



The 2020 2nd International Conference on Big Data Engineering

2020 2nd International Conference on Big Data Engineering (BDE2020)

May 29-31, 2020

Supported by



Indexed by



Scopus

<http://www.bde.net/>



The 2020 2nd International Conference on Big Data Engineering

Welcome Message from Organizing Committee

It is our great pleasure to invite you to join our international conference -2020 2nd International Conference on Big Data Engineering (BDE 2020). This event will provide a unique opportunity for editors and authors to get together and share their latest research findings and results.

We're confident that over the two days you'll get the theoretical grounding, practical knowledge, and personal contacts that will help you build long-term, profitable and sustainable communication among researchers and practitioners working in a wide variety of scientific areas with a common interest in Big Data Engineering.

On behalf of all the conference committees, we would like to thank all the authors as well as the technical program committee members and reviewers. Their high competence, their enthusiasm, their time and expertise knowledge, enabled us to prepare the high-quality final program and helped to make the conference become a successful event.

We truly hope you'll enjoy the conference and get what you expect from the conference.

Organizing Committee

May 15, 2020



The 2020 2nd International Conference on Big Data Engineering

Conference Introductions

Welcome to 2020 BDE Shanghai conference. This conference is organized by ACM Chapter Singapore. The objective of the conference is to provide a platform for researchers, engineers, academicians as well as industrial professionals from all over the world to present their research results and development activities in Big Data Engineering.

Papers will be published in the following proceeding:

International Conference Proceedings Series by ACM (ISBN: 978-1-4503-7722-5), which will be archived in the ACM Digital Library, and indexed by Ei Compendex, Scopus and submitted to be reviewed by Thomson Reuters Conference Proceedings Citation Index (ISI Web of Science).

Conference website and email: <http://www.bde.net/> and bde.conference@gmail.com



Table of Contents

Keynote Speakers Introductions.....	1
Invited Speakers Introductions.....	4
Instructions for the Online Tool “ZOOM”	7
Presentation Instructions	13
Schedule for Conference	14
Morning Session	16
Opening Remarks & Testing (9:00-9:10)	16
Keynote Speech I (9:10-9:55)	16
Keynote Speech II (9:55-10:40)	17
Take a Break: 10:40-10:55	17
Keynote Speech III (10:55-11:40)	18
Invited Speech I (11:40-12:05)	19
Lunch 12:05-15:00	19
Afternoon Session	20
Invited Speech II (15:00-15:25)	20
Invited Speech III (15:25-15:50)	21
Take a Break: 15:50- 16:05	21
Oral Presentation Abstracts.....	22
Session 1.....	22
BE0007 (16:05-16:20)	22
BE0015 (16:20-16:35)	23
BE0016 (16:35-16:50)	23
BE0017 (16:50-17:05)	24
BE0024 (17:05-17:20)	24
BE0030 (17:20-17:35)	25
BE5001 (17:35-17:50)	25
Session 2.....	26
BE0004 (10:00- 10:15)	26
BE0008 (10:15- 10:30)	27
BE0023 (10:30- 10:45)	27
BE0034 (10:45- 11:00)	28
BE5004 (11:00- 11:15)	28



The 2020 2nd International Conference on Big Data Engineering

BE0025 (11:15- 11:30)	29
Lunch: 11:30- 14:30.....	29
Session 3.....	30
BE0013 (14:30- 14:45)	30
BE0019 (14:45- 15:00)	31
BE0020 (15:00- 15:15)	31
BE0021 (15:15- 15:30)	32
BE0022 (15:30- 15:45)	32
BE0028 (15:45- 16:00)	33
BE5002 (16:00- 16:15)	33

Keynote Speakers Introductions

Keynote Speaker I



Prof. Kai Hwang, IEEE Live Fellow

Chinese University of Hong Kong, Shenzhen, China

Prof. Kai Hwang is a Presidential Chair Professor at the Chinese University of Hong Kong (CUHK), Shenzhen, where he heads the AIRS Research Center for AI Cloud and IoT Edge Computing. Dr. Hwang received his Ph.D. in EECS from UC Berkeley, and he has taught at USC and Purdue University for 44 years, prior to joining CUHK in 2018. Currently, he also serves as a Chief Scientist of the Cloud Computing Center, Chinese Academy of Sciences.

Dr. Hwang has published 10 scientific books and over 260 original papers. His Google Scholar citation has exceeded 19,000 times with an h-index of 60. An IEEE Life Fellow, he received the Outstanding Achievement Award from the Computer Federation of China (CFC) in 2005, and the Lifetime Achievement Award, IEEE CloudCom in 2012. He has also performed consulting work with IBM, MIT Lincoln Lab, Academia Sinica, and INRIA in France.

Keynote Speaker II



Prof. Yao Liang

Purdue University School of Science,
Indiana University Purdue University, USA

Prof. Yao Liang received his B.S. degree in Computer Engineering and M.S. degree in Computer Science from Xi'an Jiaotong University, Xi'an, China. He received his Ph.D. degree in Computer Science from Clemson University, Clemson, USA, in 1997. He is currently a Professor in the Department of Computer and Information Science, Purdue University School of Science, Indiana University Purdue University, Indianapolis (IUPUI), USA. His research interests include wireless sensor networks, Internet of Things, cyberinfrastructure, multimedia networking, adaptive network control and management, machine learning, neural networks, datamining, data management and integration, data engineering, and distributed systems. His research projects have been funded by NSF. Prior to joining IUPUI, he was on the faculty of Department of Electrical and Computer Engineering at Virginia Tech, USA. He also had extensive industrial R&D experiences as a Technical Staff Member in Alcatel USA. Dr. Liang has published numerous papers on various prestigious journals and international conferences, and received two US patents. He has received the 2019 Glenn W. Irwin, Jr., M.D., Research Scholar Award at IUPUI. He has served regularly on Program Committees for various major international conferences, and served as a reviewer for numerous prestigious journals. Dr. Liang has given invited talks and lectures at various universities in US, Europe and China. He is a Senior Member of IEEE, and a Member of ACM.

Keynote Speaker III

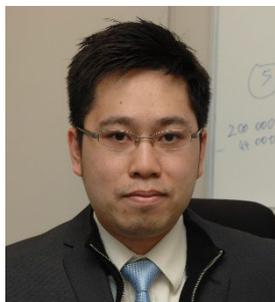


Prof. Changxu Wu
University of Arizona, USA

Prof. Changxu Wu received his Ph.D. degree in Industrial and Operational Engineering from the University of Michigan-Ann Arbor (2007). He is currently a tenured full professor of Department of Systems and Industrial Engineering University of Arizona, starting from August 2017. Dr. Wu directs the Cognitive System Lab and he is interested in integrating cognitive science and engineering system design, especially modeling human cognition system with its applications in system design, improving transportation safety, promoting human performance in human-computer interaction, and inventing innovative sustainable and smart energy systems with human in the loop. Dr. Wu has published 116 papers in the field including 80 journal papers, 36 conference papers, 1 book chapter, and 2 patents in intelligent system design authorized. The journal papers include IEEE Transactions on Systems, Man, and Cybernetics (Part A), IEEE Transactions on Intelligent Transportations Systems, Psychological Review (Impact Factor: 9.02), ACM Transactions on Computer-Human Interaction, International Journal of Human-Computer Studies, as well as several other journals. He was the Chair of Human Performance Modeling Technical Group of Human Factors and Ergonomics Society (HFES) in USA. He is also Associate Editors for IEEE Transactions on Intelligent Transportations Systems, IEEE Transaction on Human-Machine Systems, and Behaviour & Information Technology. He received the Senior Researcher of the Year Award from the Dean of School the Engineering & Applied Sciences at SUNY Buffalo and Outstanding Student Instructor Award from the American Society of Engineering Education (ASEE).

