

Apprentissage incrémental par sélection de données dans un flux pour une application de sécurité routière

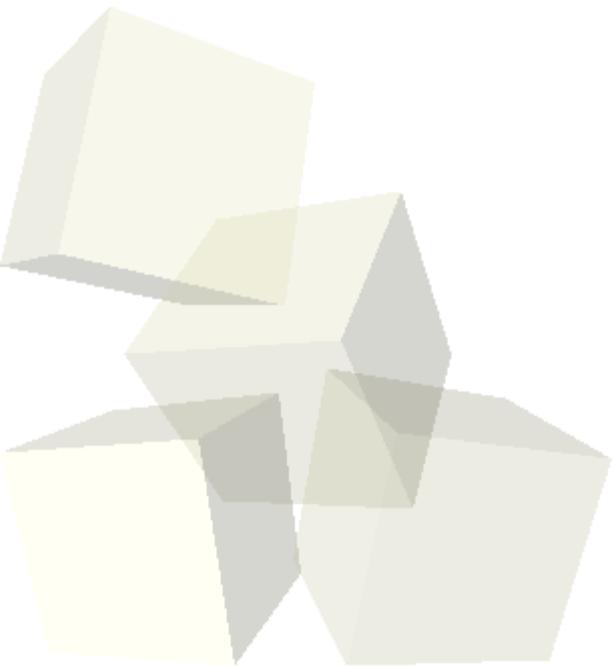
Nicolas Saunier *INRETS – Télécom Paris*
Sophie Midenet *INRETS*
Alain Grumbach *Télécom Paris*

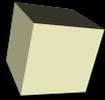
Conférence Cap 2004
14-16 juin 2004





- Objectif: application de sécurité routière.
- Le problème d'apprentissage.
- Les algorithmes.
- Résultats expérimentaux.



