World Congress 2017
www.worldcongressdsa.com

The Frontiers in Intelligent Data and Signal Analysis

New York, USA
July 8 - 20, 2017

17th Industrial Conference on Data Mining
ICDM 2017

July 12 – 16, 2017

www.data-mining-forum.de
Preface

The sixteenth event of the Industrial Conference on Data Mining ICDM was held in New York (www.data-mining-forum.de) running under the umbrella of the World Congress on “The Frontiers in Intelligent Data and Signal Analysis, DSA 2017” (www.worldcongressdsa.com).

After the peer-review process, we accepted 26 high-quality papers for oral presentation. The topics range from theoretical aspects of data mining to applications of data mining, such as in multimedia data, in marketing, in medicine, and in process control, industry, and society. Extended versions of selected papers will appear in the international journal Transactions on Machine Learning and Data Mining (www.ibai-publishing.org/journal/mldm).

A tutorial on Data Mining, a tutorial on Case-Based Reasoning, a tutorial on Intelligent Image Interpretation and Computer Vision in Medicine, Biotechnology, Chemistry and Food Industry, and a tutorial on Standardization in Immunofluorescence were held before the conference.

We would like to thank all reviewers for their highly professional work and their effort in reviewing the papers.

We also thank the members of the Institute of Applied Computer Sciences, Leipzig, Germany (www.ibai-institut.de), who handled the conference as secretariat. We appreciate the help and understanding of the editorial staff at Springer Verlag, and in particular Alfred Hofmann, who supported the publication of these proceedings in the LNAI series.

Last, but not least, we wish to thank all the speakers and participants who contributed to the success of the conference. We hope to see you in 2018 in New York at the next World Congress on “The Frontiers in Intelligent Data and Signal Analysis, DSA 2018” (www.worldcongressdsa.com), which combines under its roof the following three events: International Conferences Machine Learning and Data Mining, MLDM, the Industrial Conference on Data Mining, ICDM, and the International Conference on Mass Data Analysis of Signals and Images in Medicine, Biotechnology, Chemistry and Food Industry, MDA.

July 2017

Petra Perner
17th Industrial Conference on Data Mining
ICDM 2017

Chair
Petra Perner .................................. IBaI Leipzig, Germany

Program Committee
Ajith Abraham...............................Machine Intelligence Research Labs, USA
Brigitte Bartsch-Spörl ......................BSR Consulting GmbH, Germany
Orlando Belo ................................University of Minho, Portugal
Bernard Chen .................................University of Central Arkansas, USA
Jeroen de Bruin ..............................Medical University of Vienna, Austria
Antonio Dourado ............................University of Coimbra, Portugal
Stefano Ferilli ...............................University of Bari, Italy
Geert Gins ....................................KU Leuven, Belgium
Warwick Graco ..............................ATO, Australia
Aleksandra Gruca ...........................Silesian University of Technology, Poland
Hartmut Ilgner ..............................Council for Scientific and Industrial Research, South Africa
Pedro Isaias ..................................Universidade Aberta, Portugal
Piotr Jedrzejowicz .........................Gdynia Maritime University, Poland
Martti Juhola .................................University of Tampere, Finland
Janusz Kacprzyk ............................Polish Academy of Sciences, Poland
Mehmed Kantardzic .......................University of Louisville, USA
Eduardo F. Morales .........................INAOE, Ciencias Computacionales, Mexico
Armand Prieditris............................Newstar Labs, USA
Rainer Schmidt.................................University of Rostock, Germany
Victor Sheng........................................University of Central Arkansas, USA
Kaoru Shimada.................................Section of Medical Statistics, Fukuoka Dental College, Japan
Gero Szepannek..............................Santander Consumer Bank, Germany

Additional Reviewers

Juliane Perner.................................Cancer Research UK Cambridge Institute
ICDM 2017 – Program

July 12th, 2017

8:15 - 8:50 am  Registration

8:50 - 9:00 am  Opening
   *Petra Perner*

**First Session**  
*Session Chair: Harry Zhang*

9:00 - 9:30 am  Incorporating Positional Information into Deep Belief Networks for Sentiment Classification  
   *Yong Jin, Harry Zhang, Donglei Du*