Invited Speakers Introductions

Invited Speaker I



Dr Ka-Chun Wong

City University of Hong Kong, Hong Kong

Dr. Ka-Chun Wong finished his PhD degree in Department of Computer Science at University of Toronto (where deep learning AI was popularized in 2010s) within 3.5 years (2012-13 departmental average: 6 years after master degree) by the end of 2014, after his RGC-funded MPhil degree (his academic family tree). He was merited as the first associate editor ever outside USA and Germany for the open-peer-review journal, BioData Mining, since 2016. He is also on the editorial board of Applied Soft Computing and Computer Modeling in Engineering and Sciences since 2016 and 2019 respectively. He was invited as the plenary/keynote speakers for ICBCB 2017, ISACIT 2018, DSIT 2019, and IC-LYCS 2019. In addition, he has solely edited 2 books published by Springer and CRC Press, attracting 30 peer-reviewed book chapters around the world (i.e. Argentina, Australia, Belgium, Brazil, China, Egypt, France, Germany, Hong Kong, India, Japan, Spain, and USA). In 2017 and 2018, without his Toronto affiliation, he has single-authored DNA motif informatics published on Bioinformatics and iScience (Cell Press) with CityU affiliation, demonstrating solid examples for his PhD students at CityU. After that, he led his CityU team to pursue the direction further as published on Nucleic Acids Research (2018 IF=11.1), iScience (Cell Press), and ACM Computing Surveys with sole CityU first authorship and sole CityU corresponding authorship in 2019, resulting in an independent research team with outputs published on high-impact journals (e.g. Nucleic Acids Research (2018 IF=11.1) and IEEE Transactions on Cybernetics (2018 IF=10.4)) with sole CityU corresponding authorship in a research autonomy manner since 2020. Lastly, Ka-Chun explicitly showed his track records since the 2003 SARS outbreak (when he took HKCEE) below, encouraging all peer students under the Hong Kong education system to trust and embrace their dreams under diverse settings.

Invited Speaker II



Dr Wei Li

Central Queensland University (CQU), Australia

Dr Wei Li holds a PhD degree in computer science from the Institute of Computing Technology of Chinese Academy of Sciences China. He currently works for the School of Engineering & Technology, Central Queensland University (CQU) Australia. He has been the Head of Masters of Information Technology Programme of CQU. His research interests include dynamic software architecture, big data, P2P volunteer computing and multi-agent systems. Dr Wei Li has been a peer reviewer of a number of international journals, including IEEE Transactions on Software Engineering, ELSEVIER Journal of Systems and Software and John Wiley & Sons Journal of Software Maintenance and Evolution: Research and Practice, and a program committee member of more than 30 international conferences.

Invited Speaker III



Dr Gabriella Casalino
University of Bari, Italy

Dr Gabriella Casalino is currently an Assistant Professor at the CILab laboratory of the department of Informatics, University of Bari, working on machine learning techniques applied to Web Economy domain. This position has been funded by the Italian Ministry of Education, University and Research (M.I.U.R.) through the European funding project AIM (Attraction and International Mobility).

Her research activity is focused on Computational Intelligence with a particular interest for data analysis. Three are the main themes she is currently working on: 1) Intelligent Data Analysis 2) Computational Intelligence for eHealth, 3) Data Stream Mining.

Topics in which she has produced original contributions include: image analysis, educational data mining, text mining, e-health, bioinformatics and signal processing.

Gabriella Casalino has got the Ph.D. in Computer Science at the Doctoral School in Computer Science of the Department of Informatics, at University of Bari "A. Moro" in 2015. She defended the thesis "Non-negative factorization methods for extracting semantically relevant features in Intelligent Data Analysis". She was awarded by the Italian Ministry of Education, University and Research (M.I.U.R.) with a grant covering the three-year period of her Ph.D. Between February 2014 and June 2014, she spent five months in Mons, Belgium, for a research internship at the Department of Mathematics and Operational Research, Faculté polytechnique, Université de Mons, supervised by Prof. Nicolas Gillis.

In 2008 she got a B.Sc. in Computer Science from the University of Bari, and in 2011 she got a M.Sc. in Computer Science from the same University.

She is active in the computer science community as reviewers for international journals.

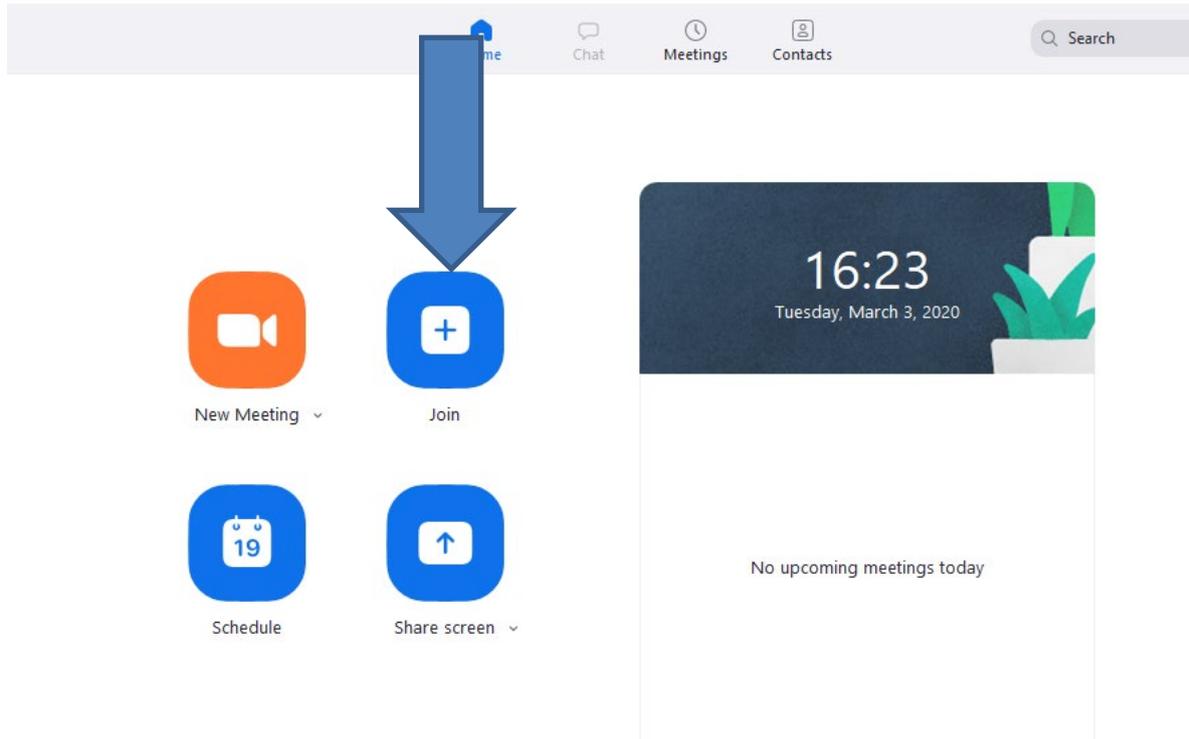
She is in the editorial board of the international journal Electronics as topic editor, and she is guest editor of a special issue in the same journal.

Parallel to her research activity, she carries out teaching activities in the Informatics Department of University of Bari.

Instructions for the Online Tool “ZOOM”

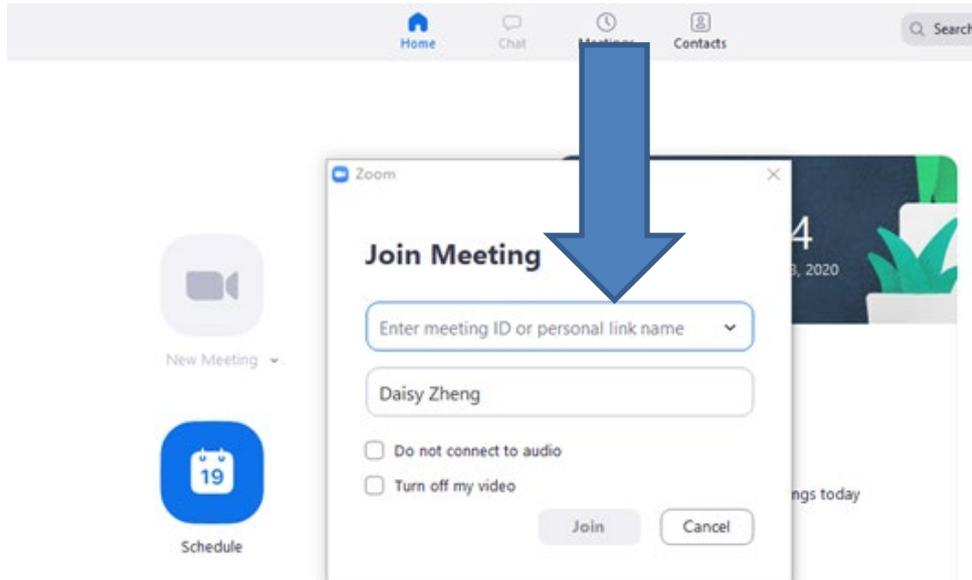
1. You can download the software “Zoom” from this URL:
<http://www.zoom.us/>

