Stream-Based Learning through Data Selection in a Road Safety Application

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

STAIRS 2004
23-24 juin 2004

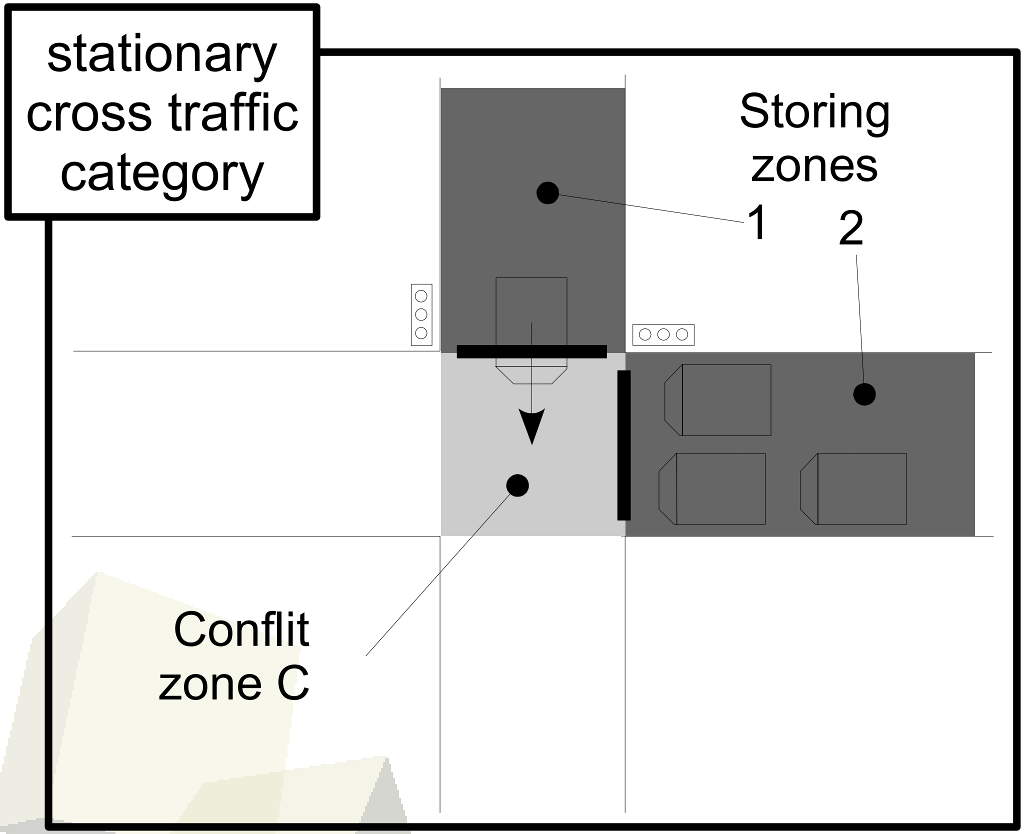
Outline

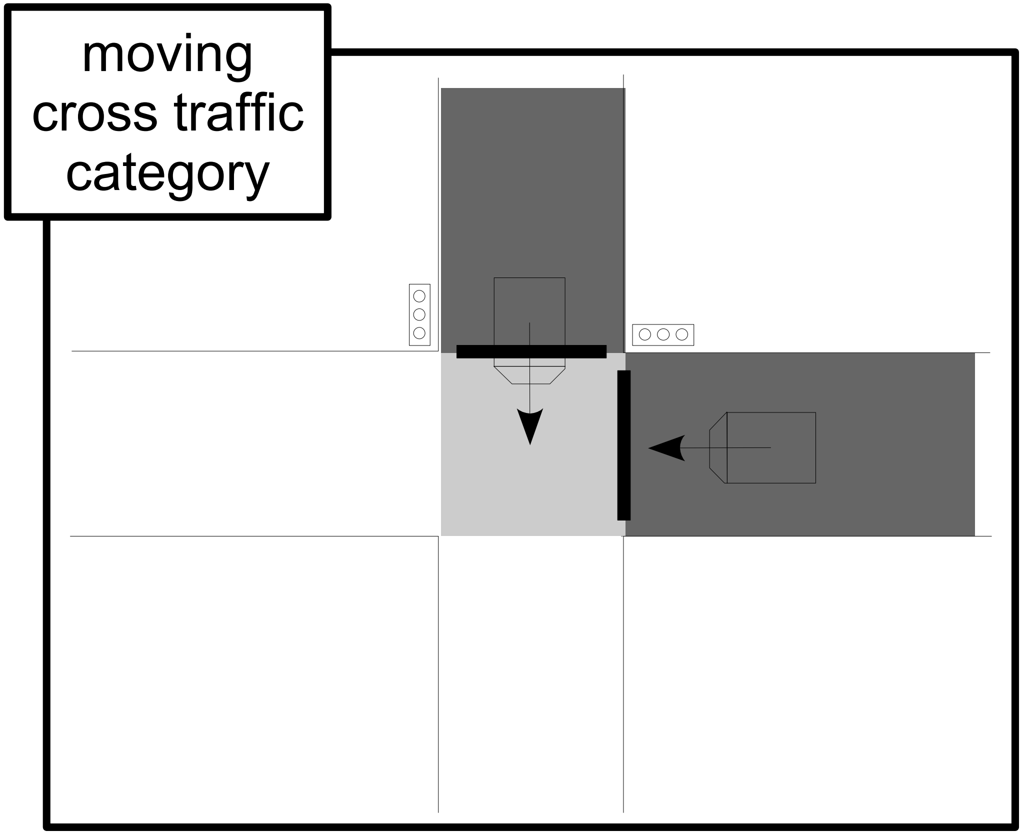
- Goal: road safety application.
- The learning problem.
- The algorithms.
- Experimental results.

