

# SCF2020

2020 Services Conference Federation

## ADVANCE PROGRAM

The 2020 International Conference on Web Services (ICWS 2020)

 [www.icws.org](http://www.icws.org)

The 2020 International Conference on Services Computing (SCC 2020)

 [www.thesc.org](http://www.thesc.org)

The 2020 International Conference on Cloud Computing (CLOUD 2020)

 [www.TheCloudComputing.org](http://www.TheCloudComputing.org)

The 2020 International Conference on AI and Mobile Services (AIMS 2020)

 [www.theMobileServices.org](http://www.theMobileServices.org) or [www.ai1000.org](http://www.ai1000.org)

The 2020 International Conference on Big Data (BigData 2020)

 [www.bigdatacongress.org](http://www.bigdatacongress.org)

The 2020 World Congress on Services (SERVICES 2020)

 [www.ServicesCongress.org](http://www.ServicesCongress.org)

The 2020 International Conference on Internet of Things (ICIOT 2020)

 [www.iciot.org](http://www.iciot.org)

The 2020 International Conference on Edge Computing (EDGE 2020)

 [www.theEdgeComputing.org](http://www.theEdgeComputing.org)

The 2020 International Conference on Cognitive Computing (ICCC 2020)

 [www.theCognitiveComputing.org](http://www.theCognitiveComputing.org)

The 2020 International Conference on Blockchain (ICBC 2020)

 [www.blockchain1000.org](http://www.blockchain1000.org)

# SCF 2020 Leadership Team

## General Chairs

<b>ICWS 2020</b>	<b>AIMS 2020</b>
<i>Prof. Cheng-Zhong Xu, University of Macau, China</i> <i>Prof. Raj Sunderraman, Georgia State University, USA</i>	<i>Prof. Teruo Higashino, Osaka University, Japan</i> <i>Prof. Rafal Angryk, Georgia State University, USA</i> <i>Prof. Youping Deng, University of Hawaii, USA</i>
<b>CLOUD 2020</b>	<b>SERVICES 2020</b>
<i>Dr. Mudhakar Srivatsa, IBM Thomas J. Watson Research Center, USA</i> <i>Prof. Song Guo, Hong Kong Polytechnic University, China</i> <i>Prof. Shi-Jinn Horng, National Taiwan University of Science and Technology (Vice-Chair)</i>	<i>Prof. Jamal Bentahar, Concordia University, Canada</i>
<b>BigData 2020</b>	<b>EDGE 2020</b>
<i>Prof. Seung-Jong Jay Park, Louisiana State University, USA</i> <i>Prof. Minyi Guo, Shanghai Jiaotong University, China</i> <i>Prof. Min Li, Central South University, China</i>	<i>Karolj Sakla, Ruđer Bošković Institute, Croatia</i>
<b>SCC 2020</b>	<b>ICBC 2020</b>
<i>Prof. Lakshmish Ramaswamy, The University of Georgia, USA</i>	<i>Prof. J. Leon Zhao, City University of Hong Kong, China</i> <i>Prof. Richard Brooks, Clemson University, USA</i>
<b>ICIOT 2020</b>	
<i>Prof. Dimitrios Georgakopoulos, Swinburne University, Australia</i> <i>Prof. Zhipeng Cai, Georgia State University, USA</i>	

## Program Chairs

<b>ICWS 2020</b>	<b>SCC 2020</b>
<i>Prof. Wei-Shinn Ku, National Science Foundation, USA</i> <i>Mr. Yasuhiko Kanemasa, Fujitsu Laboratories Ltd, Japan</i> <i>Prof. Mohamed Adel Serhani, United Arab Emirates University, UAE</i> <i>Prof. Noseong Park, George Mason University, USA (Vice-Chair)</i>	<i>Prof. Qingyang Wang, Louisiana State University--Baton Rouge, USA</i> <i>Prof. Yunni Xia, Chongqing University, China</i> <i>Prof. Sangeetha Seshadri, IBM Almaden Research Center, USA</i>
<b>CLOUD 2020</b>	<b>AIMS 2020</b>
<i>Dr. Qi Zhang, IBM Thomas J. Watson Research, USA</i> <i>Prof. Yingwei Wang, University of Prince Edward Island, Canada</i>	<i>Prof. Ruifeng Xu, Harbin Institute of Technology, Shenzhen, China</i> <i>Dr. De Wang, Sunmi US Inc., USA</i> <i>Prof. Wei Zhong, University of South Carolina Upstate, USA</i> <i>Dr. Ling Tian, University of Electronic Science and Technology of China, China</i> <i>Dr. Yongsheng Bai, Eastern Michigan University, USA</i>
<b>BigData 2020</b>	<b>SERVICES 2020</b>
<i>Prof. Surya Nepal, CSIRO, Australia</i> <i>Dr. Wenqi Cao, Facebook, Inc, USA</i> <i>Prof. Aziz Nasridinov, Chungbuk National University, Korea (South)</i> <i>Dr. MD Zakirul Alam Bhuiyan, Fordham University, USA</i> <i>Dr. Xuan Guo, University of North Texas, USA</i>	<i>Prof. Joao Eduardo Ferreira, Department of Computer Science, University of Sao Paulo, Brazil</i> <i>Prof. Balaji Palanisamy, University of Pittsburgh, USA</i> <i>Prof. Kejiang Ye, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China (Vice-Chair)</i> <i>Mr. Siva Kantamneni, Principal / Partner at Deloitte Consulting, USA (Vice-Chair)</i>
<b>ICCC 2020</b>	<b>EDGE 2020</b>
<i>Dr. Yujiu Yang, Tsinghua University, Shenzhen, China</i> <i>Dr. Lei Yu, IBM Thomas J. Watson Research Center, USA</i>	<i>Dr. Ajay Katangur, Missouri State University, USA</i> <i>Prof. Shih-Chun Lin, North Carolina State University, USA</i> <i>Dr. Jinpeng Wei, University of North Carolina at Charlotte, USA (Vice-Chair)</i> <i>Prof. Shuhui Yang, Purdue University Northwest, USA (Vice-Chair)</i>
<b>ICIOT 2020</b>	<b>ICBC 2020</b>
<i>Prof. Wei Song, North China University of Technology, China</i> <i>Prof. Kisung Lee, Louisiana State University--Baton Rouge, USA</i> <i>Prof. Zhisheng Yan, Georgia State University, USA (Vice-Chair)</i>	<i>Prof. Zhixiong Chen, PhD, Mercy College, USA</i> <i>Prof. Laizhong Cui, Shenzhen University, China (Vice-Chair)</i> <i>Dr. Balaji Palanisamy, University of Pittsburgh, USA (Vice-Chair)</i>

## **SCF 2020 General Chairs**

Prof. Yi Pan, Georgia State University, USA  
Prof. Samee U. Khan, North Dakota State University, USA  
Dr. Wu Chou, Vice President-Artificial Intelligence & Software at Essenlix Corporation, USA  
Dr. Ali Arsanjani, Amazon Web Services (AWS), USA

## **SCF 2020 Program Chair**

Dr. Liang-Jie Zhang, Kingdee International Software Group Co., Ltd, China

## **SCF 2020 Industry Track Chair**

Mr. Siva Kantamneni, Principal / Partner at Deloitte Consulting, USA

## **CFO**

Dr. Min Luo, Huawei, USA

## **Industry Exhibit and International Affairs Chair**

Dr. Zhixiong Chen, Mercy College, USA

## **Steering Committee**

Prof. Calton Pu, Georgia Tech, USA (Co-Chair)  
Dr. Liang-Jie Zhang, Kingdee International Software Group Co., Ltd, China (Co-Chair)

## Program Overview

Sept 18, 2020 (Friday)		
Services Conference Federation (SCF 2020) (ICBC 2020, CLOUD 2020, ICWS 2020, BigData 2020, SCC 2020, EDGE 2020, ICC 2020, ICIOT 2020, AIMS 2020, SERVICES 2020)		
Channel	Automated Webinar	On-Demand Webinar
0:00-8:00		<b>All Sessions: 7/24 Available</b> ICWS 2020 Sessions I-V CLOUD 2020 Sessions I-VII BIGDATA 2020 Sessions I-VI SCC 2020 Sessions I-III AIMS 2020 Sessions I-IV SERVICES 2020 Sessions I-III ICCC 2020 Sessions I-III ICIOT 2020 Sessions I-IV EDGE 2020 Sessions I-III ICBC 2020 Sessions I-V
8:00-11:00	Services Conference Federation (SCF 2020) Opening Steering Committee Chair; General Chairs; PC Chairs; Opening & Invited Plenary Session	
11:00-16:00		
16:00-19:00	Services Conference Federation (SCF 2020) Opening Steering Committee Chair; General Chairs; PC Chairs; Opening & Invited Plenary Session	
19:00-24:00		

Sept 19, 2020 (Saturday)		
Services Conference Federation (SCF 2020) (ICBC 2020, CLOUD 2020, ICWS 2020, BigData 2020, SCC 2020, EDGE 2020, ICCC 2020, ICIOT 2020, AIMS 2020, SERVICES 2020)		
Channel	Aotomated Webinar	On-Demand Webinar
0:00-8:00		<b>All Sessions: 7/24 Available</b> ICWS 2020 Sessions I-V CLOUD 2020 Sessions I-VII BIGDATA 2020 Sessions I-VI SCC 2020 Sessions I-III AIMS 2020 Sessions I-IV SERVICES 2020 Sessions I-III ICCC 2020 Sessions I-III ICIOT 2020 Sessions I-IV EDGE 2020 Sessions I-III ICBC 2020 Sessions I-V
8:00-11:00	Services Conference Federation (SCF 2020) Invited Plenary Session	
11:00-16:00		
16:00-19:00	Services Conference Federation (SCF 2020) Invited Plenary Session	
19:00-24:00		

Sept 20, 2020 (Sunday)		
Services Conference Federation (SCF 2020) (ICBC 2020, CLOUD 2020, ICWS 2020, BigData 2020, SCC 2020, EDGE 2020, ICC 2020, ICIOT 2020, AIMS 2020, SERVICES 2020)		
Channel	Automated Webinar	On-Demand Webinar
0:00-8:00		<b>All Sessions: 7/24 Available</b> ICWS 2020 Sessions I-V CLOUD 2020 Sessions I-VII BIGDATA 2020 Sessions I-VI SCC 2020 Sessions I-III AIMS 2020 Sessions I-IV SERVICES 2020 Sessions I-III ICC 2020 Sessions I-III ICIOT 2020 Sessions I-IV EDGE 2020 Sessions I-III ICBC 2020 Sessions I-V
8:00-11:00	<b>Services Conference Federation (SCF 2020) Opening Steering Committee Chair; General Chairs; PC Chairs; Opening &amp; Invited Plenary Session</b>	
11:00-16:00		
16:00-19:00	<b>Services Conference Federation (SCF 2020) Invited Plenary Session</b>	
19:00-24:00		

## About Springer

**Springer** is a leading global scientific, technical and medical portfolio, providing researchers in academia, scientific institutions and corporate R&D departments with quality content through innovative information, products and services. Springer has one of the strongest STM and HSS eBook collections and archives, as well as a comprehensive range of hybrid and open access journals. Springer is part of **Springer Nature**, a global publisher that serves and supports the research community. Springer Nature aims to advance discovery by publishing robust and insightful science, supporting the development of new areas of research and making ideas and knowledge accessible around the world.

As part of Springer Nature, Springer sits alongside other trusted brands like Nature Research, BMC and Palgrave Macmillan. Visit [www.springer.com](http://www.springer.com) and follow [@SpringerNature](https://twitter.com/SpringerNature)





## **SCF2020 Proceedings by Springer LNCS**



**ICWS 2020 – The 2020 International Conference on Web Services**

Springer LNCS Volume Number: 12406

**CLOUD 2020 – The 2020 International Conference on Cloud Computing**

Springer LNCS Volume Number: 12403

**BIGDATA 2020 – The 2020 International Conference on Big Data**

Springer LNCS Volume Number: 12402

**SCC 2020 – The 2020 International Conference on Services Computing**

Springer LNCS Volume Number: 12409

**AIMS 2020 – The 2020 International Conference on AI & Mobile Services**

Springer LNCS Volume Number: 12401

**SERVICES 2020 – The 2020 World Congress on Services**

Springer LNCS Volume Number: 12411

**ICCC 2020 – The 2020 International Conference on Cognitive Computing**

Springer LNCS Volume Number: 12408

**ICIOT 2020 – The 2020 International Conference on Internet of Things**

Springer LNCS Volume Number: 12405

**EDGE 2020 – The 2020 International Conference on Edge Computing**

Springer LNCS Volume Number: 12407

**ICBC 2020 – The 2020 International Conference on Blockchain**

Springer LNCS Volume Number: 12404

## Table of Contents

SCF 2020 Leadership Team .....	2
Program Overview.....	5
About Springer .....	8
SCF2020 Proceedings by Springer LNCS.....	9
2020 Services Conference Federation Invited Plenary Sessions .....	11
2020 International Conference on Web Services (ICWS 2020).....	12
2020 International Conference on Cloud Computing (CLOUD 2020).....	17
2020 International Conference on Big Data (BigData 2020) .....	23
2020 International Conference on Services Computing (SCC 2020) .....	29
2020 International Conference on AI & Mobile Services (AIMS 2020).....	32
2020 World Congress on Services (SERVICES 2020) .....	36
2020 International Conference on Cognitive Computing (ICCC 2020).....	39
2020 International Conference on Internet of Things (ICIOT 2020).....	42
2020 International Conference on Edge Computing (EDGE 2020) .....	46
2020 International Conference on Blockchain (ICBC 2020) .....	49
Journals on Services Innovations .....	54
Services Society .....	55
Introduction of Big Data Certification.....	56
Services Society Special Interests Groups (SIGs) .....	57
Services Conference Federation.....	58

# **2020**

## **Services Conference Federation**

### **Invited Plenary Sessions**

(09/18 Friday, 8:00-11:00)  
(09/18 Friday, 16:00-19:00)

(09/19 Saturday, 8:00-11:00)  
(09/19 Saturday, 16:00-19:00)

(09/20 Sunday, 8:00-11:00)  
(09/20 Sunday, 16:00-19:00)

# 2020 International Conference on Web Services (ICWS 2020)

---

## Session I: Web-based Services (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Wei-Shinn Ku (National Science Foundation, USA)

20 minutes for each paper.

### Web Service API Anti-patterns Detection as a Multi-Label Learning Problem (ICWS2020-0001)

*Islem Saidani and Ali Ouni (Ecole de Technologie Supérieure (ETS), University of Quebec, Canada); Mohamed Wiem Mkaouer (Rochester Institute of Technology (RIT), Rochester, NY, USA)*

**Abstract:** Anti-patterns are symptoms of poor design and implementation solutions applied by developers during the development of their software systems. Recent studies have identified a variety of Web service anti-patterns and defined them as sub-optimal solutions that result from bad design choices, time pressure, or lack of developers experience. The existence of anti-patterns often leads to software systems that are hard to understand, reuse, and discover in practice. Indeed, it has been shown that service designers and developers tend to pay little attention to their service interfaces design. Web service anti-patterns detection is a non-trivial and error-prone task as different anti-pattern types typically have interleaving symptoms that can be subjectively interpreted and hence detected in different ways. In this paper, we introduce an automated approach that learns from a set of interleaving Web service design symptoms that characterize the existence of anti-pattern instances in a service-based system. We build a multi-label learning model to detect 8 common types of Web service anti-patterns. We use the ensemble classifier chain (ECC) model that transforms multi-label problems into several single-label problems which are solved using genetic programming (GP) to find the optimal detection rules for each anti-pattern type. To evaluate the performance of our approach, we conducted an empirical study on a benchmark of 815 Web services. The statistical tests of our results show that our approach can detect the eight Web service anti-pattern types with an average F-measure of 93% achieving a better performance compared to different state-of-the-art techniques. Furthermore, we found that the most influential factors that best characterize Web service anti-patterns include the number of declared operations, the number of port types, and the number of simple and complex types in service interfaces.

### A Stochastic-Performance-Distribution-Based Approach to Cloud Workflow Scheduling with Fluctuating Performance (ICWS2020-0002)

*Yi Pan, Xiaoning Sun and Yunni Xia (Chongqing University, China); Chen Peng (Xihua University, China); Shanchen Pang (China University of Petroleum, China); Xiaobo Li (Chongqing Animal Husbandry Techniques Extension Center, China); Yong Ma (Jiangxi Normal University, China)*

**Abstract:** The cloud computing paradigm is characterized by the ability to provide flexible provisioning patterns for computing resources and on-demand common services. As a result, building business processes and workflow-based applications on cloud computing platforms is becoming increasingly popular. However, since real-world cloud services are often affected by real-time performance changes or fluctuations, it is difficult to guarantee the cost-effectiveness and quality-of-service (QoS) of cloud-based workflows at real time. Existing researches in this direction mainly consider time-invariant performance of cloud infrastructures and scheduling decision-making as a static optimization problem. In this work, instead, we consider that workflows, in terms of Directed Acyclic Graphs (DAGs), to be supported by decentralized cloud infrastructures are with time-varying performance and aim at reducing the monetary cost of workflows with the completion-time constraint to be satisfied. We tackle the performance-fluctuation workflow scheduling problem by incorporating a stochastic-performance-distribution-based framework for estimation and optimization of workflow critical paths. The proposed method dynamically generates the workflow scheduling plan according to the accumulated stochastic distributions of tasks. In order to prove the effectiveness of our proposed method, we conducted a large number of experimental case studies on real third-party commercial clouds and showed that our method was significantly better than the existing method.

### Web API Search: Discover Web API and its Endpoint with Natural Language Queries (ICWS2020-0003)

*Lei Liu (Fujitsu Laboratories of America, Inc., USA); Mehdi Bahrami (Fujitsu Laboratory of America Inc., USA); Junhee Park (Fujitsu Laboratories of America, Inc., USA); Wei-Peng Chen (Fujitsu Laboratory of America Inc., USA)*

**Abstract:** In recent years, Web Application Programming Interfaces (APIs) are becoming more and more popular with the development of the Internet industry and software engineering. Many companies provide public Web APIs for their services, and developers can greatly accelerate the development of new applications by relying on such APIs to execute complex tasks without implementing the corresponding functionalities themselves. The proliferation of web APIs, however, also introduces a challenge for developers to search and discover the desired API and its endpoint. This is a practical and crucial problem because according to ProgrammableWeb, there are more than 22,000 public Web APIs each of which may have tens or hundreds of endpoints. Therefore, it is difficult and time-consuming for developers to find the desired API and its endpoint to satisfy their development needs. In this paper, we present an intelligent system for Web API searches based on

natural language queries by using a two-step transfer learning. To train the model, we collect a significant amount of sentences from crowdsourcing and utilize an ensemble deep learning model to predict the correct description sentences for an API and its endpoint. A training dataset is built by synthesizing the correct description sentences and then is used to train the two-step transfer learning model for Web API search. Extensive evaluation results show that the proposed methods and system can achieve high accuracy to search a Web API and its endpoint.

## **Session II: Web Security and Privacy** (September 18 - September 20, 2020, 7/24 Available)

Chair: Mr. Yasuhiko Kanemasa (Fujitsu Laboratories Ltd, Japan)

20 minutes for each paper.

### **A Secure and Efficient Smart Contract Execution Scheme** (ICWS2020-0004)

Zhaoxuan Li and Rui Zhang (Institute of Information Engineering, Chinese Academy of Sciences, China); Pengchao Li (Chinese Academy of Sciences, China)

Abstract: As a core technology of the blockchain, the smart contract is receiving increasing attention. However, the frequent outbreak of smart contract security events shows that improving the security of smart contracts is essential. How to guarantee the privacy of contract execution and the correctness of calculation results at the same time is still an issue to be resolved. Using secure multi-party computation (SMPC) technology to implement smart contracts is considered to be one of the potential solutions. But in the existing SMPC based contract execution schemes, a problem has been ignored, that is, the attacker can perform the same process as the reconstructor to recover the secret, which leads to the leakage of users' privacy. Therefore, in order to solve this problem in the process of smart contract operation, an improved homomorphic encryption algorithm is proposed in this paper, which has a relatively small public key size, short ciphertext length, and high encryption efficiency. Then, a contract execution scheme integrated with SMPC and homomorphic encryption (SMPC-HE for short) is further proposed, which is able to guarantee the privacy of contract execution and the correctness of the calculation results at the same time, and also makes smart contract execution fairer. Finally, our scheme is proved secure, efficient and has low space overhead by theory and experiment results.

### **A Reputation Based Hybrid Consensus for E-commerce Blockchain** (ICWS2020-0005)

You Sun and Rui Zhang (Institute of Information Engineering, Chinese Academy of Sciences, China); Rui Xue (Chinese Academy of Sciences, China); Qianqian Su (Institute of Information Engineering, Chinese Academy of Sciences, China); Pengchao Li (Chinese Academy of Sciences, China)

Abstract: Blockchain can achieve non-tampering, non-repudiation, consistency and integrity that other data management technologies do not have. Especially in peer-to-peer networks, the decentralized nature of blockchain has drawn tremendous attention from academic and industrial communities. Recently, the field of e-commerce has also begun to realize its important role. Although blockchain technology has many advantages in achieving trust establishment and data sharing among distributed nodes, in order to make it better to be applied in e-commerce, it is necessary to improve the security of transactions and the efficiency of consensus mechanisms. In this paper, we present a reputation based hybrid consensus to solve the problem of transaction security and efficiency. Our scheme integrates the reputation mechanism into transactions and consensus, and any improper behavior of nodes will be reflected in the reputation system and fed back to a new round of transactions and consensus. We implement distributed reputation management and enable users to append new reputation evaluations to the transaction that has previously evaluated. Meanwhile, we demonstrated that the scheme can defend against existing attacks such as selfish mining attacks, double spending attacks and flash attacks. We implement a prototype and the result shows that our scheme is promising.

### **A Contract based User-Centric Computational Trust towards E-governance** (ICWS2020-0006)

Bin Hu (University of Chinese Academy of Sciences, Beijing & Institute of Computing Technology, Chinese Academy of Sciences, China); Xiaofang Zhao and Cheng Zhang (Institute of Computing Technology, Chinese Academy of Sciences, China); Yan Jin (Chinese Academy of Sciences, Institute of Computing Technology, China); Bo Wei (The First Research Institute of the Ministry of Public Security, China)

Abstract: E-Government services are persistent targets of the organized crime by hackers, which hinders the delivery of services. Computational trust is an important technique for the security work of service providers (SPs). However, it relies on data collection about users' past behaviors conventionally from other SPs, which incurs the uncertainty of data and thus impacts the quality of data. Motivated by this issue, this paper proposes a novel smart contract based user-centric computational trust framework (UCCT) which collects the behavioral data of the user. It uses smart contract as a rational trustworthy agent to automatically monitor and manage the user's behaviors on the user side, so as to provide deterministic data quality assurance services for the computational trust. Furthermore, a privacy-preserving way of the data sharing is provided for the user and a personalized security mechanism for the SP. A new ledger is also introduced to provide a user-centric and efficient search. The results of experiments conducted on a Hyperledger Fabric based blockchain platform demonstrate that the time cost of user-centric ledger in UCCT can be less than 1 s. Moreover, even if a more complicated contract is provided, the improvement of transaction per second (TPS), which is made by UCCT, is not less than 8%.

### **Session III: Social Networking Services** (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Mohamed Adel Serhani (United Arab Emirates University, UAE)

20 minutes for each paper.

