A Betting Function for addressing Concept Drift with Conformal Martingales

Eliades Charalambos, Harris Papadopoulos

Outline

- Motivation
- Data Exchangeability–Inductive Conformal Martingales
- Betting Function
- Experiments and Results
- Conclusions

Motivation

- The need to detect Concept Drift(CD) with respect to a significance level, Inductive Conformal Martingale can provide valid guarantees.
- We propose a betting function that avoids the continuous reduction of the Martingale value.
- A computationally efficient betting function with only a few parameters to tune.

Concept Drift

- Given a data stream $S = \{(x_0, y_0), (x_1, y_1), \dots\}$
  - x_i is an input vector , y_i the corresponding label

- If the set S can be divided in two sets generated by different distributions:
  $S_{0,t} = \{(x_0, y_0), \dots, (x_t, y_t)\}$ and $S_{t+1,\dots} = \{(x_{t+1}, y_{t+1}), \dots\}$

Then a Concept Drift occurred at timestamp $t + 1$.

Consequently, a violation of the exchangeability occurred.

Data Exchangeability

- Exchangeability:
  - Given an infinite sequence of random variables $(Z_1, Z_2, Z_3, \dots)$ the joint distribution $P(Z_1, Z_2, Z_3, \dots)$ is exchangeable if it is invariant under any permutation of those random variables.
  - Testing if the data is exchangeable is equivalent to testing the data for being i.i.d.
- Test Exchangeability Martingale
  - Is a sequence of random variables $S_1, S_2, S_3, \dots$ greater or equal to zero.
  - They keep the conditional expectation $\mathbb{E}(S_{n+1} | S_1, S_2, S_3, \dots S_n) = S_n$.

How a Martingale works

- Consider a fair game where a gambler with infinite wealth follows a strategy that is based on the distribution of the events in the game. The gain acquired by the gambler can be described by the value of a Martingale.
- Specifically Ville's inequality (Ville, 1939) indicates that the probability to have high profit(C) would be small, $\mathbb{P}(\exists n, S_n \geq C) \leq 1/C$

Conformal Martingales

- Is an exchangeability Martingale which is calculated as a function of p-values:
- $S_n = \prod_{i=1}^{n} f_i(p_i)$, where $f_i(p_i) = f_i(p_i | \, p_1, p_2, \ldots, p_{i-1})$ is the betting function.
- $S_n = S_{n-1} f_n(p_n)$.

- The **exchangeability assumption** is rejected with a significance level equal to $\frac{1}{M}$ if the value of the S_n is equal to M (Ville's inequality (Ville, 1939))

Pvalue Calculation

- To find the **pvalue** of the example z_j we calculate the sequence

$$H_j = \{a_{k+1},, \dots, a_j\}$$

- Then $p_j = \frac{|\{a_i \in H_j | a_i > a_j\}| + U_j |\{a_i \in H_j | a_i = a_j\}|}{j-k}$

Where U_j is a random number from the uniform distribution (0,1).

Calculating Non-conformity scores

- Given a sequence of examples $\{z_1, z_2, \dots\}$

  where $z_i = (x_i, y_i)$ with x_i an input vector and y_i the corresponding label.
- The first k examples $\{z_1, z_2, \dots, z_k\}$ will be used to train the underlying algorithm.
- The examples $\{z_{k+1}, \dots, z_n\}$ arrive one by one and a numerical value is assigned to each example called nonconformity score denoted by a_j and equal to $A\{z_i, \{z_1, z_2, \dots, z_k\}\}$ with $i \in \{1, \dots, j\}$.
- The NCS is based on the underlying algorithm and when a new example arrives a new NCS is assigned to each example.

