A Fuzzy Drift Correlation Matrix for Multiple Data Stream Regression

1st Yiliao Song
Decision System&e-Service Intelligence (DeSI) Lab
Centre for Artificial Intelligence
Faculty of Engineering and Information Technology
University of Technology Sydney Sydney, Australia
Yiliao.Song@student.uts.edu.au

2nd Guangquan Zhang
Decision System&e-Service Intelligence (DeSI) Lab
Centre for Artificial Intelligence
Faculty of Engineering and Information Technology
University of Technology Sydney Sydney, Australia
Guangquan.Zhang@uts.edu.au

3rd Haiyan Lu
Decision System&e-Service Intelligence (DeSI) Lab
Centre for Artificial Intelligence
Faculty of Engineering and Information Technology
University of Technology Sydney Sydney, Australia
Haiyan.Lu@uts.edu.au

4th Jie Lu
Decision System&e-Service Intelligence (DeSI) Lab
Centre for Artificial Intelligence
Faculty of Engineering and Information Technology
University of Technology Sydney Sydney, Australia
Jie.Lu@uts.edu.au

Abstract—How to handle concept drift problem is a big challenge for algorithms designed for the data streams. Currently, techniques related to the concept drift problem focus on single data stream. However, it normally needs to handle multiple relevant data streams in the real-world application. Current concept drift methods can not be directly used in the multi-stream setting. They can only be limitedly applied on each stream separately, which omits the drift correlation between streams. In the multi-stream scenario, when drift occurs in a stream, other streams may face or have faced a similar drift problem as well. This pattern of simultaneous or delayed occurrence of drift is critical to analyze and predict multiple streams as a whole dynamic system. To fill the gap in the multi-stream scenario, this paper proposes a fuzzy drift variance (FDV) to measure the correlated drift patterns among streams. FDA is able to present how the pattern of drift occurrence for any two streams correlates and how delayed this correlation is. Seven synthetic streams are designed to validate FDA. The experimental results show a good presentation ability of FDA for drift-correlated multiple streams.

Index Terms—fuzzy membership, concept drift, data stream, multiple streams

I. INTRODUCTION

Concept drift refers to the phenomenon that the data distribution of a data stream may change over time [1]. Unlike static batch data, data streams are the sequences of continuous arriving data instances consisting of multiple infinite and fast evolving data series [2]. In recent years, data stream mining has been extensively studied in growing fields of multidisciplinary research including data bases [3], artificial intelligence, machine learning, automated scientific discovery, statistics, decision making and so on [4].

Once drift occurs in a data stream, the off-line learning performance will be largely impaired [5]. The predictor trained with the previous data can on longer be applied to predict labels for newly arrived data that follows a different distribution [6].

A number of methods have been proposed to specially solve the concept drift problem [7]. Current research on the concept drift problem focuses on two aspects. One is how to detect the occurrence of drift, namely drift detection techniques [8]–[10]. The other is how the trained predictor swiftly adapts to newly arrived data, namely drift adaptation techniques [11], [12].

The limitation in current research on concept drift is that they are all designed for single data stream or independent data streams. However, the real-world applications are mostly of multiple relevant data streams [13]. For example, in the finance industry, the bid/offer rate in the inter-bank lending market always involves the behaviors of more than two banks.

Methods for handling concept drift problem in single data stream can not be directly used in the multi-stream scenario [14]. They can be only limitedly applied on each stream separately [15]. In the multi-stream scenario, when drift occurs in a stream, other streams may face or have faced to a similar drift problem as well. Drift can occur in two streams simultaneously or with delay. This pattern of simultaneous or delayed occurrence of drift is critical to analyze and predict multiple streams.

To fill the gap of research on concept drift in the multi-stream scenario, this paper proposes a fuzzy drift variance matrix (FDA) to measure the correlated drift patterns among streams. FDA is a three dimensional matrix $M(i, j, l)$, the first two dimensions are for streams and the third dimension is to indicate whether the drift correlation is simultaneous or delayed. Each element in FDA matrix can be 0 or 1. The value of 1 indicates that ith stream and jth stream has l-step delayed drift correlation. To compute FDA, we propose a measurement called fuzzy error deviation (fed) of single stream as well as

between two streams. fed is computed based on the variation of the predictive error of data streams. A normalization process and difference transformation are conducted to eliminate the magnitude difference between streams. A sigmoid fuzzy membership function is applied to limit the intermediate results strictly in $[0, 1]$ and distinctly differ instances before and after drift occurs. Given a specific delayed step l, if two streams had small fed between streams but large fed for each stream, these two stream is considered to have high drift correlation with l-step delay, and the corresponding value in FDA is 1.

