2019 International Conference on Big Data Engineering (BDE 2019)

June 11-13, 2019

Hong Kong, China

Supported by

Indexed by

http://www.bde.net

Welcome Message from Organizing Committee

It is our great pleasure to invite you to join our international conferences - 2019 International Conference on Big Data Engineering (BDE 2019). This event will provide a unique opportunity for editors and authors to get together and share their latest research findings and results. We look forward to welcoming you at Hong Kong.

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
June 11, 2019

Table of Contents

Keynote Speakers Introductions

Keynote Speaker I

Prof. Kai Hwang
Chinese University of Hong Kong, China

Prof. Kai Hwang is presently a Presidential Chair Professor in Computer Science and Engineering at the Chinese University of Hong Kong (CUHK), Shenzhen, China. He also serves as a Chief Scientist at the Cloud Computing Center, Chinese Academy of Sciences. He has taught at the University of Southern California and at Purdue University for many years prior joining CUHK. He received the Ph.D. in Electrical Engineering and Computer Science from UC Berkeley in 1972. Dr. Hwang has published extensively in the fields of computer architecture, parallel processing, cloud computing, and network security.

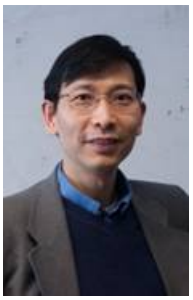
By 2018, his Google Scholar citation has exceeded 17,828 with an h-index of 55. His latest two books appeared in 2017: Cloud Computing for Machine Learning and Cognitive Applications (The MIT Press) and Big Data Analytics for Cloud/IoT and Cognitive Computing (Wiley, U.K.). The Chinese editions are both available from Huazhang Press in Beijing. An IEEE Life Fellow, he received the Outstanding Achievement Award from the Computer Federation of China in 2004. He received the Lifetime Achievement Award from the IEEE CloudCom 2012 for his pioneering work in parallel computing and distributed systems. He has delivered numerous keynotes or distinguished lectures in many Conferences or research centers. Dr. Hwang has performed consulting work with IBM, MIT Lincoln Lab, Chinese Academy of Sciences, and INRIA in France. He has served as a visiting chair professor at Tsinghua University and Hong Kong University.

Keynote Speaker II

Prof. Scott A. Klasky
Oak Ridge National Laboratory, USA

Prof. Scott A. Klasky is a distinguished scientist and the group leader for Scientific Data in the Computer Science and Mathematics Division at the Oak Ridge National Laboratory. He holds an appointment at the University of Tennessee, and Georgia Institute of technology. He obtained his Ph.D. in Physics from the University of Texas at Austin (1994), specializing in general relativity. Dr. Klasky is a world expert in scientific computing and scientific data management, co- authoring over 200 papers, and a PI on 8 multi-million dollar projects in the department of energy.

Keynote Speaker III



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 IV

Prof. Changxu (Sean) Wu
University of Arizona, USA

Dr. Changxu (Sean) 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).

Conference Introductions

Welcome to 2019 BDE Hongkong 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: 978-1-4503-6091-3), 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

Conference Venue

Conference Room I – II at 1/F, Regal Oriental Hotel

Add: 30-38 Sa Po Road, Kowloon City, Hong Kong

Regal Oriental Hotel is the only full-service hotel located in the heart of Hong Kong's heritage district Kowloon City, famous for authentic Hong Kong-style cuisine, and neighbouring the world-class Kai Tak Cruise Terminal. The hotel is easily accessible to all major transport links, shopping arcades, restaurants and tourist attractions, including Kowloon City Plaza, Festival Walk, MegaBox and the Ladies' Market in Mong Kok, and is within walking distance of the historic Kowloon Walled City Park.

Registration Guide

June 11, 2019 (Tuesday)

Time: 10:00~17:00

Venue: Regal Oriental Hotel

Registration Steps

1. Arrive at Regal Oriental Hotel;
2. Inform the conference staff of your paper ID;
3. Sign your name on the Participants list;
4. Sign your name on Lunch & Dinner requirement list;
5. Check your conference kits: (1 conference program, 1 lunch coupon, 1 dinner coupon, 1 receipt, 1 name card, 1 flash disk (papers collection), 1 laptop bag);

Finish registration.

Tips: Please arrive at the conference to upload or copy Slides (PPT) into the laptop room 10 minutes before the session begins.

Note:

(1) The organizer doesn't provide accommodation, and we suggest you make an early reservation.

(2) One Best Presentation will be selected from each presentation session, and the Certificate for Best Presentation will be awarded at the end of each session on June 12, 2019.

(3) One day tour includes lunch but does not include attractions tickets, and participants need to take care of themselves.

Presentation Instructions

Instructions for Oral Presentations

Devices Provided by the Conference Organizer:

Laptop Computer (MS Windows Operating System with MS PowerPoint and Adobe Acrobat Reader)
Digital Projectors and Screen
Laser Sticks

Materials Provided by the Presenters:

PowerPoint or PDF Files (Files should be copied to the Conference laptop at the beginning of each Session.)

