

On Data Science and Machine Intelligence Research Innovation and Translation

A SHORT INTRODUCTION TO THE DATA SCIENCE AND MACHINE INTELLIGENCE LAB

Guandong Xu, Shaowu Liu and Qian Li *

Data Science and Machine Intelligence Lab, University of Technology Sydney, Australia.*

E-mail: {Guandong.Xu, Shaowu.Liu, Qian.Li}@uts.edu.au

ABSTRACT

The Data Science and Machine Intelligence (DSMI) lab at the University of Technology Sydney (UTS) is a joint taskforce of excellent academics and researchers dedicated towards excellence and innovation across academia and industry, with research priorities in data science and artificial intelligence. The lab is led by the founding director Professor Guandong Xu and currently has 7 academics and ~20 research students. In the past few years, the lab has received over \$6M external research funding from national funding bodies, governments, corporations, and private sectors. Supported by these funds, researchers at DSMI conducted world-class research and published papers in premier journals, such as TNNLS, TSE, TKDE, and world-top conferences, such as CVPR, AAI, IJCAI, SIGIR, SIGKDD, WWW, WSDM, ICDE, ICDM, EMNLP. The team also received a number of international and national awards from Data Analytics, Computer, information industry and financial services professional bodies, such as EFMA, ACS, and iAwards.

I. INTRODUCTION

The Data Science and Machine Intelligence (DSMI)¹ lab was founded by Professor Guandong Xu in 2015. The lab has intensively researched on topics of artificial intelligence, behavioural modelling, social computing, recommender systems, natural language processing, predictive and prescriptive analytics, advanced visualization, causality discovery, and causal inference. In the past few years, the lab has received over \$6M external research funding from national funding bodies, governments, corporations, and private sectors, such as Australian Research Council Discovery Projects (ARC-DP) and Linkage Projects (ARC-LP), Cooperative Research Centre Project (CRC-P), Colonial First State Investment, NSW Ministry of Health, OnePath, and Providence Asset Group. Supported by these funds, researchers at DSMI continuously publish papers in top conferences and journals in data science and artificial intelligence, such as TNNLS, TKDE, TSE, CVPR, AAI, IJCAI, SIGIR, SIGKDD, WWW, WSDM, ICDE, ICDM, EMNLP.

In addition to its research excellence, the DSMI lab is also renowned for its translational research in positively impacting the industry and society. For example, the lab's research

outcomes have been implemented and operationalised in the following real-world businesses:

- Our AI-based models have significantly simplified the life insurance underwriting process at OnePath, extending the insurance coverage for more vulnerable Australians.
- Our data analytics solutions have optimized the policies of National Emergency Access Target (NEAT) at NSW Ministry of Health, saving lives in emergency departments.
- Our undergoing project with Providence Asset Group incorporates AI into solar farms, delivering renewable energy solutions for world-class energy efficiency.

As a result of impact and recognition, the lab, alongside its industry partners, has received many awards at national and international levels. For example, the work with OnePath was selected by CeBIT (the largest international computer expo) in 2018 to present a feature on the red-carpet showcase, receiving industry awards such as the European Finance Management Association Workforce transformation award (No.3), Australian Computer Society Digital Disruptors Award winner, Australian Information Industry Association Awards, etc.

The above only provides a glance at the research at DSMI, and more details are given in the rest of this article.

II. RESEARCH AREAS

The DSMI lab mainly focuses on data science and artificial intelligence research, though it also involves interdisciplinary research, such as FinTech, digital health, and renewable energy. The key research areas include

- Recommender systems.
- Behavioural modelling, social computing, web mining.
- Text mining and NLP, software code analysis.
- Predictive analytics, time series forecasting.
- Causality and fairness in machine learning discovery.
- Knowledge graph and representation learning.

We briefly introduce the above research areas as follows.