Nonconformity Measure

- **Underlying Algorithm**
  - Tree Classifier, Random Forest
- **NCM**
  - For each example z_j classifier will output the posterior probability $\widetilde{p_j}$

For each label y_j, therefore we define the NCM: $\boldsymbol{a_j = -\widetilde{p_j}}$

Existing Betting Functions

- **Histogram Estimator**
  - We take a fix number of bins k, this will partition the [0,1] into

$$B_1 = [0, \frac{1}{k}), B_2 = [\frac{1}{k}, \frac{2}{k}), \ldots, B_k = [\frac{k-1}{k}, 1)$$

  - When a pvalue $p_n \in B_j$ then the density estimator will be equal to $\widehat{f_n}(p_n) = \frac{n_j.k}{n-1}$, where n_j is the number of p-values belonging to B_j
- **Kernel estimator**
  - $\widehat{f_n}(x) = \frac{1}{nh} \sum_{i=i}^{n} K(\frac{x-x_i}{h})$

Where h is the bandwidth and K(z)=$\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

Proposed Betting Function

- **Theorem**: When the distribution of the p-values is uniform, for any betting function other than f= 1 then $S_\infty = 0$.
- Our betting function is built on top of any betting function f_n.
- Consider two players. Player one uses f_n and player two uses the cautious betting function.
- Cautious Betting Function: $h_n = \begin{cases} 1 & if \; \frac{S1_{n-1}}{minS1_{n-k}} \leq \varepsilon \\ f_n & if \; \frac{S1_{n-1}}{minS1_{n-k}} > \varepsilon \end{cases}$
- $With\; S1_{n-k} = \prod_{i=1}^{n} f_i(p_i)$, $\varepsilon > 0$ $\; k \in \{n-w, \dots, n-1\}$ and w is an integer

CD detection with ICM

Data: Training set $\{z_1, z_2, \ldots, z_k\}$, Test set$\{z_{k+1}, \ldots, z_n\}$, significance level δ

Initialize $S_1 = 1$

for i=1,...,n-k **do**

$\alpha_i = A(z_{k+i}, \{z1, \ldots, z_k\})$

$p_i = \frac{\#\{j:\alpha_j > \alpha_i\} + U_j \#\{j:\alpha_j = \alpha_i\}}{i}$

Calculate betting function $B_i = B(p_1, \ldots, p_{i-1})$

$S_i = S_{i-1} \cdot B_i(p_i)$

if $S_i > \frac{1}{\delta}$ **then**

Raise an Alarm

end

end

Algorithm 1: Detect CD using ICM

Now if the final value of the Martingale S_{n-k} exceeds 10 or 100 then we can reject the exchangeability assumption at a significance level equal to 10% and 1% respectively, thus an alarm is raised for CD detection.

Experiments and results - Datasets

Dataset	Number of Instances	Number of Variables	Number of labels	Number of concepts	Chunk size	Training set size
RECOVERY TIME DATASET	100100	1 numeric (ranging from 0 to 1)	2	2	10000 90000	100
STAGGER	1000000	3 categorical (3 values)	2	4	10000	200
SEA	1000000	3 numeric (ranging from 0 to 10)	2	4	250000	1000
ELEC	45312	8 numeric	2	unknown	unknown	300
AIRLINES	539383	7 numeric	2	unknown	unknown	200

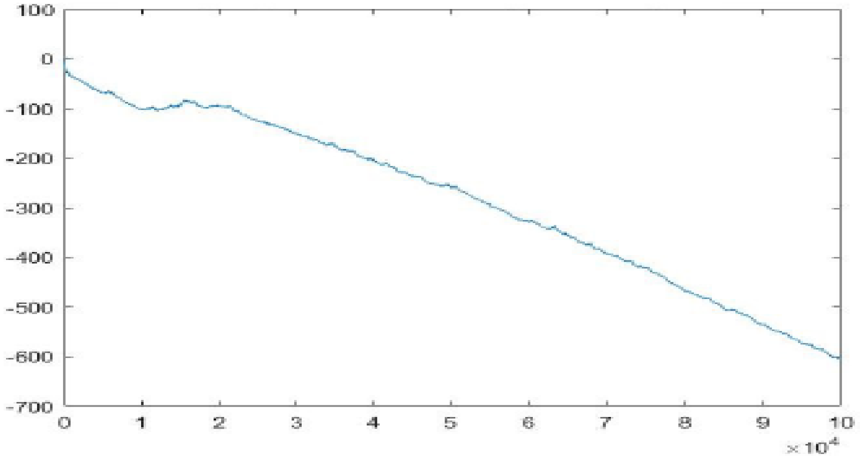
Experiments and results – Experimental Setting