Goals

- Consequences of the regulation in a signalized intersection on the behavior, the discomfort and the risk undergone by users.
- Study of vehicle interactions,
  - detections of interactions in the conflict zone,
  - severity evaluation: spatio-temporal distance between the interaction and the accident.
- Severity indicators,
  - difficult interpretation of the data,
  - labels can be obtained: learning problem.

Examples of interaction categories





IF movement(C, 1 → C) ∩ stationary(2)

THEN interaction (cat Stat. Cross)

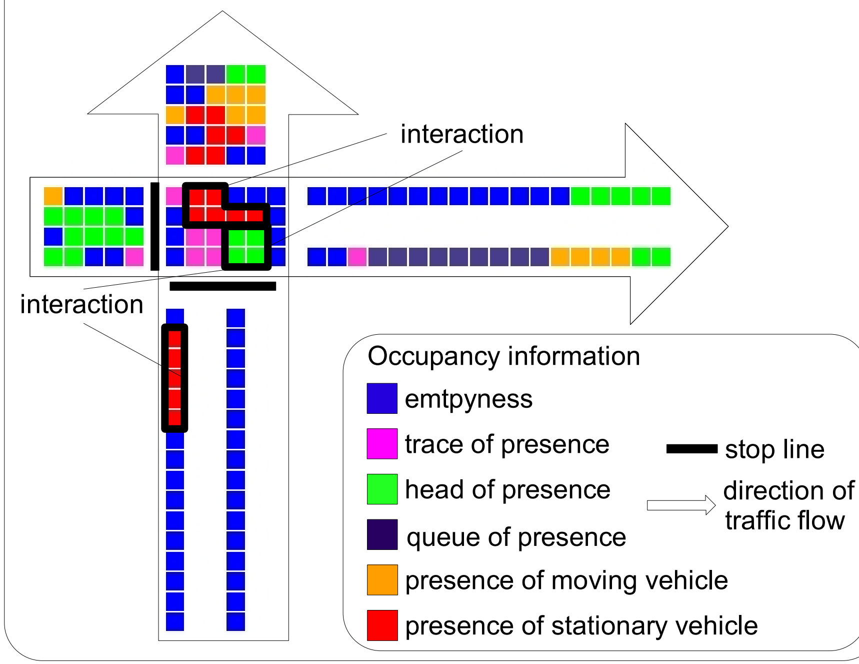
IF movement(C, 1 → C) ∩ movement(2)

THEN interaction (cat Moving Cross)

