IIT-Deakin Research Internship

DR YE ZHU

SENIOR LECTURER

YE.ZHU@DEAKIN.EDU.AU

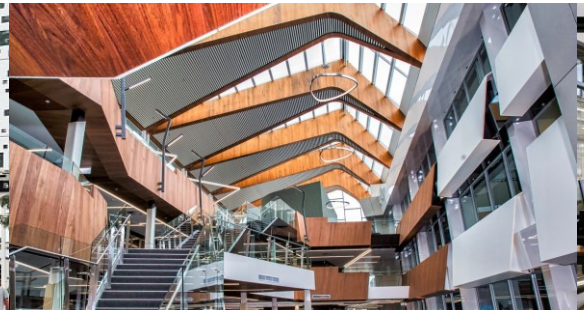
20/03/2024

Team for Universal Learning and Intelligent Processing

16 YEARS' EXPERIENCE IN AI RESEARCH

School of IT at Deakin University is ranked **No. 6 in the Australia** and **top 100 in the world**.

TULIP is the abbreviation of Team for Universal Learning and Intelligent Processing, and it belongs to Deakin University's strategic research centre – CREST. Since 2006, TULIP has started to take Honors students, High Degree by Research (Master by Research, PhD) students, as well as research interns. Our research have been funded by ARC, Indian DST Sparc, CRGS, FaST, DU-UB, HK GRF/HKPU and Chinese NSFC grants.




We conduct research in trusted intelligent techniques

Artificial Intelligence

How to automate various intelligent processing using advanced algorithms.

Business Intelligence

How to support business decision making using data science and artificial intelligence.

Privacy & Security

How to preserve the privacy of individuals while analyzing the data.

Research Internship@TULIP

- The Student Researcher Internship offers more opportunities for students to work on **high impact and critical research projects**.
- It allows opportunities **beyond the limitations** of our traditional internship program on aspects such as duration, time commitment, and working location.

Internship project in 2024

March – May 2024

Project-1: Time Series Anomaly Detection

Project-2: Mining Massive Trajectory Data

Project-3: Anomaly detection for industrial quality assurance

Project-4: Data Stream Clustering for Real-Time Data Analysis

Project leaders

Dr **Ye Zhu** is a Senior Lecturer (Associate Professor in US) of Computer Science at the School of IT, Deakin University, where he recently received both an Early Career Researcher Award and a Teaching and Learning Award. He holds a PhD from Monash University, which earned him the Mollie Holman Medal for the best doctoral thesis of the year in 2017. His research spans clustering analysis, anomaly detection, similarity learning, and pattern recognition, and has resulted in over 50 publications in top-tier conferences and journals, such as SIGKDD, ICML, IJCAI, VLDB, AAAI, ICDM, TKDE, AIJ, VLDBJ, ISJ, PRJ, JAIR, and MLJ. He has been co-chairing and serving on the program committees of prestigious international conferences.

Mr **Yang Cao** is currently a 3rd year PhD candidate at the School of IT, Deakin University, Australia. Before that, he received the B.S. and M.S. degree from Monash University and Deakin University, respectively. His research focuses on data mining, including anomaly detection, changing point detection and applications in different domains. His research has been published in Journal of Artificial Intelligence Research, Pattern Recognition, Expert Systems with Applications and PAKDD. He also obtained a Wong Swee Soon Prize from Monash University.

Project timeline

Week 1-2

- Identify the subproject topic and design the research problems
- Explore existing literatures, collecting data and source code

Week 2-4

- Understanding and preprocessing datasets
- Get familiar with the selected comparison methods