The remainder of this paper will cover the following: Section II reviews related works. Section III presents the proposed FDA matrix. Section IV reports the empirical studies of our method on the synthetic data streams. The conclusions and future works are drawn in Section V.

II. RELATED WORK

Concept drift is identified if the underlying distribution changes, i.e., $p_{t+1}(\boldsymbol{X}, y) \neq p_t(\boldsymbol{X}, y)$ [16]. Concept drift may occur in four types: *sudden* drift, *incremental* drift, *gradual* drift and *reoccurring* concept [17]. In this paper, we only consider the sudden drift and incremental drift to determined the drift correlation. This is because gradual drift and reoccurring concept need a long period to identify but the drift correlation discussed in this paper only focuses on the first training set. In addition, sudden and incremental drift are supposed to be the research foundation for other types of drift which are considered as a combination of sudden drift [18].

To build the drift correlation between stream, it needs to detect drift in each stream. Current drift detection methods have been categorised into error-based drift detection [19] or statistic-based detection [20]. This paper uses error-based measurement to present the drift severity and build the correlation based on this measurement.

Existing research on multiple streams mainly focuses on two streams and assume these two streams are independent, such as the adaptive framework for multiple stream [21], and the framework for regression task [14]. In these studies, the task of multiple streams is limited to two independent streams, and one stream has to be a source stream and the other one a target stream. Under this condition, domain adaptation techniques are applied to transfer knowledge between two streams. It needs a strong prior knowledge of two streams so that which one is a source stream or a target stream can be known. Beside, the correlation between streams is not limited to the source-and-target relationship.

III. THE FUZZY DRIFT VARIANCE MATRIX (FDA)

The FDA is a three dimensional matrix, noted by $M(i, j, l) = v_{ij}^l$ where $v_{ij}^l \in \{0, 1\}$. For $i \neq j$, v_{ij}^l measures the l-step delayed drift correlation from the ith to the jth stream. $v_{ij}^l = 1$ denotes a strong drift correlation between the ith and jth data streams. For example, drift always occurs simultaneously in ith and jth streams; $v_{ij}^1 = 1$ means if drift occurs in the ith stream at time t, drift will occur in the jth stream at time $t + 1$; $v_{ij}^{-2} = 1$ means if drift occurs in the

jth stream at time t, drift will occur in the ith stream at time $t + 2$. Clearly, $v_{ij}^l = v_{ji}^{-l}$. Therefore, we can let $l \in \mathcal{N}$ instead of $l \in \mathcal{Z}$ without any loss of drift correlation between two streams. When $i = j$, v_{ij}^l actually focuses on single stream. Considering two streams which are exactly the same, the pattern of their drift occurrence should be the same as well. Therefore, the simultaneous drift correlation should be 1 and any delayed drift correlation should be 0. Accordingly, in $M(i, j, l)$, the value of v_{ii}^l is manually set by 1 when $l = 0$ and set by 0 when $l \neq 0$.

We presented the process of how to compute $M(i, j, l)$ in Fig. 1. In the next subsections, we will explain each part step by step.

A. Preliminaries and notations

Notations for defining the fuzzy drift variance are listed as below.

- $\boldsymbol{S}_i$: the ith data streams;
- $s(i, t) = (\boldsymbol{X}_{i,t}, y_{i,t})$: the observation of ith data stream at time t where $\boldsymbol{X}_{i,t}$ is the attributes and $y_{i,t}$ is the label;
- $\hat{h}_{i,t}(\cdot)$: the estimated predictor function for ith data stream at time t;
- $\ell(\cdot)$: the loss function

B. Fuzzy error deviation for single stream

One difficulty to compute $M(i, j, l)$ is that whether drift occurs and how severe of this drift in a data stream can not be directly observed. The drift severity is a critical factor to compute $M(i, j, l)$. Therefore, it needs to firstly design a measurement to estimate how severe of drift in a data stream $\boldsymbol{S}_i$. In this paper, we propose a measurement called fuzzy error deviation (*fed*) to estimate the drift severity by the predictor error.

