# TABLE OF CONTENTS

RESPONSIBLE DATA SCIENCE ................................................................. 1
Lise Getoor

DATACOMMONS ...................................................................................... 2
Ramanathan Gaha

DECEPTION, ROBUSTNESS AND TRUST IN BIG DATA FUELED DEEP LEARNING SYSTEMS .............................................. 3
Ling Liu

THE NEW SCIENCE OF CAUSE AND EFFECT, WITH REFLECTIONS ON DATA SCIENCE AND ARTIFICIAL INTELLIGENCE ......................... 4
Judea Pearl

FEDERATED RECOMMENDATION SYSTEMS .............................................. 5
Yang Qiang

UNCERTAINTY-AWARE OPINION INFERENCE UNDER ADVERSARIAL ATTACKS ................................................................. 6
Adil Alim ; Xiujiang Zhao ; Jin-Hee Cho ; Feng Chen

MIND THE STEP-ASYNCPSGD: ADAPTIVE ASYNCHRONOUS PARALLEL STOCHASTIC GRADIENT DESCENT ........................................ 16
Karl Bäckström ; Marina Papatriantafilou ; Philippas Tsigas

SKETCHYCORESVD: SKETCHY SVD FROM RANDOM SUBSAMPLING OF THE DATA MATRIX ........................................................... 26
Chandrajit Bajaj ; Yinyi Liang ; Thomas Wang

EXPLAINABLE AUTHORSHIP VERIFICATION IN SOCIAL MEDIA VIA ATTENTION-BASED SIMILARITY LEARNING ............................................. 36
Benedikt Boeninghoff ; Steffen Hustler ; Dorothea Kolossa ; Robert M. Nickel

FINDING STABLE CLUSTERING FOR NOISY DATA VIA STRUCTURE-AWARE REPRESENTATION .......................................................... 46
Huixuan Chen ; Jing Li

RESTRICTED RECURRENT NEURAL NETWORKS ........................................ 56
Enmiao Du ; Jie Ding ; Yahiad Tarokh

PARSIMONIOUS MORPHEME SEGMENTATION WITH AN APPLICATION TO ENRICHING WORD EMBEDDINGS .......................................... 64
Ahmed El-Kishky ; Frank Xu ; Aston Zhang ; Jiawei Han

A STREAMING MODEL FOR GENERALIZED RAYLEIGH WITH EXTENSION TO MINIMUM NOISE FRACTION ........................................... 74
Soumyajit Gupta ; Chandrajit Bajaj

DETECTING MODEL CHANGES AND THEIR EARLY WARNING SIGNALS USING MDL CHANGE STATISTICS .................................................. 84
So Hirai ; Kenji Yamanishi

ON THE GLOBAL CONVERGENCE OF CONTINUOUS–TIME STOCHASTIC HEAVY–BALL METHOD FOR NONCONVEX OPTIMIZATION ....................... 94
Wenqing Hu ; Chris Junchi Li ; Xiang Zhou

MINING NEWS EVENTS FROM COMPARABLE NEWS CORPORA: A MULTI-ATTRIBUTE PROXIMITY NETWORK MODELING APPROACH .................................. 105
Hyungsul Kim ; Ahmed El-Kishky ; Xiang Ren ; Jiawei Han

ON RELAXING FAILING QUERIES OVER RDF DATABASES ...................................... 115
Wafaa Mebrey ; Badran Raddaoui ; Mohammad Albilani

INFINITE DROPOUT FOR TRAINING BAYESIAN MODELS FROM DATA STREAMS ................................................................. 125
Van-Son Nguyen ; Duc-Tung Nguyen ; Linh Ngo Van ; Khoat Than

SUBSPACE CLUSTERING WITH ACTIVE LEARNING ..................................... 135
Hankai Peng ; Nicos G. Pavlidis

CTC-ATTENTION BASED NON-PARAMETRIC INFERENCE MODELING FOR CLINICAL STATE PROGRESSION ............................................ 145
Riazat Ryan ; Handong Zhao ; Ming Shao