Duration of each Presentation (Tentatively):

Regular Oral Presentation: about 12 Minutes of Presentation and 3 Minutes of Question and Answer

Instructions for Poster Presentation

Materials Provided by the Conference Organizer:

The place to put poster

Materials Provided by the Presenters:

Home-made Posters
Maximum poster size is A1
Load Capacity: Holds up to 0.5 kg

Best Presentation Award

One Best Presentation will be selected from each presentation session, and the Certificate for Best Presentation will be awarded at the end of each session on June 12, 2019.

Dress code

Please wear formal clothes or national representative of clothing.

Schedule for Conference

Lobby of Regal Oriental Hotel, June 11 (10:00-17:00)
Arrival and Registration
Conference Room I – II at 1/F, June 12 (9:00-12:05)
Opening Remark (9:00-9:05) Prof. Changxu (Sean) Wu, University of Arizona, USA
Keynote Speech I (9:05-9:45) **Title: Fusion of Smart Clouds with Machine Learning, Artificial Intelligence and Internet of Things** Prof. Kai Hwang, Chinese University of Hong Kong, China
Keynote Speech II (9:45-10:25) **Title: Taming tsunami of scientific data** Prof. Scott A. Klasky, Oak Ridge National Laboratory, USA
Coffee Break & Group Photo (10:25-10:45)
Keynote Speech III (10:45-11:25) **Title: Networked Wireless Sensor Data Acquisition in the Big Data Era** Prof. Yao Liang, Purdue University School of Science, Indiana University Purdue University, USA
Keynote Speech IV (11:25-12:05) **Title: Human Performance Modeling and its Applications in Systems Engineering** Prof. Changxu (Sean) Wu, University of Arizona, USA
Lunch (12:05-13:00)
Conference Room I – II at 1/F, June 12 (13:00-18:10)
Session 1 Big Data Analysis (13:00-15:30) Chair: Prof. Changxu (Sean) Wu
Coffee Break (15:30-15:45)
Session 2 Big Data Applications (15:45-18:10) Chair: Prof. Yao Liang
Dinner (18:30-19:30)
Conference Room I – II at 1/F, June 12 (10:25-15:45)
Poster Session
June 13 (9:00-17:00)
One-Day Tour

Morning Session

Morning, June 12, 2019 (Wednesday)

Time: 9:00~12:05

Venue: Regal Oriental Hotel

Opening Remarks (9:00~9:05)

Addressed by Prof. Changxu (Sean) Wu, University of Arizona, USA

Keynote Speech I (9:05~9:45)

Title: Fusion of Smart Clouds with Machine Learning, Artificial Intelligence and Internet of Things

IEEE Live Fellow, Prof. Kai Hwang

Chinese University of Hong Kong, China

Abstract— In this open speech, Dr. Hwang will address critical issues surrounding the fusion of cloud architecture, AI chips, fog nodes, mobile devices, and intelligent robots in various IoT edge and cloud platforms to face the grand challenges in big-data, AI and cognitive applications. He will assess hardware, software, networking and intelligence integration techniques are illustrated with pioneering projects at Google, AWS, Chinese Academy of Sciences, Aliyun, etc. The ultimate design goal is to achieve enhanced agility in AI systems, end user mobility, global security, and scalability in public clouds and IoT platforms.

The roles of machine learning and big-data analytics are identified for edge, fog and cloud computing. He will cover the co-design of hardware mechanisms, AI system upgrade, and software support to achieve sustained performance and ensure cyber security. Intelligent robots, IoT devices, NB-IoT, 5G and satellite-based intelligent networks are presented to face the challenges in AI and HPC in cloud/IoT applications. An on-going research project is presented for integrated IoT/Cloud Computing at CUHK-SZ. This project develops a smart cloud for use in industrial design and intelligent manufacturing in the greater Bay Area of Peral River.

Keynote Speech II (9:45~10:25)

Title: TamingTsunami of Scientific Data

Prof. Scott A. Klasky

Oak Ridge National Laboratory, USA

Abstract—The USA Exascale Computing Project (ECP) is focused on accelerating the delivery of a capable exascale computing ecosystem that delivers 50 times more computational science and data analytic application power than possible with DOE HPC systems such as Titan (ORNL) and Sequoia (LLNL). As next generation applications and experiments grow in concurrency and in complexity, the data produced often grows to extreme levels, limiting scientific knowledge discovery.