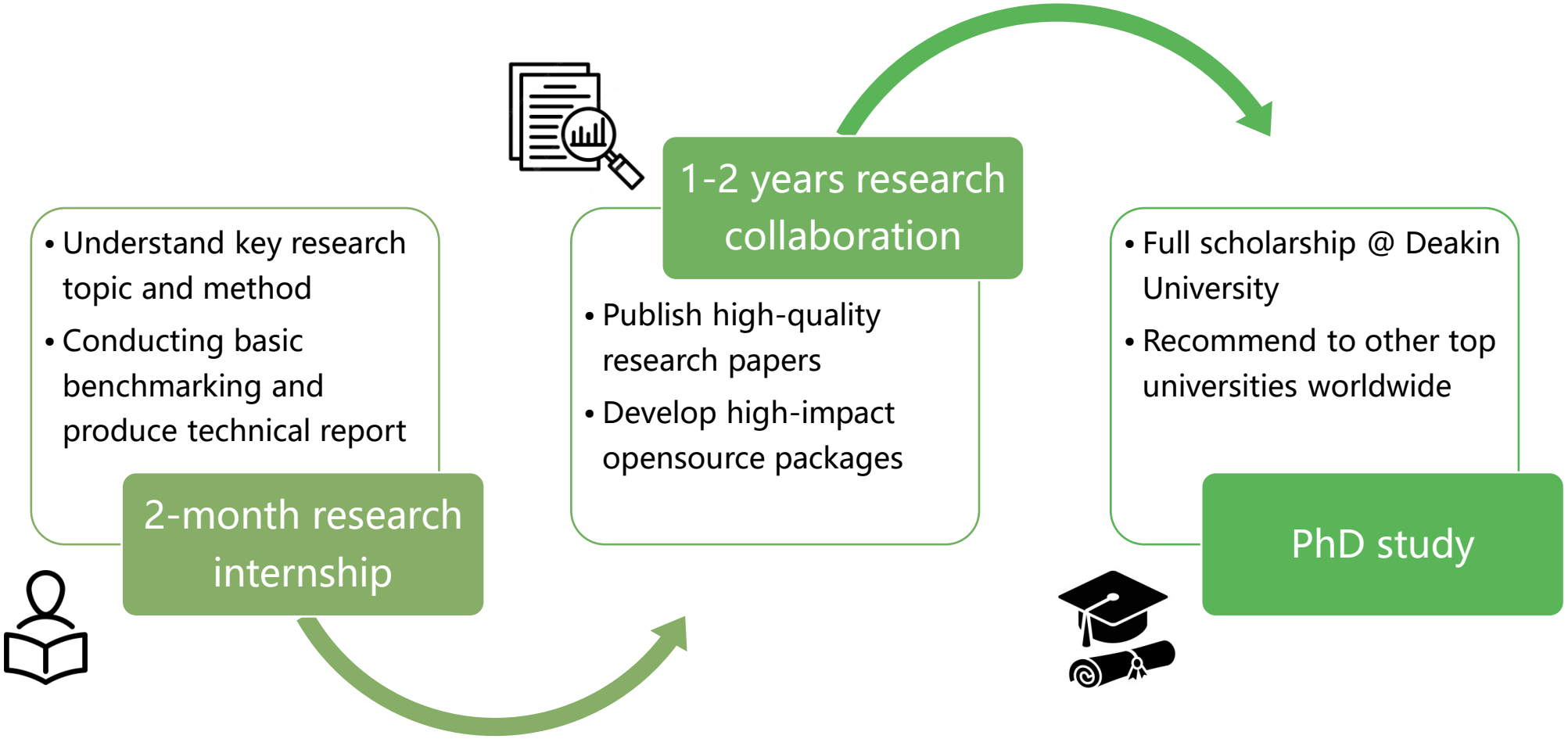
Week 4-6

- Benchmarking results using open-source packages and datasets
- Performance analysis of the evaluated models

Week 6-8

- Validating and organising all research data, code and results
- Writing technical research report

Project expectation and PhD pathway



Excellent intern

University of Pennsylvania (Wharton) is ranked global **No. 3 in Best** Business Schools.

Daniel Gao (He/Him)
Australian Delegate to G20 Youth || Student at the Wharton School

Research Intern under Dr. Gang Li
Deakin Research · Internship
Mar 2021 - Sep 2022 · 1 yr 7 mos
Melbourne, Victoria, Australia

Data-mining and data visualisation for PhD candidates at TULIP-lab

Intern
Anthony Raso & Associates
Jan 2020 - Feb 2020 · 2 mos
Melbourne, Victoria, Australia · On-site

Education

The Wharton School
Finance & Business Economics and Public Policy
Activities and societies: Penn Undergraduate Capital Partners (Project Lead/Sourcing Analyst), Moso Capital (Investment Analyst), Social Impact Consulting (Consultant)