XSVM: SCALABLE DISTRIBUTED KERNEL SUPPORT VECTOR MACHINE TRAINING .......................................................... 155
Ruchi Shah ; Shaoshuai Zhang ; Ying Lin ; Panruo Wu
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDMF: Hierarchical Data Modeling Framework for Modern Science Data</td>
<td>165</td>
</tr>
<tr>
<td>Standards</td>
<td></td>
</tr>
<tr>
<td>Andrew J. Tritt; Oliver Rubel; Benjamin Dichter; Ryan Ly; Donghe Kang; Edward F. Chang; Loren M. Frank; Kristofer Bouwland</td>
<td></td>
</tr>
<tr>
<td>Ordalia: Deep Learning Hyperparameter Search Via Generalization</td>
<td>180</td>
</tr>
<tr>
<td>Error Bound Extrapolation</td>
<td></td>
</tr>
<tr>
<td>Benedetto I. Buratti; Eli Upfal</td>
<td></td>
</tr>
<tr>
<td>Hierarchical Demand Forecasting for Factory Production of Perishable Goods</td>
<td>188</td>
</tr>
<tr>
<td>Subsampled Information Criteria for Bayesian Model Selection in the Big Data Setting</td>
<td>194</td>
</tr>
<tr>
<td>Lijiang Geng; Yishu Xue; Guangyu Hu</td>
<td></td>
</tr>
<tr>
<td>An Interactive Data Quality Test Approach for Constraint Discovery and Fault Detection</td>
<td>200</td>
</tr>
<tr>
<td>Hajar Homayouni; Sudipto Ghosh; Indrakshi Ray; Michael G Kahn</td>
<td></td>
</tr>
<tr>
<td>Activation Ensembles for Deep Neural Networks</td>
<td>206</td>
</tr>
<tr>
<td>Diego Klubjan; Mark Harmon</td>
<td></td>
</tr>
<tr>
<td>Online Federated Multitask Learning</td>
<td>215</td>
</tr>
<tr>
<td>Rui Li; Fenglong Ma; Wenjian Jiang; Jing Gao</td>
<td></td>
</tr>
<tr>
<td>Sparse Block Regression (SBR) for Big Data with Categorical Variables</td>
<td>221</td>
</tr>
<tr>
<td>Xiang Liu; Huyunling Huang; Ziyang Tang; Tonglin Zhang; Baijian Yang</td>
<td></td>
</tr>
<tr>
<td>Algorithms on Compressed Time-Evolving Graphs</td>
<td>227</td>
</tr>
<tr>
<td>Michael Nelson; Sridhar Radhakrishnan; Chandra N. Sekharan</td>
<td></td>
</tr>
<tr>
<td>Discovering Partial Periodic Spatial Patterns in Spatiotemporal Databases</td>
<td>233</td>
</tr>
<tr>
<td>R. Uday Kiran; C. Saideep; Koji Zettou; Masashi Toyoda; Masaru Kitsuregawa; P. Krishna Reddy</td>
<td></td>
</tr>
<tr>
<td>Co-representation Learning Framework for the Open-Set Data</td>
<td>239</td>
</tr>
<tr>
<td>Zhuowei Wang; Bo Dong; Yu Lin; Yigong Wang; Md Shihabul Islam; Latifur Khan</td>
<td></td>
</tr>
<tr>
<td>Mixture-Based Multiple Imputation Model for Clinical Data with a Temporal Dimension</td>
<td>245</td>
</tr>
<tr>
<td>Ye Xue; Diego Klubjan; Yuan Lao</td>
<td></td>
</tr>
<tr>
<td>A Discrete Mathematics Approach for Large Scale Improvement in Classification Training Time</td>
<td>253</td>
</tr>
<tr>
<td>Samdeh Yadav; Mathis Bode</td>
<td></td>
</tr>
<tr>
<td>Diagnosis of Neural Network Via Backward Deduction</td>
<td>260</td>
</tr>
<tr>
<td>Pejifeng Yin; Lei Huang; Sanhuwan Lee; Mu Qiao; Shahhi Ashtana; Taojiga Nakamura</td>
<td></td>
</tr>
<tr>
<td>Wiser: A Highly Available HTAP DBMS for IOT Applications</td>
<td>268</td>
</tr>
<tr>
<td>Ronald Barber; Christian Garcia-Arellano; Ronen Grossman; Guy Lohnan; C. Mohan; Rene Muller; Hamid Pirahesh; Vijayashankar Raman; Richard Silde; Adam Storm; Yuanyuan Tian; Pinar Tozun; Yingjun Wu</td>
<td></td>
</tr>
<tr>
<td>Progress-Based Container Scheduling for Short-Lived Applications in a Kubernetes Cluster</td>
<td>278</td>
</tr>
<tr>
<td>Yuqi Fu; Shaolun Zhang; Jose Terrero; Ying Mao; Guangya Liu; Sheng Li; Dingwen Tao</td>
<td></td>
</tr>
<tr>
<td>Finding Mutual X at Wechat-Scale Social Network in Ten Minutes</td>
<td>288</td>
</tr>
<tr>
<td>Conghui He; Shijie Sun; Benli Li; Xiaogang Tu; Donghai Yu</td>
<td></td>
</tr>
<tr>
<td>A Bicameralism Voting Framework for Combining Knowledge from Clients into Better Prediction</td>
<td>298</td>
</tr>
<tr>
<td>Yu-Tung Hsieh; Chuan-Yu Lee; Ching-Chi Lin; Pangfeng Liu; Jun-Jan Wu</td>
<td></td>
</tr>
<tr>
<td>Rapid Prototyping of Hierarchical Agglomerative Clustering</td>
<td>307</td>
</tr>
<tr>
<td>Algorithms for Distributed Systems</td>
<td></td>
</tr>
<tr>
<td>Saiyedul Islam; Navneet Goyal; Sundar Balasubramaniam; Poonam Goyal; Achat Agarwal; Kirtil Singh Rathore; Nischay Singh</td>
<td></td>
</tr>
<tr>
<td>General-Purpose vs. Specialized Data Analytics Systems: A Game of ML &amp; SQL Thrones</td>
<td>317</td>
</tr>
<tr>
<td>Erdokia Kaasela; Nikodimos Provatou; Ioannis Konstantinou; Avrilia Floratou; Nectarios Koziris</td>
<td></td>
</tr>
<tr>
<td>D3N: A Multi-Layer Cache for the Rest of Us</td>
<td>327</td>
</tr>
<tr>
<td>Eminen Ugur Kaynar; Mania Abdi; Mohammad Hossein Hajkazemi; Ana Tark; Raja R. Sambasivan; David Cohen; Larry Rudolph; Peter Desnoyers; Orran Krieger</td>
<td></td>
</tr>
<tr>
<td>Auptimizer - An Extensible, Open-Source Framework for Hyperparameter Tuning</td>
<td>339</td>
</tr>
<tr>
<td>Jiayi Liu; Samarth Tripathi; Unmesh Kurup; Mohak Shah</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>MULTILYRA: SCALABLE DISTRIBUTED EVALUATION OF BATCHES OF ITERATIVE</td>
<td>349</td>
</tr>
<tr>
<td>GRAPH QUERIES</td>
<td></td>
</tr>
<tr>
<td>Abbas Mazloumi ; Xiaolin Jiang ; Rajiv Gupta</td>
<td></td>
</tr>
<tr>
<td>AFRAME: EXTENDING DATAFRAMES FOR LARGE-SCALE MODERN DATA ANALYSIS</td>
<td>359</td>
</tr>
<tr>
<td>Phanuvadee Sinthong ; Michael J. Carey</td>
<td></td>
</tr>
<tr>
<td>PSPIE: PARTIAL MATCH SHERDING FOR COMPLEX EVENT PROCESSING</td>
<td>372</td>
</tr>
<tr>
<td>Ahmad Slo ; Sukanya Bhosnok ; Albert Flaug ; Kurt Rothermel</td>
<td></td>
</tr>
<tr>
<td>TAZER: HIDING THE COST OF REMOTE I/O IN DISTRIBUTED SCIENTIFIC WORKFLOWS</td>
<td>383</td>
</tr>
<tr>
<td>Joshua Suerieterlein ; Ryan D. Friese ; Nathan R. Tallent ; Malachi Schram</td>
<td></td>
</tr>
<tr>
<td>MECHANISM DESIGN FOR AN INCENTIVE-AWARE BLOCKCHAIN-ENABLED FEDERATED LEARNING PLATFORM</td>
<td>395</td>
</tr>
<tr>
<td>Kentaroh Toyoda ; Allan N. Zhang</td>
<td></td>
</tr>
<tr>
<td>NOVEL ONLINE ALGORITHMS FOR NONPARAMETRIC CORRELATIONS WITH</td>
<td>404</td>
</tr>
<tr>
<td>APPLICATION TO ANALYZE SENSOR DATA</td>
<td></td>
</tr>
<tr>
<td>Wei Xiao</td>
<td></td>
</tr>
<tr>
<td>ELASTIC EXECUTOR PROVISIONING FOR ITERATIVE WORKLOADS ON APACHE SPARK</td>
<td>413</td>
</tr>
<tr>
<td>Donglin Yang ; Wei Rang ; Daizhao Cheng ; Yu Wang ; Jiannan Tian ; Dingwen Tao</td>
<td></td>
</tr>
<tr>
<td>EIRENE: IMPROVING SHORT JOB LATENCY PERFORMANCE WITH COORDINATED</td>
<td>423</td>
</tr>
<tr>
<td>COLD DATA MIGRATION AND SCHEDULER-AWARE TASK CLONING</td>
<td></td>
</tr>
<tr>
<td>Wei Zhou ; K. Preston White ; Hongfeng Yu</td>
<td></td>
</tr>
<tr>
<td>THE OTREE: MULTIDIMENSIONAL INDEXING WITH EFFICIENT DATA SAMPLING FOR HPC</td>
<td>433</td>
</tr>
<tr>
<td>Cesare Cugnasco ; Hadrien Calmet ; Pol Santamaria ; Raúl Sirvent ; Ane Beatrix Eguzkitza ; Guillaume Houzeaux ; Tolaanda Becerra ; Jordi Torres ; Jesus Labarta</td>
<td></td>
</tr>
<tr>
<td>AVES: A DECISION ENGINE FOR ENERGY-EFFICIENT STREAM ANALYTICS ACROSS LOW-POWER DEVICES</td>
<td>441</td>
</tr>
<tr>
<td>Roshan Bharath Das ; Marc X. Makkes ; Alexandra Uta ; Lin Wang ; Henri Bal</td>
<td></td>
</tr>
<tr>
<td>SCALING OUT MULTI-WAY STREAM JOINS USING OPTIMIZED, ITERATIVE PROBING</td>
<td>449</td>
</tr>
<tr>
<td>Manuel Dossinger ; Sebastian Michel</td>
<td></td>
</tr>
<tr>
<td>DATA SERVICES WITH BINDAAS: RESTFUL INTERFACES FOR DIVERSE DATA SOURCES</td>
<td>457</td>
</tr>
<tr>
<td>Pradeebaru Kathiravel ; Yusaf Nadir Saghar ; Tushar Aggarwal ; Ashish Sharma</td>
<td></td>
</tr>
<tr>
<td>DATACALC: AD-HOC ANALYSES ON HETEROGENEOUS DATA SOURCES</td>
<td>463</td>
</tr>
<tr>
<td>Johannes Luong ; Dirk Habich ; Wolfgang Lehner</td>
<td></td>
</tr>
<tr>
<td>INVESTIGATING EDGE VS. CLOUD COMPUTING TRADE-OFFS FOR STREAM PROCESSING</td>
<td>469</td>
</tr>
<tr>
<td>Pedro Silva ; Alexandra Costan ; Gabriel Antoniu</td>
<td></td>
</tr>
<tr>
<td>EFFECTIVELY UNIFIED OPTIMIZATION FOR LARGE-SCALE GRAPH COMMUNITY DETECTION</td>
<td>475</td>
</tr>
<tr>
<td>Jianping Zeng ; Hongfeng Yu</td>
<td></td>
</tr>
<tr>
<td>SUBGRAPH2VEC: HIGHLY-VECTORIZED TREE-LIKE SUBGRAPH COUNTING</td>
<td>483</td>
</tr>
<tr>
<td>Langshi Chen ; Jiayu Li ; Cenk Sahinalp ; Madhav Marathe ; Anil Vullikanti ; Andrey Nikolaev ; Egor Smirnov ; Raslan Israfilov ; Judy Qiu</td>
<td></td>
</tr>
<tr>
<td>NIIOBE: AN INTELLIGENT I/O BRIDGING ENGINE FOR COMPLEX AND DISTRIBUTED WORKFLOWS</td>
<td>493</td>
</tr>
<tr>
<td>Kun Feng ; Harinaran Devarajan ; Anthony Kougkas ; Xian-He Sun</td>
<td></td>
</tr>
<tr>
<td>NETDYNA: MINING NETWORKED COEVOLVING TIME SERIES WITH MISSING VALUES</td>
<td>503</td>
</tr>
<tr>
<td>Hairi ; Hanghang Tong ; Lei Ying</td>
<td></td>
</tr>
<tr>
<td>G-FINDER: APPROXIMATE ATTRIBUTED SUBGRAPH MATCHING</td>
<td>513</td>
</tr>
<tr>
<td>Lihui Liu ; Boxin Du ; Jiejun Xu ; Hanghang Tong</td>
<td></td>
</tr>
<tr>
<td>EXPERIMENTAL EVALUATION OF BOUNDED-DEPTH LSM MERGE POLICIES</td>
<td>523</td>
</tr>
<tr>
<td>Qitchong Mao ; Steven Jacobs ; Waleed Amjad ; Vagelis Hristidis ; Vassilis J. Tsotras ; Neal E. Young</td>
<td></td>
</tr>
<tr>
<td>GENERALIZING DESIGN OF SUPPORT MEASURES FOR COUNTING FREQUENT PATTERNS IN GRAPHS</td>
<td>533</td>
</tr>
<tr>
<td>Jinghan Meng ; Napath Piaksiriratan ; Yicheng Tu</td>
<td></td>
</tr>
<tr>
<td>MULTIVARIATE LONG-TERM STATE FORECASTING IN CYBER-PHYSICAL SYSTEMS: A SEQUENCE TO SEQUENCE APPROACH</td>
<td>543</td>
</tr>
<tr>
<td>Nikhil Muratidhar ; Sathappan Mathiah ; Kiyoshi Nakayama ; Ratnesh Sharma ; Naren Ramakrishnan</td>
<td></td>
</tr>
<tr>
<td>SEMI-SUPERVISED LEARNING OVER STREAMING DATA USING MOA</td>
<td>553</td>
</tr>
<tr>
<td>Minh Nguyen ; Heitor Murilo Gomes ; Albert Bifet</td>
<td></td>
</tr>
<tr>
<td>UNDERSTANDING SPATIO-TEMPORAL URBAN PROCESSES</td>
<td>563</td>
</tr>
<tr>
<td>Lais M. A. Rocha ; Aline Bessa ; Fernando Chirigati ; Eugene Ofriel ; Mirella Moro ; Juliana Freire</td>
<td></td>
</tr>
</tbody>
</table>
TOWARDS INTERPRETABLE GRAPH MODELING WITH VERTEX REPLACEMENT
GRAMMARS............................................................................................................770
Justus Hibshman ; Satyaki Sikdar ; Tim Weninger

CLUSTERED HIERARCHICAL ENTROPY-SCALING SEARCH OF ASTRONOMICAL AND
BIOLOGICAL DATA ....................................................................................................780
Najib Ishaq ; George Student ; Noah M. Daniels

MULTI-STEP FORECASTING VIA MULTI-TASK LEARNING............................................790
Shayan Jawed ; Ahmed Rashid ; Lars Schmidt-Thieme

DATA CONTEXT ADAPTATION FOR ACCURATE RECOMMENDATION WITH ADDITIONAL
INFORMATION .............................................................................................................800
Hyunsik Jeon ; Rohan Koo ; U Kang

B-BIT SKETCH TRIE: SCALABLE SIMILARITY SEARCH ON INTEGER SKETCHES ..............810
Shunsuke Kanda ; Yasuo Tabei

MODELLING ONLINE COMMENT THREADS FROM THEIR START ..................................820
Rachel Krohn ; Tim Weninger

IMPROVING SCALABILITY OF PARALLEL CNN TRAINING BY ADJUSTING MINI-BATCH
SIZE AT RUN-TIME...................................................................................................830
Sanwoo Lee ; Qiao Kang ; Sandeep Madireddy ; Prassana Balaprapaksh ; Ankit Agrawal ; Alok Choudhary ;
Richard Archibald ; Wei-Keng Liao

MULTI-LAYER FACIAL REPRESENTATION LEARNING FOR EARLY PREDICTION OF
SEPTIC SHOCK ...........................................................................................................840
Chen Lin ; Julie Ivy ; Min Chi

JSCN: JOINT SPECTRAL CONVOLUTIONAL NETWORK FOR CROSS DOMAIN
RECOMMENDATION .....................................................................................................850
Zhixi Liu ; Lei Zheng ; Jiawei Zhang ; Jiaya Han ; Philip S. Yu

PREDICTING INFLUENCE PROBABILITIES USING GRAPH CONVOLUTIONAL NETWORKS ..860
Jing Liu ; Yudded Chen ; Duanshun Li ; Noseong Park ; Kisung Lee ; Dongwon Lee

HYBRID DEEP EMBEDDING FOR RECOMMENDATIONS WITH DYNAMIC ASPECT-LEVEL
EXPLANATIONS ...........................................................................................................870
Huanrui Luo ; Ning Yang ; Philip S. Yu

FAST COMPUTATION OF PERSISTENT HOMOLOGY WITH DATA REDUCTION AND DATA
PARTITIONING .............................................................................................................880
Nicholas O. Malott ; Philip A. Wilsey

HIGH DIMENSIONAL DATA CLUSTERING BY MEANS OF DISTRIBUTED DIRICHLET
PROCESS MIXTURE MODELS .......................................................................................890
Khudidja Meguelai ; Beneditte Fontez ; Nadine Hilgert ; Florent Masseglia

SENTIMENT-AWARE AND PERSONALIZED TOUR RECOMMENDATION ......................900
Prarthana Padia ; Kwan Hui Lim ; Jeffrey Chu ; Aaron Harwood

UNSUPERVISED AUTHOR DISAMBIGUATION USING HETEROGENEOUS GRAPH
CONVOLUTIONAL NETWORK EMBEDDING ...................................................................910
Ziyue Qiao ; Yi Du ; Yanjie Fu ; Pengfei Wang ; Yuanchun Zhou

NEWS RECOMMENDER SYSTEM CONSIDERING TEMPORAL DYNAMICS AND NEWS
TAXONOMY ................................................................................................................920
Shaina Raza ; Chen Ding

PATIENT-LEVEL CLASSIFICATION ON CLINICAL NOTE SEQUENCES GUIDED BY
ATTRIBUTED HIERARCHICAL ATTENTION ................................................................930
Cansu Sen ; Thomas Hartvigsen ; Xiangnan Kong ; Elke Rundensteiner

RECENT: NETWORK-BASED RECRUITMENT FOR ACTIVE FAKE NEWS CORRECTION ....940
Ajitesh Srivastava ; Rajgopal Kannan ; Charalampos Chelmini ; Viktor K. Prassana

