

SIAM International Conference on Data Mining SDM 2013

Martin Trnečka

International Center for Information and Uncertainty
Palacky University, Olomouc



europa
european
social fund in the
czech republic



EUROPEAN UNION



MINISTRY OF EDUCATION,
YOUTH AND SPORTS



INVESTMENTS IN EDUCATION DEVELOPMENT

Basic Information

- Top data mining conference.
- Place: Austin, Texas, USA.
- Date: May 2-4, 2013.
- Sponsors: IBM, Google, NFS.

Conference

- Acceptance rate: 25.57%
- Number of submissions: 348
- Number of participants: 704
- Number of presentations: 89
- Number of invited talks: 4
- Number of parallel section: 4
- Number of tutorials: 6
- Number of application papers: 50%
- Number of countries: 26 (most of papers is from USA and China)

Selected Invited Talks

- Regional Climate Informatics: A Statistical Perspective
Doug Nychka
Statistical model for climate predictions.
- Social Networks As Information Filters
Lada Adamic
Spreading and evolving of information in social networks.
- Modeling Individual-Level Data in the 21st Century
Padhraic Smyth
Role of a individual data and their meaning.

Selected Presentations

- Missing Or Inapplicable: Treatment of Incomplete Continuous Valued Features in Supervised Learning
Pang-Ning Tan, et. al.
This talk introduce new approach to missing (not applicable) values in data.
- Discriminative Feature Selection for Uncertain Graph Classification
Xiangnan Kong, et. al.
In this talk was present so-called uncertain graphs. This kind of graphs was use for analysis connections in normally working a damaged brains.
- Patient Risk Prediction Model via Top-k Stability Selection
Jiayu Zhou, et. al.
New approach to selecting stable descriptors for predictions.

Feature Selection

- Good feature.
- USPS Digit identification - graphical dataset.
- In general: decompose input matrix (ordinal) with dimension $n \times m$ to three matrices $n \times k$, $k \times k$ and $k \times m$, where k represent number of features.
- Error is represent as another matrix.
- Direct application to dimensionality reduction (for example for reduction of numbers of question in Questionary).
- It can be implemented via another decomposition method (SVD, NMF).

Probabilistic Matrix Factorization (PMF)

- Analogy to BMF method.
- In each step is selected “random cluster” represented via random set of objects and random set of attributes.
- Interesting results on random data.
- Not applicable for real datasets.
- Authors propose application in the field of recommendation.

Selected Tutorials

- Big Data Analytics for Healthcare
Jimeng Sun, Chandan Reddy
How can the datamining help improvement healthcare.
- Recent Advances in Applied Matrix Technologies
Fei Wang, Hanghang Tong
Matrix calculus and its application in the datamining.

Others

- Poster section.
- Social network analysis.
- Healthcare.
- Weather forecast.