

# PAKDD-2011 (Conference, Workshops, Tutorials) Floor Plan of Function Room



**May 24, 2011, Pre-Conference Program ( Workshops & Tutorials )**

**Registration ( 8:00 -- 9:00 )**

*For detail schedule, please refer to each workshop program*

Workshop 1: QIMIE Venue: Madrid 2	Workshop 2: BDM Venue: Madrid 3	Workshop 3: BI Venue : Madrid 5	Phd. Symposium: DSDM Venue : Madrid 6	Workshop 4: AI-TCM Venue : Madrid 7	Workshop 5: DMHM Venue : Madrid 8	Tutorial : Amazon Talk Venue : Madrid VIP (马德里贵宾厅)
--------------------------------------	------------------------------------	------------------------------------	--	--	--------------------------------------	--

**Coffee Break ( 10:30 --10:50 )**

Workshop 1: QIMIE Venue: Madrid 2	Workshop 2: BDM Venue: Madrid 3	Workshop 3: BI Venue: Madrid 5	Phd. Symposium: DSDM Venue : Madrid 6	Workshop 4: AI-TCM Venue : Madrid 7	Workshop 5: DMHM Venue: Madrid 8	Tutorial : Amazon Talk Venue: Madrid VIP (马德里贵宾厅)
--------------------------------------	------------------------------------	-----------------------------------	--	--	-------------------------------------	---

**Buffet Lunch ( 12:00 -- 14:00 )**

Workshop 1: QIMIE Venue: Madrid 2	Workshop 2: BDM Venue: Madrid 3	Workshop 3: BI Venue: Madrid 5	Phd. Symposium: DSDM Venue : Madrid 8	Tutorial 4: Data Stream Mining: Challenges and Techniques Venue: Madrid VIP (马德里贵宾厅) Time: 13:30 -- 15:00		
--------------------------------------	------------------------------------	-----------------------------------	--	--	--	--

**Coffee Break (15:40 -- 16:00)**

Workshop 1: QIMIE Venue: Madrid 2	Workshop 2: BDM Venue: Madrid 3	Workshop 3: BI Venue: Madrid 5	Phd. Symposium: DSDM Venue : Madrid 8	Tutorial 4: Data Stream Mining: Challenges and Techniques Venue: Madrid VIP (马德里贵宾厅) Time: 15:15 --17:00		
--------------------------------------	------------------------------------	-----------------------------------	--	---	--	--

**Closing**

**May 25, 2011 (Day 1)**

8:30 -- 9:00	<p align="center"><b>Opening:</b> Welcome Speech Venue: Espana 2</p>		
9:00 -- 10:00	<p align="center"><b>Keynote Speech I</b> Title: Hadoop: Industry View and Future Directions Speaker: Dr. Amr Awadallah, Co-Founder and CTO of Cloudera Venue: Espana 2 Chair: Prof. David Cheung, The University of Hong Kong</p>		
<p align="center"><b>Coffee Break ( Time: 10:00 -- 10:20, Venue: Foyer )</b></p>			
10:20 -- 12:20	<p align="center">Session 1A: Clustering 1 Venue: Madrid 2</p>	<p align="center">Session 1B: Sequence Mining Venue: Madrid 3</p>	<p align="center">Session 1C: Graph Mining Venue: Madrid 5</p>
<p align="center"><b>Lunch ( Time: 12:20 -- 14:00, Venue: Mercado )</b></p>			
14:00 -- 15:40	<p align="center">Session 2A: Classification Venue: Madrid 2</p>	<p align="center">Session 2B: Text Mining 1 Venue: Madrid 3</p>	<p align="center">Session 2C: Pattern Mining 1 Venue: Madrid 5</p>
<p align="center"><b>Coffee Break ( Time: 15:40 -- 16:00, Venue: Foyer )</b></p>			
16:00 -- 17:30	<p align="center">Session 3A: Social Networks Venue: Madrid 2</p>	<p align="center">Session 3B: Outlier Detection Venue: Madrid 3</p>	<p align="center">Session 3C: BGI Forum Venue: Madrid 5</p>
18:30 -- 20:30	<p align="center">BGI Reception Venue: Poolside</p>		