Learning the severity

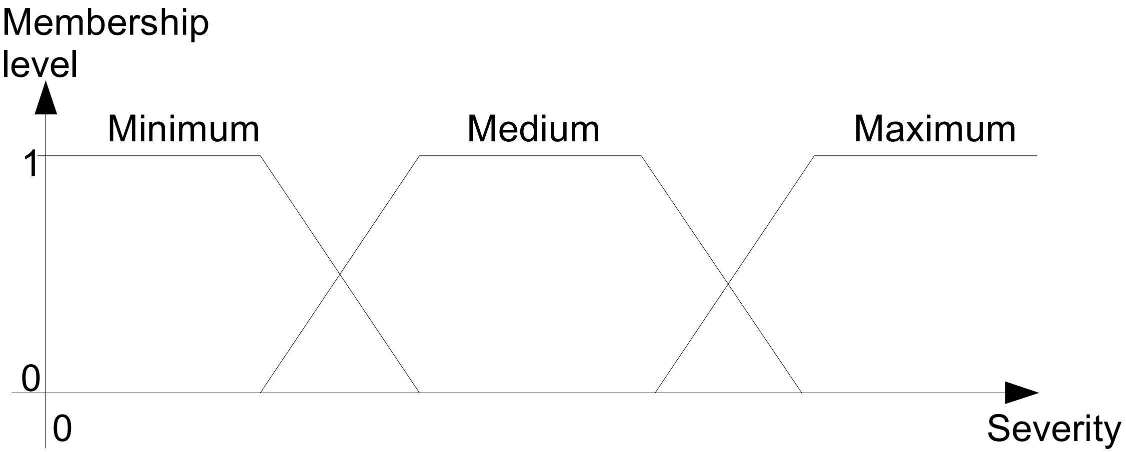
A human expert watches the video and estimates the severity of vehicle interactions.

The images resulting from video processing are used for the application.



- 8 months experiments on a real intersection.
- Multi-purpose data, dynamic information.
- Data + available labels = learning problem.

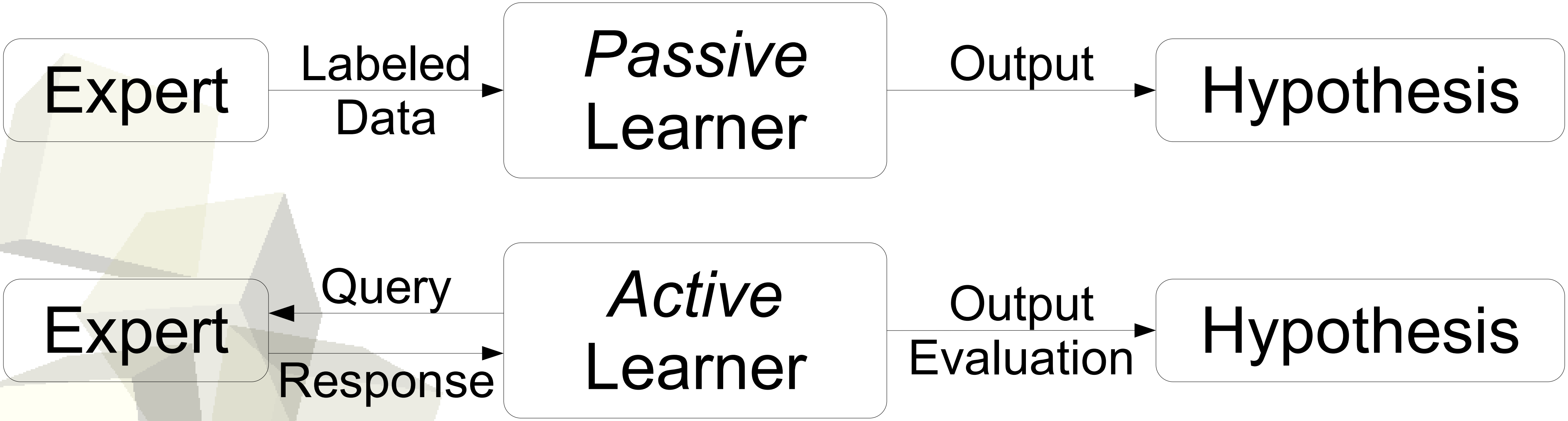
The learning problem



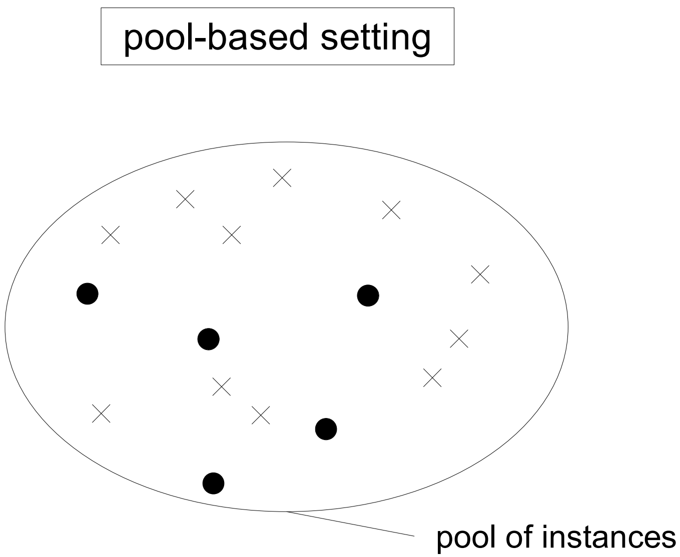
- Features:
  - sequential access,
  - expert judgement: model the uncertainty with fuzzy classes (progressive boundaries),
  - N classes and N-1 “fuzzy”,
  - closeness / overlapping of the classes,
  - unbalanced dataset.
- Difficult learning problem: poor performance with passive batch learning.