Impact Factor: **5.1** / 5-Year Impact Factor: **4.4**

Available access | Research article | First published online December 1, 2022

Value creation in wine tourism – an exploration through deep neural networks

Daniel Gao, Haiyang Xia, [...], and Rob Law, +3 View all authors and affiliations

OnlineFirst | https://doi-org.ezproxy-b.deakin.edu.au/10.1177/13567667221140605

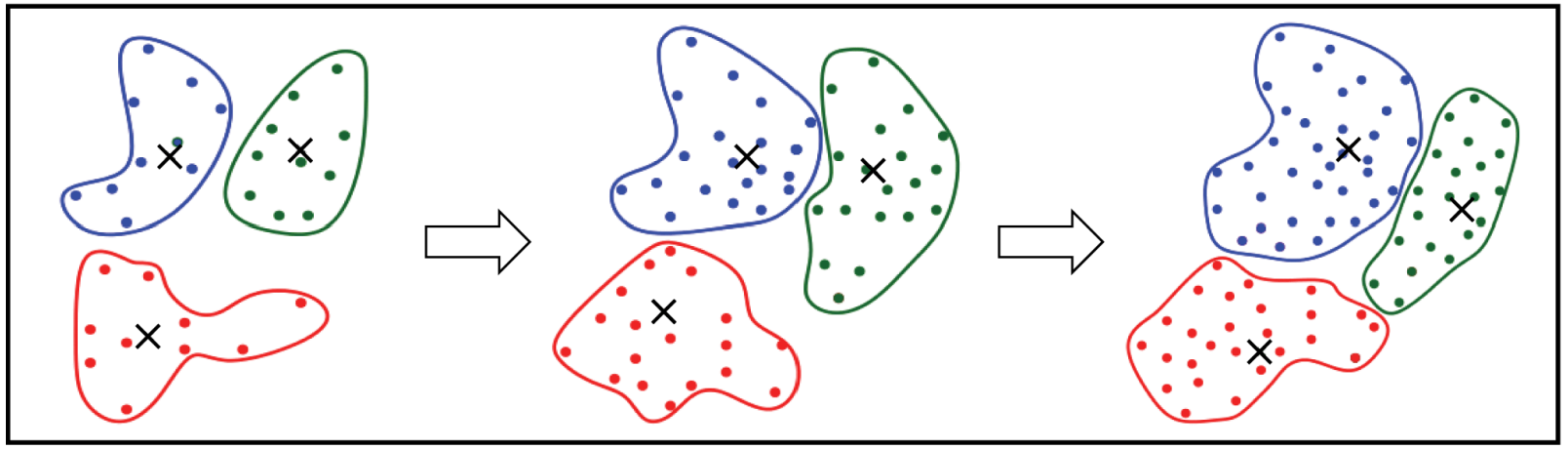
Contents | PDF / ePub | Cite article | Share options | Information, rights and permissions

Abstract

The aim of this paper is to explore what aspects create experiential value for wine tourists. We synthesize the extant literature into four dimensions for wine tourism value creation, namely, product-related aspects; sensory and affective experiential aspects; cognitive, educational experiential aspects; and social-relational experiential value-creating aspects. So far, most studies merely discuss product-related aspects whilst insights on experiential value are less known. Using online review data from wine tourists in Australia, we develop a novel deep neural network-based framework using an innovative AI-based exploratory design. Results of the case study reveal that in addition to product-related aspects, sensory-and education-related experiential aspects are also highly important for value creation in wine tourism. Theoretical and practical implications, as well as ideas for future research are discussed.

