Lecture Notes in Artificial Intelligence 9166

Subseries of Lecture Notes in Computer Science

LNAI Series Editors

Randy Goebel
  University of Alberta, Edmonton, Canada

Yuzuru Tanaka
  Hokkaido University, Sapporo, Japan

Wolfgang Wahlster
  DFKI and Saarland University, Saarbrücken, Germany

LNAI Founding Series Editor

Joerg Siekmann
  DFKI and Saarland University, Saarbrücken, Germany
More information about this series at http://www.springer.com/series/1244
Machine Learning and Data Mining in Pattern Recognition

11th International Conference, MLDM 2015
Hamburg, Germany, July 20–21, 2015
Proceedings

Springer
Preface

The tenth event of the International Conference on Machine Learning and Data Mining MLDM 2015 was held in Hamburg (www.mldm.de) running under the umbrella of the World Congress “The Frontiers in Intelligent Data and Signal Analysis, DSA2015” (www.worldcongressdsa.com).

For this edition the Program Committee received 123 submissions. After the peer-review process, we accepted 41 high-quality papers for oral presentation; from these, 40 are included in this proceedings volume. The topics range from theoretical topics for classification, clustering, association rule and pattern mining to specific data-mining methods for the different multimedia data types such as image mining, text mining, video mining, and Web mining. 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, a tutorial on Standardization in Immunofluorescence, and a tutorial on Big Data and Text Analysis were held before the conference.

We were pleased to give out the best paper award for MLDM for the fourth time this year. There are four announcements mentioned at www.mldm.de. The final decision was made by the Best Paper Award Committee based on the presentation by the authors and the discussion with the auditorium. The ceremony took place during the banquet. This prize is sponsored by ibai solutions (www.ibai-solutions.de), one of the leading companies in data mining for marketing, Web mining, and e-commerce.

The conference was rounded up by an outlook of new challenging topics in machine learning and data mining before the Best Paper Award ceremony.

We would like to thank all reviewers for their highly professional work and their effort in reviewing the papers. We would also thank 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, 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. See you in 2016 in New York at the next World Congress (www.worldcongressdsa.com) on “The Frontiers in Intelligent Data and Signal Analysis, DSA2016,” which combines under its roof the following three events: International Conference on 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 2015

Petra Perner
Organization

Chair

Petra Perner  IBaI Leipzig, Germany

Program Committee

Sergey V. Ablameyko  Belarus State University, Belarus
Patrick Bouthemy  Inria-VISTA, France
Michelangelo Ceci  University of Bari, Italy
Xiaoqing Ding  Tsinghua University, China
Christoph F. Eick  University of Houston, USA
Ana Fred  Technical University of Lisbon, Portugal
Giorgio Giacinto  University of Cagliari, Italy
Makato Haraguchi  Hokkaido University Sapporo, Japan
Dimitrios A. Karras  Chalkis Institute of Technology, Greece
Adam Krzyzak  Concordia University, Montreal, Canada
Thang V. Pham  Intelligent Systems Lab Amsterdam (ISLA), The Netherlands
Gabriella Sanniti di Baja  CNR, Italy
Linda Shapiro  University of Washington, USA
Tamas Sziranyi  MTA-SZTAKI, Hungary
Alexander Ulanov  HP Labs, Russian Federation
Patrick Wang  Northeastern University, USA

Additional Reviewers

Jeril Kuriakose  Manipal University Jaipur, India
Goce Ristanoski  NICTA, Australia
Hamed Bolandi  Islamic Azad University (IAU), Iran
# Contents

## Graph Mining

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greedy Graph Edit Distance</td>
<td>3</td>
</tr>
<tr>
<td><em>Kaspar Riesen, Miquel Ferrer, Rolf Dornberger, and Horst Bunke</em></td>
<td></td>
</tr>
<tr>
<td>Learning Heuristics to Reduce the Overestimation of Bipartite Graph</td>
<td></td>
</tr>
<tr>
<td>Edit Distance Approximation</td>
<td>17</td>
</tr>
<tr>
<td><em>Miquel Ferrer, Francesc Serratosa, and Kaspar Riesen</em></td>
<td></td>
</tr>
<tr>
<td>Seizure Prediction by Graph Mining, Transfer Learning, and Transformation Learning</td>
<td>32</td>
</tr>
<tr>
<td><em>Nimit Dhulekar, Srinivas Nambirajan, Basak Oztan, and Bülent Yener</em></td>
<td></td>
</tr>
</tbody>
</table>

## Classification and Regression

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local and Global Genetic Fuzzy Pattern Classifiers</td>
<td>55</td>
</tr>
<tr>
<td><em>Søren Atmakuri Davidsen, E. Sreedevi, and M. Padmavathamma</em></td>
<td></td>
</tr>
<tr>
<td>IKLTSa: An Incremental Kernel LTSA Method</td>
<td>70</td>
</tr>
<tr>
<td><em>Chao Tan, Jihong Guan, and Shuigeng Zhou</em></td>
<td></td>
</tr>
</tbody>
</table>