9:30 - 10:00 am  Tracking Multiple Social Media for Stock Market Event Prediction  
   *Fang Jin, Wei Wang, Prithwish Chakraborty, Nathan Self, Feng Chen, Naren Ramakrishnan*

10:00 - 10:30 am  Ensemble Sales Forecasting Study in Semiconductor Industry  
   *Qiuping Xu, Vikas Sharma*

10:30 - 10:45 am  Coffee Break

**Second Session**  
*Session Chair: Kaoru Shimada*

10:45 - 11:15 am  Towards an efficient method of modeling “Next Best Action” for Digital Buyer’s journey in B2B  
   *Anit Bhandari, Kiran Rama, Nandini Seth, Nishant Niranjan, Parag Chitalia, Stig Berg*

11:15 - 11:45 am  Association Rule-based Classifier Using Artificial Missing Values  
   *Kaoru Shimada, Takaaki Arahira, Takashi Hanioka*

11:45 - 12:15  Mining Location-based Service Data for Feature Construction in Retail Store Recommendation  
   *Tsung-Yi Chen, Lyu-Cian Chen, Yuh-Min Chen*
12:15 - 1:45 pm    Lunch

Opening
Session Chair: Francisco Díaz-Pérez

1:45 - 2:15 pm    Poster Session
Poster stay all time during conference

Posters

   Kazuhiro Kohara

2. Particle Filter in Single Object Tracking with MapReduce
   Di Wang and Min Chen

3. Edge Detection Using MapReduce Based Ant Colony Optimization
   Michael J. Papesca and Min Chen

4. Clustering aided Support Vector Machines
   Goce Ristanoski, Rahul Soni, Sutharshan Rajasegarar, James Bailey, Christopher Leckie

5. Industry 4.0: Mining Physical Defects in Production of Surface-Mount Devices
   Farshid Tavakolizadeh, José Ángel Carvajal Soto, Dávid Gyulai, and Christian Beecks

   Richa Pandey, Rajeshwari Sah, Poornima Kulkarni

7. A Big Data Analytics framework for smart grid fault prediction
   Quang Nguyen, Bernardo Hermont, Sindhu Suresh

8. Surfacing Value from Unstructured Patient Data
   Andrew Malinow
9. Deploying Advanced Data Analytics Techniques with Speech Analytics Outputs for Fraud Detection
Sunil Kappal

10. Simulation based Approach for Special Discounting Threshold
Shashank Shekhar, Kaushik Shankar, Bishwarup Bhattacharjee, Sushanta Kumar Mishra, Sam Pritchett, Parag Chitalia

11. Geographic Named Entity Recognition in Twitter
LaKeisha Williams, Sarah Darley, Lucas Overbey, Scott Batson

12. A graph-based ranking model for automatic keyphrases extraction from Arabic documents
Mohamed Salim EL BAZZI

13. MapReduce and Spark-based Analytic Framework Using Social Media Data for Earlier Flu Outbreak Detection
Ali Al Essa and Miad Faezipour

14. Deep learning from label proportions using Batch Averager
Ehsan Mohammady Ardehaly, Aron Culotta

15. A Graph Data System: Implementation, Algorithms and Applications
Yan Zhang and Yiming Wu, Tuo Zhu, Yafu Li, Zherui Lu, Qinglong Zhang, Zhi-Jie Zhang, and Zheng Xu

S. Yasrobi, J. Alston, B. Yadranjiaagham and N. Tabrizi

17. Using Classification Algorithms for Telemonitoring Parkinson’s Disease Severity
Francisco Diaz-Pérez, Alfonsa García-López, Manuel Rubio-Sanchez, and Agustín Álvarez-Marquina

18. Development of a Text-Based Classifier to Identify Immunology-Related NIH Grants
Fenglou Mao, Susan F. Cooper, Dawei Lin
2:15 - 2:30 pm  Joint Exhibition Walk

**Third Session**  
*Session Chair: Walid Atwa*

2:30 - 3:00 pm  Constraint-based Clustering Algorithm for Multi-Density Data and Arbitrary Shapes  
*Walid Atwa, Kan Li*

3:00 - 3:30 pm  Towards a Large Scale Practical Churn Model for Prepaid Mobile Markets  
*Amit Kumar Meher, Jobin Wilson, R. Prashanth*