In this presentation, Dr. Scott A. Klasky will talk about the new set of applications and experiments which push the edge of scientific data processing and simulation. He will present some of the exciting new research in this area to cope with this tsunami of data, along with the challenges in implementing these effectively on next-generation computer architectures. In his presentation he will also focus on many of the software technologies my group is developing to meet these challenges, including ADIOS (https://www.olcf.ornl.gov/center-projects/adios/) a next generation to ingest, reduce, and move data on HPC systems and over the WAN to other computational resources. His goal is to excite a new set of researchers to work in this impactful area and to attract researchers from around the world to work together on scientific challenges that they can make impact together.

Coffee Break & Group Photo Taking 10:25~10:45

Keynote Speech III (10:45~11:25)

Title: Networked Wireless Sensor Data Acquisition in the Big Data Era

Prof. Yao Liang

Purdue University School of Science,

Indiana University Purdue University, USA

Abstract—Wireless sensor networks (WSNs) and Internet of Things (IoT) are fundamentally changing today's practice of numerous scientific and engineering endeavors by enabling continuous monitoring and sensing physical variables of interest at unprecedented high spatial densities and long-time durations. This has significantly impacted broad fields such as environmental sciences, ecosystems, natural hazards, precision agriculture, smart building, and smart city. We focus on outdoor large-scale WSNs/IoT that are deployed in harsh environments such as mountainous areas, hilly watersheds, and forests, which presents great challenges in WSN/IoT data acquisition, because of the severe resource constraints (e.g., battery power, bandwidth, memory size, and CPU capacity) of tiny sensor nodes. In this talk, I will share my group's work on wireless sensor data acquisition and present our novel compressed sensing approach which can recover the sensing data at the sink with high fidelity when very few data packets need be collected, leading to a significant reduction of the network transmissions and thus an extension of the WSN/IoT lifetime. I will also illustrate our results using a real-world outdoor environmental WSN testbed deployed in Pennsylvania USA.

Keynote Speech IV (11:25~12:05)

Title: Human Performance Modeling and its Applications in Systems Engineering

Prof. Changxu (Sean) Wu

University of Arizona, USA

Abstract—This research seminar introduces the major research activities at the Cognitive System at University of Arizona, focusing on human cognition/performance modeling with its applications in systems engineering (e.g., human-in-the-loop transportation systems and human-machine interaction). Human performance modeling is a growing and challenging area in human factors and cognitive systems engineering. It builds computational models based on the fundamental mechanisms of human cognition and human-system interaction, employs both mathematical and discrete event simulation methods in industrial engineering, and predicts human performance and workload in real-world systems. It can be used to design, improve, and evaluate systems with human in the loop. Current and future research topics as well as the relation between human model and big data will also be introduced.

Lunch 12:05~13:00

Oral Presentation Abstracts

Session 1- Big Data Analysis

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, June 12, 2019 (Wednesday)

Time: 13:00-15:30

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Changxu (Sean) Wu

BD0008 Presentation 1

Smart Meter for Smart Homes: Data Mining Applications of Produced Big Data

Shashi Kant Srivastava

Jindal Global Business School, India

Abstract—We attempt to address the opportunities spawned from the voluminous data generated by the incorporation of smart technology within a building. Furthermore, we suggest a framework to discover the acceptance pattern of smart technologies. In the absence of the actual data, we simulated data to perform our research. The main objective of the research is to demonstrate the potential applications of the data obtained from the smart meter to different stakeholders involved in business and policy. Since buildings are one of the largest concerns for various business and government organizations, our research provides multiple future avenues to researchers. The paper demonstrates exploration process of technology adoption behavior of building occupants.

Oral Presentation Abstracts

Session 1- Big Data Analysis

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, June 12, 2019 (Wednesday)

Time: 13:00-15:30

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Changxu (Sean) Wu

BD0016 Presentation 2

A Free Stale Synchronous Parallel Strategy for Distributed Machine Learning

Hang Shi[1], Yue Zhao[1], Bofeng Zhang[1], Kenji Yoshigoe[2] and Athanasios V. Vasilakos[3]

1: Hang Shi, Yue Zhao, Bofeng Zhang, Kenji Yoshigoe and Athanasios V. Vasilakos; 2: Faculty, Japan; 3: Department of Computer Science, Electrical and Space Engineering, Lulea University of Technology, Sweden

Abstract— With the machine learning applications processing larger and more complex data, people tend to use multiple computing nodes to execute the machine learning tasks in distributed way. However, in real world, people always encounter a problem that a few nodes in system exhibit poor performance and drag down the efficiency of the whole system. In existing parallel strategies such as bulk synchronous parallel and stale synchronous parallel, these nodes with poor performance may not be monitored and found out in time. To address this problem, we proposed a free stale synchronous parallel (FSSP) strategy to free the system from the negative impact of those nodes. Our experimental results on some classical machine leaning algorithms and datasets demonstrated that FSSP strategy outperformed other existing parallel computing strategy.