To obtain *fed*, the first step is to compute predictive error. Given $\hat{h}_{i,t}$ the estimated predictor for ith data stream, the predictive error $\hat{e}_{i,t}$ is computed in (1).

$$\hat{e}_{i,t} = \ell(\hat{h}_{i,t}(\boldsymbol{X}_{i,t}), y_{i,t}). \tag{1}$$

A popular method to determine whether drift occurs in the single stream scenario is to obtain the value of $\hat{e}_{i,t}$, if $\hat{e}_{i,t}$ starts to increase and beyond a threshold, drift is detected. This paper develops the idea of error-based drift detection methods and applies this idea to compute *fed*.

The threshold in current error-based drift detection methods for single stream is manually set before using it. It is not preferred to manually set the threshold for each stream in a multi-stream scenario, although this is the simplest way to apply the error-based detection idea. A main reason is that the value of $\hat{e}_{i,t}$ largely depends on the magnitude of $y_{i,t}$. The thresholds for multiple streams may differ a lot because of the magnitude discrepancy. Thus, it needs complicated tuning process to balance the relationship among these thresholds to obtain a reasonable drift severity.

To solve the above-mentioned problem, we propose to introduce the *normalization process*, *difference transformation* and *fuzzy membership mapping* to measure the severity of
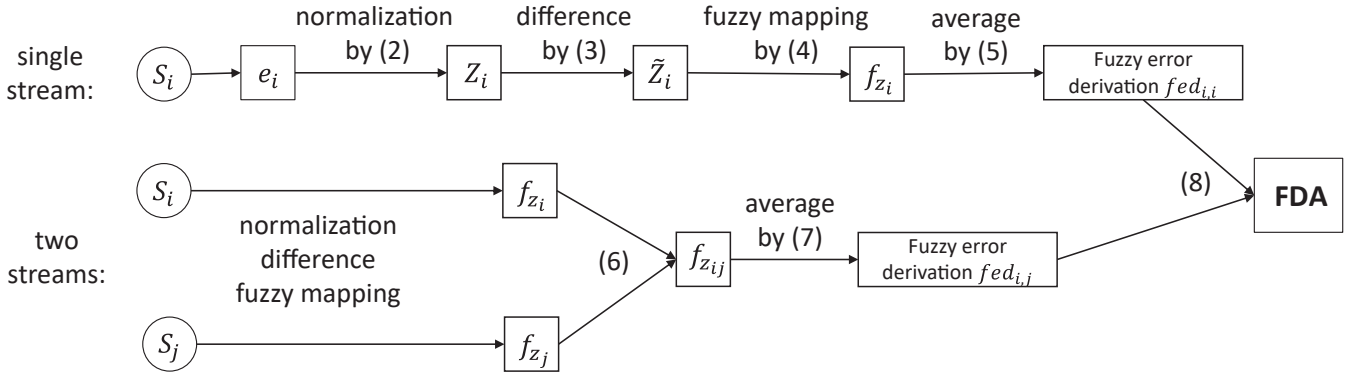
Fig. 1. Flowchart of computing FDA

the drift. Given the records of $\hat{e}_{i,t}$ for a specific period $\mathcal{T} = \{1,\ldots,T\}$, the normalization process is implemented by the Z-score, which is computed as (2), where $std(\hat{e}_{i,t})$ is the corresponding standard deviation.

$$Z_{i,t} = \frac{|\hat{e}_{i,t}|}{std(\hat{e}_{i,t})}. \tag{2}$$

The difference caused by the magnitude of streams will not affect the value of $Z_{i,t}$. Next, we propose a difference transformation on Z_{i_t} by (3). $\sum_{t'<t} Z_{i_{t'}}$ is a summation of Z-score before t, and $|t' < t|$ is the number of instances before t. Similarly, $\frac{\sum_{t' \geq t} Z_{i,t'}}{|t' \geq t|}$ is the average level of accumulated Z-value after t. The value of $\widetilde{Z}_{i,t}$ is small if the distribution of Z-score does not change during $\mathcal{T}$ while is large if the distribution changes, which exactly corresponds to the drift problem.

$$\widetilde{Z}_{i,t} = \begin{cases} \left| \frac{\sum_{t'<t} Z_{i,t'}}{|t'<t|} - \frac{\sum_{t'\geq t} Z_{i,t'}}{|t'\geq t|} \right| & t \neq 0 \\ Z_{i,t} & t = 0 \end{cases} \tag{3}$$