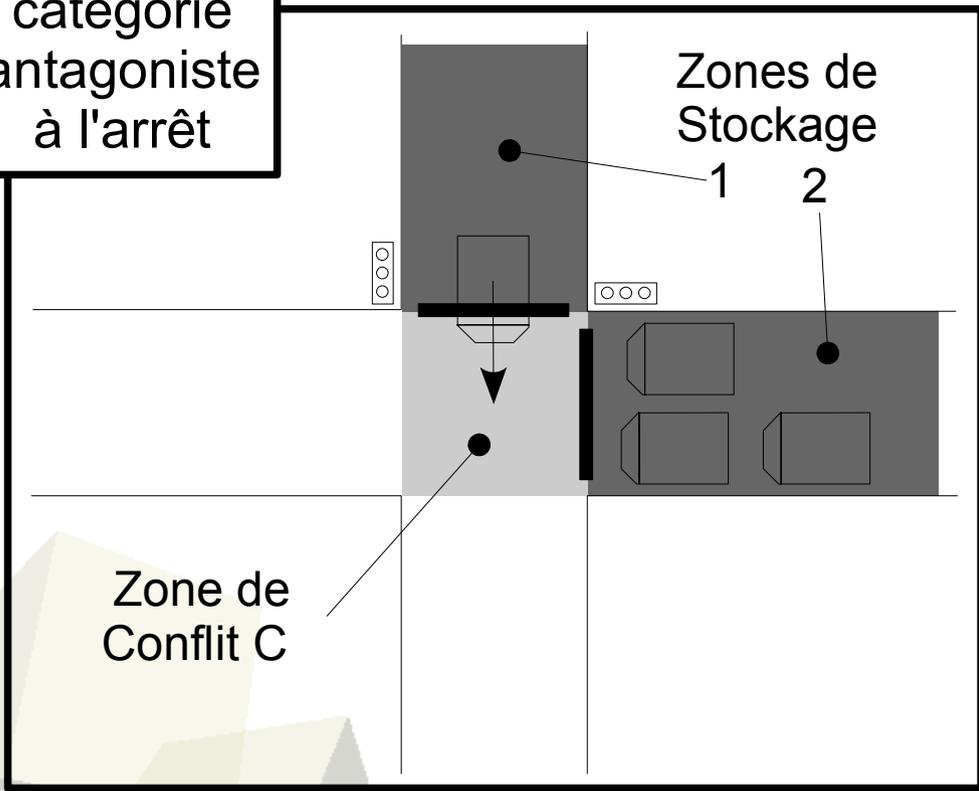


- Incidence de la régulation d'un carrefour à feux sur le comportement, la gêne et le risque subis par les usagers.
- Etude des interactions entre véhicules,
 - ◆ détection des interactions dans la zone de conflit,
 - ◆ évaluation de la sévérité: proximité spatio-temporelle de l'interaction à l'accident.
- Indicateurs de sévérité:
 - ◆ données difficiles à interpréter,
 - ◆ mais étiquetables: problème d'apprentissage.



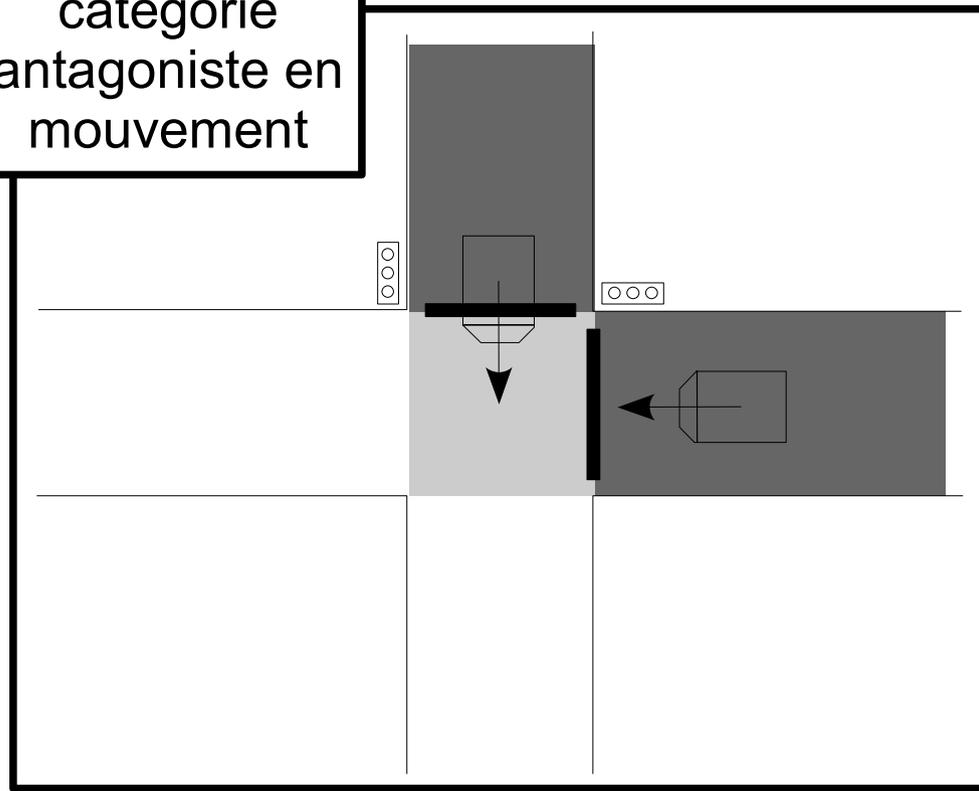
Exemples de catégories

catégorie antagoniste à l'arrêt



$SI \text{ mouvement}(C, 1 \rightarrow C) \cap \text{arrêt}(2)$
 ALORS interaction cat Antag. Arrêt

catégorie antagoniste en mouvement



$SI \text{ mouvement}(C, 1 \rightarrow C) \cap \text{mouvement}(2)$
 ALORS interaction cat Antag. Mvt