#### **CMU: Towards Cooperative Content Caching with User Device in Mobile Edge Networks** (ICWS2020-0007)

Zhenbei Guo and Fuliang Li (Northeastern University, China); Yuchao Zhang (Beijing University of Posts and Telecommunications, China); Changsheng Zhang (Northeastern University, China); Tian Pan (Beijing University of Posts and Telecommunications, China); Weichao Li and Yi Wang (Southern University of Science and Technology, China)

Abstract: Content caching in mobile edge networks has stirred up tremendous research attention. However, most existing studies focus on predicting content popularity in mobile edge servers (MESs). In addition, they overlook how the content is cached, especially how to cache the content with user devices. In this paper, we propose CMU, a three-layer (Cloud-MES-Users) content caching framework and investigate the performance of different caching strategies under this framework. A user device who has cached the content can offer the content sharing service to other user devices through device-to-device communication. In addition, we prove that optimizing the transmission performance of CMU is an NP-hard problem. We provide a solution to solve this problem and describe how to calculate the number of distributed caching nodes under different parameters, including time, energy and storage. Finally, we evaluate CMU through a numerical analysis. Experiment results show that content caching with user devices could reduce the requests to Cloud and MESs, and decrease the content delivery time as well.

#### **An FM Developer Recommendation Algorithm by Considering Explicit Information and ID Information** (ICWS2020-0008)

Xu Yu (School of Information Science and Technology, Qingdao University of Science and Technology, China); He Ya (China); Biao Xu (China University of Mining and Technology, China); Junwei Du and Feng Jiang (Qingdao University of Science and Technology, China); Dunwei Gong (China University of Mining and Technology, China)

Abstract: Recently, the developer recommendation on crowdsourcing software platform is of great research significance since an increasingly large number of tasks and developers have gathered on the platforms. In order to solve the problem of cold-start, the existing developer recommendation algorithms usually only use explicit information but not ID information to represent tasks and developers, which causes poor performance. In view of the shortcomings of the existing developer recommendation algorithms, this paper proposes an FM recommendation algorithm based on explicit to implicit feature mapping relationship modeling. This algorithm firstly integrates fully the ID information, explicit information and rating interaction between the completed task and the existing developers by using FM algorithm in order to get the implicit features related to their ID information. Secondly, for the completed tasks and existing developers, a deep regression model is established to learn the mapping relationship from explicit features to implicit features. Then, for the cold-start task or the cold-start developer, the implicit features are determined by the explicit features according to the deep regression model. Finally, the ratings in the cold-start scene can be predicted by the trained FM model with the explicit and implicit features. The simulation results on Topcoder platform show that the proposed algorithm has obvious advantages over the comparison algorithm in precision and recall.

#### **Characteristics of Similar-Context Trending Hashtags in Twitter: A Case Study** (ICWS2020-0009)

Eiman Alothali (United Arab Emirates University, United Arab Emirates); Abdul Kadhim Hayawi (Zayed University, United Arab Emirates); Hany Alashwal (United Arab Emirates University & College of Information Technology, United Arab Emirates)

Abstract: Twitter is a popular social networking platform that is widely used in discussing and spreading information on global events. Twitter trending hashtags have been one of the topics for researcher to study and analyze. Understanding the posting behavior patterns as the information flows increase by rapid events can help in predicting future events or detection manipulation. In this paper, we investigate similar-context trending hashtags to characterize general behavior of specific-trend and generic-trend within same context. We demonstrate an analysis to study and compare such trends based on spatial, temporal, content, and user activity. We found that the characteristics of similar-context trends can be used to predict future generic trends with analogous spatiotemporal, content, and user features. Our results show that more than 70% users participate in location-based hashtag belongs to the location of the hashtag. Generic trends aim to have more influence in users to participate than specific trends with geographical context. The retweet ratio in specific trends is higher than generic trends with more than 79%.

### **Session IV: Services Platform** (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Noseong Park (George Mason University, USA)

20 minutes for each paper.

#### **GraphInf: A GCN-based Popularity Prediction System for Short Video Networks** (ICWS2020-0010)

Yuchao Zhang (Beijing University of Posts and Telecommunications, China); Pengmiao Li (Beijing University of Post

*and Telecommunication, China); Zhi-Li Zhang (University of Minnesota, USA); Wendong Wang (Beijing University of Posts and Telecommunications, China); Chaorui Zhang (Huawei Technologies Co., Ltd., Hong Kong); Yishuang Ning (Tsinghua University, China); Bo Lian (Kuaishou, China)*

**Abstract:** As the emerging entertainment applications, short video platforms, such as Youtube, Kuaishou, quickly dominant the Internet multimedia traffic. The caching problem will surely provide a great reference to network management (e.g., traffic engineering, content delivery). The key to cache is to make precise popularity prediction. However, different from traditional multimedia applications, short video network exposes unique characteristics on popularity prediction due to the explosive video quantity and the mutual impact among these countless videos, making the state-of-the-art solutions invalid. In this paper, we first give an in-depth analysis on 105,231,883 real traces of 12,089,887 videos from Kuaishou Company, to disclose the characteristics of short video network. We then propose a graph convolutional neural-based video popularity prediction algorithm called GraphInf. In particular, GraphInf clusters the countless short videos by region and formulates the problem in a graph-based way, thus addressing the explosive quantity problem. GraphInf further models the influence among these regions with a customized graph convolutional neural (GCN) network, to capture video impact. Experimental results show that GraphInf outperforms the traditional Graph-based methods by 44.7%. We believe such GCN-based popularity prediction would give a strong reference to related areas.

#### **Finding Performance Patterns from Logs with High Confidence (ICWS2020-0011)**

*Joshua Kimball, Rodrigo Alves Lima and Calton Pu (Georgia Institute of Technology, USA)*

**Abstract:** Performance logs contain rich information about a system's state. Large-scale web service infrastructures deployed in the cloud are notoriously difficult to troubleshoot, especially performance bugs. Detecting, isolating and diagnosing fine-grained performance anomalies requires integrating system performance measures across space and time. To achieve scale, we present our megatables approach, which automatically interprets performance log data and outputs millibottleneck predictions along with supporting visualizations. We evaluate our method with three illustrative scenarios, and we assess its predictive ability. We also evaluate its ability to extract meaningful information from many log samples drawn from the wild.

#### **Keyphrase Extraction in Scholarly Digital Library Search Engines (ICWS2020-0012)**

*Krutarth I Patel (Kansas State University, USA); Cornelia Caragea (University of Illinois at Chicago, USA); Jian Wu (Old Dominion University, USA); C. Lee Giles (Pennsylvania State University, USA)*

**Abstract:** Scholarly digital libraries provide access to scientific publications and comprise useful resources for researchers who search for literature on specific subject areas. CiteSeerX is an example of such a digital library search engine that provides access to more than 10 million academic documents and has nearly one million users and three million hits per day. Artificial Intelligence (AI) technologies are used in many components of CiteSeerX including Web crawling, document ingestion, and metadata extraction. CiteSeerX also uses an unsupervised algorithm called noun phrase chunking (NP-Chunking) to extract keyphrases out of documents. However, often NP-Chunking extracts many unimportant noun phrases. In this paper, we investigate and contrast three supervised keyphrase extraction models to explore their deployment in CiteSeerX for extracting high quality keyphrases. To perform user evaluations on the keyphrases predicted by different models, we integrate a voting interface into CiteSeerX. We show the development and deployment of the keyphrase extraction models and the maintenance requirements.

### **Session V: Services Architecture (September 18 - September 20, 2020, 7/24 Available)**

Chair: Mr. Yasuhiko Kanemasa (Fujitsu Laboratories Ltd, Japan)

20 minutes for each paper.

#### **Reducing the Cost of Aggregation in Crowdsourcing (ICWS2020-0013)**

*Loïc Hédouët (INRIA Rennes, France); Rituraj Singh (Univ Rennes/INRIA/IRISA, France); Zoltan Miklos (University Rennes 1, France)*

**Abstract:** Crowdsourcing is a way to solve problems that need human contribution. Crowdsourcing platforms distribute replicated tasks to workers, pay them for their contribution, and aggregate answers to produce a reliable conclusion. A fundamental problem is to infer a correct answer from the set of returned results. Another challenge is to obtain a reliable answer at a reasonable cost: unlimited budget allows hiring experts or large pools of workers for each task but a limited budget forces to use resources at best. This paper considers crowdsourcing of simple boolean tasks. We first define a probabilistic inference technique, that considers difficulty of tasks and expertise of workers when aggregating answers. We then propose CrowdInc, a greedy algorithm that reduce the cost needed to reach a consensual answer. CrowdInc distributes resources dynamically to tasks according to their difficulty. We show on several benchmarks that CrowdInc achieves good accuracy, reduces costs, and we compare its performance to existing solutions.

#### **Scheduling Multi-workflows over Edge Computing Resources with Time-varying Performance, a Novel probability-mass function and DQN-based Approach (ICWS2020-0014)**

*Hang Liu and Yuyin Ma (Chongqing University, China); Peng Chen (Xihua University, China); Yunni Xia (Chongqing*



*University, China); Wanbo Zheng (Kunming University of Science and Technology, China); Xiaobo Li (Chongqing Animal Husbandry Techniques Extension Center, China)*

Abstract: The edge computing paradigm is featured by the ability to off-load computing tasks from mobile devices to edge clouds and provide high cost-efficient computing resources, storage and network services closer to the edge. A key question for workflow scheduling in the edge computing environment is how to guarantee user-perceived quality of services when the supporting edge services and resources are with unstable, time-variant, and fluctuant performance. In this work, we study the workflow scheduling problem in the multi-user edge computing environment and propose a Deep-Q-Network (DQN) -based multi-workflow scheduling approach which is capable of handling time-varying performance of edge services. To validate our proposed approach, we conduct a simulative case study and compare ours with other existing methods. Results clearly demonstrate that our proposed method beats its peers in terms of convergence speed and workflow completion time.



# 2020 International Conference on Cloud Computing (CLOUD 2020)

---

## Session I: Cloud Security and Privacy (September 18 - September 20, 2020, 7/24 Available)

Chair: Dr. Qi Zhang (IBM Thomas J. Watson Research, USA)

20 minutes for each paper.

### [A Network Intrusion Detection Approach Based on Asymmetric Convolutional Autoencoder \(CLOUD2020-1001\)](#)

*Shujian Ji (University of Chinese Academy of Sciences & SIAT, China); Chengzhong Xu (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, USA); Kejiang Ye (Chinese Academy of Sciences, China)*

Abstract: With the continuous development of Internet technology, cyberspace security protection technology is still facing many new intrusion threats. Aiming at the complex and changeable network abnormal traffic and intrusion behavior, we apply the deep learning approach to the field of intrusion detection. In this paper, we detail our proposed asymmetric convolutional autoencoder (ACAE) for feature learning. In addition, we propose a network intrusion detection model based on the combination of asymmetric convolutional autoencoder and random forest. This approach can well combine the advantages of deep learning and shallow learning. Our proposed approach is evaluated on KDD99 and NSL-KDD dataset, and compared with other approaches in the field of intrusion detection. Our model can effectively improve the classification accuracy of network abnormal traffic and it has strong robustness and scalability in current NIDSs.

### [A Replication Study to Explore Network-Based Co-Residency of Virtual Machines in the Cloud \(CLOUD2020-1002\)](#)

*Sanchay Gupta and Robert Miceli (Johns Hopkins University, USA); Joel Coffman (United States Air Force Academy & Johns Hopkins University, USA)*

Abstract: By deploying virtual machines (VMs) on shared infrastructure in the cloud, users gain flexibility, increase scalability, and decrease their operational costs compared to on-premise infrastructure. However, a cloud environment introduces new vulnerabilities, particularly from untrusted users sharing the same physical hardware. In 2009, Ristenpart et al. demonstrated that an attacker could place a VM on the same physical hardware and extract confidential information from a target using a side-channel attack. We replicated this seminal work on cloud cartography and network-based co-residency tests on Amazon Web Services (AWS) and OpenStack cloud infrastructures. Although the AWS Elastic Compute Cloud (EC2) cloud cartography remains similar to prior work, current mitigations deter the network-based co-residency tests. OpenStack's cloud cartography differs from AWS EC2's, and we found that OpenStack was vulnerable to one network-based co-residency test. Our results indicate that co-residency threats remain a concern more than a decade after their initial description.

### [Analyzing CNN Based Behavioural Malware Detection Techniques in Cloud IaaS \(CLOUD2020-1003\)](#)

*Andrew D McDole (Tennessee Technological University, USA); Mahmoud Abdelsalam (Manhattan College, USA); Maanank Gupta (Tennessee Technological University, USA); Sudip Mittal (University of North Carolina Wilmington, USA)*

Abstract: Cloud Infrastructure as a Service (IaaS) is vulnerable to malware due to its exposure to external adversaries, making it a lucrative attack vector for malicious actors. A datacenter infected with malware can cause data loss and/or major disruptions to service for its users. This paper analyzes and compares various Convolutional Neural Networks (CNNs) for online detection of malware in cloud IaaS. The detection is performed based on behavioural data using process level performance metrics including CPU usage, memory usage, disk usage etc. We have used the state of the art DenseNets and ResNets in effectively detecting malware in online cloud system. CNN are designed to extract features from data gathered from a live malware running on a real cloud environment. Experiments are performed on OpenStack (a cloud IaaS software) testbed designed to replicate a typical 3-tier web architecture. Comparative analysis is performed for different metrics for different CNN models used in this research.

## Session II: Cloud Application Development (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Yingwei Wang (University of Prince Edward Island, Canada)

20 minutes for each paper.

### [Abstracting Containerisation and Orchestration for Cloud-Native Applications \(CLOUD2020-1004\)](#)

*Jose Quenum (NUST, Namibia); Gervasius Ishuuwa (Namibia University of Science and Technology, Namibia)*

**Abstract:** Developing cloud-native applications demands a radical shift from the way we design and build traditional applications. Application designers usually divide their business logic into several business functions, each developed according to a microservices architectural style and packaged in containers. Throughout the stages of cloud-native application development (development, testing, staging and production), container orchestration helps coordinate the execution environment. Thanks to the increasing popularity of cloud-native applications, there has been a growing interest in container and orchestration technologies recently. However, despite their closeness, these two inter-related technologies are supported by different toolsets and specification formalisms, with minimal portability between them and usually a disregard for the best practices. This paper presents *velo*, a domain-specific language (DSL) that unifies containerisation and orchestration concepts. *velo* has two components: (1) a specification language that supports an abstract description of containerisation and orchestration for a complex application; and (2) a transpiler, a source-to-source compiler into concrete container manifest and orchestration description.

#### **Scalable Sketch-based Sport Video Retrieval in the Cloud (CLOUD2020-1005)**

*Ihab Al Kabary and Heiko Schuldt (University of Basel, Switzerland)*

**Abstract:** Content-based video retrieval in general and in sport videos in particular has attracted an increasing interest in the past few years, due to the growing interest in sports analytics. Especially sketch-based queries, enabling spatial search in video collections, are increasingly being demanded by coaches and analysts in team sports as an essential tool for game analysis. Although there has been great progress in the last years in the field of sketch-based retrieval in sports, most approaches focus on functional aspects and only consider just a very limited number of games. The problem is to scale these systems to allow for interactive video retrieval on a large game collection, beyond single games. In this paper, we show how SportSense, our sketch-based video retrieval system, can be deployed and scaled-out in the Cloud, allowing managers and analysts to interactively search for scenes of their choice within a large collection of games. In our evaluations, we show how the system can scale to a collection of the size of an entire season with response times that enable real-time analysis.

#### **Online Tutoring through a Cloud-based Virtual Tutoring Center (CLOUD2020-1006)**

*Xiaolin Hu, Sai Tabdil, Manisha Achhe and Yi Pan (Georgia State University, USA); Anu Bourgeois (GSU, USA)*

**Abstract:** Online tutoring has gained popularity in recent years, which allows students to get tutoring service from tutors in an online, virtual environment. To support effective online tutoring, an integrated online tutoring system is essential. This paper presents a cloud-based virtual tutoring center that supports online tutoring for college students to complement the tutoring service of physical tutoring centers on a campus. We present the overall architecture of the virtual tutoring center system and show preliminary results of using the virtual tutoring center to provide online tutoring for computer science students.

### **Session III: Cloud Infrastructure (September 18 - September 20, 2020, 7/24 Available)**

Chair: Dr. Qi Zhang (IBM Thomas J. Watson Research, USA)

20 minutes for each paper.

#### **A Resource Trend Analysis from a Business Perspective based on a Component Decomposition Approach (CLOUD2020-1007)**

*Yuji Saitoh, Tetsuya Uchiumi and Yukihiro Watanabe (Fujitsu Laboratories Ltd., Japan)*

**Abstract:** To ensure reliability for information and communication technology (ICT) systems, it is important to analyze resource usage for the purpose of provisioning resources, detecting failures, and so on. It is more useful to understand the resource usage trends for each business process because generally multiple business processes run on an ICT system. For example, we can detect an increase in resource usage for a specific business process. However, conventional methods have not been able to analyze such trends because resource usage data is usually mixed and cannot be separated. Therefore, in the previous work, we proposed an analysis method that decomposes the data into components of each business process. The method successfully analyzed only single sources of data. However, an actual system consists of multiple resources and multiple devices. Therefore, in this paper, we enhance this method so that it is able to analyze multiple sources of data by incorporating a technique for unifying multiple sources of data into a single sources of data on the basis of a workload dependency model. In addition, the proposed method can also analyze the relationship between resources and application workloads. Therefore, it can identify specific applications that cause resource usage to increase. We evaluated the proposed method by using the data of on-premise and actual commercial systems, and we show that it can extract useful business trends. The method could extract system-wide processes, such as file-copy between two servers, and identify a business event corresponding to a resource usage increase.

#### **Job Completion Time in Dynamic Vehicular Cloud Under Multiple Access Points (CLOUD2020-1008)**

*Aida Ghazizadeh (Old Dominion University, USA); Puya Ghazizadeh (St. John's University, USA); Stephan Olariu (Old Dominion University, USA)*

**Abstract:** Vehicular cloud is a group of vehicles whose corporate computing, sensing, communication and physical resources can be coordinated and dynamically allocated to authorized users. One of the attributes that set vehicular clouds

apart from conventional clouds is re- source volatility. As vehicles enter and leave the cloud, new compute resources become available while others depart, creating a volatile environment where the task of reasoning about fundamental performance metrics such as job completion time becomes very challenging. In general, predicting job completion time requires full knowledge of the probability distributions of the intervening random variables. However, the datacenter manager does not know these distribution functions. Instead, using accumulated empirical data, she may be able to estimate the first moments of these random variables. In this work we offer approximations of job completion time in a dynamic vehicular cloud model involving vehicles on a highway where jobs can be downloaded under multiple stations.

**ZTIMM: A Zero-Trust-Based Identity Management Model for Volunteer Cloud Computing** (CLOUD2020-1009)

*Abdullah Albuali (Southern Illinois University at Carbondale, USA); Tessema Mengistu (Southern Illinois University, USA); Dunren Che (Southern Illinois University at Carbondale, USA)*

**Abstract:** The availability of efficient, affordable, green computing alternatives for emerging paradigms, such as edge, mobile, and volunteer computing, is growing. But these models are plagued by security issues, such as in volunteer computing, where trust among entities is nonexistent. This paper presents a zero-trust model that assumes no trust for any volunteer node and always verifies using a server-client model. This paper proposes an adaptive behavior evaluation model that estimates each VN's trust score. Therefore, tasks are assigned to the most trusted VN in the system. A series of trust-aware mechanisms to handle VN's life cycle is also presented.

**Session IV: Cloud Performance Management** (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Yingwei Wang (University of Prince Edward Island, Canada)

20 minutes for each paper.

**PLMSys: A Cloud Monitoring System Based on Cluster Performance and Container Logs** (CLOUD2020-1010)

*Yongzhong Sun (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China); Kejiang Ye (Chinese Academy of Sciences, China); Cheng-Zhong Xu (University of Macau, China)*

**Abstract:** The cloud platforms aggregate a lot of computing resources and adopt Docker technology to achieve dynamic resource scaling and provide services on demand. Docker, a kind of lightweight virtualization technology which has the characteristics of resource isolation, rapid deployment and low cost, is widely used in the construction of the cloud services so that Docker-based containers has become the important basis of core cloud businesses. Therefore, in order to manage the large-scale cloud cluster and enforce the quality of cloud services for consumers, Monitoring mechanism for the container-based clouds are indispensable. In this paper, we design and implement a cloud monitoring system - PLMSys based on cluster performance and container logs. It provides the following functions: (1) Multi-dimensional resources monitoring. PLMSys can monitor the running state of the cluster host and container, including the utilization of CPU, memory, disk and other resources. (2) Container log collection. PLMSys can centrally collect the logs generated by all containers of the cluster. (3) Rule-based exception alerts. PLMSys allows users to define the abnormal state of the host and container by creating rules, and provides multiple alerting methods. (4) Workload analysis and prediction. PLMSys extracts the descriptive statistics from the cluster workload and uses the time series models to predict the future workload. (5) Data monitoring visualization. The system uses rich visual charts to reflect the running state of cluster host and container. By using PLMSys, users can better manage cluster hosts and containers.

**Wise Toolkit: Enabling Microservice-Based System Performance Experiments** (CLOUD2020-1011)

*Rodrigo Alves Lima and Joshua Kimball (Georgia Institute of Technology, USA); J. E. (Joao Eduardo) Ferreira (University of Sao Paulo, Brazil); Calton Pu (Georgia Institute of Technology, USA)*

**Abstract:** In this paper, we present the Wise toolkit for microservice-based system performance experiments. Wise comprises a microservice-based application benchmark with controllable workload generation; milliScope, a set of system resource and event monitoring tools; and WED-Make, a workflow language and code generation tool for the construction and execution of system experiments with automatic provenance collection. We also show a running example reproducing the experimental verification of the millibottleneck theory of performance bugs to illustrate how we have used Wise for the performance study of microservice-based benchmark applications in the cloud.

**Capestor: a Service Mesh-based Capacity Estimation Framework for Cloud Applications** (CLOUD2020-1012)

*Yao Sun (Nanjing Institute of Big Data, Jinling Institute of Technology, China); Lun Meng (Hohai University, China); Shudong Zhang (Capital Normal University, China)*

**Abstract:** Due to complex deployment configurations and application logics in cloud computing, microservice architectures are widely used to divide cloud applications into multiple independent microservices communicating with each other through APIs that are not associated with languages and platforms. However, a large amount of various microservices make it difficult for operators to predict the performance of microservices and then estimate cloud applications' capacity. This

paper proposes a capacity estimation framework called as Capestor for cloud applications based on a service mesh. Capestor employs a service mesh to place target microservices in isolated containers, simulates workloads, and collect monitoring data related with performance and resources. Then, Capestor employs a lasso regression model to correlate resources and performance, and estimates the capacity of each microservice to plan fine-grained flexible expansion. Finally, we evaluate Capestor with a typical microservice based application. The experimental results show that Capestor can estimate the capacity of microservices, and provide performance guarantee for applications with low prediction error.

#### **Session V: Cloud Storage** (September 18 - September 20, 2020, 7/24 Available)

Chair: Dr. Qi Zhang (IBM Thomas J. Watson Research, USA)

20 minutes for each paper.

