

The Role of ICT in our Daily Life Applications: Obstacles and Challenges

Streaming Data Classification with Concept Drift

Mashail Althabiti and Manal Abdullah*

Faculty of Computing and Information Technology, King Abdul Aziz University, Jeddah, Saudi Arabia

ABSTRACT

The concept of the data stream has emerged as a result of the evolution of technologies in various fields for instance: banking, electronic commerce, social media, and many others. It represents the sequence of data examples that are generated at a very high speed which can be hard to be stored in memory. Thus, it became hard to extract valuable information from the continuous data stream using traditional data mining. Data Stream Mining DSM algorithms should fulfil some requirements such as limited memory, concept drift detection, and one scan processing. Concept Drift must be tracked to avoid poor performance and inaccurate results of predictive models. It refers to the changing in the data stream distributions due to several reasons including the changes in the environment, individual preferences, or adversary activities. In this paper, we will analyze the classification algorithms handling concept drift for DSM. Also, popular concept drift datasets, data stream tools, and evaluation measures will be presented.

KEY WORDS: CONCEPT DRIFT, DATA MINING, DATA STREAM MINING, CLASSIFICATION

INTRODUCTION

Nowadays, millions of people around the world share data anywhere and anytime. The emerging technologies in telecommunications, entertainments industry, social media sites, banking services, and other applications have led to the massive growth of generated data stream. The data stream can be referred to the sequence of data examples that produced at a very high rate and arrive continuously at a potentially infinite stream. According

to the “10 Key Marketing Trends For 2017” report,” 90% of the data in the world has been produced in the last two years only, 2.5 quintillion bytes each day (Web-1). As a result, the massive amount of data cannot be stored for farther processing and mining. So, the data stream mining concept has emerged to extract the knowledge from the data stream and provide real-time processing. However, it has some constraints that must cope with such as and concept drift, limited memory, and one scan of the data. Concept Drift is referring to the changes

ARTICLE INFORMATION:

*Corresponding Author: maaabdullah@kau.edu.sa

Received 5th Nov, 2018

Accepted after revision 30th Dec, 2018

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007 CODEN: USA BBRCBA

 Thomson Reuters ISI ESC / Clarivate Analytics USA and
Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 SJIF 2017: 4.196

© A Society of Science and Nature Publication, Bhopal India
2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

DOI: 10.21786/bbrc/12.1/20

in data concepts over time. This may result in wrong predictions and inaccurate results. In non-stationary environments, learning a model from unstable data can result in inaccurate results and predictions. The underlying data distribution may change, so the model will not be consistent with the new data anymore. For example, predicting a customer's behavior toward shopping, where her/his preferences have been changed. Thus, this will produce wrong results based on old data. So, concept drift must be handled and tracked using detecting methods that can cope with the data stream.

Data stream mining with concept drift handling were highly studied last decade. There are many surveys studies have literature the data stream mining from different perspectives. The authors in (Gama et al., 2014) have surveyed the state-of-the-art adaptive learning algorithms with concept drift detection. They have addressed the concept drift through various applications and highlighted several evaluation techniques. In (Khamassi et al., 2016) the authors have presented general criteria to help researchers in designing their concept drift handling methods. They have categorized the existing concept drift algorithms according to these criteria. Also, 14 drift detectors have been evaluated in (Barros and Santos, 2018) using six artificial datasets and compared in term of accuracy and detection. The authors have used Naive Bayes (NB) and Hoeffding Tree classifiers to test the drift detectors.

In this study, we will present the data stream mining components. We will address the problem of concept drift in classification algorithms and highlight the exciting state of art handling methods. Besides, the most used datasets, tools, and evaluation methods will be presented. This rest of this research is organized as follows. In Section 2 we will discuss the concept of data stream mining. In Section 3, DSM components will be presented. The most used dataset and evaluation methods in data stream mining studies will be presented in Section 4 and Section 5 respectively. Section 6, will present DSM tools and conclude the study in Section 7.