2. How to join online conference in Zoom



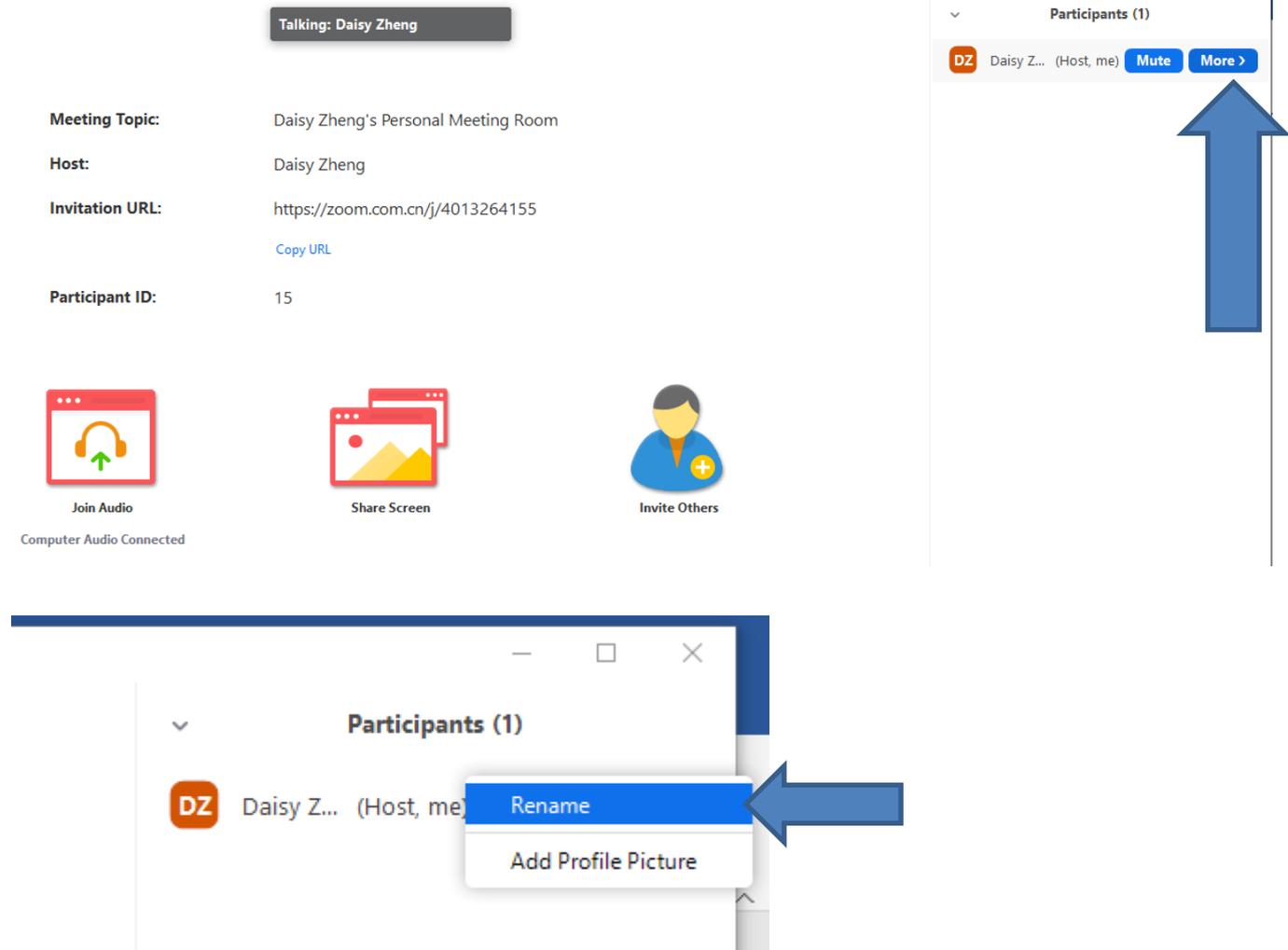
Please click **“join”**

3. Fill in the Conference ID



Please fill in the **Online-Meeting Room ID: 690-3446-0497 (Password: 198935)** and join the online conference

4. How to rename



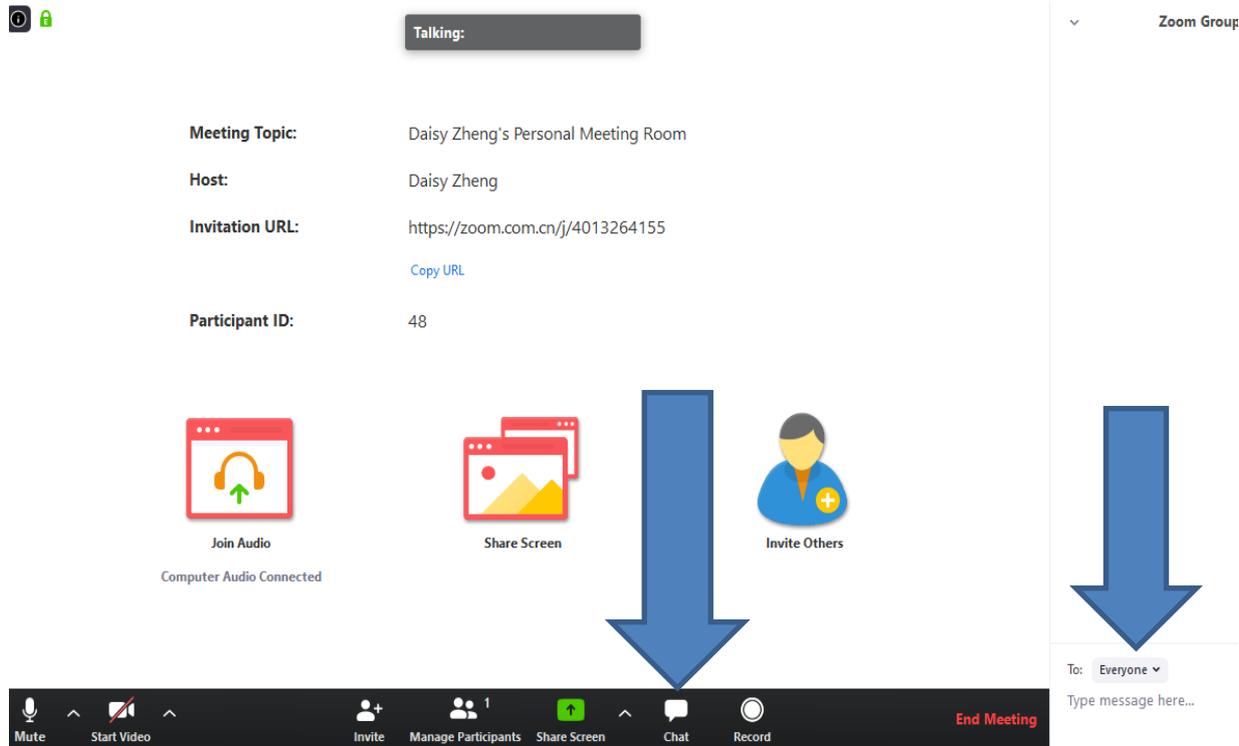
First, you can click **“More”**

Then, you can click **“rename”**.

If you are **presenter**, your name need to be renamed as **BEXXXX (your paper id)+ your name**.

If you are **listener**, your name need to be renamed as **listener + your name**.