##### **NLC: A Replacement Caching Algorithm based on Non-Critical Path Least Counts for In-Memory Computing** (CLOUD2020-1013)

*JingYa Lv (Shenzhen Institutes of Advanced Technology, China); Yang Wang (Shenzhen Institute of Advanced Technology, China)*

**Abstract:** The explosion of applications in data-parallel systems and ever-growing high-efficiency needs for task processing and data analysis have made parallel systems under enormous memory pressure when dealing with large datasets. Out-of-memory errors and excessive garbage collection can seriously affect the system performance. Generally, for those data-flow tasks with intensive in-memory computing requirements, how to achieve efficient memory caching algorithms is a primary measure to make a trade-off between performance and memory overhead. By taking advantage of existing researches on the DAG-based task scheduling, we design a new caching algorithm for in-memory computing by exploiting the critical path information of DAG, called Non-critical path least reference count (NLC). The strategy is distinct from the existing ones in that it applies the global information of the critical path to the caching replacements rather than the task scheduling as most existing works do. Through empirical studies, we demonstrated that NLC can not only effectively enhance the parallel execution efficiency, but also reduce the number of evictions, improve the hit ratio, and memory utilization rate as well. Our comprehensive evaluations based on the selected benchmark graphs indicate that our strategy can not only fulfil the parallel system requirements but also reduce the costs by as much as 19%, compared with the most advanced LRC algorithm.

##### **Manage Storage Resources Efficiently Using Predictive Analytics** (CLOUD2020-1014)

*Gazanfur A Mohammed (Cisco Systems, USA)*

**Abstract:** Univariate time series forecasting can be used to dynamically tune the resource allocation for databases. It is vital that relational databases have adequate storage for archive logs since lack of space can cause the database to hang, while overallocation can reduce the efficiency of utilization. Most of the time, storage is allocated for the peak usage and kept for the duration of its lifecycle. This paper presents a conceptual model that uses predictive analysis to dynamically scale the storage allocation for archive logs generated by databases. The framework presented in this paper does exploratory data analysis on archive logs data, compares the accuracy of various statistical models such as autoregressive integrated moving average (ARIMA), Holt damped trend, Holt linear trend, Mean, Naïve and Seasonal Naïve models, suggests the best model suited for each database, and provides forecast of storage usage. These predictions can be used as input in other automated systems to automatically provision the storage or repurpose unused storage as needed.

##### **A Utility-Based Fault Handling Approach for Efficient Job Rescue in Clouds** (CLOUD2020-1015)

*Fei Xie, Jun Yan and Jun Shen (University of Wollongong, Australia)*

**Abstract:** In recent years, many organizations face challenges when managing large amount of data and data-intensive computing tasks. Cloud computing technology has been widely-used to alleviate these challenges with its on-demand services and distributed architecture. Data replication is one of the most significant strategies to decrease the access latency and improve data availability, data reliability and resource utilization by creating multiple data copies to geographically-distributed data centers. When a fault occurs at a data center, existing jobs that require data access in this data center can be redirected to other data centers, where data replicas are available. This paper proposes a utility-based fault handling (UBFH) approach to rescue the jobs at the faulty data center. Then a fault handling algorithm is developed to determine the direction of job redirection by considering the network performance and job attributes. Our main objective is to achieve better repairability, job rescue utility and job operation profit. The simulation results show that our UBFH approach outperforms HDFS, RR and JSQ approaches in all these aspects.

#### **Session VI: Cloud and Distributed Systems** (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Yingwei Wang (University of Prince Edward Island, Canada)

20 minutes for each paper.

##### **Bioinf-PHP: Bioinformatics Pipeline for Protein Homology and Phylogeny** (CLOUD2020-1016)

*Michael Zhou (Skyline High School, USA); Yongsheng Bai (Next-Gen Intelligent Science Training, USA)*

**Abstract:** Catalase is a special category of enzyme that plays a critical role in regulating the level of harmful hydrogen peroxide in cells. There are three main families of these proteins: Typical Catalases, Catalase-Peroxidases (katG), and Manganese Catalases. In order to uncover potential evolutionary relationships between these enzymes, we have developed a bioinformatics pipeline named Bioinf-PHP to search for protein homology and phylogeny, and to compare these three catalase families at the functional level based on sequence similarity. Protein motif analysis of the sequences featured in the pipeline were conducted using the MEME algorithm. The top three significant motifs were reported for all of the catalase sequences. The Bioinf-PHP pipeline also runs BLASTP to search for homology between bacteria catalase and yeast protein sequences. The neighbor-joining phylogenetic tree was constructed with *Saccharomyces cerevisiae* to infer evolutionary relationships as a test example. The structural similarities between orthologous sequences provided further evidence of functional similarity.

**Towards multi-criteria volunteer cloud service selection (CLOUD2020-1017)**

*Yousef Alsenani, Garth V Crosby, Khaled R. Ahmed and Tomas Velasco (Southern Illinois University Carbondale, USA)*

**Abstract:** Volunteer cloud computing (VCC) have recently been introduced to provide low-cost computational resources to support the demands of the next generation IoT applications. The vital process of VCC is to provide on demand resource provisioning and allocation in response to resource failures, behavior of volunteers (donors, users) and dynamically changing workloads. Most existing work addresses each of these factors (reliability, trust, and load) independently. Finding the most reliable machine (e.g., the lowest hardware failure rate) does not guarantee that the machine is trustworthy or not loaded, and vice versa. To address these problems, this research proposed a model to select volunteer node (VN) based on three criteria: the trustworthiness of the volunteer, the reliability of the node, and the resource load. We use three different models to estimate the three factors. We used exponential distribution reliability to estimate the reliability of VN and neural network to predict VN resource usages. . In addition, we propose a new version of the beta function to estimate trustworthiness. Then we apply multiple regression to weigh each factor and decide which factor will be most effective for preventing task failure. Finally, a VN is selected based on multiple criteria decision analysis.

**Cardinality based rate limiting system for time-series data (CLOUD2020-1018)**

*Deepak K Vasthimal (eBay Inc, USA); Sudeep Kumar (eBay Inc., USA)*

**Abstract:** Massive monitoring systems that require high availability and performance for both ingestion and retrieval of data are often encountered with rogue streams of data having a high cardinality. The management of such high cardinality data sets for time-series data and a performance sensitive system is challenging. The challenges primarily arise as the time-series data sets, typically needs to be loaded onto a limited memory space before results can be returned to the client. This affects the number of incoming queries that can be supported simultaneously. Too many time-series can potentially degraded read performance and thereby affect user experience. Our proposed rate-limiting system described herein seeks to address a key availability issue on a high-volume, time-series system by using a dynamic cardinality computation in combination with a central assessment service to detect and block high cardinality data streams. As a result of this technical improvement, anomalous logging behavior is detected quickly, affected tenants are notified, and hardware resources are used optimally.

**Session VII: Cloud Resources Optimization (September 18 - September 20, 2020, 7/24 Available)**

Chair: Dr. Qi Zhang (IBM Thomas J. Watson Research, USA)

20 minutes for each paper.

**A new personalized POI recommendation based on time-aware and social influence (CLOUD2020-1019)**

*Nan Wang (Heilongjiang University, China); Yong Liu and Peiyao Han (Heilongjiang University, Harbin, China); Xiaokun Li (Postdoctoral Program of Heilongjiang Hengxun Technology Co., Ltd., China); Jinbao Li (Heilongjiang University, Harbin, China)*

**Abstract:** With the rapid growth of the location-based social networks (LBSNs), Point of Interest (POI) recommendation has become an important research topic in data mining. However, the existing works do not reasonably utilize the time sensitivity of POI recommendations and have not taken full account of the user's behavior preferences at different time, causing the POI recommendation performance poor. We propose a Time-aware and POI Recommendation model based on Tensor Factorization, named TPR-TF. Firstly, we study the POI recommendation problem of time sensitivity and propose a temporal dynamic segmentation algorithm based on hierarchical clustering. Through dividing the fine grain of time, the experiment result is more reasonable and effectively than the previous method which divided identical time empirically. Secondly, by combining the time-aware recommendation with the influence of the user's direct friendship and potential friendship, we expand the scope of users' social influence, and then further improve the POI recommendation performance. Experimental results on the two datasets indicate that our TPR-TF model is superior to the current mainstream POI recommendation models both in precision and recall.

**Toward a Decentralized Service Marketplace: The Interplay Between Blockchain and Algebraic Service Composition (CLOUD2020-1020)**

*Chen Qian (Donghua University, China); Wenjing Zhu (Institute of Scientific and Technical Information of Shanghai,*



*Shanghai Library, China)*

Abstract: Service marketplaces are supposed to guarantee an open platform for sellers and customers of cloud services. But their potentials cannot be fully released, due to the widely known shortcomings including but not limited to central power of authority, data privacy, lack of customization, rigid and complex trading procedure. We argue that decentralized marketplaces, although not mature, are the most promising solution to address these issues. In this paper, we present our work in progress, which is oriented toward a blockchain-enabled marketplace for sharing services at different levels of granularity in a flexible and trustworthy manner.

### **Leveraging Federated Clouds For Improving Blockchain Performance (CLOUD2020-1021)**

*Rkn Sai Krishna (Teradata India Pvt Ltd & Sri Sathya Sai Institute of Higher Learning, India); Chandrasekhar Tekur (Teradata India Pvt Ltd, India); Ravi Mukkamala (Old Dominion University, USA); Pallav Kumar Baruah (Sri Sathya Sai Institute of Higher Learning, India)*

Abstract: Today, blockchains are popular due to their promise of maintaining a tamper-proof and reliable distributed ledger. As more and more organizations are trying to incorporate this technology within their applications, they are finding the performance and scalability of these systems to be significant bottlenecks. The emerging federated clouds, with a conglomeration of autonomous clouds, could be a solution for this problem. Integrating federated clouds within a blockchain can help overcome these bottlenecks without undermining the autonomy and the peer-to-peer characteristic of the underlying systems. In this paper, we identify a few of the roles in which a federated cloud can play an effective role in a blockchain. In particular, we focus on its role in reducing the communication cost, in cooperation with the miners, incurred in propagating the mined blocks to all the nodes in the system. The proposed integrated approach results in reduced communication cost and improved time for synchronization of the distributed ledger at different nodes. The reduction in network traffic and latency in block synchronization is achieved without deviating from the fundamental autonomous nature of blockchains. The proposed approach results in higher transaction throughputs and lower query response times.

### **Post-cloud Computing Models and Their Comparisons (CLOUD2020-1022)**

*Yingwei Wang (University of Prince Edward Island, Canada); Parimala Thulasiraman (University of Manitoba, Canada)*

Abstract: In this paper, four computing models, CDEF (Cloudlet, Dew computing, Edge computing, and Fog computing), were portrayed and compared; the concept of post-cloud computing was examined and defined. This paper tries to clarify the connotation and denotation of each post-cloud computing model and to help users to choose the proper one for further exploration.

# 2020 International Conference on Big Data (BigData 2020)

---

## Session I: Big Data Analytics I (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Surya Nepal (CSIRO, Australia)

20 minutes for each paper.

### [A Web Application for Feral Cat Recognition through Deep Learning \(BIGDATA2020-2001\)](#)

*Richard Sinnott (University of Melbourne, Australia)*

Abstract: Deep learning has gained much attention and been applied in many different fields. In this paper, we present a web application developed to identify and detect the number of distinct feral cats of Australia using deep learning algorithms targeted to data captured from a set of remote sensing cameras. Feral cat recognition is an especially challenging application of deep learning since the cats are often similar and, in some cases, differ only in very small patterns on their fur. Given the automated, sensor-based image capture from remote cameras, further challenges relate to the limited number of images available. To tackle this, we train four neural network models to distinguish 75 classes (i.e. distinct feral cats) using 30 to 80 images for each class. Based on a range of evaluation metrics, we select Mask R-CNN model with ImageNet pre-trained weights augmented with the ResNet-50 network as the basis for the web application. Using images of cats from cameras in five different forests around Victoria, we achieved an average accuracy of identification of individual (distinct) cats of 89.4% with a maximum accuracy of 96.3%. This work is used to support ecologists in the School of Biosciences at the University of Melbourne.

### [Scalable reference genome assembly from compressed pan-genome index with Spark \(BIGDATA2020-2002\)](#)

*Altii Ilari Maarala (University of Helsinki, Finland); Ossi Arasalo (Aalto University, Finland); Daniel Valenzuela (University of Helsinki, Finland); Keijo Heljanko (University of Helsinki & HIIT, Finland); Veli Mäkinen (University of Helsinki, Finland)*

Abstract: High-throughput sequencing (HTS) technologies have enabled rapid sequencing of genomes and large-scale genome analytics with massive data sets. Traditionally, genetic variation analyses have been based on the human reference genome assembled from a relatively small human population. However, genetic variation could be discovered more comprehensively by using a collection of genomes i.e., pan-genome as a reference. The pan-genomic references can be assembled from larger populations or a specific population under study. Moreover, exploiting the pan-genomic references with current bioinformatics tools requires efficient compression and indexing methods. To be able to leverage the accumulating genomic data, the power of distributed and parallel computing has to be harnessed for the new genome analysis pipelines. We propose a scalable distributed pipeline, PanGenSpark, for compressing and indexing pan-genomes and assembling a reference genome from the pan-genomic index. We experimentally show the scalability of the PanGenSpark with human pan-genomes in a distributed Spark cluster comprising 448 cores distributed to 26 computing nodes. Assembling a consensus genome of a pan-genome including 50 human individuals was performed in 215 minutes and with 500 human individuals in 1468 minutes. The index of 1.41 TB pan-genome was compressed into a size of 164.5 GB in our experiments.

### [MCF: Towards Window-based Multiple Cuckoo Filters in Stream Computing \(BIGDATA2020-2003\)](#)

*Ziyue Hu and Menglu Wu (Shenzhen Institutes of Advanced Technology, China); Xiaopeng Fan (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China); Yang Wang (Shenzhen Institute of Advanced Technology, China); Cheng-Zhong Xu (University of Macau, China)*

Abstract: In this paper, we present a new stream-oriented filter, named Multiple Cuckoo Filter (MCF), to support concise representation and approximate membership queries for multiple sets with unpredicted cardinality. MCF is composed of a group of standard cuckoo filters with the same set dimension, in which the representation and membership query are decomposed into single set representation and query. MCF allows each cuckoo filter to be configured dynamically with changing sliding windows, and compare the fingerprints for the same key with each other, in order to find out whether a given item exists in multiple sets simultaneously. It is proved that MCF outperforms CF in false positive theoretically. Experiments demonstrate that the query delay of MCF grows linearly with the number of cuckoo filters, decreases gradually with the increments of the window size, and increases with the growth of set cardinality piecewise.

## Session II: Big Data Analytics II (September 18 - September 20, 2020, 7/24 Available)

Chair: Dr. Wenqi Cao (Facebook, Inc, USA)

20 minutes for each paper.

### **Spatial Association Pattern Mining using In-Memory Computational Framework (BIGDATA2020-2004)**

*Jin Soung Yoo, Wentao Shao and Kanika Binzani (Purdue University Fort Wayne, USA)*

Abstract: Spatial association pattern mining is a useful spatial data mining task for discovering interesting relationship patterns of spatial features based on spatial proximity. Spatial data mining is known as data-intensive computing. The explosive growth of spatial data demands computationally efficient methods for analyzing large complex data. Parallel and distributed computing is effective for large-scale data mining algorithms. This paper presents spatial association pattern mining with Spark which is a specially-designed in-memory parallel computing platform. The performance of Spark-based method proposed is compared with its corresponding MapReduce-based method.

### **A Performance Prediction Model for Spark Applications (BIGDATA2020-2005)**

*Muhammad Usama Javaid (Eura Nova, Belgium); Florian Demesmaeker, Amir Kanoun, Sabri Skhiri and Amine Ghrab (Co-author, Belgium)*

Abstract: Apache Spark is a popular open-source distributed processing framework that enables efficient processing of massive amounts of data. It has a large number of parameters that need to be tuned to get the best performance. However, tuning these parameters manually is a complex and time-consuming task. Therefore, a robust performance model to predict applications execution time could greatly help in accelerating the deployment and optimization of big data applications relying on Spark. In this paper, we ran extensive experiments on a selected set of Spark applications that cover the most common workloads to generate a representative dataset of execution time. In addition, we extracted application and data features to build a machine learning based performance model to predict Spark applications execution time. The experiments show that boosting algorithms achieved better results compared to the other algorithms.

### **Chemical XAI to Discover Probable Compounds' Spaces based on Mixture of Multiple Mutated Exemplars and Bioassay Existence Ratio (BIGDATA2020-2006)**

*Takashi Isobe (Hitachi High Tech America, Inc., USA)*

Abstract: Chemical industry pays much cost and long time to develop a new compound having aimed biological activity. On average, 10,000 candidates are prepared for each successful compound. The developers need to efficiently discover initial candidates before actual synthesis, optimization and evaluation. We developed a similarity-based chemical XAI system to discover probable compounds' spaces based on mixture of multiple mutated exemplars and bioassay existence ratio. Our system piles up 4.4k exemplars and 100M public DB compounds into vectors including 41 features. Users input two biologically active sets of exemplars customized with differentiated features. Our XAI extracts compounds' spaces simultaneously similar to multiple customized exemplars using vectors' distances and predicts their biological activity and target with the probability shown as existence ratio of bioassay that is the information of biological activity and target obtained from public DB or literature including related specific text string. The basis of prediction is explainable by showing biological activity and target of similar compounds included in the extracted spaces. The mixture of multiple mutated exemplars and bioassay existence ratio shown as probability with the basis of prediction can help the developers extract probable compounds' spaces having biological activity from unknown space. The response time to extract the spaces between two sets of 128 exemplars and 100M public DB compounds was 9 minutes using single GPU with HDD read and 1.5 minutes on memory. The bioassay existence ratio of extracted spaces was 2 - 9 times higher than the average of public ones. The correlation coefficient between predicted and actual pIC50 intensity of biological activity was 0.85 using randomly selected 64 compounds. Our XAI discovered probable compounds' spaces from large space at high speed and probability.

### **Session III: Big Data Modeling (September 18 - September 20, 2020, 7/24 Available)**

Chair: Prof. Aziz Nasridinov (Chungbuk National University, Korea (South))

20 minutes for each paper.

### **Validating Goal-Oriented Hypotheses of Business Problems Using Machine Learning: An Exploratory Study of Customer Churn (BIGDATA2020-2007)**

*Sam Supakkul (NCR Corporation, USA); Robert Sungsoo Ahn and Ronaldo Goncalves (University of Texas at Dallas, USA); Diana Villarreal (NCR Corporation, USA); Liping Zhao (University of Manchester, USA); Tom Hill and Lawrence Chung (University of Texas at Dallas, USA)*

Abstract: Organizations are investing in Big Data and AI, but the majority of these projects were predicted to fail. A study shows that one of the biggest obstacles is the lack of understanding how to use analytics to improve business. This paper presents Metis, a method for ensuring that business goals and problems are explicitly traceable to Machine Learning projects and potential (or hypothesized) complex problems can be properly validated before investing in costly solutions. Using this method, business goals are captured to provide context for hypothesizing business problems, which can be further refined into more detailed problems to identify features of data that are suitable for machine learning (ML). A supervised ML algorithm is then used to generate a prediction model that captures the underlying patterns and insights about the business



problems in the data. A ML Explainability model is then used to extract from the prediction model the individual features and the degree of which contribute to the problems. The extracted feature contributions are then fed back to the goal-oriented problem model to validate the most important business problems. Our experiment results shows that Metis is able to detect the most influential problem root cause when it was not apparent through data analysis. This approach is illustrated using a real-world customer churn problem for a bank and a publicly available customer churn dataset.

**The collaborative influence of multiple interactions on successive POI recommendation** (BIGDATA2020-2008)

*Nan Wang (Heilongjiang University, China); Yong Liu and Peiyao Han (Heilongjiang University, Harbin, China); Xiaokun Li (Postdoctoral Program of Heilongjiang Hengxun Technology Co., Ltd., China); Jinbao Li (Heilongjiang University, Harbin, China)*

**Abstract:** With the rapid development of social networks, users hope to obtain more accurate and personalized services. In general, POI recommendation often uses the historical behaviors of a user to recommend the top N POIs and rarely consider the current state of the user. Unlike POI recommendation, successive POI recommendation is more sensitive to user preferences and changes in time and space. In order to alleviate the data sparsity, we make full use of the interaction of time, space and user interest preferences, and propose a successive POI recommendation model called UTeSp. The UTeSp model uses the collaborative influence of multiple interactions to build a model, which can well adapt to the needs of users at different times and different locations. And it can change dynamically. Furthermore, we associate the user's inherent interest preference with the user's friend's influence on the target user, and propose a user-level interest preference based on attention mechanism, which can obtain more accurate user preference results. In addition, a novel TDP\_HC algorithm is designed to segment time dynamically. Based on the partial order relationship, we propose two interpretable methods to enhance the learning ability of the model. The two methods can be used in other similar successive POI recommendation models. Experimental results show that the F1-score of UTeSp model on the two real datasets is better than that of several mainstream successive POI recommendation models, and the two partial order methods also show the effectiveness of our model.

**A Data-Driven Method for Dynamic OD Passenger Flow Matrix Estimation in Urban Metro Systems** (BIGDATA2020-2009)

*Jiexia Ye (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China); Juanjuan Zhao (Shenzhen Institutes of Advanced Technology, China); Liutao Zhang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China); Cheng-Zhong Xu (University of Macau, China); Jun Zhang (Institute of Computing Technology, Chinese Academy of Sciences, China); Kejiang Ye (Chinese Academy of Sciences, China)*

**Abstract:** Dynamic O-D flow estimation is the basis of metro network operation, such as transit resource allocation, emergency coordination, strategy formulation in urban rail system. It aims to estimate the destination distribution of current inflow of each origin station. However, it is a challenging task due to its limitation of available data and multiple affecting factors. In this paper, we propose a practical method to estimate dynamic OD passenger flows based on long-term AFC data and weather data. We first extract the travel patterns of each individual passenger based on AFC data. Then the passengers of current inflows based on these patterns are classified into fixed passengers and stochastic passengers by judging whether the destination can be inferred. Finally, we design a K Nearest Neighbors (KNN) and Gaussian Process Regression (GPR) combined hybrid approach to dynamically predict stochastic passengers' destination distribution based on the observation that the distribution has obvious periodicity and randomness. We validate our method based on extensive experiments, using AFC data and weather data in Shenzhen, China over two years. The evaluation results show that our approach with 85% accuracy surpasses the results of baseline methods and the estimation precision reaches 85%.

**Session IV: Big Data Applications I** (September 18 - September 20, 2020, 7/24 Available)

Chair: Dr. MD Zakirul Alam Bhuiyan (Fordham University, USA)

20 minutes for each paper.

**Cross-Cancer Genome Analysis on Cancer Classification Using Both Unsupervised and Supervised Approaches** (BIGDATA2020-2010)

*Jonathan Zhou (Horace Greeley High School, USA); Baldwin Chen (Ardsley High School, USA); Nianjun Zhou (IBM, USA)*

**Abstract:** Many problems exist within the current cancer diagnosis pipeline, one of which is alarmingly high over-diagnosis rates in breast, prostate, and lung cancer. Through quantifying gene expression levels, next-generation sequencing techniques such as RNA-Seq offer an opportunity for researchers and clinicians to gain a more complete view of a cell's transcriptome. With the adoption of this new data source, cross-cancer methods for cancer diagnosis have become more viable. We utilize mutual information in conjunction with a Gaussian mixture model and t-SNE to evaluate the separability of cancer and non-cancer tissue samples from RNA-Seq expression data. The Gaussian mixture and t-SNE combination produced clear clustering without supervision, suggesting the ability to separate tissue samples algorithmically. Afterwards,

we use a collection of deep neural networks to classify tissue origin and status from tissue sample gene expressions. We use genes selected based on the prior mutual information technique. First, we select the top 500 genes from candidate genes without considerations for overlap in the predictability of those genes. We then applied Recursive Feature Elimination (RFE) to select 200 genes, thus accounting for covariation. We find that the performance using the top 500 genes is only slightly better than the 200 genes selected using RFE, and the two approaches achieved similar performance overall, indicating that only a small subset of genes is required for the identification of status and origin. This work indicates that RNA sequencing data is a useful tool for cross-cancer studies. Next steps include the implementation of a greater amount of non-cancer data from other datasets to decrease bias in model training.