DATA STREAM MINING

Learning a model is considered as an essential step in data mining and machine learning (Mittal and Kashyap, 2016). Previously, it was done in static environments where the whole datasets are available, stored and can be accessed many times. On the contrary, learning from massive datasets in non-stationary environments which has become a challenging area. The huge generation of continuous data in everyday applications has emerged the concept of data stream.

The data stream is a sequence of potentially non-stop data instances that can be read and processed only

once. As technologies evolving, traditional data mining has become hard to deal with the stream of data on the Internet of Things IoT, web searches, banking transactions and many more (Gaber, 2011). Thus, the data stream mining has become an attractive research area. Data stream mining refers to the process of finding knowledge and valuable pattern in continuous, potentially infinite, and high-volume data streams. It plays an essential role in predictive modeling and decision-making. Data stream mining has several challenges that must be overcome including the following:

- Resource constraints: the data stream is potentially infinite, huge and comes in high speed, so it is hard to be stored in a memory. Also, the processing time must be as shorter as possible (Kantardzic, 2011).
- One scan: data stream cannot be accessed randomly or many times (Kantardzic, 2011).
- Data preprocessing: Since data is continuously arriving, it is not feasible to use manual data preprocessing methods. It should be fully automated and automatically updated as data evolving (Krempel et al., 2014).
- Privacy and confidentiality: the data stream is infinite and comes in portions, so the information will not be incomplete. In this case, it is hard to judge the privacy of a model that has a data stream as input (Krempel et al., 2014).
- Concept drift: data instance may change over time.

DATA STREAM MINING COMPONENTS

The massive data stream can be generated from social media applications such as Twitter, Facebook, Pinterest (Mohanty et al., 2015). Also, the health, economic, financial industry, and many others. The process of the data stream mining involves several components as shown in Figure 1 including the streaming data as input, estimator, data mining algorithm, drift detection and the extracted knowledge as output.

Input

The data stream can be generated from different sources such as web searches, social media posts, real-time surveillance systems, banking activities, and other non-stationary environments.

Estimators

Estimation is a critical step that prepares the data stream to the knowledge extraction process. Data stream examples must be processed in real time due to its high speed. Also, it cannot all be stored in a memory. So, Estimation methods are needed to select a subset of the arriving

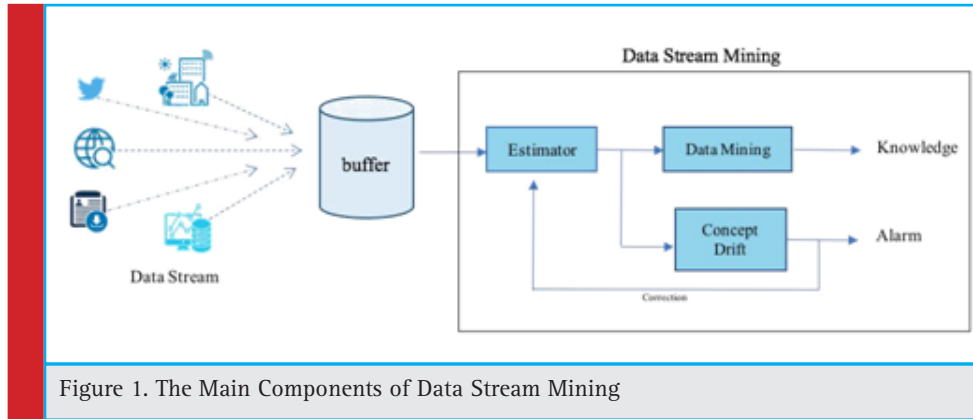


Figure 1. The Main Components of Data Stream Mining

data stream. It can be categorized into data-based and task-based techniques as shown in Table 1.

