Unsupervised Assessment of Landscape Shifts Based on Persistence Entropy and Topological Preservation

Sebastián Basterrech[0000−0002−9172−0155]

Department of Applied Mathematics and Computer Science
Technical University of Denmark, Kongens Lyngby, Denmark
sebbas@dtu.dk

Abstract. Concept drift typically refers to the analysis of changes in data distribution. A drift in the input data can have negative consequences on a learning predictor and the system's stability. The majority of concept drift methods emphasize the analysis of statistical changes in non-stationary data over time. In this context, we consider a slightly different perspective, where concept drift also integrates significant changes in the topological characteristics of the data stream.
In this article, we introduce a novel framework for monitoring changes in multi-dimensional data streams. We explore a generalization of the standard concept drift focusing on the changes in the topological characteristics of the data stream. Our developed approach is based on persistence entropy and topology-preserving projections in a continual learning scenario. The framework operates in both unsupervised and supervised environments. To demonstrate the utility of the proposed framework, we analyze the model across three scenarios using data streams generated with MNIST samples. The obtained results reveal the potential of applying topological data analysis for shift detection and encourage further research in this area.

Keywords: Distribution shifts · Persistence entropy · Dimensionality reduction · Concept Drift · Self-organizing maps

1 Introduction

In continual learning scenarios, designing a machine learning (ML) model that is robust to distribution shifts is a crucial objective. Traditional ML methods are susceptible to data perturbations, and shifts in input data distribution can significantly affect a model's performance. Concept drift detectors encompass a family of techniques developed to analyze and detect distribution changes in the context of streaming data and time series. The concept is based on changes in the statistical characteristics of the data over time [18]. In this study, we explore a generalization of the previous definition of concept drift that emphasizes changes in the topological characteristics of the data. There are objects that are essentially equivalent to each other if we consider "equivalence" in the sense that it is possible to define a simple continuous transformation that approximates one object to the other. The essence of an object remains unchanged under simple transformations, such as rotation, translation, scaling, and other types of continuous transformations [12]. On the other hand, there are objects that are essentially different, as it is not feasible to find a continuous transformation to transform one object into another [10]. Or at least, it is not easy to find such a transformation with low computational resources. The field of Topological Data Analysis (TDA), specifically through the mathematical formalism of algebraic topology, defines these concepts of equivalences and differences between objects. Persistent Entropy (PE) is a measure based in Shannon entropy that provides a summary of the geometric information derived from the topological features of a cloud of points [2,30]. It has been successfully used to effectively distinguish chaotic and periodic time series [30].

In this work, we introduce an extension of the classic concept drift that integrates algebraic topology. We extend the concept of drift, which is based on statistical and geometrical information, to another concept that incorporates the notion of *essential sameness* and *essential difference* [10]. We observe drift when the significant geometric characteristics of a cloud of points *essentially* change, becoming different from those of another cloud of points. Figure 1 illustrates examples of objects that can be deformed using simple continuous transformations such as rotation, stretching, bending, and scaling. The first row of the figure displays three digits that can be considered topologically equivalent, while the three objects in the bottom row can also be considered equivalent. However, it is not possible to transform any object in the first row into a digit in the second row. To quantify these types of topological changes, we take into account concepts provided by TDA. We empirically investigate the changes in the distribution of high-dimensional data in an unsupervised scenario, using tools from persistence homology. We develop a general-purpose framework that projects the input data into a low-dimensional space using a dimensionality reduction technique that preserves the topological features of the data. We then apply metrics of persistent homology to evaluate significant changes. The projection is made using Self-Organizing Maps (SOM), which is a mapping technique to reduce dimensionality while preserving the topological characteristics of the input space [25]. In the latent space, we explore the potential of persistent homology to find significant differences among data coming from consecutive chunks. We use the metric of persistent entropy, which summarizes the analysis of persistent homology in a single value [1]. In summary, this work offers the following contributions.

(i) We introduce a general-purpose framework for concept drift detection that operates in both supervised and unsupervised environments. This framework utilizes dimensionality reduction through a topological preserving mapping and evaluates significant changes using persistent homology.

(ii) The framework delivers results using a p-value score. When each chunk of data arrives, a non-parametric statistical test is performed, facilitating easy monitoring of drifts. The hypothesis test is conducted on the values of persistent entropy. Since the framework provides a p-value score, the decision regarding the absence or presence of a drift is both robust and fast.

(iii) We provide an initial experimental evaluation with promising results across three case studies. We compare three dimensionality reduction techniques to evaluate the impact of preserving topological features when projecting data from the input space to the latent space. The results show the benefits of combining a topology-preserving mapping with information regarding persistent homology.