**May 26, 2011 (Day 2)**

**9:00--10:00**

**Keynote Speech II**

Title: Supernetworks: Opportunities and Challenges for Decision-Making in the 21<sup>st</sup> Century

Speaker: Prof. Anna Nagurney, University of Massachusetts Amherst, USA

Venue: Espana 2

Chair: Prof. Joshua Zhexue Huang, Shenzhen Institutes of Advanced Technology, CAS, China

**Coffee Break ( Time: 10:00 -- 10:20, Venue: Foyer )**

**10:20--12:20**

Session 4A:  
Machine Learning 1  
Venue: Madrid 2

Session 4B:  
Text Mining 2  
Venue: Madrid 3

Session 4C:  
Feature Extraction  
Venue: Madrid 5

Session 4D:  
Evaluation  
Venue: Madrid 6

Session 4E:  
Time Series and Feature  
Selection Venue: Madrid 7

**Lunch ( Time: 12:20 -- 14:00, Venue: Mercado )**

**14:00--17:30**

Excursion

**18:30--21:30**

Banquet  
Venue: Espana 2

**May 27, 2011 (Day 3)**

<p align="center"><b>9:00 -- 10:00</b></p>	<p align="center"><b>Keynote Speech III</b>                  Title: A Learning Framework for Data Objects with Complex Semantics                  Speaker: Prof. Zhi-Hua Zhou, Nanjing University, China                  Venue: Espana 2                  Chair: Prof. Longbing Cao, University of Technology Sydney, Australia</p>				
<p align="center"><b>Coffee Break ( Time:10:00 -- 10:20, Venue: Foyer)</b></p>					
<p align="center"><b>10:20 -- 12:20</b></p>	<p align="center">Session 5A: Machine Learning 2 Venue: Madrid 2</p>	<p align="center">Session 5B : Applications Venue: Madrid 3</p>	<p align="center">Session 5C: Imbalanced Data Venue: Madrid 5</p>	<p align="center">Session 5D: Text Mining 3 Venue: Madrid 6</p>	<p align="center">Session 5E: Evaluation 2 Venue: Madrid 7</p>
<p align="center"><b>Lunch ( Time: 12:20 -- 13:50, Venue: Mercado)</b></p>					
<p align="center"><b>14:00 -- 15:40</b></p>	<p align="center">Tutorial 1: Handling Concept Drift: Importance, Challenges and Solutions Venue: Madrid 2</p>	<p align="center">Tutorial 2: Privacy-Preservation of Graphs and Social Networks Venue: Madrid 3</p>	<p align="center">Session 6C: Pattern Mining 2 Venue: Madrid 5</p>	<p align="center">Tutorial 3: Data Mining for Healthcare Management Venue: Madrid 6</p>	<p align="center">Tutorial 5: Visualization and Browsing of Large Image Databases Venue: Madrid 7</p>
<p align="center"><b>Coffee Break ( Time:15:40 -- 16:00, Venue: Foyer)</b></p>					
<p align="center"><b>16:00 -- 17:30</b></p>	<p align="center">Tutorial 1: Handling Concept Drift: Importance, Challenges and Solutions Venue: Madrid 2</p>	<p align="center">Tutorial 2: Privacy-Preservation of Graphs and Social Networks Venue: Madrid 3</p>	<p align="center">Session 7C: Clustering 2 Venue: Madrid 5</p>	<p align="center">Tutorial 3: Data Mining for Healthcare Management Venue: Madrid 6</p>	<p align="center">Tutorial 6: Social Image Tags: From Motivation To Mining Venue: Madrid 7</p>

## PAKDD-2011 Conference

### SESSION 1A: Clustering 1 ( 120 mins )

**Session Chair:** Jie Tang

**Time:** 10:20 – 12:20

**Venue:** Madrid 2

1. Liping Jing, Jiali Yun, Jian Yu and Joshua Huang. High-Order Co-Clustering Text data on Semantics-based Representation Model **(Regular)**
2. Nenad Tomasev, Milos Radovanovic, Dunja Mladenic and Mirjana Ivanovic. The Role of Hubness in Clustering High-Dimensional Data **(Regular)**
3. Baijie Wang and Xin Wang. Spatial Entropy-based Clustering for Mining Data with Spatial Correlation **(Short)**
4. Guangzhi Qu. Self-Adjust Local Connectivity Analysis for Spectral Clustering **(Short)**
5. Saurav Sarmah, Rosy Das (Sarmah) and Dhruva Bhattacharyya. An Effective Density-based Hierarchical Clustering Technique to identify Coherent Patterns from Gene Expression Data **(Short)**

