Pruning AdaBoost for Continuous Sensors Mining Applications

Mojdeh Rastgoo, Guillaume Lemaitre, Xavier Rafael, Felip Miralles and Pier Luigi Casale

eHealth R&D, BDigital Technology Centre, Barcelona, Spain

Introduction

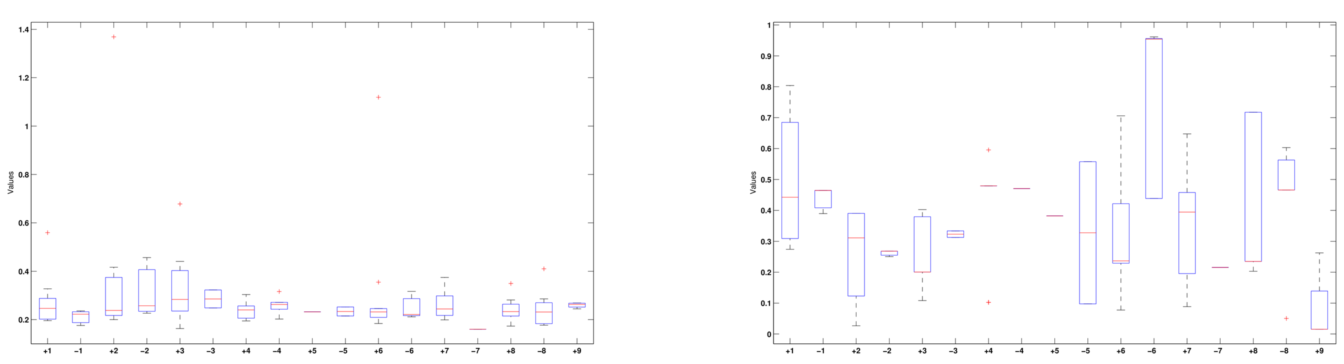
- Crucial task of mining sensors data in continuous learning environment
- AdaBoost is a well known framework for mining continuous streams, however when many subsequent batches of data are provided, AdaBoost tends to create large ensembles that suffer of two main drawbacks:
  - Increasing memory
  - Over-fitting
- Pruning techniques for AdaBoost classifier considering continuous learning environment could provide:
  - Least memory consuming ensembles for each stage of learning
  - A first attempt to retain only the significant information acquired from previous knowledge

Reduce Error Algorithm (RE)

- New version of reduced error method introduced first by *Margineantu & Dietterich et el*
- Our implementation is different from the initial version since:
  - No back-fitting is performed after each selection
  - The selected classifiers are re-weighted based on the pruning distribution
- RE algorithm:
  - Training AdaBoost on training set with 100 base classifiers $\mathbf{H}$
  - The pruning distribution $\mathbf{W_t}$ is initialized
  - The weak learner $\mathbf{h_t}$ with minimum error on $\mathbf{W_t}$ is added as first base classifier to $\mathbf{P}$ while its weight $\alpha_\mathbf{t}$ is updated. Pruning distribution is updated as well $\mathbf{W_{t+1}}$
  - For the desired number of base classifier,
    - The error of combination of each remaining classifier $\mathbf{h_t}$ and $\mathbf{P}$ is evaluated on $\mathbf{W_{t+1}}$. The base classifier which leads to minimum error is added to $\mathbf{P}$ and its relative weight $\alpha_\mathbf{t}$ is updated

Learner Weights Analysis (LWA)

- Distribution of weights (right) and learners (left) based on different dimension provided by AdaBoost for breast datasets



- Selection of weak learners based on the weight distribution considering:
  - Learners with higher weights have more impact
  - An ensemble is better when more diversified the classifiers forming it are
- LWA Algorithm:
  - The AdaBoost is trained on training set, providing $\mathbf{T}$ decision stump classifier
  - The weights of decision stumps $\alpha_\mathbf{t}$ are grouped using their dimension parameters (Matrix $\mathbf{M}$ of size $\mathbf{n} \times \mathbf{D}$, when $\mathbf{n}$ is the number of elements for each of $\mathbf{D}$ dimensions)
  - $\mathbf{M}$ is sorted by row in descendant order, then sorted by column in a descendent order
  - Sorted $\mathbf{M}$ is transformed to vector $\mathbf{V}$ by concatenating the column of $\mathbf{M}$
  - $\mathbf{t}$ first classifier corresponding to the first $\mathbf{t}$ element of $\mathbf{V}$ are selected for final ensemble

Pareto Analysis (PA)