Oral Presentation Abstracts

Session 1- Big Data Analysis

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, June 12, 2019 (Wednesday)

Time: 13:00-15:30

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Changxu (Sean) Wu

BD0018 Presentation 3

Indonesian Automatic Text Summarization Based on a New Clustering Method in Sentence Level

Zefeng Cai[1], Nankai Lin[1], Chuyu Ma[1] and Shengyi Jiang[2]

1: School of Information Science and Technology of Guangdong University of Foreign Studies, China

2: Eastern Language Processing Center, China

Abstract—With the development of the Internet, the amount of information grows exponentially, and the automatic text summarization technology becomes more and more important. At present, the majority of researches on automatic summarization techniques are applied to common languages such as Chinese and English, but it is few in low resource language. In this paper, we constructed an automatic summary dataset of Indonesian language and conducts related research on Indonesian automatic abstracts. And in this paper, we propose a new and efficient extraction-based automatic text summarization method based on sentence similarity clustering. Based on the idea of clustering, this paper considers the semantics of sentences and we clusters sentences according to the similarity between sentences and sentences. According to the rules we extracts the abstracts and finally obtains the summarization results. This method not only ensures the integrity, criticality and importance of the summary, but also reduces the information redundancy of the summary. In the evaluation, our method achieved good results and exceeded all the baselines in the indexes of F_1 score of ROUGE-1、ROUGE-2、ROUGE-3.

Oral Presentation Abstracts

Session 1- Big Data Analysis

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, June 12, 2019 (Wednesday)

Time: 13:00-15:30

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Changxu (Sean) Wu

BD0019 Presentation 4

CACV-tree: A New Computational Approach for Sentence Similarity Modeling

Jingwei Wang, Wenxin Hu and Wen Wu
East China Normal University, China

Abstract—Sentence similarity modeling plays an important role in Natural Language Processing (NLP) tasks, and thus has received much attention. In recent years, due to the success of word embedding, the neural network method has achieved sentence embedding,obtaining attractive performance. Nevertheless, most of them focused on learning semantic information and modeling it as a continuous vector, while the syntactic information of sentences has not been fully exploited. On the other hand, prior works have shown the benefits of structured trees that include syntactic information, while few methods in this branch utilized the advantages of sentence compression. This paper makes the first attempt to absorb their advantages by merging these techniques in a unified structure, dubbed as CACV-tree. The experimental results, based on 14 widely used datasets, demonstrate that our model is effective and competitive, compared against state-of-the-art models.

Oral Presentation Abstracts

Session 1- Big Data Analysis

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, June 12, 2019 (Wednesday)

Time: 13:00-15:30

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Changxu (Sean) Wu

BD0020 Presentation 5

A Heterogeneous and Interactive Big Earth Data Framework

Haixia Bi, Yong Xue, Patrick Merritt, Chris Windmill and Bradley Davis
University of Derby, United Kingdom

Abstract—With the rapid development of Earth science and its implementations, massive data-sets are collected from satellite observations, ground sensor networks and other sources. However, the characteristics of the derived Earth data, such as huge volume, multi-source, multi-resolution, and multi-temporal, bring significant challenges to the storage, processing and visualization of the big Earth data. To address the problems caused by the huge Earth data-sets, this paper presents a heterogeneous and interactive big Earth data framework. It integrates the raster-vector Earth data cloud storage into one framework, and introduces data processing procedures based on work ow and machine learning techniques. In addition, we design a real-time rendering and interactive visualization of the Earth data. The framework provides a theoretical reference for future implementations of the system.

Oral Presentation Abstracts

Session 1- Big Data Analysis

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, June 12, 2019 (Wednesday)

Time: 13:00-15:30

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Changxu (Sean) Wu

BD0023 Presentation 6

The Role of Big Data, Data Science and Data Analytics in Financial Engineering

Venkamaraju Chakravaram[1], Vidya Sagar Rao G[2] and Srinivas Jangirala[1]

1:O. P. Jindal Global University, India; 2: Osmania University, India

Abstract—Financial engineering is the process of creating innovative solutions for the existing financial problems of a company by using applications of mathematical methods. Financial engineering uses tools and knowledge from the fields of computer science, big data, data science, data analytics, statistics, economics and applied mathematics to address current financial issues as well as to devise new and innovative financial products. Financial Engineering is helpful in derivative pricing, financial regulation, execution, corporate finance, portfolio management, risk management, trading of structured products. Therefore, financial engineering is used by Commercial Banks, Investment Banks, Insurance companies and other fund hedging agencies. The present study focus on the role of big data, data science and data analytics in financial engineering as a successful tool at all stages of insurance business management practices. How these insurance companies are using said three data tools effectively as fasteners of financial engineering for the successful design, development and implementation of innovative business processes and products in this competitive and ever changing insurance market with innovative product features and strategies.

