2018 IEEE International Conference on Information Reuse and Integration (IRI 2018)

Salt Lake City, Utah, USA
6-9 July 2018
# 2018 IEEE International Conference on Information Reuse and Integration for Data Science

**IRI 2018**

## Table of Contents

Message from the IRI 2018 Program Co-Chairs  xiv

IRI 2018 Foreword  xvi

IRI 2018 Organizing Committee  xvii

IRI 2018 Program Committee  xviii

### Session A1.1: Security and Privacy

Adversarial Anomaly Detection Using Centroid-Based Clustering  1

*Imrul Chowdhury Anindya (University of Texas at Dallas) and Murat Kantarcıoglu (University of Texas at Dallas)*

A Survey of Medicare Data Processing and Integration for Fraud Detection 9

*Richard Bauder (Florida Atlantic University) and Taghi Khoshgoftaar (Florida Atlantic University)*

Decentralized IoT Data Management Using BlockChain and Trusted Execution Environment 15

*Gbadebo Ayoade (University of Texas at Dallas), Vishal Karande (University of Texas at Dallas), Latifur Khan (University of Texas at Dallas), and Kevin Hamlen (University of Texas at Dallas)*

### Session A1.2: Modeling and Mining

Large-Scale Realistic Network Data Generation on a Budget 23

*Brian Ricks (University of Texas at Dallas), Patrick Tague (Carnegie Mellon University), and Bhavani Thuraisingham (University of Texas at Dallas)*

Towards an Architecture Model for Dynamic Software Product Lines Engineering 31

*Edilton Santos (Federal University of Bahia) and Ivan Machado (Federal University of Bahia)*

Measuring the Growth in Complexity of Models from Industrial Control Networks 39

*Mustafa Amir Faisal (University of Texas at Dallas) and Alvaro Cardenas (University of Texas at Dallas)*
Session A2: Social Media and Users

Movie Recommendations Using the Deep Learning Approach 47
Jeffrey Lund (Brigham Young University) and Yiu-Kai Ng (Brigham Young University)

Recommending Scientific Collaboration Based on Topical, Authors and Venues Similarities 55
Abdulrhman M. Alshareef (University of Ottawa and King Abdulaziz University), Mohammed F. Alhamid (King Saud University), and Abdulmotaleb El Saddik (University of Ottawa)

On the Use of URLs and Hashtags in Age Prediction of Twitter Users 62
Abhinay Pandya (University of Oulu), Mourad Oussalah (University of Oulu), Paola Monachesi (Utrecht University), Panos Kostakos (University of Oulu), and Lauri Lovén (University of Oulu)

Session A3: Classification and Processing

The Effects of Random Undersampling with Simulated Class Imbalance for Big Data 70
Tawfiq Hasanin (Florida Atlantic University) and Taghi Khoshgoftaar (Florida Atlantic University)

Medicare Fraud Detection Using Random Forest with Class Imbalanced Big Data 80
Richard Bauder (Florida Atlantic University) and Taghi Khoshgoftaar (Florida Atlantic University)

New Techniques for Coding Political Events across Languages 88
Yan Liang (University of Oklahoma), Khaled Jabr (University of Oklahoma), Christian Grant (University of Oklahoma), Jill Irvine (University of Oklahoma), and Andrew Halterman (Massachusetts Institute of Technology)

Applications 1: Detection

Topic Diffusion Discovery Based on Sparseness-Constrained Non-Negative Matrix Factorization 94
Yihuang Kang (National Sun Yat-sen University), Keng-Pei Lin (National Sun Yat-sen University), and I-Ling Cheng (National Sun Yat-sen University)

Activity Detection and Analysis Using Smartphone Sensors 102
Amari Vaughn (North Carolina A&T State University), Paul Biocco (North Carolina A&T State University), Yang Liu (North Carolina A&T State University), and Mohd Anwar (North Carolina A&T State University)

Utilizing Netflow Data to Detect Slow Read Attacks 108
Clifford Kemp (Florida Atlantic University), Chad Calvert (Florida Atlantic University), and Taghi Khoshgoftaar (Florida Atlantic University)
Poster Session 1

On Wordsense Disambiguation through Morphological Transformation and Semantic Distance and Domain Link Knowledge
Mourad Oussalah (University of Oulu), Mikko Väisänen (University of Oulu), and Ekaterina Gilman (University of Oulu)

Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study
Sahil Dhankhad (Lakehead University), Emad Mohammed (Lakehead University), and Behrouz Far (Calgary University)

How Perceptions of Web Resource Boundaries Differ for Institutional and Personal Archives
Faryaneh Poursardar (Texas A&M University) and Frank Shipman (Texas A&M University)

Oculomotor Plant Feature Extraction from Human Saccadic Eye Movements
Sobiga Shanmugathasan (California State Polytechnic University) and Sampath Jayarathna (California State Polytechnic University)

Prioritized Situation Awareness for Soccer Robots Using Timed Transition Petri Nets
Max Weissman (The College of New Jersey), Daniel Ponsini (The College of New Jersey), and Seung-yun Kim (The College of New Jersey)

Session B1: Knowledge Representation and Management

Handling Hyperknowledge Representations through an Interactive Visual Approach
Marcio Ferreira Moreno (IBM Research), Rodrigo Santos (IBM Research), Wallas Santos (IBM Research), Rafael Brandão (IBM Research), Patricia Carrion (IBM Research), and Renato Cerqueira (IBM Research)

Knowledge-Based Randomization for Amplification
Thouraya Bouabana-Tebibel (Ecole Nationale Supérieure d’Informatique), Stuart H. Rubin (Space and Naval Warfare Systems Center Pacific), Lydia Bouzar-Benlabiod (Ecole Nationale Supérieure d’Informatique), Miled Basma Bentaiiba-Lagrid (Ecole Nationale Supérieure d’Informatique), and Maria Roumaissa Hanini (Ecole Nationale Supérieure d’Informatique)

Knowledge Amplification Using Randomization in Case-Based Reasoning -- Case Study: Severity of Mammography Mass
Miled Basma Bentaiiba-Lagrid (Ecole Nationale Supérieure d’Informatique), Lydia Bouzar-Benlabiod (Ecole Nationale Supérieure d’Informatique), Stuart H. Rubin (Space and Naval Warfare Systems Center Pacific), Thouraya Bouabana-Tebibel (Ecole Nationale Supérieure d’Informatique), and Maria Roumaissa Hanini (Ecole Nationale Supérieure d’Informatique)

Randomization-Based Knowledge Discovery with Application to Weather Prediction
Lydia Bouzar-Benlabiod (Ecole Nationale Supérieure d’Informatique), Stuart H. Rubin (Space and Naval Warfare Systems Center Pacific), and Meziani Lila (Ecole Nationale Supérieure d’Informatique)
Session B2: Learning

A Complex Task Scheduling Scheme for Big Data Platforms Based on Boolean Satisfiability

Problem 170

Huang Hong (Southwest University of Science and Technology), Latifur Khan (University of Texas at Dallas), Ayoade Gbadebo (University of Texas at Dallas), Zhou Shaohua (Southwest University of Science and Technology), and Wei Yong (Southwest University of Science and Technology)

Deep Transfer Learning for Traffic Sign Recognition 178

Grant Rosario (Florida Atlantic University), Thomas Sonderman (Florida Atlantic University), and Xingquan Zhu (Florida Atlantic University)

Partition-Aware Scalable Outlier Detection Using Unsupervised Learning 186

Pallabi Parveen (AT&T), Melissa Lee (AT&T), Austin Henslee (AT&T), Matt Dugan (AT&T), and Brad Ford (AT&T)

Applications 2: Social Media and Users

Web-Based Recommendation System Architecture for Knowledge Reuse in MOOCs Ecosystems 193

Rodrigo Campos (Universidade Federal do Rio de Janeiro), Rodrigo Pereira dos Santos (Universidade Federal do Estado do Rio de Janeiro), and Jonice Oliveira (Universidade Federal do Rio de Janeiro)

Forecasting Weekly Crude Oil Using Twitter Sentiment of U.S. Foreign Policy and Oil Companies Data 201

Mourad Oussalah (University of Oulu) and Ahmed Zaidi (University of Cambridge)

The Impact of Live Tweeting on Social Movements 209

Brandeis Marshall (Spelman College), Takeria Blunt (Spelman College), and Tayloir Thompson (Spelman College)

Applications 3: Information Reuse

Automated Playlist Generation from Personal Music Libraries 217

Diana Lin (California State Polytechnic University Pomona) and Sampath Jayaratna (California State Polytechnic University Pomona)

Poster Session 2

Design-Computational Thinking, Transfer and Flavors of Reuse: Scaffolds to Information and Data Science for Sustainable Systems in Smart Cities 225

Chien-Sing Lee (Sunway University) and K. Daniel Wong (Daniel Wireless Software Pte. Ltd.)