In the next step, $\widetilde{Z}_{i,t}$ is mapped to $fz_{i,t}$ by a sigmoid fuzzy membership with a,b its parameters. The sigmoid fuzzy membership function strictly limits the $fz_{i,t}$ in $[0,1]$, and distinctly differ instances with small $\widetilde{Z}_{i,t}$ from instances with large $\widetilde{Z}_{i,t}$.

$$fz_{i,t} = \frac{1}{1 + e^{-a(\widetilde{Z}_{i,t}-b)}}. \tag{4}$$

In (5), $fed_{i,i}$ for the ith stream during a specific period $\mathcal{T}$ is defined as the summation of $fz_{i,t}$ divided by the length of $\mathcal{T}$. $fed_{i,i}$ denotes the general drift severity for single stream. A larger value of $fed_{i,i}$ means that the drift occurs more frequently in ith stream.

$$fed_{i,i} = \frac{\sum_{t \in \mathcal{T}} fz_{i,t}}{|\mathcal{T}|}. \tag{5}$$

C. Fuzzy error deviation between two streams

To compute $fed_{i,j}(i \neq j)$, the first step is to compute $\widetilde{Z}_{i,t}$ and $\widetilde{Z}_{j,t}$. After that, $fz_{ij,t}(l)$ is computed in (6) and $fed_{ij,t}(l)$ is calculated by (7) respectively. Compared to the single stream

case, $fz_{ij,t}(l)$ and $fed_{ij,t}(l)$ are dependent with the delay parameter l in $M(i,j,l)$.

$$fz_{ij,t}(l) = |fz_{i,t} - fz_{j,t+l}| \tag{6}$$

$$fed_{i,j}(l) = \frac{\sum_{t \in \mathcal{T}} fz_{ij,t}(l)}{|\mathcal{T}|} \tag{7}$$

If the pattern of the occurrence of drift in ith and jth stream has correlation, for example, drift always occurs simultaneously in these two stream. $fed_{i,i}$ and $fed_{j,j}$ will have large values because drift occurs in both of these two streams. However, $fed_{i,j}(l = 0)$ will be small because $fz_{i,t}$ and $fz_{j,t}$ share the same increasing and decreasing pattern, and the item of $|fz_{i,t} - fz_{j,t}|$ will be small.

D. Simultaneous or delayed drift-correlation

According to the previous two subsection, we know that if a data streams $\boldsymbol{S}_i$ has a small value of $fed_{i,i}$, it is not likely to have drift problem, and therefore, for $j \neq i$, $v_{ij}^l = 0$. If $\boldsymbol{S}_i$ and $\boldsymbol{S}_j$ have large values of $fed_{i,i}$ and $fed_{j,j}$ respectively but $fed_{i,j}(l)$ is small, it is possible that ith and jth stream have an l-step delayed drift correlation.

The last difficulty is to determine how small the $fed_{i,j}(l)$ is denoting a drift correlation. As $fz_{i,t}$ is a positive scalar, $|fz_{i,t} - fz_{j,t+l}|$ is definitely smaller than $\max(fz_{i,t}, fz_{j,t+l})$, and also very likely to be smaller than $\min(fz_{i,t}, fz_{j,t+l})$ if both $fz_{i,t}$ and $fz_{j,t+l}$ are large.