Oral Presentation Abstracts

Session 1- Big Data Analysis

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, June 12, 2019 (Wednesday)

Time: 13:00-15:30

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Changxu (Sean) Wu

BD0027 Presentation 7

An Efficient Method for Text Classification Task

Qiancheng Liang, Ping Wu and Chaoyi Huang

East China Normal University Department of Computer Science & Technology, China

Abstract—Text classification has always been a research hot-spot in the field of natural language processing. For the neural network input matrix, only the word vector of the word level is extracted, which ignores the expression of the overall semantic features of the text level, resulting in insufficient representation of text features and affecting accurate classification. In this paper, a text representation matrix combining word2vec and LDA topic models is proposed. Combining word meaning and semantic features, inputting LSTM for text classification, and introducing Attention mechanism to improve LSTM model, LSTM-Attention model is designed. The experimental results show that the LSTM classification model has better classification result than the traditional machine learning model, and the LSTM model with the Attention mechanism has a certain degree of improvement compared with the classical text classification models.

Oral Presentation Abstracts

Session 1- Big Data Analysis

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, June 12, 2019 (Wednesday)

Time: 13:00-15:30

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Changxu (Sean) Wu

BD0035 Presentation 8

Using Machine Learning for Prediction of Factors Affecting Crimes in Saudi Arabia

Anadil Alsaqabi, Fatimah Aldhubayi and Saleh Albahli
Qassim University, Saudi Arabia

Abstract—Crime rates are expected to increase in the world as the growth of many complex factors like: unemployment, poverty, weather, violent ideologies and religion and etc. Obviously crimes have negatively influenced the development of society, economic progress and reputation of a nation. Hence, Analyzing large volume of data with machine learning algorithms can be used to predict the crime distribution over an area to provide indicators of specific areas which may become a criminal hotspot. The aim of this paper is to predict factors that most affected crimes in Saudi Arabia by developing a machine learning model to predict an acceptable output value. Our results show that Factor Analysis of Mixed Data (FAMD) as features selection methods showed more accurate on machine learning classifiers rather than Principal Component Analysis (PCA) method. Naïve Bayes classifier perform better than other classifiers on both features selections methods with accuracy 97.53% for FAMD and PCA equals to 97.10%.

Oral Presentation Abstracts

Session 1- Big Data Analysis

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, June 12, 2019 (Wednesday)

Time: 13:00-15:30

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Changxu (Sean) Wu

BD0051 Presentation 9

Application Strategy of Big Data in the Development of Complex Industrial Products (CIPs)

Baolei Zhang, Fuquan Zhao and Zongwei Liu

Tsinghua University, China

Abstract—The product development is the key business of manufacturing and determines the competitive advantage of manufacturing enterprises, and has high difficulty in implementation. The product development of Complex Industrial Products (CIPs) is a great practical challenge for most enterprises. The demand for mass customization products makes enterprises to face more complicated product development situation. The deep integration of information technology and manufacturing technology makes big data an important value source for enterprises. Full application of big data to promote product development of CIPs has become a feasible approach for product development of enterprises. The value of big data needs to be applied through the knowledge-based application of data. The core work is to develop the functional data model. The application of big data in product development will eventually move towards knowledge-based intelligent. The case study provides the mechanism verification for the application strategy of big data in the development of CIPs.

Oral Presentation Abstracts

Session 1- Big Data Analysis

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, June 12, 2019 (Wednesday)

Time: 13:00-15:30

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Changxu (Sean) Wu

BD0060 Presentation 10

Construction and Application of Data Standard in Big Data Environment

Jia Haitian[1] and Jia Chun[2]

1: Suzhou Institute of Trade & Commerce, China; 2: Yellow River Conservancy Technical Institute, China

Abstract—The school information construction has gone through more than 10 years, the existing system has reached a certain number, each system runs independently. At the same time, the role of security monitoring, online course resources and online behavior unstructured data resources in the construction of intelligent campus has become more and more important. According to the current situation of the university, this paper puts forward a data model for the construction of university information. The integration of structured data and unstructured data is solved by Hadoop distributed system architecture, which provides the basis for data analysis and decision-making. Big data will be the direction of intelligent campus development in the next 1-2 years, and its construction and implementation will effectively promote the construction of school information.

Coffee Break 15:30~15:45

Oral Presentation Abstracts

Session 2- Big Data Applications

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, June 12, 2019 (Wednesday)

Time: 15:45-18:10

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Yao Liang

BD0009 Presentation 11

Review and Investigate the Mapping Knowledge Domain of Financial Big Data Research

Wei Zhou, Danxue Luo and Jin Chen

Yunnan University of Finance and Economics, China

Abstract—In today's society, the big data is widely used in various fields owing to its quantitative and objective characteristics. Therefore, there has been an increasing number of investigations on financial issues related to big data in recent years. It is believed that analyzing the status quo and the emerging trends of financial research and big data research and the beginner who are interested in financial and big data research. To do so, this paper provides the mapping knowledge domain of financial and big data research based on 724 papers on the web of science from 1992 to 2018 by using Cite Space, which is an effective tool for scientometric studies. The visualization analyses of cited reference cluster, collaborations networks, author co-citation, and timeline view are presented in this study to show the research streams and the papers that made significant theoretical contributions. Also, the authors in this research area are analyzed in detail. Besides, the specific hot spots and emerging trends can be identified. There are two contributions to this study. Firstly, we give a comprehensive investigation of the status quo and the emerging trends of financial and big data research in the recent 26 years. Secondly, we make the development of financial research and big data easier and direct to learn for beginners.