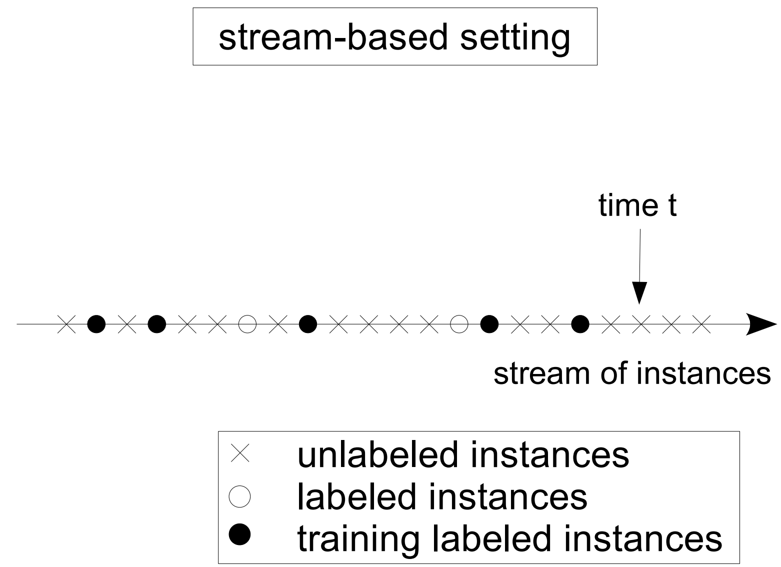
Ideas

- Incremental algorithm:
  - “intelligent” data selection of instances, in order to specify the boundaries: distortion of the real data distribution.
- Active learning:



Active learning





- Criterion for data selection:
  - uncertainty sampling,
  - query by comittee,
  - version space,
  - expected future error.

[Schohn et al. 2000, Tong 2001, Freund et al. 1997]

Generic algorithm

- initialization: hypothesis h.
- for each instance x_t, if *selection criterion* satisfied
  - update of hypothesis h.
- until *stopping criterion*.

- Main elements:
  - Selection criterion,
  - Stopping criterion and choice of the final hypothesis.

Selection criterion

- Selection
  - of unlabeled instances: adaptation of criteria used in the pool-based setting ?
  - of labeled instances: misclassified instances (Windowing). [Fürnkranz 98]
- Labeling of all instances,
  - misclassified instances by the current hypothesis h,
  - no use of fuzzy-labeled instances.

Stopping criterion

- Difficult to estimate the quality of the learnt hypotheses (validation set).
- Improvement of the quality of learnt hypotheses (robustness, stability),
  - combination of hypotheses (Bagging, Boosting): Vote of the last learnt hypotheses.
  - parameter: number of combined hypotheses.

Our algorithm (MC)

Let i be the number of selected instances,
Let h_i be the hypothesis learnt after the selection of i instances,
Let $Vote_{i,j}$ be the hypothesis obtained by taking majority vote over the hypotheses $\{h_k, i<k\leq j\}$.

- initialization: hypothesis h_0, $i=0$
- for each instance x_t, ask for its label y_t
- if (y_t is not fuzzy) and ($Vote_{max(0,i-n),i}(x_t) \neq y_t$)
  - update of hypothesis h_i in h_{i+1}
  - $i=i+1$
- while the expert is willing to label.