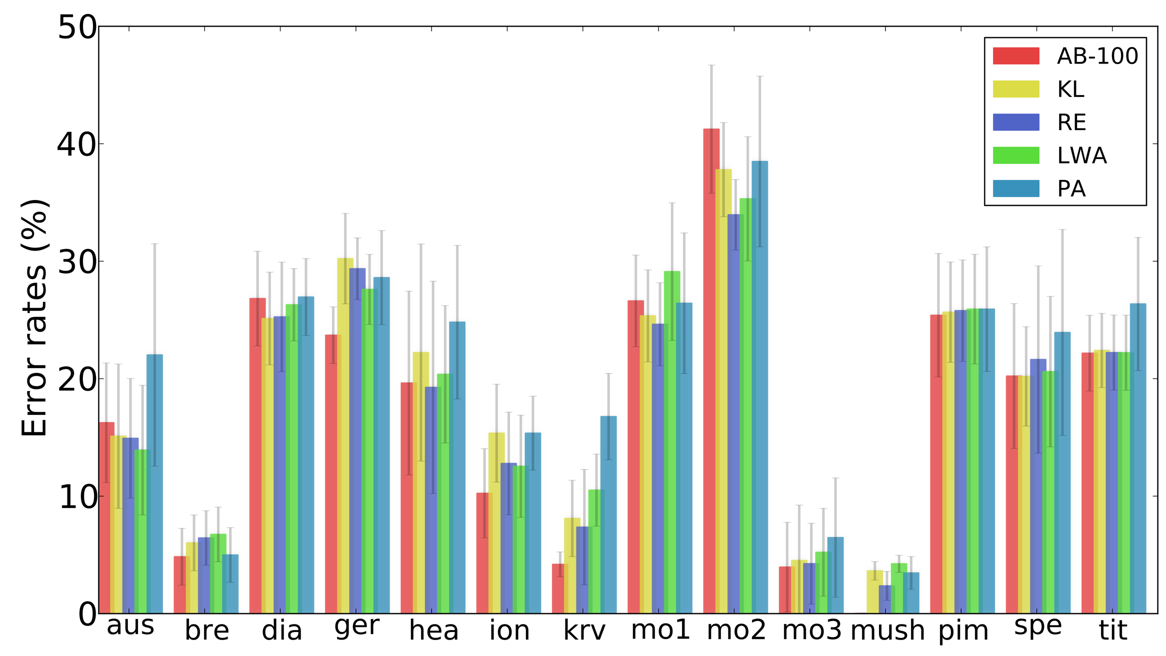
- PA is based on the assumption that few key actions will produce significant overall effects
- This technique estimates the effectiveness of each feature dimension and accordingly selects the classifier from feature dimensions with higher impact
- The effectiveness could be adjusted based on threshold
- This method could work in an automatic way, where only few classifier will be selected based on their effectiveness, or in an non-automatic way the number of desired classifiers are fixed

Pareto Analysis (PA)

- PA Algorithm:
  - The features are grouped based on total number of weight outliers in each of them
  - The frequency distribution of outliers are sorted in descendant order
  - The cumulative percentage of the sorted group is computed
  - All the dimensions with lower cumulative error than the specified threshold are considered