Tackling Class Imbalance in Cyber Security Datasets 229

Charles Wheelus (Florida Atlantic University), Elias Bou-Harb (Florida Atlantic University), and Xingquan Zhu (Florida Atlantic University)
Session C1: Diagnosis and Prediction

Misclassification Cost-Sensitive Software Defect Prediction  256..............................................................
   Ling Xu (Chongqing University), Bei Wang (Chongqing University), Ling Liu (Georgia Institute of Technology), Mo Zhou (Chongqing University), Shengping Liao (Chongqing University), and Meng Yan (Zhejiang University)

Correlating Refugee Border Crossings with Internet Search Data  264..............................................................
   Panos Kostakos (University of Oulu), Abhinay Pandya (University of Oulu), Mourad Oussalah (University of Oulu), Simo Hosio (University of Oulu), Christoph Breidbach (University of Melbourne), Vassilis Kostakos (University of Melbourne), Niels van Berkel (University of Melbourne), Olga Kyriakouli (Harokopio University of Athens), and Arash Sattari (University of Oulu)

Applications 4: AI and Machine Learning 1

Classification for Authorship of Tweets by Comparing Logistic Regression and Naïve Bayes Classifiers  269................................................................................................................
   Opeyemi Aborisade (North Carolina A&T State University) and Mohd Anwar (North Carolina A&T State University)

A Graduate School Recommendation System Using the Multi-Class Support Vector Machine and KNN Approaches  277..............................................................
   Alisha Baskota (Brigham Young University) and Yiu-Kai Ng (Brigham Young University)
Identifying Medicare Provider Fraud with Unsupervised Machine Learning 285
Richard Bauder (Florida Atlantic University), Raquel da Rosa (Florida Atlantic University), and Taghi Khoshgoftaar (Florida Atlantic University)

Applications 4: AI and Machine Learning 2

Context-Sensitive Neural Sentiment Classification 293
Shekoofeh Mokhtari (Florida International University), Tao Li (Florida International University), and Ning Xie (Florida International University)

Speaker Verification Using Adapted Bounded Gaussian Mixture Model 300
Muhammad Azam (Concordia University) and Nizar Bouguila (Concordia University)

Toward a Dynamic K in K-Nearest Neighbor Fingerprint Indoor Positioning 308
Jiusong Hu (Hunan University), Hongli Liu (Hunan University), and Dawei Liu (Xi’an Jiaotong-Liverpool University)

Soccer Ball Recognition and Distance Prediction Using Fuzzy Petri Nets 315
Patrick Hansen (The College of New Jersey), Philip Franco (The College of New Jersey), and Seung-yun Kim (The College of New Jersey)

Applications 5: Data Analytics 1

RDF Data Storage Techniques for Efficient SPARQL Query Processing Using Distributed Computation Engines 323
Mahmudul Hassan (Arizona State University) and Srividya K. Bansal (Arizona State University)

Efficient Processing of Range Queries over Distributed Relational Databases 331
Richard Price (Georgia Gwinnett College), Lakshmish Ramaswamy (University of Georgia), and Seyedamin Pouriyeh (University of Georgia)

Compositional Data Analysis with PLS-DA and Security Applications 338
Divya Ankam (Concordia University) and Nizar Bouguila (Concordia University)

Session C2: Decision Support

Is Gene Selection Enough for Imbalanced Bioinformatics Data? 346
Ahmad Abu Shanab (Florida Atlantic University) and Taghi Khoshgoftaar (Florida Atlantic University)

An Approach for Schema Extraction of JSON and Extended JSON Document Collections 356
Angelo Augusto Frozza (Federal Institute Catarinense), Ronaldo dos Santos Mello (Federal University of Santa Catarina), and Felipe de Souza da Costa (Federal University of Santa Catarina)
Applications 5: Data Analytics 2

Semantic Data Querying over NoSQL Databases with Apache Spark .364 .................................................................
Mahmudul Hassan (Arizona State University) and Srividya K. Bansal (Arizona State University)