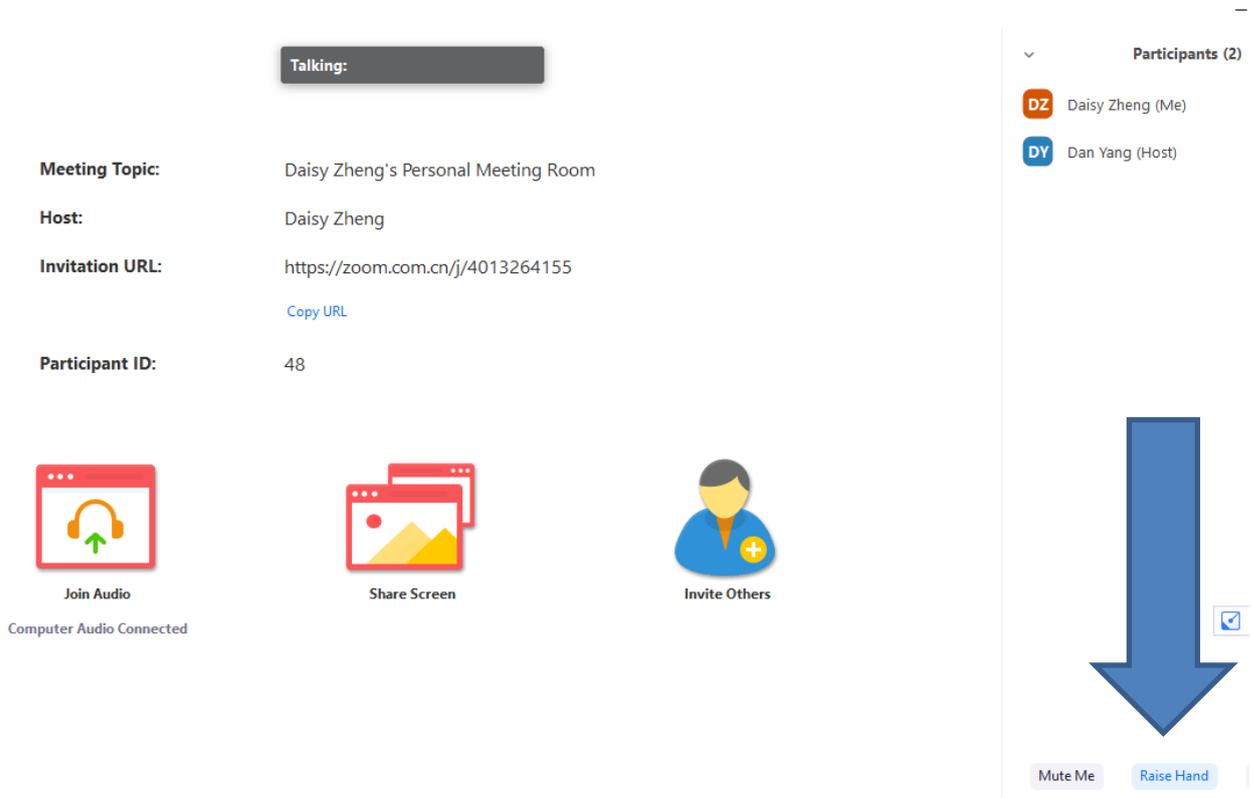
5. How to Chat with Others in Zoom:



You can click **“Chat”** first.

Then, you can click **“everyone”** to choose who you want to talk with.

6. How to Use Raise Your Hands and Ask Questions in Zoom:



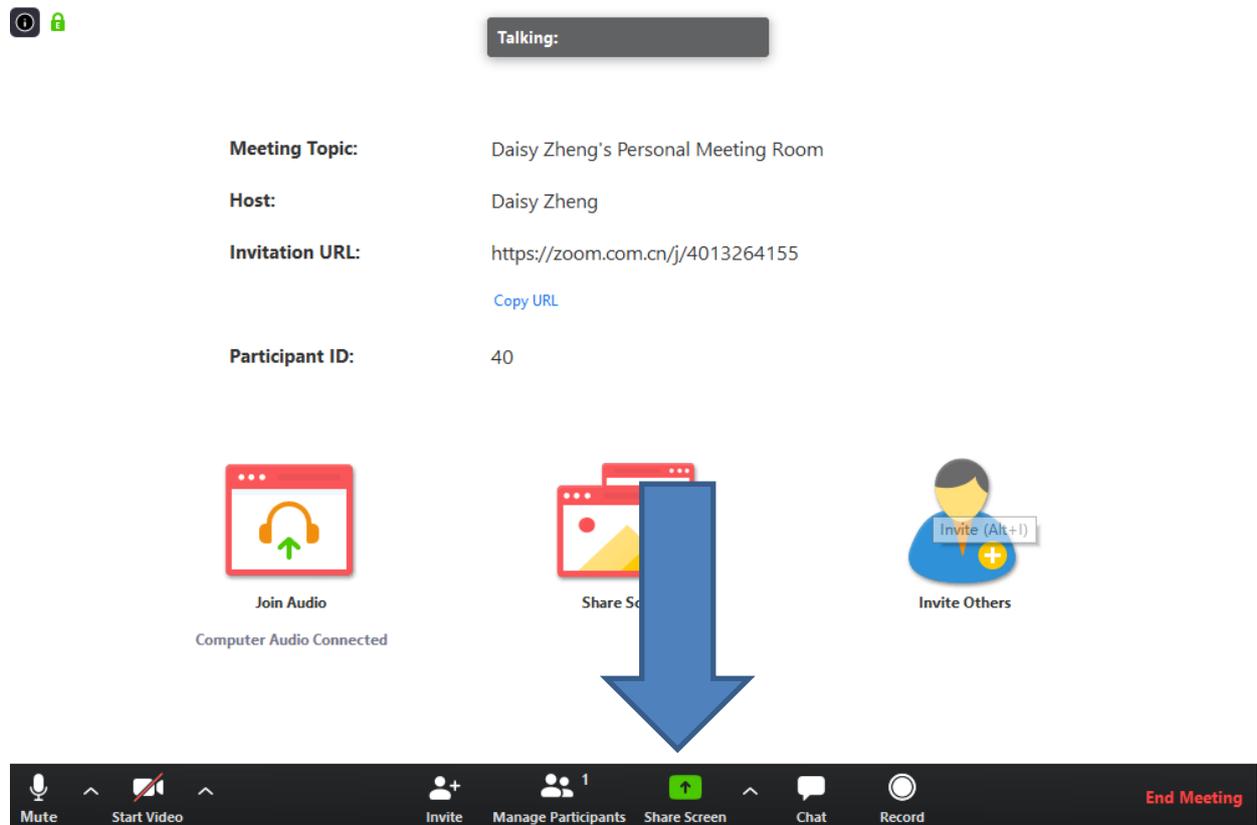
If you have any problems during the conference, you can click “**raise your hands**” or use “**chat**” to communicate with the conference secretary and the conference secretary will help you.

When you have questions about keynote speeches, you can also use “**raise your hands**” function.

After the keynote speech, keynote speakers will answer your questions.

7. How to Share Your Screen

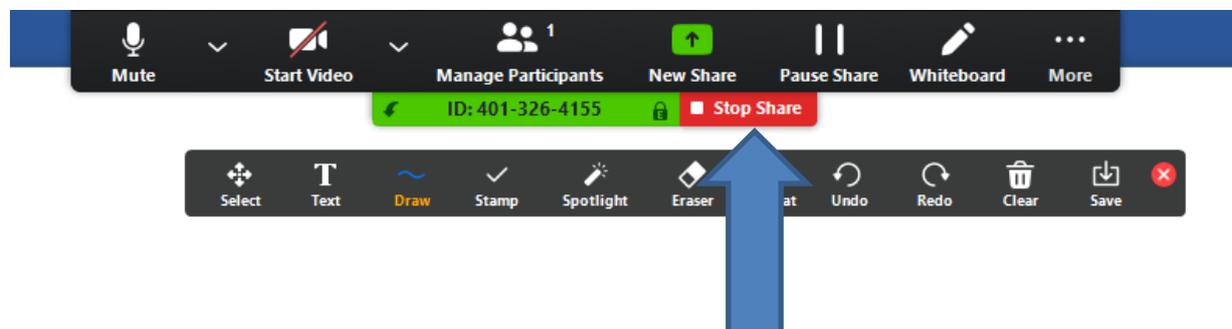
Zoom Meeting ID: 401-326-4155



When you do your presentation, you need to share your screen.

You can click “**share screen**”.

8. How to stop sharing your screen



After your oral presentation, you can click “**stop share**”.



The 2020 2nd International Conference on Big Data Engineering

Presentation Instructions

Instructions for Oral Presentations

Materials Provided by the Presenters:

PowerPoint or PDF Files

Duration of each Presentation (Tentatively):

Regular Oral Presentation: about **10 Minutes** of Presentation and **5 Minutes** of Question and Answer.