We solve this difficulty from a different aspect. If ith stream and jth stream have strong l-step delayed drift correlation, $fz_{i,t} - fz_{j,t+l}$ is close to 0 for all t. We assume the distribution of $fz_{ij,t}(l)$ given ith stream and jth stream have strong l-step delayed drift correlation is $\mathcal{P}_{fz}$. The expectation of $\mathcal{P}_{fz}$ is 0 and the variance σ_{fz}^2 is unknown. $fed_{i,j}(l)$ is the sum of $fz_{ij,t}(l)$. According to the central limit theorem, $fed_{i,j}(l) \sim \mathcal{N}(0, \sigma_{fz}^2/|\mathcal{T}|)$. We use $\frac{\sum_{t \in \mathcal{T}} fz_{ij,t}^2(l)}{|\mathcal{T}|-1}$ to estimate σ_{fz}^2, denoted by $\hat{\sigma}_{fz}^2$. If the value of $fed_{i,j}(l)$ is less than $1.96 \times \hat{\sigma}_{fz}$, the $fed_{i,j}(l)$ is considered small enough to demonstrate a strong drift correlation between ith and jth stream and its corresponding $v_{ij}^l = 1$.
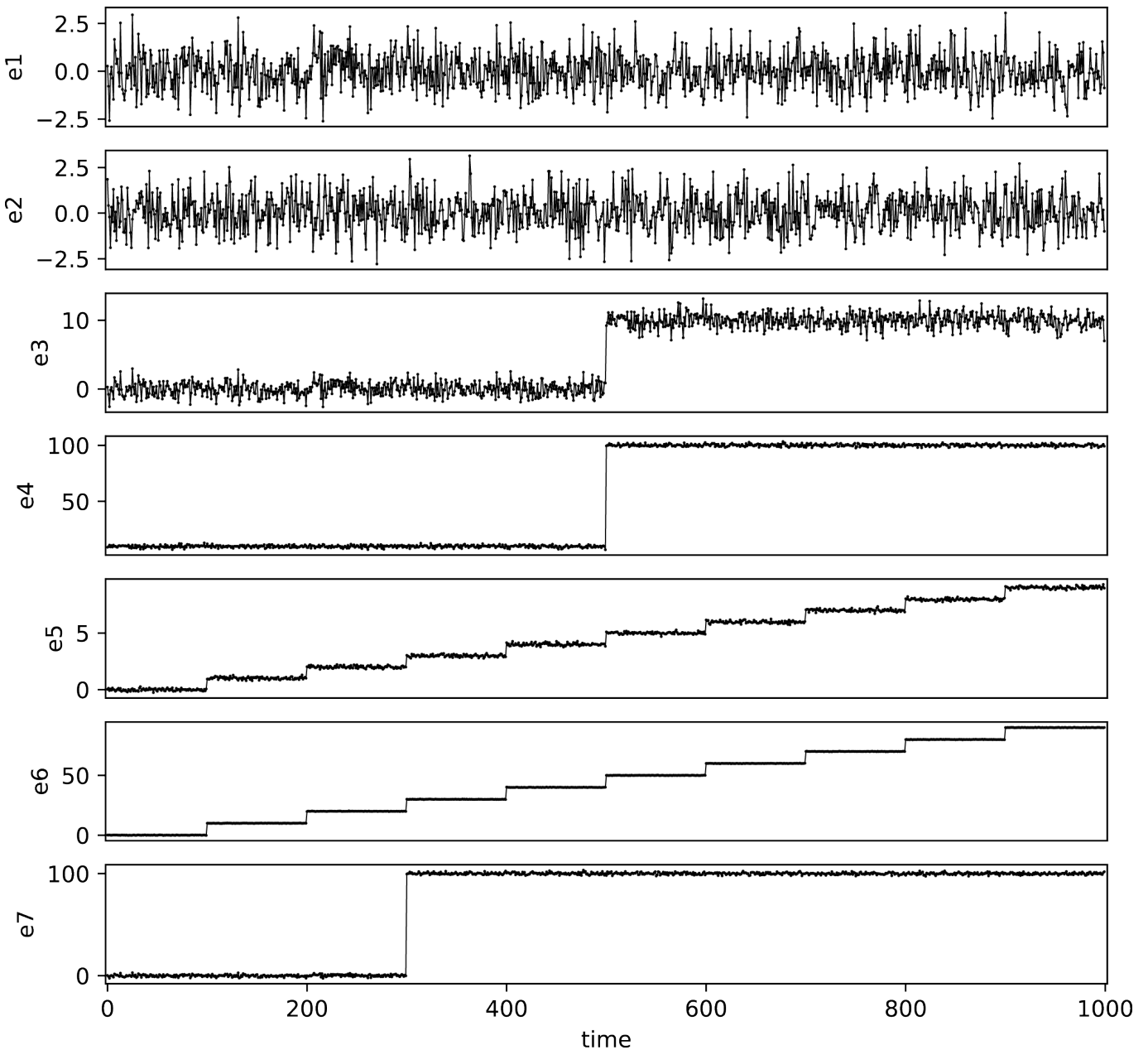
Fig. 2. Simulated error series multiple data streams.

