Unsupervised Concept Drift Detection based on Parallel Activations of Neural Network

Joanna Komorniczak, Paweł Ksieniewicz

Department of Computer Systems and Networks
Wroclaw University of Science and Technology

Discovering Drift Phenomena in Evolving Landscape
ACM KDD 2024 | Barcelona, Spain

Motivation

data streams and concept drift

Data streams

Characteristics of data stream processing:

- Potentially infinite inflow of data samples:
  - in the form of individual samples (online processing),
  - **in the form of data chunks** (batch processing).
- Necessity of incremental learning and continuous adaptation.
- Maintaining low computational complexity of methods.
- Resistance to **concept drifts**.
- Possibility of **delayed labeling**.

Concept drift

Three main taxonomy axes:

- → In terms of **impact**: *virtual* and *real* drift.
- → In terms of **dynamics**: *sudden*, *gradual*, *incremental*.
- → In terms of **recurrence**: concept may re-appear after some time period.

Why should we care about concept changes?

Real changes usually result in degeneration of recognition quality.

What to do about them?

Detect and **act** upon a change.

Contribution

Proposition of *Parallel Activations Drift Detection* (PADD):

- → fully **unsupervised** and **implicit** drift detection method,
- → utilizes a deterministic, untrained Neural Network and statistical analysis of its activations to recognize changes in posterior data distribution,
- → experimental evaluation of the proposed approach:
  - analysis of hyperparameters determining the sensitivity of the method,
  - comparison with six state-of-the-art drift detection methods from supervised and unsupervised families.

Parallel Activations Drift Detector

the proposed approach to drift detection

Parallel Activations Drift Detector

For each data chunk:

- get activations at the NN output for current data
- for each NN output and for each replications:

  randomly sample s values from past activations and current activations
- calculate p-value of Student's T-Test for sampled data
- check if p-value *subceeds* alpha

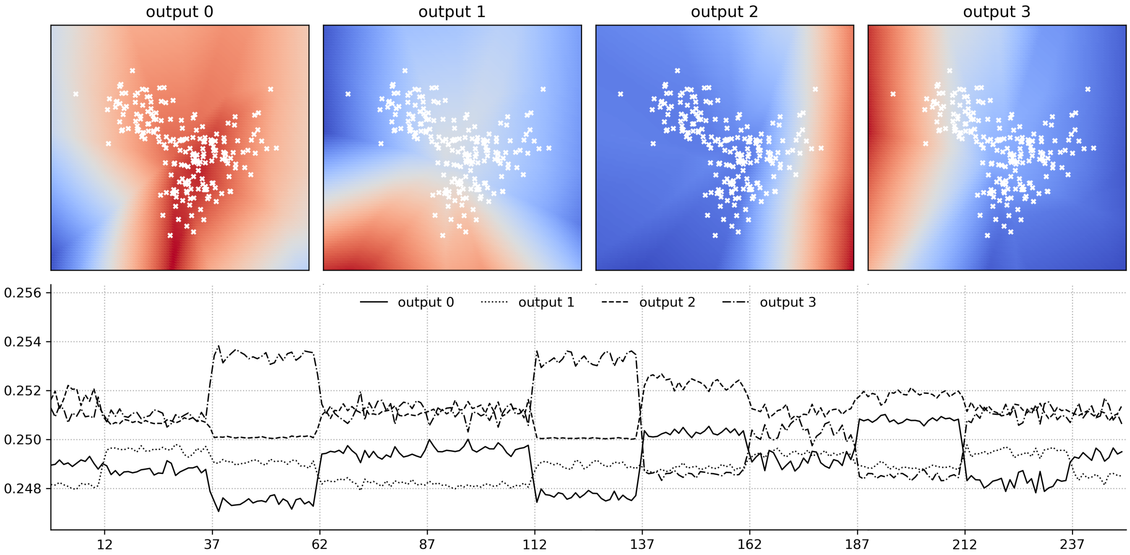
  if so, **increment a counter**
- check if counter *exceeds* threshold-based value of required test passes

  if so, **indicate drift**
- after detection, *forget* past activations

Pseudocode of the proposed approach:

for all $\mathcal{DS}_k \in \mathcal{DS}$ **do**
  $c \leftarrow \mathcal{NN}(\mathcal{DS}_k)$
  if $\mathcal{C}$ is not empty **then**
    $a \leftarrow 0$
    for all $e_i \in e$ **do**
      for all $r_i \in r$ **do**
        $cc \leftarrow$ random s from $c[e_i]$
        $pc \leftarrow$ random s from $C[e_i]$
        $p \leftarrow S(pc, cc)$
        if $p < \alpha$ **then**
          increment a
        end if
      end for
    end for
    if $a > \theta \times e \times r$ **then**
      $\mathcal{C} \leftarrow \emptyset$
      indicate drift in chunk k
    end if
  end if
  store c in $\mathcal{C}$
end for

