

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

**Salt Lake City, Utah, USA
6-9 July 2018**



**IEEE Catalog Number: CFP18IRI-POD
ISBN: 978-1-5386-2660-3**

**Copyright © 2018 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

| | |
|-------------------------|-------------------|
| IEEE Catalog Number: | CFP18IRI-POD |
| ISBN (Print-On-Demand): | 978-1-5386-2660-3 |
| ISBN (Online): | 978-1-5386-2659-7 |

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

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..... | |
| <i>Imrul Chowdhury Anindya (University of Texas at Dallas) and Murat Kantarcioglu (University of Texas at Dallas)</i> | |
| A Survey of Medicare Data Processing and Integration for Fraud Detection .9..... | |
| <i>Richard Bauder (Florida Atlantic University) and Taghi Khoshgoftaar (Florida Atlantic University)</i> | |
| Decentralized IoT Data Management Using BlockChain and Trusted Execution Environment .15..... | |
| <i>Gbadebo Ayode (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)</i> | |

Session A1.2: Modeling and Mining

| | |
|---|--|
| Large-Scale Realistic Network Data Generation on a Budget .23..... | |
| <i>Brian Ricks (University of Texas at Dallas), Patrick Tague (Carnegie Mellon University), and Bhavani Thuraisingham (University of Texas at Dallas)</i> | |
| Towards an Architecture Model for Dynamic Software Product Lines Engineering .31..... | |
| <i>Edilton Santos (Federal University of Bahia) and Ivan Machado (Federal University of Bahia)</i> | |
| Measuring the Growth in Complexity of Models from Industrial Control Networks .39..... | |
| <i>Mustafa Amir Faisal (University of Texas at Dallas) and Alvaro Cardenas (University of Texas at Dallas)</i> | |

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), Christan 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 .117.....
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 .122.....
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 .126.....
Faryaneh Poursardar (Texas A&M University) and Frank Shipman (Texas A&M University)
- Oculomotor Plant Feature Extraction from Human Saccadic Eye Movements .130.....
Sobiga Shanmugathan (California State Polytechnic University) and Sampath Jayarathna (California State Polytechnic University)
- Prioritized Situation Awareness for Soccer Robots Using Timed Transition Petri Nets .134.....
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 .139.....
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 .147.....
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 Bentaiba-Lagrid (Ecole Nationale Supérieure d'Informatique), and Maria Roumaïssa Hanini (Ecole Nationale Supérieure d'Informatique)
- Knowledge Amplification Using Randomization in Case-Based Reasoning -- Case Study: Severity of Mammography Mass .155.....
Miled Basma Bentaiba-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 Roumaïssa Hanini (Ecole Nationale Supérieure d'Informatique)
- Randomization-Based Knowledge Discovery with Application to Weather Prediction .163.....
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..... | |
| <i>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)</i> | |
| Deep Transfer Learning for Traffic Sign Recognition .178..... | |
| <i>Grant Rosario (Florida Atlantic University), Thomas Sonderman (Florida Atlantic University), and Xingquan Zhu (Florida Atlantic University)</i> | |
| Partition-Aware Scalable Outlier Detection Using Unsupervised Learning .186..... | |
| <i>Pallabi Parveen (AT&T), Melissa Lee (AT&T), Austin Henslee (AT&T), Matt Dugan (AT&T), and Brad Ford (AT&T)</i> | |

Applications 2: Social Media and Users

| | |
|---|--|
| Web-Based Recommendation System Architecture for Knowledge Reuse in MOOCs Ecosystems .193 | |
| <i>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)</i> | |
| Forecasting Weekly Crude Oil Using Twitter Sentiment of U.S. Foreign Policy and Oil Companies Data .201..... | |
| <i>Mourad Oussalah (University of Oulu) and Ahmed Zaidi (University of Cambridge)</i> | |
| The Impact of Live Tweeting on Social Movements .209..... | |
| <i>Brandeis Marshall (Spelman College), Takeria Blunt (Spelman College), and Tayloir Thompson (Spelman College)</i> | |

Applications 3: Information Reuse

| | |
|---|--|
| Automated Playlist Generation from Personal Music Libraries .217..... | |
| <i>Diana Lin (California State Polytechnic University Pomona) and Sampath Jayarathna (California State Polytechnic University Pomona)</i> | |