Best Presentation Award

One Best Presentation will be selected from each presentation session, and the Certificate for Best Presentation will be sent after the conference.



Schedule for Conference

Tips: The time in the schedule is according to local time in China.

*Online test is for testing the Internet connection and helping participants get familiar with software Zoom.

Please make sure that you will attend online test.

May 29, 2020 (Friday) Room ID: 690-3446-0497 (Password: 198935)		
Online Test		
Morning Test (9:00-11:30)	9:00-9:30 (21:00-21:30 May 28 in USA)	Test for Prof. Yao Liang
	9:30-10:00	Test for Prof. Kai Hwang
	10:00-10:30	Test for Prof. Changxu Wu
	10:30-10:50	Test for Dr. Wei Li
	10:50-11:10	Test for Dr. Ka-Chun Wong
	11:10-11:30	Test for Dr. Gabriella Casalino
Afternoon Test (13:30- 17:30)	13:30- 14:30	Test for Presenters of Session 1 (BE0007, BE0015, BE0016, BE0017, BE0024, BE0030, BE5001)
	14:30- 15:30	Test for Presenters of Session 2 (BE0004, BE0008, BE0023, BE0034, BE5004, BE0025)
	15: 30-16:30	Test for Presenters of Session 3 (BE0013, BE0019, BE0020, BE0021, BE0022, BE0028, BE5002)
	flex time	Any questions about the test, please contact the staff at other time on May 29 Afternoon before 17:30.
May 30, 2020 (Saturday) Room ID: 690-3446-0497 (Password: 198935)		
Morning Session for Keynote Speeches		
9:00- 11:40 KN Speeches (China Time)	9:00- 9:10	Opening Remarks
	9:10-9:55	Keynote Speech I Title: Cloud, AIoT and Edge Computing in 5G Mobile Core Environment Prof. Kai Hwang, Chinese University of Hong Kong, Shenzhen, China



	9:55-10:40 (21:55-22:40 May 29 in USA)	Keynote Speech II Title: Multiresolution Learning: A 20-Year Perspective Prof. Yao Liang, Purdue University School of Science & Indiana University Purdue University, USA
	10:40-10:55	Take a break
	10:55-11:40	Keynote Speech III Title: Big Data and Human-AI System Interaction Prof. Changxu Wu, University of Arizona, USA
Invited Speech I	11:40-12:05	Invited Speech I Title: Disruptive AI Technologies for Molecular Biology and Medicine: DNA Motifs, CRISPR-Cas9 Off-Targets, and Cancer Screening from Blood Dr Ka-Chun Wong, City University of Hong Kong, Hong Kong
Lunch time: 12:05- 15:00		
Afternoon Session: Invited speeches and Session 1		
15:00- 15:50 Invited Speech II and III (China Time)	15:00-15:25	Invited Speech II Title: Big Data: Algorithms and Ecosystems Dr. Wei Li, Central Queensland University (CQU), Australia
	15:25-15:50	Invited Speech III Title: Dynamic Incremental Semi-Supervised Fuzzy Clustering for Data Stream Classification Dr. Gabriella Casalino, University of Bari, Italy
Break Time: 15:50- 16:05		
16:05- 17:50	Session 1 (BE0007, BE0015, BE0016, BE0017, BE0024, BE0030, BE5001)	

May 31, 2020 (Sunday) Room ID: 690-3446-0497 (Password: 198935)

Morning Session: Session 2		
Morning Session	10:00- 11:30	Session 2 (BE0004, BE0008, BE0023, BE0034, BE5004, BE0025)
Lunch time: 11:30- 14:30		
Afternoon Session: Session 3		
Afternoon Session	14:30- 15:45	Session 3 (BE0013, BE0019, BE0020, BE0021, BE0022, BE0028, BE5002)



The 2020 2nd International Conference on Big Data Engineering

Morning Session

Morning, May 30, 2020 (Saturday)

Time: 9:00-12:05

Online-Meeting Room ID: 690-3446-0497 (Password: 198935)

Opening Remarks & Testing (9:00-9:10)

Addressed by Chair Prof. Changxu Wu from University of Arizona, USA

Keynote Speech I (9:10-9:55)

Title: Cloud, AIoT and Edge Computing in 5G Mobile Core Environment

Prof. Kai Hwang

Chinese University of Hong Kong, China

Abstract—In this talk, Prof. Hwang will assess recent advances in big data, Internet clouds, Artificial Intelligence (AI), Internet of Things (IoT) and edge computing. The impacts of 5G mobile core networks, Inspur AI stations, Beidou satellite system, and Huawei 5G network slicing techniques are evaluated for specific industrial and societal applications. Dr. Hwang will report the very first 5G cloud/AIoT research platform ever built on a university campus in China.

This research cloud was designed and constructed at the Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS) on the campus of CUHK in Shenzhen. This AIRS cloud system is a large-scale AI-oriented cloud supported by 5G edge core. The system was specifically designed for big data analytics, AIoT (for AI+IoT), machine learning, and industrial robotics applications. Initially, we aim at serving the Greater Bay area of the Pearl River.

Keynote Speech II (9:55-10:40)

Title: Multiresolution Learning: A 20-Year Perspective

Prof. Yao Liang

Purdue University School of Science & Indiana University Purdue University, USA

Abstract—In machine learning field, two crucial challenges for modeling are the generalization and robustness. Multiresolution learning paradigm has been introduced more than 20 years ago as a systematic approach to improve the generalization and robustness of neural network modeling. In this talk, I will present and review the original idea and work of multiresolution learning, its continuing development and extension, and its applications. I will provide insights on why the multiresolution learning offers a novel and systematic paradigm for constructing predictive modeling with significantly improved generalization performance and robustness for both regression and classification problems. We show that the introduced paradigm is very general and can be applied to tasks with either signal-based input space or feature-based input space. Applications in various domains are given to illustrate the underlying idea of the multiresolution learning and to demonstrate its superior performance in models' generalization and robustness, especially for very difficult tasks.



Take a Break: 10:40-10:55

Keynote Speech III (10:55-11:40)

Title: Big Data and Human-AI System Interaction

Prof. Changxu Wu

University of Arizona, USA

Abstract— Prof. Changxu Wu will introduce the new research in big data of digital world human-AI system interaction in this talk. Giving the rapid development of big data, 5G, internet of things (IoT), and other related technology, there are lots of big-data based AI technology can interact with human. Dr. Wu will introduce the recent research in human-AI interaction, including example AI systems e.g., autonomous vehicles, V2V/V2I intelligent transportation systems, and dialogue systems, and how they interact with human. Dr. Wu will also talk about the future of human-AI interaction.

Invited Speech I (11:40-12:05)

Title: Disruptive AI Technologies for Molecular Biology and Medicine: DNA Motifs, CRISPR-Cas9 Off-Targets, and Cancer Screening from Blood

Dr Ka-Chun Wong

City University of Hong Kong, Hong Kong

Abstract—In this talk, I will present my research group contributions in bioinformatics and health informatics in recent years. In particular, the unconventional and disruptive AI technologies are focused.

Firstly, the DNA binding of transcription factors is central to gene regulation and stem cell development. The DNA binding pattern (i.e. DNA motif) elucidation of transcription factors forms the basis for downstream research. Therefore, I will present our breakthroughs in elucidating DNA binding patterns from the protein-coding sequences of transcription factors using AI as well as our synthetic biology approach to synthesize a heterodimeric DNA motif from two monomeric DNA motifs. A DNA motif published on Nature has been rescued.

Secondly, CRISPR-Cas9 is the predominant tool for gene editing and raised substantial concerns on its clinical implications. To avoid any side effect, its off-target predictions are fundamentally essential. I will present our recent work in predicting CRISPR-Cas9 off-targets using deep learning, the latest AI technology.

Finally, I will present our very recent work in screening cancers from blood. I will demonstrate how our proposed AI approach (CancerA1DE) can outperform the existing approach (CancerSEEK) proposed in John Hopkins University. In particular, our approach can double the existing sensitivity from 38% to 77% for the earliest cancer detection (i.e., Stage I) at the 99% specificity level.