In summary, the FDA is computed in (8), where $i \neq j$, ρ is the threshold for the small $fed_{i,i}$ and $\lambda = 1.96 \times \frac{\sum_{t \in \mathcal{T}} fz_{ij,t}^2(l)}{|\mathcal{T}|-1}$.

$$v_{ij}^l = \begin{cases} 1 & \min(fed_{i,i}, fed_{j,j}) > \rho, fed_{i,j}(l) < \lambda \\ 0 & else \end{cases} \quad (8)$$

IV. EXPERIMENTAL EVALUATION

To test whether the proposed FDA can present drift correlation well, we simulate seven error series for multiple streams.

A. Data Simulation

We assume to mimic a scenario of the multi-stream containing seven streams, each of them is denoted by $S_1, \ldots, S_7$

TABLE I
EXPECTED FDA

$l=0$	S_1	S_2	S_3	S_4	S_5	S_6	S_7
S_1	0	0	0	0	0	0	0
S_2	0	0	0	0	0	0	0
S_3	0	0	1	1	0	0	0
S_4	0	0	1	1	0	0	0
S_5	0	0	0	0	1	1	0
S_6	0	0	0	0	1	1	0
S_7	0	0	0 1 ($l=200$)	0 1 ($l=200$)	0	0	1

TABLE II
COMPUTED FUZZY-DRIANCE MATRIX

$l=0$	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$	$\mathcal{S}_5$	$\mathcal{S}_6$	$\mathcal{S}_7$
$\mathcal{S}_1$	0	0	0	0	0	0	0
$\mathcal{S}_2$	0	0	0	0	0	0	0
$\mathcal{S}_3$	0	0	1	1	0	0	0
$\mathcal{S}_4$	0	0	1	1	0	0	0
$\mathcal{S}_5$	0	0	0	0	1	1 1 ($l=150$) 1 ($l=200$) 1 ($l=250$)	0
$\mathcal{S}_6$	0	0	0	0	1 1 ($l=150$) 1 ($l=200$) 1 ($l=250$)	1	0
$\mathcal{S}_7$	0	0	0 1 ($l=200$) 1 ($l=250$)	0 1 ($l=200$) 1 ($l=250$)	0	0	1

separately. Instead of generating the full data of each stream, we directly simulate the error series during a time period of 1000. It is assumed that each stream has at most one type of drift, but the type of the drift can be sudden or incremental.

The error series $e_1, \ldots, e_7$ are generated as follows:

- e_1 and e_2 are two samples of size 1000 independently drawn from the normal distribution $\mathcal{N}(0,1)$.
- e_3 contains samples from two distributions. The first 500 instances of e_3 are samples from the normal distribution $\mathcal{N}(0,1)$, and the last 500 instances of e_3 are samples from another normal distribution $\mathcal{N}(10,1)$.
- e_4 contains instances from two distributions. The first 500 instances of e_4 are samples from $\mathcal{N}(0,1)$, and the last 500 instances are from $\mathcal{N}(100,1)$.
- e_5 contains instances from ten distributions to mimic an incremental drift. The first 100 instances are from $\mathcal{N}(0,1)$, the following 100 instances are from $\mathcal{N}(1,1)$ and so on, and the last 100 instances are from $\mathcal{N}(9,1)$.
- e_6 also mimic an incremental drift but with different magnitude compared to e_5. The first 100 instances are from $\mathcal{N}(0,1)$, the following 100 are from $\mathcal{N}(10,1)$ and so on, and the last 100 instances are from $\mathcal{N}(90,1)$.
- e_7 contains instances from two distributions. The first 300 instances are from $\mathcal{N}(0,1)$, and the rest 700 instances are from $\mathcal{N}(100,1)$.

The simulated error series are presented in Fig. 2. e_1 and e_2 are assumed to simulate the predictive error of two streams with the same magnitude that do not have drift problems. As for e_3, e_4 and e_7, all of them contains the sudden drift problem. The occurrence of sudden drift of e_3 and e_4 are at the 500th instance while that of e_7 is earlier at the 300th instance. e_4 and e_7 have same magnitude while the magnitude of e_3 is smaller after drift occurs. e_5 and e_6 have simultaneous drift pattern but with different magnitude.

B. Evaluation and results

In this section, the final computational results of FDA is presented as well as the results of fuzzy error deviation. To evaluate the results, we manually analyze the seven error series and consider the analyzed FDA as the ground truth. The effectiveness of the FDA is determined by comparison it to the analyzed FDA.