Data Stream Classification and Concept Drift

Data Stream Classification

Classification is a supervised machine learning algorithm. It uses the past data (training set) to build a model and then (2) use it to predict class labels (testing set) (Han et al., 2012). Classification algorithms in non-stationary environments must fulfill the data stream requirements regarding the processing time, limited memory and one-time scan. In the dynamic environment, some instances in streaming data may change over time because of

the high velocity and limited memory, and this called concept drift. It is the changes in data distribution of the output given the input, while the distribution of the input may stay unchanged (Gama et al., 2014). For example, predicting a customer’s behavior toward shopping, where her/his preferences have been changed. Thus, this will produce wrong results based on old data.

Concept Drift

Concept drift occurs between two points of time t_0 and t_1 when the joint distribution of x (independent variable) and y (target variable) at time t_0 is not equal to the joint

Table 1. Estimation Method (Žliobaite et al., 2016).

| Estimation Methods | Method type | Description | Key points |
|------------------------------|-------------------|---|---|
| Sampling | Data-based method | An old statistical method that concerned with the probability processing a data item or not | The main obstacle with sampling is that it is not possible to determine the size of the dataset |
| Load shedding | Data-based method | The process of dropping a series of examples in the streaming data | It is not easy to be implemented in the data stream algorithms because the dropped data might be used in structuring the models |
| Sketching | Data-based method | The method of sampling the stream vertically | The main disadvantage of sketching is that of accuracy |
| Synopsis data structures | Data-based method | The method of converting the summary of the streaming data into data structures and use it for analysis purposes | The main synopsis data structures are frequency moments, wavelet analysis, quantiles, and histograms |
| Aggregation | Data-based method | The process of summarizing the data stream using computing and statistical measures | It is not recommended to be used with distributions that have high fluctuating data |
| Approximation | Task-based method | The process of designing stream mining algorithms for computationally hard problems | It considered a good solution for data stream mining problems |
| Sliding window | Task-based method | The process of analyzing recent data streams and summarizing the old versions. | It emphasizes recent data and easily understood |
| Algorithm output granularity | Task-based method | The process of mining the data stream, adapting resources and merging the extracted knowledge when memory is full | It performs well with very high fluctuating data distribution |

distribution of (x,y) in t_1 (Gama et al., 2014). It can be represented as:

$$p_{t_0}(X, y) \neq p_{t_1}(X, y). \quad (1)$$

Concept drifts may occur if there is a change in: (1) the prior probabilities of classes $p(y)$, (2) the class-conditional probability distributions $p(X,y)$, or (3) the posterior probabilities $p(y|X)$.

Concept drift in the data stream may happen due to different reasons such as the changes in environment, individual preferences, or adversary activities (Z'liobaite et al., 2016). It may happen in different forms as shown in Figure 3: (1) sudden drift: when a concept changes to another abruptly, (2) incremental drift: when there are many intermediate concepts in between, (3) gradual drift: when the concepts emerge in an un-sudden way over time, and (4) reoccurring drift: when old concepts reappear after some time.

Data stream classification algorithms can be categorized into single classifiers and ensemble algorithms. Regarding concept drift adaptation, some algorithms update their classifiers continuously in the occurrence of drift or not. Others algorithms trigger changes in the classifier whenever a drift is detected (Gama et al., 2014).

Data Stream Classification Algorithms with concept drift

Data stream classification algorithms with concept drift can be classified into single classifier algorithms and ensemble algorithms.

Single classifier

Some single classifier algorithms observe and detect the drift in the data distribution by using statistical methods and keep track of the base classifier performance (Mittal and Kashyap, 2016). In case of discovering drift, it will alarm the base classifier to update it or rebuild it such as:

- The authors in (Gama et al., 2004) have designed the Drift Detection Method (DDM) that monitors the classifier error-rate. If the error rate reaches the warning and drift level, then we can observe that a data distribution has been changed.
- DDM detector has been modified into an improved version named Early Drift Detection Method (EDDM) in (Baena-Garcia et al., 2006). EDDM used to detect gradual drift that emerge slowly by considering the distances between the classification errors.
- Reactive Drift Detection Method (RDDM) is another modified version of DDM (Barros et al., 2017). The proposed Algorithm has overcome the problem of performance loss of DDM by discarding the older examples. It periodically recalculates the DMM calculations that determine the alarm and drift levels. Also, the drift occurs whenever the number of examples in the alarm level reached the threshold.