Lunch 12:05-15:00



The 2020 2nd International Conference on Big Data Engineering

Afternoon Session

May 30, 2020 (Saturday)

Time: 15:00 -17:50

Online-Meeting Room ID: 690-3446-0497 (Password: 198935)

Invited Speech II (15:00-15:25)

Title: Big Data: Algorithms and Ecosystems

Dr Wei Li

Central Queensland University (CQU), Australia

Abstract—The goal of big data processing is centered in quality results and fulfilment of the constraints on computing resources. This talk is about the transformation of traditional algorithms to fit for big data processing. The transformation is considered for distributed features and resources constraints of big data processing and quality of results. The big data ecosystems are introduced in terms of batch processing and stream processing. A case study is presented throughout this talk to demonstrate a typical big data algorithm, the integration of big data systems to apply the algorithm and programming of the algorithm to fit for distributed environments.

Invited Speech III (15:25-15:50)

Title: Dynamic Incremental Semi-Supervised Fuzzy Clustering for Data Stream Classification

Dr Gabriella Casalino

University of Bari Aldo Moro, Italy

Abstract—Data stream mining is an active research area that has recently emerged to discover knowledge from large amount of continuously generated data. Data streams are a manifestation of Big Data, which are characterized by the four ‘V’ dimensions, namely Volume, Velocity, Variety and Veracity. Particularly, the term “data stream” defines a sequence of data that arrives in a continuous and changing manner. Managing and processing data streams raises new challenges because it is not feasible to store the arriving data in a traditional database management system and needs to be processed in an online manner. The intrinsic nature of stream data requires the development of algorithms capable of performing fast and incremental processing of data objects, suitably addressing time and memory limitations. For this purpose a new algorithm called Dynamic Incremental Semi-Supervised Fuzzy C-Means is proposed [1]. The algorithm incrementally generates informational patterns, which represent a synthesized view of all data records analyzed in past and progressively evolves as new data records are available. The algorithm is able to learn a model from semi-supervised data since it is based on Semi-Supervised Fuzzy C-Means algorithm (SSFCM) by Pedrycz; in fact, labeled samples may be difficult or expensive to obtain in some real-world scenarios, meanwhile unlabeled data are relatively easy to collect. Moreover, the fuzzy representation of data relationships enhances patterns discovery in them. The algorithm has been shown to be effective to solve problems from different domains, such as students’ assessments prediction, credit card fault detection, bipolar disorder classification and occupancy detection.



Take a Break: 15:50- 16:05



The 2020 2nd International Conference on Big Data Engineering

Oral Presentation Abstracts

Session 1

Tips: The schedule for each presentation is for reference only. In order not to miss your presentation, we strongly suggest that you attend the whole session.

Afternoon, May 30, 2020 (Saturday)

Time: 16:05- 17:50

Online-Meeting Room ID: 690-3446-0497 (Password: 198935)

Session Chair: Dr Gabriella Casalino

BE0007 (16:05-16:20)

SiteResearch on Spatiotemporal Behavior Changes of Pedestrians Based on Intelligent Image Analysis Data
——A Case Study of Shanghai Binjiang Green Space

LI Liang and WANG Xin

1: Shanghai University of Engineering Science, China

2:Tongji University, China

Abstract— Nowadays, in high-density cities, the full and reasonable use of public space has become an important research topic. This paper applies the population heat data analysis and intelligent image analysis technology to investigate the pedestrian activities in Binjiang Green Space of Xuhui in Shanghai by interviewing, coming to the conclusion of spatiotemporal behavior changes characteristics of pedestrians in Binjiang Green Space. At the same time, some adaptive suggestions are put forward for the planning and management of Binjiang Green Space.

BE0015 (16:20-16:35)

Name-Nationality Classification Technology under Keras Deep Learning

Yu Kang

Chengdu University of Information Technology, China

Abstract— To improve the classification efficiency of personnel nationality information, the classification method of personnel nationality information is explored. First, a new classification method is proposed using the Keras model and deep learning theory. Two methods based on support vector machine (SVM) and convolutional neural network classification are proposed. (1) The personal name is input and a set of numbers corresponding to the positions in the alphabet are output orderly. (2) The personal name is input and the number output relies on the number of occurrences of each character of a name, regardless of order. Second, for the problem that the classification accuracy of nationality information by name is not high, the adaptive moment estimation (Adam) algorithm is used to optimize it. Finally, to prove the reliability of the proposed methods, these methods are used to verify the nationality information of Olympic personnel. The results show that comparing the two methods, the classification method that relies on the number of occurrences of characters in the name gets good grades, and the ultimate average score is 90.35. The score of the first method is only 79.34. Through this investigation, it is found that the second method proposed can effectively use the name of the person to determine the nationality information. Applying it in real life can improve the classification efficiency of personnel nationality information.

BE0016 (16:35-16:50)

Adaptive Weighted Loss Makes Brain Tumors Segmentation More Accurate in 3D MRI Volume

Xukun Zhang, Wenxin Hu and Jianhua Cai

East China Normal University, China

Abstract— Accurately segmenting brain tumors in Magnetic Resonance Imaging (MRI) volume can benefit the diagnosis, monitoring, and surgery planning of the disease. Manual delineation practices require anatomical knowledge, are expensive, time consuming and can be inaccurate due to human error. In the data era, machines with learning capabilities can achieve automatic brain tumor segmentation in MRI volume with promising performance on large region. However, it is not very effective for tumor segmentation on small region. Although more and more methods use pixel-level loss to guide algorithms to pay more attention to accurate segmentation of small regions, this problem still exists. In this paper, we propose an adaptive weighted loss, which can automatically adjust the proportion of loss generated by different region segmentation, thereby making small region segmentation more accurate. We added the adaptive weighted loss to a 3D MRI brain tumor segmentation network using auto-encoder regularization (3D-AE), and performed extensive validation on the MICCAI Brain Tumor Segmentation Challenge 2018 dataset (BRATS 2018). The achieved dice score is 0.769 for core tumor, 0.904 for the whole tumor and 0.887 for enhanced tumor. The overall results show better performance than the state-of-the-art in both dice score and precision on BRATS 2018.

BE0017 (16:50-17:05)

Demographic Prediction Based on User's Movie-Watching Behavior

Xueying Liu, Wenxin Hu and Wei Peng

East China Normal University, China

Abstract—Demographic attributes play an important role in adjusting the market strategy of film companies and customizing recommendation for their user. However, it's difficult to obtain full information of demographic attributes of online user and previous work only adapt single-source data (user historical record) for prediction, which is not enough for user representation. In this paper, we proposed a novel approach named CNN-ETNA to predict user demographic attributes from their movie-watching behavior. Our model CNN-ETNA is further improved by extracting semantic feature from movie to learn informative user representations. Experiments conducted on MovieLens dataset demonstrate that our approach effectively improves the performance of demographic prediction compared to baseline algorithms.

BE0024 (17:05-17:20)

Using 3D Convolutional Networks with Shortcut Connections for Improved Lung Nodules Classification

Zhifeng Lin, Jun Zheng and Wenxin Hu

East China Normal University, China

Abstract— Lung cancer is the leading cause of cancer-related death worldwide. Due to the difficulty of artificial extraction of medical image features and the development of artificial intelligence in the field of medical image, various deep learning methods for lung nodules classification have been proposed to help doctors diagnose and detect lung cancer in the early stage. The traditional 2D CNN cannot make use of the 3D spatial characteristics of CT data, while the 3D CNN has many parameters, which leads to low model efficiency. Residual Network (ResNet) is a residual structure using skip connection that makes deep classification network easier to train. Therefore, motivated by the work of 3D CNN and ResNet, in this paper, a VGG based 3D residual connection network, called VGG+ResCon, is proposed to mine the vertical information of tumor CT images and accelerate the training efficiency of the model. Besides, after enhancing the dataset, focal loss is used to replace the traditional cross-entropy as the loss function, which solves the problem of uneven distribution of positive and negative samples of medical data (more negative samples than positive samples). And it also makes the model more focused on difficult-to-classify samples. This methodology was evaluated on the LUNG Nodule Analysis 2016 (LUNA16) dataset, with the best precision of 93.62%, recall of 92.48%, specificity of 96.83% and f1-score of 93.04%. Experimental results demonstrate the effectiveness of the proposed method in classifying malignant and benign pulmonary nodules.