#### **Heavy Vehicle Classification through Deep Learning (BIGDATA2020-2011)**

*Pei-Yun Sun, Wan-Yun Sun, Yicheng Jin and Richard Sinnott (University of Melbourne, Australia)*

Abstract: Understanding the flow of traffic on road networks is increasingly important especially with the continued urbanization of the global population. Numerous hardware and software technologies have been applied to measure traffic volumes by Government agencies and/or organizations such as Google, however they are either expensive to deploy; limited in their ability to disambiguate the kinds of vehicles on the road network, or of increasing importance, they infringe on the privacy of individuals, e.g. tracking phones. In this paper we describe work applying deep learning technologies to identify and classify different vehicles on the road network of Victoria with specific focus on heavy goods vehicles (trucks and trailers). Specifically, we present an approach to automatically detect, classify and count the unique classes of trucks and trailers that are found on the road network and the direction of travel. We apply and compare leading deep learning approaches including You Only Look Once version 3 (YOLOv3) and Single Shot Multi-Box Detector (SSD). This paper builds upon earlier work which focused on data (video) from a single traffic junction in Melbourne. This work is based on a wider range of data (videos) from locations reflecting the diversity of road use including multi-lane motorways, rural roads and city roads.

#### **Fake News Classification of Social Media through Sentiment Analysis (BIGDATA2020-2012)**

*Richard Sinnott (University of Melbourne, Australia)*

Abstract: The impacts of the Internet and the ability for information to flow in real-time to all corners of the globe has brought many benefits to society. However, this capability has downsides. Information can be inexact, misleading or indeed downright and deliberately false. Fake news has now entered the common vernacular. In this work, we consider fake news with specific regard to social media. We hypothesise that fake news typically deals with emotive topics that are deliberated targeted to cause a reaction and encourages the spread of information. As such, we explore sentiment analysis of real and fake news as reported in social networks (Twitter). Specifically, we develop an AWS-based Cloud platform utilising news contained in the untrustworthy resource FakeNewsNet and a more trusted resource CredBank. We train algorithms using Naive Bayes, Decision Tree and Bi-LSTM for sentiment classification and feature selection. We show how social media sentiment can be used to improve the accuracy in identification of fake news from real news.

### **Session V: Big Data Applications II (September 18 - September 20, 2020, 7/24 Available)**

Chair: Dr. Xuan Guo (University of North Texas, USA)

20 minutes for each paper.

#### **Clinical Trials Data Management in the Big Data Era (BIGDATA2020-2013)**

*Martha O. Perez-Arriaga (Department of Veterans Affairs & Cooperative Studies Program, Clinical Research Pharmacy Coordinating Center, USA); Krishna A Poddar (Department of Veterans Affairs)*

Abstract: The Department of Veterans Affairs (VA) Cooperative Studies Program (CSP), Clinical Research Pharmacy Coordinating Center (Center) has supported clinical trials for more than four decades. Managing information from clinical trials and published results in the Big Data era presents new challenges and opportunities. These include and are not limited to data attribution, aggregation, adaptability, and prompt analysis. Hence, the Center has created a dynamic application to present a broad understanding of the clinical trials' achievements. To collect crucial information from clinical trials, this application includes 1) data attribution to identify provenance and to preserve relationships between trials and resulting publications, 2) data normalization to deal with variety of formats and concepts, 3) data aggregation to integrate information from different trials, and 4) data analysis with a friendly interface to consult aggregated information promptly. This work establishes a Semantic Data Model for each clinical trial to create a summary of key information in a machine-readable format, and to enrich each summary with semantic information. In addition, it allows the union of these models to represent a global knowledge source from a set of clinical trials. The organized models offer compatibility and interoperability within and among clinical trials.

#### **Big Data Applications on Large-Scale Infrastructures (BIGDATA2020-2014)**

*Verena Kantere (University of Ottawa, Canada)*

Abstract: For a range of major scientific computing challenges that span fundamental and applied science, the deployment of Big Data Applications on a large-scale system, such as an internal or external cloud, a cluster or even distributed public

resources ("crowd computing"), needs to be offered with guarantees of predictable performance and utilization cost. Currently, however, this is not possible, because scientific communities lack the technology, both at the level of modelling and analytics, which identifies the key characteristics of BDAs and their impact on performance. There is also little data or simulations available that address the role of the system operation and infrastructure in defining overall performance. Our vision is to fill this gap by producing a deeper understanding of how to optimize the deployment of Big Data Applications on hybrid large-scale infrastructures. Our objective is the optimal deployment of BDAs that run on systems operating on large infrastructures, in order to achieve optimal performance, while taking into account running costs. We describe a methodology to achieve this vision. The methodology starts with the modeling and profiling of applications, as well as with the exploration of alternative systems for their execution, which are hybridization's of cloud, cluster and crowd. It continues with the employment of predictions to create schemes for performance optimization with respect to cost limitations for system utilization. The schemes can accommodate execution by adapting, i.e. extend or change, the system.

**Predicting the DJIA with news headlines and historic data using hybrid genetic algorithm/support vector regression and BERT (BIGDATA2020-2015)**

*Benjamin Warner, Aaron Crook and Renzhi Cao (Pacific Lutheran University, USA)*

Abstract: One important application of Artificial Intelligence(AI) is forecasting stock price in the stock market, as such knowledge is highly useful for investors. We first examined several state-of-the-art AI techniques, including hybrid symbiotic organisms search/feedforward neural network, hybrid genetic algorithm/regression, neuro-fuzzy systems, reasoning neural networks, bidirectional encoder representations from transformers (BERT), and recurrent neural networks. After that, we proposed a new AI model that uses hybrid genetic algorithm/support vector regression and BERT to predict daily closes in the Dow Jones Industrial Average. We found that using headline data may provide an improvement accuracy with root-mean-squared-errors ranging between 0 and 10.4% lower than without headline data depending on the model used, although further analysis may reveal significant improvements. The code and data used in our model can be found at: <https://github.com/bcwarner/djia-gasvr-bert>.

**Session VI: Big Data Cleaning (September 18 - September 20, 2020, 7/24 Available)**

Chair: Prof. Surya Nepal (CSIRO, Australia)

20 minutes for each paper.

**Ensemble learning for heterogeneous missing data imputation (BIGDATA2020-2016)**

*Andre L Costa Carvalho (Universit  du Qu bec, Canada); Darine Ameyed (ETS, Canada); Mohamed Cheriet (Ecole de technologie sup rieure (University of Quebec), Canada)*

Abstract: Missing values can significantly affect the result of analyses and decision making in any field. Two major approaches deal with this issue: statistical and model-based methods. While the former brings bias to the analyses, the latter is usually designed for limited and specific use cases. To overcome the limitations of the two methods, we present a stacked ensemble framework based on the integration of the adaptive random forest algorithm, the Jaccard index, and Bayesian probability. Considering the challenge that the heterogeneous and distributed data from multiple sources represents, we build a model in our use case, that supports different data types: continuous, discrete, categorical, and binary. The proposed model tackles missing data in a broad and comprehensive context of massive data sources and data formats. We evaluated our proposed framework extensively on five different datasets that contained labelled and unlabelled data. The experiments showed that our framework produces encouraging and competitive results when compared to statistical and model-based methods. Since the framework works for various datasets, it overcomes the model-based limitations that were found in the literature review.

**Entropy-based Approach to Efficient Cleaning of Big Data in Hierarchical Databases (BIGDATA2020-2017)**

*Eugene Levner (Holon Institute of Technology, Israel); Boris Kriheli (Ashkelon Academic College, Israel); Arriel Benis (Holon Institute of Technology, Israel); Alexander Ptuskin (Moscow Bauman Technical University, Kaluga Branch, Russia); Amir Elalouf and Sharon Hovav (Bar Ilan University, Israel); Shai Ashkenazi (Ariel University, Israel)*

When databases are at risk of containing erroneous, redundant, or obsolete data, a cleaning procedure is used to detect, correct or remove such undesirable records. We propose a methodology for improving data cleaning efficiency in a large hierarchical database. The methodology relies on Shannon's information entropy for measuring the amount of information stored in databases. This approach, which builds on previously-gathered statistical data regarding the prevalence of errors in the database, enables the decision maker to determine which components of the database are likely to have undergone more information loss, and thus to prioritize those components for cleaning. In particular, in cases where the cleaning process is iterative (from the root node down), the entropic approach produces a scientifically motivated stop-ping rule that determines the optimal (i.e. minimally required) number of tiers in the hierarchical database that need to be examined. This stopping rule defines a more streamlined representation of the database, in which less in-formative tiers are eliminated.

**Dissecting Biological Functions for BRCA Genes and their Targeting MicroRNAs within Eight Clusters**  
(BIGDATA2020-2018)

*Yining Zhu (British International School of Houston & Next-Gen Intelligent Science Training, USA); Ethan Sun (Seven Lakes High School, USA); Yongsheng Bai (Next-Gen Intelligent Science Training, USA)*

Abstract: The Cancer Genome Atlas (TCGA) estimated that 12.4 percent of women born in the US will develop breast cancer in their lives. The goal of this study is to identify common biological signature (ex. Gene Ontology or GO terms) for breast cancer candidate genes and miRNAs among these eight clusters identified through bioinformatics method in our recent study. The eight "communities" or clusters for Breast Invasive Carcinoma (BRCA) generated in our previous study are performed for functional annotation in this study. We observed that among all the GO terms enriched in top five groups of these eight clusters, Transcription, Lumen, Nucleolus, and Nucleoplasm are enriched the most, followed by Nucleoside binding, Nucleotide binding, and ATP binding. Two clusters among eight contain three or more previously reported breast cancer risk genes. We examined these clusters in terms of pathway association for miRNAs targeting the breast cancer risk genes, and found that 18 of 26 targeting miRNAs are involved in breast cancer related pathway PI3K-Akt. Our study showed that many miRNAs targeting breast cancer cluster genes are also associated with breast cancer related pathways, which provide evidence that some miRNA and genes pairs likely contribute to breast cancer in the context of their targeting relationship.

# 2020 International Conference on Services Computing (SCC 2020)

---

## Session I: Service Composition (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Qingyang Wang (Louisiana State University--Baton Rouge, USA)

20 minutes for each paper.

### Web Service Composition by Optimizing Composition-Segment Candidates (SCC2020-3001)

*Fang-Yuan Zuo and Yu-Bin Yang (Nanjing University, China)*

Abstract: Web service composition has been increasingly challenging in recent years due to the escalating number of services and the diversity of task objectives. Despite many researches have already addressed the optimization of multiple Quality of Service(QoS) attributes, most of the currently available methods have to build a large web service dependency graph, which may incur excessive memory consumption and extreme inefficiency. To address these issues, we present a novel web service composition method by optimizing composition-segment candidates. Firstly, we formalize the web service composition problem as a Mixed-Integer Linear Programming(MILP) model and introduce some effective techniques for complex cases, and then a standard solver can be applied to this model. Afterwards, a candidate optimization method is proposed to solve the MILP model efficiently, which runs sharply fast without building a web service dependency graph. Experimental results on both Web Service Challenge 2009's datasets and substantial datasets randomly generated show that the proposed method outperforms the state-of-art while achieving a much ideal tradeoff among all the objectives with better performance.

### Collaborative Learning using LSTM-RNN for Personalized Recommendation (SCC2020-3002)

*Benjamin A Kwapong, Richard Anarfi and Kenneth K Fletcher (University of Massachusetts Boston, USA)*

Abstract: Today, the ability to track users' sequence of online activities, makes identifying their evolving preferences for recommendation practicable. However, despite the myriad of available online activity information, most existing time-based recommender systems either focus on predicting some user rating, or rely on information from similar users. These systems, therefore, disregard the temporal and contextual aspects of users preferences, revealed in the rich and useful historical sequential information, which can potentially increase recommendation accuracy. In this work, we consider such rich, user online activity sequence, as a complex dependency of each user's consumption sequence, and combine the concept of collaborative filtering with long short-term memory recurrent neural network (LSTM-RNN), to make personalized recommendations. Specifically, we use encoder-decoder LSTM-RNN, to make sequence-to-sequence recommendations. Our proposed model builds on the strength of collaborative filtering while preserving individual user preferences for personalized recommendation. We conduct experiments using Movielens dataset to evaluate our proposed model and empirically demonstrate that it improves recommendation accuracy when compared to state-of-the-art recommender systems.

### Microservices Backlog - A model of granularity specification and microservice identification (SCC2020-3003)

*Fredy Humberto Vera Rivera (Universidad Francisco de Paula Santander & Universidad del Valle, Colombia); Eduard Puerto-Cuadros (Universidad Francisco de Paula Santander, Colombia); Hernan Astudillo (Universidad Técnica Federico Santa María, Chile); Carlos Gaona-Cuevas (Universidad del Valle, Colombia)*

Abstract: Microservices are a software development approach where applications are composed of small independent services that communicate through well-defined APIs. A major challenge of designing these applications is determining the appropriate microservices granularity, which is currently done by architects using their judgment. This article describes Microservice Backlog (MB), a fully automatic genetic-programming technique that uses the product backlog's user stories to (1) propose a set of microservices for optimal granularity and (2) allow architects to visualize at design time their design metrics. Also, a new Granularity Metric (GM) was defined that combines existing metrics of coupling, cohesion, and associated user stories. The MB-proposed decomposition for a well-known state-of-the-art case study was compared with three existing methods (two automatics and one semi-automatic); it had consistently better GM scoring and fewer average calls among microservices, and it allowed to identify critical points. The wider availability of techniques like MB will allow architects to automate microservices identification, optimize their granularity, visually assess their design metrics, and identify at design time the system critical points.

## Session II: Service Management (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Yunni Xia (Chongqing University, China)

20 minutes for each paper.



### **QoS Time Series Modeling and Forecasting for Web Services: A Comprehensive Survey of Subsequent Applications and Experimental Configurations (SCC2020-3004)**

*Yang Syu (Academia Sinica, Taiwan)*

**Abstract:** Time-aware (time series-based) Web service QoS modeling and forecasting have been investigated and addressed for over a decade and a large number of studies and approaches have been produced. However, these existing efforts lack a comprehensive and detailed review that profoundly and systematically organizes, analyzes, and discusses this body of work. Thus, to fill this gap, the authors offered the paper QoS Time Series Modeling and Forecasting for Web Services: A Comprehensive Survey, in which four essential research concerns of this area, namely, problems, approaches, performance measures, and QoS datasets, have been recognized and reviewed in detail. However, aside from these essential research concerns, we also identified two optional research concerns from the current studies, namely, the subsequent applications and experimental configurations. Due to space restrictions, these two optional research concerns were only briefly mentioned in the above survey article, and thus, in this supplementary paper, the authors thoroughly present and review these two optional research concerns. The primary purpose of performing QoS time series modeling and forecasting is to obtain accurate future QoS estimations for subsequent usage (application), such as QoS-aware service composition and proactive service replacement for SLA/QoS management. Therefore, in the section addressing the first optional research concern, the application of each surveyed study is identified first, and then these current applications are introduced in detail. However, to comprehensively and rigorously observe and evaluate the performance of a proposed or employed approach under different conditions, a set of configuration settings must be varied to run experimentation. Thus, in the second part of this paper, we first define and discuss the identified experimental configuration parameters in this research area and then list the parameters and settings that have been considered in each surveyed study.

### **An Attention Model for Mashup Tag Recommendation (SCC2020-3005)**

*Kenneth K Fletcher (University of Massachusetts Boston, USA)*

**Abstract:** Mashups have emerged as a popular technique to compose value-added web services/APIs, to fulfill some complicated business needs. This has increased the number of available mashups over the internet. The increase however, poses a new requirement of organizing and managing these mashups for better understanding and discovery. For this reason, tags have become highly important because they describe items and allows for easy discovery. Most existing tag recommendation methods typically follow a manual process based on controlled vocabulary, or consider tags as words in isolation contained in mashup descriptions. Such methods therefore fail to characterize the diverse functional features of mashups. This work proposes an attention model to automatically recommend mashup tags. Specifically, our proposed model has two levels of attention mechanisms applied at the word- and sentence-levels and subsequently recommend top-N words with highest attention weights as tags. Our model is based on the intuition that not every word in a mashup description is equally relevant in identifying its functional aspects. Therefore, determining the relevant sections involves modeling the interactions of the words, not just their presence in isolation. We demonstrate the effectiveness of our method by conducting extensive experiments on a real-world dataset crawled from [www.programmableweb.com](http://www.programmableweb.com). We also compare our method with some baseline tag recommendation methods for verification.

### **Automated Web Service Specification Generation through a Transformation-based Learning (SCC2020-3006)**

*Mehdi Bahrami and Wei-Peng Chen (Fujitsu Laboratory of America Inc., USA)*

**Abstract:** Web Application Programming Interface(API) allows third-party and subscribed users to access data and functions of a software application through the network or the Internet. Web APIs expose data and functions to the public users, authorized users or enterprise users. Web API providers publish API documentations to help users to understand how to interact with web-based API services, and how to use the APIs in their integration systems. The exponential raise of the number of public web service APIs may cause a challenge for software engineers to choose an efficient API. The challenge may become more complicated when web APIs updated regularly by API providers. In this paper, we introduce a novel transformation-based approach which crawls the web to collect web API documentations (unstructured documents). It generates a web API Language model from API documentations, employs different machine learning algorithms to extract information and produces a structured web API specification that compliant to Open API Specification (OAS) format. The proposed approach improves information extraction patterns and learns the variety of structured and terminologies. In our experiment, we collect a sheer number of web API documentations. Our evaluation shows that the proposed approach find RESTful API documentations with 75% accuracy, constructs API endpoints with 84%, constructs endpoint attributes with 95%, and assigns endpoints to attributes with an accuracy 98%. The proposed approach were able to produces more than 2,311 OAS web API Specifications.

### **Session III: Quality of Services (QoS) (September 18 - September 20, 2020, 7/24 Available)**

Chair: Prof. Sangeetha Seshadri (IBM Almaden Research Center, USA)

20 minutes for each paper.

**Midiag: a Sequential Trace-based Fault Diagnosis Framework for Microservices** (SCC2020-3007)

*Yao Sun (Nanjing Institute of Big Data, Jinling Institute of Technology, China); Lun Meng (Hohai University, China); Shudong Zhang (Capital Normal University, China)*

Abstract: Cloud applications are often deployed in shared data centers to optimize re-source allocation and improve management efficiency. However, since a cloud application often has a large amount of different microservices, it is difficult for operators to analyze these microservices with a unified model. To deal with the above problem, this paper proposes a sequential trace-based fault diagnosis framework called as Midiag by mining the patterns of microservices' system call sequences. Midiag collects system calls with a non-invasive lightweight tool, and then uses k-means to cluster system call sequences as patterns with the longest common subsequence. The GRU-based neural network is employed to model the patterns of system call sequences to predict the next system call, and thus Midiag diagnoses faults by comparing the predicted system call and the actual one in a specific pattern. We have validated Midiag with many different types of applications deployed in containers. The results demonstrate that Midiag can well classify these applications as different types and accurately diagnose the applications injected with faults.

**An Empirical Study of Web API Quality Formulation** (SCC2020-3008)

*Esi Adeborna (University of Massachusetts Lowell, USA); Kenneth K Fletcher (University of Massachusetts Boston, USA)*

Abstract: This paper presents an empirical study on one of the popular web API repositories, www.programmableweb.com. The study is to ascertain the impact of the structure and formulation of external web API quality factors on the overall web API quality. The study is based on the hypothesis that, in such a multi-factor quality measurement, the structure and formulation of the quality factors can make a substantial difference in its quantification. Specifically, we employ statistical tools such as exploratory factor analysis, to determine the latent factors that contributes to web API quality. We subsequently determine the loading of each latent factors to propose a new quality model for web API quality computation.

**On the Diffusion and Impact of Code Smells in Web Applications** (SCC2020-3009)

*Narjes Bessghaier and Ali Ouni (Ecole de Technologie Supérieure (ETS), University of Quebec, Canada); Mohamed Wiem Mkaouer (Rochester Institute of Technology (RIT), Rochester, NY, USA)*

Abstract: Web applications (web apps) have become one of the largest parts of the current software market over years. Modern web apps offer several business benefits over other traditional and standalone applications. Mainly, cross-platform compatibility and data integration are some of the critical features that encouraged businesses to shift towards the adoption of Web apps. Web apps are evolving rapidly to acquire new features, correct errors or adapt to new environment changes especially with the volatile context of the web development. These ongoing amends often affect software quality due to poor coding and bad design practices, known as code smells or anti-patterns. The presence of code smells in a software project is widely considered as form of technical debt and makes the software harder to understand, maintain and evolve, besides leading to failures and unforeseen costs. Therefore, it is critical for web apps to monitor the existence and spread of such anti-patterns. In this paper, we specifically target web apps built with PHP being the most used server-side programming language. We conduct the first empirical study to investigate the diffuseness of code smells in Web apps and their relationship with the change proneness of affected code. We detect 12 types of common code smells across a total of 223 releases of 5 popular and long-lived open-source web apps. The key findings of our study include: 1) complex and large classes and methods are frequently committed in PHP files, 2) smelly files are more prone to change than non-smelly files, and 3) Too Many Methods and High Coupling are the most associated smells with files change-proneness.

**Performance Evaluation on Blockchain Systems: A Case Study on Ethereum, Fabric, Sawtooth and Fisco-Bcos** (SCC2020-3010)

*Rui Wang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China); Kejiang Ye (Chinese Academy of Sciences, China); Tianhui Meng (Shenzhen Institutes of Advanced Technology, China); Chengzhong Xu (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, USA)*

Abstract: Blockchain technology is currently receiving increasing attention with widely used in many fields such as finance, retail, Internet of Things, and intelligent manufacturing. Although many blockchain applications are still in the early stage, this technique is very promising and has great potential. Blockchain is considered as one of the core technologies to trigger a new round of disruptive changes after Internet. In the future, it is expected to change the development prospects of many industries. However, the current blockchain systems suffer from poor performance which affects large-scale application. In order to better understand the performance of the blockchain systems, in this paper, we analyze four mainstream blockchain systems (Ethereum, Fabric, Sawtooth and Fisco-Bcos), and then perform a performance comparison through open source blockchain benchmarking tools. After that, we propose several optimization methods and discuss the future development of blockchain technique.

# 2020 International Conference on AI & Mobile Services (AIMS 2020)

---

## Session I: Neural Networks (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Ruifeng Xu (Harbin Institute of Technology, Shenzhen, China)

20 minutes for each paper.

### [A Neural Framework for Chinese Medical Named Entity Recognition \(AIMS2020-4001\)](#)

*Zhengyi Zhao, Weichuan Xing, Junlin Wu, Xurui Sun, Yuan Chang and Binyang Li (University of International Relations, China)*

Abstract: Named Entity Recognition (NER) in the medical field targets to extract names of disease, surgery, and the organ location from medical texts, which is considered as the fundamental work for medical robots and intelligent diagnosis systems. It is very challenging to recognize the named entities in Chinese medical texts, because (a) one single Chinese medical named entity is usually expressed with more characters/words than other languages, i.e. 3.2 words and 7.3 characters in average; (b) different types of medical named entities are usually nested together. To address the above issue, this paper presents a neural framework that is constructed by two modules: a pre-trained module to distinguish each individual entity from the nested expressions, while a modified Bi-LSTM module to effectively identify long entities. We conducted the experiments based on the CCKS 2019 dataset, our proposed method can identify the medical entity in Chinses, especially for those nested entities embodied in long expressions, and 88.84% was achieved in terms of F1-score, and 15.09% improvement was achieved compared to the baseline models.