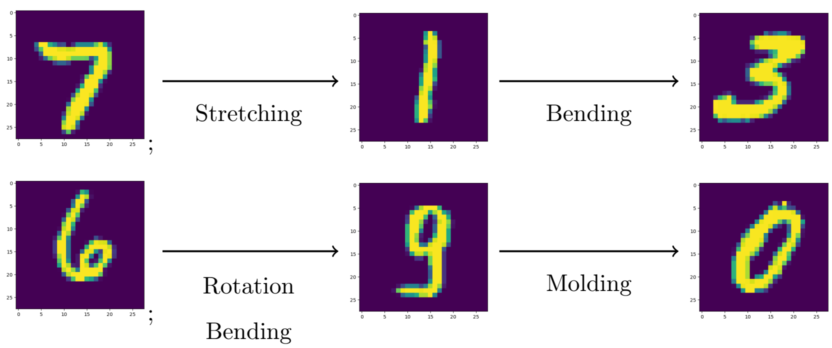

Fig. 1: Examples of objects that can be deformed and transformed into another object equivalent in terms of topology. However, some shapes cannot be considered equivalent because it is not possible to define a sequence of simple transformations to deform an original shape into another one while maintaining the original structure. For instance, from any of the digits at the top of the figure, we cannot form any of the digits at the bottom of the figure.

2 Background

2.1 Concept Drift

Drift detection techniques constitute a family of computational methods for detecting changes in the distribution of time series and data streams. Shifts in data distribution can occur in different forms, probably the most accepted categories are: sudden, gradual, and incremental shifts [42,41]. A majority of drift detection techniques employ a classifier to categorize incoming instances, and the predictor generates a class label for each input instance, which is then compared to the actual class label. Subsequently, the accuracy is assessed and utilized as a tool to determine whether a drift has occurred. When the classifier's accuracy significantly decreases, then it is assumed that the data distribution has changed [23,17,19]. In this scenario, the effectiveness of various ensemble classifiers has been examined [29,26,13,27]. However, this approach can be applied only in a supervised context, and it requires the presence of ground truth labels, which are not always available. Another set of methods relies on statistical tests directly over the raw data [40,7,20,11]. A lot of distribution shift detection techniques depend on the computation of the empirically estimated distributions. These approaches are sensitive to outliers and noise, and raw data analysis (e.g. applying density estimation) may be affected by the curse of dimensionality [4,5]. Several studies propose to compare summary of statistics and aggregation metrics of the raw data, for instance, Cumulative Sum and Exponentially Weighted Moving Average [34]. For a more comprehensive review of the latest advancements in the use of data descriptors for concept drift detection, see [14,32,22].

2.2 Persistent Entropy

Persistent homology is a key instrument in TDA as it may be used to describe the inherent structure of complex objects such as manifolds [33]. Specifically, persistent homology studies the evolution of k-dimensional topological features (often referred to as *holes*) along a sequence of high-dimensional complex objects (named *simplicial complexes*) [2,10]. We understand topological features as shapes or data that remain unchanged under certain continuous transformations, such as connected components, independent cycles, and holes [10]. Persistent homology tracks changes of topological features of data that persist across multiple scales, following a specific algorithm that analyzes the connectivity information among the data points [10]. Persistent Entropy, based in Shannon entropy, provides a summary of the information derived from persistent homology [2]. It is a measure for finding significant differences in the geometrical distribution of data points [2,31]. For a comprehensive and detailed exploration about TDA and persistent homology, see [10,31].

3 Methodology

This section outlines the contributions made in this brief article. First, we discuss the approach for transforming the input patterns into a different landscape that simplifies the analysis of drifts. Next, we introduce the process for estimating geometric changes between data points in different chunks. Finally, we present the sequence of modules that compose the developed framework.

3.1 Creation of the latent space

Monitoring and detecting distribution shifts is specially harder in the case of high-dimensional data. Even if some attempts have been introduced in the literature for sparse multivariate time series [43,38], many algorithms are not properly scalable, for instance the methods based in probability mass distribution. In addition, the norm computation also has limitations in a high-dimensional space (e.g. Euclidean norm) [39]. For this reason, instead of working directly over the points in the original space, we analyze an approach that project the input points into a latent space. The projection is made using dimensionality reduction (DR) techniques, which is a common approach for handling the data in high dimensions. We assume a context that performing a similarity analysis in the original input space may be computationally expensive. Consequently, it is often more resource-efficient to first convert the data into a latent space, and then carry out the similarity analysis. Here, we investigate the projections generated by Self Organizing Maps (SOMs) (also called Kohonen networks) and we compare the results with other two popular DR models (a linear projection (PCA) and the Kernel PCA). The selection of an adequate data descriptor is crucial for ensuring a proper geometry in the latent space preserving the main features of the original space.

SOM is a bio-inspired method that combines concepts from Hebbian learning, vector quantization, and competitive learning [16,25]. Real-world data most often contain redundancies and inherent correlations among the variables. SOM is a two-layered neural network that transforms intricate relationships among high-dimensional data into straightforward geometric relationships on a standard lattice, typically a two-dimensional grid [15]. Despite its simplicity, the SOM method is effective as a Dimensionality Reduction (DR) technique, a clustering method, and a visualization tool for high-dimensional data [35]. Another advantage is that the method is applicable to unsupervised problems and has the capability to preserve the most important topological features of the reference data [25].

3.2 Assessing shifts in the latent space

Recently, it was introduced a clustering method based in self-organizing maps for assessing distribution changes in data streams with high dimensional data [6]. Self-organizing map is used for projecting the input data into a latent space, then the analysis is done in the latent space, where the authors computed a distance matrix between the input pattern and cluster centers. The assessment of the distribution shifts is done by applying a statistical summary. This approach of using a data descriptor was also applied in [20,21], and it is commonly used in methods based on kernel projections [37].