Poster Session 2

| | |
|---|--|
| Design-Computational Thinking, Transfer and Flavors of Reuse: Scaffolds to Information and Data Science for Sustainable Systems in Smart Cities .225..... | |
| <i>Chien-Sing Lee (Sunway University) and K. Daniel Wong (Daniel Wireless Software Pte. Ltd.)</i> | |
| Tackling Class Imbalance in Cyber Security Datasets .229..... | |
| <i>Charles Wheelus (Florida Atlantic University), Elias Bou-Harb (Florida Atlantic University), and Xingquan Zhu (Florida Atlantic University)</i> | |

| | |
|---|---|
| Efficient All-and-One Support Vector Machines Based on One-versus-All Data Inseparability .233..... | |
| | <i>Patoomsiri Songsiri (Chulalongkorn University) and Boonserm Kijirikul (Chulalongkorn University)</i> |
| RevMap: A Visualized Framework for Holistic View of Reviews .240..... | |
| | <i>Shekoofeh Mokhtari (Florida International University), Tao Li (Florida International University), and Ning Xie (Florida International University)</i> |
| Visualization of Server Log Data for Detecting Abnormal Behaviour .244..... | |
| | <i>Reeta Suman (University of Calgary), Behrouz Far (University of Calgary), Emad A. Mohammed (Lakehead University), Ashok Nair (City of Calgary), and Sanaz Janbakhsh (City of Calgary)</i> |
| Modeling CDC Data for Real Time Feature Estimation .248..... | |
| | <i>Tannaz Rezaei Damavandi (California State Polytechnic University Pomona), Sampath Jayarathna (California State Polytechnic University Pomona), and Yu Sun (California State Polytechnic University Pomona)</i> |
| Deep Learning Based vs. Markov Chain Based Text Generation for Cross Domain Adaptation for Sentiment Classification .252..... | |
| | <i>Omar Abdelwahab (University of Louisville) and Adel Elmaghraby (University of Louisville)</i> |

Session C1: Diagnosis and Prediction

| | |
|--|--|
| Misclassification Cost-Sensitive Software Defect Prediction .256..... | |
| | <i>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)</i> |
| Correlating Refugee Border Crossings with Internet Search Data .264..... | |
| | <i>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)</i> |

Applications 4: AI and Machine Learning 1

| | |
|---|---|
| Classification for Authorship of Tweets by Comparing Logistic Regression and Naive Bayes Classifiers .269..... | |
| | <i>Opeyemi Aborisade (North Carolina A&T State University) and Mohd Anwar (North Carolina A&T State University)</i> |
| A Graduate School Recommendation System Using the Multi-Class Support Vector Machine and KNN Approaches .277..... | |
| | <i>Alisha Baskota (Brigham Young University) and Yiu-Kai Ng (Brigham Young University)</i> |

| | |
|---|--|
| Identifying Medicare Provider Fraud with Unsupervised Machine Learning .285..... | |
| <i>Richard Bauder (Florida Atlantic University), Raquel da Rosa (Florida Atlantic University), and Taghi Khoshgoftaar (Florida Atlantic University)</i> | |

Applications 4: AI and Machine Learning 2

| | |
|--|--|
| Context-Sensitive Neural Sentiment Classification .293..... | |
| <i>Shekoofeh Mokhtari (Florida International University), Tao Li (Florida International University), and Ning Xie (Florida International University)</i> | |
| Speaker Verification Using Adapted Bounded Gaussian Mixture Model .300..... | |
| <i>Muhammad Azam (Concordia University) and Nizar Bouguila (Concordia University)</i> | |
| Toward a Dynamic K in K-Nearest Neighbor Fingerprint Indoor Positioning .308..... | |
| <i>Jiusong Hu (Hunan University), Hongli Liu (Hunan University), and Dawei Liu (Xi'an Jiaotong-Liverpool University)</i> | |
| Soccer Ball Recognition and Distance Prediction Using Fuzzy Petri Nets .315..... | |
| <i>Patrick Hansen (The College of New Jersey), Philip Franco (The College of New Jersey), and Seung-yun Kim (The College of New Jersey)</i> | |