Oral Presentation Abstracts

Session 2- Big Data Applications

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, June 12, 2019 (Wednesday)

Time: 15:45-18:10

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Yao Liang

BD0011-A Presentation 12

Quality Control Framework of Big Data for Early Warning of Agricultural Meteorological Disasters

Shunbao Liao and Jiale Li

Institute of Disaster Prevention, China

Abstract— Agricultural meteorological disasters, including floods, droughts, dry hot winds, low temperature chills, typhoons, hail and continuous rain, can lead to significant reduction in agricultural output. The big data platform for early warning of agricultural meteorological disaster is the basis of business operation system for early warning of agricultural meteorological disasters, and the data quality is an important guarantee for success of the early warning. Quality control of big data for early warning of agricultural meteorological disaster involves names of data sets, metadata, data documents and content of data sets. The quality control for contents of data sets is divided into quality control of attribute data and that of spatial data, and quality control of spatial data is divided into quality control of vector data and that of raster data. Methods for data quality control are divided into full automatic, semi-automatic and full manual control methods.

Oral Presentation Abstracts

Session 2- Big Data Applications

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, June 12, 2019 (Wednesday)

Time: 15:45-18:10

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Yao Liang

BD0013 Presentation 13

Hotspot Detection in Social Media Based on Improved Strategy Clustering

Ruixin Ma, Peng Xu, Lili Li and Chuang Wang

Dalian University of Technology, China

Abstract— With the rapid growth of social media platform, hotspot detection has become an increasing important issue in the We media era. As for researchers, the emergence of all kinds of public opinion crises is one of the greatest challenges. In order to make a response to public opinion precisely, a more effective approach for hotspot detection is necessary. Clustering algorithm is one of the best choices for distinguishing hotspots. However, the existing methods still need some improvement in filtering data and calculating text similarity. Therefore, we propose a method which combines the two strategies to improve the performance of clustering algorithm for hotspot detection. One strategy is message importance which is used for measuring the value of information from social media. The other strategy is an improved text similarity based on feature words which is applied to solve sparse vector from the short texts. Through improving the accuracy of clustering, we can obtain more accurate hotspots. We validate the proposed method is feasible through the experiment based on the datasets from Weibo online.

Oral Presentation Abstracts

Session 2- Big Data Applications

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, June 12, 2019 (Wednesday)

Time: 15:45-18:10

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Yao Liang

BD0021 Presentation 14

Big Data Quality Metrics for Sentiment Analysis Approaches

Imane El Alaoui and Youssef Gahi

University of Ibn Tofail, Morocco

Abstract—In a world increasingly connected, and in which information flows quickly and affects a very large number of people, sentiment analysis has seen a spectacular development over the past ten years. This is due to the fact that the explosion of social networks has allowed anyone with internet access to publicly express his opinion. Moreover, the emergence of big data has brought enormous opportunities and powerful storage and analytics tools to the field of sentiment analysis. However, big data introduces new variables and constraints that could radically affect the traditional models of sentiment analysis. Therefore, new concerns, such as big data quality, have to be addressed to get the most out of big data. In this paper, we first highlight the most important big data quality metrics to consider in any big data project. Then, we show how these metrics should be specifically considered in SA approaches and this for each phase in the big data value chain.

Oral Presentation Abstracts

Session 2- Big Data Applications

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, June 12, 2019 (Wednesday)

Time: 15:45-18:10

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Yao Liang

BD0033-A Presentation 15

AI-based Big-data Decision-making Support Systems for Korean Contractor's Life-cycle Project Management of Overseas Engineering-Procurement-Construction (EPC)

Jaemin Cho[1], Hoyoung Roh[1], Eulbum Lee[1] and Hyunsoo Kim[2]

1: Pohang university of science and technology (POSTECH), South Korea; 2: Sang-Ah management consulting, South Korea

Abstract—Due to the failure of data management by engineering & construction companies, project managers do not receive reliable data analysis to make timely decisions. To solve these problems, we are developing AI-based engineering big data integration analysis support system. The system is based on a big data-based knowledge base that collects and builds various EPC engineering commercial and public data through Enterprise Resource Planning (ERP) or Project Management Information System (PMIS), and an engineering machine learning platform with various algorithms. We are developing intelligent decision-making applications such as Predicting design costs, analyzing design errors, analyzing change order, analyzing bidding documents, and Plant equipment prediction maintenance. This has the advantage of leading to project managers' preemptive response through project risk management and dashboards. The research team used ERP, PMIS, commercial data (RS Means, Richardson, CESK) and public data from the engineering & construction companies as the data source. The expertise (Lessons Learned) gained through this platform can open up to the public or sell to other companies.