### SESSION 1B: Sequence Mining ( 120 mins )

**Session Chair:** Latifur Khan

**Time:** 10:20 – 12:20

**Venue:** Madrid 3

1. Yasuhiro Urabe, Kenji Yamanishi, Ryota Tomioka and Hiroki Iwai. Real-time Change-point Detection Using Sequentially Discounting Normalized Maximum Likelihood Coding **(Regular)**
2. Xiang-yun Qing and Xing-yu Wang. A Sequential Dynamic Multi-class Model and Recursive Filtering by Variational Bayesian Methods **(Regular)**
3. Muhammad Muzammal and Rajeev Raman. Mining Sequential Patterns from Probabilistic Databases **(Short)**
4. Xu Sun, Hisashi Kashima and Naonori Ueda. Large Scale Real-life Action Recognition Using Conditional Random Fields and a New Training Method **(Short)**
5. Atsuyoshi Nakamura and Mineichi Kudo. Packing Alignment: Alignment for Sequences of Various Length Events **(Short)**

### SESSION 1C: Graph Mining ( 120 mins )

**Session Chair:** Murat Kantarcioglu

**Time:** 10:20 – 12:20

**Venue:** Madrid 5

1. Leting Wu, Xiaowei Ying, Xintao Wu, Aidong Lu and Zhi-Hua Zhou. Spectral Analysis of k-balanced Signed Graphs **(Regular)**
2. U Kang, Brendan Meeder and Christos Faloutsos. Spectral Analysis for Billion-Scale Graphs: Discoveries and Implementation **(Regular)**
3. Yasuo Tabei, Daisuke Okanohara, Shuichi Hirose, and Koji Tsuda. LGM: Mining Frequent Subgraph from Linear Graphs **(Short)**
4. Rajul Anand and Chandan K. Reddy. Graph-based Clustering with Constraints **(Short)**
5. Di Jin, Dayou Liu, Bo Yang, Jie Liu and Dongxiao He. Ant Colony Optimization with Markov Random Walk for Clustering in Complex Networks **(Short)**

**SESSION 2A:** Classification ( 100 mins )

**Session Chair:** Liping Jing

**Time:** 14:00 – 15:40

**Venue:** Madrid 2

1. Indre Zliobaite. Identifying Hidden Contexts in Classification **(Regular)**
2. Junfeng Pan, Gui-Rong Xue and Yong Yu. Cross-Lingual Sentiment Classification via Bi-view Non-negative Matrix Tri-Factorization **(Regular)**
3. Xiaojun Wan. Collaborative Data Cleaning for Sentiment Classification with Noisy Training Corpus **(Short)**
4. Maya Wardeh, Frans Coenen and Trevor Bench-Capon. Multiagent Based Classification Using Argumentation From Experience **(Short)**

**SESSION 2B:** Text Mining 1 ( 100 mins )

**Session Chair:** Dat Tran

**Time:** 14:00 – 15:40

**Venue:** Madrid 3

1. Weidong Yang, Hao Zhu, Nan Li and Guansheng Zhu. Adaptive and Effective Keyword Search for XML **(Regular)**
2. Sangeetha Kutty, Richi Nayak and Yuefeng Li. XML Documents Clustering using Tensor Space Model **(Regular)**
3. Ying Liu. An Efficient Pre-processing Method to Identify Logical Components from PDF Documents **(Short)**
4. Xianling Mao, Xiaobing Liu, Xiaoming Li, Hongfei Yan, Jing He, Nan Di and Dongdong Shan. SizeSpotSigs: an Effective Deduplicate Algorithm Considering the Size of Page Content **(Short)**
5. Jie Tang, Wim De Smet and Marie-Francine Moens. Knowledge Transfer across Multilingual Corpora via Latent Topics **(Short)**