- For algorithm 1 δ={0.01}
- To calculate the histogram estimator we have used the last 1000 observations.
- To calculate the kernel estimator we have used the p-values of the last 500 observations for the recovery dataset and for the rest datasets the p-values of the last 100 observations.
- For the Cautious betting function we have used a $W = 5000$ and $\varepsilon = 100$

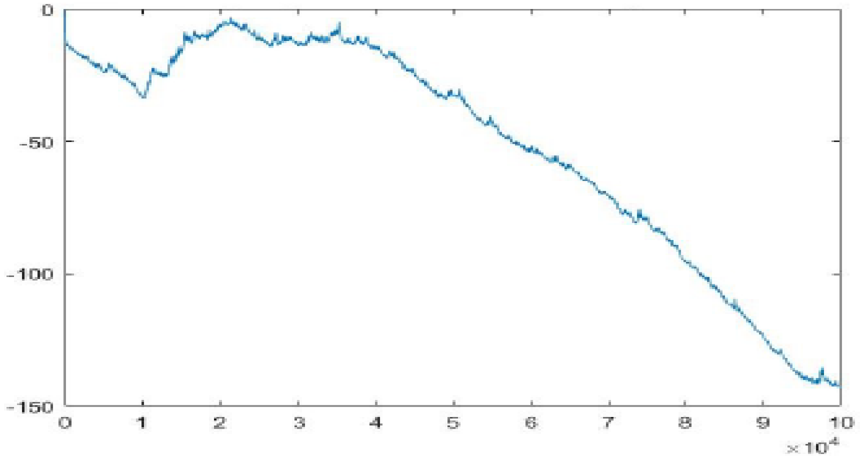
Experiments and results – Performance Measures

- ***Accuracy***: Average accuracy of the classifier (excluding the training set).
- ***Mean delay***: Average number of observations before detecting a CD after it has occurred.
- **True alarm rate(TAR)**: Average rate of CDs that have been correctly detected per chunk.
- **False alarm rate(FAR):** Average rate of CDs erroneously detected per chunk.
- **Number of CDs detected:** Total number of CDs detected in a real-world dataset.

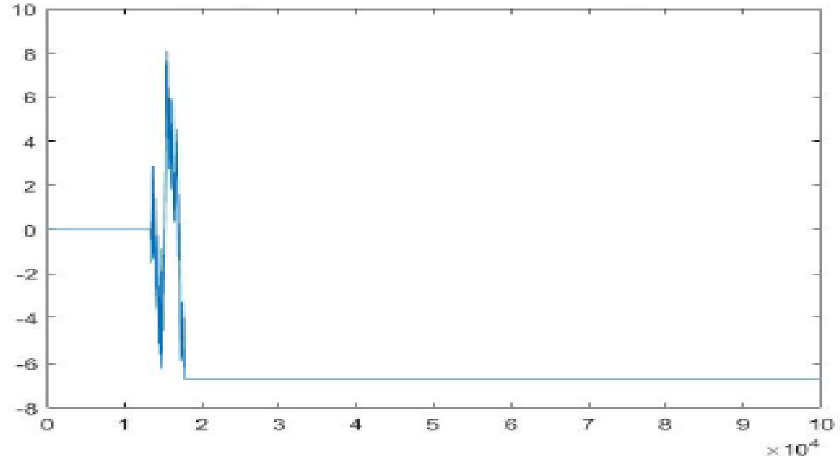
Experiments and results – Recovery Time dataset log Martingale growth