Intuition behind the proposed approach



Experiment design

data streams, evaluation criteria, planned experiments

Experiment design

- Implementation in *Python* programming language.
- Libraries: *stream-learn*, *numpy*.
- Synthetic non-stationary data streams generated using *stream-learn* generator.

Two planned experiments

- Analysis of method's hyperparameters influencing its sensitivity (*alpha* and *threshold*).
- Comparison with reference approaches.

Parameter	Configuration
Number of chunks	250
Chunk size	200
Number of features	30, 60, 90 (30% informative)
Drift frequency	3, 5, 10, 15 drifts
Drift dynamics	sudden, gradual
Replications	10

Drift Detection Errors

The evaluation of drift detection:

- → Classification quality is not a good indicator of detector's performance.
- → Hence, the evaluation was based on three **drift detection error** measures:
  - **D1** – The average distance of each detection to the nearest drift,
  - **D2** – The average distance of each drift to the nearest detection,
  - **R** – The adjusted ratio of the number of drifts to the number of detections.

Results of experiments

Hyperparameter selection

- The evaluated hyperparameters were *alpha* and *threshold*, responsible for method's sensitivity.
- The results of three measures are interpreted as an RGB image, where red channel shows D1, green channel – D2 and blue channel – R error.

  The best results across three measures are indicated as dark regions.
- There's a range of two hyperparameters offering a optimal drift detection across all three measures.
- The following parameter combinations were selected for the comparison experiment: for *gradual* drifts $\alpha = 0.13$, $\theta = 0.26$; for *sudden* drifts $\alpha = 0.07$, $\theta = 0.0.19$.

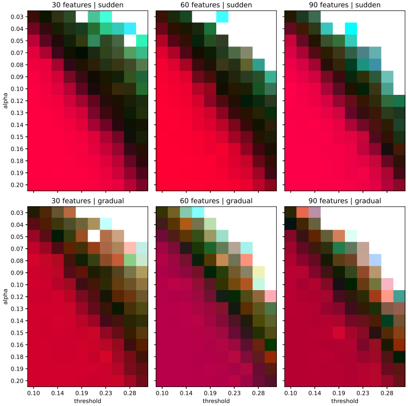


Figure: Results of hyperparameter selection experiment for all data streams, presented as an RGB image

Comparison with reference approaches

Acronym	Name and reference	Category
MD3	**Margin Density Drift Detector** [T. Sethi and M. Kantardzic, 2015]	Unsupervised with label request
OCDD	**One-class Drift Detector** [Ö. Gözüaçık and F. Can, 2021]	Unsupervised
CDDD	**Centroid Distance Drift Detector** [J. Klikowski, 2022]	Unsupervised
ADWIN	**Adaptive Windowing** [A. Bifet and R. Gavalda, 2007]	Supervised
DDM	**Drift Detection Method** [J. Gama et al., 2004]	Supervised
EDDM	**Early Drift Detection Method** [M. Baena-Garcia et al., 2006]	Supervised

Comparison with reference approaches

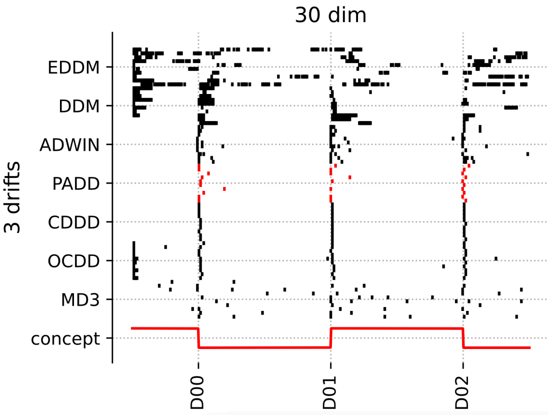


Figure: Results for a single stream type – the moments of drift detection across all ten replications of evaluation. Ideal detection is indicated with vertical lines overlapping with ticks on x-axis.