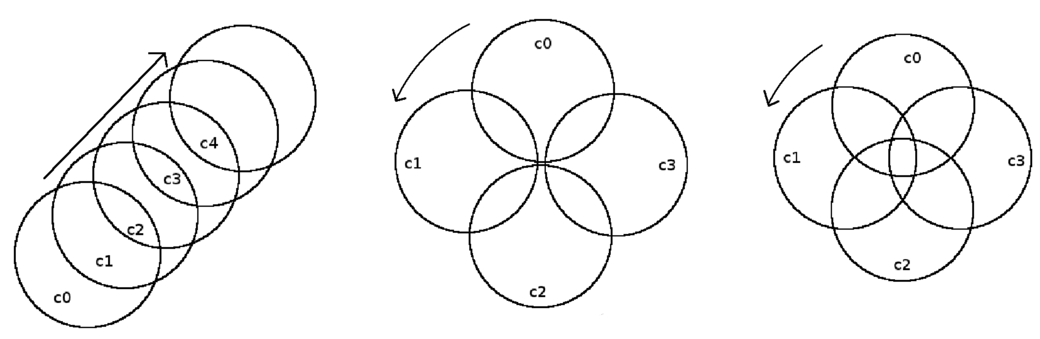
Results: UCI Datasets Repository

- 10-fold cross-validation on 14 UCI datasets - 90% pruning

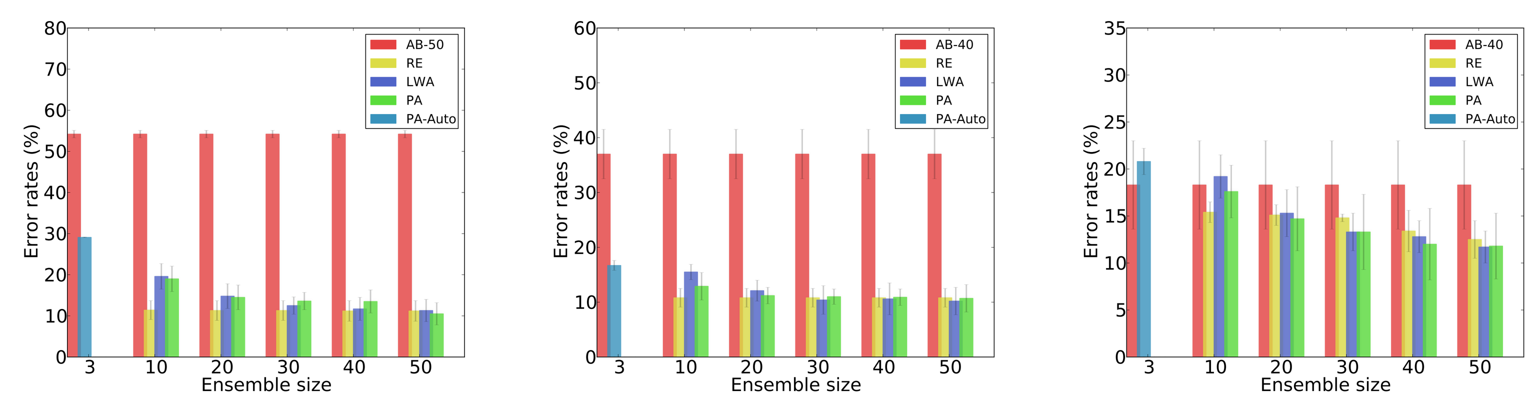


Results: Simulated Drifting Datasets

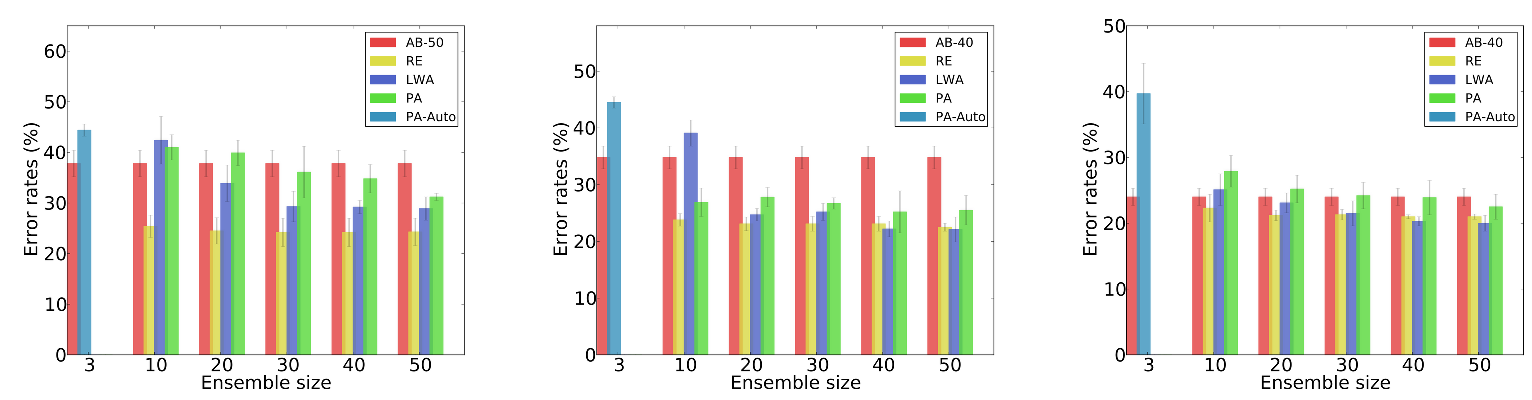
- Simulated drifts datasets, Dataset 1, linear drift (right), Dataset 2, circular drift (middle), Dataset 3, circular drift (left)



- Exp. 1, Goal: Classifying final stage of data



- Exp. 2, Goal: Classifying whole datasets (keeping memory of previous stages)



Results: Real World Datasets

- Classifying real datasets with concept drifts, while the memory of previous stages are considered

	# Ensemble Size			
	10			50
Real Data	*LWA*	*PA*	*RE*	*AdB*
Sensor Streams	11.1 ± 5.0	19.2 ± 9.7	**6.4 ± 3.6**	17.3 ± 9.7
Power Supply	28.3±0.2	28.3±0.2	**28.1±0.0**	28.2±0.0
Elec2	28.4±0.8	28.9±1.7	**27.3±0.2**	30.5±3.9

Conclusions

- The selection of weak classifiers from different pools of sub-sampled data may improve the final ensemble in terms of accuracy, diversity and adaptation ability to drift.
- The pruning method introduced in this research specially RE and LWA could improve the performance of AdaBoost in continuous learning and easily could be adapted for large datasets with high quantity of information.

http://bdigital.org ; https://www.udg.edu/ mojdeh.rastgoo@gmail.com