**SESSION 2C:** Pattern Mining 1 ( 100 mins )

**Session Chair:** Junbin Gao

**Time:** 14:00 – 15:40

**Venue:** Madrid 5

1. Michael Steinbach, Haoyu Yu, Gang Fang and Vipin Kumar. Using Constraints to Generate and Explore Higher Order Discriminative Patterns **(Regular)**
2. Bo Wang, Zhaonan Li, Jie Tang, Kuo Zhang and Songcan Chen. Learning to Advertising-How many ads are enough? **(Regular)**
3. Jin Soung Yoo and Mark Bow. Mining Maximal Co-location Patterns **(Regular)**
4. Yun Sing Koh, Russel Pears and Gill Dobbie. Automatic Assignment of Item Weights for Pattern Mining on Data Streams **(Short)**

**SESSION 3A:** Social Networks ( 90 mins )

**Session Chair:** Ee Peng Lim

**Time:** 16:00 – 17:30

**Venue:** Madrid 2

1. Chao Li, Zhongying Zhao, Jun Luo and Jianping Fan. Info-Cluster Based Regional Influence Analysis in Social Networks (**Short**)
2. Yasuhiro Fujiwara, Makoto Onizuka and Masaru Kitsuregawa. Efficient Centrality Monitoring for Time-evolving Graphs (**Short**)
3. Richi Nayak. Utilizing Past Relations and User Similarities in a Social Matching System (**Short**)
4. Jhao-Yin Li and Mi-Yen Yeh. On Sampling Type Distribution from Heterogeneous Social Networks (**Short**)
5. Pei Li, Jeffrey Xu Yu, Hongyan Liu, Jun He and Xiaoyong Du. Ranking Individuals and Groups by Influence Propagation (**Short**)

**SESSION 3B:** Outlier Detection (90 mins)

**Session Chair:** Ho Tu Bao

**Time:** 16:00 – 17:30

**Venue:** Madrid 3

1. Dan He. Learning the Funding Momentum of Research Projects (**Regular**)
2. Hao Huang, Qinming He, Jiangfeng He and Lianhang Ma. RADAR: Rare Category Detection via Computation of Boundary Degree (**Short**)
3. Jun Gao, Weiming Hu, Zhongfei Zhang, Xiaoqin Zhang and Ou Wu. RKOF: Robust Kernel-Based Local Outlier Detection (**Short**)
4. Flora Tsai. Retrieval and Novelty Mining of Chinese Sentences (**Short**)
5. Timothy Hospedales, Tao Xiang and Shaogang Gong. Finding Rare Classes: Adapting Generative and Discriminative Models in Active Learning (**Short**)

**SESSION 3C:** BGI Forum: An Universal Platform for Trans-omics----The Future of Biology (90 mins)

**Session Chair:** Yong Zhang

**Time:** 16:00 – 17:30

**Venue:** Madrid 5

1. Yong Zhang. Introduction of trans-omics
2. Tin-Lap Lee. An integrated platform on genomics analysis
3. Siuming Yiu. Indexing in alignment optimization
4. Defa Xu. Supercomputer in bioinformatics applications

**SESSION 4A:** Machine Learning 1 ( 120 mins )

**Session Chair:** Xiaoyuan Su

**Time:** 10:20 – 12:20

**Venue:** Madrid 2

1. Daisuke Kimura, Tetsuji Kuboyama, Tetsuo Shibuya and Hisashi Kashima. A subpath kernel for rooted unordered trees **(Regular)**
2. Victor Cheng and C.H. Li. Classification Probabilistic PCA with Application in Domain Adaptation **(Regular)**
3. Bin Wang, Harry Zhang, Bruce Spencer and Yuanyuan Guo. The Unsymmetrical-Style Co-training **(Short)**
4. Jianxin Wu. Balance Support Vector Machines Locally Using the Structural Similarity Kernel **(Short)**
5. Xiaoyuan Su, Russell Greiner, Taghi M. Khoshgofttar and Amri Napolitano. Using Classifier-Based Nominal Imputation to Improve Machine Learning **(Short)**

