

ICDAR'22: Intelligent Cross-Data Analysis and Retrieval

Minh-Son Dao
dao@nict.go.jp
National Institute of Information and
Communications Technology
Tokyo, Japan

Cathal Gurrin
Dublin City University
Dublin, Ireland
cgurrin@computing.dcu.ie

Michael Alexander Riegler
Simula Metropolitan Center for
Digital Engineering and University of
Tromsø, Norway
michael@simula.no

Yuta Nakashima
Osaka University
Osaka, Japan
n-yuta@ids.osaka-u.ac.jp

Duc-Tien Dang-Nguyen
University of Bergen and Kristiania
University College
Bergen, Norway
ductien.dangnguyen@uib.no

Mianxiong Dong
Muroran Institute of Technology
Muroran, Japan
mx.dong@csse.muroran-it.ac.jp

ABSTRACT

We have witnessed the rise of cross-data against multimodal data problems recently. The cross-modal retrieval system uses a textual query to look for images; the air quality index can be predicted using lifelogging images; the congestion can be predicted using weather and tweets data; daily exercises and meals can help to predict the sleeping quality are some examples of this research direction. Although vast investigations focusing on multimodal data analytics have been developed, few cross-data (e.g., cross-modal data, cross-domain, cross-platform) research has been carried on. In order to promote intelligent cross-data analytics and retrieval research and to bring a smart, sustainable society to human beings, the specific article collection on "Intelligent Cross-Data Analysis and Retrieval" is introduced. This Research Topic welcomes those who come from diverse research domains and disciplines such as well-being, disaster prevention and mitigation, mobility, climate change, tourism, healthcare, and food computing.

CCS CONCEPTS

• **Theory of computation** → **Design and analysis of algorithms**; • **Information systems**; • **Security and privacy** → **Privacy protections**; • **Computing methodologies** → **Machine learning**; **Knowledge representation and reasoning**; **Artificial intelligence**; **Computer vision**;

KEYWORDS

cross-data analysis; privacy-preserving; prediction model; time-series data forecast; cross-modal models; image clustering and segmentation; social multimedia; hidden knowledge reveal; transformer; federated learning

ACM Reference Format:

Minh-Son Dao, Michael Alexander Riegler, Duc-Tien Dang-Nguyen, Cathal Gurrin, Yuta Nakashima, and Mianxiong Dong. 2022. ICDAR'22: Intelligent

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICMR '22, June 27–30, 2022, Newark, NJ, USA

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9238-9/22/06...\$15.00

<https://doi.org/10.1145/3512527.3531441>

Cross-Data Analysis and Retrieval. In *Proceedings of the 2022 International Conference on Multimedia Retrieval (ICMR '22)*, June 27–30, 2022, Newark, NJ, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3512527.3531441>

1 INTRODUCTION

Data have played a critical role in human life. In the digital era, where data can be collected almost anywhere, at any time, and by anything, people can own a vast volume of real-time data reflecting their living environment in various granularity. From these data, people can extract the necessary information to gain knowledge towards becoming wise. Since data do not come from a sole source, they only reflect a small part of a massive puzzle of life. Hence, the more pieces of data can be collected and filled into a canvas, the faster the puzzle can be solved. If we consider a puzzle piece as single-modal data, the puzzle game becomes a multimodal data analytic problem. If we consider a group of puzzle pieces assembled as a segment of the puzzle as one domain (e.g., mountain, house, animal), the puzzle game becomes a multi-domain problem. If we consider a 3D puzzle game, we are talking of a multi-platform problem. Finally, the bidirectional mapping between puzzle pieces and the frame (e.g., sample picture of a puzzle) during the game can be considered as cross-data/domain/platform problem. In other words, we can use a set of data (i.e., multimodal data) from certain domains with analytic models built on one platform to infer (e.g., prediction, interpolation, query) data from another domain(s) and vice versa.

2 SCOPE

The domain of the research can vary from well-being, disaster prevention and mitigation, mobility, climate change, tourism, healthcare, and food computing, and urban management. Example topics of interest include but is not limited to the following

- Event-based cross-data retrieval.
- Data mining and AI technology.
- Multimodal/Crossmodal complex event processing
- Transfer Learning and Transformers.
- Multimodal/Crossmodal data associations hypotheses.
- Intelligent cross-data analysis and retrieval applications.
- Cross-datasets for Repeatable Experimentation.
- Federated Analytics and Federated Learning for cross-data.
- Privacy-public data collaboration.

- Integration of diverse multimodal data.

3 OBJECTIVE

Followed by the success of the ICMR-ICDAR workshops, the third ICDAR workshop aims to provide a playground to people interested in the workshop's topics. In this playground, people share their experiences and brave new ideas towards making cross-data more intelligent by compensating each type of data's strengths and propose a new way to analyze and retrieve cross-data under different perspectives.