A. Recommender Systems

Recommender System, an active domain of information provision, focuses on modelling user-centric information (e.g., access-logs, purchase history, rating records, and product reviews) to recommend new items. The recommendation aims

¹lab website: www.dsmi.tech

at improving user experience and loyalty, facilitating decision-making for users and creating more revenues for online businesses and merchants, and so on. Our strengths include

- Cold-start and long-tail recommendation.
- Social network and trust-based recommendation.
- Sequential-based recommendation.
- Geo-based recommendation.
- Explainable recommendation.

See (1; 2; 3) for examples of our research in this area.

B. Behaviour Modelling and Social Computing

Behaviour modelling and social computing are two fundamental research directions at DSMI. Behaviour modelling is to derive user behavioural patterns, preferences, profiles from user behaviours and user-generated textual data, while social computing is to analyze the social characteristics and trends demonstrated from intra-human or human-computer interactions. We continuously publish papers on these two topics and apply the developed techniques to help our industry partners with their business problems. Our strengths include

- Identifying misbehaviours or abnormal behaviours.
- Discovering underlying behaviour patterns.
- Techniques: topic modelling, representation learning, and embedding, rule mining, sequential pattern mining, anomaly detection, time series analysis.

See (4; 5; 6) for examples of our research in this area.

C. Text Mining and Natural Language Processing

Our research advanced the state-of-the-art in devising new algorithms in analysing meaning, topics, patterns from user-generated content, such as user comments and reviews, product descriptions, social media posts, and customer spoken language, e.g. call-logs for actionable business insights. For example, DSMI has developed analytic models using open-source software to identify risk factors in mental health and customer behaviour patterns in the R&D area, as well as applying to enterprise business operations for identifying high-risks to enable a preventative management approach. Those models have been successfully deployed in Colonial First State Investment (one Australian top-tier superannuation company) for customer churn prediction, and OnePath life Insurance for mental insurance insight analysis.

Our strengths include

- Term indexing, retrieval and search and ranking.
- Sentiment analysis and figurative detection.
- Topic modelling and summarisation.
- Word and context embedding algorithms.
- Latent semantic analysis.

See (7; 8; 9; 10) for examples of our research in this area.

D. Predictive Analytics

The lab has extensive experience in designing customised predictive analytical tools within business operating frameworks to develop automated assistive AI tools to augment the performance of staff (including improving operational resource

efficiency, quality assurance and fraud identification), as well as through developing more target and improved personalised customer experiences. Partnered companies include the Australian Taxation Office, OnePath, NSW Health, Colonial First State, Providence Asset Group. Our strengths include

- Domain-specific feature engineering.
- Structured and unstructured data integration.
- Time-series forecasting.
- Multi-label supervised learning.
- What-if analysis, prescriptive analytics.

See (11; 12; 5; 13) for examples of our research in this area.

E. Causality and Fairness in Machine Learning

Our research advances the state-of-the-art causal analysis theory and its interplay with interpretable machine learning (specifically deep learning, transfer learning, and supervised learning). Our research pursues the goal of enabling machine learning methods with causality and human-level intelligence. Although current interpretable models have greatly improved the interpretability landscape, they are unable to provide causal explanations for human-level intelligence. The causality is a clear and mathematically sound mechanism for users to understand the core of machine learning. How to empower the interpretable models with advanced causal explanations is largely unexploited. Fairness becomes a common concern by public applications in various areas, especially some regulated areas, such as online job-seeking and recruitment systems. Such fairness issue is caused by either training data bias, or limitation of machine learning algorithms. Our ongoing research along this line will research devising effective algorithms to alleviate the fundamental fairness challenge in machine learning research.

Our strengths include

- Causal inference and causal discovery.
- Causal reasoning for machine learning.
- Causality in the explainable recommendation.
- Visual causality analysis.
- Fairness in machine learning.

See (14; 15) for examples of our research in this area.