Other algorithms detect concept drift using windowing techniques by comparing the distributions the windows such as:

- ADWIN (Bifet and Gavald, 2007) detects the different types of changes using sliding windows with the most recent examples. Concept drift can be observed if the means between two sub-windows is greater than the threshold. As the window grows, the processing time becomes longer. Thus, authors have proposed a developed version called ADWEN2 to satisfy the memory and time requirements.
- Concept adapting very fast decision tree is another algorithms that use single classifier (Hulten et al., 2001). It extended the VFDT algorithm with the ability to detect the concept drift. Also, employs sliding window to keep the classifier updated with the recent instances.
- Authors in (Du et al., 2014) have proposed a window-based algorithm called ADDM where the size of the window is dynamically determined. It detects the concept drift by keeping track the entropy of the window. It reports a concept drift when the entropy is equal one. ADDM has been evaluated using seven datasets containing different types of concept drift. It showed good performance in detecting drift comparing to other methods. Also, it obtained high accuracy.
- The authors in (Liu et al., 2017) have proposed a fuzzy windowing method to adapt concept drift, named FW-DA. The proposed algorithm reports a drift when there is a significant difference between the test statistics of the current window and the old window.
- The authors in (Nishida and Yamauchi, 2007) have proposed STEP algorithm that considers the accuracy of two windows recent and old. Drift is discovered if there is a significant difference between the two windows which calculated through a statistical test.
- The authors in (Pesaranghader and Viktor, 2016) have proposed Fast Hoeffding Drift Detection Method (FHDDM). The proposed method monitors the probabilities of correct predictions over the sliding window. It compares the maximum and the most recent probabilities and observes the change if the differences between these probabilities equal or exceed the threshold.

Ensemble Classifiers

In this approach, the algorithms use a set of classifiers where each classifier is assigned a weight and adapting to the changes by updating its components and its associated weights (Mittal and Kashyap, 2016) such as:

- DWM (Kolter and Maloof, 2003) maintains a set of experts, each of them assigned to a weight. When an instance arrives, it passed to an expert and then returned with a local prediction. DWM determine the global prediction using the local predictions and expert weights.
- The AUE2 algorithm (Brzezinski and Stefanowski, 2014) is another method that partition the data stream into chunks, and each chunk contains a set of examples. For every arriving chunk, a classifier associated with a weight will be created.
- Each classifier performance is evaluated by calculating the error rate on data chunk to determine the worst performing classifiers.
- In addition, two classifiers can be ensembled to form a detection system to detect both sudden and gradual drift (Jadhav and Deshpande, 2017). It composed of two classifiers online classifier and block-based classifier. Whenever data instance arrives, the online classifier updates itself, so any occurrence of sudden changes can be detected. While, block-based classifier work on blocks of data instances, which can observe the gradual changes. The classifiers error rate will be calculated to detect the changes. The drift can be observed if the value of the error rate is the same for the next blocks of the data stream.
- Double-Window-based Classification Algorithm DWCDs is another window-based method used to detect changes in data stream (Zhu et al., 2010). It detects the concept drift by checking the data distributions periodically. The proposed algorithm starts with generating decision trees using the data in the sliding window. If a concept drift is observed, then the model of DWCDs will be updated.
- Moreover, the authors in (Bach and Maloof, 2008) have proposed paired learner (PL) algorithm that ensemble two classifier; stable and reactive. The stable classifier used to predict based on its overall experience, while the reactive classifier predicts based on the recent window. PL observes the distributional changes by comparing the performance of these two classifiers.

