POOL AND ACCURACY BASED STREAM CLASSIFICATION:
A new ensemble algorithm on data stream classification using recurring concepts detection

Mohammad Javad Hosseini, Zahra Ahmadi, Hamid Beigy
{mjhosseini,z_ahmadi}@ce.sharif.edu
beigy@sharif.edu
Outline

- Proposed learning algorithm
- Evaluation and comparison of the method
- Conclusion and future works
Introduction

- We want to classify data having these properties
  - Streaming data
  - Concept drift
  - Recurring Concepts

- Previous approaches
  - The most similar method to ours is the CCP Framework [1]
Proposed Method (PASC)

- Pool of classifiers, one for each concept
- Batches of k instances
  \[ B_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,k}) \]
  - Label prediction
  - Revealing true labels
    \[ L_t = (l_{t,1}, l_{t,2}, \ldots, l_{t,k}) \]
  - Pool update
The main framework of PASC

- PASC has three phases shown in the main loop

```
Input: an infinite stream of batches of instances B_i.
After classification of each instance B_i,j,
its label is revealed to the algorithm.
Output: Predicted labels of instances B_i,j.

1. Pool = Ø; // the pool of classifiers
2. C = make_classifier(B_1,L_1);
3. RDC = new classifier(); //only used in Bayesian
4. //method
5. ac = 1; // active classifier
6. W_1=1;
7. Pool = Pool U {C};
8. X_1=sum_data(B_1);
9. RDC.update(X_1,1); //1 is the label of X_1
10. for i=2 to infinity do
    11. Classify B_i.
    12. Update Pool with B_i and L_i.
    13. determine active classifier (classifier weights);
14. end for
```

The main framework of PASC.
Three phases of PASC

- Phase 1: Classifying the batch
  - Active classifier method
  - Weighted classifiers method
- Phase 2: Updating the classifiers’ pool
  - Bayesian method
  - Heuristic method
- Phase 3: Determining the active classifier (or classifiers’ weights)
  - Needed for the next iteration
Phase 1: Classifying the Batch
Active classifier method

- **Active classifier:**
  - Classifier updated using the previous batch of data
    - A new classifier
    - The most relevant classifier

- If there is no concept drift
  - Active classifier would be the best choice

- When a sudden concept drift occurs
  - Performance decreases significantly
Weighted classifiers method

- A positive weight is assigned to each classifier in the beginning of processing the batch
  - According to the performance of the classifier on the previous batch

- After receiving true label of the classified instance, classifiers’ weights can be updated:
  \[ w'(j) = w(j) \times \beta^{M(j,i)} \]
  - \( w(j) \) and \( w'(j) \) are the current and new weights of the \( j^{th} \) classifier
  - \( \beta \) is a parameter in [0,1)
  - \( M(j,i) = 0 \) showing that \( j^{th} \) classifier’s prediction for \( i^{th} \) instance is true and 1, otherwise
Using weighted majority method with these weights
- Guaranties performance near the best classifier [2]

But we slightly modify it to improve the efficiency
- Uses only the classifier with the highest weight
- Uses only a subsample of instances for updating
Phase 2: Updating the classifiers’ pool
Phase 2: Updating the classifiers’ pool

- After receiving $L_t$, the true labels of the batch $B_t$:
  - Update a classifier in the batch
    - If any of the classifiers describe the batch well enough
  - Create a new classifier using this batch
    - Otherwise

- So we need to find:
  - The concept (classifier) which best describes $B_t$ and $L_t$
  - A measure of the best classifier’s correspondence to this batch

- This phase can be accomplished in two ways:
  - Bayesian method
  - Heuristic method
Bayesian method for Updating the classifiers’ pool

- The relevance probability of the \(i^{th}\) hypothesis (\(h_i\)) to \(B_t\) and \(L_t\) is estimated as:

\[
P(h_i|B_t, L_t) = \frac{P(B_t, L_t|h_i) * P(h_i)}{P(B_t, L_t)}
\]

- \(P(B_t,L_t)\) is equal for all hypotheses and the prior probability of all hypotheses are the same, so the best classifier is:

\[
\arg\max_i P(h_i|B_t, L_t) = \arg\max_i P(B_t, L_t|h_i) = \arg\max_i P(L_t|B_t, h_i) * P(B_t|h_i)
\]

- Conditional probability that the labels of the instances \((x_{t,1}, x_{t,2}, \ldots, x_{t,k})\) be \((l_{t,1}, l_{t,2}, \ldots, l_{t,k})\) using \(i^{th}\) hypothesis

- Probability that the batch is produced in an environment described by the \(i^{th}\) concept
Bayesian method for Updating the classifiers’ pool (Cont.)

