

Clustering automatique de trajectoires avec des HMMs

Application à la sécurité du trafic routier

N. Saunier, en collaboration avec T. Sayed

Department of Civil Engineering
University of British Columbia
Vancouver

saunier@civil.ubc.ca
saunier@aist.enst.fr

Séminaire Machine Learning ENST
09 / 01 / 2006

Outline

- 1 Introduction
- 2 Apprentissage semi-supervisé de conflits de trafic
- 3 Résultats expérimentaux

Outline

- 1 Introduction
- 2 Apprentissage semi-supervisé de conflits de trafic
- 3 Résultats expérimentaux

Diagnostic de sécurité routière

- Approche traditionnelle réactive : données historiques d'accidents (traitement de points noirs).
- Approche proactive : "Don't wait for accident to happen" (ICTCT).
- Besoin de données complémentaires, faciles à recueillir, basées sur des événements du trafic plus fréquents et en rapport avec la sécurité (accidents).

Les conflits de trafic

Definition

"A conflict is an observational situation in which two or more road users approach each other in space and time to such an extent that a collision is imminent if their movements remain unchanged." (1st International Traffic Conflicts Workshop, 1977)

- Etude de la fréquence et des caractéristiques des conflits sur une intersection par une équipe d'observateurs pendant plusieurs heures.
- Ensemble de méthodes pour le recueil et l'analyse de données de conflits de trafic.

Automatiser les méthodes

- Les TCTs constituent le cadre de nos méthodes de diagnostic détaillé de la sécurité.
- Limitations des TCTs :
 - Coût de recueil des données.
 - Subjectivité des observateurs.
- Objectif : système robuste et générique de détection des conflits de trafic dans des données vidéo,
 - à partir des trajectoires (bruitées) des véhicules.

Apprentissage supervisé

- Beaucoup de systèmes à base de règles, fondés sur une analyse manuelle de la tâche :
 - Problème de généralité de la méthode et des modèles appris.
 - Données (positions et vitesses) trop bruitées pour une analyse cinématique directe.
- Besoin de méthodes d'apprentissage.
 - Exemples : modèles de Markov cachés (HMM), réseaux Bayésiens dynamiques (DBN).

Des événements rares

- Difficulté de rassembler assez de données étiquetées pour faire de l'apprentissage.
- Complexité des événements rares : plusieurs classes, nécessitant d'autant plus de données.
- Abondance de données vidéo, qui ne peuvent être traitées par des êtres humains.

Apprentissage semi-supervisé

- Besoin de méthodes non-supervisées pour explorer l'espace des données.
- Mais supervision indispensable pour détecter les conflits de trafic parmi toutes les interactions du trafic.
- Contributions :
 - Une méthode de clustering de type k-moyennes utilisant des HMMs.
 - Une heuristique simple pour déterminer le nombre de classes.
 - Une méthode de détection des conflits de trafic.

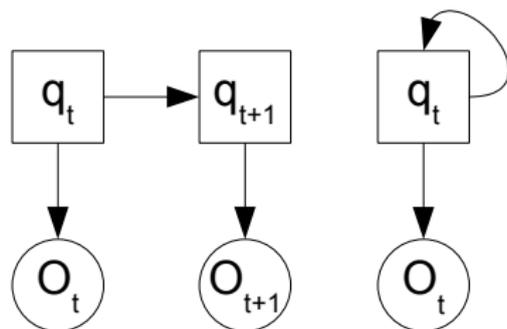
Outline

- 1 Introduction
- 2 Apprentissage semi-supervisé de conflits de trafic**
- 3 Résultats expérimentaux

Clustering de données séquentielles

- Comparaison des deux séquences comme un processus de transformation d'une séquence en une autre.
 - Exemple : distance d'édition pour la reconnaissance de caractères manuscrits.
- Clustering indirect de séquences, par extraction d'un ensemble de caractéristiques, pour revenir au cadre classique de vecteurs de taille fixe.
- Clustering statistique de séquences, utilisant des modèles comme les HMMs et les DBNs pour décrire la dynamique de chaque classe de séquences.
 - Des séquences sont similaires si elles sont produites par le même modèle.

Modèles de Markov cachés



- Probabilités a priori : $\pi_j = P(q_1 = S_j)$.
- Probabilités de transition : $P(q_{t+1} = S_j | q_t = S_i)$.
- Distribution des observations :
 $b_j(o_t) = P(O_t = o_t | q_t = S_j) = \mathcal{N}(O_t, \mu_j, \Sigma_j)$.

Sélection de modèle

- Comment déterminer le nombre de classes ?
- Optimisation de critères comme le "critère d'inférence Bayésienne" (BIC) ou la longueur de description minimale (MDL).
- Heuristique : élimination des classes auxquelles sont assignées trop peu d'exemples.