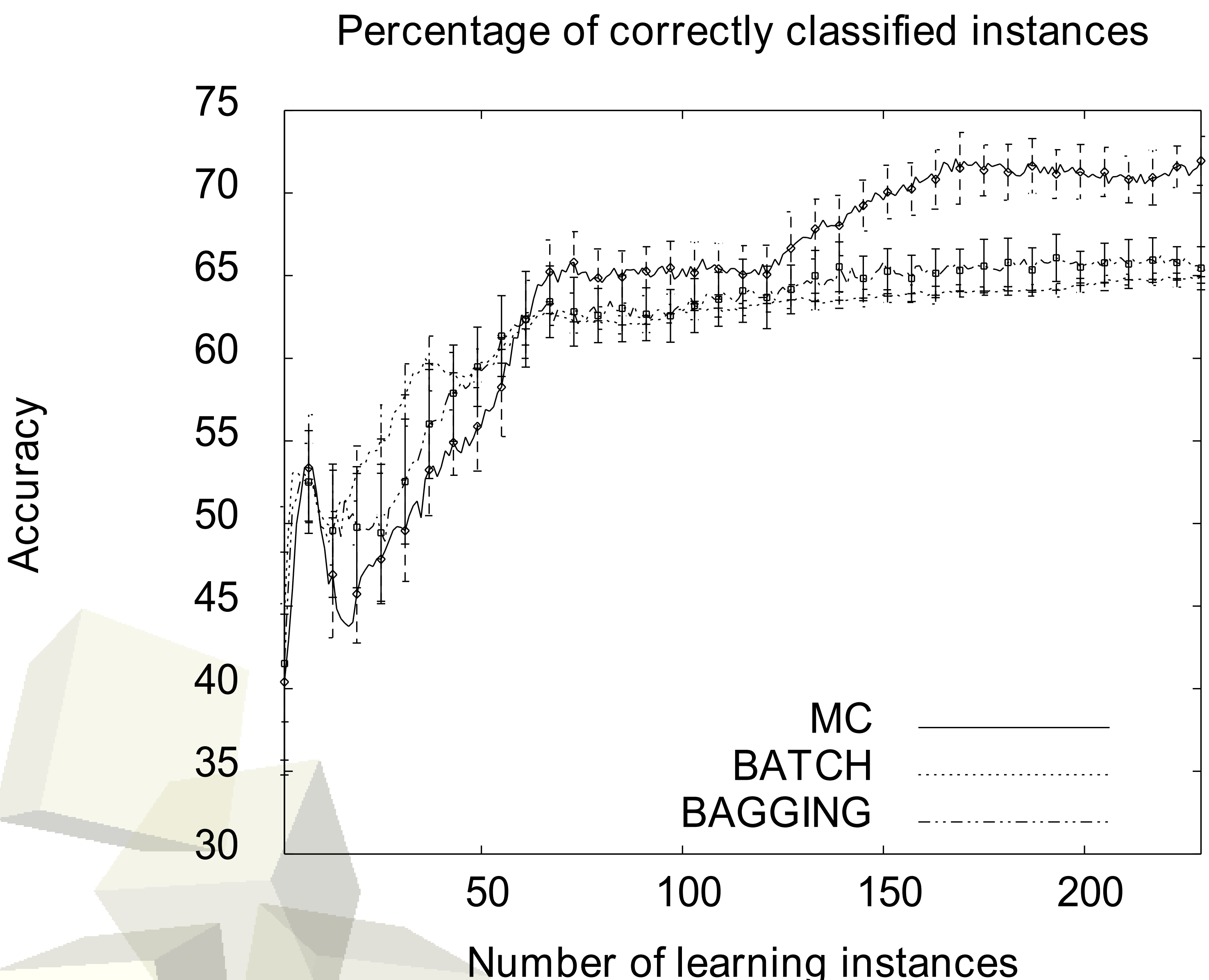
Results on benchmarks

Base	Batch	MC	Number of selected instances
Soybean	93,9	90,2	93 / 596
Vote	90,3	95,2	24 / 390
Spambase	84,9	82,0	852 / 4139
Iris disc	96,0	93,3	17 / 132

UCI repository of machine learning databases
naïve bayes classifiers (estimate conditional probabilities, assuming the independence of attributes)
10-fold cross-validation

- like Windowing in a random order.

Results on severity (1/2)



Learning curves (averaged over 50 trials, n=7)

- MC (our algorithm),
- BATCH (classical batch learning),
- BAGGING (vote of hypotheses learnt on random subsets; here n hypotheses and subsets of the same size as the learning set chosen by MC).

3 classes, naive bayes classifiers
Initialization with 3 instances randomly drawn from a separate set.
52 minutes of stream: 828 instances in the data stream.
4 x 10 minutes (2 traffic conditions): 371 exemples for test.

Results on severity (2/2)

- Final performance:

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

$$\textit{Precision for class } A = \frac{\textit{Number of instances correctly classified in class } A}{\textit{Number of instances classified in class } A} = \frac{1}{1+2}$$

predicted \ true	A	B
A	1	2
B	3	4

Conclusion

- Promising incremental algorithm.
- Future work:
  - intelligent combination of hypotheses: better than Vote ?
  - extension to longer periods to process the database: detection of concept drift, performance monitoring.