ICDAR'23: Intelligent Cross-Data Analysis and Retrieval

Guillaume Habault
KDDI Research, Inc.
Fujimino, Japan
xgu-habault@kddi.com

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

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

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

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

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

ABSTRACT

Recently, there has been an increased interest in cross-data research problems, such as predicting air quality using life logging images, predicting congestion using weather and tweets data, and predicting sleep quality using daily exercises and meals. Although several research focusing on multimodal data analytics have been performed, few studies have been conducted on cross-data research (e.g., cross-modal data, cross-domain, cross-platform). The article collection "Intelligent Cross-Data Analysis and Retrieval" aims to encourage research in intelligent cross-data analytics and retrieval and contribute to the creation of a sustainable society. Researchers from diverse domains such as well-being, disaster prevention and mitigation, mobility, climate, tourism and healthcare are welcome to contribute to this Research Topic.

CCS CONCEPTS

• **Theory of computation** → **Design and analysis of algorithms**; • **Information systems** → **Spatial-temporal systems**; • **Computing methodologies** → **Machine learning**; **Knowledge representation and reasoning**; **Artificial intelligence**; **Computer vision**; **Natural language processing**; **Learning latent representations**;

KEYWORDS

cross-data analysis; prediction model; cross-modal models; multimodal models; image clustering and segmentation; transformer; cheapfakes detection; out-of-context detection; enterprise social networks; feature extraction; graph neural networks; spatio-temporal prediction; congestion prediction; cross-domain recommendation; association analysis; social analytics; public mood;

1 INTRODUCTION

Data plays a critical role in human life. In the digital era, where data can be collected almost anywhere and anytime, people have access to a vast volume of real-time data that reflects their living environment differently. People can extract necessary information from these data to gain knowledge and become wiser. Nevertheless, data often comes from multiple sources and only reflects a small part of the big puzzle of life. Despite potentially missing some pieces, the goal is to capture the puzzle's image with the available pieces. The more pieces of data we can collect and assemble within a given frame, the faster we can solve the puzzle. The challenge becomes even more significant when dealing with multimodal data, cross-domain and cross-platform problems. A multimodal data puzzle would be one where pieces have different shapes and sizes. A cross-domain puzzle would be one where the pieces come from distinct sub-pictures. Finally, a cross-platform puzzle would be one where the pieces that need to be assembled come from different puzzles. But in all these scenarios, you still have to put the pieces of this multidimensional puzzle together to unveil the entire picture.

2 SCOPE

Research covers various fields such as well-being, disaster prevention and mitigation, mobility, climate, healthcare, urban management and more. Example topics include, but are not limited to:

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

3 OBJECTIVE

Followed by the success of the ICMR-ICDAR workshops, the fourth ICDAR workshop aims to provide a discussion space for people pushing the boundaries in handling multimodal and cross-domain data. During this workshop, people share their experiences and brave new ideas towards making cross-data more intelligent, which, in the end, will contribute to the development of a more acute and sustainable society.

4 KEYNOTE SPEAKER

Yan Liu received the M.Sc. and Ph.D. degrees from Carnegie Mellon University, in 2004 and 2007, respectively. She is currently a Professor in the Computer Science Department, Viterbi School of Engineering at University of Southern California (USC). Before joining USC, she was a research staff member at the IBM T.J. Watson Research Center. Her research interests include machine learning with applications to health, sustainability, and social media.

5 ACCEPTED PAPERS

Out of the 22 papers that were submitted, the workshop accepted a total of 9 papers, which included 4 long papers, 4 short papers, and 1 invited paper.

Macháček et al. propose a conditional diffusion probabilistic model to generate synthetic gastrointestinal polyp images. Authors investigate the usefulness at training phase of such synthetic images for binary image segmentation. Their paper "*Mask-conditioned latent diffusion for generating gastrointestinal polyp images*" demonstrates that the effectiveness of their approach is influenced by the chosen model architectures.

Niu et al. propose in their work entitled "*User-irrelevant Cross-domain Association Analysis for Cross-domain Recommendation with Transfer Learning*" an association analysis method for cross-domain recommendation without having to share user-relevant data between domains. They demonstrate the effectiveness of their method compared to the common baseline on existing public datasets, despite the difficulty in finding public datasets that include different domains with overlapping data.

Luo et al. introduce a novel multimodal architecture for classifying protest demonstration images in "*A Joint Scene Text Recognition and Visual Appearance Model for Protest Issue Classification*". The approach includes extracting text from placards appearing in the image to create a text modality and enhance image classification accuracy. The authors evaluate their proposal on a newly created dataset and preliminary results are promising with the potential for future advancements.

Gan et al. investigate the importance of coaching for improving driving behaviors and promoting road safety. In this survey paper entitled "*Procedural Driving Skill Coaching from More Skilled Drivers to Safer Drivers: A Survey*", authors provide a comprehensive review of existing research on coaching methods for drivers, including pre-drive, on-board, and post-drive coaching. They also identify the challenges and opportunities for providing continuous feedback to drivers, and offer insights into future research directions.

Ung et al. explore in their paper, "*Towards Multimodal Spatio-Temporal Transformer-based Models for Traffic Congestion Prediction*", the use of dynamic people-flow and rainfall data to improve traffic congestion prediction. The authors enhance a transformer-based model with an early fusion method in order to combine the multimodal data. This work in progress highlights the potential of a multimodal prediction model that includes people-flow data when compared to a unimodal version.

Nguyen et al. propose a cross-modal approach for detecting Cheapfakes by combining Natural Language Processing and Computer Vision techniques. As demonstrated in their paper, "*Leveraging Cross-Modals for Cheapfakes Detection*", the proposed method enhances the robustness of previous approaches. This advancement highlights the potential of the cross-modal approach to tackle real-world problems like Cheapfakes detection.

Pham et al. focus on detecting out-of-context media, which involves identifying context alterations. In this paper titled "*Detecting Cheapfakes using Self-Query Adaptive-Context Learning*", authors propose a Self-Query Adaptive-Context Learning method that can adapt to new contexts at inference time by enriching its knowledge through image search engine queries. Although this work is still in progress, it opens up new directions for detecting out-of-context misuses.

Konishi et al. propose a feature extraction method for Enterprise Social Networks that captures enterprise-specific features while learning their common characteristics. This approach, described in this paper titled "*CG-GNN: A Novel Compiled Graphs-based Feature Extraction Method for Enterprise Social Networks*", can also assist in revitalizing such social networks.

Prokopiou et al. introduce in their paper, "*MTSS: Movie Trailers Surveillance System using Social Media Analytics and Public Mood*", a movie trailer surveillance platform. The proposed approach gathers metrics from social media platforms and analyzes them for sentiment and emotion extraction. Such information monitors the popularity of movie trailers and support marketing campaigns.

6 SUMMARY

Overall, the workshop brought together a wide range of research areas, including medical and traffic congestion, as well as surveys and applications. Most of the research presented at the workshop focused on Artificial Intelligence (AI) that incorporates prior domain knowledge. There was a strong interest in leveraging multimodal data and cross-modal AI models to gain insights and improve performance. As a result, the workshop promises to foster fruitful discussions and drive exciting research in the future.