C. Expected results of FDA

According to $e_1, \ldots, e_7$, it can be inferred that $\mathcal{S}_1$ and $\mathcal{S}_2$ do not contain drift problem, and therefore we expect $fed_{1,1}$ and $fed_{2,2}$ are small. The other streams all contain drift problem, and therefore we expect $fed_{3,3}, \ldots, fed_{7,7}$ are large. $\mathcal{S}_3$ and $\mathcal{S}_4$ have different magnitude but drift occurs at the same time in these two streams, so we expect $v_{34}^0 = 1$. Similarly, we expect $v_{56}^0 = 1$. $\mathcal{S}_7$ contains sudden drift but drift occurs 200-step earlier than $\mathcal{S}_3$ and $\mathcal{S}_4$. Accordingly, we expect $v_{73}^{200} = 1$. Based on the above analysis, the expected FDA is presented in Table I.

TABLE III
FUZZY ERROR DEVIATION FOR SINGLE STREAM

	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$	$\mathcal{S}_5$	$\mathcal{S}_6$	$\mathcal{S}_7$
$fed_{i,i}$	1.4E-03	2.6E-03	6.1E-01	6.2E-01	8.9E-01	8.9E-01	5.4E-01

D. The computational results of FDA

Fuzzy error deviation for each single stream $fed_{i,i}$ is the basis of computing fuzzy-driance. It measures the severity of drift in a data stream. The fuzzy error deviation of $\mathcal{S}_1, \ldots, \mathcal{S}_7$ is given in Table III. In this paper, we set $\rho = 0.3$. $fed_{1,1}$ and $fed_{2,2}$ is small which indicates that $\mathcal{S}_1$ and $\mathcal{S}_2$ do not contain drift. The other five streams all have large fed, indicating drift occurs in these streams.

Evaluation of FDA for the stream which does not contain drift: as $\mathcal{S}_1$ and $\mathcal{S}_2$ do not contain drift, $M(i = 1, j, l)$ and

$M(i = 2, j, l)$ are 0 for all j and l in Table II, which is consistent with the expected results in Table I.

Evaluation of FDA for simultaneous drift correlation: in Table II, the value of $M(3, 3, 0)$, $M(3, 4, 0)$, $M(4, 4, 0)$, $M(4, 3, 0)$, $M(5, 5, 0)$, $M(5, 6, 0)$, $M(6, 6, 0)$, $M(6, 5, 0)$, and $M(7, 7, 0)$ are 1. According to these results, $\mathcal{S}_3$ and $\mathcal{S}_4$ have simultaneous drift correlation, and $\mathcal{S}_5$ and $\mathcal{S}_6$ have simultaneous drift. This result is also consistent with the expected result.

Evaluation of FDA for delayed drift correlation: the value of $M(7, 3, 200)$ and $M(7, 4, 200)$ are 1, indicating a delayed drift correlation between $\mathcal{S}_7$ and $\mathcal{S}_3$, and $\mathcal{S}_7$ and $\mathcal{S}_4$. This is consistent with the expected results.

However, there are also some inconsistent results. One inconsistence is that the simultaneous occurred incremental drift may lead to a delayed drift correlation. As is in Table II, when $l \neq 0$, $M(5, 6, l)$ and $M(6, 5, l)$ are 1, but they are supposed to be 0. Unlike sudden drift, the incremental drift occurs in a period, inducing the delayed drift correlation.

The other inconsistence is about the value of $M(7, 3, l)$ and $M(7, 4, l)$. When $l = 200$, they equal to 1, which is in our expectation. However, when $l = 250$ they are still 1. Therefore, FDA is able to identify the delayed drift correlation but it can not precisely identify the delayed step.

E. Discussion

The comparison between expected FDA and computed FDA shows the advantages and drawbacks of the proposed FDA.

Advantages of FDA: it can effectively identify the non-drift stream; it can effectively identify the simultaneous drift correlation with regard to different drift types.

Drawbacks of FDA: it has limited ability to identify delayed drift correlation. FDA is also likely to identify delayed drift correlation even streams only contain simultaneous correlation.

V. Conclusion

In this paper, we proposed a error-based measurement termed FDA to measure the correlated drift patterns in multi-stream scenario. Seven error series are generated to test FDA. The experimental results show that our FDA performs good for non-drift correlations as well as simultaneous drift correlations with regards to different types of drift. It has drawbacks when identifying delayed drift correlations. Our future work is to improve its performance for the delayed drift correlation.