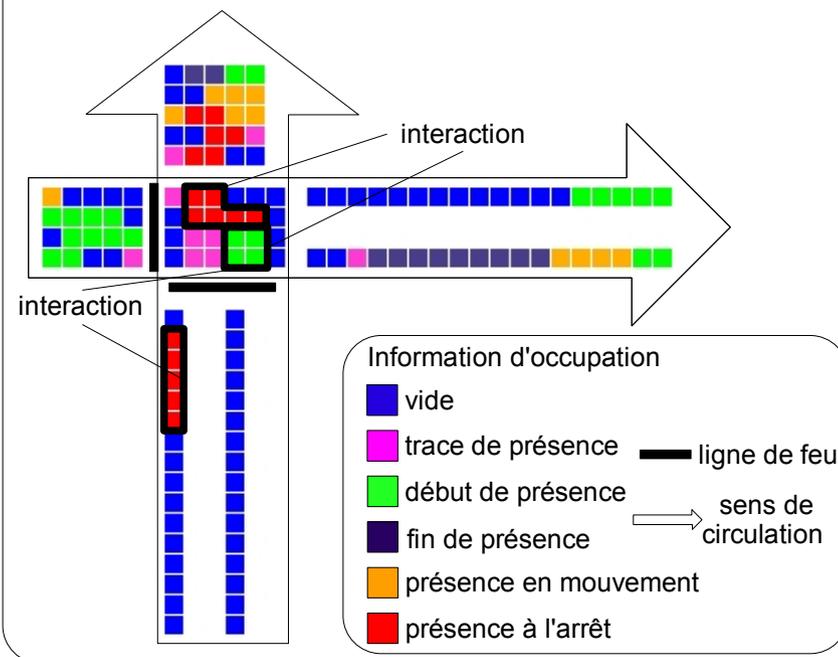


Le système expérimental

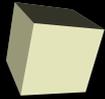
Un expert humain regarde la vidéo et estime la sévérité des interactions entre véhicules.



Les images produites par le traitement d'image sont utilisées pour l'application (1 image / seconde).



- Expérience sur site réel, pendant 8 mois.
- Données génériques, information dynamique.
- Accès séquentiel.



Le problème d'apprentissage

■ Caractéristiques:

- ♦ jugement expert: modélisation de l'incertitude par des classes floues (frontières graduelles),
- ♦ N classes et N-1 étiquettes "floues",
- ♦ proximité / chevauchement des classes,
- ♦ tailles déséquilibrées.

■ Problème de classification difficile: performances faibles avec apprentissage passif batch.





Apprentissage actif dans un flux

- initialisation: hypothèse h .
- pour tout exemple x_t , si *critère de sélection* vérifié
 - mise à jour de l'hypothèse h .
- jusqu'à *critère d'arrêt*.

■ Éléments centraux, à préciser:

- ♦ Critère de sélection,
- ♦ Critère d'arrêt et choix de l'hypothèse finale.

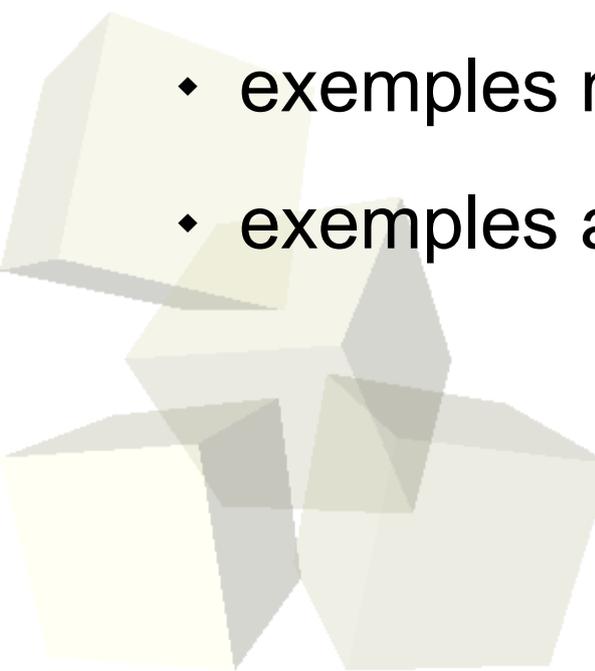


■ Sélection

- ♦ des exemples avant étiquetage: adaptation des critères existants ?
- ♦ des exemples après étiquetage: exemples mal-classés (Windowing). [Fürnkranz 98]