SCALEBLY DISTRIBUTED SUBTRAJECTORY CLUSTERING.............................................950
Panagiotos Tampakis ; Nikos Pelekos ; Christos Doukeridis ; Yannis Theodoridis

EFFECTIVE UTILIZATION OF EXTERNAL KNOWLEDGE AND HISTORY CONTEXT IN
MULTI-TURN SPOKEN LANGUAGE UNDERSTANDING MODEL ....................................960
Yufan Wang ; Tingting He ; Rui Fan ; Wenji Zhou ; Xinhui Yu

DEEPBLUE: BI-LAYERED LSTM FOR TWEET POPULARITY ESTIMATION .....................968
Jian Wen ; Zhongbao Zhang ; Zichuang Yin ; Li Sun ; Sen Su ; Philip S. Yu

NON-LOCAL ATTENTION LEARNING ON LARGE HETEROGENEOUS INFORMATION
NETWORKS ................................................................................................................978
Yuxin Xiao ; Zecheng Zhang ; Carl Yang ; Chengxiang Zhai

DETECT: DEEP TRAJECTORY CLUSTERING FOR MOBILITY-BEHAVIOR ANALYSIS ........988
Mingsuan Yue ; Yaguang Li ; Haoze Yang ; Ritesh Ahuja ; Yao- Yi Chiang ; Cyrus Shulabi
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORIGIN: NON-RIGID NETWORK ALIGNMENT</td>
<td>998</td>
</tr>
<tr>
<td>Si Zhang ; Hanghang Tong ; Jiejun Xu ; Yifan Hu ; Ross Maciejewski</td>
<td></td>
</tr>
<tr>
<td>TOWARDS REAL TIME TEAM OPTIMIZATION</td>
<td>1008</td>
</tr>
<tr>
<td>Qinghai Zhou ; Liangye Li ; Hanghang Tong</td>
<td></td>
</tr>
<tr>
<td>INTEGRATING LOCAL VERTEX/EDGE EMBEDDING VIA DEEP MATRIX FUSION AND SIAMESE MULTI-LABEL CLASSIFICATION</td>
<td>1018</td>
</tr>
<tr>
<td>Yang Zhou ; Chao Jiang ; Zijie Zhang ; Dejing Dou ; Ruoming Jin ; Pengwei Wang</td>
<td></td>
</tr>
<tr>
<td>BIGMAT: A DISTRIBUTED AFFINITY-PRESERVING RANDOM WALK STRATEGY FOR INSTANCE MATCHING ON KNOWLEDGE GRAPHS</td>
<td>1028</td>
</tr>
<tr>
<td>Ali Assi ; Hamid Mecheick ; Wajdi Dhibi</td>
<td></td>
</tr>
<tr>
<td>MULTIMODAL, CONTEXT-AWARE, FEATURE REPRESENTATION LEARNING FOR CLASSIFICATION AND LOCALIZATION</td>
<td>1034</td>
</tr>
<tr>
<td>Sreyasee Das Bhattacharjee ; William J. Tolome ; Roy Cheria ; Urmimala Sarka</td>
<td></td>
</tr>
<tr>
<td>ARBITRATED DYNAMIC ENSEMBLE WITH ABSTAINING FOR TIME-SERIES</td>
<td>1040</td>
</tr>
<tr>
<td>Dihia Boulegane ; Albert Bifet ; Giyysarpuram Madhasudan</td>
<td></td>
</tr>
<tr>
<td>DEEP TENSOR FACTORIZATION FOR MULTI-CRITERIA RECOMMENDER SYSTEMS</td>
<td>1046</td>
</tr>
<tr>
<td>Zhengyu Chen ; Sibo Gai ; Donglin Wang</td>
<td></td>
</tr>
<tr>
<td>MOTIF2VEC: MOTIF AWARE NODE REPRESENTATION LEARNING FOR HETEROGENEOUS NETWORKS</td>
<td>1052</td>
</tr>
<tr>
<td>Manoj Reddy Dareddy ; Mahashweta Das ; Hao Yang</td>
<td></td>
</tr>
<tr>
<td>GRAPH MATCHING ON SOCIAL NETWORKS WITHOUT ANY SIDE INFORMATION</td>
<td>1060</td>
</tr>
<tr>
<td>Charalampos Davalas ; Dimitrios Michail ; Iraklis Varlamis</td>
<td></td>
</tr>
<tr>
<td>PYRAMID: A GENERAL FRAMEWORK FOR DISTRIBUTED SIMILARITY SEARCH ON LARGE-SCALE DATASETS</td>
<td>1066</td>
</tr>
<tr>
<td>Shiyuan Deng ; Xiao Yan ; K. W. Ng Kelvin ; Chenyu Jiang ; James Cheng</td>
<td></td>
</tr>
<tr>
<td>VISUALIZATION FOR QUALITY HEALTHCARE: PATIENT FLOW EXPLORATION</td>
<td>1072</td>
</tr>
<tr>
<td>Veronika Domova ; Shiva Sander-Tava lay</td>
<td></td>
</tr>
<tr>
<td>REFINING IMAGE SEARCH RESULTS USING MULTIPLE ATTRIBUTES</td>
<td>1080</td>
</tr>
<tr>
<td>Tingting Dong ; Shoji Nishimura ; Jianquan Liu</td>
<td></td>
</tr>
<tr>
<td>THE LINK REGRESSION PROBLEM IN GRAPH STREAMS</td>
<td>1088</td>
</tr>
<tr>
<td>Bowen Dong ; Charu C. Aggarwal ; Philip S. Yu</td>
<td></td>
</tr>
<tr>
<td>HIERARCHICAL-DOCUMENT-STRUCTURE-AWARE ATTENTION WITH ADAPTIVE COST SENSITIVE LEARNING FOR BIOMEDICAL DOCUMENT CLASSIFICATION</td>
<td>1096</td>
</tr>
<tr>
<td>Dandan Fang ; Jiyong Zhang ; Welichong Zhao ; Xianwei Xu ; Xingpeng Jiang ; Xiaohua Hu ; Tingting He</td>
<td></td>
</tr>
<tr>
<td>METAPATH ENHANCED GRAPH ATTENTION ENCODER FOR HINS REPRESENTATION LEARNING</td>
<td>1103</td>
</tr>
<tr>
<td>Iechao Guo ; Huo Yu Wang ; Haiying Shen</td>
<td></td>
</tr>
<tr>
<td>TASK FAILURE PREDICTION IN CLOUD DATA CENTERS USING DEEP LEARNING</td>
<td>1111</td>
</tr>
<tr>
<td>Jiechao Guo ; Huo Yu Wang ; Haiying Shen</td>
<td></td>
</tr>
<tr>
<td>GOSSIP LEARNING: OFF THE BEATEN PATH</td>
<td>1117</td>
</tr>
<tr>
<td>Lodovicio Giaretta ; Sarnas Girdzijauskas</td>
<td></td>
</tr>
<tr>
<td>VISUAL ANOMALY DETECTION IN EVENT SEQUENCE DATA</td>
<td>1125</td>
</tr>
<tr>
<td>Shunan Guo ; Zhaocen Jin ; Qing Chen ; David Gotz ; Hongyuan Zha ; Nan Cao</td>
<td></td>
</tr>
<tr>
<td>MULTI-SPATIAL SCALE EVENT DETECTION FROM GEO-TAGGED TWEET STREAMS VIA POWER-LAW VERIFICATION</td>
<td>1131</td>
</tr>
<tr>
<td>Yi Han ; Shanika Karunasekera ; Christopher Leckie ; Aaron Harwood</td>
<td></td>
</tr>
<tr>
<td>FROM TEXT CLASSIFICATION TO KEYPHRASE EXTRACTION FOR SHORT TEXT</td>
<td>1137</td>
</tr>
<tr>
<td>Song-Eun Lee ; Kang-Min Kim ; Woo-Jong Ryu ; Jemin Park ; Sangkeun Lee</td>
<td></td>
</tr>
<tr>
<td>ENHANCING ITEMSET TREE RULES AND PERFORMANCE</td>
<td>1143</td>
</tr>
<tr>
<td>Jay Lewis ; Ryan G. Benton ; David Bourrie ; Jennifer Lavergne</td>
<td></td>
</tr>
<tr>
<td>MC2: UNSUPERVISED MULTIPLE SOCIAL NETWORK ALIGNMENT</td>
<td>1151</td>
</tr>
<tr>
<td>Gen Li ; Li Sun ; Zhonghao Zhang ; Pengxin Ji ; Sen Su ; Philip S. Yu</td>
<td></td>
</tr>
<tr>
<td>GRAPH MATCHING VIA MULTI-SCALE HEAT DIFFUSION</td>
<td>1157</td>
</tr>
<tr>
<td>Lin Li ; Daniel L. Sussman</td>
<td></td>
</tr>
<tr>
<td>COMMUNITY-PRESERVING GRAPH CONVOLUTIONS FOR STRUCTURAL AND FUNCTIONAL JOINT EMBEDDING OF BRAIN NETWORKS</td>
<td>1163</td>
</tr>
<tr>
<td>Jiahaot Liu ; Guixiang Ma ; Fei Jiang ; Chun-Ta Lu ; Philip S. Yu ; Ann B. Ragon</td>
<td></td>
</tr>
<tr>
<td>LATTE: APPLICATION ORIENTED SOCIAL NETWORK EMBEDDING</td>
<td>1169</td>
</tr>
<tr>
<td>Lin Meng ; Jiayang Bai ; Jianwei Zhang</td>
<td></td>
</tr>
<tr>
<td>D2D-TM: A CYCLE VAE-GAN FOR MULTI-DOMAIN COLLABORATIVE FILTERING</td>
<td>1175</td>
</tr>
<tr>
<td>Linh Nguyen ; Tsukasa Ishigaki</td>
<td></td>
</tr>
</tbody>
</table>
APPLICATION OF MULTI-COLOR 3D-CUBE SUCCESSIVE CONVOLUTION NETWORK FOR CONVECTIVE STORM NOWCASTING ................................................................. 1705
Wei Zhang ; Lei Han ; Juan Zhang ; Han Yang Guo ; Jie Dai
AD BLOCKING WHITELIST PREDICTION FOR ONLINE PUBLISHERS ................................................................. 1711
Shuai Zhao ; Achir Kalra ; Chong Wang ; Cristian Borcea ; Yi Chen
A CONGESTION DIFFUSION MODEL WITH INFLUENCE MAXIMIZATION FOR TRAFFIC BOTTLENECKS IDENTIFICATION IN METROPOLITAN SCALES ................................................................. 1717
Baoshan Zhao ; Chengzhong Xu ; Siyu Liu ; Juanjuan Zhao ; Li Li
ATTRIBUTED SEQUENCE EMBEDDING ................................................................................................................. 1723
Zhongfang Zhuang ; Xiangqun Kong ; Rundensteiner Elke ; Jihane Zouaoui ; Aditya Arora
FEDERATED MULTIMODAL BIG DATA STORAGE & ANALYTICS PLATFORM FOR ADDITIVE MANUFACTURING ................................................................................................................. 1729
Kareem S. Aggour ; Vijay S. Kumar ; Paul Cuddihy ; Jenny Weisenberg Williams ; Vipul Gupta ; Laura Dsyl ; Tim Hanlon ; Justin Gambone ; Joseph Vinciquerra
DELOG: A HIGH-PERFORMANCE PRIVACY PRESERVING LOG FILTERING FRAMEWORK ................................................................. 1739
Amei Agrawal ; Abhishek Dixit ; Namrata A Shettar ; Darshil Kapadia ; Vikram Agrawal ; Rajat Gupta ; Rohit Karulpia
VERTICA FLATTENED TABLES AND LIVE AGGREGATE PROJECTIONS: A COLUMN-BASED ALTERNATIVE TO MATERIALIZED VIEWS FOR ANALYTICS ................................................................................................................................. 1749
Yuanzhe Bei ; Thao Pham ; Akshay Aggarwal ; Nga Tran ; Jaimin Dave ; Chuck Bear ; Michael Leuchtenburg
VISUAL ANALYTICS FOR DEEP EMBEDDINGS OF LARGE SCALE MOLECULAR DYNAMICS SIMULATIONS ................................................................................................................................. 1759
Junghoon Chae ; Dobsindhu Bhowmik ; Heng Ma ; Arvind Ramanathan ; Chad Steed
INFECT: A LARGE SCALE GRAPH-BASED FRAUD DETECTION SYSTEM FOR E-COMMERCE INSURANCE ................................................................................................................................. 1765
Cen Chen ; Chen Liang ; Jiaxin Lin ; Li Wang ; Qizi Liu ; Xincing Yang ; Jun Zhou ; Yang Shuang ; Yuan Qi
SEQ2GRAPH: DISCOVERING DYNAMIC NON-LINEAR DEPENDENCIES FROM MULTIVARIATE TIME SERIES ................................................................................................................................. 1774
Xuan-Hong Dang ; Syed Tousafah Shah ; Petros Zerfos
DATA-CENTRIC HELICOPTER FAILURE ANTICIPATION: THE MGB OIL PRESSURE VIRTUAL SENSOR CASE ................................................................................................................................. 1784
Nassir Dauquay ; Ammar Meouchche ; Pierre-Loïc Matsonneau ; Vasilie-Marian Scuturici ; Jean-Marc Petit
SCALING UP HETEROGENEOUS WAVEFORM CLUSTERING FOR LONG-DURATION MONITORING SIGNAL ACQUISITION, ANALYSIS, AND INTERACTION: BRIDGING BIG DATA ANALYTICS WITH MEASUREMENT INSTRUMENT USAGE PATTERN ................................................................................................................................. 1794
Masaharu Goto ; Naoki Kobayashi ; Gang Ren ; Mitsunori Ogihara
HIGH-RESOLUTION ROAD VEHICLE COLLISION PREDICTION FOR THE CITY OF MONTREAL ................................................................................................................................. 1804
Antoine Helbert ; Timothée Guédon ; Tristan Glataud ; Brigitte Jaumard
TOWARDS POWER EFFICIENCY IN DEEP LEARNING ON DATA CENTER HARDWARE ................................................................................................................................. 1814
Miro Hodak ; Masha Gorkovenko ; Ajay Dholakia
LARGE DATA FLOW GRAPHS IN LIMITED GPU MEMORY ................................................................................................................................. 1821
Geert Janssen ; Vladimir Zolotov ; Tung D. Le
INTELLIGENT EXPLORATION OF LARGE-SCALE DATA: WHAT CAN WE LEARN IN TWO PASSES? ................................................................................................................................. 1831
Chandrika Kamath
AN EFFICIENT CLOUD-BASED FRAMEWORK FOR DIGITAL MEDIA KNOWLEDGE EXTRACTION ................................................................................................................................. 1841
Chaitanya Kanchibhotla ; Prathviraj Venkatesh ; Dvivn Somayajula ; P Raddha Krishna
CONTEXTUAL PRICE FEATURES FOR E-COMMERCE SEARCH RANKING ................................................................................................................................. 1851
Ishita Khan ; Aritra Mandal ; Prathyusha Senthil Kumar
A FRAMEWORK FOR EXPLAINABLE TEXT CLASSIFICATION IN LEGAL DOCUMENT REVIEW ................................................................................................................................. 1858
Christian J. Mahoney ; Jianping Zhang ; Nathaniel Huber-Fliffett ; Peter Gronwall ; Haozheng Zhao
TARGETED DISPLAY ADVERTISING: THE CASE OF PREFERENTIAL ATTACHMENT ................................................................................................................................. 1868
Saurav Manchanda ; Pranjul Yadav ; Khoa Doan ; S. Sathiya Keerthi
MNP INSIDE OUT: A GAME THEORY ASSISTED MACHINE LEARNING MODEL TO DETECT SUBSCRIBER CHURN BEHAVIORS UNDER CHINA'S MOBILE NUMBER PORTABILITY POLICY ................................................................................................................................. 1878
Ye Ouyang ; Aidong Yang ; Shuming Zeng ; Fanyu Meng
BRIDGING THE GAP BETWEEN COMMUNITY AND NODE REPRESENTATIONS: GRAPH EMBEDDING VIA COMMUNITY DETECTION .............................................................................................................................. 2681
Artem Lutov ; Dingqi Yang ; Philippe Cugrê-Mauroux