### [Infant Sound Classification on Multi-stage CNNs with Hybrid Features and Prior Knowledge \(AIMS2020-4002\)](#)

*Chunyan Ji, Sunitha Basodi, Xueli Xiao and Yi Pan (Georgia State University, USA)*

Abstract: We propose an approach of generating hybrid feature set and using prior knowledge in a multi-stage CNNs for robust infant sound classification. The dominant and auxiliary features within the set are beneficial to enlarge the coverage as well as keeping a good resolution for modeling the diversity of variations within infant sound. The novel multi-stage CNNs method work together with prior knowledge constraints in decision making to overcome the limited data problem in infant sound classification. Prior knowledge either from rules or from statistical results provides a good guidance for searching and classification. The effectiveness of proposed method is evaluated on commonly used Dustan Baby Language Database and Baby Chillanto Database. It gives an encouraging reduction of 4.14% absolute classification error rate compared with the results from the best model using one-stage CNN. In addition, on Baby Chillanto Database, a significant absolute error reduction of 5.33% is achieved compared to one-stage CNN and it outperforms all other existing related studies.

### [An Annotated Chinese Corpus for Rumor Veracity Detection \(AIMS2020-4003\)](#)

*Bo Yan (The 54th Research Institute of China Electronics Technology Group Corporation, China); Yan Gao, Shubo Zhang, Yi Zhang, Yan Du and Binyang Li (University of International Relations, China)*

Abstract: With the popularity of social media, Twitter, Facebook, and Weibo etc. platforms have become an indispensable part of people's life, where users can freely release and spread information. Meanwhile, the information credibility cannot be guaranteed and there exist a great amount of rumors in social media. These information will usually bring negative impact, and even affect the real society. To solve this problem, there are some work on the rumor corpora construction for automatic rumor detection. However, existing work focused on political domain and most of them were limited in English texts. As a result, these corpora cannot be well applied into other domains with resource-poor languages. This paper proposes a Chinese rumor detection corpus, named CRDC. This corpus consists of 10,000 rumors and 14,472 non-rumors from Weibo. Moreover, other information including language-independent features are also acquired, including rumors' retweet and like information, which can effectively help rumor detection and rumor propagation research in other languages. To better demonstrate the corpus, we also conducted some initial experiments to show details and statistics of our corpus.

## Session II: Machine Learning (September 18 - September 20, 2020, 7/24 Available)

Chair: Dr. De Wang (Sunmi US Inc., USA)

20 minutes for each paper.

### [Answer Selection Based on Mixed Embedding and Composite Features \(AIMS2020-4004\)](#)

*Mingli Wu, Xianwei Cui, Jiale Li and Jianyong Duan (North China University of Technology, China)*



**Abstract:** With the rapid growth of text information, intelligent question answering has gained more attention than ever. In this paper, we focus on answer selection, one kind of question answering tasks. In this field, deep neural networks and attention mechanism have brought encouraging results. To improve the performance further, we investigate mixed embedding (word embedding and character embedding) representation for sentences to encode rich meaning. At the same time, we introduce a convolutional neural network (CNN) to compensate the loss of the max pooling layer in our attention based bidirectional Long Short-Term Memory (biLSTM) model. CNN features and the features from max pooling form final composite features, which are employed to select correct answers. Experimental results show that we can obviously improve the Mean Reciprocal Rank (MRR) performance by 6.0% with the help of mixed embedding and composite features. The MRR and ACC@1 score are 79.63% and 69.60% respectively.

#### **Review Spam Detection Based on Multidimensional (AIMS2020-4005)**

*Liming Deng (Fuzhou University, China); Jingjing Wei (Fujian Jiangxia University, China); Shaobin Liang, Yuhuan Wen and Xiangwen Liao (Fuzhou University, China)*

**Abstract:** Review spam detection aims to detect the reviews with false information posted by the spammers on social media. The existing methods of review spam detection ignore the importance of the information hidden in the user interactive behaviors and fail to extract the indistinct contextual features caused by irregular writing style of reviews. In this paper, a new review spam detection method based on multi-dimensional features is proposed. The method utilizes the principal component analysis to get low-dimensional features to characterize the user-product relationship. Then, a neural network constructed with nested LSTM and capsule network is trained to extract textual context features and spatial structure features. Finally, the model combines the text and user behavioral features as the overall features, which are used as the input to the classification module to detect spam reviews. Experimental results show that the F1 value of our proposed method is 1.6%~3.5% higher than the existing methods, indicating the efficiency and effectiveness of our model, especially on the natural distribution datasets.

#### **Attention-based Interaction Trajectory Prediction (AIMS2020-4006)**

*Zhe Liu (University Of Electronic Science And Technology Of China, China); Lizong Zhang (University of Electronic Science and Technology of China, China); ZhiHong Rao (China Electronic Technology Cyber Security Co., Ltd & Cyberspace Security Key Laboratory of Sichuan Province, China); Guisong Liu (UESTC, China)*

**Abstract:** Trajectory prediction is a hot topic in the field of computer vision and has a wide range of applications. Trajectory prediction refers to predicting the future trajectory of a target based on its past trajectory. This paper proposes a method based on graph neural network and attention mechanism, in order to update trajectory characteristics by implement global pedestrian interaction. And, a direct relationship between history and future is introduced with the attention module for reducing error propagation. The method was evaluated on several real-world crowd datasets, the results demonstrate the effectiveness of our method.

### **Session III: AI Solutions (September 18 - September 20, 2020, 7/24 Available)**

Chair: Prof. Wei Zhong (University of South Carolina Upstate, USA)

20 minutes for each paper.

#### **Attention-based Asymmetric Fusion Network for Saliency Prediction in 3D Images (AIMS2020-4007)**

*Xinyue Zhang and Ting Jin (Hainan University, China)*

**Abstract:** Nowadays the visual saliency prediction has become a fundamental problem in 3D imaging area. In this paper, we proposed a saliency prediction model from the perspective of addressing three aspects of challenges. First, to adequately extract features of RGB and depth information, we designed an asymmetric encoder structure on the base of U-shape architecture. Second, to prevent the semantic information between salient objects and corresponding contexts from diluting in cross-modal distillation stream, we devised a global guidance module to capture high-level feature maps and deliver them into feature maps in shallower layers. Third, to locate and emphasize salient objects, we introduced a channel-wise attention model. Finally we built the refinement stream with integrated fusion strategy, gradually refining the saliency maps from coarse to fine-grained. Experiments on two widely-used datasets demonstrate the effectiveness of the proposed architecture, and the results show that our model outperforms six selective state-of-the-art models.

#### **Candidate Classification and Skill Recommendation in a CV Recommender System (AIMS2020-4008)**

*Adrian S Kurdija, Petar Afric and Lucija Šikić (University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia); Boris Plejić (Ericsson Nikola Tesla, Zagreb, Croatia); Marin Šilić (University of Zagreb, Croatia); Goran Delac, Klemo Vladimir and Sinisa Srblić (University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia)*

**Abstract:** In this paper, we describe a CV recommender system with a focus on two properties. The first property is the ability to classify candidates into roles based on automatic processing of their CV documents. The second property is the ability to recommend skills to a candidate which are not listed in their CV, but the candidate is likely to have them. Both features are based on skills extraction from a textual CV document. A spectral skill clustering is precomputed for the purpose of candidate classification, while skill recommendation is based on various similarity-based strategies.

Experimental results include both automatic experiments and an empirical study, both of which demonstrate the effectiveness of the presented methods.

**A Novel Method to Estimate Students' Knowledge Assessment (AIMS2020-4009)**

*Phuoc Hung Pham and Javed I. Khan (Kent State University, USA)*

**Abstract:** Performance of learning can be enriched with proper and timely feedback. This paper proposes a solution based on a Bayesian network in machine learning that can examine and judge students' written response to identify evidences that students fully comprehend concepts being considered in a certain knowledge domain. In particular, it can estimate probabilities that a student has known concepts in computer science at different cognitive ability levels in a sense. Thus, the method can offer learners personalized feedbacks on their strengths and shortcomings, as well as advising them and instructors of supplementary education actions that may help students to resolve any lacks to improve their knowledge and exam score.

**Session IV: AI Scenarios (September 18 - September 20, 2020, 7/24 Available)**

Chair: Dr. Ling Tian (University of Electronic Science and Technology of China, China)

20 minutes for each paper.

**OSAF\_e: One-Stage Anchor Free Object Detection Method Considering Effective Area (AIMS2020-4010)**

*Z. Yong and Lizong Zhang (University of Electronic Science and Technology of China, China); ZhiHong Rao (China Electronic Technology Cyber Security Co., Ltd & Cyberspace Security Key Laboratory of Sichuan Province, China); Guiduo Duan and Chunyu Wang (University of Electronic Science and Technology of China, China)*

**Abstract:** The task of object detection is to identify the bounding box of the object and its corresponding category in images. In this paper, we propose a new one-stage anchor free object detection algorithm OSAF\_e, with the consideration of effective mapping area. A feature extraction network is used to obtain high level feature, and the true bounding box of the object in the original image is mapped to the grid of feature map, in order to perform category prediction and bounding box regression. The proposed algorithm is evaluated with the Pascal Voc dataset, and the experiments indicate that it has a better result.

**A Mobile Application using Deep Learning to Automatically Classify Adult-only Images (AIMS2020-4011)**

*Richard Sinnott (University of Melbourne, Australia)*

**Abstract:** The Internet has become an essential part of everyday life. It links people with enormous amounts of information covering almost any topic imaginable. However harmful or inappropriate information such as pornography can also be easily found on the web which should not always be available, especially to minors. Internet filters are typically used to block such inappropriate content. These are largely based on the metadata related to the websites or by directly blocking the URLs related to those websites. However seemingly innocuous websites can contain undesirable images that should not be accessible to children. In this paper, we describe how images and videos can automatically be identified (classified) without any human supervision based on their subject matter. To achieve this, we apply deep learning methods to detect and classify adult-only image content from both images and live videos. We use the TensorFlow library and two pre-trained models: MobileNet\_v1 and Inception\_v3, with an official (academic) pornography dataset including associated labelling. The performance of each model was investigated. The final solution was delivered as an iOS application to detect and classify photos and live videos based on their adult-only content. The app achieved an accuracy of over 92%.

**Building Vector Representations for Candidates and Projects in a CV Recommender System (AIMS2020-4012)**

*Adrian S Kurdija, Petar Afric and Lucija Šikić (University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia); Boris Plejić (Ericsson Nikola Tesla, Zagreb, Croatia); Marin Šilić (University of Zagreb, Croatia); Goran Delac, Klemo Vladimir and Sinisa Srblić (University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia)*

**Abstract:** We describe a CV recommender system built for the purpose of connecting candidates with projects that are relevant to their skills. Each candidate and each project is described by a textual document (CV or a project description) from which we extract a set of skills and convert this set to a numeric representation using two known models: Latent Semantic Indexing (LSI) and Global Vectors for Word Representation (GloVe) model. Indexes built from these representations enable fast search of similar entities for a given candidate/project and the empirical results demonstrate that the obtained 12 distances correlate with the number of common skills and Jaccard similarity.

**Rehabilitation XAI to Predict Outcome with Optimal Therapies (AIMS2020-4013)**

*Takashi Isobe (Hitachi High Tech America, Inc., USA)*

**Abstract:** Value-based payment is becoming general in healthcare. In rehabilitation medicine, medical services are becoming to be paid depending on the outcome obtained from hospitalization period and dependency score called as FIM (Functional Independent Measurement). The optimal therapies to maximize the outcome differs by each patient's age, sex, disease, handicap, FIM and therapies. Non-experienced hospitals have a difficulty in improving the outcome. Therefore, there are

needs to maximize the outcome by optimizing therapies. We developed a rehabilitation XAI system to predict outcome with optimal therapies. Our system piles up medical records into vectors and predicts the outcome with optimal therapies using machine learning based on vector distance that can explain the basis of prediction in the same way as doctors suggesting optimal therapies to patients based on similar past cases. The interface not only displays optimal therapies but also predicts outcome by each patient. We used data from multiple hospitals and evaluated the adaptability of our system. In case of using the data from one hospital, the pattern achieving high outcome, which was most important because it was used to suggest optimal therapies, occupied the proportion of 31.1% in the actual record while the precision and recall were 64.5% and 73.4%. In case of using the data from another hospital, they were 64.4% and 66.1% against the actual proportion of 35.7%. In case of using the data from both hospitals, they were 63.6% and 71.0% against the actual proportion of 33.3%. Our system achieved similar performance and adaptability between two hospitals. Correlation coefficient between actual and predicted outcome were 0.681 using 204 patients' record. We compared the accuracy to predict outcome between our XAI and humans. Average outcomes of top 70% patients predicted at hospitalization by our XAI and humans were 43.0 and 42.4. Our XAI could predict outcome at higher accuracy than humans.

# 2020 World Congress on Services (SERVICES 2020)

---

## Session I: Industry Services (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Joao Eduardo Ferreira (Department of Computer Science, University of Sao Paulo, Brazil)  
20 minutes for each paper.

### Blockchain driven Three Domain Secure 2.x in Digital Payment Services Architecture (SERVICES2020-5001)

*Vikas S Shah (Knights of Columbus, USA)*

**Abstract:** Due to the recent advancements in digital commerce, consumers expect real-time digital payment convenient and available across channels as more connected devices become payment devices. It offers consumers to pay in-store or online purchases in many diversified ways. The three domains secure protocol evolved to version 2.x (3DS2) supporting the development in digital payment domain and its rapid adaptation. The specification includes the provisioning of the application-based purchases enabling risk-based decisions to authenticate the consumer transactions. 3DS2 enhances consumers' checkout experiences through out-of-band authentication. It eliminates the need for enrollment process and static password supporting non-payment activities and native mobile. The primary challenges to implement 3DS2 are dimensioning the risks, real-time variability in the risk factors, and precision to compute the accumulative risk associated with the individuals. Financial services, merchants, and consumers are enabled to connect into the blockchain network using application programming interfaces (APIs). It alleviates participants of Blockchain network from having to build out their own distributed transactions' server nodes. This paper proposes a blockchain-driven 3DS2 service architecture framework that integrates the risk-based decisions and provides a secure communication platform in digital commerce. We illustrate the increased level of authenticity, maintainability, extendibility, and flexibility in the digital payment ecosystem with the industry case study of membership-based in-store or online charitable contribution campaigns during point-of-sale.

### Decision support for patient discharge in hospitals - analyzing the relationship between length of stay and readmission risk, cost, and profit (SERVICES2020-5002)

*Isabella Eigner (FAU Erlangen Nuremberg, Germany); Freimut Bodendorf (University of Erlangen-Nuremberg, Germany)*

**Abstract:** Determining the optimal time for patient discharge is a challenging and complex task that involves multiple opposing decision perspectives. On the one hand, patient safety and the quality of healthcare service delivery and on the other hand, economic factors and resource availability need to be considered by hospital personnel. By using state-of-the-art machine learning methods, this paper presents a novel approach to determine the optimal time of patient discharge from different viewpoints, including a cost-centered, an outcome-centered, and a balanced perspective. The proposed approach has been developed and tested as part of a case study in an Australian private hospital group. For this purpose, unplanned readmissions and associated costs for episodes of admitted patient care are analyzed with regards to the respective time of discharge. The results of the analyses show that increasing the length of stay for certain procedure groups can lead to reduced costs. The developed approach can aid physicians and hospital management to make more evidence-based decisions to ensure both sufficient healthcare quality and cost-effective resource allocation in hospitals.

### Educational application of Big Data Research: a comparison of China and US (SERVICES2020-5003)

*Ting Zhang (Beijing Jiaotong University & Information Technology Center, China); Clara Elizabeth and Renzhi Cao (Pacific Lutheran University, USA)*

**Abstract:** Huge amount of data is generated each day, and big data research can be applied in various fields to analyze and extract new knowledge from this big amount of data, with education being one of the most important applications. However, few researchers are focused on comparative analysis of big data application in education between China and the US. In this paper, we analyze the differences between these two countries in their applications of big data theory to the education field. In addition, we also discuss the ethical challenges of these applications, such as how China and US protect privacy in educational field when applying big data research. Finally, this paper also discusses the lessons learned from the development of big data applications to the education field. This includes privacy protection, which could be used to help policy makers of countries around the world regulate the application of big data theory in the education field.

## Session II: Business Services (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Balaji Palanisamy (University of Pittsburgh, USA)  
20 minutes for each paper.

### **Case Study on Key Influencing Factors of Modern Service Industry Development (SERVICES2020-5004)**

*Zhu Xiangbo (Shenzhen Polytechnic & School of Construction and Environment Engineering, China); Ou Guoliang (Shenzhen Polytechnic, China); Li Yi (Shenzhen Institute of Information Technology, China); Zhou Zhigang (Shenzhen Polytechnic, China)*

**Abstract:** In the process of industrial structure adjustment and economic development model innovation, the role and status of modern service industry are becoming increasingly important. Shenzhen Municipal People's Government had paid great attention on the development of modern service industry. On one hand, modern service industry has been quite large and mature, and on the other hand, there are problems of insufficient in total volume, relatively narrow industry field and somehow inadequate industry structure. Based on rational choosing and building an influencing indicator system, this paper tries to analysis the development of modern service industry in Shenzhen. Through using methods such as cointegration analysis, Granger causality test and multiple linear regression, combined with the relevant data collected, we found that the foreign direct investment and the investment level of modern service industry are the key factors. Therefore, this paper suggested that the development of modern service industry in Shenzhen might improve on expanding foreign investment, strengthening the investment level of modern service industry, and upgrading professional talents.

### **Review for Influence of 5G on Industry Internet (SERVICES2020-5005)**

*Yang Liu (Kingdee International Software Group Co. & Research Institute of Information Technology, Tsinghua University, China)*

**Abstract:** In recent year, the downward pressure on the global economy is increasing, and the market is gradually changing. The urgent problem for entrepreneurs is how to keep their companies growing continuously. Industry Internet is a new direction that can help them by promoting business growth of enterprises. And the growth should be realized by exploring innovative business scenarios. The core of industry internet is not the internet, but the traditional industry. It forms the industry value chain through the mutual connection of traditional industries. The enterprises in each link of the value chain will become more powerful because of their value-added data. This paper solves several key problems in the development of enterprises: what economic era are they in? What is industry internet? How does industry Internet empower enterprises? How can 5G related technologies help traditional industries transform into industry internet?

### **DLchain: Blockchain with Deep Learning as Proof-of-Useful-Work (SERVICES2020-5006)**

*Changhao Chenli, Boyang Li and Taeho Jung (University of Notre Dame, USA)*

**Abstract:** Blockchains based on Proof-of-Work can maintain a distributed ledger with a high security guarantee but also lead to severe energy waste due to the useless hash calculation. Proof-of-Useful-Work (PoUW) mechanisms are alternatives, but finding hard puzzles with easy verification and useful results is challenging. Recent popular deep learning algorithms require large amount of computation resources due to the large-scale training datasets and the complexity of the models. The work of deep learning training is useful, and the model verification process is much shorter than its training process. Therefore, in this paper, we propose DLchain, a PoUW-based blockchain using deep learning training as the hard puzzle. Theoretical analysis shows that ours can achieve a security level comparable to existing PoW-based cryptocurrency when the miners' best interest is to maximize their revenue. Notably, this is achieved without relying on common assumptions made in existing PoUW-based blockchain such as globally synchronized timestamps.

### **Session III: Technology Services (September 18 - September 20, 2020, 7/24 Available)**

**Chair:** Prof. Kejiang Ye (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China)  
20 minutes for each paper.

### **AIServicex: A Knowledge Graph-based Intelligent Question-Answering System for Personal Services (SERVICES2020-5007)**

*Yao Sun (Nanjing Institute of Big Data, Jinling Institute of Technology, China); Wenming Gui (School of Software Engineering, Jinling Institute of Technology, China); Cheng Han (School of Computer Engineering, Jinling Institute of Technology, China); Yan Zhang (Jinling Institute Of Technology, China); Shudong Zhang (Capital Normal University, China)*

**Abstract:** Knowledge graph-based question-answering systems are widely used in e-commerce enterprises. They can reduce the costs of customer services and improve service capabilities. The description of questions is often ambiguous, and the knowledge graph's update in online personal services always has a high overhead. To address the above issues, by augmenting domain semantics, this paper proposes a knowledge graph-based intelligent question-answering system called as AIServicex. It employs a gate recurrent unit model to identify entities and assertions, and then gets the most related semantic augmentation contents from existing external domain knowledge via topic comparison. Then, it ranks all the candidates to get optimal answers by discovering several heuristic rules. Periodically, it augments the global knowledge graph with minimized updating costs with an Integer linear programming resolving model. This mechanism can recognize question entities precisely, and map domain knowledge to the KG automatically, which achieves a high answering precision with a low overhead. Experiments with a production e-commerce data show that AIServicex can improve the precision.

**A Fuzzy AHP and TOPSIS Approach for Web Service Selection (SERVICES2020-5008)**

*Sandile T Mhlana, Manoj Lall and Ojo Sunday (Tshwane University of Technology, South Africa)*

Abstract: Multi criteria decision making (MCDM) model is proposed for determining the most suitable web service from a collection of functionally-equivalent web services with different non-functional properties. This paper presents an evaluation approach that combines fuzzy analytic hierarchy process (AHP) and Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) to solve the MCDM selection problems with conflicting criteria. Fuzzy AHP method determines subjective weights by dealing with vagueness and uncertainty in subjective user's judgment while, TOPSIS algorithm ranks the different alternatives. A numerical example is based on a real-world dataset is presented to illustrate the procedural matters of the web service selection model. The numerical results show that the proposed approach can effectively select an appropriate web service based on user preference. WeatherStationService performed better than other web services under the selected QoS requirements.

**Blockchain Federation Enabled Trustable Internet of Things (SERVICES2020-5009)**

*Zhitao Wan (University of Nottingham Ningbo China and Ningbo Free Trade Zone Blockchain Laboratory, China); Minqiang Cai (Institute of Advanced Technology Research, Ge Lian Corporation, China); Xiuping Hua (University of Nottingham Ningbo China, China); Jinqing Yang and Xianghua Lin (Institute of Advanced Technology Research, Ge Lian Corporation, China)*

Abstract: The Internet of Things (IoT) provides an infrastructure enabling advanced functions based on interoperable information and communication technologies by interconnecting physical and virtual things. IoT has gained more and more attention from both academy and industry. However, the design, maintenance, and governance of IoT are facing challenges because of an enormous amount of heterogeneous devices involved. IoT requires strengthening its trustable features. The Blockchains are immutable distributed ledger systems supporting trustable features including security, scalability, privacy, safety, and connectivity usually without a central authority. Due to the complexity of IoT system any current blockchain cannot satisfy all demands from heterogeneous devices, networks, and higher level functionalities. This paper proposes a novel reference architecture combines Blockchain Federation with IoT to make IoT capable of being trusted. The architecture enables trustable features by integrating blockchains with different IoT tiers and layers. Proof of concept system is implemented to demonstrate the correctness and feasibility of the architecture. Furthermore, the future directions of Blockchain Federation and IoT evolution are discussed.



# 2020 International Conference on Cognitive Computing (ICCC 2020)

---

## Session I: Cognitive Modeling (September 18 - September 20, 2020, 7/24 Available)

Chair: Dr. Yujiu Yang (Tsinghua University, Shenzhen, China)

20 minutes for each paper.