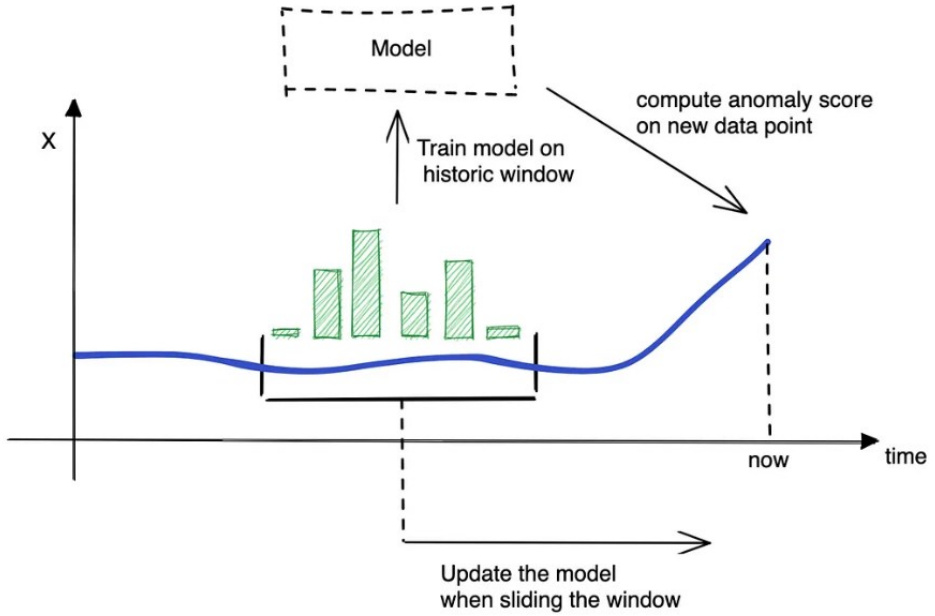
Sub research topics 1

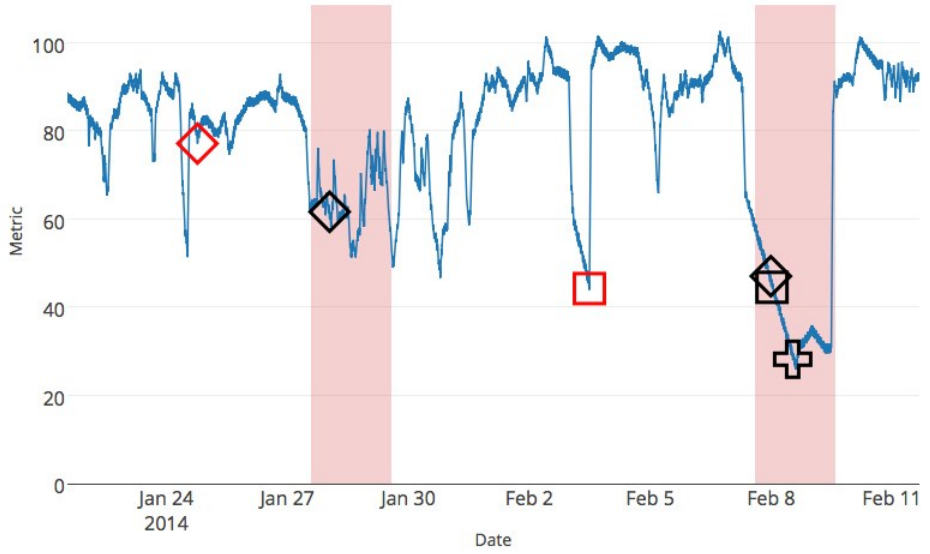
- **Streaming clustering:** explores evolving clustering representations that can dynamically update cluster models as new data arrives, capturing the underlying patterns and structures in the data stream.



Sub research topics 2

- **Streaming anomaly detection:** detecting anomalies in continuous data streams, where data arrives in real-time and needs to be processed sequentially.





Tasks for week 1 and 2

Identify the latest methods and **validate the source code** provided by the authors, for example:

https://hoanganhngo610.github.io/river-clustering.kdd.2022/related-materials.html

https://www.jmlr.org/papers/volume22/20-1380/20-1380.pdf

https://github.com/online-ml/river/

https://github.com/Stream-AD/MemStream?tab=readme-ov-file

https://github.com/Stream-AD/MStream

https://github.com/Stream-AD/MIDAS

https://github.com/Fengrui-Liu/StreamAD

Validate the source code

- Run the source code on the dataset to get the same results shown in the original paper