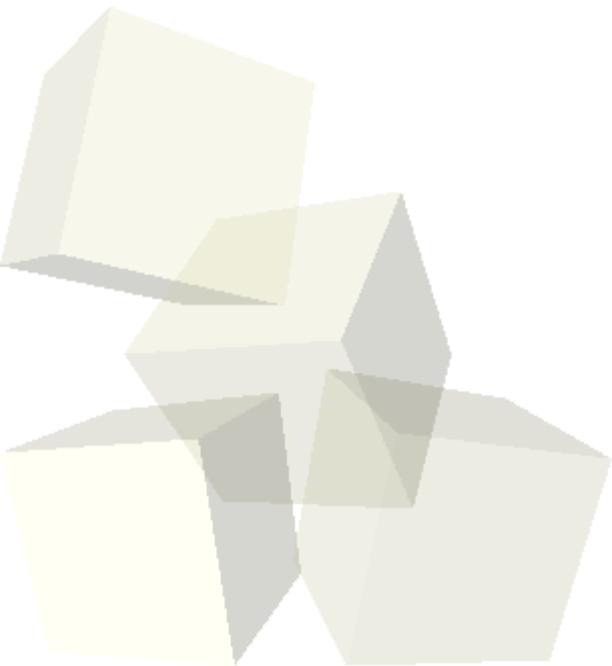
■ Etiquetage de tous les exemples,

- ♦ exemples mal-classés par l'hypothèse courante h ,
- ♦ exemples aux étiquettes floues non utilisés.





- Difficile: ensemble de validation.
- Amélioration de la qualité des hypothèses apprises (robustesse, stabilité),
 - ♦ combinaison d'hypothèses (Bagging, Boosting): Vote des derniers classifieurs appris.
 - ♦ paramètre: nombre de classifieurs combinés.



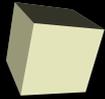


Résultats benchmarks

Base	Batch	MC	Nombre d'exemples sélectionnés
Soybean	93,9	90,2	93,0 / 596
Vote	90,3	95,2	23,8 / 390
Spambase	84,9	82,0	851,7 / 4139
Iris disc	96,0	93,3	17,3 / 132

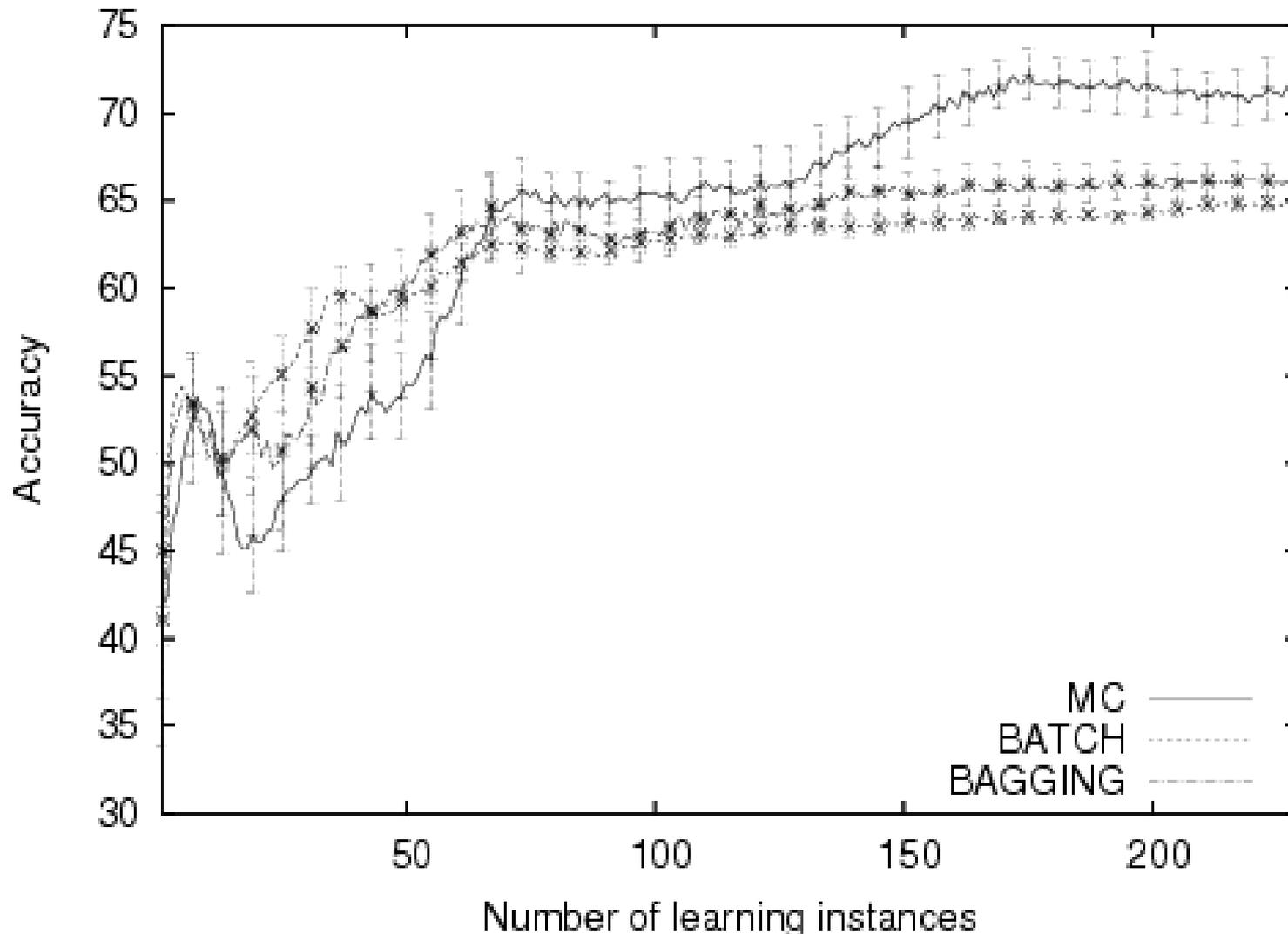
*classifieurs bayésiens naïfs
validation croisée en 10 ensembles.*

