New ensemble methods for evolving data streams

A. Bifet, G. Holmes, B. Pfahringer, R. Kirkby, and R. Gavaldà

Laboratory for Relational Algorithmics, Complexity and Learning **LARCA**
UPC-Barcelona Tech, Catalonia

University of Waikato
Hamilton, New Zealand

Paris, 29 June 2009
15th ACM SIGKDD International Conference on
Knowledge Discovery and Data Mining 2009

New Ensemble Methods For Evolving Data Streams

Outline

- a new experimental data stream framework for studying concept drift
- two new variants of Bagging:
  - `ADWIN` Bagging
  - Adaptive-Size Hoeffding Tree (ASHT) Bagging.
- an evaluation study on synthetic and real-world datasets

Outline

1. MOA: Massive Online Analysis
2. Concept Drift Framework
3. New Ensemble Methods
4. Empirical evaluation

What is MOA?

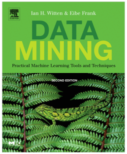
{M}assive {O}nline {A}nalysis is a framework for online learning from data streams.

- It is closely related to WEKA
- It includes a collection of offline and online as well as tools for evaluation:
  - boosting and bagging
  - Hoeffding Trees

  with and without Naïve Bayes classifiers at the leaves.

WEKA

- Waikato Environment for Knowledge Analysis
- Collection of state-of-the-art machine learning algorithms and data processing tools implemented in Java
  - Released under the GPL
- Support for the whole process of experimental data mining
  - Preparation of input data
  - Statistical evaluation of learning schemes
  - Visualization of input data and the result of learning

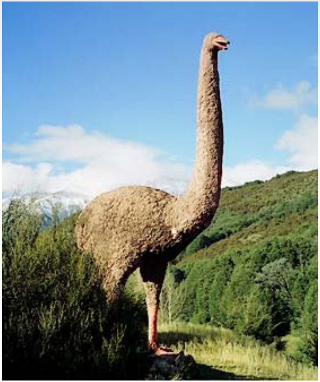


- Used for education, research and applications
- Complements “Data Mining” by Witten & Frank

WEKA: the bird

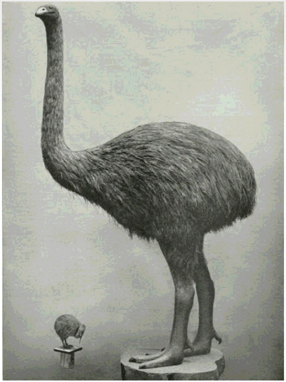
MOA: the bird

The Moa (another native NZ bird) is not only flightless, like the Weka, but also extinct.



MOA: the bird

The Moa (another native NZ bird) is not only flightless, like the Weka, but also extinct.

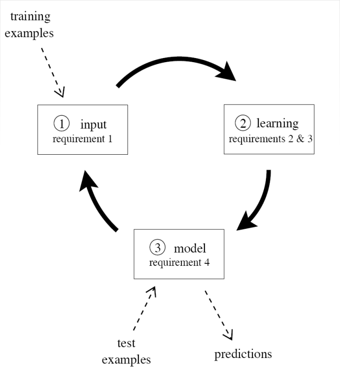


MOA: the bird

The Moa (another native NZ bird) is not only flightless, like the Weka, but also extinct.

Data stream classification cycle

1. Process an example at a time, and inspect it only once (at most)
2. Use a limited amount of memory
3. Work in a limited amount of time
4. Be ready to predict at any point



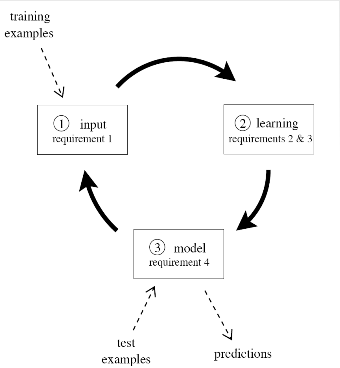
Experimental setting

Evaluation procedures for Data Streams

- Holdout
- Interleaved Test-Then-Train or Prequential

Environments

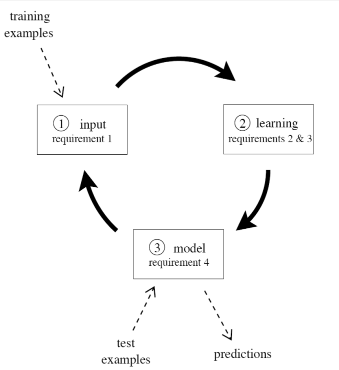
- Sensor Network: 100Kb
- Handheld Computer: 32 Mb
- Server: 400 Mb



Experimental setting

Data Sources

- Random Tree Generator
- Random RBF Generator
- LED Generator
- Waveform Generator
- Function Generator



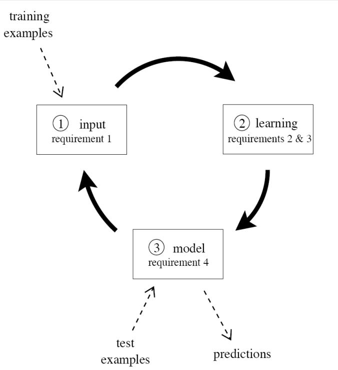
Experimental setting

Classifiers

- Naive Bayes
- Decision stumps
- Hoeffding Tree
- Hoeffding Option Tree
- Bagging and Boosting