AN EFFECTIVE AND SCALABLE DATA MODELING FOR ENTERPRISE BIG DATA PLATFORM .............................................................................................................................. 2691
Jayesh Patel

ESTIMATION OF DISEASE CODE FROM ELECTRONIC PATIENT RECORDS .............................................................................................................................. 2698
Shusaku Tsomoto ; Tomohiro Kimura ; Harako Iwata ; Shoji Hirano

A GA-BASED FRAMEWORK FOR MINING HIGH FUZZY UTILITY ITEMSETS .............................................................................................................................. 2708
Jimmy Ming-Tai Wu ; Jerry Chun-Wei Lin ; Philippe Fournier-Viger ; Tomasz Witkowski ; Tsung-Pei Hong ; Matin Pirouz

A DATA-DRIVEN APPROACH FOR CONTINUOUS ADHERENCE PREDICTIONS IN SLEEP APNEA THERAPY MANAGEMENT .............................................................................................................................. 2716
Mathew Araujo ; Louis Kazaglis ; Conrad Iber ; Jaideep Srivastava

THE SERUMS TOOL-CHAIN: ENSURING SECURITY AND PRIVACY OF MEDICAL DATA IN SMART PATIENT-CENTRIC HEALTHCARE SYSTEMS .............................................................................................................................. 2726
V. Janicj ; J. K. F. Bowles ; A. F. Vermeulen ; A. Silvina ; M. Belk ; C. Fidas ; A. Pitillides ; M. Kumari ; M. Rossbory ; M.Vinov ; T. Given-Wilson ; A. Legay ; E. Blackledge ; R. Arredouani ; G. Stylianou ; W. Huang

CLASSIFICATION MODELS AND SURVIVAL ANALYSIS FOR PROSTATE CANCER USING RNA SEQUENCING AND CLINICAL DATA .............................................................................................................................. 2736
Md Faisal Kabir ; Simone A. Ludwig

REGIONAL ANALYSIS OF DEATH RATE DUE TO AIR POLLUTION IN TURKEY AND ITS NEIGHBORS .............................................................................................................................. 2746
Yunus Emre Karazac ; Sultan Turhan ; Ozgun Pinarer ; A. Teoman Naskali

USING HOSPITAL ADMINISTRATIVE DATA TO INFERENCE PATIENT-PATIENT CONTACT VIA THE CONSISTENT CO-PRESENCE ALGORITHM .............................................................................................................................. 2756
Jeffrey Lienert ; Felix Reed-Tochas ; Laura Koehly ; Christopher Steven Marcum

ANALYZING PUBLIC OUTLOOK TOWARDS VACCINATION USING TWITTER .............................................................................................................................. 2763
Ratnaja Mahajan ; William Romine ; Michele Miller ; Tanvi Banerjee

A HYBRID MODEL USING LSTM AND DECISION TREE FOR MORTALITY PREDICTION AND ITS APPLICATION IN PROVIDER PERFORMANCE EVALUATION .............................................................................................................................. 2773
Peichang Shi ; Ariya Gangopadhyay ; Carolyn Owens ; Brenda Blunt ; Christine Grogan

SINGLE-CELL REGULATORY NETWORK INFECTION AND CLUSTERING FROM HIGH-DIMENSIONAL SEQUENCING DATA .............................................................................................................................. 2782
Aristidis G. Vrahatis ; Georgios N. Dimitrakopoulos ; Sotiris K. Tasoulis ; Spiros V. Georgakopoulos ; Vassili P. Plagianakos

THE DEVELOPMENT OF MACHINE LEARNING INFUSED OUTPATIENT PROGNOSTIC MODELS FOR TACKLING IMPACTS OF CLIMATE CHANGE AND ENSURING DELIVERY OF EFFECTIVE POPULATION HEALTH SERVICES .............................................................................................................................. 2790
Jaya Shankar Vuppalapati ; Santosh Kedari ; Anitha Ilapakurti ; Chandrasekar Vuppalapati ; Sharat Kedari ; Rajasekar Vuppalapati

ENHANCING CLINICAL INFORMATION RETRIEVAL THROUGH CONTEXT-AWARE QUERIES AND INDICES .............................................................................................................................. 2800
Andrew Wen ; Yanshan Wang ; Vinod C. Kaggal ; Sijia Liu ; Hongfang Liu ; Jungwei Fan

TRANSPOSE-BASED INTEGRATED DATA REDUCTION TECHNIQUES FOR SPEEDING UP CLASSIFIER TRAINING .............................................................................................................................. 2808
Reham M. Alamro ; Abdou S. Youssef

HIERARCHICAL TRANSFER CONVOLUTIONAL NEURAL NETWORKS FOR IMAGE CLASSIFICATION .............................................................................................................................. 2817
Xishuang Dong ; Hsiang-Huang Wu ; Yuchong Yan ; Lijun Qian

RECONSTRUCTION OF AGENTS’ CORRUPTED TRAJECTORIES OF COLLECTIVE MOTION USING LOW-RANK MATRIX COMPLETION .............................................................................................................................. 2826
Kelun Gajamannage ; Randy Paffenroth

AUTONOMOUS WORKLOAD CHANGE CLASSIFICATION AND PREDICTION FOR BIG DATA WORKLOADS .............................................................................................................................. 2835
Mikhael Genkin ; Frank Dehne

PARALLEL GRADIENT BOOSTING BASED GRANGER CAUSALITY LEARNING .............................................................................................................................. 2845
Pei Guo ; Chen Lian ; Yan Tang ; Jianwu Wang

MODELLING WEALTH FROM CALL DETAIL RECORDS AND SURVEY DATA WITH MACHINE LEARNING: EVIDENCE FROM PAPUA NEW GUINEA .............................................................................................................................. 2855
Muhammad Rizal Khafi ; Hendrik ; Dhurani Dhar Burra ; Rio Fandi Dianco ; Dikara Maitri Pradipta Alkariyaya ; Muhammad Rheza Mutzahid ; Anissara Zahara ; George Hodge ; Rajius Idzalika
ARCHCONTRACT: USING SMART CONTRACTS FOR DISPOSITION

Danielle Alves Batista; Tim Weingaertner

TOWARDS A FLEXIBLE SYSTEM ARCHITECTURE FOR AUTOMATED KNOWLEDGE BASE CONSTRUCTION FRAMEWORKS

Ozcan Din

AUTOMATED INTERPRETABILITY OF LINKED DATA ONTOLOGIES: AN EVALUATION WITHIN THE CULTURAL HERITAGE DOMAIN

 Nano Freire; Sjors De Valk

BUILDING THE NATIONAL RADIO RECORDINGS DATABASE: A BIG DATA APPROACH TO DOCUMENTING AUDIO HERITAGE

Emily Goodman; Mark A. Mattenzo; Shawn Vancour; William Vanden Dries

EXTENDING THE SCOPE OF COMPUTATIONAL ARCHIVAL SCIENCE: A CASE STUDY ON LEVERAGING ARCHIVAL AND ENGINEERING APPROACHES TO DEVELOP A FRAMEWORK TO DETECT AND PREVENT “FAKE VIDEO”