In this work, we modify the analysis based on distances and statistical summaries of the points in the latent space, to an approach that assesses topological changes according to the homological characteristics of the points in the latent space. Once the dimensionality reduction is done, a distance matrix is computed. The distance matrix has the information between the projected point and the cluster centers. Instead of working directly with the coordinates of points in the latent space generated by the DR method, these are summarized using relative locations. Coordinates are arbitrarily selected, and more often than not, they don't consider any property of the data itself. There are even problems in cases where the coordinates are not natural in any sense [9]. Therefore, a relative location of the latent space points is computed, calculating the distance between the mass centers and the projected points. Hence, our focus is on the geometric properties of the latent space, independent of the chosen coordinates in the latent space. The methodological approach is illustrated in Figure 2. Note that, the approach is general in the sense that can be used any type of DR technique.

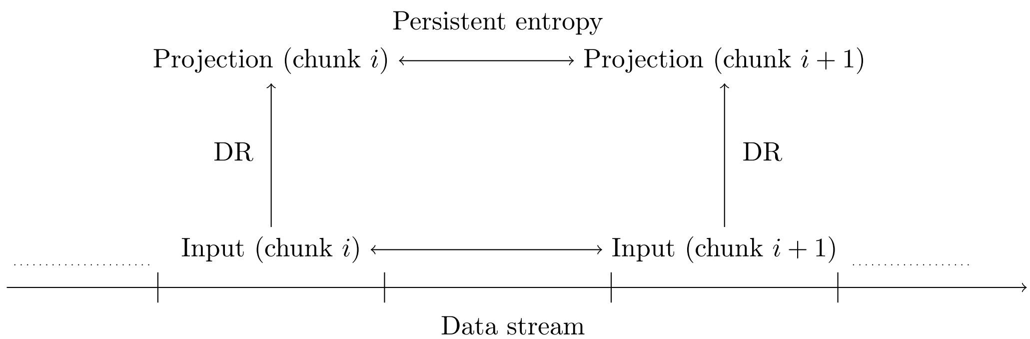

Fig. 2: Assessing the topological changes in the latent space: comparing the persistent entropy of projected points using a Dimensionality Reduction (DR) technique.

3.3 Pipeline of the proposed framework

The proposed framework for assessing topological changes based in persistent entropy involves the following steps:

- Dimensionality reduction: When a chunk of data comes, then the input data points are projected using a DR method.
- Embedding of the geometrical properties in the latent space: For each projected data point, the distance matrix is computed using the data representative of each cluster (centroid distance matrix).
- Representation of topological features: For each chunk of data, a persistent diagram is computed using the cloud of points created by the centroid distance matrices. Computation of the persistence entropy per chunk (the infinity bar in the computation of persistent entropy is ignored [36].
- Statistical analysis. A final index is computed between the persistent entropy in one chunk of data and its subsequent chunk. Finally, we generate a sequence of p-value scores computed with the non-parametric Mann-Whitney U test. This sequence of p-value provides the information about significant changes in the topological properties of the data.

Note, an initial training phase is performed for computing the initial clustering and its representative mass centers. An initial time-window of the data stream is used for training the SOM weights and other global parameters. After this initial training phase, we continue the learning process following a usual continual learning scenario.

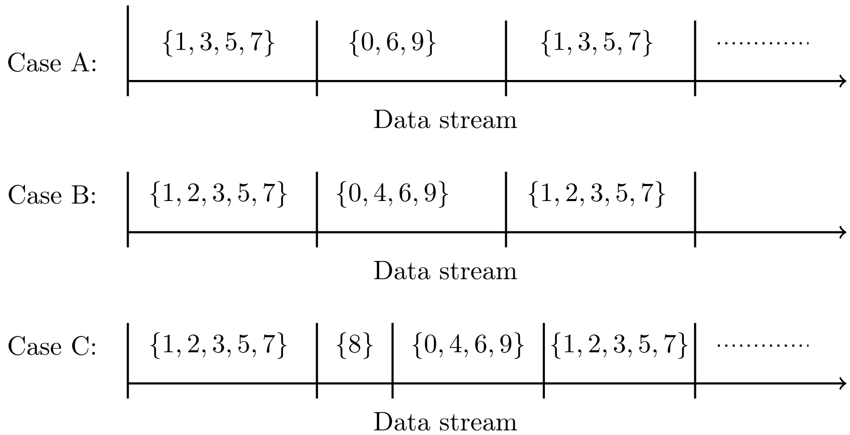

Fig. 3: Generation data streams. Data streams were generated with the MNIST samples interchanged among the different topological types. The graphics illustrate the transition between the sequence of images from one topological type to another type.

4 Experimental results

In this section, we demonstrate the utility of the proposed approach for monitoring and detecting topological changes in a data stream in the context of continual learning. We designed the experiments to evaluate and contrast the efficacy of the dimensionality reduction methods previously discussed: PCA, Kernel PCA, and SOM. In this ongoing work, we analyze three problems with annotated data streams. There exists a deficiency in the availability of extensive and diverse real-world data streams within high-dimensional spaces for analyzing the impact of distribution changes [41]. This inconvenience is more relevant in the domain of unsupervised analysis of streaming data. As a consequence, we created three synthetic data streams with annotated shifts. The created stream has samples from MNIST [28]. The annotations indicate the time-stamps where was injected the topological differences between two cloud of points.