3:30 - 4:00 pm  Smart Stores: A scalable foot trace collection and prediction system  
*Soheila Abrishami, Piyush Kumar, and Wickus Nienaber*

4:00 - 4:15 pm  Coffee Break

**Fourth Session**  
*Session Chair: Yunlong Wang*

4:15 - 4:45 pm  Multivariate Time Series Representation and Similarity Search Using PCA  
*Aminata Kane, Nematollaah Shiri*

4:45 - 5:15 pm  Message Passing on Factor Graph: A Novel Approach for Orphan Drug Physician Targeting  
*Yunlong Wang, Yong Cai*

5:15 - 5:45 pm  Fast GPU-based Influence Maximization within Finite Deadlines via Node-level Parallelism  
*Koushik Pal, Zisis Poulos, Edward Kim, and Andreas Veneris*

5:45 pm  End of First Day

6:00 - 9:00 pm  Banquet
July 13th, 2017

**Invited Talk**  
*Session Chair:* Petra Perner  

9:00 - 10:00 am  
**Intelligent Data Processing in Detecting Landmines to Save Human Lives**  
*Prof. Lakhmi Jain*  
*University of Canberra, Australia*

10:00 - 10:15 am  
**Coffee Break**

**Fifth Session**  
*Session Chair:* Stefano Ferilli

10:15 - 10:45 am  
**Visual Scenes Mining for Agent Awareness Module**  
*Gang Ma, Zhentao Tang, Xi Yang, Zhongzhi Shi, Kun Yang*

10:45 - 11:15 am  
**Predicting Hospital Re-admissions from Nursing Care Data of Hospitalized Patients**  
*Muhammad K Lodhi, Rashid Ansari, Yingwei Yao, Gail M Keenan, Diana Wilkie, Ashfaq A Khokhar*

11:15 - 11:45 am  
**Activity Prediction in Process Management using the WoMan Framework**  
*Stefano Ferilli, Domenico Redavid, and Sergio Angelastro*

11:45 - 12:15 pm  
**Hierarchical Text Classification of Autopsy Reports to Determine MoD and CoD through Term-Based and Concepts-Based Features**  
*Ghulam Mujtaba, Liyana Shuib, Ram Gopal Raj, Mohammed Ali Al-Garadi, Retnagowri Rajandram, Khairunisa Shaikh*

12:15 - 1:45 pm  
**Lunch Break**
Sixth Session
Session Chair: Miad Faezipour

1:45 - 2:15 pm Collaborative Filtering Fusing Label Features Based on SDAE
Huan Huo, Xiufeng Liu, Deyuan Zheng, Zonghan Wu, Shengwei Yu, Liang Liu

2:15 - 2:45 pm Interestingness Classification of Association Rules for Master Data
Wei HAN, Julio BORGES, Peter NEUMAYER, Yong DING, Till RIEDEL, Michael BEIGL

2:45 - 3:15 pm MapReduce and Spark-based Analytic Framework Using Social Media Data for Earlier Flu Outbreak Detection
Al Essa, Ali, Faezipour, Miad

3:15 - 3:30 pm Coffee Break

Seventh Session
Session Chair: Francis Tay

3:30 - 4:00 pm An Integrated Approach using Data Mining and System Dynamics to Policy Design: Effects of Electric Vehicle Adoption on CO2 Emissions in Singapore
Bohao Zhang, Francis E. H. Tay

4:00 - 4:30 pm Using the Results of Capstone Analysis to Predict a Weather Outcome
Anthony G Nolan, Warwick J. Graco

5:00 - 5:30 pm Classification of Network Traffic Using Fuzzy Clustering for Network Security 277
Terrence P. Fries

5:30 - 6:00 pm Machine Learning and Pattern Recognition Techniques for Information Extraction to Improve Production Control and Design Decisions
Carlos Alberto Escobar Diaz, Ruben Morales-Menendez
6:00 - 6:15 pm  Break

**Eighth Session**  
*Session Chair: Petra Perner*

6:15 - 6:45 pm  Real-time Prediction of Styrene Production Volume based on Machine Learning Algorithms  
*Yikai Wu, Fang Hou, Xiaopei Cheng*