Hoda Hamouda; Jessica Bushey; Victoria Lemieux; James Stewart; Corinne Rogers; James Cameron; Ken Thibodeau; Chen Feng

USING DATA PARTITIONS AND STATELESS SERVERS TO SCALE UP FEDORA REPOSITORIES

Gregory Jansen; Aaron Coburn; Adam Soroka; Richard Marciano

DIGITAL LEGACIES ON PAPER: READING PUNCHCARDS WITH COMPUTER VISION

Gregory Jansen

PRELIMINARY ANALYSIS OF A LARGE-SCALE DIGITAL ENTERTAINMENT DEVELOPMENT ARCHIVE: A CASE STUDY OF THE ENTERTAINMENT TECHNOLOGY CENTER’S PROJECTS

Eric Kaltman

ENTERPRISE ARCHITECTURE — A VALUE PROPOSITION FOR RECORDS PROFESSIONALS

Shadreck Katsa

REFRAMING DIGITAL CURATION PRACTICES THROUGH A COMPUTATIONAL THINKING FRAMEWORK

Richard Marciano; Sarah Agarrat; Hannah Frisch; Margaret Rose Hunt; Kanishk Jain; Genevieve Kocienda; Hannah Krauss; Chenxi Liu; Mary McKinley; Danish Mir; Connor Mullane; Emery Patterson; Debashish Pradhan; James Santos; Britton Schams; Huda Hamouda; Jessica Bushey; Victoria Lemieux; James Stewart; Corinne Rogers; James Cameron; Ken Thibodeau; Chen Feng

AN INTELLIGENT CLASS: THE DEVELOPMENT OF A NOVEL CONTEXT CAPTURING FRAMEWORK SUPPORTING THE FUNCTIONAL AUTO-CLASSIFICATION OF RECORDS

Nathaniel Payne

COMPUTATIONAL THINKING IN ARCHIVAL SCIENCE RESEARCH AND EDUCATION

William Underwood; Richard Marciano

WHAT COMPUTATIONAL ARCHIVAL SCIENCE CAN LEARN FROM ART HISTORY AND MATERIAL CULTURE STUDIES

Lyneise Williams

AUTOMATIC EXTRACTION OF PERSONALITY FROM TEXT: CHALLENGES AND OPPORTUNITIES

Nazma Akram; Johan Fernqvist; Tim Ishib:er; Lisa Kaati; Björn Pelzer

EXAMINATION OF DOUBLE ARBITER PUFs ON SECURITY AGAINST MACHINE LEARNING ATTACKS

Mezhah A. Alamro; Yu Zhuang; Ahmad O. Aseeri; Mohammed Saeed Alkatheiri

MALICIOUS URL LINKAGE ANALYSIS AND COMMON PATTERN DISCOVERY

Shin-Ying Huang; Tzu-Hsien Chuang; Shi-Meng Huang; Tao Ban

LEARNING TO GENERATE DIVERSE AND AUTHENTIC REVIEWS VIA AN ENCODER-DECODER MODEL WITH TRANSFORMER AND GRU

Kaifu Jin; Xi Zhang; Jia Yuan Zhang

BYAKKO: AUTOMATIC WHITELIST GENERATION BASED ON OCCURRENCE DISTRIBUTION OF FEATURES OF NETWORK TRAFFIC

Nobuyuki Kanaya; Yu Tsuda; Yuuki Takano; Daisuke Inoue

A FRAMEWORK FOR CYBER THREAT INTELLIGENCE EXTRACTION FROM RAW LOG DATA

Max Landauer; Florian Skopik; Markus Wurzenberger; Wolfgang Hotwagner; Andreas Rauber

DETECTING ADVERSARY USING WINDOWS DIGITAL ARTIFACTS

Seng Pei Liew; Satoshi Beda

DETECTION OF PHISHING WEBSITES USING GENERATIVE ADVERSARIAL NETWORK

Pierrick Robic-Butez; Thu Yen Win
PARALLELIZED TOPOLOGICAL RELAXATION ALGORITHM .......................................................... 3406
Guangchen Ruan ; Hui Zhang
PARALLEL R COMPUTING ON THE WEB .................................................................................. 3416
Ranjini Sabramanian ; Hui Zhang
SPATIAL-TEMPORAL SCIENTIFIC DATA CLUSTERING VIA DEEP CONVOLUTIONAL NEURAL NETWORK ........................................................................................................... 3424
Jianxin Sun ; Chunchun Wu ; Yafeng Ge ; Yusong Li ; Hongfeng Yu
PARALLEL TRAINING VIA COMPUTATION GRAPH TRANSFORMATION ..................................... 3430
Fei Wang ; Guoyang Chen ; Weifeng Zhang ; Taek Romg
A METHODOLOGY FOR CROSS-PLATFORM, EVENT-DRIVEN BIG DATA ANALYTICS-AS-A-SERVICE ...................................................................................................................... 3440
Claudio A. Ardagna ; Valerio Bellandi ; Paolo Ceravolo ; Ernesto Damiani ; Rino Fiuazz
CONTEXT-AUGMENTED SOFTWARE DEVELOPMENT IN TRADITIONAL AND BIG DATA PROJECTS: LITERATURE REVIEW AND PRELIMINARY FRAMEWORK ........................................... 3449
Glauco Melo ; Paulo Alencar ; Don Cowan
A FRAMEWORK FOR IDENTIFYING AND PRIORITIZING DATA ANALYTICS OPPORTUNITIES IN ADDITIVE MANUFACTURING ................................................................. 3458
Hyunseo Park ; Hyunwoo Ko ; Yung-Tsuan T. Lee ; Hyunho Cho ; Paul Witherell
SKI: AN AGILE FRAMEWORK FOR DATA SCIENCE ................................................................. 3468
Jeffrey Saltz ; Alex Sutherland
ACHIEVING AGILE BIG DATA SCIENCE: THE EVOLUTION OF A TEAM'S AGILE PROCESS METHODOLOGY .................................................................................................................. 3477
Jeffrey S. Saltz ; Ivan Shamsharin
A HYBRID APPROACH TO DYNAMIC ENTERPRISE DATA PLATFORM ..................................... 3486
Mehmet Selman Sezgin ; Ahmet Tugrul Bayrak ; Olcay Taner Yildiz
REAL-TIME MACHINE LEARNING COMPETITION ON DATA STREAMS AT THE IEEE BIG DATA 2019 .......................................................................................................................... 3493
Dhiya Boulegane ; Nedeljko Radulovic ; Albert Bijet ; Ghislain Fievet ; Jimin Sohn ; Yeonwoo Nam ; Seojeong Yu ; Dong-Wan Choi
COLLABORATIVE STREAMING: TRUST REQUIREMENTS FOR PRICE SHARING ......................... 3498
Tobias Grubenmann ; Daniele Dell’Aglio ; Abraham Bernstein
MLK SMART CORRIDOR: AN URBAN TESTBED FOR SMART CITY APPLICATIONS .......... 3506
Austin Harris ; Jose Stovall ; Mina Saritip
SCALABLE AND RELIABLE MULTI-DIMENSIONAL AGGREGATION OF SENSOR DATA STREAMS .................................................................................................................. 3512
Sören Henning ; Wilhelm Hasselbring
TV AD EVENTS AND DIGITAL SEARCH: ON THE SELECTION OF OUTCOME MEASURES .......... 3518
Shavendra Hill ; Anthony Colas ; H. Andrew Schwartz ; Gordon Bartch
IMAGE MINING FOR REAL TIME QUALITY ASSURANCE IN RAPID PROTOTYPING .................. 3529
Sebastian Trinks ; Carsten Felden
A DEEP LEARNING APPROACH TO TRESPASSING DETECTION USING VIDEO SURVEILLANCE DATA ..................................................................................................................... 3535
Mazanmnl Bashir ; Elke A. Rundensteiner ; Rameza Ahsan
CLASSIFYING IN-VEHICLE NOISE FROM MULTI-CHANNEL SOUND SPECTRUM BY DEEP BEAMFORMING NETWORKS ........................................................................................................ 3545
Seok-Jun Bu ; Sung-Bae Cho
AGENT-BASED MODELING TO SIMULATE ROAD TRAVEL USING BIG DATA FROM SMARTPHONE GPS: AN APPLICATION TO THE CONTINUENTAL UNITED STATES .............................................................................................................................. 3553
Sashikanth Gurram ; Vijayaraghavan Sivaraman ; Jonathan T. Apple ; Abdul R. Pinjari
TRAVEL PATTERN EXTRACTION FROM SMART CARD DATA USING DATA POLISHING .......... 3563
Mio Hose ; Masashi Kawano ; Taku Moriyama ; Kosuke Miyazaki ; Masaki Ito
USING TIMED SEQUENTIAL PATTERNS IN THE TRANSPORTATION INDUSTRY ...................... 3573
Somayah Karsoun ; Le Gruenwald ; Clark Barrus ; Eleazar Leal
PERSONALIZED POI EMBEDDING FOR SUCCESSIVE POI RECOMMENDATION WITH LARGE-SCALE SMART CARD DATA .......................................................................................... 3583
Jin-Young Kim ; Kyung-Hyun Lim ; Sung-Bae Cho
MODELING DYNAMIC SPATIAL-TEMPORAL CLUSTER RELATIONSHIPS .................................. 3590
Ivens Portugal ; Paulo Alencar ; Donald Cowan
BLOCKCHAIN TECHNOLOGY AS A MECHANISM FOR DIGITAL RAILWAY TICKETING .................. 3599
J. D. Preece ; J. M. Easton
CPSITRES: A COLLABORATIVE SYSTEM FOR ANALYSIS OF BIG DATA ON SEA ICE
Vinit VeerendraVeer Singh; Scott Sorensen; Chandra Kambhamettu
4134
SEXTANT: A COMPUTATIONAL FRAMEWORK FOR SCALABLE AND EFFICIENT CORRELATION OF SPATIO-TEMPORAL TRAJECTORIES
Brian Thompson; Dave Czedel; Jeremy Martin; Kristen Snee; Alex Cheung
4138
PTOLEMAIC INDEXING FOR MANAGING AND QUERYING INTERNET OF THINGS (IOT) DATA
Christian Beecks; Fabian Berns; Kjeld Willy Schmidt
4148
A NEW APPROACH FOR EFFICIENT STRUCTURE DISCOVERY IN IOT
Fabian Berns; Kjeld Schmidt; Alexander Grass; Christian Beecks
4152
EFFECTIVELY TESTING SYSTEM CONFIGURATIONS OF CRITICAL IOT ANALYTICS PIPELINES
Morgan K. Geldenhuys; Lauritz Thamsen; Kain Kordian Gonturskay; Felix Lorenz; Olej Kao
4157
DETECTING ERRORS IN SHORT-TERM ELECTRICITY DEMAND FORECAST USING PEOPLE DYNAMICS
Guillaume Hubault; Yasutaka Nishimura; Kiyohito Yoshihara; Chihiro Ono
4163
KSENSOR: COORDINATED INTELLIGENCE FROM CO-LOCATED SENSORS
Oliivera Kotevska; Kalyan Perumalla; Juan Lopez
4169
TOWARDS AN INTERACTIVE VISUALIZATION FRAMEWORK FOR IOT DEVICE DATA FLOW
Ebelechukwu Nwafor; Habeeb Oluwofobi
4175
IOT BASED URBAN NOISE MONITORING IN DEEP LEARNING USING HISTORICAL REPORTS
Sayed Khushal Shah; Zeenat Tariq; Yuyong Lee
4179
SMART HOME IOT ANOMALY DETECTION BASED ON ENSEMBLE MODEL LEARNING FROM HETEROGENEOUS DATA
Sihai Tang; Zhaochen Gu; Qing Yang; Song Fu
4185
SPEECH EMOTION DETECTION USING IOT BASED DEEP LEARNING FOR HEALTH CARE
Zeenat Tariq; Sayed Khushal Shah; Yuyong Lee
4191
ASSOCIATION MODEL BETWEEN VISUAL FEATURE AND AQI RANK USING LIFEOF DATA
Phuong-Binh Vo; Trong-Dat Phan; Minh-Son Dao; Koji Zettzu
4197
AUTOMATIC HYPERPARAMETER TUNING METHOD FOR LOCAL OUTLIER FACTOR, WITH APPLICATIONS TO ANOMALY DETECTION
Zekun Xu; Debrjit Kukre; Arin Chaudhuri
4201
COST-ADAPTIVE NEURAL NETWORKS FOR PEAK VOLUME PREDICTION WITH EMM FILTERING
Bin Yu; Giovanna Graciani; Anderson Naucimento; Juhua Bu
4208
SCALABILITY ANALYSIS OF BLOCKCHAIN ON A SERVERLESS CLOUD
Alex Kaplanovich; Karuna P. Joshi; Yelena Yesh
4214
PREDICTING TRANSACTION LATENCY WITH DEEP LEARNING IN PROOF-OF-WORK BLOCKCHAINS
Enrico Tedeschi; Tor-Arne S. Nordmo; Dag Johansen; Håvard D. Johansen
4223
AONT-NZZD: A SECURE AND EFFICIENT DISPERSAL SCHEME IN DISTRIBUTED STORAGE SYSTEMS
Pengcheng Xie; Hui Li; Haiyang Yu; Zequn Chen
4232
SECURITY VALIDATION TESTING ENVIRONMENT IN THE CLOUD
Cristian Bucur; Eduard Babulak
4240
THE IMPACT OF PREPROCESSING IN NATURAL LANGUAGE FOR OPEN SOURCE INTELLIGENCE AND CRIMINAL INVESTIGATION
Jan William Johnsen; Katrin Franke
4248
BASIC FORENSIC PROCEDURES FOR CYBER CRIME INVESTIGATION IN SMART GRID NETWORKS
Igor Kotsiuba; Inna Skarga-Bandurova; Alkiviadis Giannakoulias; Oksana Budia
4255
TECHNIQUE FOR FINDING AND INVESTIGATING THE STRONGEST COMBINATIONS OF CYBERATTACKS ON SMART GRID INFRASTRUCTURE
Igor Kotsiuba; Inna Skarga-Bandurova; Alkiviadis Giannakoulias; Mykhailo Chaikin; Aleksandar Jevremovic
4265
DEEP IN THE DARK: A NOVEL THREAT DETECTION SYSTEM USING DARKNET TRAFFIC
Sanjay Kumar; Harald Vranken; Joost Van Dijk; Timo Hamalainen
4273
PACE: PLATFORM FOR ANDROID MALWARE CLASSIFICATION AND PERFORMANCE EVALUATION
Ajit Kumar; Vinit Agarwal; Shishir K. Shandilya; Andrii Shulaginov; Saket Upadhyay; Bhavnna Yadav
4280
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>PURCHASE PREDICTION IN FREE ONLINE GAMES VIA SURVIVAL ANALYSIS</td>
<td>4289</td>
</tr>
<tr>
<td>SYSTEM</td>
<td>4294</td>
</tr>
<tr>
<td>COMPARATIVE STUDY OF WEAR-LEVELING IN SOLID-STATE DRIVE WITH NTFS FILE SYSTEM</td>
<td>4299</td>
</tr>
<tr>
<td>DETECTING WEB SPAM IN WEBGRAPHS WITH PREDICTIVE MODEL ANALYSIS</td>
<td>4309</td>
</tr>
<tr>
<td>CYBERCRIME INVESTIGATIONS IN THE ERA OF SMART APPLICATIONS: WAY FORWARD THROUGH BIG DATA</td>
<td>4315</td>
</tr>
<tr>
<td>FORECASTING OF TRENDS IN LEGAL SPEND MANAGEMENT</td>
<td>4320</td>
</tr>
<tr>
<td>EXPERIMENTAL EVALUATION OF CNN PARAMETERS FOR TEXT CATEGORIZATION IN LEGAL DOCUMENT REVIEW</td>
<td>4325</td>
</tr>
<tr>
<td>IMAGE ANALYTICS FOR LEGAL DOCUMENT REVIEW : A TRANSFER LEARNING APPROACH</td>
<td>4329</td>
</tr>
<tr>
<td>NORMALISATION OF SWIFT MESSAGE COUNTERPARTIES WITH FEATURE EXTRACTION AND CLUSTERING</td>
<td>4337</td>
</tr>
<tr>
<td>QUALITATIVE MAPPING MODELING OF CRIMINALS' SENSE OF SECURITY IN THEFT CASES</td>
<td>4342</td>
</tr>
<tr>
<td>JAPANESE MISTAKABLE LEGAL TERM CORRECTION USING INFREQUENCY-AWARE BERT CLASSIFIER</td>
<td>4352</td>
</tr>
<tr>
<td>SUPERVISED KEY TERMS CLUSTERING FOR REGULATORY MONITORING</td>
<td>4357</td>
</tr>
<tr>
<td>FAILING &amp; FALLING (F&amp;F): LEARNING TO CLASSIFY ACCIDENTS AND INCIDENTS IN AIRCRAFT DATA</td>
<td>4366</td>
</tr>
<tr>
<td>PRESCRIPTIVE EQUIPMENT MAINTENANCE: A FRAMEWORK</td>
<td>4375</td>
</tr>
<tr>
<td>MACHINE LEARNING USE CASES FOR SMART MANUFACTURING KPIS</td>
<td>4381</td>
</tr>
<tr>
<td>SPATIOTEMPORAL REAL-TIME ANOMALY DETECTION FOR SUPERCOMPUTING SYSTEMS</td>
<td>4390</td>
</tr>
<tr>
<td>FORECASTING CROSS-BORDER POWER EXCHANGES THROUGH AN HVDC LINE USING DYNAMIC MODELLING</td>
<td>4395</td>
</tr>
<tr>
<td>APPLICATION OF MACHINE LEARNING AND SPATIAL BOOTSTRAPPING TO IMAGE PROCESSING FOR PREDICTIVE MAINTENANCE</td>
<td>4402</td>
</tr>
<tr>
<td>SELF-SUPERVISED MULTI-STAGE ESTIMATION OF REMAINING USEFUL LIFE FOR ELECTRIC DRIVE UNITS</td>
<td>4412</td>
</tr>
<tr>
<td>DATA IMUTATION METHOD BASED ON PROGRAMMING BY EXAMPLE: AREP-S</td>
<td>4422</td>
</tr>
<tr>
<td>WIND TURBINE OPERATIONAL STATE PREDICTION: TOWARDS FEATURELESS, END-TO-END PREDICTIVE MAINTENANCE</td>
<td>4431</td>
</tr>
<tr>
<td>SUBJECT-ORIENTED DATA RETRIEVAL AND ANALYSIS ON SINA WEIBO</td>
<td>4435</td>
</tr>
<tr>
<td>SEASONALITY-ADJUSTED CONCEPTUAL-RELEVANCY-AWARE RECOMMENDER SYSTEM IN ONLINE GROCERIES</td>
<td>4444</td>
</tr>
<tr>
<td>PURCHASE PREDICTION IN FREE ONLINE GAMES VIA SURVIVAL ANALYSIS</td>
<td>4449</td>
</tr>
</tbody>
</table>