Prediction strategies

- Majority class
- Naive Bayes Leaves
- Adaptive Hybrid



Easy Design of a MOA classifier

- `void resetLearningImpl ()`
- `void trainOnInstanceImpl (Instance inst)`
- `double[] getVotesForInstance (Instance i)`
- `void getModelDescription (StringBuilder out, int indent)`

Outline

Extension to Evolving Data Streams

New Evolving Data Stream Extensions

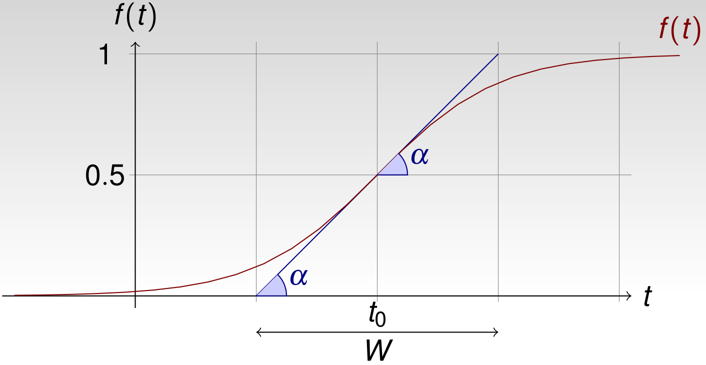
- New Stream Generators
- New UNION of Streams
- New Classifiers

Extension to Evolving Data Streams

New Evolving Data Stream Generators

- Random RBF with Drift
- LED with Drift
- Waveform with Drift
- Hyperplane
- SEA Generator
- STAGGER Generator

Concept Drift Framework

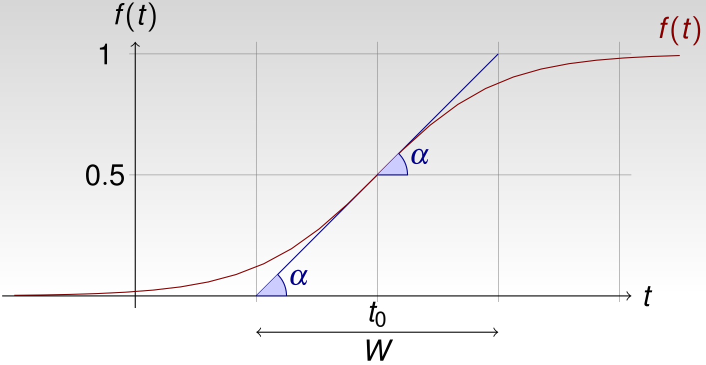


Definition

Given two data streams a, b, we define $c = a \oplus_{t_0}^{W} b$ as the data stream built joining the two data streams a and b

- $\Pr[c(t) = b(t)] = 1/(1 + e^{-4(t-t_0)/W})$.
- $\Pr[c(t) = a(t)] = 1 - \Pr[c(t) = b(t)]$

Concept Drift Framework



Example

- $(((a \oplus_{t_0}^{W_0} b) \oplus_{t_1}^{W_1} c) \oplus_{t_2}^{W_2} d) \ldots$
- $(((SEA_9 \oplus_{t_0}^{W} SEA_8) \oplus_{2t_0}^{W} SEA_7) \oplus_{3t_0}^{W} SEA_{9.5})$
- $\text{CovPokElec} = (\text{CoverType} \oplus_{581{,}012}^{5{,}000} \text{Poker}) \oplus_{1{,}000{,}000}^{5{,}000} \text{ELEC2}$

Extension to Evolving Data Streams

New Evolving Data Stream Classifiers

- Adaptive Hoeffding Option Tree
- DDM Hoeffding Tree
- EDDM Hoeffding Tree
- OCBoost
- FLBoost

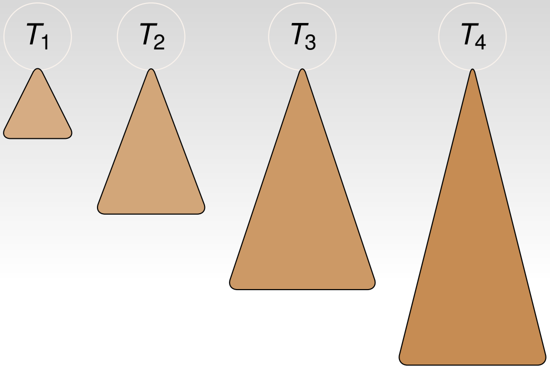
Outline

Ensemble Methods

New ensemble methods:

- Adaptive-Size Hoeffding Tree bagging:
  - each tree has a maximum size
  - after one node splits, it deletes some nodes to reduce its size if the size of the tree is higher than the maximum value
- `ADWIN` bagging:
  - When a change is detected, the worst classifier is removed and a new classifier is added.