4 INVITED AND KEYNOTE SPEAKERS

Hugo Lewi Hammer received the M.Sc. and Ph.D. degrees from the Norwegian University of Science and Technology, in 2003 and 2008, respectively. He is currently a Professor with the Department of Computer Science, Oslo Metropolitan University, Oslo, Norway and an Adjunct Chief Research Scientist at Simula Metropolitan Center, Oslo, Norway. Before joining OsloMet, he worked as a researcher with Norwegian Computer Center, Oslo, Norway. His research interests include computer intensive statistical methods, uncertainty quantification and transparent machine learning methods.

5 ACCEPTED PAPERS

In total 5 long, 5 short and once invited paper out of 21 papers were accepted for the workshop.

Salehi et al. present in their work entitled "*Is More Realistic Better? A Comparison of Game Engine and GAN-based Avatars for Investigative Interviews of Children*" a comprehensive evaluation of a system that utilizes different data sources to provide a realistic avatar that can be used to train police and child protection services in how to conduct interviews of maltreated children. In their study they focus on the visual aspects and explore if more visual realistic avatars lead to a better user experience.

Sharma et al. explore in their paper the usefulness of explainable artificial intelligence in a medical use case. In their paper "*Explainable Artificial Intelligence for Human Embryo Cell Cleavage Stages Analysis*" they study how different methods such as Grad-CAM, LIME and SHAP can be used to identify important characteristics in embryo images associated with specific embryo cleavage stages.

Tran et al. provide insights into this new federated learning field and sleep quality prediction from a boarder perspective. In their paper "*FedMCRNN: Federated Learning using Multiple Convolutional Recurrent Neural Networks for Sleep Quality Prediction*". It provides the necessary background for potential future research initiatives in sleep quality prediction. Thus, their findings have implications for the development of Artificial Intelligence (AI) doctors.

Gan et al. present an interesting survey paper titled "*IoT-based Multimodal Analysis for Smart Education: Current Status, Challenges, and Opportunities*". The authors discuss the current status, challenges, and opportunities and provide empirical evidence on IoT-based multimodal analysis capabilities for smart education.

La et al. propose a multimodal method for extracting image-text matching at the object level, image captioning for the global description of an image, and textual entailment between captions. They utilize boosting algorithm to improve the model predictions.

The approach is presented in "*Multimodal cheap fakes Detection by Utilizing Image Captioning for Global Context*" and evaluated on the MMSys'21 Grand Challenge dataset.

Hamza et al examine the challenges of intellectual property rights in Big Data applications. The paper, titled "*Towards intellectual property rights protection in big data*," addresses the emergence of proprietary Big Data sets that require an alternative conceptual framework along with security policies and regulations. The authors also address fair use of information assets, particularly on the xData cross-database platform. The authors emphasize the profound challenges of copyright on cross-database platforms and the paradigm shift from ownership to access control.

Nguyen, Hoang et al. propose three algorithms, random scheduling (RS), round-robin (RR), and proportional fair (PF), to allocate resources using federated learning with Poisson Point Process (PPP) distributed in the cellular networks. The title is "*Efficient Resource Allocation using Federated Learning in Cellular Networks*". This study demonstrates that FL is receiving attention in various applications, including wireless communication technology.

Chi-Hao, Quoc-Thang, et al. investigate existing algorithms from various DIF techniques and incorporate this knowledge into our DIF web tool, DeDigi. For DeDigi's design, based on the Design Science system, each design iteration considers user feedback on how the tool's design is perceived from their perspective. The outcome of the privacy-by-design tool is five approaches from various categories of the focused field, which is introduced in the paper "*DeDigi: A Privacy-by-Design Platform for Image Forensics*".

Tran and Saleem introduce a Rainfall and Traffic Congestion Prediction System (RTC-PS) to predict and discover congestion patterns in the future time, presented in "*Predicting High-risk Congestion Areas During Heavy Rain Using Multi Prediction Model and Maximum Periodic Frequent Pattern Algorithms*" paper. The system is extended to develop the navigator application to help people find the optimal traveling route during heavy rain and help urban managers have enough time to handle their resources.

Mallick et al. present in their paper, "*A Hybrid Transformer Network for Detection of Risk Situations on Multimodal Life-Log Health Data*", the architecture of a hybrid transformer to detect risk situations of fragile people living at home using visual and time-series data. Good evidence of cross-modal data models is introduced.

Vo and Nakashima show how a multi-modal deep neural network classifies the tone in political advertising videos in their paper "*Tone Classification for Political Advertising Video Using Multi-modal Cues*". Their work creates a fully automated analysis of political advertising videos with a cutting-edge pre-processing pipeline. When compared to a single-modal classifier, the results highlighted the potential of a multi-modal classification model.

6 SUMMARY

Overall, the workshop attracted research from several exciting and diverse domains ranging from medical applications to social sciences, from surveys to applications. Artificial Intelligence with domain-like prior knowledge is the majority research trend in this workshop. The significant common interest among papers in this workshop is leveraging multimodal data and cross-modal AI models to get substantial insights, especially with imbalanced datasets, and promises fruitful research in the future.