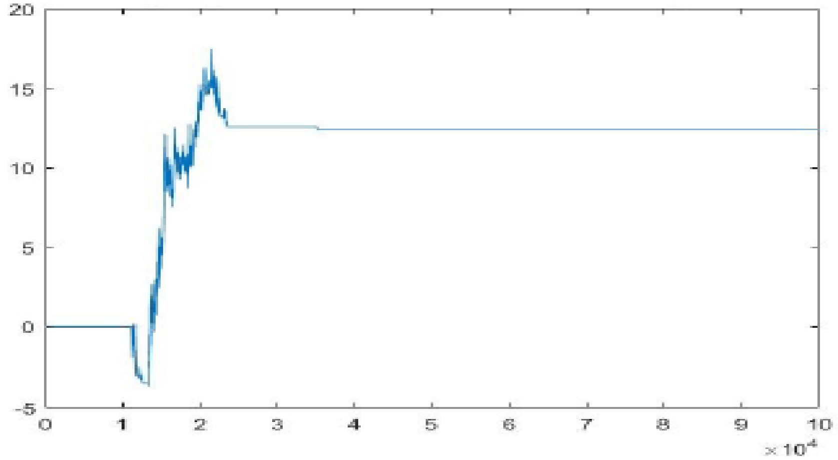
(a) *Histogram*



(b) Kernel



(c) Histogram with Cautious



(d) Kernel with Cautious

Experiments and results – Recovery Time dataset

Betting Function	FAR	TAR	Mean Delay
Histogram	0	0	—
Kernel	0	0.3	14835
Cautious with Histogram	0	0.7	12335
Cautious with Kernel	0	0.8	12339

Experiments and results - STAGGER

Betting function	No of Bins	Accuracy	Mean delay	TAR	FAR
Histogram	5	0.93690	227.6	1	0.002
Histogram	10	0.93604	236.7	1	0.004
Histogram	15	0.93535	250.8	1	0.008
Histogram with Cautious	5	0.94225	86.4	1	0.002
Histogram with Cautious	10	0.94440	77.2	1	0
Histogram with Cautious	15	0.94438	70.2	1	0
Kernel	—	0.94290	113.8	1	0.002
Kernel with Cautious	—	0.94550	48.4	1	0

Experiments and results - SEA

Betting Function	No of Bins	Accuracy	Mean delay	TAR	FAR
Histogram	5	0.8610	-	0	0
Histogram	10	0.8809	6023.6	0.998	0
Histogram	15	0.8606	-	0	0
Histogram with Cautious	5	0.8606	-	0	0
Histogram with Cautious	10	0.9149	1367.2	1	0
Histogram with Cautious	15	0.9140	441.1	1	0
Kernel	—	0.8607	-	0	0
Kernel with Cautious	—	0.9151	518.3	1	0

Experiments and results - Elec

Betting function	No of Bins	Accuracy	Number of CD detected
Histogram	5	0.73470	116.2
Histogram	10	0.73711	103.8
Histogram	15	0.73472	83.6
Histogram with Cautious	5	0.75429	102.2
Histogram with Cautious	10	0.75101	107.4
Histogram with Cautious	15	0.74836	108.8
Kernel	—	0.75178	128.8
Kernel with Cautious	—	0.75929	146.2

Experiments and results - AIRLINES

Betting Function	Number of bins	Accuracy	Number of cd detected
Histogram	5	0.57177	21.2
Histogram	10	0.56229	5.8
Histogram	15	0.55354	3
Histogram with Cautious	5	0.57377	39.6
Histogram with Cautious	10	0.59633	59.6
Histogram with Cautious	15	0.59088	51
Kernel	—	0.56597	11.8
Kernel with Cautious	—	0.60184	71.4

Experiments and results – Comparison with two state of the art algorithms

Dataset	CAUTIOUS	AWE	DWE-NB
STAGGER	0.946	0.948	0.901
SEA	0.915	0.879	0.876
ELEC	0.759	0.756	0.800
AIRLINES	0.602	0.618	0.640

Conclusions

- We propose a new BF called Cautious.
- It addresses the problem that Martingale get values close to zero.
- It improves existing betting functions especially when the change occurs after a big-time interval.
- Experiments show that it can detect cases which the other two betting functions failed.
- The proposed approach has similar accuracy to the two state of the art algorithms

Future Directions

- Combine the Cautious betting function with more than one uniformity test.
- Employ strategies for selecting representative training set

Thank you!!!