Comparison with reference approaches

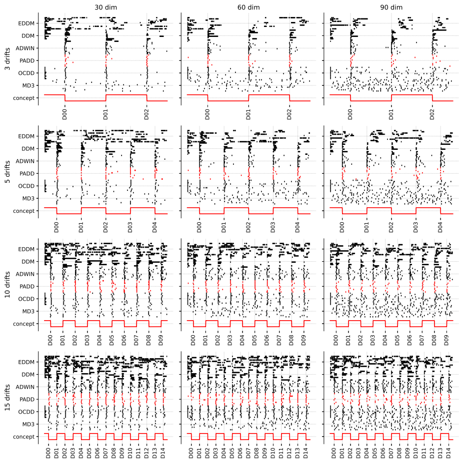


Figure: All results for *sudden* drift

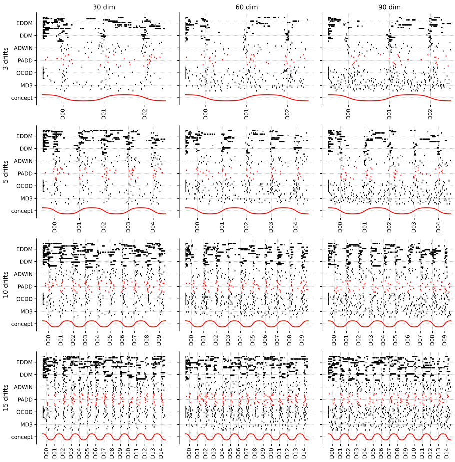
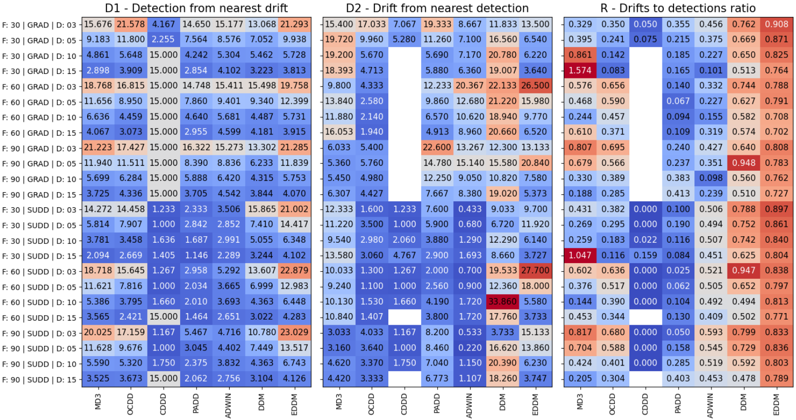


Figure: All results for *gradual* drift

Comparison with reference approaches



D1 - Detection from nearest drift

	MD3	OCDD	CDDD	PADD	ADWIN	DDM	EDDM
F: 30 | GRAD | D: 03	15.676	21.578	4.167	14.650	15.177	13.068	21.293
F: 30 | GRAD | D: 05	9.183	11.800	2.255	7.564	8.576	7.052	9.938
F: 30 | GRAD | D: 10	4.861	5.648	15.000	4.242	5.304	5.462	5.728
F: 30 | GRAD | D: 15	2.898	3.909	15.000	2.854	4.102	3.223	3.813
F: 60 | GRAD | D: 03	18.768	16.815	15.000	14.748	15.411	15.498	19.758
F: 60 | GRAD | D: 05	11.656	8.950	15.000	7.860	9.401	9.340	12.399
F: 60 | GRAD | D: 10	6.636	4.459	15.000	4.640	5.681	4.487	5.731
F: 60 | GRAD | D: 15	4.067	3.073	15.000	2.955	4.599	4.181	3.915
F: 90 | GRAD | D: 03	21.223	17.427	15.000	16.322	15.273	13.302	21.285
F: 90 | GRAD | D: 05	11.940	11.511	15.000	8.390	8.836	6.233	11.839
F: 90 | GRAD | D: 10	5.699	6.284	15.000	5.888	6.420	4.315	5.753
F: 90 | GRAD | D: 15	3.725	4.336	15.000	3.705	4.542	3.844	4.070
F: 30 | SUDD | D: 03	14.272	14.458	1.233	2.333	3.506	15.865	21.002
F: 30 | SUDD | D: 05	5.814	7.907	1.000	2.842	2.852	7.410	14.417
F: 30 | SUDD | D: 10	3.781	3.458	1.636	1.687	2.991	5.055	6.348
F: 30 | SUDD | D: 15	2.094	2.669	1.405	1.146	2.289	3.244	4.102
F: 60 | SUDD | D: 03	18.718	15.645	1.267	2.958	5.292	13.607	22.879
F: 60 | SUDD | D: 05	11.621	7.816	1.000	2.034	3.665	6.999	12.983
F: 60 | SUDD | D: 10	5.386	3.795	1.660	2.010	3.693	4.363	6.448
F: 60 | SUDD | D: 15	3.565	2.421	15.000	1.464	2.651	3.022	4.283
F: 90 | SUDD | D: 03	20.025	17.159	1.167	5.467	4.716	10.780	23.029
F: 90 | SUDD | D: 05	11.628	9.676	1.000	3.045	4.402	7.449	13.517
F: 90 | SUDD | D: 10	5.590	5.320	1.750	2.375	3.832	4.363	6.743
F: 90 | SUDD | D: 15	3.525	3.673	15.000	2.062	2.756	3.104	4.126