4.1 Benchmark data

We generated three synthetic datasets using the MNIST dataset, following the methodology outlined in [32,3,6]. The procedure consists of creating a stream with chunks of samples that follow a specific distribution, and then alternating these chunks with chunks of instances from a different distribution. By construction, the timing of the distribution changes is predefined. As a consequence, we have the exact time-stamps wherein a drift was injected. We have created the data to check if the method is able to detect "essential" changes between sequences of digits. We divided the digits into three groups: without any holes (without any enclosed space), with one hole, and with two holes. We denote the three experimental studies as: A, B and C. For all the cases, we analyze 20000 samples. For case studies A and B, the drift is injected every 1000 samples, and for scenario C, the drift occurs every 500 samples.

In case study A, we analyze a data stream where changes occur between chunks with digits in $\{1, 3, 5, 7\}$ and chunks with digits $\{0, 6, 9\}$ (digits with one hole). Case study B also includes the digits 2 and 4. These digits are problematic due to variations in handwriting. Some people write the numbers 2 and 4 without any hole, while others write them with one hole, it depends on handwriting style. We decided to create the data stream considering the following exchanges between points in $\{1, 2, 3, 5, 7\}$ and $\{0, 4, 6, 9\}$. Finally, we also evaluate in the study case C a data stream with the number 8, which is not topologically equivalent to any of the other digits. Case C exchanges samples from the three subsets 1,2,3,5,7, 0,4,6,9, and 8. Figure 3 depicts how the streaming data was generated, in each of the three scenarios. Case study A and B have 20 exchanges between subset of digits with holes and without holes, and case study C has 40 exchanges between subsets of digits without hole, with one hole, and two holes.

4.2 Experimental settings

For each of the three case study, we used the first 20% of samples for training the parameters of the SOM algorithm. This training was made offline, as a pre-phase of the continual learning process. The SOM algorithm has a grid with 10×10 neurons. We also analyzed three values for the chunk size parameter $\{50, 100, 250\}$. The quality assessment of the monitoring for the shifts was done using a p-value computed with the non-parametric Mann-Whitney U test. We also use the same MNIST data to evaluate PCA and Kernel-PCA as drift detector tools. Also for PCA and Kernel-PCA an initial time window comprising 20% of the data stream was used for training both clustering methods. Subsequently, we apply the trained methods to project the data points into a latent space, following a CL setting.

4.3 Implementation details

The implementation of the methods developed during our investigation, as well as the experimental environment, was carried out using the *Python* v3.9 programming language. Several libraries were utilized to facilitate this process, including *NumPy* v1.19.5 for numerical computations, *stream-learn* v0.8.16 for handling data streams, *Sklearn* v1.0.2 for machine learning tasks [8], and the *Persim* v0.3.6 package for operations related to persistent homology [36]. The source code of our investigation and the datasets are available in the git repository[1].

[1] https://github.com/sebabaster/Drift-persistence

4.4 Results

An example of the characteristics of the data stream in the case C is illustrated in Figure 4. The figure illustrate the evolution of the mean value of each centroid distance matrices. In addition, we show the application of Pruned Exact Linear Time (PELT) algorithm [24], which is an offline approach (it is only as a visualization of the complexity of the problem). The PELT method is recognized for its computational efficiency, when compared with other change-point detection techniques.

Figure 4 provides an illustration of the data stream characteristics for benchmark problem C. This figure depicts the evolution of the mean values computed for each centroid distance matrix. Furthermore, it shows the detected points using the PELT technique in an offline context. The background colors represent the changes detected by PELT, and the vertical green lines indicate the injected shifts. This visualization serves as a visual representation of the problem's complexity. The other experiments simulate a continual learning environment, making them even more complex than this problem using offline settings. Figure 6 presents the results for the three DR techniques. These techniques were evaluated using benchmark data C, which was split into chunks of 250 samples each. The vertical lines (represented by green dashed lines) indicate the injected shifts. The horizontal line represents a p-value of 0.05. According to the results, the linear projection is unable to accurately predict the shifts. This finding aligns with other studies in the literature that discuss the limitations of linear projections in detecting distribution shifts [22,6]. The performance of SOM appears to be slightly better than that of Kernel-PCA. For instance, refer to the p-values in the chunks between 30 and 35. The chunk size is a key parameter at the moment of working in on-line settings. The influence of the chunk size is shown in Figures 7, 8, and 9, as well as in Table 1. Figure 7 shows the results of using linear projections before the analysis of the persistence entropy. There are two graphs: the top graph shows the results for chunks with 50 instances, and the bottom graph shows the results for chunks of 100 samples. This figure also highlights the limitations of linear projections for solving this specific problem, as the method provides few alarms and detects only a low number of drifts. Figure 8 presents the results of Kernel PCA for chunks with 50 and 100 samples. Similar graphics are depicted in Figure 9 where are presented the results of SOM for chunks with 50 and 100 samples. Table 1 summarizes the experiments conducted to evaluate the impact of chunk size. The last two columns show the flags generated by the model using p-values with significance levels of 0.05 and 0.1. Additionally, we show the number of injected drifts in the experiment. Note that this number is an approximation due to the anomalies that can exist in the datasets (digits with a different number of holes than expected). According to the table, it seems that SOM may provide results using p-values with a 0.05 level of significance, while Kernel-PCA obtains better results when a p-value with a 0.1 level of significance is considered. As one might intuitively expect, smaller chunks decrease the quality of the geometric pattern analysis. On the other hand, larger chunks can encompass more than one shift. The optimal chunk size should be determined experimentally, taking this trade-off into account.