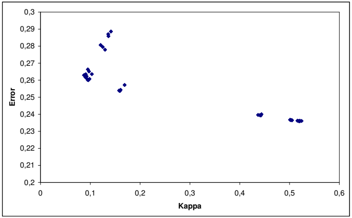
Adaptive-Size Hoeffding Tree



Ensemble of trees of different size

- smaller trees adapt more quickly to changes,
- larger trees do better during periods with little change
- diversity

Adaptive-Size Hoeffding Tree



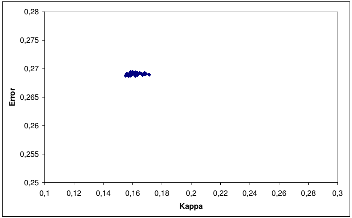


Figure: Kappa-Error diagrams for ASHT bagging (left) and bagging (right) on dataset RandomRBF with drift, plotting 90 pairs of classifiers.

ADWIN Bagging

ADWIN

An adaptive sliding window whose size is recomputed online according to the rate of change observed.

ADWIN has rigorous guarantees (theorems)

- On ratio of false positives and negatives
- On the relation of the size of the current window and change rates

ADWIN Bagging

When a change is detected, the worst classifier is removed and a new classifier is added.

Outline

Empirical evaluation

Dataset	**Most Accurate Method**
Hyperplane Drift 0.0001	Bag10 ASHT W+R
Hyperplane Drift 0.001	Bag10 ASHT W+R
SEA W = 50	Bag10 ASHT W+R
SEA W = 50000	BagADWIN 10 HT
RandomRBF No Drift 50 centers	Bag 10 HT
RandomRBF Drift .0001 50 centers	BagADWIN 10 HT
RandomRBF Drift .001 50 centers	Bag10 ASHT W+R
RandomRBF Drift .001 10 centers	BagADWIN 10 HT
Cover Type	Bag10 ASHT W+R
Poker	OzaBoost
Electricity	OCBoost
CovPokElec	BagADWIN 10 HT

Empirical evaluation

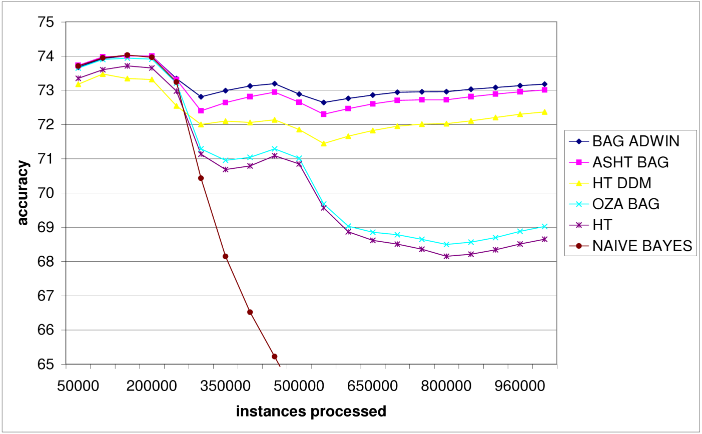


Figure: Accuracy on dataset LED with three concept drifts.

Empirical evaluation

	SEA W = 50		
	Time	Acc.	Mem.
Bag10 ASHT W+R	33.20	**88.89**	0.84
BagADWIN 10 HT	54.51	88.58	1.90
Bag5 ASHT W+R	19.78	88.55	0.01
HT DDM	8.30	88.27	0.17
HT EDDM	8.56	87.97	0.18
OCBoost	59.12	87.21	2.41
OzaBoost	39.40	86.28	4.03
Bag10 HT	31.06	85.45	3.38
AdaHOT50	22.70	85.35	0.86
HOT50	22.54	85.20	0.84
AdaHOT5	11.46	84.94	0.38
HOT5	11.46	84.92	0.38
HT	6.96	84.89	0.34
NaiveBayes	5.32	83.87	0.00

Empirical evaluation

	SEA W = 50000		
	Time	Acc.	Mem.
BagADWIN 10 HT	53.15	**88.53**	0.88
Bag10 ASHT W+R	33.56	88.30	0.84
HT DDM	7.88	88.07	0.16
Bag5 ASHT W+R	20.00	87.99	0.05
HT EDDM	8.52	87.64	0.06
OCBoost	60.33	86.97	2.44
OzaBoost	39.97	86.17	4.00
Bag10 HT	30.88	85.34	3.36
AdaHOT50	22.80	85.30	0.84
HOT50	22.78	85.18	0.83
AdaHOT5	12.48	84.94	0.38
HOT5	12.46	84.91	0.37
HT	7.20	84.87	0.33
NaiveBayes	5.52	83.87	0.00

Summary

`http://www.cs.waikato.ac.nz/∼abifet/MOA/`

Conclusions

- Extension of MOA to evolving data streams
- MOA is easy to use and extend
- New ensemble bagging methods:
  - Adaptive-Size Hoeffding Tree bagging
  - `ADWIN` bagging

Future Work

- Extend MOA to more data mining and learning methods.