D2 - Drift from nearest detection

	MD3	OCDD	CDDD	PADD	ADWIN	DDM	EDDM
F: 30 | GRAD | D: 03	15.400	17.033	7.067	19.333	8.667	11.833	13.500
F: 30 | GRAD | D: 05	19.720	9.960	5.280	11.260	7.100	16.560	6.540
F: 30 | GRAD | D: 10	19.200	5.670		5.690	7.170	20.780	6.220
F: 30 | GRAD | D: 15	18.393	4.713		5.880	6.360	19.007	3.640
F: 60 | GRAD | D: 03	9.800	4.333		12.233	20.367	22.133	26.500
F: 60 | GRAD | D: 05	13.840	2.580		9.860	12.680	21.220	15.980
F: 60 | GRAD | D: 10	11.880	2.140		6.570	10.620	18.940	9.770
F: 60 | GRAD | D: 15	16.053	1.940		4.913	8.960	20.660	6.520
F: 90 | GRAD | D: 03	6.033	5.400		22.600	13.267	12.300	13.133
F: 90 | GRAD | D: 05	5.360	5.760		14.780	15.140	15.580	20.840
F: 90 | GRAD | D: 10	5.450	4.980		12.250	9.050	10.820	7.580
F: 90 | GRAD | D: 15	6.307	4.427		7.667	8.380	19.020	5.373
F: 30 | SUDD | D: 03	12.333	1.600	1.233	7.600	0.433	9.033	9.700
F: 30 | SUDD | D: 05	11.220	3.500	1.000	5.900	0.680	6.720	11.920
F: 30 | SUDD | D: 10	9.540	2.980	2.060	3.880	1.290	12.290	6.140
F: 30 | SUDD | D: 15	13.580	3.060	4.767	2.900	1.693	8.660	3.727
F: 60 | SUDD | D: 03	10.033	1.300	1.267	2.000	0.700	19.533	27.700
F: 60 | SUDD | D: 05	9.240	1.100	1.000	2.560	0.900	12.360	18.000
F: 60 | SUDD | D: 10	10.130	1.530	1.660	4.190	1.720	33.860	5.580
F: 60 | SUDD | D: 15	10.840	1.407		3.800	1.720	17.760	3.733
F: 90 | SUDD | D: 03	3.033	4.033	1.167	8.200	0.533	3.733	15.133
F: 90 | SUDD | D: 05	3.160	3.640	1.000	8.460	0.220	16.620	13.860
F: 90 | SUDD | D: 10	4.620	3.370	1.750	7.040	1.150	20.390	6.230
F: 90 | SUDD | D: 15	4.420	3.333		6.773	1.107	18.260	3.747