All the mentioned algorithms are summarized in Table 2.

Output

This component represents the knowledge and valuable pattern extracted from the data stream.

DATASETS

Several well-known datasets have been used to evaluate the effectiveness of the classification algorithms in

detecting Concept Drift. Datasets can be real or artificial where contains one type of drift or various types. Table 3 shows the most used dataset in data stream mining studies with the presence of Concept Drift.

EVALUATION MEASURES

The following list presents the well-known evaluation measure for classification data stream algorithms:

- Accuracy score: It is calculated by dividing the number of correct predictions by the total classifier's predictions (Han et al., 2012).
- Recall: It refers to as the true positive rate (Han et al., 2012).
- CPU Time: It measures the total runtime of the CPU in training and testing the classifier (Dhaliwal and Bhatia, 2017).
- Memory: It measures the total memory consumed to run the classifier and store the running statistics (Dhaliwal and Bhatia, 2017).
- Kappa Statistic: It measures the homogeneity among the classifiers (Dhaliwal and Bhatia, 2017).

The concept drift detection can be assessed through different measures such as:

- The probability of true change detection: It measures the algorithms ability to discover drifts when they occur (Gama et al., 2014).
- Delay time of detection: It measures the time would be passed before the change is detected (Gama et al., 2014).

DATA STREAM MINING TOOLS

The most popular tools used in data stream environment are listed below:

- Weka: it provides a set of data mining and machine learning algorithms. These algorithms can be implemented directly on datasets through Weka GUI or Import Weka Java library (Web-1).
- MOA: is a project developed in University of Waikato, New Zealand. It provides an environment to deal with data stream, run experiments, and implement data stream mining algorithms (Bifet et al., 2010).
- SAMOA: Scalable Advanced Massive Online Analysis is an open source tool provides the well-known data stream and machine learning algorithms (Morales and Bifet, 2015).
- Apache Storm: an open source platform for processing infinite streams of data. It is scalable and fast which make it suitable to produce immediate analytics, and perform online machine learning (Web-6).

Table 2. Classification Algorithms with Concept Drift.