6:45 - 7:15 pm  A graph-based ranking model for automatic keyphrases extraction from Arabic documents  
*Mohamed Salim EL BAZZI, Driss MAMMASS, Taher ZAKI, Abdelatif ENNAJI*

7:15 - 7:45 pm  Mining Frequent Subgraph Pattern over a Collection of Attributed-Graphs and Construction of a Relation Hierarchy for Result Reporting  
*Petra Perner*

7:45 pm  **End of Conference**

Fair Well
ICDM 2017 – Tutorials

July 8th, 2017
9:00 am – 5:30 pm  Standardization in Immunofluorescence
Prof. Dr. Petra Perner

July 11th, 2017
9:00 am – 4:00 pm  Intelligent Image Interpretation and Computer Vision in Medicine, Biotechnology, Chemistry & Food Industry
Prof. Dr. Petra Perner

July 15th, 2017
9:00 am – 4:00 pm  Case-Based Reasoning
Prof. Dr. Petra Perner

July 16th, 2017
9:00 am – 4:00 pm  Data Mining
Prof. Dr. Petra Perner
Dear Author,

to ensure that the proceedings and a particular paper from the proceedings is properly cited, we would highly recommend that you follow the citation style given below.
If you do not properly follow the citation style given below you might risk that your paper or any of the other papers is not properly recognized by the citation index.
Thank you for taking care of this important issue.

Kind regards,  
Prof. Dr. Petra Perner  
Chair of the Conference

Recommended Citation Style of the Proceedings and the Papers


A particular paper:

Example for Poster Proceedings:

Example for Workshop Proceedings:

Example for Journal:
A Note on Citation Policy

As you might know, the value of a publication, a conference or a workshop is judged based on the citation rate. American publications and conferences are highly ranked based on this statistical measure. American publishers make sure that people accept this policy and publish papers that cite the respective publications published in the journal or at the conference where the researchers submit their paper.

If you want your publication to be considered a valuable one, you should follow this policy when you publish at the conferences ICDM, MLDM or MDA. The papers of ICDM, MLDM and MDA are published in the Springer LNCS or LNAI Series and by ibai-publishing Prof. Dr. Petra Perner and most of the libraries worldwide have a copy of these proceedings. Also, the library index at http://www.informatik.uni-trier.de/~ley/db/carefully lists all publications that have been published at these conferences.

All OPEN ACCESS journals of ibai-publishing are also listed there:

- **International Journal Transactions on Machine Learning and Data Mining**
  [http://www.informatik.uni-trier.de/~ley/db/journals/mldm/index.html]

- **International Journal Transactions on Case-Based Reasoning**
  [http://www.informatik.uni-trier.de/~ley/db/journals/cbrmd/index.html]

- **International Journal Transactions on Mass Data Analysis of Signals and Images**
  [http://www.informatik.uni-trier.de/~ley/db/journals/mda/index.html].

Any researcher can access this index for free. It shows title, authors and their affiliation, and abstract. It provides a researcher with a fast and complete overview of all publications presented at a conference or a workshop or published in a journal.

So please, before you submit a paper, check the index for relevant publications. If you have only cited publications from outside the main stream of ICDM, MLDM or MDA or any of the journals, the question will be raised why you then did not publish your paper outside the main stream.

If the reviewer suggests to cite a publication, please do so! It will make your publication more valuable.
Publication of Papers as Journal Papers and OPEN ACCESS Papers

For journal publication, please consider ibai publishing’s OPEN ACCESS journals (www.ibai-publishing.org). The paper in the journals will be provided as hardcopy journal to the authors and as pdf-files on the web.

There will be a special issue for this conference in the Transactions on Machine Learning and Data Mining (Hardcopy ISSN: 1865-6781, Online ISSN: 2509-9337)
http://www.ibai-publishing.org/journal/mldm/about.php
The impact factor of the journal was 0.612 in 2015.

There will be a special issue for this conference in the Transactions on Case-Based Reasoning (Hardcopy ISSN:1867-366X, Online E-ISSN: 2509-9345). http://www.ibai-publishing.org/journal/cbr/about.php
The impact factor of the journal was 0.574 in 2015.
Note
Note
Note
Note