### [End-to-End Nested Multi-Attention Network for 3D Brain Tumor Segmentation \(ICCC2020-6001\)](#)

*Xinrui Zhuang (Tsinghua University, China); Yujiu Yang (Tsinghua University, Shenzhen, China)*

Abstract: Utilizing the powerful feature learning ability of deep learning, researchers have proposed a variety of effective methods for brain tumor segmentation in three-dimensional medical images. However, the existing approaches often need to be processed in stages pipeline, without considering the anatomical nested structural characteristics of brain tumors, thus affecting the accuracy and efficiency of tumor segmentation. In this paper, we propose the Nested Multi-Attention mechanism Network (NMA-Net) tailored for brain tumors, which can effectively connect the sub-segmentation tasks of different organizations, and can directly conduct end-to-end training. By using the segmentation result of the tumor peripheral tissue as a kind of soft attention to the tumor segmentation task, it can make the subsequent network focus more on the region of interest, which makes it possible to obtain more accurate segmentation results. Besides, we transform multi-class segmentation tasks into multiple binary sub-segmentation tasks. Experiments on the BraTS'2017 Challenge Dataset show that the proposed NMA-Net framework is very suitable for organ tissue segmentation with nested anatomical structures. Here, our single-view model achieves the best segmentation performance compared with the exiting approaches, and the multi-view fusion model also achieves the state-of-the-art performance on the TC and ET sub-regions.

### [Improve Fake Products Detection with Aspect-based Sentiment Analysis \(ICCC2020-6002\)](#)

*Jiaming Li (Harbin Institute of Technology, Shenzhen & South China University of Technology, China); Yonghao Fu (Harbin Institute of Technology (Shenzhen), China); Daoxing Liu (Harbin Institute of Technology at Shenzhen, China); Ruifeng Xu (Harbin Institute of Technology, Shenzhen, China)*

Abstract: With the development of e-commerce, the number of counterfeit products is increasing and the rights and interests of customers have been seriously infringed. A product can be evaluated by reviews and rating objectively. However, the topics of reviews are diverse while customers tend to focus on only a few aspects, and many reviews have wrong scores that are inconsistent with the content. Natural language processing (NLP) is helpful to mining the opinion of reviews automatically. In this paper, the goal is to improve fake products detection through text classification technology. Precisely, we use CNN and LSTM model to judge whether the review is quality related or not, which can remove useless reviews, and aspect-based sentiment analysis with attention mechanism to determine the sentiment polarity of the concerned aspect to get ratings for different aspects. We experiment on the Self-Annotated dataset and results show that by using text classification technology, the performance of fake product detection can be greatly improved.

### [HFF: Hybrid Feature Fusion Model for Click-Through Rate Prediction \(ICCC2020-6003\)](#)

*Yunzhou Shi (Tsinghua University, China); Yujiu Yang (Tsinghua University, Shenzhen, China)*

Abstract: Deep neural network (DNN) which is applied to extract high-level features plays an important role in the Click Through Rate (CTR) task. Although the necessity of high-level features has been recognized, how to integrate high-level features with low-level features has not been studied well. There are some works fuse low- and high-level features by simply sum or concatenation operations. We argue it is not an effective way because they treat low- and high-level features equally. In this paper, we propose a novel hybrid feature fusion model named HFF. HFF model consists of two different layers: feature interaction layer and feature fusion layer. With feature interaction layer, our model can capture high-level features. And the feature fusion layer can make full use of low- and high- level features. Comprehensive experiments on four real-world datasets are conducted. Extensive experiments show that our model outperforms existing the state-of-the-art models.

## Session II: Sensing Intelligence (September 18 - September 20, 2020, 7/24 Available)

Chair: Dr. Lei Yu (IBM Thomas J. Watson Research Center, USA)

20 minutes for each paper.

### [A Dual Layer Regression Model for Cross-border E-commerce Industry Sale and Hot Product prediction \(ICCC2020-6004\)](#)

*Wangda Luo (Harbin Institute of Technology (Shenzhen), China); Hang Su (Harbin Institute of Technology, Shenzhen, China); Yuhao Liu (Harbin Institute of Technology, Shenzhen & Chinese Nankai University, China); Ruifeng Xu (Harbin*

*Institute of Technology, Shenzhen, China)*

**Abstract:** We introduce a new regression model for time series forecasting in cross-border e-commerce domain. We focus on industry sale prediction and hot product prediction. E-commerce products contain many attributes, and to fully utilize these features for better results, our model employs a novel dual layer regression architecture to improve the generalization. The first layer is a time series regression model which aims at capturing correlation between the history data and future data. The second layer is a features regression model which can enhance the relationship of features and target value in each time points. To verify the effectiveness of our proposed model, we establish two cross-border e-commerce datasets about imported lipsticks and shoes. Then we conduct extensive experiments for industry sale prediction and hot product prediction. The experimental performance demonstrates that our proposed model achieves impressive results compared to other strong baselines and the precision of hot product prediction reached 90%.

**PRTransE: Emphasize more important facts based on Pagerank for Knowledge Graph Completion (ICCC2020-6005)**

*Zhongwen Li (Harbin Institute of Technology, China); Bin Zhang (Peng Cheng Laboratory, China); Yang Liu (Harbin Institute of Technology, Shenzhen & Peng Cheng Laboratory, China); Qing Liao (Harbin Institute of Technology (Shenzhen), China)*

**Abstract:** Knowledge graph is a hot research field in the direction of artificial intelligence. The task of knowledge graph completion is to predict the links between entities. Translation-based models (such as TransE, TransH, and TransR) are a class of well-known knowledge graph completion methods. However, most existing translation-based models ignore the importance of triples in the completion process. In this paper, we propose a novel knowledge graph completion model PRTransE, which considers the importance information of triples based on PageRank and combines the importance information of triples with knowledge graph embedding. Specifically, PRTransE integrates the entity importance and relationship importance of the triplet at the same time, and adopts different processing methods for the importance information of the positive and negative tuples, so that pay adaptive attention to different triplet information in the learning process and improve learning performance to achieve better completion effect. Experimental results show that, in two real-world knowledge graph datasets, PRTransE has the best overall performance compared to the five comparison models.

**ALBERT-based Chinese Named Entity Recognition (ICCC2020-6006)**

*Haifeng Lv (Kingdee International Software Group Company Limited, China); Yishuang Ning (Kingdee International Software Group Company Limited); Ke Ning (Kingdee International Software Group Company Limited, China)*

**Abstract:** Chinese named entity recognition (NER) has been an important problem in natural language processing (NLP) field. Most existing methods mainly use traditional deep learning models which cannot fully leverage contextual dependencies that are very important for capturing the relations between words or characters for modeling. To address this problem, various language representation methods such as BERT have been proposed to learn the global context information. Although these methods can achieve good results, the large number of parameters limited the efficiency and application in real-world scenarios. To improve both of the performance and efficiency, this paper proposes an ALBERT-based Chinese NER method which uses ALBERT, a Lite version of BERT, as the pre-trained model to reduce model parameters and to improve the performance through sharing cross-layer parameters. Besides, it uses conditional random field (CRF) to capture the sentence-level correlation information between words or characters to alleviate the tagging inconsistency problems. Experimental results demonstrate that our method outperforms the comparison methods over 4.23-11.17% in terms of relative F1-measure with only 4% of BERT's parameters.

**Session III: Cognitive Computing Technologies and Frameworks (September 18 - September 20, 2020, 7/24 Available)**

**Chair:** Dr. Yujiu Yang (Tsinghua University, Shenzhen, China)  
20 minutes for each paper.

**Context Based Quantum Language Model with Application to Question Answering (ICCC2020-6007)**

*Qin Zhao (Harbin Institute of Technology, Shenzhen, China); Chenguang Hou (National University of Singapore, CRISP, Singapore); Ruifeng Xu (Harbin Institute of Technology, Shenzhen, China)*

**Abstract:** Equipped with quantum probability theory, quantum language models (QLMs) aimed at a principled approach to modeling term dependency have drawn increasing attention. However, even though they are theoretically more general and have effective performance, current QLMs do not take context information into account. The most important element, namely density matrix, is constructed as a summation of word projectors, whose representation is independent of context information. To address this problem, we propose a Context based Quantum Language Model (C-QLM). Between word embedding and sentence density matrix, a bidirectional long short term memory network is adopted to learn the hidden context information. Then a set of vectors is utilized to extract density matrices's features for question and answer sentences which can be used to calculate the matching score. Experiment results on TREC-QA and WIKI-QA datasets demonstrate the effectiveness of our proposed model.



**Semantic Enhancement based Dynamic Construction of Domain Knowledge Graph** (ICCC2020-6008)

*Yao Sun (Nanjing Institute of Big Data, Jinling Institute of Technology, China); Lun Meng (Hohai University, China); Yan Zhang (Jinling Institute Of Technology, China)*

Abstract: Knowledge graph (KG) is one of key technologies for intelligently answering questions, which can reduce customer service's costs and improve its self-service capabilities. However, the description of questions is often ambiguous, and the operation and maintenance of online KG based QA services introduces a high cost. To address the above issues, this paper proposes a semantic enhancement based dynamic construction of domain knowledge graph for answering questions. We first employ a model combining LSTM and CRF to identify entities, and then propose a semantic enhancement method based on topic comparison to introduce external knowledge. We employ heuristic rules to get optimal answers, and then periodically update the global KG according to the integer linear programming solver's results. Our approach can achieve a high precise answering results with a low response delay by accurately recognizing entities, automatically mapping domain knowledge to the KG, and online updating the KG. The experimental results show that our approach compared with the traditional method improves the precision, recall and F-measure by 6.41%, 16.46% and 11.17%, respectively.

**Traffic Incident Detection from Massive Multivariate Time-Series Data** (ICCC2020-6009)

*Nicholas A Sterling and John Miller (University of Georgia, USA)*

Abstract: Smart-city infrastructure has the potential to improve the lives of anyone who finds themselves in an urban environment. However, Vehicle Traffic Management (VTM) which ameliorates traffic congestion to improve productivity and reduce commuter stress is hampered by traffic incidents which introduce unexpected and chaotic variability into the traffic network. Automatic Incident Detection (AID) mechanisms aim to quickly and reliably identify vehicle traffic incidents to reduce the effect of traffic incidents on VTM. This paper shows that widely available magnetic traffic sensor data can be used with the AdaBoost Machine Learning (ML) model to produce a reliable, light-weight AID mechanism to assist with VTM. A comparison with other ML models is also presented as well as ideas for future work.

**Cognitive and Predictive Analytics on Big Open Data** (ICCC2020-6010)

*Carson K. Leung (University of Manitoba, Canada)*

Abstract: Nowadays, big data are everywhere because huge amounts of valuable data can be easily generated and collected from a wide variety of data sources at a rapid rate. Embedded into these big data are implicit, previously unknown and potentially useful information and valuable knowledge that can be discovered by data science. Due to their value, these big data are often considered as the new oil. In recent years, many governments make their collected big data freely available to their citizens, who could then gain some insights about services available in the city from these open data. In this paper, we present a cognitive and predictive analytic approach to analyze open data for discovering interesting patterns such as tipping patterns. In general, tipping is a voluntary action conceived as social norm that is valuable to service workers in many countries. With the introduction of ride hailing services, traditional taxi services have began facing increased competition. As such, there are increasing interests in factors that are associated with generous tips. Hence, to evaluate the practicality of our approach, we conduct a case study on applying our approach to transportation data (e.g., taxi trip records) from New York City (NYC) to examine and predict tip generosity. Although we conducted the case study on NYC data, our presented approach is expected to be applicable to perform cognitive and predictive analytics on big open data from other cities.

# 2020 International Conference on Internet of Things (ICIOT 2020)

---

## Session I: IoT Services (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Wei Song (North China University of Technology, China)

20 minutes for each paper.

### Combining Individual and Joint Networking Behavior for Intelligent IoT Analytics (ICIOT2020-7001)

*Jeya Vikranth Jeyakumar (University of California, Los Angeles, USA); Ludmila Cherkasova (ARM Research, USA); Mani B. Srivastava (University of California, Los Angeles, USA)*

**Abstract:** The IoT vision of a trillion connected devices over the next decade requires reliable end-to-end connectivity and automated device management platforms. While we have seen successful efforts for maintaining small IoT testbeds, there are multiple challenges for the efficient management of large-scale device deployments. With Industrial IoT, incorporating millions of devices, traditional management methods do not scale well. In this work, we address these challenges by designing a set of novel machine learning techniques, which form a foundation of a new tool, IoTelligent, for IoT device management, using traffic characteristics obtained at the network level. The design of our tool is driven by the analysis of 1-year long networking data, collected from 350 companies with IoT deployments. The exploratory analysis of this data reveals that IoT environments follow the famous Pareto principle, such as: (i) 10% of the companies in the dataset contribute to 90% of the entire traffic; (ii) 7% of all the companies in the set own 90% of all the devices. We designed and evaluated CNN, LSTM, and Convolutional LSTM models for demand forecasting, with a conclusion of the Convolutional LSTM model being the best. However, maintaining and updating individual company models is expensive. In this work, we design a novel, scalable approach, where a general demand forecasting model is built using the combined data of all the companies with a normalization factor. Moreover, we introduce a novel technique for device management, based on autoencoders. They automatically extract relevant device features to identify device groups with similar behavior to flag anomalous devices.

### BWCNN: Blink to Word, a Real-Time Convolutional Neural Network Approach (ICIOT2020-7002)

*Albara Ah Ramli, Rex Liu, Rahul Krishnamoorthy and Vishal I b (University of California Davis, USA); Xiaoxiao Wang (University of California, Davis, USA); Ilias Tagkopoulos and Xin Liu (University of California Davis, USA)*

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease of the brain and the spinal cord, which leads to paralysis of motor functions. Patients retain their ability to blink, which can be used for communication. Here, We present an Artificial Intelligence (AI) system that uses eye-blinks to communicate with the outside world, running on real-time Internet-of-Things (IoT) devices. The system uses a Convolutional Neural Network (CNN) to find the blinking pattern, which is defined as a series of Open and Closed states. Each pattern is mapped to a collection of words that manifest the patient's intent. To investigate the best trade-off between accuracy and latency, we investigated several Convolutional Network architectures, such as ResNet, SqueezeNet, DenseNet, and InceptionV3, and evaluated their performance. We found that the InceptionV3 architecture, after hyper-parameter fine-tuning on the specific task led to the best performance with an accuracy of 99.20% and 94ms latency. This work demonstrates how the latest advances in deep learning architectures can be adapted for clinical systems that ameliorate the patient's quality of life regardless of the point-of-care.

### Risk Assessment of Vehicle Sensor Data as a Vending Object or Service (ICIOT2020-7003)

*Frank Bodendorf (Friedrich-Alexander-University of Erlangen-Nuremberg, Germany); Jörg Franke (FAU Erlangen-Nuremberg, Germany)*

**Abstract:** Connected cars generate a huge amount of vehicle data during operation. In the future, the amount of sensor-generated data will continue to increase. The connectivity of the cars, more powerful processors, and improved telematics and navigation systems will allow the amount of data to grow further. Vehicle data provides a basis for a large number of business models. In addition to selling vehicles, automobile manufacturers can generate additional revenue by selling vehicle-generated data as goods or services. First, a typology of vehicle data is described in order to derive value potentials of data products. Motivated by the value perspective, risks in data transfer to third parties are often neglected. In order to assess these risks, a new risk management model for intangible products like data is presented. The main phases of the risk assessment procedure are walked through, outlining possible criteria and metrics in each phase. Finally, the model is demonstrated by evaluating risks of data transfer to third parties in the automotive industry, using the example of vehicle-generated road segment data (RSD).

## Session II: IoT Applications (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Kisung Lee (Louisiana State University--Baton Rouge, USA)

20 minutes for each paper.

### [NACK-based reliable multicast communication for Internet of Things firmware update \(ICIOT2020-7004\)](#)

*Jiye Park (University of Duisburg-Essen, Germany); Dongha Lee (Germany); Markus Jung (OSRAM GmbH, Germany); Erwin P Rathgeb (University of Duisburg-Essen, Germany)*

Abstract: The demand for efficient IoT firmware update protocols is increasing. Especially in scenarios with a large number of constrained devices, transferring a big amount of data, like a firmware image file, over a constraint network takes a long time to complete. During this time the functionality of the devices may be reduced. Therefore the firmware update is a critical use case for IoT. Multicast group communication can shorten the transmission time and use the network bandwidth efficiently. However, current IoT protocols using multicast cannot guarantee reliability which is most important for firmware file transmission. Furthermore existing solutions for reliable multicast cause a significant network overhead which can be prohibitive for a constrained IoT environment. To address these problems, we propose a reliable multicast solution employing a Negative ACK (NACK) mechanism that can be integrated with the Constraint Application Protocol (CoAP) widely used in IoT. Our protocol mitigates network congestion by reducing the number of packets that have to be sent while keeping the message size small making it a suitable solution even with a large number of devices. What is more, our proposal does not require an additional stack. In order to demonstrate the feasibility and effectiveness of our proposal, we carried out a real-world evaluation in a wireless mesh network testbed.

### [On Burial Depth of Underground Antenna in Soil Horizons for Decision Agriculture \(ICIOT2020-7005\)](#)

*Abdul Salam and Usman Raza (Purdue University, USA)*

Abstract: Decision agriculture is the practice of accurately capturing the changing parameters of the soil including water infiltration and retention, nutrients supply, acidity, and other time changing phenomena by using the modern technologies. Using decision agriculture, fields can be irrigated more efficiently hence conserving water resources and increasing productivity. Internet of Underground Things (IOUT) is being used to monitor the soil for smart irrigation. Communication in wireless underground sensor networks is affected by soil characteristics such as soil texture, volumetric water content (VWC) and bulk density. These soil characteristics vary with soil type and soil horizons within the soil. In this paper we have investigated the effects of these characteristics by considering Holdrege soil series and homogeneous soil. It is shown that consideration of soil characteristics of different soil horizons leads to 6 dB improved communication in wireless underground communications.

### [A WiVi based IoT Framework for Detection of Human Trafficking Victims kept in Hideouts \(ICIOT2020-7006\)](#)

*Amir H Gandomi (University of Technology Sydney, Australia)*

Abstract: Human trafficking is the trade of humans for the purpose of forced labor, sexual slavery, or commercial sexual exploitation for the trafficker or others. The traffickers often trick, defraud, or physically force victims into selling sex and forced labor. In others, victims are lied to, assaulted, threatened, or manipulated into working under inhumane, illegal, or otherwise unacceptable conditions. According to the estimation of the International Labor Organization, there are more than 40.3 million victims of human trafficking globally. It is a threat to the Nation as well as to humanity. There have been many efforts by government agencies NGOs to stop human trafficking and rescuing victims, but the traffickers are getting smarter day by day. From multiple sources, it is observed that the traffickers generally hide humans in hidden rooms, sealed containers, and boxes disguised as goods. This congestion results in Critical mental and physical damages in some cases. It is practically impossible to physically go and check each box, containers or rooms. So in this paper, we propose a Wireless Vision-based IoT framework, which uses the reflection of WiFi radio waves generated by WiFi to detect the presence of humans inside a cement or metal enclosure from outside.

## Session III: Industrial IoT (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Zhisheng Yan (Georgia State University, USA)

20 minutes for each paper.

### [RelIoT: Reliability Simulator for IoT Networks \(ICIOT2020-7007\)](#)

*Kazim Ergun, Xiaofan Yu and Nitish Nagesh (University of California San Diego, USA); Ludmila Cherkasova (ARM Research, USA); Pietro Mercati and Raid Ayoub (Intel Corporation, USA); Tajana Simunic Rosing (University of California, San Diego, USA)*

Abstract: The next era of the Internet of Things (IoT) calls for a large-scale deployment of edge devices to meet the growing demands of applications such as smart cities, smart grids, and environmental monitoring. From low-power sensors to multi-core platforms, IoT devices are prone to failures due to the reliability degradation of electronic circuits, batteries, and other components. As the network of heterogeneous devices expands, maintenance costs due to system failures become unmanageable, making reliability a major concern. Prior work has shown the importance of automated reliability

management for meeting lifetime goals for individual devices. However, state-of-the-art network simulators do not provide reliability modeling capabilities for IoT networks. In this paper, we present an integrated reliability framework for IoT networks based on the ns-3 simulator. The lack of such tools restrained researchers from doing reliability-oriented analysis, exploration, and predictions early in the design cycle. Our contribution facilitates this, which can lead to the design of new network reliability management strategies. The proposed framework, besides reliability, incorporates three other interrelated models - power, performance, and temperature - which are required to model reliability. We validate our framework on a mesh network with ten heterogeneous devices, of three different types. We demonstrate that the models accurately capture the power, temperature, and reliability dynamics of real networks. We finally simulate and analyze two examples of energy-optimized and reliability-optimized network configurations to show how the framework offers an opportunity for researchers to explore trade-offs between energy and reliability in IoT networks.

**Spatio-Temporal Coverage Enhancement in Drive-By Sensing Through Utility-Aware Mobile Agent Selection (ICIOT2020-7008)**

*Navid Hashemi Tonekaboni (College of Charleston, USA); Lakshmish Ramaswamy, Deepak Mishra, Omid Setayeshfar and Soroush Omidvar (University of Georgia, USA)*

**Abstract:** In recent years, the drive-by sensing paradigm has become increasingly popular for cost-effective monitoring of urban areas. Drive-by sensing is a form of crowdsensing wherein sensor-equipped vehicles (aka, mobile agents) are the primary data gathering agents. Enhancing the efficacy of drive-by sensing poses many challenges, an important one of which is to select non-dedicated mobile agents on which a limited number of sensors are to be mounted. This problem, which we refer to as the mobile-agent selection problem, has a significant impact on the spatio-temporal coverage of the drive-by sensing platforms and the resultant datasets. The challenge here is to achieve maximum spatiotemporal coverage while taking the relative importance levels of geographical areas into account. In this paper, we address this problem in the context of the SCOUTS project, the goal of which is to map and analyze the urban heat island phenomenon accurately. Our work makes several major technical contributions. First, we delineate a model for representing the mobile agents selection problem. This model takes into account the trajectories of the vehicles (public transportation buses in our case) and the relative importance of the urban regions, and formulates it as an optimization problem. Second, we provide an algorithm that is based upon the utility (coverage) values of mobile agents, namely, a hotspot-based algorithm that limits the search space to important sub-regions. Third, we design a highly efficient coverage redundancy minimization algorithm that, at each step, chooses the mobile agent, which provides maximal improvement to the spatio-temporal coverage. This paper reports a series of experiments on a real-world dataset from Athens, GA, USA, to demonstrate the effectiveness of the proposed approaches.

**Review for Message-Oriented Middleware (ICIOT2020-7009)**

*Yang Liu (Kingdee International Software Group Co. & Research Institute of Information Technology, Tsinghua University, China)*

**Abstract:** With the rapid development of 5G technology and micro-service software development technology, a growing number of intelligent devices are connected to the Internet, and it has become a trend to provide services to customers in a collaborative way. As the information exchange center of collaborative services, the role of message-oriented middleware becomes more and more important. Firstly, this paper describes the definition, main characteristics, core technology, main products and existing problems of middleware. Then, it takes one of the problems as an example, designs a IoT gateway architecture based on a message middleware, and introduces its application in the Internet of things in detail.

**Session IV: IoT Device Innovations (September 18 - September 20, 2020, 7/24 Available)**

Chair: Prof. Kisung Lee (Louisiana State University--Baton Rouge, USA)

20 minutes for each paper.

