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}nalysys 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
  - \text{Hoeffding Trees}
- with and without \text{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
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

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

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

Example

- \(((a \oplus_{t_0} W_0 b) \oplus_{t_1} c) \oplus_{t_2} d)\) …
- \(((SEA_9 \oplus_{t_0} W \text{SEA}_8) \oplus_{2t_0} W \text{SEA}_7) \oplus_{3t_0} W \text{SEA}_{9.5})
- CovPokElec = (CoverType \oplus_{5,000}^{5,000} 581,012 \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

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

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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Most Accurate Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperplane Drift 0.0001</td>
<td>Bag10 ASHT W+R</td>
</tr>
<tr>
<td>Hyperplane Drift 0.001</td>
<td>Bag10 ASHT W+R</td>
</tr>
<tr>
<td>SEA W = 50</td>
<td>Bag10 ASHT W+R</td>
</tr>
<tr>
<td>SEA W = 50000</td>
<td>Bag_{ADWIN} 10 HT</td>
</tr>
<tr>
<td>RandomRBF No Drift 50 centers</td>
<td>Bag 10 HT</td>
</tr>
<tr>
<td>RandomRBF Drift .0001 50 centers</td>
<td>Bag_{ADWIN} 10 HT</td>
</tr>
<tr>
<td>RandomRBF Drift .001 50 centers</td>
<td>Bag10 ASHT W+R</td>
</tr>
<tr>
<td>RandomRBF Drift .001 10 centers</td>
<td>Bag_{ADWIN} 10 HT</td>
</tr>
<tr>
<td>Cover Type</td>
<td>Bag10 ASHT W+R</td>
</tr>
<tr>
<td>Poker</td>
<td>OzaBoost</td>
</tr>
<tr>
<td>Electricity</td>
<td>OCBoost</td>
</tr>
<tr>
<td>CovPokElec</td>
<td>Bag_{ADWIN} 10 HT</td>
</tr>
</tbody>
</table>
Empirical evaluation

**Figure:** Accuracy on dataset LED with three concept drifts.
Empirical evaluation

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Time</th>
<th>Acc.</th>
<th>Mem.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bag10 ASHT W+R</td>
<td>33.20</td>
<td>88.89</td>
<td>0.84</td>
</tr>
<tr>
<td>BagADWIN 10 HT</td>
<td>54.51</td>
<td>88.58</td>
<td>1.90</td>
</tr>
<tr>
<td>Bag5 ASHT W+R</td>
<td>19.78</td>
<td>88.55</td>
<td>0.01</td>
</tr>
<tr>
<td>HT DDM</td>
<td>8.30</td>
<td>88.27</td>
<td>0.17</td>
</tr>
<tr>
<td>HT EDDM</td>
<td>8.56</td>
<td>87.97</td>
<td>0.18</td>
</tr>
<tr>
<td>OCSBoost</td>
<td>59.12</td>
<td>87.21</td>
<td>2.41</td>
</tr>
<tr>
<td>OzaBoost</td>
<td>39.40</td>
<td>86.28</td>
<td>4.03</td>
</tr>
<tr>
<td>Bag10 HT</td>
<td>31.06</td>
<td>85.45</td>
<td>3.38</td>
</tr>
<tr>
<td>AdaHOT50</td>
<td>22.70</td>
<td>85.35</td>
<td>0.86</td>
</tr>
<tr>
<td>HOT50</td>
<td>22.54</td>
<td>85.20</td>
<td>0.84</td>
</tr>
<tr>
<td>AdaHOT5</td>
<td>11.46</td>
<td>84.94</td>
<td>0.38</td>
</tr>
<tr>
<td>HOT5</td>
<td>11.46</td>
<td>84.92</td>
<td>0.38</td>
</tr>
<tr>
<td>HT</td>
<td>6.96</td>
<td>84.89</td>
<td>0.34</td>
</tr>
<tr>
<td>NaiveBayes</td>
<td>5.32</td>
<td>83.87</td>
<td>0.00</td>
</tr>
</tbody>
</table>
### Empirical evaluation

![Table of results]

**SEA**

\[ W = 50000 \]

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