| | Algorithm | Classifier | Type | Estimator | Dataset | Key points |
|----|--|--|-------------------------------------|--|--|---|
| 1 | FHDDM (Pesaranghader and Viktor, 2016) | Naive Bayes (NV) and Hoeffding Tree | Single classifier | Slide window | Sine1, Mixed, Circles, Airlines, Poker Hand, and Electricity | The detection delay is shorter than other detectors |
| 2 | EDDM (Baena-Garcia et al., 2006) | Decision tree and two nearest-neighbourhood learning algorithm | Single classifier | Window | 4 Artificial Datasets (SINE1, CIRCLES, GAUSS, MIXED and SINE1G) and 1 real dataset | It can detect slow gradual drift and deal with noisy datasets |
| 3 | DDM (Gama et al., 2004) | Neural network, decision tree and perceptron | Single classifier | Window | 8 Artificial Datasets and 1 real dataset | It detects sudden changes and gradual changes (changes that do not emerge very slowly) |
| 4 | DWM (Kolter and Maloof, 2003) | Incremental Tree Inducer (ITI) and NV | Ensemble classifiers | Not available | Stagger and Sea | According to the performance changes, the classifiers are added |
| 5 | PL (Bach and Maloof, 2008) | Naive Bayes | Ensemble classifiers | Slide window | Malware detection dataset, meeting scheduling dataset, electricity prediction dataset, and two synthesis dataset including Stagger and Sea | The number of the trained learners used are less than other ensemble methods |
| 6 | DWCDS (Zhu et al., 2010) | Random Decision Trees | Ensemble classifiers | Slide window | SEA, HyperPlane, KDDCup99, Yahoo shopping data and LED | It detects drift better than single window-based algorithms. |
| 7 | AUE2 (Brzezinski and Stefanowski, 2014). | Hoeffding Tree | Data stream is Ensemble classifiers | partitioned into chunks | Synthetic datasets generated by the MOA tool and 4-real datasets (Elec, Poker, Airline, and COV) | It consumes less memory comparing to other ensemble approaches. It detects different types of drift including sudden, gradual, recurring. |
| 8 | ADWIN (Bifet and Gavald, 2007). | Naïve Bayes | Single classifier | Slide window | Electricity Market dataset and synthetic dataset | ADWIN works only for one-dimensional data |
| 9 | Ensemble classification system (Jadhav and Deshpande, 2017). | J48, Naive Bayes, and Random Forest | Ensemble classifiers | Data stream is partitioned into fixed blocks | Census income and Spam email datasets | It detects both sudden and gradual drift. It handles missing values |
| 10 | CVFDT (Hulten et al., 2001) | Hoeffding Tree | Single classifier | Slide window | hyperplane and web data | CVFDT can keep its model up-to date with streaming data that contains concept drift |
| 11 | ADDM (Du et al., 2014) | IB1, j48, NNge, and SVM | Single classifier | Slide window | Five artificial datasets (Gauss, Mixed, Stagger, Sine1, and Sine1g) (Elist and Elec2). | ADDM used Hoeffding bound to determine the sliding window. In the evaluation, ADDM has lost upon one out of seven datasets |
| 12 | STEPD (Nishida and Yamauchi, 2007). | IB1 classifier and Naive Bayes | Single classifier | Slide window | Five artificial datasets (STAGGER, GAUSS, MIXED2, CIRCLES, and HYPERP) | It can discover sudden and gradual drift with the presence of noise |
| 13 | FW-DA (Liu et al., 2017). | DDM and ECDD | Single classifier | Slide window | SEA, Elec, Airline, and Spam | FW_DA performed well in detecting and adapting concept drift |
| 14 | RDDM (Barros et al., 2017) | Naive Bayes | Single classifier | Window | Agrawal, Mixed, Sin, Led, Airlines, pokerhand, and Electricity | RDDM has the higher accuracy among the others methods in detecting |

| | Dataset Name | Type | #instants | #Attribute | Key points |
|----|--|------------|-----------|------------|--|
| 1 | Census Income (web-2) | Real | 48842 | 14 | <ul style="list-style-type: none"> It has different independent variables including name, marital status, education, occupation, and many others. These variables are used to predict the person's income which represents the dependent variable. it includes only gradual drifts. |
| 2 | Sine1 (Gama et al., 2004) | Artificial | - | 2 | <ul style="list-style-type: none"> It involves sudden concept drift it has two features x and y, each feature distributed in zero and one. It uses $[y = \sin(x)]$ to draw the curve. The data is labeled to positive and negative according to its position to the curve. |
| 3 | Elec 2 (Gama et al., 2004) | Real | 45,312 | 8 | <ul style="list-style-type: none"> It consists of five features used to predict target variable whether it's up or down. It is collected from the Australian Electricity Market of Australia at 30-minute intervals from 1996 to 1998. |
| 4 | Airline (Web-3) | Real | 539,3837 | | <ul style="list-style-type: none"> It includes the flight information such as the departure date, time of flight, destination, distance, and many others. It used to predict if a given flight will delay. |
| 5 | Cover type (Web-4) | Real | 581, 012 | 54 | <ul style="list-style-type: none"> It includes information of four wilderness areas in Colorado including Elevation, Aspect, Slope, Soil Type, and others variables. it used to predict the forest cover type. |
| 6 | Stagger Generator (Du et al., 2014) | Artificial | 240 | 3 | <ul style="list-style-type: none"> The three features are shape, size, and color. it contains sudden drift. |
| 7 | Sine 2 (Gama et al., 2004) | Artificial | - | 2 | <ul style="list-style-type: none"> It involves two features x and y, each feature distributed in zero and one. It uses $[y < 0.5 + 0.3 \sin(3\pi x)]$ to draw the curve. The data is labeled to positive and negative according to its position to the curve. |
| 8 | SEA Generator (Street and Kim, 2001) | Artificial | - | 3 | <ul style="list-style-type: none"> It is used to generate data stream containing sudden drift. |
| 9 | LED Generator (Brzezinski and Stefanowski, 2014) | Artificial | - | 24 | <ul style="list-style-type: none"> The 24 attributes used to predict the seven-digit showed over the LED display. |
| 10 | Mixed (Gama et al., 2004) | Artificial | - | 4 | <ul style="list-style-type: none"> It contains sudden concept drift. The example is labeled as positive if two of the three conditions are fulfilled: $v, w, y < 0.5 - 0.3 \sin(3\pi x)$. If not, it will be negative. |