F. Knowledge Graph and Representation Learning

The multi-relationship of Knowledge Graph (KG) brings new challenges for Knowledge Graph analysis, it also makes the research on KG more attractive, because, with this kind of automatically extracted structured human knowledge, we have an opportunity to reveal the human knowledge reasoning patterns with analysis methodologies. As a result of KG analysis, KG can be used as a semantic enhancement for downstream application scenarios, such as Recommender System (RS). In our research, both KG completion and KG based downstream applications are studied. Although a huge amount of human knowledge facts have been collected from multiple open resources, existed KG is still incomplete. Part of our current research focuses on the following proper embedding models for KG completion, 1) entity & relation embedding model, 2) conceptual taxonomy integrated embedding model,

and 3) multi-relational graph sub-structure embedding. See (16; 17; 18; 19) for examples of our research in this area.

III. RESEARCH WITH REAL-WORLD IMPACT

This section summarises how the lab's research is translated into business solutions, improving business outcomes and benefiting society.

A. Using AI and a hybrid ESS solution to fully integrate solar generation into the distribution system

Funded by Australia Cooperative Research Centres Projects (CRC-P), in this project, we use the combination of AI and hydrogen to unleash the power of solar energy produced at solar farms for better energy productivity, efficiency, operation, and maintenance. DSMI provides the expertise in AI for a hybrid Energy Storage System (ESS) solution to integrate solar generation fully into the distribution system. The project aims to develop a widely applicable integrated package for small-scale solar farming, focusing not just on photovoltaic technologies and solutions, but on the monitoring, control, integration and optimisation of distributed solar farming.

B. Smart Personalized Privacy Preserved Information Sharing in Social Networks

Funded by Australia Research Council Discovery Projects (ARC-DP), this project aims to create a novel and effective method for privacy protection at the individual level, which is now a great concern of persons, businesses, and government agencies in this big data age. The project expects to build an automatic smart and practical personalised privacy-preserving system through removing the fundamental obstacles. The project will significantly advance human knowledge of privacy, and push Australia to the front line of the research field, and protect Australia better.

C. Reshaping Australian superannuation practice via big data analytics

Funded by Australia Research Council Linkage Projects (ARC-LP), this project aims to reform superannuation investment practices in Australia. Using sophisticated data analytics and machine-learning techniques, combined with economic modelling and quantitative finance. The project will try to understand the broad characteristics of Australian superannuation investors and their practice from a 'big data' perspective. The expected outcomes of this project are the identification of key determinants for successful superannuation behaviour to inform decision-making for better superannuation practices and policies. It is expected that this project will contribute to safeguarding the future of Australia's superannuation schemes, and to better financial security at retirement.

D. AI-enhanced Underwriting systems and insights of Mental Health

Funded by OnePath-Zurich, this project aims to optimise the AI-based underwriting risk engine, anti-selection detection, and conduct a pilot study of mental health disorder analysis and visualisation. This project has received the international EFMA workforce transformation award in 2020.

E. AI-enhanced Life Insurance underwriting automation and optimization for ANZ Wealth

Funded by OnePath/ANZ Wealth, this project aims to develop an AI-based underwriting risk engine to improve the current manual underwriting process in life insurance. The data-driven model provides personalised, efficient service with improved quality assurance for customers when they apply for insurance. This project is partnered with ANZ wealth.

F. Longitudinal Study on Taxpayer Behavioural Analysis

Funded by Australia Research Council Linkage Projects (ARC-LP), this project is to investigate the taxpayer behavioural patterns in Australian. Specific analytics lens is developed to reveal the characteristics of interested cohorts, e.g. debtor from longitudinal point of view. This project was jointly funded by ARC and Australian Taxation Office.

G. Personality mining via call log analysis

Funded by Colonial First State, this project is to devise a data model of big-five personality scores based on customer call centre logs. Big data analytics on textual, audio, and video data is used to train the personality analysis engine. The predicted personality traits will create value for various business applications.