**Authors**: Wanshan Yang; Ting Huang; Junlin Zeng; Yan Tang; Lijun Chen; Shivakant Mishra; Youjian Eugene Liu
AN ENERGY-EFFICIENT RECONFIGURABLE LSTM ACCELERATOR FOR NATURAL
LANGUAGE PROCESSING ..........................................................4450
Elham Azari; Sarma Vrudhala

EXPLORATION OF OPENCL 2D CONVOLUTION KERNELS ON INTEL FPGA, CPU, AND GPU
PLATFORMS ..................................................................................4460
Zheming Jin; Hal Finkel

A CASE STUDY OF K-MEANS CLUSTERING USING SYCL .............................4466
Zheming Jin; Hal Finkel

BAYESIAN-BASED HYPERPARAMETER OPTIMIZATION FOR SPIKING NEUROMORPHIC
SYSTEMS .......................................................................................4472
Maryam Parsa; J. Parker Mitchell; Catherine D. Schuman; Robert M. Patton; Thomas E. Potok; Kaushik Roy

EVOLVING ENERGY EFFICIENT CONVOLUTIONAL NEURAL NETWORKS ..........................4479
Steven R. Young; Pravallika Devineni; Maryam Parsa; J. Travis Johnston; Bill Kay; Robert M. Patton;
Catherine D. Schuman; Derek C. Rose; Thomas E. Potok

QUANTUM GROVER SEARCH-BASED OPTIMIZATION FOR INNOVATIVE MATERIAL
DISCOVERY ......................................................................................4486
Sima Esfandiarpour Borujeni; Ramkumar Harikrishnakumar; Saideep Nannapaneni

LEARNING TO PREDICT MATERIAL STRUCTURE FROM NEUTRON SCATTERING DATA ..........4490
Cristina Garcia-Cardona; Ramakrishnan Kannan; Travis Johnston; Thomas Proffen; Katharine Page; Sudip K. Seal

VISUALIZATION SYSTEM FOR EVOLUTIONARY NEURAL NETWORKS FOR DEEP
LEARNING ........................................................................................4498
Jungloo Chae; Catherine D. Schuman; Steven R. Young; J. Travis Johnston; Derek C. Rose; Robert M. Patton;
Thomas E. Potok

CLUSTERED LATENT DIRICHLET ALLOCATION FOR SCIENTIFIC DISCOVERY ..................4503
Christopher Gropp; Alexander Herzog; Ilya Safro; Paul W. Wilson; Amy W. Apon

REALISTIC TRANSPORT SIMULATION: TACKLING THE SMALL DATA CHALLENGE
WITH OPEN DATA ........................................................................4512
Guina Guo; Jalal Majed Khalil; Du Yan; Virginia Stiopiku

MACHINE LEARNING FOR PREDICTION OF MID TO LONG TERM HABITUAL
TRANSPORTATION MODE USE .......................................................4520
Alina Lazar; Alexandra Ballow; Ling Jin; C. Anna Spurlock; Alexander Sim; Kesheng Wu

UNDERSTANDING DATA SIMILARITY IN LARGE-SCALE SCIENTIFIC DATASETS .................4525
Payton Linton; William Melodia; Alina Lazar; Deborah Agarwal; Ludovico Bianchi; Devarshi Ghoshal;
Gilbert Pastorello; Lavanya Ramakrishnan; Kesheng Wu

DETECTING DEPENDENCY BETWEEN DISCRETE RANDOM VARIABLES AND
APPLICATION .................................................................................4532
Edgar Llamas; Ivan Garcia; Andres Mendez

EXPLORATION OF WORKFLOW MANAGEMENT SYSTEMS EMERGING FEATURES FROM
USERS PERSPECTIVES ......................................................................4537
Ryan Mitchell; Loic Pottier; Steve Jacobs; Rafael Ferreira Da Silva; Mats Rynge; Karan Vahi; Ewa Deelman

EMPOWERING AGROECOSYSTEM MODELING WITH HTC SCIENTIFIC WORKFLows:
THE CYCLES MODEL USE CASE .....................................................4545
Rafael Ferreira Da Silva; Rajiv Mayani; Yuping Shi; Armen R. Kemanian; Mats Rynge; Ewa Deelman

EVALUATING SCIENTIFIC WORKFLOW ENGINES FOR DATA AND COMPUTE INTENSIVE
DISCOVERIES ..................................................................................4553
Rina Singh; Jeffrey A. Graves; Valentine Anantharaj; Sreenivas R. Sakumara

INFORMATION EXTRACTION FROM CANCER PATHOLOGY REPORTS WITH GRAPH
CONVOLUTION NETWORKS FOR NATURAL LANGUAGE TEXTS .........................4561
Hong-Jun Yoon; John Guanley; M. Todd Young; Georgia Tourassi

HUMAN-MACHINE INFORMATION EXTRACTION SIMULATOR FOR BIOLOGICAL
COLLECTIONS ..................................................................................4565
Icaro Alcântara; Aditi Mailadi; Andréa Matsumaga; Mauricio Tsugawa; Fortes José A. B.