- I.I.D condition holds for the instances in a batch
  - Size of the batch is small enough
- First term:
  \[
P(L_t \mid B_t, h_i) = \prod_{j=1}^{j=k} P(l_{t,j} \mid x_{t,j}, h_i)
\]
  - Can be estimated using posterior probabilities computed by classifiers
- Second term:
  \[
P(B_t \mid h_i) = \prod_{j=1}^{j=k} P(x_{t,j} \mid h_i)
\]
Bayesian method for Updating the classifiers’ pool (Cont.)

- For all instances $j$, assume that:
  
  $$P(x_{t,j} | h_i) = p_i$$

- So
  
  $$P(B_t | h_i) = \prod_{j=1}^{j=k} P(x_{t,j} | h_i) = p_i^k$$

- To estimate $p_i$:
  
  - Construct a classifier which has a summarization of $B_{k<t}$, i.e. $X_{k<t}$ and the concept numbers of each of the batches seen so far
    - The classifier is trained incrementally to estimate the probability $P(B_t | h_i)$ by $P(X_t | h_i)$
To prevent underflow of the products, logarithm of the probabilities is used

\[
\begin{align*}
\arg\max_{i,j} P(h_i|B_t, L_t) \\
= \arg\max_i k \times \log p_i + \sum_{j=1} \log P(l_{t,j}|x_{t,j}, h_i)
\end{align*}
\]

If the pool is not full, the result of this expression is compared with a parameter \(\theta_1\)

- To decide whether a new classifier should be added.
- To improve the efficiency, only a subsample of the instances are used to estimate the best concept.
Heuristic method for Updating the classifiers’ pool

- Accuracy of classifiers on the new batch is used to determine the best concept.

- If the pool is not full, the best accuracy is compared with a parameter $\theta_2$.
  - To decide whether a new classifier should be added.

- To improve the efficiency, only a subsample of the instances are used to estimate the best concept.
Phase 3: Determining the active classifier (or classifier weights)
Phase 3: Determining the active classifier (or classifier weights)

- Using active classifier method in phase 1
  - The active classifier should be set

- Using weighted classifiers method in phase 1
  - Classifiers’ weights should be set.
  - Each classifier is tested on a subsample of the batch and weights are set by:
    \[ w_0(i) = \beta(2^{A(i)}) \]
  - \( A(i) \) is the accuracy of the \( i^{th} \) classifier
Experiments

- Datasets:
  - Emailing list Dataset [1]
  - Spam Filtering Dataset [3]
  - Hyperplane Dataset

- Updatable naïve Bayes classifier was used

- Parameter setting:
  - $\beta = 0.1$, used in weighted classifiers method
  - Maximum number of classifiers in the pool = 10
  - Heuristic and Bayesian methods’ thresholds are 0.95 and $2\times \log(0.75) \times m$, respectively
    - $m$ is the subsample size used in the Bayesian method
  - Batch size = 50 for emailing list and spam filtering data sets and 500 for the hyperplane data set

- Comparisons are done in terms of accuracy, precision, recall and running time with the CCP Framework method [1]
Experiments (Cont.)

- Hyperplane data set

<table>
<thead>
<tr>
<th>Classification Method</th>
<th>Batch assignment Method</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>Time(ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active Classifier</strong></td>
<td>Bayesian</td>
<td>0.777</td>
<td>0.742</td>
<td>0.852</td>
<td>1154</td>
</tr>
<tr>
<td><strong>Weighted Classifiers</strong></td>
<td>Bayesian</td>
<td>0.827</td>
<td>0.781</td>
<td>0.909</td>
<td>1265</td>
</tr>
<tr>
<td><strong>Active Classifier</strong></td>
<td>Heuristic</td>
<td>0.760</td>
<td>0.726</td>
<td>0.838</td>
<td>1239</td>
</tr>
<tr>
<td><strong>Weighted Classifiers</strong></td>
<td>Heuristic</td>
<td>0.841</td>
<td>0.814</td>
<td>0.885</td>
<td>1332</td>
</tr>
<tr>
<td><strong>Active Classifier</strong></td>
<td>CCP (Leader, follower)</td>
<td>0.758</td>
<td>0.724</td>
<td>0.836</td>
<td>1141</td>
</tr>
<tr>
<td><strong>Weighted Classifiers</strong></td>
<td>CCP (Leader, follower)</td>
<td>0.831</td>
<td>0.807</td>
<td>0.871</td>
<td>1377</td>
</tr>
</tbody>
</table>
Spam filtering data set