5 Discussion

As far as we know, this is the first work that attempts to define drifts in terms of topology. Our research hypothesis was to investigate whether the use of TDA is helpful for monitoring topological changes and detecting "essential" differences between chunks of objects. In other words, if we have a sequence of different types of donuts and then suddenly the sequence receives coffee cups, then a good drift detector would detect a drift (using the traditional notion of drift). However, if we consider purely topological information, both shapes are considered equivalent. Therefore, it would not be appropriate to consider this type of change as a drift. To answer this research question, we implemented a set of experiments. We defined three case studies based on changes in the number of holes in the digits. The digits were not checked individually (they were sampled using a uniform distribution). Consequently, there may be instances where a digit was expected to have a specific number of holes, but the sample has a different number of holes. For examples, see Figure 5. Therefore, to generate new datasets with better quality labels would be beneficial for future analysis. In this study, we did not evaluate different strategies for mitigating the catastrophic forgetting problem. When a drift is detected, it is necessary to define an appropriate strategy to fine-tune the framework. This is especially important for SOM. We left this evaluation for future research. Another limitation of our experimental analysis is that the framework is composed of two modules: dimensionality reduction and persistence entropy. Further investigation is required to analyze the relevance of each of these modules and how they affect drift detection.

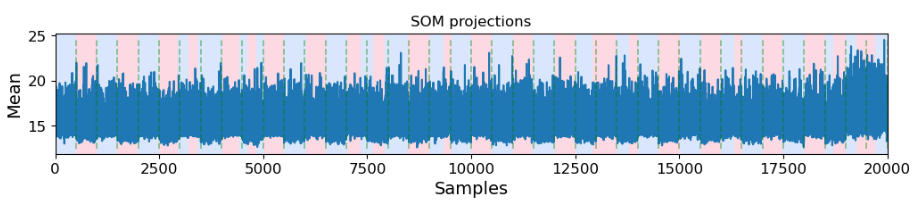

Fig. 4: Example of the latent space. Off-line analysis of the latent space generated by the SOM projections, and applying change point detection over the distance matrix.

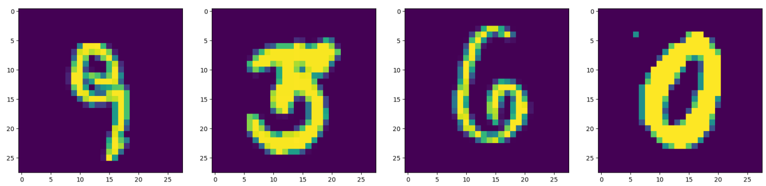

Fig. 5: An example of digits with a structure different from the one assumed in the case studies. The first image has two holes instead of one. The second image, which can be a number 3 or 5, has a hole. The third image doesn't have any holes. The last image is not a connected component (it has an isolated pixel in the top-left).

6 Conclusions and future work

We introduced a novel approach to concept drift detection, leveraging algebraic topology and persistent entropy. We broaden the scope of concept drift, which is typically associated solely with statistical distribution changes in incoming data. In the explored approach, we also integrate the definition of drift to detect significant changes in topological features. The framework uses the SOM algorithm to transform the input space, reducing its dimensionality while preserving its topological characteristics. Then, the analysis of data drifts is conducted in the latent space. We explore the potential of persistent entropy to identify significant differences among data from consecutive chunks. We showed the performance of the method over three study cases (based on the MNIST dataset), and we compared the performance with PCA and Kernel-PCA. The proposed method does not make any assumption about the data distribution, and it can be applied in both supervised and unsupervised problems. We believe that this work is an initial step towards applying TDA and topographic maps in the area of concept drift.

A potential direction for future research could involve evaluating the framework with different types of data streams, where topological changes are driven by other characteristics. Additionally, it would be interesting to compare persistent entropy with other measures.

Table 1: Sensitive analysis of the chunk size for the three study cases, the three methods, and two levels of significance of the hypothesis test (0.05 and 0.1). The table shows the number of potential drifts (flags) detected by each method. The fourth column shows the number of injected drifts. This is an approximate number because, during the construction of the dataset, each image wasn't individually verified. Consequently, digits like 2, 4, 6, and 9 may sometimes have a hole and sometimes not.