WATAPI: COMPOSING WEB API SPECIFICATION FROM API DOCUMENTATIONS
THROUGH AN INTELLIGENT AND INTERACTIVE ANNOTATION TOOL ..............................4573
Pedro Cardinhas; Bogaslaw Obura; Georgios Theodoropoulos; Ibad Kareshi

ANALYSING SOCIAL MEDIA AS A HYBRID TOOL TO DETECT AND INTERPRET LIKELY
RADICAL BEHAVIOURAL TRAITS FOR NATIONAL SECURITY ........................................4579

CLUSTERCLEAN: A WEAK SEMI-SUPERVISED APPROACH FOR CLEANING DATA
LABELS .............................................................................................4589
Kyriaki Dimitriadou; Rahul Manghwani; Timothy C. Hoad
A LOCATION INDEPENDENT MACHINE LEARNING APPROACH FOR EARLY FAKE NEWS DETECTION ................................................................. 4740
Haoxian Liu

IOCMINER: AUTOMATIC EXTRACTION OF INDICATORS OF COMPROMISE FROM TWITTER ........................................................................ 4747
Amirreza Niamatallah; Lida Safarinejad; Reginald Harper; Bei-Tseng Chu

PRAT - A TOOL FOR ASSESSING RISK IN WRITTEN COMMUNICATION .......................................................... 4755
Amendra Shrestha; Lisa Kautt; Nazar Akrami

NETWORK EMBEDDING: ON COMPRESSION AND LEARNING ............................................................................................ 4763
Eva Akbas; Mehmet Emin Aktas

DISTRIBUTED COMMUNITY DETECTION IN LARGE NETWORKS USING AN INFORMATION-THEORETIC APPROACH ........................................ 4773
Md Abdul Motaleb Faysal; Shaikh Arifuzzaman

A SCALABLE GRAPH ANALYTICS FRAMEWORK FOR PROGRAMMING WITH BIG DATA IN R (PBDR) ................................................................................. 4783
S. M. Shamimul Hasan; Drew Schmidt; Ramakrishnan Kannan; Neena Imam

EFFICIENT SIMILARITY-BASED ALIGNMENT OF TEMPORALLY-SITUATED GRAPH NODES WITH APACHE SPARK ........................................ 4793
Hubert Naaeke; Ke Li; Bernd Amann; Olivier Caré

GRAPEHEO: CHARACTERIZING AND UNDERSTANDING SOFTWARE EVOLUTION USING CALL GRAPHS ........................................................................ 4799
Vijay Walunj; Gharib Gharibi; Duy H. Ho; Yuyang Lee

COMPUTING COMPLEX GRAPH PROPERTIES WITH SQL QUERIES ......................................................................................... 4808
Xiantian Zhou; Carlos Ordonez

PREDICTING POST-STROKE HOSPITAL DISCHARGE DISPOSITION USING INTERPRETABLE MACHINE LEARNING APPROACHES ........................................................................ 4817
Jin Cho; Alnoor Alharrin; Zhen Hu; Nancy Fell; Mina Sartipi

RECURRENT NEURAL NETWORK BASED FEATURE SELECTION FOR HIGH DIMENSIONAL AND LOW SAMPLE SIZE MICRO-ARRAY DATA .................................................. 4823
Shanta Chowdhury; Xishuang Dong; Xianglang Li

REINFORCEMENT LEARNING FRAMEWORK TO IDENTIFY CAUSE OF DISEASES - PREDICTING ASTHMA ATTACK CASE ................................................................................. 4829
Quan Do; Son Tran; Alexa Doig

COMPUTER-AIDED CLINICAL SKIN DISEASE DIAGNOSIS USING CNN AND OBJECT DETECTION MODELS ................................................................................................. 4839
Xin He; Shihao Wang; Shaohua Shi; Zhenheng Tang; Yuxin Wang; Zhihao Zhao; Jing Dai; Ronghao Ni; Xiaofeng Zhang; Xiaoming Liu; Zhihui Wu; Wu Yu; Xiaowen Chu

AUTOMATED MACHINE LEARNING FOR EEG-BASED CLASSIFICATION OF PARKINSON’S DISEASE PATIENTS ......................................................................................... 4845
Milan Koch; Victor Geraedts; Hao Wang; Martijn Tannemaat; Thomas Bäck

TOWARDS EXPLAINABLE MELANOMA DIAGNOSIS: PREDICTION OF CLINICAL INDICATORS USING SEMI-SUPERVISED AND MULTI-TASK LEARNING ........................................................................ 4853
Seiya Murabayashi; Hitoshi Iyatomi

STOCHASTIC GASTRIC IMAGE AUGMENTATION FOR CANCER DETECTION FROM X-RAY IMAGES ......................................................................................... 4858
Hideaki Okamoto; Quan Hua Cap; Takakiyo Nomura; Hitoshi Iyatomi; Jan Hashimoto

EXPLOITING ANTI-MONOTONIC CONSTRAINTS IN MINING PALINDROMIC MOTIFS FROM BIG GENOMIC DATA ................................................................................. 4864
Olawafemi A. Sarumi; Carson K. Leung

BAYESIAN NON-LINEAR SUPPORT VECTOR MACHINE FOR HIGH-DIMENSIONAL DATA WITH INCORPORATION OF GRAPH INFORMATION ON FEATURES ........................................................................ 4874
Wenli Sun; Changgee Chang; Qi Long

EXPLAINABLE DEEP LEARNING APPLIED TO UNDERSTANDING OPIOID USE DISORDER AND ITS RISK FACTORS ................................................................................. 4883
T. Elizabeth Workman; Yijun Shao; Joel Kupersmith; Friedhelm Sandbrink; Joseph L. Goulet; Nawar M. Shaar; Christopher SPEVAK; Cynthia Brandt; Marc R. Blackman; Qing Zeng-Treitler

DISCOVERING SUBLANGUAGES IN A LARGE CLINICAL CORPUS THROUGH UNSUPERVISED MACHINE LEARNING AND INFORMATION GAIN ........................................................................ 4889
T. Elizabeth Workman; Guy Divita; Qing Zeng-Treitler

NOISY SUBGRAPH ISOMORPHISMS ON MULTIPLEX NETWORKS ......................................................................................... 4899
Hui Jin; Xie He; Yanghai Wang; Hao Li; Andrea L. Bertozzi
SUSPICIOUS NETWORK EVENT RECOGNITION LEVERAGING ON MACHINE LEARNING

AN IDENTITY PRIVACY PRESERVING IOT DATA PROTECTION SCHEME FOR CLOUD BASED ANALYTICS

KRATOS: A SECURE, AUTHENTICATED AND PUBLICLY VERIFIABLE SYSTEM FOR EDUCATIONAL DATA USING THE BLOCKCHAIN

UTILITY AND PRIVACY ASSESSMENTS OF SYNTHETIC DATA FOR REGRESSION TASKS

RIBS: RISKY BLIND-SPOTS FOR ATTACK CLASSIFICATION MODELS

THE ANTI-DATA-MINING (ADM) FRAMEWORK - BETTER PRIVACY ON ONLINE SOCIAL NETWORKS AND BEYOND

DROID-NNET: DEEP LEARNING NEURAL NETWORK FOR ANDROID MALWARE DETECTION

PRIVACY-PRESERVING TOP-K DOMINATING QUERIES IN DISTRIBUTED MULTI-PARTY DATABASES

IMPROVING K-NEAREST NEIGHBOR PATTERN RECOGNITION MODELS FOR PRIVACY-PRESERVING DATA ANALYSIS

EXPERIENTIAL LEARNING: CASE STUDY-BASED PORTABLE HANDS-ON REGRESSION LABWARE FOR CYBER FRAUD PREDICTION

LOCAL DIFFERENTIAL PRIVACY WITH K-ANONYMITY FOR FREQUENCY ESTIMATION

A DEEP LEARNING MODEL WITH MULTI-SCALE SKIP CONNECTIONS FOR SOLAR FLARE PREDICTION COMBINED WITH PRIOR INFORMATION

WINDOW-BASED FEATURE EXTRACTION METHOD USING XGBOOST FOR TIME SERIES CLASSIFICATION OF SOLAR FLARES

SOLAR FLARE PREDICTION USING TWO-TIER ENSEMBLE WITH DEEP LEARNING AND GRADIENT BOOSTING MACHINE

AN APPROACH FOR SCALE SUSPICIOUS NETWORK EVENTS DETECTION

AUTOMATED EVENT PRIORITIZATION FOR SECURITY OPERATION CENTER USING DEEP LEARNING

SUSPICIOUS NETWORK EVENT RECOGNITION USING MODIFIED STACKING ENSEMBLE MACHINE LEARNING

IEEE BIGDATA 2019 CUP: SUSPICIOUS NETWORK EVENT RECOGNITION

AN APRIORI-BASED DATA ANALYSIS ON SUSPICIOUS NETWORK EVENT RECOGNITION

NAIVE TRANSFER LEARNING APPROACHES FOR SUSPICIOUS EVENT PREDICTION

APPLICATION OF XGBOOST TO THE CYBER-SECURITY PROBLEM OF DETECTING SUSPICIOUS NETWORK TRAFFIC EVENTS

IEEE BIGDATA 2019 CUP: BINARY CLASSIFICATION VIA TENSOR COMPLETION

SUSPICIOUS NETWORK EVENT RECOGNITION LEVERAGING ON MACHINE LEARNING
INTERDEPENDENCY BETWEEN THE STOCK MARKET AND FINANCIAL NEWS

COMPARING IMPLEMENTATION VARIANTS OF DISTRIBUTED SPATIAL JOIN ON SPARK

TOPOLOGICAL DEFORMATIONS IN BIOLOGICAL SYSTEMS

A STATISTICAL CAUSAL INFERENCE METHOD FOR EXPLORING ULTRASONICS AND REALISTIC TRANSPORT SIMULATION WITH OPEN DATA

HARDENING DGA CLASSIFIERS UTILIZING IVAP

ON UNDERSTANDING DEGRADATION KINETICS OF VIDEO INTERPOLATION

ON UNDERSTANDING BIOSONAR DEFORMATIONS USING DEEP LEARNING

CORANK: SIMULTANEOUSLY RANKING STATISTICAL CORRECTION OF AVERAGE CUSTOMER RATINGS FOR PRODUCT TEMPERATURE SAFELY DURING PILGRIMAGE

AN OPTIMIZED SCHEDULING PROCESS FOR A LARGE CROWD TO PERFORM SPATIO-TEMPORAL MOVEMENTS SAFELY DURING PILGRIMAGE

STUDYING THE RELATIONSHIP BETWEEN SELF-ASSEMBLY DEGRADATION AND NON-VISUAL PERCEPTION BASED ON THE CONCEPT OF ATOMIC EMBEDDING