CONCLUSION

Concept Drift is one of the main challenges of DSM. It must be detected and handled to avoid inaccurate results of learning models. In this research, we have discussed the concept of the data stream. The DSM components including the input/output, estimation methods, and classification algorithms with concept drift have been presented. Also, we have highlighted the most used DSM tools, datasets, and evaluations measures in data stream experiments.

REFERENCES

- [1] Bach, S. H. & Maloof, M. A. Paired Learners for Concept Drift. Proceedings of the 8th IEEE International Conference on Data Mining 2008 Pisa, Italy.
- [2] Baena-Garcia, M., Campo-Ávila, J. D., Fidalgo, R., Bifet, A., Gavalda, R. & Morales-Bueno, R. 2006. Early Drift Detection Method. Fourth international workshop on knowledge discovery from data streams, 6, 77-86.
- [3] Barros, R. S. M., Cabral, D. R. L., Jr, P. M. G. A. & Santos, S. G. T. C. 2017. RDDM: Reactive drift detection method. Expert Systems with Applications, 90, 344-355.
- [4] Barros, R. S. M. & Santos, S. G. T. C. 2018. A large-scale comparison of concept drift detectors. Information Sciences, 451-452, 348-370.
- [5] Bifet, A. & Gavald, R. Learning from Time-Changing Data with Adaptive Windowing. Proceedings of the 2007 SIAM International Conference on Data Mining, 2007. 443-448.
- [6] Bifet, A., Holmes, G., Kirkby, R. & Pfahringer, B. 2010. MOA: Massive Online Analysis. Journal of Machine Learning Research, 11, 1601-1604.