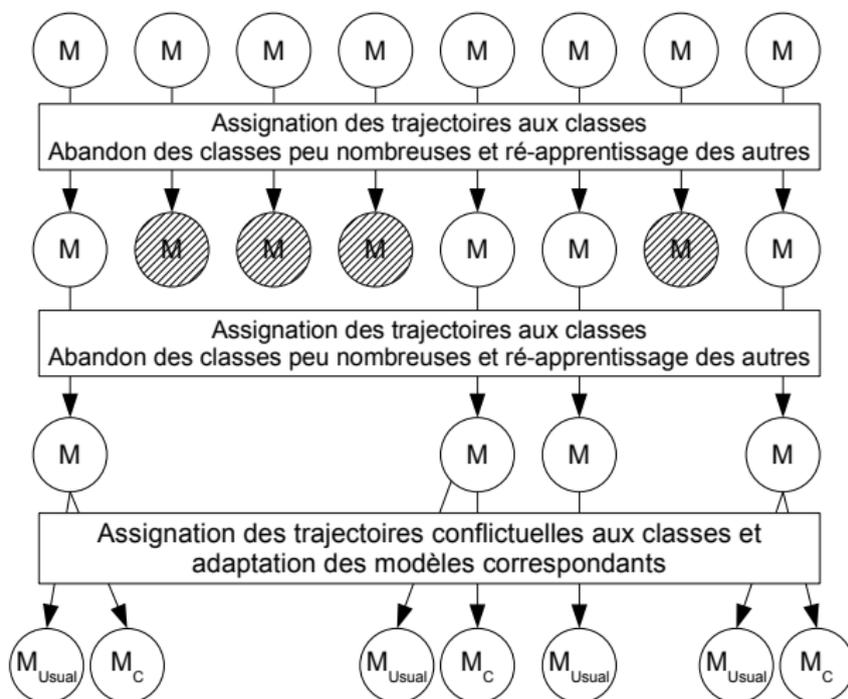
Algorithme de clustering de données séquentielles

- 1 Initialisation aléatoire d'un ensemble de $K = K_i$ HMMs (nombre de classes).
- 2 Assignation des exemples aux HMMs ("tout ou rien")
 $\text{argmax}_m P(O|\lambda_m)$.
- 3 Apprentissage de nouveaux HMMs avec les exemples de chaque classe (si le nombre d'exemples est suffisant).
- 4 Retour à 2 (si le nombre de classes K a changé).

Détection des conflits

- Observations constituées de la position et des vecteurs-vitesses à chaque instant.
- Détection des interactions, pré-conditions des conflits de trafic.
- Utilisation d'exemples de conflits de trafic pour identifier les trajectoires "conflictuelles".
- Adaptation des HMMs des classes auxquelles appartiennent les trajectoires conflictuelles :
 - Adaptation des paramètres des Gaussiennes, moyennes et co-variances.