Case study	Chunk size	Method	Drifts	Flags (p-value at 0.05 level)	Flags (p-value at 0.1 level)
A	50	SOM	20	13	30
		PCA		22	51
		KernelPCA		27	58
	100	SOM	20	9	16
		PCA		10	27
		KernelPCA		14	22
	250	SOM	20	3	5
		PCA		6	11
		KernelPCA		4	8
B	50	SOM	20	18	36
		PCA		17	34
		KernelPCA		17	39
	100	SOM	20	10	26
		PCA		6	12
		KernelPCA		10	21
	250	SOM	20	4	12
		PCA		3	4
		KernelPCA		9	14
C	50	SOM	40	53	75
		PCA		21	44
		KernelPCA		27	51
	100	SOM	40	43	53
		PCA		13	26
		KernelPCA		26	37
	250	SOM	40	40	43
		PCA		8	10
		KernelPCA		25	35

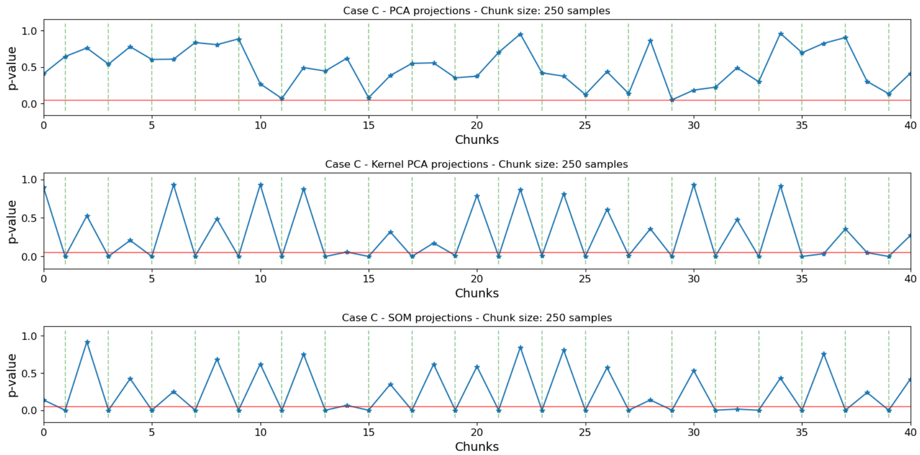

Fig. 6: Comparison among the three projections. The comparison was made over the dataset C with chunk size of 250 samples.

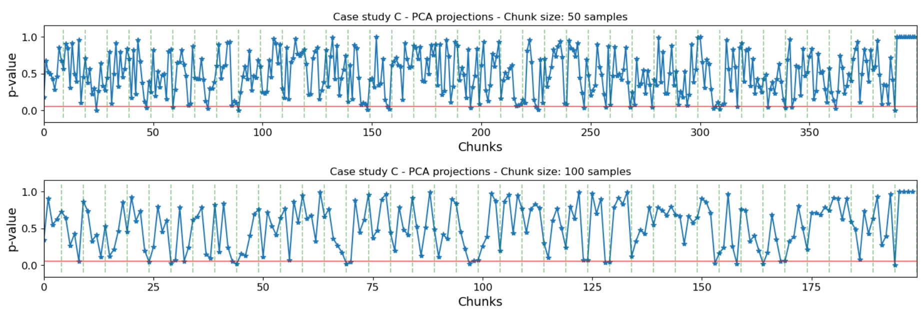

Fig. 7: Results of the method using PCA projections in case study C for chunk sizes of 50 and 100 samples.

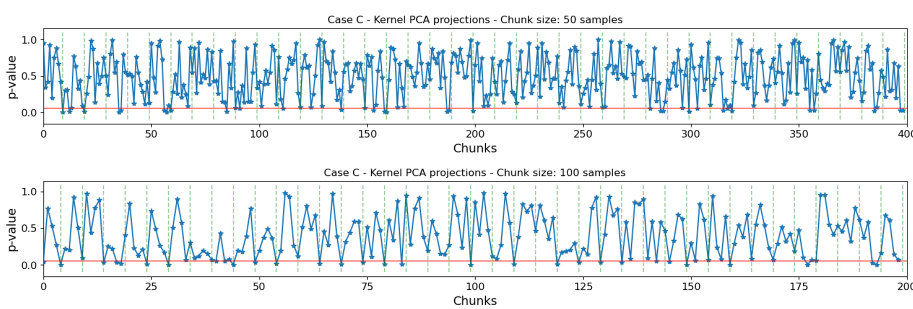

Fig. 8: Results of the method using Kernel PCA projections in case study C for chunk sizes of 50 and 100 samples.

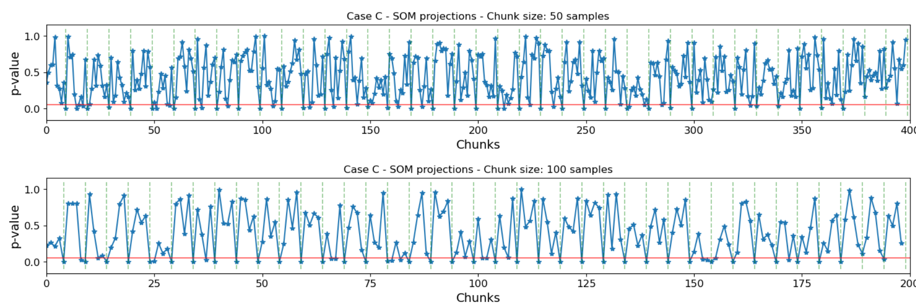

Fig. 9: Results of the method using SOM projections in case study C for chunk sizes of 50 and 100 samples.