<table>
<thead>
<tr>
<th>Classification Method</th>
<th>Batch assignment Method</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>Time(ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Classifier</td>
<td>Bayesian</td>
<td>0.908</td>
<td>0.923</td>
<td>0.956</td>
<td>7357</td>
</tr>
<tr>
<td>Weighted Classifiers</td>
<td>Bayesian</td>
<td>0.905</td>
<td>0.938</td>
<td>0.934</td>
<td>8349</td>
</tr>
<tr>
<td>Active Classifier</td>
<td>Heuristic</td>
<td>0.908</td>
<td>0.931</td>
<td>0.947</td>
<td>3676</td>
</tr>
<tr>
<td>Weighted Classifiers</td>
<td>Heuristic</td>
<td>0.909</td>
<td>0.934</td>
<td>0.944</td>
<td>3838</td>
</tr>
<tr>
<td>Active Classifier</td>
<td>CCP (Leader follower)</td>
<td>0.915</td>
<td>0.921</td>
<td>0.969</td>
<td>3772</td>
</tr>
<tr>
<td>Weighted Classifiers</td>
<td>CCP (Leader follower)</td>
<td>0.899</td>
<td>0.932</td>
<td>0.933</td>
<td>4758</td>
</tr>
</tbody>
</table>
Conclusions and future works

- The general framework used in this paper
  - Maintains a pool of classifiers
  - Updates them according to consecutive batches of data
  - Uses them to classify new batches of data

- The most similar method to ours is the CCP framework
  - Our method improves the accuracy while its parameter tuning is simpler

- Some future directions include
  - Better Management of the classifiers in the pool
  - Dynamically set the parameters according to the data sets
  - Testing more real data sets
References


Thank You
You can contact us for any questions

{mjhosseini,z_ahmadi}@ce.sharif.edu
beigy@sharif.edu
## Experiments (Cont.)

<table>
<thead>
<tr>
<th>Classification Method</th>
<th>Batch assignment Method</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>Time(ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Classifier</td>
<td>Bayesian</td>
<td>0.759</td>
<td>0.735</td>
<td>0.762</td>
<td>1862</td>
</tr>
<tr>
<td>Weighted Classifiers</td>
<td>Bayesian</td>
<td>0.832</td>
<td>0.828</td>
<td>0.812</td>
<td>2411</td>
</tr>
<tr>
<td>Active Classifier</td>
<td>Heuristic</td>
<td>0.748</td>
<td>0.718</td>
<td>0.768</td>
<td>1490</td>
</tr>
<tr>
<td>Weighted Classifiers</td>
<td>Heuristic</td>
<td><strong>0.828</strong></td>
<td><strong>0.811</strong></td>
<td><strong>0.830</strong></td>
<td><strong>1440</strong></td>
</tr>
<tr>
<td>Active Classifier</td>
<td>CCP (Leader follower)</td>
<td>0.771</td>
<td>0.748</td>
<td>0.775</td>
<td>1311</td>
</tr>
<tr>
<td>Weighted Classifiers</td>
<td>CCP (Leader follower)</td>
<td>0.816</td>
<td>0.791</td>
<td>0.828</td>
<td>1860</td>
</tr>
</tbody>
</table>
Experiments (Cont.)

- **Comparison of methods’ accuracies**
  - Using weighted classifiers method leads to much better accuracies (about 8%) in data sets with sudden concept drift
  - Our batch assignment methods (Bayesian and heuristic) have results similar to the CCP framework method while having simpler parameter tuning

- **Comparison of methods’ run times**
  - Active classifier method
    - Heuristic and CCP methods are similar
    - Bayesian takes the most time
  - Weighted classifier method
    - Heuristic takes the least and Bayesian takes the most time