Applications 5: Data Analytics 1

| | |
|--|--|
| RDF Data Storage Techniques for Efficient SPARQL Query Processing Using Distributed Computation Engines .323..... | |
| <i>Mahmudul Hassan (Arizona State University) and Srividya K. Bansal (Arizona State University)</i> | |
| Efficient Processing of Range Queries over Distributed Relational Databases .331..... | |
| <i>Richard Price (Georgia Gwinnett College), Lakshmesh Ramaswamy (University of Georgia), and Seyedamin Pouriyeh (University of Georgia)</i> | |
| Compositional Data Analysis with PLS-DA and Security Applications .338..... | |
| <i>Divya Ankam (Concordia University) and Nizar Bouguila (Concordia University)</i> | |

Session C2: Decision Support

| | |
|--|--|
| Is Gene Selection Enough for Imbalanced Bioinformatics Data? .346..... | |
| <i>Ahmad Abu Shanab (Florida Atlantic University) and Taghi Khoshgoftaar (Florida Atlantic University)</i> | |
| An Approach for Schema Extraction of JSON and Extended JSON Document Collections .356..... | |
| <i>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)</i> | |

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. Eisengberg (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 .436.....
*Saad Sadiq (University of Miami), Marina Zmиеva (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 .444.....
*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 .452.....
*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
Lakshmesh Ramaswamy (University of Georgia)*

Identifying Tweets with Fake News .460.....
*Saranya Krishnan (University of Washington Bothell) and Min Chen
(University of Washington Bothell)*

Diabetic Retinopathy Stage Classification Using Convolutional Neural Networks .465.....
*Xiaoliang Wang (Virginia State University), Yongjin Lu (Virginia State
University), Yujuan Wang (Sun Yat-sen University), and Wei-Bang Chen
(Virginia State University)*

7th IEEE International Workshop on Empirical Methods for Recognizing Inference in Text (EM-RITE 2018)

Preliminary Modeling of Gender Identification in the Social Community .472.....
*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 .478.....
Min-Yuh Day (Tamkang University) and Chao-Yu Chen (Tamkang University)

Automatic Dialogue Template Synthesis for Chatbot by Story Information Extraction .485.....
*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)*

Automatic Question Generation from Children’s Stories for Companion Chatbot .491.....
Che-Hao Lee (National Central University), Tzu-Yu Chen (National Central University), Liang-Pu Chen (Institute for Information Industry), Ping-Che Yang (Institute for Information Industry), and Richard Tzong-Han Tsai (National Central University)

6th IEEE International Workshop on Formal Methods Integration (FMI 2018)

Tally Keeping-LTL: An LTL Semantics for Quantitative Evaluation of LTL Specifications .495.....
Raphaël Khoury (Université du Québec à Chicoutimi) and Sylvain Hallé (Université du Québec à Chicoutimi)

From Petri Nets to UML Model: A New Transformation Approach .503.....
Lila Meziani (Ecole Nationale Supérieure d’Informatique), Thouraya Bouabana-Tebibel (Ecole Nationale Supérieure d’Informatique), and Lydia Bouzar-Benlabiod (Ecole Nationale Supérieure d’Informatique)

A Graph Transformation Approach to Generate Analysable Maude Specifications from UML Interaction Overview Diagrams .511.....
Chafika Djaoui (University Mohamed Seddik Benyahia), Elhillali Kerkouche (University Mohamed Seddik Benyahia), Allaoua Chaoui (University Constantine 2), and Khaled Khalfaoui (University Mohamed Seddik Benyahia)

Context-Dependent Reachability Analysis for Hybrid Systems .518.....
Stefan Schupp (RWTH Aachen University), Justin Winkens (RWTH Aachen University), and Erika Ábrahám (RWTH Aachen University)

2nd International Workshop on Social Media Networks Analytics and Applications (SMA2 2018)

Trusted Environments for Volunteer Computing .526.....
Michele Ianni (DIMES-UNICAL) and Elio Masciari (ICAR-CNR)

Using Semantic Tools to Represent Data Extracted from Mobile Devices .530.....
Giovanni Cozzolino (Università degli Studi di Napoli "Federico II")

How to Exploit Recommender Systems in Social Media .537.....
Fabio Persia (Free University of Bozen-Bolzano), Mouzhi Ge (Masaryk University), and Daniela D’Auria (University of Naples)

Author Index 543