Schéma de l'algorithme

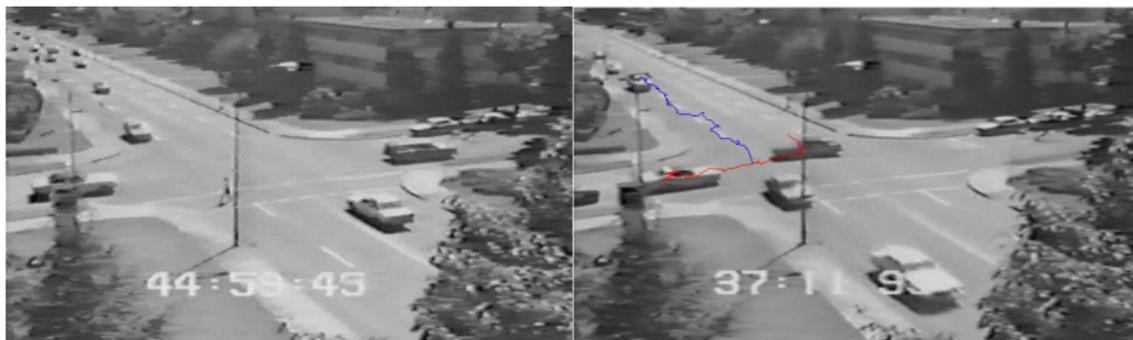


Outline

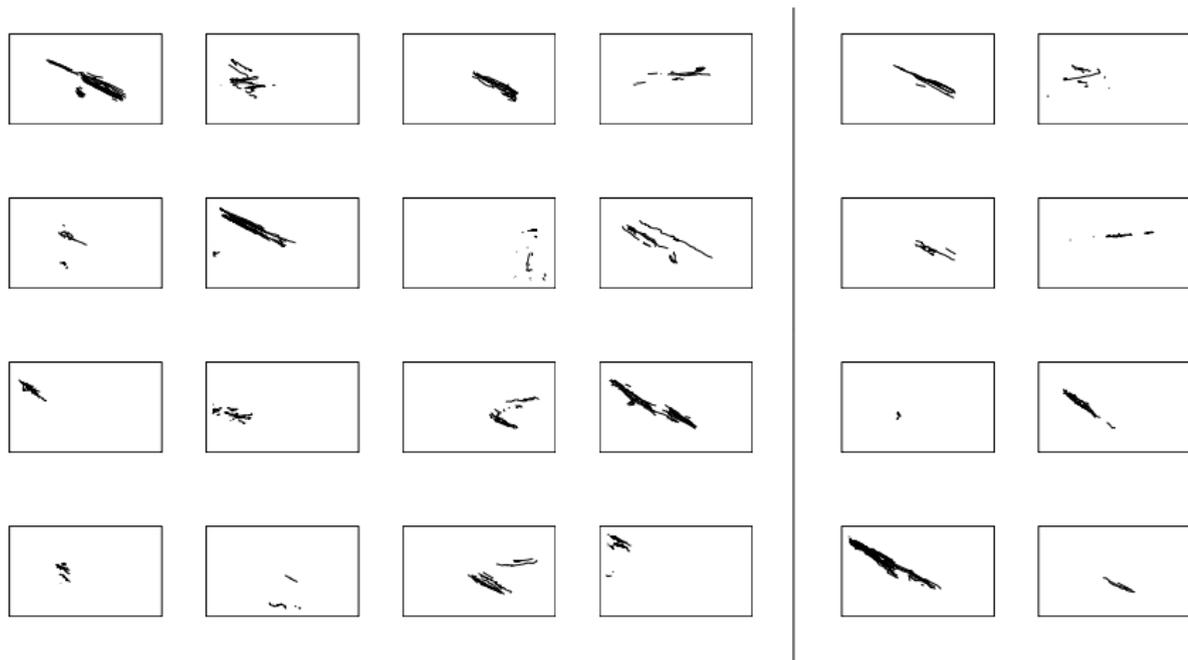
- 1 Introduction
- 2 Apprentissage semi-supervisé de conflits de trafic
- 3 Résultats expérimentaux**

Les données

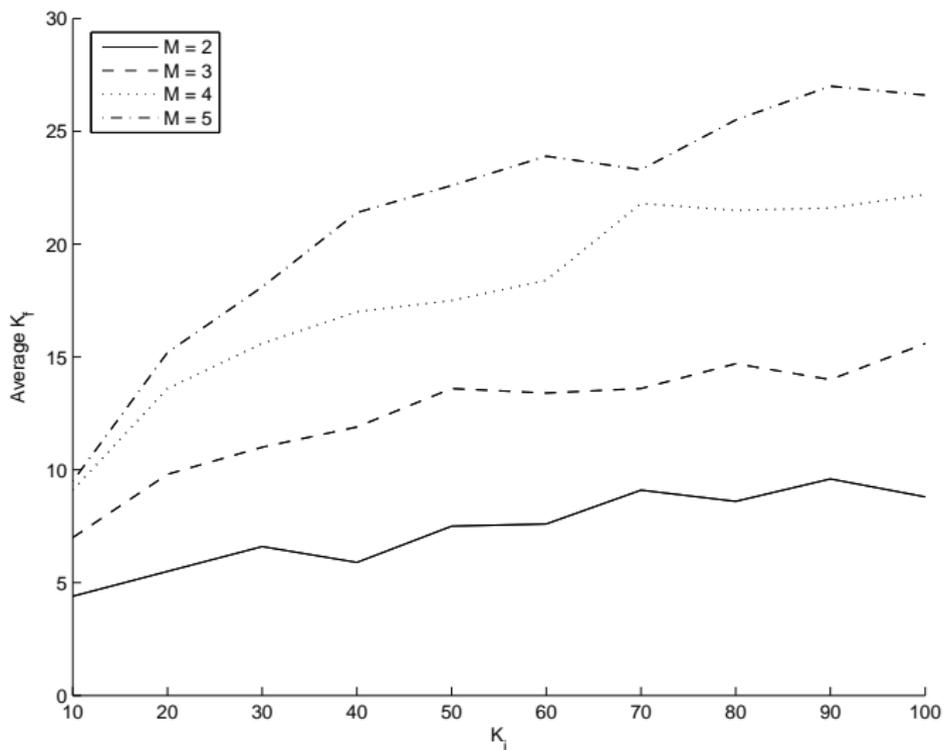
- Séquences vidéo d'entraînement des observateurs de conflits de trafic.
- 10 séquences sur un même site, comprenant 5 conflits de trafic identifiés et "utilisables".
- 560 trajectoires dans 8 séquences.



Exemples de partition de l'ensemble des trajectoires



Etude du nombre de classes en fonction de K_i



Résultats en détection

Adaptation	Corrects	Incertains	Fausse Alarmes
pas d'adaptation	10	17	38
$\alpha = 0.05$	10	13	6
$\alpha = 0.10$	10	13	10
$\alpha = 0.15$	10	12	6
$\alpha = 0.20$	10	3	3
$\alpha = 0.25$	10	5	2
$\alpha = 0.30$	10	5	2
$\alpha = 0.35$	10	4	1
$\alpha = 0.40$	10	4	0
$\alpha = 0.45$	10	4	0
$\alpha = 0.50$	10	3	0

Perspectives

- Données artificielles pour tester la méthode.
- Amélioration du suivi des véhicules, par utilisation des coordonnées réelles.
- Utilisation de DBNs pour modéliser les trajectoires.
- Algorithmes alternatifs de clustering des trajectoires, qui adapte au cours de l'apprentissage les classes aux exemples des conflits de trafic.