**Development of an Electronic System for the Analysis and Integration of Data on Water Care (ICIOT2020-7010)**

*Febe Hernandez and Martha S Lopez-de la Fuente (Universidad de Monterrey, Mexico)*

**Abstract:** This work exposes the development of an electronic system for the detection of pollutant agents in water and integrates the design of the system with methods for collection of information and the detection of polluting particles in deposits. For this purpose, the stimulation is applied by dielectrophoresis, which manipulates the particles suspended in water, so that it is possible to perform a detection, extraction and accounting process to obtain data of the contaminants by doing experiment on liquid samples. The ultimate purpose is the anticipated detection of possible problems related to contaminants in the water. The finished design and the functional prototype show experiments in the detection of live and inert particles in the range of 5  $\mu\text{m}$  and 10  $\mu\text{m}$ , and the methodology to carry out the proposed analysis. The elements of the system integrate the use of an LM3S6965 processor, which is responsible for processing the input parameters that pass through a signal conditioner, also it is responsible for the electrical stimulation necessary for the experiment and then using a microscope it's possible to generate images of the results. For the experiments shown there is the generation of signals of 5Vpp with a frequency of 28kHz, applied to a mixture of 10um polystyrene particles and 5um yeast cells, obtaining results about the behavior of these particles. The main objective of the research carried out is to develop a solution to the problem

that exists with the current methods that are not portable and to implement a system based on the System on Chip (SoC) architecture. The results of this work presents a functional prototype of the system, and shows the detection of contaminants and a data transmitter for the collection of information. Also, the experiments made show the handling of inert and living particles with the use of polystyrene and yeast beads. One of the main characteristics of the system is that it could be adjusted to read parameters in other liquid containers to detect different particles.

**Image Privacy Protection by Particle Swarm Optimization based Pivot Pixel Modification** (ICIOT2020-7011)

*Jishen Yang (Georgia State University, USA); Yan Huang (Kennesaw State University, USA); Zhenzhen Xie (Jilin University, China); Junjie Pang (Qingdao University, China); Wei Li (Georgia State University, USA)*

**Abstract:** The image classification models based on neural networks recently have outperformed most of the traditional models, and by a very high speed, have been developed and implemented by industry because of the capability of qualifying almost any computer vision tasks. Hence, the exposure of users' image data to unauthorized powerful models causes more information leaks in a shorter time. Through experiments, we find that for one input image, the sensitivity of the change of image classification model prediction scores to each pixels' RGB value change is different and the pattern of the sensitivity on each pixel is highly related to the category and composition of the input image. By utilizing this feature, we present Pivot Pixel Noise Generator by Particle Swarm Optimization to generate noise points on original images to lower the target model's accuracy of correctly predicting the target image's label, so to protect the information contained in the target image from the image classification models. The model performs in a half-black-box manner and balances the number of queries to the target and the total number of modified points. We also propose an initialization strategy for the model, PSO Knowledge Transfer, which initializes the model's parameters with experience learned from previous runs to further reduce the number of query times and noise points. The model is evaluated using the image classification benchmark model \$ResNet50\$ and shows an advantage compared to the baseline algorithm.

**Deriving Interpretable Rules for IoT Discovery through Attention** (ICIOT2020-7012)

*Franck Le (IBM T. J. Watson, USA); Mudhakar Srivatsa (IBM T.J. Watson Research Center, USA)*

**Abstract:** Due to their high vulnerability, IoT has become a primary target for cybercriminals (e.g., botnets, network infiltration). As a result, many solutions have been developed to help users and administrators identify IoT devices. While solutions based on deep learning have been shown to outperform traditional approaches in other domains, their lack of explanation and their inference latency present major obstacles for their adoption in network traffic analysis, where throughputs of Gbps are typically expected. Extracting rules from a trained neural network presents a compelling solution, but existing methods are limited to feedforward networks, and RNN/LSTM. In contrast, attention-based models are a more recent architecture, and are replacing RNN/LSTM due to their higher performance. In this paper, we therefore propose a novel efficient algorithm to extract rules from a trained attention-based model. Evaluations on actual packet traces of more than 100 IoT devices demonstrate that the proposed algorithm reduces the storage requirements and inference latency by 4 orders of magnitude while still achieving an average f1-score of 0.995 and a fidelity score of 98.94%. Further evaluation on an independent dataset also shows improved generalization performance: The extracted rules achieve better performance, especially thanks to their inherent capability to identify unknown devices.



# 2020 International Conference on Edge Computing (EDGE 2020)

---

## Session I: Edge Analytics (September 18 - September 20, 2020, 7/24 Available)

Chair: Dr. Ajay Katangur (Missouri State University, USA)

20 minutes for each paper.

### Survey of Edge Computing Framework and Some Recommendation (EDGE2020-8001)

*Yiwen Sun and Bihai Zhang (Georgia Institute of Technology, China); Min Luo (Georgia Institute of Technology, USA)*

Abstract: The fast adoption and success of IoT and 5G related technology, accompanied by the ever-increasing critical demand for better QoS, revolutionized the paradigm shift from centralized cloud computing to some combination of distributed edge computing and traditional cloud computing. There are substantial researches and reviews on edge computing, and several industry-specific frameworks were proposed, but general purpose frameworks that could enable speedy utilization of millions of innovated business/IT services worldwide across the entire spectrum of the current computing paradigm is not yet properly addressed. We first proposed a generalized and service-oriented edge computing framework, based on a relatively complete survey of recent publications, then we conducted an in-depth analysis of selected works from both academia and industry aimed to access the maturity, and the gaps in this arena. Finally, we summarize the challenges and opportunities in edge computing, and we hope that this paper can inspire significant future improvements.

### IoT Digital Forensics Readiness in the Edge: A Roadmap for Acquiring Digital Evidences from Intelligent Smart Applications (EDGE2020-8002)

*Andrii Shalaginov (Norwegian University of Science and Technology, Norway); Asif Iqbal (KTH Royal Institute of Technology & ACFE, ISACA, IIA, ISC2, Sweden); Johannes Olegård (KTH Royal Institute of Technology, Sweden)*

Abstract: Entering the era of the Internet of Things, the traditional Computer Forensics is no longer as trivial as decades ago with a rather limited pool of possible computer components. It has been demonstrated recently how the complexity and advancement of IoT are being used by malicious actors attack digital and physical infrastructures and systems. The investigative methodology, therefore, faces multiple challenges related to the fact that billions of interconnected devices generate tiny pieces of data that easily comprehend the Big Data paradigm. As a result, Computer Forensics is no longer a simple methodology of the straightforward process. In this paper, we study the complexity and readiness of community-accepted devices in a smart application towards assistance in criminal investigations. In particular, we present a clear methodology and involved tools related to Smart Applications. Relevant artefacts are discussed and analysed using the prism of the Digital Forensics Process. This research contributes towards increased awareness of the IoT Forensics in the Edge, corresponding challenges and opportunities.

### Edge Architecture for Dynamic Data Stream Analysis and Manipulation (EDGE2020-8003)

*Orpaz Goldstein (UCLA, USA); Anant Shah (Verizon, USA); Derek Shiell, Mehrdad Arshad Rad and William Pressly (Verizon Digital Media, USA); Majid Sarrafzadeh (UCLA, USA)*

Abstract: The exponential growth in IoT and connected devices featuring limited computational capabilities requires the delegation of computation tasks to cloud compute platforms. Edge compute tasks largely involve sending data from an edge compute device to a central location where data is processed and returned to the edge device as a response. Since most edge network infrastructure is restricted in its ability to dynamically delegate computation while retaining context, these events are commonly limited to a predefined task that the edge function is modeled to process and respond to. Edge functions traditionally handle isolated events or periodic updates, making them ill-suited for continuous tasks on streaming data. We propose a decentralized, massively scalable architecture of modular edge compute components which dynamically defines computation channels in the network, with emphasis on the ability to efficiently process data streams from a large amount of producers and support a large amount of consumers in real time. We test this architecture on real-world tasks, involving chaining of edge functions, context retention, and machine learning models on the edge, demonstrating its viability.

## Session II: Edge AI and Industry-Specific Edges (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Shih-Chun Lin (North Carolina State University, USA)

20 minutes for each paper.

### Small, Accurate, and Fast Re-ID on the Edge: the SAFR Approach (EDGE2020-8004)

*Abhijit Suprem and Calton Pu (Georgia Institute of Technology, USA); Joao Ferreira (Instituto de Matemática e Estatística, Brazil)*



**Abstract:** We propose a Small, Accurate, and Fast Re-ID (SAFR) design for flexible vehicle re-id under a variety of compute environments such as cloud, mobile, edge, or embedded devices by only changing the re-id model backbone. Through best-fit design choices, feature extraction, training tricks, global attention, and local attention, we create a re-id model design that optimizes multi-dimensionally along model size, speed, & accuracy for deployment under various memory and compute constraints. We present several variations of our flexible SAFR model: SAFR-Large for cloud-type environments with large compute resources, SAFR-Small for mobile devices with some compute constraints, and SAFR-Micro for edge devices with severe memory & compute constraints. SAFR-Large delivers state-of-the-art results with mAP 81.34 on the VeRi-776 vehicle re-id dataset (15% better than related work). SAFR-Small trades a 5.2% drop in performance (mAP 77.14 on VeRi-776) for over 60% model compression and 150% speedup. SAFR-Micro, at only 6MB and 130MFLOPS, trades 6.8% drop in accuracy (mAP 75.80 on VeRi-776) for 95% compression and 33x speedup compared to SAFR-Large.

#### **A Microservice-based Industrial Control System Architecture Using Cloud and MEC (EDGE2020-8005)**

*Yu Kaneko (Toshiba, Japan)*

**Abstract:** Cloud computing has been adapted for various application areas. Several research projects are underway to migrate Industrial Control Systems (ICSs) to the public cloud. Some functions of ICSs require real-time processing that is difficult to migrate to the public cloud because network latency of the internet is unpredictable. Fog computing is a new computing paradigm that could address this latency issue. In particular, Multi-access Edge Computing (MEC) is a fog computing environment integrated with the 5G network, and therefore the real-time processing requirement of ICSs could be satisfied by using MEC. In this paper, we propose a microservice-based ICS architecture using the cloud and fog computing. In the architecture, each function of an ICS is implemented as a microservice and its execution locations are determined by an algorithm minimizing the total usage fee for cloud and fog computing while satisfying the real-time processing requirement. The proposed architecture and placement algorithm are evaluated by simulation under the scenario of a virtual power plant that manages distributed energy resources. The simulation result shows the proposed placement algorithm suppresses VM usage fee while satisfying the requirement of a real-time control function.

#### **Big k-Space Data in the Imaging Process Using Mathematical and Statistical Analysis (EDGE2020-8006)**

*Siwoo Kim (Rise Research Group, USA); Andrew Kyung (NVRHS, USA)*

**Abstract:** In this research, computational simulations and statistical analysis were performed with several modified mathematical functions to improve the resolution of digital images. Proposed alternative functions as new low pass filters were proved to save operation time and process. In the imaging process, when the domain of the proposed function over the frequency domain is narrow, it showed that the resolution of the final image was low due to the insufficient amount of frequency data from K-space. The main purpose of this research was to find a better low pass filter that would both improve the quality of the resolution of an image using mathematical, statistical and computational analysis. Result shows work time was decreased by a substantial amount of time to produce the final image. When the width of the function over the K space domain was narrow, low quality of image was produced due to the insufficient amount of frequency data from K-space. Higher quality of image was obtained using the proposed LPF with a certain width of domain. Also non-traditional function and its behavior of image statistics were studied in this analytic analysis.

### **Session III: Edge Storage and Processing Engine (September 18 - September 20, 2020, 7/24 Available)**

Chair: Dr. Jinpeng Wei (University of North Carolina at Charlotte, USA)

20 minutes for each paper.

#### **Godec: An open-source data processing framework for deploying ML data flows in edge-computing environments (EDGE2020-8007)**

*Ralf Meermeier (Formerly Raytheon BBN Technologies, USA); Le Zhang, Francis Keith and William Hartmann (Raytheon BBN Technologies, USA); Stavros Tsakalidis (BBN Technologies, USA); Andrew Tabarez (Raytheon BBN Technologies, USA)*

**Abstract:** We present Godec, a C++-based framework that allows easy transition of complex machine learning (ML) data flows to edge-computing environments where common data processing frameworks do not apply. Godec allows for free mixing of technologies such as Kaldi, TensorFlow and custom modules, all wrapped into a single OS process, making it easy to deploy inference engines on constrained environments like Android, iOS or embedded Linux. Godec achieves this by connecting the components into an arbitrary graph specified by a simple JSON file during startup. Despite being multithreaded, results between runs are guaranteed identical, allowing for immediate transition from offline experiments to deployment. The source code is released under the MIT license, with the authors hoping that the community will find it a useful tool to create their own components for it, in turn enabling others to mix and merge disparate technologies into applications of their own.

#### **Edge Storage Solution Based on Computational Object Storage (EDGE2020-8008)**

*Lijuan Zhong (Seagate Technology, USA)*

**Abstract:** Emerging computing/storage architecture provides new opportunities and requirements for multimedia data storage, especially at the edge (close to where the data is captured). Computational storage, defined as an architecture that conducts data processing at the storage layer so as to offload host processing or reduce data movement, allows raw data to be analyzed as the data are stored. As a consequence, the data to be stored may intrinsically carry richer metadata. Meanwhile object storage is a data storage architecture that organizes data into flexible-sized data containers, named objects. Combining object and computational storage, this paper described an edge data storage platform built on a representative computational object storage with content indexed object keys. The platform provides both computing and storage scalability for Edge applications while concurrently managing the richer metadata generated in a structured way to promote future information retrieval. Using video data as a sample use case, the concept of object key design is illustrated.

**fogcached: DRAM-NVM Hybrid Memory-Based KVS server for Edge Computing (EDGE2020-8009)**

*Midori Sugaya (Shibaura Institute of Technology, Japan); Takahiro Hirofuchi (National Institute of Advanced Industrial Science and Technology, Japan); Ryousei Takano (National Institute of Advanced Industrial Science and Technology (AIST), Japan)*

**Abstract:** With the widespread use of sensors in smart devices and robots, there is a growing expectation for edge computing that processes data not on distant cloud servers but also on or near interactive devices to store their data with low latency access. To satisfy these requirements, we consider a new edge computing system that consists of a hybrid main memory with a KVS (Key-Value-Store) server utilizing the DRAM and nonvolatile main memory (NVM). It provides large-capacity cache memory in a server, supporting high-speed processing and quick response for sensor nodes. However, since existing KVS servers are not designed for NVM, there are less satisfactory implementations that achieve low response time and high throughputs. We propose a novel hybrid KVS server that is designed and implemented on the Memcached distributed memory-caching system, which dynamically moves cached data between two types of memory devices according to access frequency in order to achieve a low latency compared to the existent approaches. We developed a Dual-LRU (Least Recently Used) structure for it. Evaluation was performed using a real machine equipped with NVM. The result showed the proposed method successfully reduced the response time and improves access throughputs.

**Preserving Patients' Privacy in Medical IoT using blockchain (EDGE2020-8010)**

*Bandar Alamri (Lero- The Irish Software Research Center & University of Limerick, Ireland); Ibrahim Tariq Javed (Lero - The Irish Software Research Centre, University of Limerick & Bahria University Islamabad Campus, Ireland); Tiziana Margaria (University of Limerick & Lero, The Irish Software Research Center, Ireland)*

**Abstract:** Medical IoT is a collection of devices and applications that are connected to healthcare systems via the Internet. Wearable devices and body sensors are used to track individuals' medical conditions. The collected data is processed, analyzed, and stored in the cloud platforms to provide healthcare services. The data does not only include personal information like users' identity and location but also consists of sensitive information such as mental status, drug addiction, sexual orientation, and genetics. Therefore, preserving an individual's privacy remains a huge challenge for IoT service providers. The existing techniques significantly reduce the originality of data which affects the application's efficiency. Therefore, in this paper, we propose the idea of using blockchains and smart contract to preserve privacy while obtaining data usability.

# 2020 International Conference on Blockchain (ICBC 2020)

---

## Session I: Fundamental Research for Blockchain (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Zhixiong Chen, PhD (Mercy College, USA)

20 minutes for each paper.

### Comparison of Decentralization in DPoS and PoW Blockchains (ICBC2020-9001)

*Chao Li (Beijing Jiaotong University, China); Balaji Palanisamy (University of Pittsburgh, USA)*

Abstract: Decentralization is a key indicator for the evaluation of public blockchains. In the past, there have been very few studies on measuring and comparing the actual level of decentralization between Proof-of-Work (PoW) blockchains and blockchains with other consensus protocols. This paper presents a new comparison study of the level of decentralization in Bitcoin and Steem, a prominent Delegated-Proof-of-Stake (DPoS) blockchain. Our study particularly focuses on analysing the power that decides the creators of blocks in the blockchain. In Bitcoin, miners with higher computational power generate more blocks. In contrast, blocks in Steem are equally generated by witnesses while witnesses are periodically elected by stakeholders with different voting power weighted by invested stake. We analyze the process of stake-weighted election of witnesses in DPoS and measure the actual stake invested by each stakeholder in Steem. We then compute the Shannon entropy of the distribution of computational power among miners in Bitcoin and the distribution of invested stake among stakeholders in Steem. Our analyses reveal that neither Bitcoin nor Steem is dominantly better than the other with respect to decentralization. Compared with Steem, Bitcoin tends to be more decentralized among top miners but less decentralized in general. Our study is designed to provide insights into the current state of the degree of decentralization in DPoS and PoW blockchains. We believe that the methodologies and findings in this paper can facilitate future studies of decentralization in other blockchain systems employing different consensus protocols.

### On the verification of Smart Contract: A Systematic review (ICBC2020-9002)

*Layth Sliman (EFREI- Paris, France)*

Abstract: Ensuring the correctness of smart contracts is of paramount importance to achieve trust and continuity in the Blockchain-based business process execution. Due to the immutable nature of distributed ledger technology on the blockchain, a smart contract should work as intended before using it. Any bugs or errors will become permanent once published and could lead to huge economic losses. To avoid such problems, verification is required to check the correctness and the security of the smart contract. In this paper, we consider the smart contracts and we investigate the verification of the correctness of the Blockchain-based smart contracts using formal verification methods. We provide an overview of the formal verification (FV) of smart contracts and we present the used methods, tools and approaches. We show a description of each method as well as its advantages and limitations.

### Understanding and Handling Blockchain Uncertainties (ICBC2020-9003)

*Xiwei Xu (Data61-CSIRO & UNSW, Australia); Herath Mudiyansele Nelanga Dilum Bandara (Commonwealth Scientific and Industrial Research Organisation, Australia & University of Moratuwa, Sri Lanka); Qinghua Lu (CSIRO, Australia); Dawen Zhang (UNSW, Australia); Liming Zhu (Data61, CSIRO, Australia)*

Abstract: Blockchain is an immutable and transparent ledger hosted by a peer-to-peer network. However, the unique design of blockchain necessitates managing network-level uncertainties, including complex multi-component interactions and asynchronous messages exchanged between services and the blockchain. It also instigates further uncertainties such as probabilistic immutability of a proof-of-work blockchain and transactions getting discarded due to low transaction fees. These uncertainties need to be understood and properly handled to build reliable applications on top of a blockchain. In this paper, we investigate the uncertainties in blockchains from the point of view of an application or service. We then present empirical evidence on uncertainties from two well-known blockchain platforms (namely Ethereum and Bitcoin) and some of the practices for handling them. Finally, we present a set of conversation patterns to design robust interactions between the application/service and blockchain to handle such uncertainties.

## Session II: Development Platforms for Blockchain (September 18 - September 20, 2020, 7/24 Available)

Chair: Prof. Laizhong Cui (Shenzhen University, China)

20 minutes for each paper.

### ACCTP: Cross Chain Transaction Platform for High-Value Assets (ICBC2020-9004)

*Minfeng Qi, Ziyuan Wang, Donghai Liu and Yang Xiang (Swinburne University of Technology, Australia); Butian*

*Huang (Hangzhou Yunphant Network Technology Co. Ltd, China); Feng Zhou (Vnet Foundation, China)*

Abstract: Recently, with the development of blockchain technology, people not only regard blockchain as a tradable cryptocurrency but also pay attention to other areas where blockchain technology can be applied, such as high-value assets (art, diamonds, etc.). Through the study of some existing blockchain-based platforms for high-value assets, we find that the blockchain ecosystem in this emerging industry presents a state of blooming but isolated from each other. In view of the problem, we discuss some projects based on cross-chain technology and believe that using cross-chain technology is one of the best solutions. Therefore, we propose the ACCTP (Asset Cross Chain Transaction Platform) model to connect different blockchain-based transaction platforms. The model is divided into three cross-chain service layers, called C2IE, C2AE, and C2TS, which are designed to implement the functions of information exchange, asset exchange and transaction exchange respectively.

#### **MULTAV: A Multi-Chain Token Backed Voting Framework for Decentralized Blockchain Governance (ICBC2020-9005)**

*Xinxin Fan, Qi Chai and Zhi Zhong (IoTeX, USA)*

Abstract: Governance is a critical component in cryptocurrency systems for their sustainable development and evolution. In particular, on-chain governance has attracted a lot of attention in cryptocurrency communities after the hard forks of Bitcoin and Ethereum. The on-chain governance mechanisms offered by the existing cryptocurrencies have been implemented on their own blockchains. This approach, while working well for the established cryptocurrencies, raises certain security concerns for newly launched cryptocurrency projects with small market capitalization and initial circulating supply. To mitigate potential attacks against on-chain governance, we present a multi-chain token backed voting framework named MULTAV in this contribution. The MULTAV framework is able to enhance security of the current on-chain governance practices by enabling token holders to vote on multiple established cryptocurrency systems. The instantiation of the MULTAV framework for electing block producers on the IoTeX network demonstrates its feasibility and effectiveness in practice.

#### **Dynamic Role-Based Access Control for Decentralized Applications (ICBC2020-9006)**

*Arnab Chatterjee and Yash Pitroda (Robert Bosch Engineering and Business Solutions Private Ltd., India); Manojkumar Somabhai Parmar (Robert Bosch Engineering and Business Solutions Private Limited, India & HEC Paris, France)*

Abstract: Access control management is an integral part of maintaining the security of an application. For access control management on the Distributed Ledger Technology (DLT), available traditional access control frameworks are inadequate. Existing access control management mechanisms are tightly coupled with the business logic, resulting in an adverse impact on the overall software quality for DLT based Decentralized Applications (dApps). In this paper, we propose a novel framework to implement dynamic role-based access control for dApps. The framework completely decouples the access control logic from the business logic and provides seamless integration with any dApp. The smart contract architecture allows for the independent management of business logic and execution of access control policies. The framework also facilitates secure, low cost, and a high degree of flexibility of access control management. Additionally, it promotes decentralized governance of access control policies and efficient smart contract upgrades. In this paper, we evaluate the framework on relevant various software quality attributes to understand its more profound implications on access control techniques. The framework can be implemented in any smart contract programming language exhibiting Turing completeness. We use the Solidity programming language to implement the framework and discuss the results.

### **Session III: Blockchain Application Reference Architectures and Instances (September 18 - September 20, 2020, 7/24 Available)**

Chair: Dr. Balaji Palanisamy (University of Pittsburgh, USA)

20 minutes for each paper.