Filter-Based Subset Selection for Easy, Moderate, and Hard Bioinformatics Data .372 ........................................
Ahmad Abu Shanab (Florida Atlantic University) and Taghi Khoshgoftaar (Florida Atlantic University)

Invited Papers Session I

Extracting Entity-Specific Substructures for RDF Graph Embedding .378 .................................................................
Muhammad Rizwan Saeed (University of Southern California) and Viktor K. Prasanna (University of Southern California)

Deep Spatio-Temporal Representation Learning for Multi-Class Imbalanced Data Classification .386 .................................................................
Samira Pouyanfar (Florida International University), Shu-Ching Chen (Florida International University), and Mei-Ling Shyu (University of Miami)

TwoRavens for Event Data .394 .................................................................
Vito D’Orazio (University of Texas at Dallas), Marcus Deng (University of Texas at Dallas), and Michael Shoemate (University of Texas at Dallas)

Invited Papers Session II

Ontology-Based Supervised Concept Learning for the Biogeochemical Literature .402 .................................................................
Deya Banisakher (Florida International University), Maria E. Presa Reyes (Florida International University), Joshua D. Eisenberg (Florida International University), Joshua Allen (Florida International University), Mark A. Finlayson (Florida International University), Rene Price (Florida International University), and Shu-Ching Chen (Florida International University)

Trustworthy Computing in the Dynamic IoT Cloud .411 .................................................................
I-Ling Yen (University of Texas at Dallas), Farokh Bastani (University of Texas at Dallas), Nidhiben Solanki (University of Texas at Dallas), and Yongtao Huang (University of Texas at Dallas)

A Multi-Model Based Approach for Driver Missense Identification .419 .................................................................
Ahmed Soliman (University of Miami) and Mei-Ling Shyu (University of Miami)

Invited Papers Session III

Accelerating Analytics for Clinical Pathways to Drive Cost Reduction and Quality Improvement .426 .................................................................
James McGlothlin (Fusion Consulting Inc.)
Reduced Residual Nets (Red-Nets): Low Powered Adversarial Outlier Detectors  
Saad Sadiq (University of Miami), Marina Zmieva (University of Miami), Mei-Ling Shyu (University of Miami), and Shu-Ching Chen (Florida International University)

Automatic Convolutional Neural Network Selection for Image Classification Using Genetic Algorithms  
Haiman Tian (Florida International University), Samira Pouyanfar (Florida International University), Jonathan Chen (Miami Palmetto Senior High School), Shu-Ching Chen (Florida International University), and Sitharama S. Iyengar (Florida International University)

7th IEEE International Workshop on Data Integration and Mining (DIM 2018)

Universal Metadata Repository: Integrating Data Profiles Across an Organization  
Neda Abolhassani (University of Georgia), Colin Puri (Accenture Technology Labs), Sonali Parthasarathy (Accenture Technology Labs), Zhijie Wang (Accenture Technology Labs), Matthew Kujawinski (Accenture Technology Labs), Teresa Tung (Accenture Technology Labs), and Lakshmish Ramaswamy (University of Georgia)

Identifying Tweets with Fake News  
Saranya Krishnan (University of Washington Bothell) and Min Chen (University of Washington Bothell)

Diabetic Retinopathy Stage Classification Using Convolutional Neural Networks  
Xiaoliang Wang (Virginia State University), Yongjin Lu (Virginia State University), Yujuan Wang (Sun Yat-sen University), and Wei-Bang Chen (Virginia State University)


Preliminary Modeling of Gender Identification in the Social Community  
Kuo-Chung Chu (National Taipei University of Nursing and Health Sciences), Peng-Hua Jiang (National Taipei University of Nursing and Health Sciences), Huan-Yu Syu (National Taipei University of Nursing and Health Sciences), and Yan-Ling Chen (National Taipei University of Nursing and Health Sciences)

Artificial Intelligence for Automatic Text Summarization  
Min-Yuh Day (Tamkang University) and Chao-Yu Chen (Tamkang University)

Automatic Dialogue Template Synthesis for Chatbot by Story Information Extraction  
Shih-Hung Wu (Chaoyang University of Technology), Liang-Pu Chen (Institute for Information Industry), Ping-Che Yang (Institute for Information Industry), and Tsun Ku (Institute for Information Industry)