- [7] Brzezinski, D. & Stefanowski, J. 2014. Reacting to Different Types of Concept Drift: The Accuracy Updated Ensemble Algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, 25, 81-94.
- [8] Dhaliwal, P. & Bhatia, M. P. S. 2017. Effective Handling of Recurring Concept Drifts in Data Streams. *Indian Journal of Science and Technology*, 10.
- [9] Du, L., Song, Q. & Jia, X. 2014. Detecting concept drift: An information entropy based change-detector using an adaptive sliding window. *Intelligent Data Analysis*, 18, 337-364.
- [10] Gaber, M. M. 2011. *Advances in data stream mining*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2, 79-85.
- [11] Gama, J., Žliobaite, I., Bifet, A., Pechenizkiy, M. & Bouchachia, A. 2014. A survey on concept drift adaptation. *ACM Computing Surveys*, 46, 1-37.
- [12] Gama, J. O., Medas, P., Castillo, G. & Rodrigues, P. Learning with Drift Detection. 17th Brazilian Symposium on Artificial Intelligence, 2004 Brazil. 286-295.
- [13] Han, J., Kamber, M. & Pei, J. 2012. *Data Mining Concepts and Techniques*, Amsterdam, Elsevier/Morgan Kaufmann.
- [14] Hulten, G., Spencer, L. & Domingos, P. Mining time-changing data streams. the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 2001 San Francisco, California. 97-106.
- [15] Jadhav, A. & Deshpande, L. 2017. An Efficient Approach to Detect Concept Drifts in Data Streams. 2017 IEEE 7th International Advance Computing Conference (IACC). Hyderabad, India.
- [16] Kantardzic, M. 2011. *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley & Sons, Inc.
- [17] Khamassi, I., Mouchaweh, M. S. & Hammami, M. 2016. Discussion and review on evolving data streams and concept drift adapting. *Evolving Systems*, 9, 1-23.
- [18] Kolter, J. Z. & Maloof, M. A. 2003. Dynamic weighted majority: a new ensemble method for tracking concept drift. Third IEEE International Conference on Data Mining. Melbourne, FL, USA, USA: IEEE.
- [19] Kreml, G., Žliobaite, N., Brzezinski, D., Hüllermeier, E., Last, M., Lemaire, V., Noack, T., Shaker, A., Sievi, S., Spiliopoulou, M. & Stefanowski, J. 2014. open challenges for data stream mining research. *SIGKDD Explorations*, 16, 1-10.
- [20] Liu, A., Zhang, G. & LU, J. 2017. Fuzzy time windowing for gradual concept drift adaptation. 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). Naples, Italy: IEEE.
- [21] Mittal, V. & Kashyap, I. 2016. Empirical Study of Impact of Various Concept Drifts in Data Stream Mining Methods. *International Journal of Intelligent Systems and Applications*, 8, 65-72.
- [22] Mohanty, H., Bhuyan, P. & Chenthati, D. 2015. *Big Data A Primer*, Springer.
- [23] Morales, G. D. F. & Bifet, A. 2015. SAMOA: Scalable Advanced Massive Online Analysis. *Journal of Machine Learning Research* 16, 149-153.
- [24] Nishida, K. & Yamauchi, K. Detecting Concept Drift Using Statistical Testing. *International Conference on Discovery Science*, 2007. Springer, Berlin, Heidelberg, 264-269.
- [25] Pesaranghader, A. & Viktor, H. L. Fast Hoeffding Drift Detection Method for Evolving Data Streams. *ECML PKDD 2016 Machine Learning and Knowledge Discovery in Databases 2016 Italy*. Springer, Cham, 96-111.
- [26] Street, W. N. & Kim, Y. A Streaming Ensemble Algorithm (SEA) for Large-Scale Classification. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001. 377-382.
- [27] Žliobaite, I., Pechenizkiy, M. & Gama, J. A. An overview of concept drift applications. *Big Data Analysis: New Algorithms for a New Society*, 2016. Springer, Cham.
- [28] Zhu, Q., Hu, X., Zhang, Y., Li, P. & Wu, X. 2010. A Double-Window-Based Classification Algorithm for Concept Drifting Data Streams. 2010 IEEE International Conference on Granular Computing. San Jose, CA, USA: IEEE.

Web sites

- Web-1: Ten Key Marketing Trends for 2017 and Ideas for Exceeding Customer Expectations. [Online]. Available: http://comsense.consulting/wpcontent/uploads/2017/03/10_Key_Marketing_Trends_for_2017_and_Ideas_for_Exceeding_Customer_Expectations.pdf [Accessed 23-12 2018].
- Web-2: UCI Machine Learning Repository: Census Income Data Set [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/census+income> [Accessed 21-12 2018].
- Web-3: Elena Ikonomovska's Web page [Online]. Available: https://kt.ijs.si/elena_ikonomovska/data.html [Accessed 3/11/2018].
- Web-4: UCI Machine Learning Repository: Covertype Data Set [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/covertype> [Accessed 28/11 2018].
- Web-5: Weka 3 - Data Mining with Open Source Machine Learning Software in Java [Online]. Available: <https://www.cs.waikato.ac.nz/ml/weka/> [Accessed 3/11/2018].
- Web-6: Apache Storm [Online]. Available: <http://storm.apache.org/index.html> [Accessed 3/11/2018].