

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188		
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) 08-08-2023		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 31-Jul-2020 - 30-Jul-2023	
4. TITLE AND SUBTITLE Final Report: Event Detection for Streaming Analytics: An Intelligent Mathematical Paradigm			5a. CONTRACT NUMBER W911NF-20-1-0249		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER 060122		
6. AUTHORS			5d. PROJECT NUMBER 106012		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Xavier University of Louisiana 1 Drexel Drive New Orleans, LA 70125 -1098			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 76086-NC-REP.7		
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Kun Zhang
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			19b. TELEPHONE NUMBER 504-520-6700

RPPR Final Report

as of 08-Aug-2023

Agency Code: 21XD

Proposal Number: 76086NCREP
INVESTIGATOR(S):

Agreement Number: W911NF-20-1-0249

Name: Andrea M. Edwards
Email: aedwards@xula.edu
Phone Number: 5044867411
Principal: N

Name: PhD Kun Zhang
Email: kzhang@xula.edu
Phone Number: 5045206700
Principal: Y

Organization: **Xavier University of Louisiana**

Address: 1 Drexel Drive, New Orleans, LA 701251098

Country: USA

DUNS Number: 020857876

EIN: 720635884

Report Date: 30-Oct-2023

Date Received: 08-Aug-2023

Final Report for Period Beginning 31-Jul-2020 and Ending 30-Jul-2023

Title: Event Detection for Streaming Analytics: An Intelligent Mathematical Paradigm

Begin Performance Period: 31-Jul-2020

End Performance Period: 30-Jul-2023

Report Term: 0-Other

Submitted By: PhD Kun Zhang

Email: kzhang@xula.edu

Phone: (504) 520-6700

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees:

STEM Participants:

Major Goals: The objective of the proposed research is twofold. First, from the scientific or mathematical aspect, we propose to create an innovative, reliable and scalable event detection prototype with theoretical guarantees for streaming data. The proposed work begins from conceptually differentiating “novelties” from “anomalies”, then proceeds to substantially extending our newly-developed streaming data mining system to real-time event detection, and further creating advanced detection algorithms for more pressing yet understudied challenges in mining streaming data. To this end, we developed the following key techniques. 1). An adaptive online kernel density estimation based algorithm to accurately pinpoint isolated anomalies and cohesive novel patterns from unlabeled, concept-drifting data streams with noise; 2). A dynamically evolving recurrent neural network to reveal suspicious rare events either as a semi-supervised model in a finite label latency context or as a supervised model in an infinitely delayed label scenario; 3). An online margin-based learning method to effectively handle evolving feature spaces while performing scalable interpretable event detection over noisy streaming data without expert labels; and 4). A multi-task learning framework to collaboratively conduct reliable and stable event detection when facing multi-source asynchronous raw data streams. All these advances are embodied in an intelligent mathematical paradigm that offers big data practitioners an open-source toolbox of qualitatively different approaches. The ability to learn, distinguish and characterize anomalous and novel signals in (multiple) streaming data would enhance the effectiveness of online risk and threat assessment, prevention, and neutralization, enabling the DoD to renovate a considerable knowledge-centric real-time flow of information to assist in critical strategic decision making.

Second, from the educational perspective, we intend to offer a vital opportunity to involve African-American students in modern data science and big data analytics research. Educational goals of establishing “The Streaming Analytics Research Team (SART)”, offering “machine learning”, “high-performance/mathematical computing” and “predictive analytics” classes at the undergraduate level, and recruiting, advising and training undergraduate researchers are also integral to the proposed project.

Accomplishments: B. Project Accomplishments

B.1 Intellectual Accomplishment

In order to achieve this objective, our group conducted very active research with 5 peer-reviewed publications [1-5]

RPPR Final Report as of 08-Aug-2023

and 3 submitted manuscripts [6-8]. Zhang also received the following awards or recognitions. She is selected as Program committee member of ICDM(2020-2023) and IJCAI (2020-2023) ; Meta-Chair of the AAAI Conference on Artificial Intelligence (2020 – 2023) and Distinguished Program committee member of IJCAI 2022.

B.2 Educational Accomplishment

From 2020 to 2023, funded by this agreement, Severn Xavier STEM undergraduate researchers were involved in the proposed project. All of those students were African Americans, and three of them were female. During this period, three students parlayed their research into top-tech summer internship opportunities. Additionally, three of our multi-year SART students received STEM job offers.