BE0030 (17:20-17:35)

Rubbing Image Retrieval Using Deep Convolutional Neural Network

Ziyang Wang, Youguang Chen, Xuanqi Wu and Peng Ren

East China Normal University, China

Abstract— With the widespread use of digital imaging data, the size of image collections about Bronze Inscriptions rubbing is increasing rapidly. It becomes difficult to manage and query a specific image from these large databases, which motivates the need for image retrieval. In this paper, we proposed an effective content-based rubbing image retrieval (CBRIR) framework based on deep convolutional neural network (DCNN) for Bronze Inscriptions rubbing. Specifically, we extract discriminative local features for image retrieval using the activations of convolutional neural networks. We use cosine metric, Euclidean metric, and Hamming metric to measure similarity in the CBRIR framework. Experimental results show that our framework has an excellent accuracy of rubbing retrieval.

BE5001 (17:35-17:50)

TIP: (17:35-17:50 China time/10:35-10:50 UK time)

Designing Interfaces for Creative Learning Environments using the Transreality Storyboarding Framework

Jazz Rasool¹, Judith Molka-Danielsen² and Carl H. Smith¹

1. Ravensbourne University London, UK
2. Molde University College, Norway

Abstract — Modern creatives, producing different media within multiple mediums, could benefit from transmedia-based methodologies that ensure core standards in effective interface design and storyboarding adopted in the production of creative works. This is relevant for students learning creative disciplines and professionals prototyping briefs within the creative industries. Students and professionals must learn to creatively fulfil a brief, express a story or narrative in a transmedia experience space that employs three components that work with different realities, senses and interactions. An approach addressing these components is formalised by the authors into the Transreality Storyboarding Framework (TSF). The first component maps eight realities that students must learn to tell stories across, the Eight Realities Design Methodology (ERDM). The second component, (PAIRS), focuses on the design of user interfaces and experiences (UI/UX) connecting people into creative works, exploring the inclusion of Passive, Active, Interactive, Responsive and Shared interactions. The third component is Contextology, a methodology for exploring how different kinds of senses engage with the narrative within those creative realities. Case studies in Architecture, Art & Design, Engineering, Training as well as Fashion are reviewed showing how workflows used aspects of these approaches for effective storytelling and engaging creative and educational experiences.

Oral Presentation Abstracts

Session 2

Tips: The schedule for each presentation is for reference only. In order not to miss your presentation, we strongly suggest that you attend the whole session.

Afternoon, May 31, 2020 (Sunday)

Time: 10:00- 11:30

Online-Meeting Room ID: 690-3446-0497 (Password: 198935)

Session Chair: Prof. Yao Liang

BE0004 (10:00- 10:15)

A Random Forest Regression Model Predicting the Winners of Summer Olympic Events

Mengjie Jia¹, Yue Zhao¹, Furong Chang¹, Bofeng Zhang¹ and Kenji Yoshigoe²

1: Shanghai University, China;

2: Toyo University, Japan

Abstract—From the past Olympic medal lists, we can find that the number of medals of China has been increasing steadily in recent years while we also observe that some countries always occupy the top positions of the Olympic medal list, such as the United States, Britain and Germany. In this work we take the data of the medal lists from the 18th to 31st Summer Olympic Games as a sample and selects GDP, the population, the size of national team and the home advantage as the characteristic parameters to build a random forest regression model to predict the number of medals. The FP-growth algorithm is used to analyze the association rules of the data. And the winners of some events in the 2020 Tokyo Olympic Games are predicted.

BE0008 (10:15- 10:30)

A Fake News Detection Framework Using Social User Graph

Yi Xie¹, Xixuan Huang¹, Xiaoxuan Xie¹ and Shengyi Jiang²

1:Guangdong University of Foreign Studies,China

2: Eastern Language Processing Center, China

Abstract—In the contemporary era of information explosion, all kinds of false and misleading information are flooding our lives, which often impact politics and even our real world. How to detect these potential fake news from massive news texts is of great significance for maintaining social stability. The current fake news detection is mainly researched from supervised classification based on text content. However, it does not always work. Because fake news is often latent, they can imitate true news from the perspective of language style, etc. Therefore, we need to capture more effective feature information to improve the performance of fake news detection models. For example, the user connections on social networks. At present, some researches have used the characteristics of users in social networks, but the user's characteristic representation is often independent of the task of fake news detection, and some researches ignored the different influences of different users in fake news detection, such as the news released by official news agency is often more authentic. In this paper, we propose a fake news detection framework that makes full use of user characteristics. Firstly, the user's news dissemination behavior was classified based on GraphSage. Then, using multi-head attention to give the user representation that obtained through GraphSage to different weights, which was combined with the news text representation was used for classification. Compared with previous studies, we have achieved the state-of-the-art performance on two real datasets BuzzFeed and PolitiFact, which show the strength of our method.

BE0023 (10:30- 10:45)

A BERT-based Ensemble Model for Chinese News Topic Prediction

Jingang Liu¹, Chunhe Xia¹, Xiaojian Li², Haihua Yan¹ and Tengting Liu¹

1: Beihang University, China

2: Guangxi Normal University, China

Abstract—With the rapid development of big data mining technology in the Chinese commercial field, the news topic prediction becomes increasingly important. Since the accuracy of Chinese news topic classification can directly affect the personalized recommendation effect of the Chinese news system and then affect business profits, the news category prediction performance needs to be higher as possible. With the great success of the BERT model in the past two years, using the BERT model alone has achieved extremely good performance on Chinese text classification tasks. Therefore, using the advantages of the BERT to study more effective methods for the Chinese news classification will become more meaningful. In this paper, we propose a model that combines the advantages of both BERT and the long short-term memory (LSTM) network, named BERT ensemble LSTM-BERT(BERT-LB). Our method is more effective than using BERT alone. This model uses a three-step method to calculate and integrate Chinese news text features. Besides, we use two datasets to evaluate our method and other baseline methods. We demonstrate that the proposed method has the promising ability to predict Chinese news topics and prove its generalization ability.

BE0034 (10:45- 11:00)

A New Fusion Method on Machine Reading Comprehension

Wenfang Tang, Hong Jiang and Kejia Xu

East China Normal University, China

Abstract— Machine reading comprehension, as one of the core tasks of natural language processing, has been a research focus in this field. This paper is mainly aimed at Dureader, a Chinese-oriented multi-document reading comprehension task dataset, and an end-to-end machine reading comprehension model based on neural network is proposed. In this paper, a combination of context-to-question attention and self-attention, and a new fusion method are used to fuse context and attention information. The comparison of experimental results shows that compared with the baseline model, by our methods, BLEU-4 index has been improved by 6% and ROUGE-L index has been improved by 3%.

BE5004 (11:00- 11:15)

System Design of Safety Road Network in Urban Morphology Prevention During COVID-19 Based on Digital Simulation Technology

Niu Zhiming and Qiu Song

Tsinghua University, China

Abstract — In December 2019, Wuhan City, Hubei Province, China first discovered a new type of unexplained pneumonia case, which began to spread rapidly in the urban area of Wuhan and Hubei Province within a short period of time. The Chinese government has taken active anti-epidemic measures. With the effect of the Wuhan closure of cities, the number of new cases in the country has been significantly reduced. However, the epidemic situation in foreign cities has just begun, and the secondary infection has become the main way of spreading COVID-19 in other cities. How to curb the secondary infection of the urban epidemic situation has become an important target for epidemic prevention.

In this paper, the digital platform Rhino & Grasshopper algorithm is used to simulate the possible flow trajectory of urban people, and the most widely used virus transmission models such as SEIR are visually combined, and a new M-SEIR model method for controlling urban epidemic spread is proposed. By exploring the relationship between the form of urban personnel activity flow and the development of the epidemic situation, it can provide a reference for the design of public transportation trips and the prevention and control of epidemic situations in urban morphology.

BE0025 (11:15- 11:30)

A Framework for Arabic Tweets Multi-label Classification Using Word Embedding and Neural Networks Algorithms

Abdullah M. Bdeir and Farid Ibrahim

Abu Dhabi University, UAE

Abstract—The need for classifying tweets is essential for many people like tourists, tourism companies and governments. In this paper, we propose a framework for Arabic Tweets multi-label classification using word embedding technique and deep learning algorithms. We built our dataset using 160k Arabic tweets gathered from Twitter. We compared two deep learning methods, Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). Our results show that it is possible to classify tweets using our methodology without any significant difference in results of accuracy scores and hamming loss for both types of networks. The accuracy scores and hamming loss were nearly 90% and 0.02, respectively.