Acknowledgment

References

[1] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM computing surveys*, vol. 46, no. 4, p. 44, 2014.

[2] M. Jaworski, P. Duda, and L. Rutkowski, "New splitting criteria for decision trees in stationary data streams," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 6, pp. 2516–2529, 2018.

[3] H. Ghomeshi, M. M. Gaber, and Y. Kovalchuk, "Eacd: evolutionary adaptation to concept drifts in data streams," *Data Mining and Knowledge Discovery*, vol. 33, no. 3, pp. 663–694, 2019.

[4] G. S. Gurjar and S. Chhabria, "A review on concept evolution technique on data stream," in *2015 International Conference on Pervasive Computing (ICPC)*. Pune, India, Jan. 8-10: IEEE, 2015, pp. 1–3.

[5] M. B. Harries, C. Sammut, and K. Horn, "Extracting hidden context," *Machine learning*, vol. 32, no. 2, pp. 101–126, 1998.

[6] S. Wang, L. L. Minku, and X. Yao, "A systematic study of online class imbalance learning with concept drift," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, pp. 4802–4821, 2018.

[7] I. Žliobaitė, "Learning under concept drift: an overview," *arXiv preprint arXiv:1010.4784*, 2010.

[8] N. Lu, G. Zhang, and J. Lu, "Concept drift detection via competence models," *Artificial Intelligence*, vol. 209, pp. 11–28, 2014.

[9] A. Liu, G. Zhang, and J. Lu, "Concept drift detection via equal intensity k-means space partitioning," *IEEE Transactions on Cybernetics*, vol. Early Access, doi:10.1109/TCYB.2020.2983962, pp. 1–14, 2020.

[10] D. Zambon, C. Alippi, and L. Livi, "Concept drift and anomaly detection in graph streams," *IEEE transactions on neural networks and learning systems*, no. 99, pp. 1–14, 2018.

[11] Y. Song, J. Lu, H. Lu, and G. Zhang, "Fuzzy clustering-based adaptive regression for drifting data streams," *IEEE Transactions on Fuzzy Systems*, pp. 1–1, 2019, doi:10.1109/TFUZZ.2019.2910714.

[12] A. Liu, G. Zhang, and J. Lu, "Region drift disagreement-based diverse instances weighting ensemble for concept drift adaptation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. Early Access, doi:10.1109/TNNLS.2020.2978523, pp. 1–16, 2020.

[13] J. Lu, A. Liu, Y. Song, and G. Zhang, "Data-driven decision support under concept drift in streamed big data," *Complex & Intelligent Systems*, pp. 1–7, 2019.

[14] A. Haque, H. Tao, S. Chandra, J. Liu, and L. Khan, "A framework for multistream regression with direct density ratio estimation," in *Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, Louisiana, USA, Feb 2-7, 2018, pp. 3149–3156.

[15] O. Koller, C. Camgoz, H. Ney, and R. Bowden, "Weakly supervised learning with multi-stream cnn-lstm-hmms to discover sequential parallelism in sign language videos," *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[16] J. Gama, R. Sebastião, and P. P. Rodrigues, "On evaluating stream learning algorithms," *Machine learning*, vol. 90, no. 3, pp. 317–346, 2013.

[17] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: A review," *IEEE Transactions on Knowledge and Data Engineering*, 2018, doi:10.1109/TKDE.2018.2876857.

[18] R. J. C. Bose, W. M. Van Der Aalst, I. Žliobaitė, and M. Pechenizkiy, "Dealing with concept drifts in process mining," *IEEE transactions on neural networks and learning systems*, vol. 25, no. 1, pp. 154–171, 2013.

[19] E. Lughofer, E. Weigl, W. Heidl, C. Eitzinger, and T. Radauer, "Recognizing input space and target concept drifts in data streams with scarcely labeled and unlabelled [sic] instances," *Information Sciences*, vol. 355-356, pp. 127–151, 2016.

[20] T. Dasu and S. Krishnan, "An information-theoretic approach to detecting changes in multi-dimensional data streams," in *the 38th Symposium on the Interface of Statistics, Computing Science, and Applications*, Pasadena California, USA, May 24-27, 2006, pp. 1–24.

[21] S. Chandra, A. Haque, L. Khan, and C. Aggarwal, "An adaptive framework for multistream classification," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. Indianapolis, Indiana, USA, Oct 24 - 28: ACM, 2016, pp. 1181–1190.