EASYRAIN: A USER-FRIENDLY PLATFORM FOR COMPARING PRECIPITATION NOWCASTING MODELS

PARTS CATALOG OBJECT RECOGNITION TECHNOLOGY FOR EFFICIENT DRAWING DISTRIBUTION MANAGEMENT

HAPPY TOILET: A SOCIAL ANALYTICS APPROACH TO THE STUDY OF PUBLIC TOILET CLEANLINESS

PRESERVING COMPOSITION AND CRystal STRUCTURES OF CHEMICAL COMPOUNDS IN ATOMIC EMBEDDING

WEB SCRAPING: STATE-OF-THE-ART AND AREAS OF APPLICATION

A PROGRAMMING FRAMEWORK FOR HETEROGENEOUS STREAM ANALYTICS

BIG FEDERAL DATA CENTERS IMPLEMENTING FAIR DATA PRINCIPLES: ARM DATA CENTER EXAMPLE

PSIDB: A FRAMEWORK FOR BATCHED QUERY PROCESSING AND OPTIMIZATION

AN OPTIMIZED SCHEDULING PROCESS FOR A LARGE CROWD TO PERFORM SPATIO-TEMPORAL MOVEMENTS SAFELY DURING PILGRIMAGE

STATISTICAL CORRECTION OF AVERAGE CUSTOMER RATINGS FOR PRODUCT RANKING

CORANK: SIMULTANEOUSLY RANKING PUBLICATION VENUES AND RESEARCHERS

ON UNDERSTANDING BIOSONAR DEFORMATIONS USING DEEP LEARNING-BASED VIDEO INTERPOLATION

ON UNDERSTANDING DEGRADATION KINETICS OF PHARMACEUTICAL GELATIN MATRICES FOR PRECISION MEDICINE: A DEEP LEARNING APPROACH

HARDENING DGA CLASSIFIERS UTILIZING IVAP

REALISTIC TRANSPORT SIMULATION WITH OPEN DATA

A STATISTICAL CAUSAL INFERENCE METHOD FOR EXPLORING ULTRASONICS AND TOPOLOGICAL DEFORMATIONS IN BIOLOGICAL SYSTEMS

COMPARING IMPLEMENTATION VARIANTS OF DISTRIBUTED SPATIAL JOIN ON SPARK

INTERDEPENDENCY BETWEEN THE STOCK MARKET AND FINANCIAL NEWS
CLUSTERED HIERARCHICAL ENTROPY-SCALING SEARCH OF ASTRONOMICAL AND BIOLOGICAL DATA ..........................................................................................................................6077
Najib Ishaq ; George Student ; Noah M. Daniels

ACCELERATING AN ON-LINE APPROXIMATION MINING FOR LARGE CLOSED ITEMSETS ..................6078
Koji Iwamura ; Takumi Nishina ; Yoshiyuki Yamamoto

VECTOR-BASED CHURN PREDICTION USING NEURAL NETWORKS IN MOBILE GAMES ..................6081
Khoon Jang ; Junwhan Kim ; Byunggu Yu

DETECTION OF HIJACKED AUTHORITATIVE DNS SERVERS BY NAME RESOLUTION TRAFFIC CLASSIFICATION ..................................................................................6084
Yong Jin ; Masahiko Tonoishi ; Satoshi Matsura

CITATION-BASED AUTHOR CONTRIBUTION MEASURE FOR BYLINE-INDEPENDENCY ..................6086
Sukhwan Jung ; Wan Chul Yoon

ESTABLISHING THE SAFETY OF A SMART HEART HEALTH MONITORING SERVICE THROUGH VALIDATION ..................................................................................6089
Murtadha Kareem ; Oliver Faust

BORDER GATEWAY PROTOCOL ANOMALY DETECTION USING NEURAL NETWORK .................6092
Mohsen Karimi ; Ali Jahanasahi ; Abbas Mazoumi ; Hadi Zamani Sabz

TOWARDS A MULTI-ENGINE QUERY OPTIMIZER FOR COMPLEX SQL QUERIES ON BIG DATA .....................................................................................................................6095
Evdokia Kassela ; Ioannis Konstantinou ; Nectarios Koziris

BIGOPTIBASE: BIG DATA ANALYTICS FOR BASE STATION ENERGY CONSUMPTION OPTIMIZATION ..................................................................................................................6098
Evdokia Kassela ; Nikodimos Provatou ; Asterios Tsouvaras ; Ioannis Konstantinou ; Nectarios Koziris

A STUDY ON SOLUTION ORIENTED SMART FACTORY DIAGNOSTIC SYSTEM FOR SME ........6101
Seung-Beom Son ; Jun-Yeong Kwon ; Chae-Soo Kim

APPLICATION OF BIG DATA ANALYTICS TO SUPPORT POWER NETWORKS AND THEIR TRANSITION TOWARDS SMART GRIDS ........................................................................6104
Sylvie Koziel ; Patrik Hilber ; Ryutaro Ichise

RECURRENT PATH INDEX FOR EFFICIENT GRAPH TRAVERSAL ..........................................................................................................................................................6107
Kazuma Kusu ; Kenji Hatanoda

MODEL MANAGEMENT AND HANDWRITTEN CHARACTER RECOGNITION: AN ENTERPRISE SOLUTION ..............................................................................................................6110
Choudur Lakshminarayana ; Thigagarajan Ramakrishnan ; Awny Al-Omari ; Khaled Bouaziz ; Faraz Ahmad ; Sri Raghavan ; Prama Agarwal

AI-BASED ONLINE P2P LENDING RISK ASSESSMENT ON SOCIAL NETWORK DATA WITH MISSING VALUE ........................................................................................................6113
Warren Landis ; Sanghwan Cha ; Majid Shuaibi

ON OPTIMIZATION OF STOCK MARKET PREDICTION METHODS ......................................................6116
Lok Ting Lam ; Shan-Wen Hsiao

HOW DOES FAKE NEWS SPREAD: RAISING AWARENESS & EDUCATING THE PUBLIC WITH A SIMULATION TOOL ..........................................................................................6119
Cheng L. Lee ; Joel-David J. J. Wong ; Zi Y. Lim ; Belinda S. T. Th o ; Sean S. W. Kwek ; Kyong Jin Shin

DEVELOPMENT OF A CONTINUOUS COMPLEX EVENT PROCESSING PLATFORM FOR REAL-TIME TACTICAL MOVING OBJECTS ........................................................................6122
Jiwon Lee ; Bonghee Hong ; Chunsu Kim ; Woo Chan Kim

UTILIZING MULTIVARIATE TIME SERIES FOR SEMANTIC SEGMENTATION ..................................6125
Frederique Van Leeuwen

MUSIC SEQUENCE PREDICTION WITH MIXTURE HIDDEN MARKOV MODELS ................................6128
Tao Li ; Minsoo Choi ; Kaiming Fu ; Lei Lin

A JOINT MODEL OF CLINICAL DOMAIN CLASSIFICATION AND SLOT FILLING BASED ON RCNN AND BIGRU-CRF .....................................................................................6133
Yuming Li ; Pin Ni ; Junkun Peng ; Jiayi Zhu ; Zhenjin Dai ; Gangmin Li ; Xuming Bai

LEARNING A FACETED CUSTOMER SEGMENTATION FOR DISCOVERING NEW BUSINESS OPPORTUNITIES AT INTEL ..................................................................................6136
Itay Lieder ; Meirav Segal ; Eran Avidan ; Asaf Cohen ; Tom Hope

A HIERARCHICAL CLASSIFICATION MODEL FOR SOLAR FLARE PREDICTION .................................6139
Cenru Liu

A MODERN APPROACH TO BIG PROVENANCE ....................................................................................6142
Crisrael Lucero ; Phuong H. Nguyen

FEDERATED QUERY PROCESSING FOR BIG DATA IN DATA SCIENCE ...........................................6145
Manoj Maniswamiah ; Tilak Agrawala ; Charles C. Tappert
PLUGIN TO A HEALTHIER LIFE: A WEB BROWSER PLUGIN FOR MENTAL HEALTH MONITORING

Jane H. K. Seah ; Kyong Jin Shin

AIRLINE MILES REDEMPTION

Joseph Sebastian ; Thilanka Munasinghe

BIG DATA ANALYSIS ON SOCIAL NETWORKING

Zhengwu Sun ; Dan Chiu-Tien Lo ; Yong Shi

MITIGATING DDOS WITH POW AND GAME THEORY

Kun-Yuan Sung ; Shau-Wen Hsiao

THE INTERNET OF RESPONSIBILITIES - CONNECTING HUMAN RESPONSIBILITIES

Xuejiao Tang ; Jiong Qiu ; Wenbin Zhang ; Ibrahim Toure ; Mingli Zhang ; Enza Messina ; Xueping Xie ; Xuebing Wang ; Sheng Yu

USING DEEP CONVOLUTIONAL NEURAL NETWORK FOR MOUSE BRAIN SEGMENTATION IN DT-MRI

Anh Thai ; Vy Bui ; Laura Reyes ; Lin-Ching Chang

USER AUTHENTICATION THROUGH KEYSTROKE DYNAMICS BY MEANS OF MODEL CHECKING: A PROPOSAL

Fabio Di Tommaso ; Michele Guerra ; Fabio Martinelli ; Francesco Merlaldo ; Massimo Piedimonte ; Giovanni Rosa ; Antonella Santone

WORD SENSE DISAMBIGUATION USING WIKIPEDIA LINK GRAPH

Hai-Lun Tu ; Pei-Chen Ho ; Jason S. Chang ; Li-Guang Chen

ROBUST FEATURES FOR ACTIVITIES RECOGNITION

Kashyap Venkatesh ; Bashar Barmada ; Veronica Liesaputra ; Guillermo Ramirez-Prado

LARGE SCALE TIME SERIES ANALYSIS FOR INFRASTRUCTURE RELIABILITY

Zhichao Wang ; Shashank Singh ; Arnold Pereira

DATA-DRIVEN CONTROL OF A CLASS OF DISCRETE-TIME LINEAR COMPLEX DYNAMICAL NETWORKS

Wei Wang ; Chungsang Feng ; Wei Quan

MAP ADJUSTMENT AS A BASE FOR PRIVACY ASSURANCE IN SEMANTIC SPATIAL TRAJECTORIES RELEASE

Xingang Wang ; Renshi Yu ; Taixian Liu ; Wenhua Li ; Xiaoling Sun

CLASSIFYING GENUINE FACE IMAGES FROM DISGUISED FACE IMAGES

Junyaup Kim ; Silho Han ; Simon S. Woo

LAUNCHING A USER-GENERATED CONTENT CAMPAIGN TO PROMOTE PUMA’S SOCIAL CONTRIBUTIONS

Lili Xiang ; Alice Mello ; Ryan Ackerman

A GRAPH CONSTRUCTION STUDY FOR GRAPH-BASED SEMI-SUPERVISED LEARNING: CASE STUDY ON UNSTRUCTURED TEXT DATA

Sumedh Yadav ; Gautam Kumar ; Shivam Kumar

EFFECTIVENESS OF USABILITY & PERFORMANCE FEATURES FOR WEB CREDIBILITY EVALUATION

Kenta Yamada ; Hayato Yamana

METHOD AND DATASET MINING IN SCIENTIFIC PAPERS

Rujing Yao ; Linlin Hou ; Yingchun Ye ; Ji Zhang ; Jian Wu

AN WORD2VEC BASED ON CHINESE MEDICAL KNOWLEDGE

Jiayi Zhu ; Pin Ni ; Yuming Li ; Junkun Peng ; Zhenjin Dai ; Gangmin Li ; Xuming Bai

Author Index