H. Develop Deep Insights in Customer Retention

Funded by Colonial First State, this project incorporated big data of customers, e.g. demography, transaction, interaction, and behaviour information into predictions of customer churn. Machine learning-based prediction models were developed for various business products.

I. NSW Emergency Department Treatment Performance Analysis

Funded by NSW Ministry of Health, this project works on Emergency Department (ED) Performance Assessment initiative for review and adjustment of '4-hour' ED discharge policy, which is currently applied across Australian hospital ED by patient cohort mining, and patient in-hospital journey and re-admission prediction model based on linked GP data. These analytical results and models are adapted to policy change and decision-making support.

IV. RESEARCH FACILITIES

Researchers at DSMI have access to world-leading facilities at UTS, such as the Data Arena – a 360-degree interactive data visualisation facility set to change the way we view and interact with data, as shown in Figure 2. The lab also has access to world-class computing and storage infrastructure, including a total of 2000+ Intel Xeon cores, 20TB+ of RAM, and 500TB+ of SSD hard drive.



Fig. 1: DSMI is in the UTS Central building in Sydney's CBD.



Fig. 2: A large cylindrical screen, four metres high and ten metres in diameter. A high performance computer graphics system drives six 3D-stereo video projectors, edge-blended to create a seamless three-dimensional panorama.

V. PUBLIC RECOGNITION

The lab's research has drawn attention from both academia and industry, evidenced by awards and media release.

National and international awards:

- 2020 Global Efma-Accenture Insurance Innovation Award in Workforce Transformation.
- 2019 Digital Disruptors Winner Award of Skills Transformation of Small Work Teams by Australian Computer Society (Figure 3).



Fig. 3: 2019 Digital Disruptors Winner Award of Skills Transformation of Small Work Teams by Australian Computer Society

- 2019 Digital Disruptors Gold Award of Service Transformation for the Digital Consumer – Corporate by Australian Computer Society.
- 2019 NSW State Merit iAwards in Category of Research & Development Project of the year, Business Service Markets, and Data Insights Innovation of the year

by Australian Information Industry Association (AIIA) (Figure 4).



Fig. 4: 2019 NSW State Merit iAwards in Category of Research & Development Project of the year.

- 2018 Digital Disruptors Gold Award for Skill Transformation in Work Team by Australian Computer Society.
- 2018 Best Industry Application of Data Analytics Award by BigInsight Data and Ai Innovation Award (Figure 5).
- 2018 Best Industry Application of AI Award by BigInsight Data and Ai Innovation Award.



Fig. 5: 2018 Best Industry Application of Data Analytics Award by BigInsight Data and Ai Innovation Award.

Media release:

- Global Award for AI Transforming Insurance Underwriting, UTS FEIT News, 2020.
- AAI awards for fintech partnerships, UTS FEIT News, 2018.
- AI-led insurance innovation wins Awards, UTS Newsroom, 2018.
- Super future, UTS Newsroom, 2018.
- AI the future of insurance and underwriting, CeBIT, 2018 (Figure 6).



Fig. 6: 2018 CeBIT Sydney red carpet show.

- Colonial First State and UTS use machine learning to predict investors, Financial Review News, 2017.
- Early wins for OnePath's AI insurance underwriting project, ComputerWorld, 2017.
- ANZ Wealth exploring AI for insurance underwriting, ComputerWorld, 2017.
- ANZ and UTS seek AI underwriting, FinanceCareer, 2017.

VI. LOOKING INTO THE FUTURE

The vision of the lab is to become a world-class research lab, delivering high-quality publications and industry innovation to unleash the potential of our partners and benefit society. To maintain and lift our profile in innovative research and industry engagement, we keep recruiting creative, passionate, and self-motivated talented students to grow our team. We also welcome research collaborations internationally.