In addition, the PI also participated in establishing a new academic program - "Health Informatics at Xavier. The PI also offered two new courses ("Machine Learning" and "Data Analytics") that introduced 30 students to data science studies. Those activities all aim to give the African-American students professional and career Development Edge.

In summary, this DOD agreement not only allowed us to perform a comprehensive study on data stream mining and its promising application on biomedical domains, but also foster and enrich the research experiences of the under-represented minority students at Xavier, and open opportunities for them in graduate schools or future careers in IT industries. Moreover, this support has motivated us to explore other important more challenging aspects of mining streaming data, such as novelty detection, which is concerned with identifying an unobserved pattern in new observations not included in training data.

Training Opportunities: From 2020 to 2023, funded by this agreement, Severn Xavier STEM undergraduate researchers were involved in the proposed project. All of those students were African Americans, and three of them were female. During this period, three students parlayed their research into top-tech summer internship opportunities. Additionally, three of our multi-year SART students received STEM job offers.

In addition, the PI also participated in establishing a new academic program - "Health Informatics at Xavier. The PI also offered two new courses ("Machine Learning" and "Data Analytics") that introduced 30 students to data science studies. Those activities all aim to give the African-American students professional and career Development Edge.

Results Dissemination: Publications

1. An Effective Cost-sensitive Sparse Online Learning Framework for Imbalanced Streaming Data Classification and Its Application to Online Anomaly Detection, Zhong Chen, Victor Sheng, Andrea Edwards, and Kun Zhang, Knowledge and Information Systems (KAIS), 65(1), pp. 59–87, 2023.
2. Proximal Cost-sensitive Sparse Group Online Learning, Zhong Chen, Huixin Zhan, Victor Sheng, Andrea Edwards, and Kun Zhang, The 2022 IEEE International Conference on Big Data (IEEE Big Data'22), pp. 495-504, Osaka, Japan, 2022.
3. Projection Dual Averaging based Second-order Online Learning, Zhong Chen, Huixin Zhan, Victor Sheng, Andrea Edwards, and Kun Zhang, The 22nd IEEE International Conference on Data Mining (ICDM'22), pp. 51-60, Orlando, Florida, United States, 2022.
4. CSRDA: Cost-sensitive Regularized Dual Averaging for Handling Imbalanced and High-dimensional Streaming Data, Zhong Chen, Zhide Fang, Victor Sheng, Andrea Edwards, and Kun Zhang, The 12th IEEE International Conference on Big Knowledge (ICBK'21), pp. 164-173, Auckland, New Zealand, 2021.
5. Adaptive Robust Local Online Density Estimation for Streaming Data, Zhong Chen, Zhide Fang, Victor S Sheng, Jiabin Zhao, Wei Fan, Andrea Edwards, and Kun Zhang, International Journal of Machine Learning and Cybernetics (JMLC), 12(6), pp. 1803-1824, 2021.

Submitted Manuscripts:

6. Sparse Online Learning in Open Feature Spaces with ℓ_1, ℓ_2 -Norm Constraint, submitted to IEEE ICDM 2023
7. Robust Sparse Online Learning in Open Feature Spaces with ℓ_1, ℓ_2 -Norm Constraint, submitted to IEEE ICDM 2023
8. Cost-sensitive Sparse Group Online Learning for Imbalanced Data Streams, submitted to the Journal of Machine Learning

RPPR Final Report

as of 08-Aug-2023

Honors and Awards: Zhang received the following awards or recognitions. She is selected as Program committee member of ICDM(2020-2023) and IJCAI (2020-2023) ; Meta-Chair of the AAAI Conference on Artificial Intelligence (2020 – 2023) and Distinguished Program committee member of IJCAI 2022.