**SESSION 4B:** Text Mining 2 ( 120 mins )

**Session Chair:** Richi Nayak

**Time:** 10:20 – 12:20

**Venue:** Madrid 3

1. Zhongwu Zhai, Bing Liu, Hua Xu and Peifa Jia. Constrained LDA for Grouping Product Features in Opinion Mining **(Regular)**
2. Tian-Jie Zhan and Chun-Hung Li. Semantic Dependent Word Pairs Generative Model for Fine-grained Product Feature Mining **(Regular)**
3. Harish Bhat and Daniel Zaelit. Predicting Private Company Exits using Qualitative Data **(Short)**
4. Ying Huang, Bingquan HUANG and Tahar Kechadi. A Rule-based Method For Customer Churn Prediction in Telecommunication Services **(Short)**
5. Rathany Chan Sam, Huong Thanh Le, Thuy Thanh Nguyen and Thien Huu Nguyen. Combining Proper Name-Coreference with CRF for Semi-supervised Named Entity Recognition in Vietnamese Text **(Short)**

**SESSION 4C:** Feature Extraction ( 120 mins )

**Session Chair:** Aixin Sun

**Time:** 10:20 – 12:20

**Venue:** Madrid 5

1. Shingo Takamatsu, Issei Sato and Hiroshi Nakagawa. Probabilistic Matrix Factorization Leveraging Contexts for Unsupervised Relation Extraction **(Regular)**
2. Bi-Ru Dai and Shu-Ming Hsu. An Instance Selection Algorithm Based on Reverse Nearest Neighbor **(Short)**
3. Dinesh Garg, Sundararajan Sellamanickam and Shirish Shevade. A Game Theoretic Approach for Feature Clustering and Its Application to Feature Selection **(Short)**
4. Gabriel Pui Cheong Fung, Fred Morstatter and Huan Liu. Feature Selection Strategy in Text Classification **(Short)**
5. Jiali Yun, Liping Jing, Jian Yu and Houkuan Huang. Unsupervised Feature Weighting based on Local Feature Relatedness **(Short)**

**SESSION 4D:** Evaluation 1 ( 120 mins )

**Session Chair:** Mohan Saravanan

**Time:** 10:20 – 12:20

**Venue:** Madrid 6

1. Tias Guns, Siegfried Nijssen and Luc De Raedt. Evaluating Pattern Set Mining Strategies in a Constraint Programming Framework **(Regular)**
2. Wei Luo and Marcus Gallagher. Faster and Parameter-Free Discord Search in Quasi-Periodic Time Series **(Regular)**
3. Jun Du and Charles Ling. Asking Generalized Queries with Minimum Cost **(Short)**
4. Claudio Sa, Carlos Soares, Alípio Jorge, Paulo Azevedo and Joaquim Costa. Mining Association Rules for Label Ranking **(Short)**

**SESSION 4E:** Time Series and Feature Selection ( 120 mins )

**Session Chair:** Philip S. Yu

**Time:** 10:20 – 12:20

**Venue:** Madrid 7

1. Krisztian Antal Buza, Alexandros Nanopoulos and Lars Schmidt-Thieme. INSIGHT: Efficient and Effective Instance Selection for Time Series Classification **(Regular)**
2. Harya Widiputra, Russel Pears and Nikola Kasabov. Multiple Time-series Prediction Through Multiple Time-series Relationships Profiling and Clustered Recurring Trends **(Short)**
3. Michal Lewandowski, Dimitrios Makris and Jean-Christophe Nebel. Probabilistic Feature Extraction from Multivariate Time Series using Spatio-Temporal Constraints **(Short)**
4. Yi Guo and Junbin Gao. Learning Image Manifold through Tensor Kernels based on Local Features **(Short)**

**SESSION 5A: Machine Learning 2 ( 120 mins )**

**Session Chair:** David Cheung

**Time:** 10:20 – 12:20

**Venue:** Madrid 2

1. Rohit Parimi and Doina Caragea. Predicting Friendship Links in Social Networks Using a Topic Modeling Approach **(Regular)**
2. Jing Yang. A Partial Correlation-Based Bayesian Network Structure Learning Algorithm under SEM **(Regular)**
3. Dijun Luo, chris ding and Heng Huang. On the global convergence of HOSVD and ParaFac algorithms **(Short)**
4. Sanparith Marukatat. Improved Spectral Hashing **(Short)**
5. Dayou Liu. An Adaptive Fuzzy k-Nearest Neighbor Method Based on Parallel Particle Swarm Optimization for Bankruptcy Prediction **(Short)**
6. Yan Zhou, Meador Inge and Murat Kantarcioglu. Compression for Anti-Adversarial Learning **(Short)**