References

1. On the stability of persistent entropy and new summary functions for topological data analysis. Journal of Intelligent Information Systems **52**(3), 637–655 (2019). https://doi.org/10.1007/s10844-017-0473-4
2. Atienza, N., Gonzalez-Diaz, R., Rucco, M.: Persistent entropy for separating topological features from noise in vietoris-rips complexes. Journal of Intelligent Information Systems **52**(3), 637–655 (2019). https://doi.org/10.1007/s10844-017-0473-4
3. Basterrech, S., Kasprzak, A., Platos, J., 0001, M.W.: A Continual Learning System with Self Domain Shift Adaptation for Fake News Detection. In: 10th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2023, Thessaloniki, Greece, October 9-13, 2023. pp. 1–10. IEEE (2023). https://doi.org/10.1109/DSAA60987.2023.10302539, https://doi.org/10.1109/DSAA60987.2023.10302539
4. Basterrech, S., Platoš, J., Rubino, G., Woźniak, M.: Experimental Analysis on Dissimilarity Metrics and Sudden Concept Drift Detection. In: Abraham, A., Pllana, S., Casalino, G., Ma, K., Bajaj, A. (eds.) Intelligent Systems Design and Applications. pp. 190–199. Springer Nature Switzerland, Cham (2023)
5. Basterrech, S., Woźniak, M.: Tracking changes using Kullback-Leibler divergence for the continual learning. In: 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 3279–3285 (2022). https://doi.org/10.1109/SMC53654.2022.9945547
6. Basterrech, S., Clemmensen, L., Rubino, G.: A Self-Organizing Clustering System for Unsupervised Distribution Shift Detection (2024), preprint: 2023 IEEE International Joint Conference on Neural Networks (IJCNNC)
7. Blanco, I.I.F., del Campo-Avila, J., Ramos-Jimenez, G., Bueno, R.M., Diaz, A.A.O., Mota, Y.C.: Online and non-parametric drift detection methods based on hoeffding's bounds. IEEE Trans. Knowl. Data Eng. **27**(3), 810–823 (2015)
8. Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J.: Api design for machine learning software: experiences from the scikit-learn project. In: ECML PKDD Workshop: Languages for Data Mining and Machine Learning. pp. 108–122 (2013)
9. Carlsson, E., Carlsson, G., Silva, V.: An algebraic topological method for feature identification. Int. Journal of Computational Geometry and Applications **16**, 291–314 (08 2006). https://doi.org/10.1142/S021819590600204X
10. Carlsson, G.: Topology and data. Bulletin of the American Mathematical Society (2), 255–308 (2009)
11. Clemmensen, L.H., Kjærsgaard, R.D.: Data Representativity for Machine Learning and AI Systems. arXiv preprint arXiv:2203.04706 (2023)
12. Dai, M., Kurtek, S., Klassen, E., Srivastava, A.: Statistical Shape Analysis, pp. 1197–1211. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-63416-2_778, https://doi.org/10.1007/978-3-030-63416-2_778
13. Du, L., Song, Q., Zhu, L., Zhu, X.: A selective detector ensemble for concept drift detection. The Computer Journal p. bxu050 (2014)
14. Faber, K., Corizzo, R., Sniezynski, B., Baron, M., Japkowicz, N.: WATCH: Wasserstein Change Point Detection for High-Dimensional Time Series Data. In: 2021 IEEE Int. Conf. on Big Data (Big Data). pp. 4450–4459 (2021). https://doi.org/10.1109/BigData52589.2021.9671962
15. Ferles, C., Papanikolaou, Y., Naidoo, K.J.: Denoising Autoencoder Self-Organizing Map (DASOM). Neural Networks **105**, 112–131 (2018). https://doi.org/https://doi.org/10.1016/j.neunet.2018.04.016