#### **Upgradeability Concept for Collaborative Blockchain-based Business Process Execution Framework (ICBC2020-9007)**

*Philipp Klinger, Long Nguyen and Freimut Bodendorf (University of Erlangen-Nuremberg, Germany)*

Abstract: Inter-organizational business processes involve different independent participants to interact with each other to run a collaborative business process. To date, central, trusted third parties mediate between non-trusting participants adding additional process complexity as well as administrative and run costs. Recent research showed that Blockchain and Smart Contracts can replace the role of a central trusted authority in collaborative execution of processes. Smart Contracts, that represent business process logic, cannot be altered due to immutability constraints of Blockchain systems. Yet, with the help of advanced technical upgradeability concepts, upgrading contracts, therefore enabling versioning of processes on a Blockchain is possible. This paper analyzes and implements three different upgradeability concepts. The implemented patterns are evaluated with regard to an existing blockchain-based execution framework for inter-organizational business processes. Our findings suggest the Unstructured Eternal Storage pattern to be the most promising for practical use,

especially regarding cost-effectiveness and added complexity. The findings are derived from simulations on a real-world use case stemming from a large German electronics manufacturing company.

#### **Analysis of Models for Decentralized and Collaborative AI on Blockchain (ICBC2020-9008)**

*Justin D. Harris (Microsoft Montreal, Canada)*

**Abstract:** Machine learning has recently enabled large advances in artificial intelligence, but these results can be highly centralized. The large datasets required are generally proprietary; predictions are often sold on a per-query basis; and published models can quickly become out of date without effort to acquire more data and maintain them. Published proposals to provide models and data for free for certain tasks include Microsoft Research's Decentralized and Collaborative AI on Blockchain. The framework allows participants to collaboratively build a dataset and use smart contracts to share a continuously updated model on a public blockchain. The initial proposal gave an overview of the framework omitting many details of the models used and the incentive mechanisms in real world scenarios. For example, the Self-Assessment incentive mechanism proposed in their work could have problems such as participants losing deposits and the model becoming inaccurate over time if the proper parameters are not set when the framework is configured. In this work, we evaluate the use of several models and configurations in order to propose best practices when using the Self-Assessment incentive mechanism so that models can remain accurate and well-intended participants that submit correct data have the chance to profit. We have analyzed simulations for each of three models: Perceptron, Naïve Bayes, and a Nearest Centroid Classifier, with three different datasets: predicting a sport with user activity from Endomondo, sentiment analysis on movie reviews from IMDB, and determining if a news article is fake. We compare several factors for each dataset when models are hosted in smart contracts on a public blockchain: their accuracy over time, balances of a good and bad user, and transaction costs (or gas) for deploying, updating, collecting refunds, and collecting rewards. A free and open source implementation for the Ethereum blockchain of these models is provided at <https://github.com/microsoft/0xDeCA10B>.

#### **Blockchain Applications in Healthcare - A review and Future Perspective (ICBC2020-9009)**

*Antara Debnath Antu, Anup Kumar and Naga Ramya Sravanthi Narikimilli Narikimilli (University of Louisville, USA); Bin Xie (University of Cincinnati, USA)*

**Abstract:** A digital transformation in health care is the positive impact of technology in health care. Wearable fitness technology, telemedicine, and AI-enabled medical devices are concrete examples of digital transformation in health care. And these are supposed to revolutionize the health care industry by improving patient care, streamline operations and reducing costs but instead, it is facing significant challenges on cybersecurity and privacy of patient data, invoicing and payment processing, medical supply chain, drug integrity. Blockchain technology can absolve the healthcare industry from facing these challenges, it can establish a blockchain of medical records. Blockchain is considered to be a highly secure, transparent and immune to hackers due to its digital encryption, it also plays a prominent role in reducing the intermediate fees as it is entirely decentralized. This review paper scrutinized the potential of blockchain technology to refine the security, privacy, and interoperability of healthcare data and after the detailed analysis of the current significant challenges in healthcare sector, we proposed few advanced uses of blockchain in health care domain like Blockchain consortium, Smart contract-based health care intelligent claim processing and prior authorization and Wearable fitness device integration and monitoring health.

#### **Session IV: Blockchain and Cloud Computing (September 18 - September 20, 2020, 7/24 Available)**

Chair: Prof. Zhixiong Chen, PhD (Mercy College, USA)

20 minutes for each paper.

#### **Lekana - Blockchain Based Archive Storage for Large-scale Cloud Systems (ICBC2020-9010)**

*Eranga Bandara (Old Dominion University, USA); Xueping Liang (University of North Carolina Greensboro, USA); Sachin Shetty (Old Dominion University, USA); Wee-Keong Ng (Nanyang Technological University, Singapore); Peter Foytik (Old Dominion University, USA); Nalin Ranasinghe and Kasun De Zoysa (University of Colombo School of Computing, Sri Lanka); Bård Langøy (CTO, Sweden); David Larsson (Architect, USA)*

**Abstract:** Blockchain is a form of a distributed storage system that stores a chronological sequence of transactions in a tamper-evident manner. Due to the decentralized trust ecosystem in blockchain, various industries have adopted blockchain to build their applications. This paper presents a novel approach to building a blockchain-based document archive storage platform, "Lekana". The Lekana platform can be adopted by service providers that require frequent document operations, such as "Pageronline", a cloud-based e-invoicing provider in Europe. With Lekana we introduce a novel approach to store an immutable hash chain of archived data which are owned by the customers in the blockchain. The proposed Lekana platform is built on top of Mystiko which is a highly scalable blockchain storage platform targeted for big data. We have integrated real-time data analytics and machine learning techniques into the Lekana platform by using the Mystiko-MI machine learning service on Mystiko blockchain. By integrating the Lekana platform with blockchain technology, we have addressed major issues in most cloud-based, centralized storage platforms (e.g. lack of data privacy, lack of data immutability, lack of traceability and lack of data provenance). As a case study in the paper, we present how "Pageronline"



cloud-based e-invoicing provider stores their archived document data and archived document hash chain in the blockchain-based Lekana platform.

**Reducing Storage Requirement in Blockchain Networks Using Overlapping Data Distribution (ICBC2020-9011)**

*Md. Touhidul Islam (Bangladesh University of Engineering and Technology, Bangladesh); Muhammad Abdullah Adnan (University of California San Diego, USA)*

**Abstract:** Blockchain technology first gained attention via public blockchain platforms like Bitcoin and Ethereum. Over the years, researchers have continuously explored the potential of blockchains in a more restricted environment, which in turn has paved the way for the creation of many private blockchain platforms. In a private blockchain system, the identity of every entity is known and thus, the issue of trust is less prominent. In addition to the added customizability of permissions regarding who gets to do what, private blockchains also reduce the resource requirement as no proof-of-work is needed here. However, each node in a private blockchain network still needs to store the whole blockchain consisting of all the transactions from the beginning. If the number of transactions in a private blockchain network rises to a very large number, the storage requirement can rise proportionately. In addition to that, public blockchain platforms like Bitcoin and Ethereum are already in need of storage optimizations because of their size. We propose a method to divide the whole blockchain of transactions into some non-overlapping shards and make multiple copies of them. Then, we distribute these shards uniformly across the nodes in the network. We show theoretically that this approach not only improves the storage requirement but also ensures the integrity of the data in blockchain in case of node failures.

**BBM: A Blockchain-based model for open banking via self-sovereign identity (ICBC2020-9012)**

*Chengzu Dong and Ziyuan Wang (Swinburne University of Technology, Australia); Shiping Chen (CSIRO Data61 & UNSW, Australia); Yang Xiang (Swinburne University of Technology, Australia)*

**Abstract:** Open banking technology is an emerging data-sharing paradigm that can facilitate and inspire new businesses via efficient data-sharing with banks, such as quick approval of loan applications and finding a better investment return. However, the majority of customers are reluctant to use the open banking service, because they feel unsafe to share data with third-party service providers. In this paper, we proposed (BBM) a blockchain-based self-sovereign identity system model for open banking. The model provides a reliable secured communication network between users and third-party service providers and enables users to control their own identities and data. Through a comparison with the existing related work, the superiority of the BBM model has been demonstrated and analysed.

**Session V: Blockchain Platforms for Applications (September 18 - September 20, 2020, 7/24 Available)**

Chair: Prof. Laizhong Cui (Shenzhen University, China)

20 minutes for each paper.

**ProvNet: Networked Blockchain for Decentralized Secure Provenance (ICBC2020-9013)**

*Changhao Chenli and Taeho Jung (University of Notre Dame, USA)*

**Abstract:** Data sharing is increasingly popular especially for scientific research and business fields where large volume of data are usually required. People benefit from data sharing, but keeping track of the history of the shared data (i.e., provenance records) for monitoring and protecting them is not easy. On one hand, due to the decentralized nature of data sharing nowadays, it is impractical to have a centralized entity who collects all the provenance records. Some previous works that combined the data sharing with blockchain have been proposed. Such systems can store the provenance records of data sharing in a blockchain system. On the other hand, previous works also focus on the scenario where malicious users will not modify the shared datasets or re-share them as the owner of these changed datasets. Besides, maintaining the correctness of collected records in the presence of malicious attackers is challenging. In this paper, we present ProvNet, a decentralized data sharing platform which can provide a secure and correct provenance record using a networked blockchain. All the valid sharing records will be collected and stored in a "blocknet", which is a tamper-proof networked blockchain. Furthermore, ProvNet can also discover and detect the misbehaviors by leveraging the stored provenance records.

**Blockchain Based Full Privacy Preserving Public Procurement (ICBC2020-9014)**

*Prem Ratan Baranwal (Talentica Software (India) Pvt Ltd, India)*

**Abstract:** Public procurement is one of the government activities most prone to corruption, and e-procurement systems have been recommended to increase transparency, outreach, and competition. Other benefits include ease of access to public tenders and easier detection of irregularities. One of the main challenges with the existing e-procurement/auction systems is ensuring bid privacy of the losing bidders and collusion between bidders and auctioneer. Most of the auction systems, proposed, depend either upon auctioneer(s) or on the trusted third party, which, according to us, is the biggest problem for addressing corruption. We propose a blockchain based solution for Public Procurement, which eliminates auctioneers/third-parties using secure multi-party computation (MPC). Our solution fully preserves bid privacy and is secure against malicious bidders.



**Have I Been Exploited? A Registry of Vulnerable Ethereum Smart Contracts** (ICBC2020-9015)

*Daniel S Connelly and Wu-chang Feng (Portland State University, USA)*

Abstract: Ethereum Smart Contracts, also known as Decentralized Applications (DApps), are small programs which orchestrate financial transactions. Though beneficial in many cases, such contracts can and have been exploited, leading to a history of financial losses in the millions of dollars for those who have invested in them. It is critical that users be able to trust the contract code they place their money into. One way for verifying a program's integrity is Symbolic Execution. Unfortunately, while the information derived from symbolic execution is beneficial, performing it is often financially and technically infeasible for users to do. To address this problem, this paper describes the design and implementation of a registry of vulnerable Ethereum contracts. The registry compiles the results of exhaustive application of symbolic analysis to deployed contracts and makes it available to users seeking to understand the risks associated with contracts they are intending to utilize.

## Journals on Services Innovations

The *Services Transactions* series journals only consider extended papers of reputable conferences and invited articles from leading reserachers and practitioners in the field.



### **Services Transactions on Services Computing (STSC, ISSN 2330-4472)**

<http://www.hipore.com/stsc>

The *Services Transactions on Services Computing (STSC)* covers state-of-the-art technologies and best practices of Services Computing, as well as emerging standards and research topics which would define the future of Services Computing.

### **Services Transactions on Cloud Computing (STCC, ISSN 2326-7550)**

<http://www.hipore.com/stcc>

The *Services Transactions on Cloud Computing (STCC)* is a refereed, international publication featuring the latest research findings and industry solutions involving all aspects of cloud computing. STCC's major topics include but not limited to architecture, modeling, development, deployment, analytics, optimization, outsourcing, delivery model, consulting, and solutions of Cloud Computing.

### **Services Transactions on Big Data (STBD, ISSN 2326-442X)**

<http://www.hipore.com/stbd>

The *Services Transactions on Big Data (STBD)* aims to provide the first Public Access publication channel for all authors working in the field of all aspects of Big Data. STBD's major topics include but not limited to architecture, modeling, development, deployment, analytics, optimization, outsourcing, delivery model, consulting, and solutions of big data.

### **Services Transactions on Internet of Things (STIOT, ISSN 2577-1310)**

<http://www.hipore.com/stiot>

The *Services Transactions on Internet of Things (STIOT)* aims to be a core resource providing leading technologies, development applications, empirical studies, and future trends in the field of Internet of Things (IOT).

### **Services Transactions on Blockchain (STBC, ISSN 2577-7238 Print, 2577-7246 Online)**

<http://www.hipore.com/stbc>

The *Services Transactions on Blockchain (STBC)* aims to provide the first Public Access publication channel for all authors working in the field of all aspects of Blockchain. STBC's major topics include but not limited to enabling technologies, applications, and standards of blockchain.

### **International Journal of Web Services Research (JWSR, EI & SCI indexed)**

<http://www.igi-global.com/journal/international-journal-web-services-research/1079>

### **International Journal of Business Process Integration and Management (IJBPIIM)**

<http://www.inderscience.com/jhome.php?jcode=ijbpim>

## Services Society



The Services Society (S2) is a non-profit professional organization that has been created to promote worldwide research and technical collaboration in services innovations among academia and industrial professionals. Its members are volunteers from industry and academia with common interests. S2 is registered in the USA as a "501(c) organization", which means that it is an American tax-exempt nonprofit organization. S2 collaborates with other professional organizations to sponsor or co-sponsor conferences and to promote an effective services curriculum in colleges and universities. The S2 initiates and promotes a "Services University" program worldwide to bridge the gap between industrial needs and university instruction. The Services Society has formed 10 Special Interest Groups (SIGs) to support technology and domain specific professional activities.

The Services Society has been sponsoring various international conferences such as the International Conference on Web Services (ICWS), International Conference on Services Computing (SCC), World Congress on Services (SERVICES), International Conference on Cloud Computing (CLOUD), International Conference on AI and Mobile Services (AIMS), International Congress on Big Data (BigData Congress), International Conference on Internet of Things (ICIOT), International Conference on Cognitive Computing (ICCC), International Conference on Edge Computing (EDGE) and International Conference on Blockchain (ICBC).

The Services Society is a technical sponsor of a series of international journals such as Services Transactions on Big Data (STBD, <http://www.hipore.com/stbd>), Services Transactions on Cloud Computing (STCC, <http://www.hipore.com/stcc>), Services Transactions on Services Computing (STSC, <http://www.hipore.com/stsc>), and Services Transactions on Blockchain (STBC, <http://www.hipore.com/stbc>).

S2 also offers volunteer services to support conferences and builds professional communities like special interest groups, local chapters, Services Society Young Scientist Forum (SSYSF). The SSYSFs have been proactive in helping organizing Services Conference Federation (SCF) satellite sessions and other professional activities.

In 2013, the *YouthLancer Foundation* was founded to proof read papers for professors or researchers whose native language is not English. As an organization composed entirely of volunteers, there is no charge for correcting papers.

The Services Society opens to everyone, 7/24!



## Introduction of Big Data Certification

<http://www.servicessociety.org/>

Big Data includes but not limited to Data as A Service (DaaS), massive data, related technologies, and applications. Nowadays, professionals from academia and industry are exploring potential values of big data in various scenarios. New computing platforms such as Cloud Computing, Mobile Internet, and Social Networking are leveraged to create innovations around big data.

Services Society (S2) has been leading the platform of exchanging information and knowledge on the latest service-centric technologies and solutions. In order to address the shortage of Big Data professionals, Services Society led the creation of Body of Knowledge on Big Data (BoK-BD) in 2014.

Based on the first version of BoK-BD, Services Society launched International Certification Program for Big Data professionals (ICP4BD), which includes the following two certification categories.

### >>Category 1: Big Data for Job Positions

#### 1.1 Big Data Business Analyst (BDBA)

Key Skills: data warehouse, data mining, machine learning, data analysis algorithm, requirement-driven data modeling, data-driven business decision making

#### 1.2 Big Data Engineer (BDE)

Key Skills: distributed system (e.g. Hadoop, Hive, Storm, Spark), scalable database programming, Web programming, mobile programming

#### 1.3 Big Data Architect (BDA)

Key Skills: data warehouse design, cloud storage, computing architecture (e.g. SOA, Cloud), ETL architecture, distributed system architecture (e.g. Hadoop, Hive, Storm, Spark), system redesign, and independent product design.

#### Key Requirements for Category 1:

- 1). 5 years working experience in related areas;
- 2). Completed related scores (e.g. scores from selected MOCC courses, scores derived from published research journals);

### >>Category 2: Big Data for Professional Grade

#### 2.1 Certified Big Data Professional (CBDP)

With this certificate, you can serve as a chief architect, or data scientist in an organization.

#### Key Requirements for Category 2:

- 1). 5 years working experience;
- 2). Completed at least two certificates defined in Category 1;
- 3). Completed related scores (e.g. scores from selected MOCC courses, scores derived from published research journals);
- 4). 3 references from CBDP holders or at least 2 references from the program committee of ICWS, SCC, CLOUD, BigData Congress, SERVICES, S2 ICIOT, or editorial board members from the Services Transactions on Big Data, Services Transactions on Cloud Computing, Services Transactions on Services Computing, Services Transactions on Internet of Things, and Services Transactions on Blockchain.

## **Services Society Special Interests Groups (SIGs)**

The Services Sector has account for 79.5% of the GDP of United States in 2016. The world's most services-oriented economy, with services sectors accounting for more than 90% of GDP. To rapidly respond to the changing economy, the Technical Activities Board at the Services Society (<http://ServicesSociety.org/>) has created the following 10 Special Interest Groups (SIGs) for our worldwide Services Innovations Community members.

Special Interest Group on Web Services ([SIG-WS](#))

Special Interest Group on Services Computing ([SIG-SC](#))

Special Interest Group on Services Industry ([SIG-SI](#))

Special Interest Group on Big Data ([SIG-BD](#))

Special Interest Group on Cloud Computing ([SIG-CLOUD](#))

Special Interest Group on Artificial Intelligence ([SIG-AI](#))

Special Interest Group on Edge Computing ([SIG-EC](#))

Special Interest Group on Cognitive Computing ([SIG-CC](#))

Special Interest Group on BlockChain ([SIG-BC](#))

Special Interest Group on Internet of Things ([SIG-IOT](#))

## Services Conference Federation

As the founding member of the Services Conference Federation (SCF), the first International Conference on Web Services (ICWS) was held in June 2003 in Las Vegas, USA. Meanwhile, the First International Conference on Web Services - Europe 2003 (ICWS-Europe'03) was held in Germany in Oct, 2003. ICWS-Europe'03 is an extended event of the 2003 International Conference on Web Services (ICWS 2003) in Europe. In 2004, ICWS-Europe was changed to the European Conference on Web Services (ECOWS), which was held at Erfurt, Germany. SCF 2019 was held successfully in San Diego, USA. To celebrate its 18-year-old birthday, SCF 2020 will be held on June 25 - June 30, 2020, Hawaii, USA.

The past 17 years, ICWS community has been expanded from Web engineering innovations to scientific research for the whole services industry. The service delivery platforms have been expanded to mobile platforms, Internet of Things, cloud computing, and edge computing. The services ecosystem is gradually enabled, value added, and intelligence embedded through enabling technologies such as big data, artificial intelligence, and cognitive computing. In the coming years, all the transactions with multiple parties involved will be transformed to blockchain.



Based on the technology trends and best practices in the field, the Services Conference Federation (SCF) will continue serving as the conference umbrella's code name for all services-related conferences. SCF 2020 defines the future of New ABCDE (AI, Blockchain, Cloud, BigData & IOT). We are very proud to announce that SCF 2020's 10 co-located theme topic conferences will all center around "services", while each focusing on exploring different themes (web-based services, cloud-based services, Big Data-based services, services innovation lifecycle, AI-driven ubiquitous services, blockchain driven trust service-ecosystems, industry-specific services and applications, and emerging service-oriented technologies).

### Some highlights of SCF 2020 are shown below:

**- Bigger Platform:** The 10 collocated conferences (SCF 2020) get sponsorship from the Services Society which is the world-leading not-for-profits organization (501 c(3)) dedicated for serving more than 30,000 worldwide Services Computing researchers and practitioners. Bigger platform means bigger opportunities to all volunteers, authors and participants. Meanwhile, Springer provides sponsorship to best paper awards and other professional activities. All the 10 conference proceedings of SCF 2020 will be published by Springer and indexed in ISI Conference Proceedings Citation Index (included in Web of Science), Engineering Index EI (Compendex and Inspec databases), DBLP, Google Scholar, IO-Port, MathSciNet, Scopus, and ZBIMath.

**- Brighter Future:** While celebrating 2020 version of ICWS, SCF 2020 highlights the Second International Conference on Blockchain (ICBC 2020) to build the fundamental infrastructure for enabling secure and trusted services ecosystems. It will also lead our community members to create their own brighter future.

**- Better Model:** SCF 2020 will continue to leverage the invented Conference Blockchain Model (CBM) to innovate the organizing practices for all the 10 theme conferences. Senior researchers in the field are welcome to submit proposals to serve as CBM Ambassador for an individual conference to start better interactions during your leadership practices for organizing SCF 2020.



===== **Member of SCF 2020** =====

Services Conference Federation (SCF) includes 10 service-oriented conferences: ICWS, CLOUD, SCC, BigData Congress, AIMS, SERVICES, ICIOT, EDGE, ICCC and ICBC.

[1] 2020 International Conference on Web Services (ICWS 2020, <http://icws.org/2020>) is the flagship theme-topic conference for Web-centric services, enabling technologies and applications.

[2] 2020 International Conference on Cloud Computing (CLOUD 2020, <http://thecloudcomputing.org/2020>) is the flagship theme-topic conference for resource sharing, utility-like usage models, IaaS, PaaS, and SaaS.

[3] 2020 International Conference on Big Data (BigData Congress 2020, <http://bigdatacongress.org/2020>) is the theme-topic conference for data sourcing, data processing, data analysis, data-driven decision making, and data-centric applications.

[4] 2020 International Conference on Services Computing (SCC 2020, <http://thescc.org/2020>) is the flagship theme-topic conference for leveraging the latest computing technologies to design, develop, deploy, operate, manage, modernize, and redesign business services.

[5] 2020 International Conference on AI & Mobile Services (AIMS 2020, <http://ai1000.org/2020>) is the theme-topic conference for artificial intelligence, neural networks, machine learning, training data sets, AI scenarios, AI delivery channels, and AI supporting infrastructure as well as mobile internet services. The goal of AIMS is to bring AI to any mobile devices and other channels.

[6] 2020 World Congress on Services (SERVICES 2020, <http://servicescongress.org/2020>) puts its focus on all innovations of services industry that includes financial services, education services, transportation services, energy services, government services, manufacturing services, consulting services, and other industry services.

[7] 2020 International Conference on Cognitive Computing (ICCC 2020, <http://thecognitivecomputing.org/2020>) puts its focus on leveraging the latest computing technologies to simulate, model, implement, and realize of cognitive sensing and brain operating systems.

[8] 2020 International Conference on Internet of Things (ICIOT 2020, <http://iciot.org/2020>) puts its focus on the science, technology, and applications of IOT device innovations as well as IOT services in various solution scenarios.

[9] 2020 International Conference on Edge Computing (EDGE 2020, <http://theedgecomputing.org/2020/>) is a theme-topic conference for leveraging the latest computing technologies to enable localized device connections, edge gateways, edge applications, edge-cloud interactions, edge-user experiences, and edge business models.

[10] 2020 International Conference on Blockchain (ICBC 2020, <http://blockchain1000.org/2020/>) concentrates on all aspects of blockchain that includes digital currency, distributed application development, industry-specific blockchains, public blockchains, community blockchains, private blockchains, blockchain-based services, and enabling technologies.