- Windowing dans un ordre fixé aléatoirement.



Résultats Sévérité (1/2)

Percentage of correctly classified instances



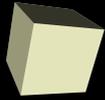
- Courbes d'apprentissage (moyennées sur 50 essais, n=7)
- MC (notre algorithme),
 - BATCH (apprentissage classique batch),
 - BAGGING (vote de plusieurs classifieurs appris sur des sous-ensembles tirés aléatoirement)

3 classes, classifieurs bayésiens naïfs

Initialisation avec 3 exemples tirés aléatoirement dans un autre ensemble.

52 minutes de flux: 828 exemples dans le flux de données.

4 x 10 minutes (2 conditions de trafic): 371 exemples en test.



Résultats Sévérité (2/2)

■ Performances finales:

		MC	BATCH	BAGGING	BATCH-EQ	BATCH-EQ-MC
MIN	Exemples bien classés	71,7 ± 1,6	64,9 ± 0,5	66,2 ± 1,0	64,3 ± 1,0	61,7 ± 1,4
	Exemples bien classés	78,3 ± 2,5	84,0 ± 1,0	82,2 ± 1,7	82,8 ± 2,0	83,8 ± 1,4
	Précision	75,4 ± 3,5	53,8 ± 1,0	58,1 ± 2,9	56,5 ± 2,6	50,2 ± 2,1
MOY	Exemples bien classés	71,2 ± 2,3	58,9 ± 0,5	60,8 ± 1,7	57,0 ± 1,5	52,8 ± 2,6
	Précision	78,2 ± 1,8	77,7 ± 0,4	77,8 ± 1,2	77,5 ± 0,9	78,3 ± 2,0
MAX	Exemples bien classés	68,5 ± 3,9	65,3 ± 0,9	67,2 ± 2,9	67,8 ± 1,7	66,1 ± 3,4
	Précision	59,2 ± 2,3	57,0 ± 0,7	56,8 ± 2,1	54,2 ± 1,4	53,1 ± 1,9



- Apprentissage prometteur dans le cadre incrémental.
- Influence initialisation.
- Perspectives:
 - ◆ combinaison intelligente d'hypothèses: on peut faire mieux qu'un Vote,
 - ◆ extension à des périodes de traitement plus longues pour le traitement de la base de données: détection de changements de concept, suivi des performances.