**SESSION 5B: Applications ( 120 mins )**

**Session Chair:** Graham Williams

**Time:** 10:20 – 12:20

**Venue:** Madrid 3

1. Trung Le, Dat Tran, Wanli Ma and Dharmendra Sharma. Multiple Distribution Data Description Learning Algorithm for Novelty Detection **(Regular)**
2. Peipei Li, Xindong Wu, Xuegang Hu and Yuhong Zhang. Random Ensemble Decision Trees for Learning Concept-Drifting Data Streams **(Short)**
3. Ning Gao, Zhihong Deng, Hang Yu and Jia-Jian Jiang. ListOPT: A Learning to Optimize Method for XML Ranking **(Short)**

**SESSION 5C: Imbalanced Data ( 120 mins )**

**Session Chair:** Chengqi Zhang

**Time:** 10:20 – 12:20

**Venue:** Madrid 5

1. Xiannian Fan, Ke Tang and Thomas Weise. Margin-Based Over-Sampling Method for Learning From Imbalanced Datasets **(Regular)**
2. Yuxuan Li and Xiuzhen Zhang. Improving k Nearest Neighbor with Exemplar Generalization for Imbalanced Classification **(Regular)**
3. Pengyi Yang, Zili Zhang, Bing Zhou and Albert Zomaya. Sample subsets optimization for classifying imbalanced biological data **(Regular)**
4. Wei Liu and Sanjay Chawla. Class Confidence Weighted kNN Algorithms for Imbalanced Data Sets **(Regular)**

**SESSION 5D: Text Mining 3 (120 mins)**

**Session Chair:** Ming Li

**Time:** 10:20 – 12:20

**Venue:** Madrid 6

1. Tomonari Masada, Atsuhiko Takasu, Yuichiro Shibata and Kiyoshi Oguri. Steering Time-Dependent Estimation of Posteriors with Hyperparameter Indexing in Bayesian Topic Models **(Regular)**
2. Dat Huynh, Dat Tran, Wanli Ma and Dharmendra Sharma. Grammatical Dependency-Based Relations for Term Weighting in Text Classification **(Regular)**
3. Hiroshi Fujimoto, Minoru Etoh, Akira Kinno and Yoshikazu Akinaga. Topic Analysis of Web User Behavior using LDA Model on Proxy Logs **(Short)**
4. Xipeng Qiu, Jinlong Zhou and Xuanjing Huang. An Effective Feature Selection Method for Text Categorization **(Short)**

**SESSION 5E: Evaluation 2 (120 mins)**

**Session Chair:** Harish Bhat

**Time:** 10:20 – 12:20

**Venue:** Madrid 7

1. Ghita Berrada and Ander de Keijzer. A IFS-based Similarity Measure to Index Electroencephalograms **(Short)**
2. Aditya Desai, Himanshu Singh and Vikram Pudi. DISC: Data-Intensive Similarity Measure for Categorical Data **(Short)**
3. Marc Segond and Christian Borgelt. Item Set Mining Based on Cover Similarity **(Short)**
4. Yifeng Zeng. Dynamic Ordering-based Search Algorithm for Markov Blanket Discovery **(Short)**

**SESSION 6C: Pattern Mining 2 ( 100 mins )**

**Session Chair:** Qingshan Jiang

**Time:** 14:00 – 15:40

**Venue:** Madrid 5

1. Colin DeLong, Kendrick Erickson, Eric Perrino, Nishith Pathak, Kyong Shim and Jaideep Srivastava. TeamSkill: Modeling Team Chemistry in Online Multi-Player Games **(Regular)**
2. Sunil Kumar Gupta, Dinh Phung, Brett Adams and Svetha Venkatesh. A Bayesian Framework for Learning Shared and Individual Subspaces from Multiple Data Sources **(Short)**
3. Xujuan Zhou. Pattern Mining for a Two-stage Information Filtering System **(Short)**
4. Guangyan Huang, Yanchun Zhang, Jing He and Zhiming Ding. Efficiently Retrieving Longest Common Route Patterns of Moving Objects By Summarizing Turning Regions **(Short)**