REFERENCES

- [1] D. Wang, X. Zhang, D. Yu, G. Xu, and S. Deng, "Came: Content-and context-aware music embedding for recommendation," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.
- [2] X. Wang, Q. Li, W. Zhang, G. Xu, S. Liu, and W. Zhu, "Joint relational dependency learning for sequential recommendation," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2020, pp. 168–180.
- [3] H. Ying, F. Zhuang, F. Zhang, Y. Liu, G. Xu, X. Xie, H. Xiong, and J. Wu, "Sequential recommender system based on hierarchical attention network," in *IJCAI International Joint Conference on Artificial Intelligence*, 2018.
- [4] N. N. Vo, X. He, S. Liu, and G. Xu, "Deep learning for decision making and the optimization of socially responsible investments and portfolio," *Decision Support Systems*, vol. 124, p. 113097, 2019.
- [5] Z. Saeed, R. A. Abbasi, I. Razzak, O. Maqbool, A. Sadaf, and G. Xu, "Enhanced heartbeat graph for emerging event detection on twitter using time series networks," *Expert Systems with Applications*, vol. 136, pp. 115–132, 2019.
- [6] J. Yin, Z. Zhou, S. Liu, Z. Wu, and G. Xu, "Social spammer detection: A multi-relational embedding approach," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2018, pp. 615–627.
- [7] R. Biddle, A. Joshi, S. Liu, C. Paris, and G. Xu, "Leveraging sentiment distributions to distinguish figurative from literal health reports on twitter," in *Proceedings of The Web Conference 2020*, 2020, pp. 1217–1227.
- [8] W. Wang, Y. Zhang, Y. Sui, Y. Wan, Z. Zhao, J. Wu, P. Yu, and G. Xu, "Reinforcement-learning-guided source code summarization via hierarchical attention," *IEEE Transactions on Software Engineering*, 2020.
- [9] N. N. Vo, S. Liu, J. Brownlow, C. Chu, B. Culbert, and G. Xu, "Client churn prediction with call log analysis," in *International Conference on Database Systems for Advanced Applications*. Springer, 2018, pp. 752–763.
- [10] Y. Wan, Z. Zhao, M. Yang, G. Xu, H. Ying, J. Wu, and P. S. Yu, "Improving automatic source code summarization via deep reinforcement learning," in *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*, 2018, pp. 397–407.
- [11] Y. Shu, Q. Li, S. Liu, and G. Xu, "Learning with privileged information for photo aesthetic assessment," *Neurocomputing*, 2020.
- [12] I. Razzak, R. A. Saris, M. Blumenstein, and G. Xu, "Integrating joint feature selection into subspace learning: A formulation of 2dpca for outliers robust feature selection," *Neural Networks*, vol. 121, pp. 441–451, 2020.
- [13] R. Biddle, S. Liu, P. Tilocca, and G. Xu, "Automated underwriting in life insurance: Predictions and optimisation," in *Australasian Database Conference*. Springer, 2018, pp. 135–146.
- [14] G. Xu, T. D. Duong, Q. Li, S. Liu, and X. Wang, "Causality learning: A new perspective for interpretable machine learning," *arXiv preprint arXiv:2006.16789*, 2020.
- [15] G. X. Y. Wu, J. Cao, "Fast: A fairness assured service recommendation strategy considering service capacity constraints," in *Proceedings of the 2020 ICSOC*, 2020.
- [16] Z. Zhou, S. Liu, G. Xu, and W. Zhang, "On completing sparse knowledge base with transitive relation embedding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 3125–3132.
- [17] Z. Zhou, S. Liu, G. Xu, X. Xie, J. Yin, Y. Li, and W. Zhang, "Knowledge-based recommendation with hierarchical collaborative embedding," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2018, pp. 222–234.
- [18] Z. Wang, Q. Li, G. Li, and G. Xu, "Polynomial representation for persistence diagram," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [19] C. Zheng, Y. Cai, J. Xu, H.-f. Leung, and G. Xu, "A boundary-aware neural model for nested named entity recognition," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 357–366.