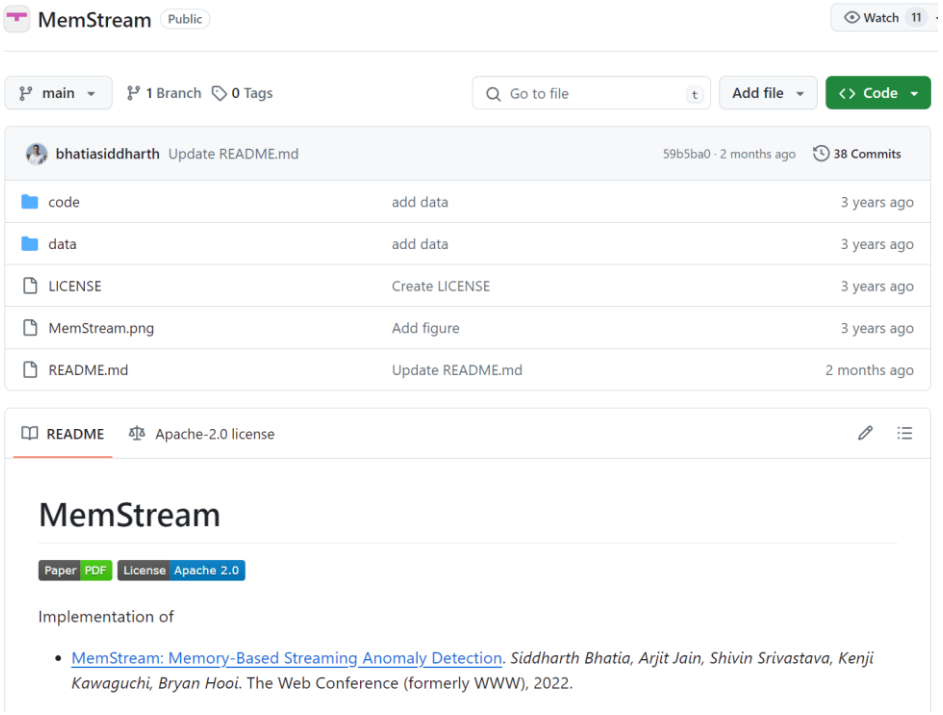


Table 2: AUC of MemStream and Streaming Baselines. Averaged over 5 runs.

Method	KDD99	NSL	UNSW	DoS	Syn.	Ion.	Cardio	Sat.	Sat.-2	Mamm.	Pima	Cover
STORM (CIKM'07)	0.914	0.504	0.810	0.511	0.910	0.637	0.507	0.662	0.514	0.650	0.528	0.778
HS-Tree (IJCAI'11)	0.912	0.845	0.769	0.707	0.800	0.764	0.673	0.519	0.929	0.832	0.667	0.731
iForestASD (ICONS'13)	0.575	0.500	0.557	0.529	0.501	0.694	0.515	0.504	0.554	0.574	0.525	0.603
RS-Hash (ICDM'16)	0.859	0.701	0.778	0.527	0.921	0.772	0.532	0.675	0.685	0.773	0.562	0.640
RCF (ICML'16)	0.791	0.745	0.512	0.514	0.774	0.675	0.617	0.552	0.738	0.755	0.571	0.586
LODA (ML'16)	0.500	0.500	– – –	0.500	0.506	0.503	0.501	0.500	0.500	0.500	0.502	0.500
Kitsune (NDSS'18)	0.525	0.659	0.794	0.907	– – –	0.514	0.966	0.665	0.973	0.592	0.511	0.888
DILOF (KDD'18)	0.535	0.821	0.737	0.613	0.703	**0.928**	0.570	0.561	0.563	0.733	0.543	0.688
xStream (KDD'18)	0.957	0.552	0.804	0.800	0.539	0.847	0.918	0.677	**0.996**	0.856	0.663	0.894
MStream (WWW'21)	0.844	0.544	0.860	0.930	0.505	0.670	**0.986**	0.563	0.958	0.567	0.529	0.874
Ex. IF (TKDE'21)	0.874	0.767	0.541	0.734	– – –	0.872	0.921	0.716	0.995	0.867	0.672	0.902
MemStream	**0.980**	**0.978**	**0.972**	**0.938**	**0.955**	0.821	0.884	**0.727**	0.991	**0.894**	**0.742**	**0.952**

Tips

1

Collect the code/papers published in good conferences/journals last 5 years

2

DO NOT try to fix the code if it cannot run correctly, just skip this method and try to find another paper

3

Follow the original paper to set or search the parameters of the model, in order to get its best performance.