R - Drifts to detections ratio

	MD3	OCDD	CDDD	PADD	ADWIN	DDM	EDDM
F: 30 | GRAD | D: 03	0.329	0.350	0.050	0.355	0.456	0.762	0.908
F: 30 | GRAD | D: 05	0.395	0.241	0.075	0.215	0.375	0.669	0.871
F: 30 | GRAD | D: 10	0.861	0.142		0.185	0.227	0.650	0.825
F: 30 | GRAD | D: 15	1.574	0.083		0.165	0.101	0.513	0.764
F: 60 | GRAD | D: 03	0.576	0.656		0.140	0.332	0.744	0.788
F: 60 | GRAD | D: 05	0.468	0.590		0.067	0.227	0.627	0.791
F: 60 | GRAD | D: 10	0.244	0.457		0.094	0.155	0.582	0.708
F: 60 | GRAD | D: 15	0.610	0.371		0.109	0.319	0.574	0.702
F: 90 | GRAD | D: 03	0.807	0.695		0.240	0.427	0.640	0.808
F: 90 | GRAD | D: 05	0.679	0.566		0.237	0.351	0.948	0.783
F: 90 | GRAD | D: 10	0.330	0.389		0.383	0.098	0.560	0.762
F: 90 | GRAD | D: 15	0.188	0.285		0.413	0.239	0.510	0.727
F: 30 | SUDD | D: 03	0.431	0.382	0.000	0.100	0.506	0.788	0.897
F: 30 | SUDD | D: 05	0.269	0.295	0.000	0.190	0.494	0.752	0.861
F: 30 | SUDD | D: 10	0.259	0.183	0.022	0.116	0.507	0.742	0.840
F: 30 | SUDD | D: 15	1.047	0.116	0.159	0.084	0.451	0.625	0.804
F: 60 | SUDD | D: 03	0.602	0.636	0.000	0.025	0.521	0.947	0.838
F: 60 | SUDD | D: 05	0.376	0.517	0.000	0.062	0.505	0.652	0.797
F: 60 | SUDD | D: 10	0.144	0.390	0.000	0.104	0.492	0.494	0.813
F: 60 | SUDD | D: 15	0.453	0.344		0.130	0.409	0.502	0.771
F: 90 | SUDD | D: 03	0.817	0.680	0.000	0.050	0.593	0.799	0.833
F: 90 | SUDD | D: 05	0.704	0.588	0.000	0.158	0.545	0.729	0.836
F: 90 | SUDD | D: 10	0.424	0.401	0.000	0.285	0.519	0.592	0.803
F: 90 | SUDD | D: 15	0.205	0.304		0.403	0.453	0.478	0.789

Figure: Color-coded table showing averaged results for all methods and data streams. Blue cells indicate low errors, while red cells – a high error.

Comparison with reference approaches

→ In the case of $D1$ error, the presented method is significantly the **best lone leader** of comparison.

→ For the R error, the presented solution – while having the best ranking value – is statistically significantly **co-dependent to the ADWIN and OCDD**.

→ For the $D2$, the best results were obtained by the OCDD method, and the **PADD results are statistically dependent on the second method in terms of quality** (ADWIN) with the minimal difference in ranks.

→ It is worth emphasizing that those criteria **should not be used independently** to evaluate methods, and it is the juxtaposition of all three that describes the proper and effective method operation.

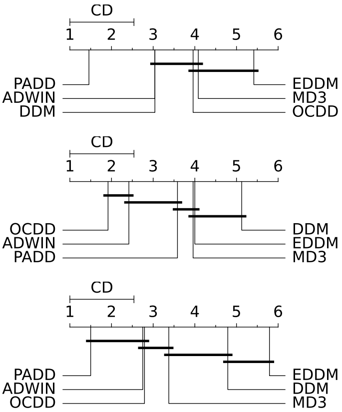


Figure: Results of a post-hoc *Nemenyi* test based on *Wilcoxon Signed Rank* across all three errors: D1 (top), D2 (center), R (bottom).

Conclusions and future works

- This work proposed an unsupervised *Parallel Activations Drift Detector* that utilizes an **untrained neural network** to recognize significant changes in the posterior distribution of the data stream to signal concept drift according to the result of a **statistical test** stabilized by **a pool of replications**.
- The conducted experimental evaluation allowed to demonstrate that the proposed PADD method states a valuable tool in the context of reference methods.

- As part of future work, it is planned to expand the hyperparameter calibration study – considering the introduction of a non-parametric version of the method.

Unsupervised Concept Drift Detection based on Parallel Activations of Neural Network

Joanna Komorniczak, Paweł Ksieniewicz

Department of Computer Systems and Networks
Wroclaw University of Science and Technology

Discovering Drift Phenomena in Evolving Landscape
ACM KDD 2024 | Barcelona, Spain