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: PD/PI

Participant: Kun Zhang

Person Months Worked: 2.00

Project Contribution:

National Academy Member: N

Funding Support:

Participant Type: Co PD/PI

Participant: Andrea Edwards

Person Months Worked: 2.00

Project Contribution:

National Academy Member: N

Funding Support:

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)

Participant: Zhong Chen

Person Months Worked: 12.00

Project Contribution:

National Academy Member: N

Funding Support:

Participant Type: Undergraduate Student

Participant: Robert Daniels

Person Months Worked: 2.00

Project Contribution:

National Academy Member: N

Funding Support:

Participant Type: Undergraduate Student

Participant: Joshua Green

Person Months Worked: 2.00

Project Contribution:

National Academy Member: N

Funding Support:

Participant Type: Undergraduate Student

Participant: Gabriel Allen

Person Months Worked: 1.00

Project Contribution:

National Academy Member: N

Funding Support:

RPPR Final Report

as of 08-Aug-2023

Participant Type: Undergraduate Student

Participant: Alicia Simon

Person Months Worked: 2.00

Project Contribution:

National Academy Member: N

Funding Support:

Participant Type: Other Professional

Participant: Bo Cao

Person Months Worked: 1.00

Project Contribution:

National Academy Member: N

Funding Support:

ARTICLES:

Publication Type: Journal Article

Peer Reviewed: Y

Publication Status: 1-Published

Journal: 2021 IEEE International Conference on Big Knowledge (ICBK)

Publication Identifier Type: DOI

Publication Identifier: 10.1109/ICKG52313.2021.00031

Volume: Issue:

First Page #:

Date Submitted: 7/23/22 12:00AM

Date Published: 10/21/21 5:00AM

Publication Location:

Article Title: CSRDA: Cost-sensitive Regularized Dual Averaging for Handling Imbalanced and High-dimensional Streaming Data

Authors: Zhong Chen; Zhide Fang; Victor Sheng; Andrea Edwards; Kun Zhang

Keywords: Online learning; Streaming data; Imbalance ratio; Cost-sensitive metrics; Sparsity

Abstract: Class-imbalance is one of the most challenging problems in online learning due to its impact on the prediction capability of data stream mining models. Most existing approaches for online learning lack an effective mechanism to handle high-dimensional streaming data with skewed class distributions, resulting in insufficient model interpretation and deterioration of online performance. In this paper, we develop a cost-sensitive regularized dual averaging (CSRDA) method to tackle this problem. Our proposed method substantially extends the influential regularized dual averaging (RDA) method by formulating a new convex optimization function. Specifically, two R_1 -norm regularized cost-sensitive objective functions are directly optimized, respectively. We then theoretically analyze CSRDA's regret bounds and the bounds of primal variables. Thus, CSRDA benefits from achieving a theoretical convergence of balanced cost and sparsity for severe imbalanced and high-dimensional streaming data min

Distribution Statement: 1-Approved for public release; distribution is unlimited.

Acknowledged Federal Support: Y

RPPR Final Report as of 08-Aug-2023

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 4-Under Review

Journal: Knowledge and Information Systems

Publication Identifier Type:

Publication Identifier:

Volume:

Issue:

First Page #:

Date Submitted: 7/23/22 12:00AM

Date Published:

Publication Location:

Article Title: An Effective Cost-sensitive Sparse Online Learning Framework for Imbalanced Streaming Data Classification and Its Application to Online Anomaly Detection

Authors: Zhong Chen, Victor Sheng, Andrea Edwards, Kun Zhang,

Keywords: Online learning, Streaming data, Imbalance ratio, Cost-sensitive learning, Sparsity, Regularized dual averaging, Online anomaly detection

Abstract: Effective streaming feature selection in dynamic online environments is essential in numerous applications. However, most of the existing methods evaluate high-dimensional features individually and ignore the potentially pertainable group structures. Moreover, the class imbalance underlying streaming data may further impair the discriminative efficacy of the selected features, resulting in deteriorated classification performance. Motivated by this observation, we propose a cost-sensitive sparse group online learning (CSGOL) framework and its proximal version (PCSGOL) to handle imbalanced and high-dimensional streaming data. Specifically, we formulate this issue as a new cost-sensitive online optimization problem by imposing the ℓ_2 -norm, ℓ_1 -norm, and group-wise sparsity constraints in the dual averaging regularization. Inspired by the proximal optimization, we further introduce the average weighted distance in CSGOL and develop the PCSGOL method to achieve stable prediction results. We

Distribution Statement: 5-Distribution authorized to DoD Components only

Acknowledged Federal Support: Y

Publication Type: Journal Article

Peer Reviewed: Y

Publication Status: 4-Under Review

Journal: Machine Learning

Publication Identifier Type:

Publication Identifier:

Volume:

Issue:

First Page #:

Date Submitted: 7/23/22 12:00AM

Date Published:

Publication Location:

Article Title: Cost-sensitive Sparse Group Online Learning for Imbalanced Data Streams

Authors: Zhong Chen¹, Victor Sheng², Andrea Edwards¹ and Kun Zhang¹

Keywords: Imbalanced data streams, cost-sensitive learning, online learning, group sparsity, online anomaly detection

Abstract: Effective streaming feature selection in dynamic online environments is essential in numerous applications. However, most of the existing methods evaluate high-dimensional features individually and ignore the potentially pertainable group structures. Moreover, the class imbalance underlying streaming data may further impair the discriminative efficacy of the selected features, resulting in deteriorated classification performance. Motivated by this observation, we propose a cost-sensitive sparse group online learning (CSGOL) framework and its proximal version (PCSGOL) to handle imbalanced and high-dimensional streaming data. Specifically, we formulate this issue as a new cost-sensitive online optimization problem by imposing the ℓ_2 -norm, ℓ_1 -norm, and group-wise sparsity constraints in the dual averaging regularization. Inspired by the proximal optimization, we further introduce the average weighted distance in CSGOL and develop the PCSGOL method to achieve stable prediction results. We

Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors

Acknowledged Federal Support: Y

CONFERENCE PAPERS:

RPPR Final Report

as of 08-Aug-2023

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 2021 IEEE International Conference on Big Knowledge (ICBK)
Date Received: 23-Jul-2022 Conference Date: 07-Dec-2021 Date Published:
Conference Location: Auckland, New Zealand
Paper Title: CSRDA: Cost-sensitive Regularized Dual Averaging for Handling Imbalanced and High-dimensional Streaming Data
Authors: Zhong Chen¹, Victor Sheng², Andrea Edwards¹ and Kun Zhang^{1*}
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 0-Other
Conference Name: IEEE ICDM 2022
Date Received: 23-Jul-2022 Conference Date: 30-Nov-2022 Date Published:
Conference Location: Orlando, US
Paper Title: Proximal Cost-sensitive Sparse Group Online Learning
Authors: Zhong Chen¹, Victor Sheng², Andrea Edwards¹, and Kun Zhang¹
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 4-Under Review
Conference Name: IEEE ICDM
Date Received: 23-Jul-2022 Conference Date: 30-Nov-2022 Date Published:
Conference Location: Orlando, US
Paper Title: Projection Dual Averaging based Second-order Online Learning
Authors: Zhong Chen¹, Victor Sheng², Andrea Edwards¹ and Kun Zhang^{1*}
Acknowledged Federal Support: **Y**

Partners

,

I certify that the information in the report is complete and accurate:

Signature: kun zhang

Signature Date: 8/8/23 2:29PM

Final Progress Report (Aug. 2020 – Jul. 2023)

Submitted to the US Army Research Office

Contract #: W911NF-20-1-0249

PI: Kun Zhang, PhD

Co-PI: Andrea Edwards, PhD

Computer Science Department

Xavier University of Louisiana

July 28th, 2023

A. Summary of Project Objective

From Aug. 2020 to Jul. 2023, Dr. Kun Zhang (the PI) and Dr. Andrea Edwards (the Co-PI) were funded by the ARO award (W911NF-20-1-0249) to work on a research project entitled “Event Detection for Streaming Analytics: An Intelligent Mathematical Paradigm”.

The objective of the proposed research is twofold. **First**, from the scientific or mathematical aspect, we propose to create an innovative, reliable and scalable event detection prototype with theoretical guarantees for streaming data. The proposed work begins from conceptually differentiating “novelties” from “anomalies”, then proceeds to substantially extending our newly-developed streaming data mining system to real-time event detection, and further creating advanced detection algorithms for more pressing yet understudied challenges in mining streaming data. To this end, we developed the following key techniques. **1).** An adaptive online kernel density estimation based algorithm to accurately pinpoint isolated anomalies and cohesive novel patterns from unlabeled, concept-drifting data streams with noise; **2).** A dynamically evolving recurrent neural network to reveal suspicious rare events either as a semi-supervised model in a finite label latency context or as a supervised model in an infinitely delayed label scenario; **3).** An online margin-based learning method to effectively handle evolving feature spaces while performing scalable interpretable event detection over noisy streaming data without expert labels; and **4).** A multi-task learning framework to collaboratively conduct reliable and stable event detection when facing multi-source asynchronous raw data streams. All these advances are embodied in an intelligent mathematical paradigm that offers big data practitioners an open-source toolbox of qualitatively different approaches. The ability to learn, distinguish and characterize anomalous and novel signals in (multiple) streaming data would enhance the effectiveness of online risk and threat assessment, prevention, and neutralization, enabling the DoD to renovate a considerable knowledge-centric real-time flow of information to assist in critical strategic decision making.