## Sentiment Analysis

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>SentiSAIL: Sentiment Analysis in English, German and Russian</td>
<td>87</td>
</tr>
<tr>
<td><em>Gayane Shalunts and Gerhard Backfried</em></td>
<td></td>
</tr>
<tr>
<td>Sentiment Analysis for Government: An Optimized Approach</td>
<td>98</td>
</tr>
<tr>
<td><em>Angelo Corallo, Laura Fortunato, Marco Matera, Marco Alessi, Alessio Camillò, Valentina Chetta, Enza Giangreco, and Davide Storelli</em></td>
<td></td>
</tr>
</tbody>
</table>

## Data Preparation and Missing Values

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Novel Algorithm for the Integration of the Imputation of Missing Values and Clustering</td>
<td>115</td>
</tr>
<tr>
<td><em>Roni Ben Ishay and Maya Herman</em></td>
<td></td>
</tr>
<tr>
<td>Improving the Algorithm for Mapping of OWL to Relational Database Schema</td>
<td>130</td>
</tr>
<tr>
<td><em>Chien D.C. Ta and Tuoi Phan Thi</em></td>
<td></td>
</tr>
<tr>
<td>Robust Principal Component Analysis of Data with Missing Values</td>
<td>140</td>
</tr>
<tr>
<td><em>Tommi Kärkkäinen and Mirka Saarela</em></td>
<td></td>
</tr>
</tbody>
</table>
## Contents

### Association and Sequential Rule Mining

Efficient Mining of High-Utility Sequential Rules  
*Souleymane Zida, Philippe Fournier-Viger, Cheng-Wei Wu, Jerry Chun-Wei Lin, and Vincent S. Tseng*  
157

MOGACAR: A Method for Filtering Interesting Classification Association Rules  
*Diana Benavides Prado*  
172

### Support Vector Machines

Classifying Grasslands and Cultivated Pastures in the Brazilian Cerrado Using Support Vector Machines, Multilayer Perceptrons and Autoencoders  
*Wanderson Costa, Leila Fonseca, and Thales Körting*  
187

Hybrid Approach for Inductive Semi Supervised Learning Using Label Propagation and Support Vector Machine  
*Aruna Govada, Pravin Joshi, Sahil Mittal, and Sanjay K. Sahay*  
199

### Frequent Item Set Mining and Time Series Analysis

Optimizing the Data-Process Relationship for Fast Mining of Frequent Itemsets in MapReduce  
*Saber Salah, Reza Akbarinia, and Florent Masseglia*  
217

Aggregation-Aware Compression of Probabilistic Streaming Time Series  
*Reza Akbarinia and Florent Masseglia*  
232

### Clustering

Applying Clustering Analysis to Heterogeneous Data Using Similarity Matrix Fusion (SMF)  
*Aalaa Mojahed, Joao H. Bettencourt-Silva, Wenjia Wang, and Beatriz de la Iglesia*  
251

On Bicluster Aggregation and its Benefits for Enumerative Solutions  
*Saullo Oliveira, Rosana Veroneze, and Fernando J. Von Zuben*  
266

Semi-Supervised Stream Clustering Using Labeled Data Points  
*Kritsana Treechalong, Thanawin Rakthanmanon, and Kitsana Waiyamai*  
281

Avalanche: A Hierarchical, Divisive Clustering Algorithm  
*Paul K. Amalaman and Christoph F. Eick*  
296
Text Mining

Author Attribution of Email Messages Using Parse-Tree Features ............ 313
Jagadeesh Patchala, Raj Bhatnagar, and Sridharan Gopalakrishnan

Query Click and Text Similarity Graph for Query Suggestions ............. 328
D. Sejal, K.G. Shailesh, V. Tejaswi, Dinesh Anvekar, K.R. Venugopal,
S.S. Iyengar, and L.M. Patnaik

Offline Writer Identification in Tamil Using Bagged Classification Trees .... 342
Sudarshan Babu

Applications of Data Mining

Data Analysis for Courses Registration ........................................ 357
Nada Alzahrani, Rasha Alsulim, Nourah Alaseem, and Ghada Badr

Learning the Relationship Between Corporate Governance and Company
Performance Using Data Mining .............................................. 368
Darie Moldovan and Simona Mutu

A Bayesian Approach to Sparse Learning-to-Rank for Search Engine
Optimization ........................................................................ 382
Olga Krasotkina and Vadim Mottl

Data Driven Geometry for Learning .......................................... 395
Elizabeth P. Chou

Mining Educational Data to Predict Students’ Academic Performance ...... 403
Mona Al-Saleem, Norah Al-Kathiry, Sara Al-Osimi, and Ghada Badr

Patient-Specific Modeling of Medical Data .................................. 415
Guilherme Alberto Sousa Ribeiro, Alexandre Cesar Muniz de Oliveira,
Antonio Luiz S. Ferreira, Shyam Visweswaran, and Gregory F. Cooper

A Bayesian Approach to Sparse Cox Regression in High-Dimensional
Survival Analysis ................................................................ 425
Olga Krasotkina and Vadim Mottl

Data Mining in System Biology, Drug Discovery, and Medicine

Automatic Cell Tracking and Kinetic Feature Description of Cell Paths
for Image Mining ................................................................. 441
Petra Perner

Author Index ......................................................................... 453