l'ensemble de données d'apprentissage.

Les travaux devraient donner lieu à plusieurs publications dans des revues internationales et à une ou deux contributions logicielles (pour Matlab, R ou Python).

Description of the thesis

At a macroscopic scale, traffic is usually represented as a fluid flowing on the road network. In order to characterize the evolution of the system, we define aggregated quantities such as the flow or density of vehicles. The challenge is to be able to understand, reproduce and anticipate the evolution of density and flux in space and time, relying both on this mathematical modeling and on traffic data.

There exists nowadays a variety of road traffic data sources (magnetic loop detectors, video cameras, floating car data, Bluetooth, etc), which can be used for model calibration. That is, finding plausible values for parameters (possibly non physical) of the model of traffic. In particular, the heterogeneity of traffic conditions in congested regimes makes it hard to obtain a good matching between simulations and reality, thus preventing from obtaining reliable traffic state predictions beyond short time horizons (15-30 min). Indeed, traffic models are usually set on a unique choice of the fundamental diagram (flux-density relation), independently of the time of the day and the day of the week. Yet, model parameters may be time dependent, as some traffic conditions and driver behavior may be correlated to specific time ranges. Besides, data analysis techniques may help in opening new perspectives in the interpretation of data and the modeling of road traffic.

As part of a collaboration between the ACUMES project-team at Inria Sophia Antipolis and the LJAD laboratory in Nice, we would like to analyze information derived from traffic data using innovative machine learning methods and exploit them within deterministic PDE models. The targeted statistical methods rely in particular on model-based functional co-clustering and Gaussian process modeling. The traffic data will be seen in this context as time functions and the statistical methods will identify groups of homogeneous loops (road network types) as well as groups of time moments (days of the week). This analysis could allow us to refine the mathematical models developed by ACUMES by identifying different standard days or a spatial typology of networks. The analysis would also allow us to calibrate certain fundamental parameters of our models.

In this perspective, we aim to calibrate and validate first and second order macroscopic traffic flow models on suitable data-sets. The study will be based on the following steps:

* Data preparation. Before analysis, data need to be sorted by sensor and direction, and aggregated over time periods (usually 1 or 6 minutes) to obtain averaged quantities measuring density, mean velocity and flow.

* Model calibration. The model parameters, in particular related to the choice of the mean speed function, will be calibrated on a subset of the data-set, possibly distinguishing

week days and time slots. It will require to develop new data-driven calibration methods, involving learning algorithms, and able to cope with the functional nature of the data.

* Model validation. The resulting parameters will be used to run simulations that will be compared to traffic evolution described by real data belonging to a different subset of the available data-set.

It is expected that the work will yield several publications in international journals and one or two software packages (for Matlab, R or Python).

Informations complémentaires

Le/la doctorant(e) sera basé(e) à INRIA Sophia-Antipolis, sous la supervision conjointe des Drs Paola Goatin et Mickaël Binois (INRIA) et du Pr Charles Bouveyron (UCA).

The Ph.D. student will be based at INRIA Sophia-Antipolis, under the joint supervision of Drs Paola Goatin and Mickael Binois (INRIA) and Prof. Charles Bouveyron (UCA).