Lunch: 11:30- 14:30



The 2020 2nd International Conference on Big Data Engineering

Oral Presentation Abstracts

Session 3

Tips: The schedule for each presentation is for reference only. In order not to miss your presentation, we strongly suggest that you attend the whole session.

Afternoon, May 31, 2020 (Sunday)

Time: 14:30- 16:15

Online-Meeting Room ID: 690-3446-0497 (Password: 198935)

Session Chair: Prof. Changxu Wu

BE0013 (14:30- 14:45)

Mining High Utility Co-location Patterns Using the Maximum Clique and the Subsume Index

Wei Song and **Qian Qiao**

North China University of Technology, China

Abstract— Mining high utility co-location patterns (HUCP) is a promising technique in spatial data mining because it treats different features with different levels of importance. Existing HUCP mining (HUCPM) algorithms are based on row instances and table instances, leading to high computational cost. To overcome this problem, an HUCPM algorithm based on the subsume index (HUCPM-SI) is developed using a typical high utility itemset mining model. Maximal cliques are first discovered, enabling the spatial database to be transformed into a clique transaction database. The subsume index is then used to mine HUCPs. Tests conducted on synthetic and real datasets demonstrate the advantages of the HUCPM-SI algorithm.

BE0019 (14:45- 15:00)

A Multidimensional Approach of Evaluating Developers

Changqiang Zhang and Ming Chen

National University of Defense Technology, China

Abstract— In this paper, we propose an approach to assess the ability of developers based on their behavior data from OSS. Specifically, we classify developers' ability into code ability, project management ability, and social ability. Code efficiency is related to the developer's commit record and the pull-request record. The developer's project management ability is achieved by tracking the developer's commit record. We use regular matching to map the commit behavior to the project management behavior and calculate the developer's project management ability according to the proportion of different behaviors. The social ability of developers is related to the data that developers interact with in the open-source community. We dug for developer reviews on commit, issue, and gist fragments. By calculating the proportion of positive emotions in developer reviews and the proportion of developers interacting with others in the reviews, the social ability of developers is obtained. We get behavioral data from 50 random developers. Twitter's data is used to test the effect of different machine learning algorithms on the accuracy of developer comment polarity judgments. It is found that the combination of SVM, xgboost and random forest have the highest prediction accuracy. Finally, we select 5 students to use Likert scale to score the results. Our score shows that the results are basically in line with expectations.

BE0020 (15:00- 15:15)

Predicting Chinese Bond Market Turbulences: Attention-BiLSTM Based Early Warning System

Peiwan Wang¹, Lu Zong¹ and Yurun Yang²

1: Xi'an Jiaotong-Liverpool University, China

2: Shanghai Clearing House, China

Abstract— The study aims to construct an effective early warning system (EWS) to predict the crisis triggered turbulence in Chinese bond market by integrating the volatility regime switching model, SWARCH, to improve the crisis classifying precision, and the stylized predictive model, Attention-BiLSTM of attention mechanism based deep neural networks, to resolve the predicting hysteresis. The model versatility and comparability are investigated and testified by applying multiple prominent EWS models to bonds with different credit rating levels. The hybrid EWS also specifies the leading factors relating to the bond credit rating, that will practically instruct governors and market participants to focus on either the national economy associated or the corporate finance concerned factors according to the bond varying credit risks to make more effective predictions.

BE0021 (15:15- 15:30)

Overview of Blockchain Consensus Mechanism

Changqiang Zhang, Cangshuai Wu and Xinyi Wang

National University of Defense Technology, China

Abstract—The advent of the Bitcoin system has brought another boom in the Internet era. In a very short time, many Blockchain systems come into being successively, whose decentralization, consensus mechanisms, intelligent contract, and other characteristics make them applicable to various fields such as finance, education, medical, technology, etc. The consensus mechanism is the core of Blockchain technology. And a good consensus mechanism plays a very important role in the stable operation of the Blockchain system. The continuous improvement of consensus mechanisms such as PoW (proof-of-work), PoS (proof-of-stake), DPoS (delegated-proof-of-stake), and PBFT (Practical Byzantine Fault Tolerance) has led to the evolution of Blockchain technology to Blockchain 3.0. Starting from the issue of Byzantine generals, this article analyzes common consensus mechanisms based on existing Blockchain applications and then evaluates their consistency. This paper has a unique perspective, selection and comment on the latest progress and important literature in the field of block chain. At last, it also has a high degree of summary and prospect to the existing problems and the future development trend.

BE0022 (15:30- 15:45)

Visualization Research of Internet Information Security in China in Recent 20 Years

Chunlei Liu¹ and Mi Shi²

1: Tianjin University, China

2: Tiangong University, China

Abstract—[Purpose / Significance] In the era of big data, Internet information security is the engine and motive force driving the steady progress of society. In order to speed up the theoretical innovation of Internet information security and promote the practical steps of network security governance, it is necessary to make a systematic and quantitative study of the research literature in this field in order to deeply analyze the hot topics and academic frontiers of this research direction.

[Method / Process] In this study, 1027 documents collected from core journals of Chinese CNKI database, CSSCI and CSCD source journals from 1997 to 2018 were used as data sources. CiteSpace 5.3.R4, a citation visualization software, was used to explore and map the scientific knowledge maps in the field of network information security in China based on time, space and knowledge base.

[Result/ Conclusion] The results show that the research hotspots focused on public opinion management, risk management, information dissemination and cyber violence in the context of the Internet information security in China. While the important academic frontier is the network risk assessment and large data closely related to the Internet.

BE0028 (15:45- 16:00)

Gene Big Data Analysis of Differentially Expressed lncRNA and miRNA in Liver Cancer with Different Gender

Jianzhi Deng¹, Yuehan Zhou² and Weixian Tang²

1: Guilin University of Technology, China

2: Guilin Medical University, China

Abstract—In this paper, we try to find the differentially expressed lncRNA and differentially expressed miRNA of Liver cancer, especially between the different gender's patients. The differentially expressed genes were screened from TCGA liver data. Based on the extracted differentially expressed lncRNAs and differentially expressed miRNAs, we reveal an 8-lncRNA (TTTY14, UCA1, LINC00162, TTTY10, XIST, ERVH48-1, ZFY-AS1 and TTTY15) to 3-miRNA (hsa-mir-506, hsa-mir-508, hsa-mir-205) regulatory network of 13 pairs inter-regulatory between male and female. The 8 differentially expressed lncRNAs in the lncRNA-miRNA regulatory network were analyzed by the multivariable COX regression model, and LINC00162 and TTTY10 were found as the co-expression differentially expressed lncRNAs. After survival kmplot analysis and receiver operating characteristic analysis of the co-expression differentially expressed lncRNAs, TTTY10 was selected and proved as the potential biomarker of liver cancer for the diagnose and therapy.

BE5002 (16:00- 16:15)

Mini-Program Design and Evaluation to Create Online-Merge-Offline Service for Retailers

Kuan-Hua Liu, Ke-Di Cao and Sheng-Ming Wang

National Taipei University of Technology, Taiwan

Abstract — As the new retail has come, a lot of retailers try to make a digital transformation. The new retail begins with the development of an online to offline (O2O) mechanism to extend their online e-commerce to offline retailers. Nowadays, leading new retail players have been actively investing in offline retailers. Furthermore, they also evolve around technology-driven sales and cost efficiency to create online merge offline (OMO) services. This study used the Buynow Shanghai flagship store as an example and presents how to design and evaluate mini-programs for creating new retail OMO services for retailers. This study begins with the focus group interview to the three main business departments of the Buynow Shanghai flagship shopping mall. Then, we implemented a quality function deployment (QFD) analysis by referring to the interview results. Thirdly, we proposed the main functions, which can create the OMO service, of the mini-program prototypes by referring to the development priority proposed in QFD analysis. Finally, we design the user interface of the mini-programs and edit a simulated operational video to provide comprehensive introduction of the mini-programs and implemented the system usability scale (SUS) test to sample users. The results of this study show that the mini-programs can enable offline retailers to create omni-channel to customers. They also simplify and save time for the consumer's shopping experience. Furthermore, it allows for better data collection and data analytics for big retailers. The future studies will focus on using Kano's model to evaluate the customers' satisfaction of the services provided by the mini-programs.