16. Fyfe, C.: Hebbian Learning and Negative Feedback Networks, Advanced Information and Knowledge Processing, vol. XVIII. Springer-Verlag London, first edn. (2005)
17. Gama, J., Medas, P., Castillo, G., Rodrigues, P.: Learning with drift detection. In: In SBIA Brazilian Symposium on Artificial Intelligence. pp. 286–295. Springer Verlag (2004)
18. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. ACM Computing Surveys (CSUR) **46**(4), 44 (2014)
19. Gonzalvez, P.M., de Carvalho Santos, S., Barros, R., Vieira, D.: A comparative study on concept drift detectors. Expert Systems with Applications **41**(18), 8144–8156 (2014)
20. Hinder, F., Artelt, A., Hammer, B.: Towards Non-Parametric Drift Detection via Dynamic Adapting Window Independence Drift Detection (DAWIDD). In: Proceedings of the 37th International Conference on Machine Learning. ICML'20, JMLR.org (2020)
21. Hinder, F., Vaquet, V., Brinkrolf, J., Hammer, B.: Model-based explanations of concept drift. Neurocomputing **555**, 126640 (2023). https://doi.org/10.1016/j.neucom.2023.126640
22. Hinder, F., Vaquet, V., Hammer, B.: Suitability of Different Metric Choices for Concept Drift Detection. In: Bouadi, T., Fromont, E., Hüllermeier, E. (eds.) Advances in Intelligent Data Analysis XX. pp. 157–170. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-01333-1_13
23. Igor, G., Webb, G.: Survey of distance measures for quantifying concept drift and shift in numeric data. Knowledge and Information Systems pp. 591–615 (2019). https://doi.org/10.1007/s10115-018-1257-z
24. Killick, R., Fearnhead, P., Eckley, I.A.: Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association **107**(500), 1590–1598 (oct 2012). https://doi.org/10.1080/01621459.2012.737745
25. Kohonen, T.: Essentials of the self-organizing map. Neural Networks **37**, 52–65 (2013). https://doi.org/https://doi.org/10.1016/j.neunet.2012.09.018
26. Kolter, J., Maloof, M.: Dynamic weighted majority: a new ensemble method for tracking concept drift. In: Data Mining, 2003. ICDM 2003. Third IEEE International Conference on. pp. 123 – 130 (nov 2003)
27. Lapinski, A., Krawczyk, B., Ksieniewicz, P., Wozniak, M.: An Empirical Insight Into Concept Drift Detectors Ensemble Strategies. pp. 1–8 (07 2018). https://doi.org/10.1109/CEC.2018.8477962
28. LeCun, Y., Cortes, C.: MNIST handwritten digit database (2010), http://yann.lecun.com/exdb/mnist/
29. Maciel, B.I.F., Santos, S.G.T.C., Barros, R.S.M.: A lightweight concept drift detection ensemble. In: 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI). pp. 1061–1068 (Nov 2015)
30. Myers, A., Munch, E., Khasawneh, F.A.: Persistent homology of complex networks for dynamic state detection. Physical Review E **100**, 022314 (Aug 2019). https://doi.org/10.1103/PhysRevE.100.022314
31. Oner, D., Garin, A., Koziński, M., Hess, K., Fua, P.: Persistent Homology With Improved Locality Information for More Effective Delineation. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(8), 10588–10595 (2023). https://doi.org/10.1109/TPAMI.2023.3246921
32. Rabanser, S., Gunnemann, S., Lipton, Z.C.: Failing loudly: An empirical study of methods for detecting dataset shift. In: 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada (2019)

33. Rieck, B., Togninalli, M., Bock, C., Moor, M., Horn, M., Gumbsch, T., Borgwardt, K.: Neural Persistence: A Complexity Measure for Deep Neural Networks Using Algebraic Topology. In: International Conference on Learning Representations (2019), https://openreview.net/forum?id=ByxkijC5FQ
34. Ross, G.J., Adams, N.M., Tasoulis, D.K., Hand, D.J.: Exponentially weighted moving average charts for detecting concept drift. Pattern Recognition Letters **33**(2), 191 – 198 (2012)
35. Saraswati, A., Nguyen, V.T., Hagenbuchner, M., Tsoi, A.C.: High-resolution Self-Organizing Maps for advanced visualization and dimension reduction. Neural Networks **105**, 166–184 (2018). https://doi.org/https://doi.org/10.1016/j.neunet.2018.04.011
36. Saul, N., Tralie, C.: Persim: A python package for persistence diagrams (2019), https://persim.scikit-tda.org/
37. Schölkopf, B., Smola, A.: Learning with kernels: support vector machines, regularization, optimization, and beyond. Adaptive computation and machine learning, MIT Press (2002)
38. Shang, M., Yuan, Y., Luo, X., Zhou, M.: An α-β-Divergence-Generalized Recommender for Highly Accurate Predictions of Missing User Preferences. IEEE transactions on cybernetics **PP** (2022). https://doi.org/10.1109/TCYB.2020.3026425
39. Sobczyk, A., Luisier, M.: Approximate Euclidean lengths and distances beyond Johnson-Lindenstrauss. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (eds.) Advances in Neural Information Processing Systems. vol. 35, pp. 19357–19369. Curran Associates, Inc. (2022)
40. Sobolewski, P., Wozniak, M.: Concept drift detection and model selection with simulated recurrence and ensembles of statistical detectors. Journal of Universal Computer Science **19**(4), 462–483 (feb 2013)
41. Souza, V.M.A., Reis, D.M., Maletzke, A.G., Batista, G.E.A.P.A.: Challenges in Benchmarking Stream Learning Algorithms with Real-world Data. Data Mining and Knowledge Discovery **34**, 1805–1858 (2020). https://doi.org/10.1007/s10618-020-00698-5
42. Wozniak, M., Zyblewski, P., Ksieniewicz, P.: Active weighted aging ensemble for drifted data stream classification. Information Sciences **630**, 286–304 (2023). https://doi.org/https://doi.org/10.1016/j.ins.2023.02.046
43. Zhang, J., Wei, Z., Yan, Z., Zhou, M., Pani, A.: Online Change-Point Detection in Sparse Time Series With Application to Online Advertising. IEEE Transactions on Systems, Man, and Cybernetics: Systems **49**(6), 1141–1151 (2019). https://doi.org/10.1109/TSMC.2017.2738151