Oral Presentation Abstracts

Session 2- Big Data Applications

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, June 12, 2019 (Wednesday)

Time: 15:45-18:10

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Yao Liang

BD0037 Presentation 16

Application of Online/Offline Sales Big Data in Household Medical Device Industry

Liu Lixia[1], Hu Gang[2] and Zhang Ming[1]

1: Hunan Cofoe Medical Science and Technology Development Co., Ltd, China; 2: Computer School, National University of Defense Technology, China

Abstract—In this paper, we discuss the current situation of the application of big data in the sales of household medical devices industry, put forward the sales target of big data in this area, and finally analyze the path to achieve this goal.

Oral Presentation Abstracts

Session 2- Big Data Applications

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, June 12, 2019 (Wednesday)

Time: 15:45-18:10

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Yao Liang

BD0046 Presentation 17

A Secure Cloud-Assisted Certificateless Group Authentication Scheme for VANETs in Big Data Environment

Haowen Tan and Ilyong Chung

Chosun Univeristy, South Korea

Abstract—Nowadays, the construction of efficient intelligent transportation system (ITS) has become a new trend for metropolitan cities with increasingly large populations. As one of the most significant component of ITS, the vehicular ad hoc networks (VANETs) are capable of building temporary vehicular sensor networks for efficient and dynamic information exchange between vehicles and road side units (RSUs). As a matter of fact, the traditional VANETs have limited computing and storing capabilities, which restrict the rapid development VANETs services provided to the drivers. Hence, with the rapid development of big data facilities, the cloud-assisted VANETs structure is proposed in order to enhance the capabilities of VANETs. In addition, due to the inherent wireless communication characteristics, data transmissions of VANETs suffer from charted and uncharted security risks and attacks. Thus proper security strategies should be adopted to guarantee secure communication and driver privacy. Emphasizing on the above issues, we develop an efficient cloud-assisted certificateless grouping authentication scheme for VANETs. In our design, vehicle anonymity is provided during the entire communication process. Note that most of the current authenticating schemes assume the secure channel between the RSU and vehicles in order for initial key message transmission, which is not necessary in our scheme.

Oral Presentation Abstracts

Session 2- Big Data Applications

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, June 12, 2019 (Wednesday)

Time: 15:45-18:10

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Yao Liang

BD0059 Presentation 18

A Sentiment Analysis Model for Faculty Comments Evaluation Using Ensemble Machine Learning Algorithms

Jay-Ar Lalata

FEU Institute of Technology, Philippines

Abstract—Teacher evaluation is the systematic procedure done in educational institutions to review the performance of the teachers in a classroom. It also aims to provide constructive feedback for teacher's professional growth which will benefit students in their education. Opinions of the students through textual comments are contributory factors when evaluating teachers. In this study, sentiment analysis or opinion mining was used to analyze the students' comments. An ensemble approach integrating five individual machine algorithms namely Naive Bayes, Logistic Regression, Support Vector Machine (SVM), Decision Tree (CART) and Random Forest algorithms were applied to classify the comments based on Majority Voting Principle. The result shows that ensemble classification system outperforms these individual classifiers based from the evaluation measures. This approach could improve the overall accuracy of classification process of the students' comments which will help teacher in the improvement of courses.

Oral Presentation Abstracts

Session 2- Big Data Applications

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, June 12, 2019 (Wednesday)

Time: 15:45-18:10

Venue: Conference Room I – II at 1/F

Session Chair: Prof. Yao Liang

BD0061 Presentation 19

Spatial-temporal Data Association Based Ontology Alignment Research in High Education Context

Wei Wang, Wenxin Mu and Juanqiong Gou

Beijing Jiaotong University, China

Abstract—In the process of practicing smart campus, student behavior is highly concerned. Student activities produce spatial-temporal data, which records the daily life of students and contains the potential regulations of student behavior. The complexity of these data brings challenges for data collection and data analysis. The key to solve these problems is data fusion. In addition, ontology alignment is an important method for exploring the association between different ontology in different fields. It can solve the problem of data fusion in practice and maximize the value of data. At present, the research methods of ontology alignment are mostly mathematical similarity algorithms, and not considering the uncertainty of spatial-temporal data. Ontology is better way to deal with spatial-temporal data. In order to solve this problem effectively, this paper proposes a method based on spatial-temporal data association, which establishes fuzzy ontology and formulates a series of fuzzy rules for fuzzy reasoning, and mines the relationship between data; then it connects the different concepts between ontologies to realize the alignment of ontologies. Finally, the fuzzy ontology modeling and fuzzy reasoning are implemented and tested by using high education context. The rationality and effectiveness of the method are verified.