Second, from the educational perspective, we intend to offer a vital opportunity to involve African-American students in modern data science and big data analytics research. Educational goals of establishing “The Streaming Analytics Research Team (SART)”, offering “machine learning”, “high-performance/mathematical computing” and “predictive analytics” classes at the undergraduate level, and recruiting, advising and training undergraduate researchers are also integral to the proposed project.

B. Project Accomplishments

B.1 Intellectual Accomplishment

In order to achieve this objective, our group conducted very active research with 5 peer-reviewed publications [1-5] and 3 submitted manuscripts [6-8]. Zhang also received the following awards or recognitions. She is selected as Program committee member of ICDM(2020-2023) and IJCAI (2020-2023) ; Meta-Chair of the AAAI Conference on Artificial Intelligence (2020 – 2023) and Distinguished Program committee member of IJCAI 2022.

B.2 Educational Accomplishment

From 2020 to 2023, funded by this agreement, **Severn Xavier STEM undergraduate researchers** were involved in the proposed project. All of those students were African Americans, and three of them were female. During this period, three students parlayed their research into top-tech summer internship opportunities. Additionally, three of our multi-year SART students received STEM job offers.

In addition, the PI also participated in establishing a new academic program - “Health Informatics at Xavier. The PI also offered two new courses (“Machine Learning” and “Data Analytics”) that introduced 30 students to data science studies. Those activities all aim to give the African-American students professional and career Development Edge.

In summary, this DOD agreement not only allowed us to perform a comprehensive study on data stream mining and its promising application on biomedical domains, but also foster and enrich the research experiences of the under-represented minority students at Xavier, and open opportunities for them in graduate schools or future careers in IT industries. Moreover, this support has motivated us to explore other important more challenging aspects of mining streaming data, such as novelty detection, which is concerned with identifying an unobserved pattern in new observations not included in training data.

Publications

1. An Effective Cost-sensitive Sparse Online Learning Framework for Imbalanced Streaming Data Classification and Its Application to Online Anomaly Detection, Zhong Chen, Victor Sheng, **Andrea Edwards, and Kun Zhang**, Knowledge and Information Systems (KAIS), 65(1), pp. 59–87, 2023.
2. Proximal Cost-sensitive Sparse Group Online Learning, Zhong Chen, Huixin Zhan, Victor Sheng, **Andrea Edwards, and Kun Zhang**, The 2022 IEEE International Conference on Big Data (IEEE Big Data'22), pp. 495-504, Osaka, Japan, 2022.
3. Projection Dual Averaging based Second-order Online Learning, Zhong Chen, Huixin Zhan, Victor Sheng, **Andrea Edwards, and Kun Zhang**, The 22nd IEEE International Conference on Data Mining (ICDM'22), pp. 51-60, Orlando, Florida, United States, 2022.
4. CSRDA: Cost-sensitive Regularized Dual Averaging for Handling Imbalanced and High-dimensional Streaming Data, Zhong Chen, Zhide Fang, Victor Sheng, Andrea Edwards, and Kun Zhang, The 12th IEEE International Conference on Big Knowledge (ICBK'21), pp. 164-173, Auckland, New Zealand, 2021.
5. Adaptive Robust Local Online Density Estimation for Streaming Data, Zhong Chen, Zhide Fang, Victor S Sheng, Jiabin Zhao, Wei Fan, **Andrea Edwards, and Kun Zhang**, International Journal of Machine Learning and Cybernetics (JMLC), 12(6), pp. 1803-1824, 2021.

Submitted Manuscripts:

6. Sparse Online Learning in Open Feature Spaces with $\ell_{1,\infty}$ -Norm Constraint, submitted to IEEE ICDM 2023
7. Robust Sparse Online Learning in Open Feature Spaces with $\ell_{1,2}$ -Norm Constraint, submitted to IEEE ICDM 2023
8. Cost-sensitive Sparse Group Online Learning for Imbalanced Data Streams, submitted to the Journal of Machine Learning