Dinner 18:30-19:30

Poster session

June 11, 2019 (Tuesday)

Time: 10:25~15:50

Venue: Conference Room I – II at 1/F

BD0014 Poster 1

Study on Standard System of Aerospace Quality Data Resources Integration under the Background of Big Data

Fengsheng Jia, Yang Gao and Yuming Wang

China Astronautics Standards Institute, China

Abstract— The integration and application of aerospace product quality data resources is an important way to carry out quality improvement, quality evaluation and precise management. Standardization is the basis for promoting quality data resources integration. The unified and normative standard system is the guarantee for efficient development of integration standards. Firstly, we analyzed the features of quality data resources according to the status quo of integration. Integration structure of quality data resources in terms of vertical and horizontal integration was proposed by adopting the methods of “decomposition-integration” and “classification-association”. Secondly, we constructed a three-dimensions architecture of quality data resource integration using the method of system engineering methodology, from the layer dimension (basis, common, special), technical dimension (description, collection, storage, transmission, processing, comprehensive management) and category dimension (rocket, spacecraft). Thirdly, we worked out 20 lists about basis and common standard by adopting the top-down approach. Finally, we proposed some standard development suggestions according to the characteristics of quality data resources and standard research strategies.

Poster session

June 11, 2019 (Tuesday)

Time: 10:25~15:50

Venue: Conference Room I – II at 1/F

BD0028 Poster 2

Survey of Intrusion Detection Methods Based on Data Mining Algorithms

Zichuan Jin, Yanpeng Cui and Zheng Yan

Xidian university, China

Abstract—With the development of data mining learning algorithms, such as One-class SVM, Fuzzy Clustering, K-means, Apriori and so on, they are more and more widely used in the field of security log analysis. For example, the combination of time series algorithm and association algorithm can be used to mine frequent item sets in transaction databases, and then generate association rules to discover the intrinsic relationship of security logs and find out the potential attack patterns of hackers. The combination of dimensionality reduction algorithm and clustering algorithm can speed up the distinction between normal log data and abnormal log data, and improve the efficiency. This paper discusses the latest security log analysis methods based on different data mining algorithms at home and abroad, lists the contribution and role of each research method for security analysis, and compares the advantages and disadvantages of the combination of different data mining algorithms for security analysis. According to the current demand of network security research, this paper puts forward the improvement direction of combining data mining algorithm with security log in the future.

Poster session

June 11, 2019 (Tuesday)

Time: 10:25~15:50

Venue: Conference Room I – II at 1/F

BD0031 Poster 3

Mining Private Vehicle Hot Routes Using Electronic Registration Identification Data

Chen Cui, Linjiang Zheng and Dihua Sun

Chongqing University, China

Abstract—Hot routes refer to routes that massive vehicles pass through in a period of time. Mining hot routes of private vehicles can help us understand the travel behavior of private vehicles, which is of great help to urban traffic management and construction. This study aims to mine hot routes of private vehicles using Electronic Registration Identification (ERI) data, which is huge amount of traffic data. In this paper, we propose a mining algorithm, Prefix-projected Sequence Pattern Mining based on Successor Set (PSSS), which is based on the idea of PrefixSpan algorithm to mine hot routes. Firstly, we extract private vehicle trips from ERI data. Then we transform trips into string sequences. We use the PSSS algorithm to mine hot routes of private vehicles. Finally, we analyze the hot routes of private vehicles and compare efficiency of two algorithms. The experimental results are of guiding significance to the traffic management and construction of intelligent transportation.

One Day Visit

June 13, 2019 (Thursday) 9:00~17:00

PRICEINCLUDES:

- air-conditioned sightseeing coach
- guided service tram ticket to the Victoria Peak admission to Sky Terrace

This classic tour includes riding the Peak Tram to visit Victoria Peak, follow by entering Sky Terrace 428 – the highest viewing platform in Hong Kong. Afterwards, tour will visit Aberdeen Fishing Village (old Hong Kong) and you may choose to ride the sampan (own expense). A visit to great bargain place Stanley Market is included.

- Round Trip Hotel Transfer with tour guide
- Ride the 130 years old Peak Tram to the Peak
- Sky Terrace 428 viewing platform

The highest viewing platform in Hong Kong

- Visit Old Hong Kong - Aberdeen Fishing Village Take a closer look at the old fishing village by taking a sampan ride (at own expense)
- Access to QTS award TSL jewellery factory

The itinerary includes a visit to TSL jewellery factory, the receiptian of the Outstanding QTS Merchant Award for more than 10 years and counting.

- Visit Stanley Market

Shoppers' Paradise for Chinses style souvenirs

- Complimentary classic Rickshaw Open Top Bus Ticket will be provided

Explore heritages sites at your convenience

The exact tour route will be adjusted according to the situation on June 13, 2019

Note

Note