**SESSION 7C: Clustering 2 ( 90 mins )**

**Session Chair:** Tao Chen

**Time:** 16:00 – 17:30

**Venue:** Madrid 5

1. Tengke Xiong, Shengrui Wang, André Mayers and Ernest Monga. Semi-Supervised Parameter-free Divisive Hierarchical Clustering of Categorical Data **(Short)**
2. Stephan Günnemann, Hardy Kremer, Charlotte Laufkötter and Thomas Seidl. Tracing Evolving Clusters by Subspace and Value Similarity **(Short)**
3. Chao Luo, Yanchang Zhao, Dan Luo and Chengqi Zhang. Agent-based Subspace Clustering **(Short)**
4. Yubin Zhan and Jianping Yin. Nonlinear Discriminative Embedding for Clustering via Spectral Regularization **(Short)**

**PAKDD-2011 Pre-Conference Program (Workshops & Tutorials)**  
**May 24, 2011**

**Workshop 1: Quality Issues, Measures of Interestingness and Evaluation of Data Mining Models (QIMIE)**

**Venue: Madrid 2**

9:00 -- 9:05

**Opening Welcome**

9:05 -- 9:55

**Invited Talk**

Speaker: *Prof. Longbing Cao*

Topic: *Knowledge Actionability: Evaluation and Practices.*

**Session 1: Clustering**

10:00 -- 10:25

- ClasSi: Measuring Ranking Quality in the Presence of Object Classes with Similarity Information.

*Anca Ivanescu, Marc Wichterich and Thomas Seidl*

10:25 -- 10:50

- The Instance Easiness of Supervised Learning for Cluster Validity.

*Vlad Estivill-Castro*

**Coffee Break ( 10:50 -- 11:05 )**

11:05 -- 11:30

- A new efficient and unbiased approach for clustering quality evaluation.

*Jean-Charles Lamirel, Pascal Cuxac and Raghvendra Mall*

**Session 2: Data Structure**

11:30 -- 11:55

- A Structure Preserving Flat Data Format Representation for Tree-Structured Data.

*Fedja Hadzic*

11:55 -- 12:20

- A fusion of algorithms in near duplicate document detection.

*Fan Jun and Huang Tiejun*

**Lunch ( 12:20 -- 13:20 )**

13:20 -- 14:10

**Invited Talk**

Speaker: *Prof. Nitesh Chawla*

Topic: *Evaluation Conundrum in Machine Learning/Data Mining.*

**Session 3: Patterns & Rules**

14:20 -- 14:45

- Searching Interesting Association Rules based on Evolutionary Computation.

*Guangfei Yang, Yanzhong Dang, Shingo Mabu, Kaoru Shimada and Kotaro Hirasawa*

14:45 -- 15:10

- An Efficient Approach to Mine Periodic-Frequent Patterns in Transactional Databases.

*Akshat Surana, Uday Kiran and Polepally Krishna Reddy*

**Coffee Break ( 15:10 -- 15:25 )**

15:25 -- 15:50

- Algorithms to Discover Complete Frequent Episodes In Sequences

*Wu Jianjun, Wan Li and Xu Zeren*

15:50 -- 16:15

- Certainty upon Empirical Distributions.

*Joan Garriga*

16:15 -- 16:40

**Closing**

**PAKDD-2011 Pre-Conference Program (Workshops & Tutorials)**  
**May 24, 2011**

**Workshop 2: Biologically Inspired Techniques for Data Mining (BDM)**

**Venue: Madrid 3**

9:00 -- 9:05	<b>Opening Welcome</b>
9:05 -- 9:40	<p><b>Invited Talk</b>            Speaker: <i>Shafiq Alam</i>            Topic: Introduction to Biologically inspired techniques for data mining</p>
9:40 -- 10:30	<p><b>Invited Talk</b>            Speaker: <i>Gillian Dobbie</i>            Topic: Swarm Intelligence and new challenges</p>
<b>Coffee Break ( 10:30 -- 10:45 )</b>	
10:45 -- 12:15	<ul style="list-style-type: none"> <li>• A Measure Oriented Training Scheme for Imbalanced Classification Problems <i>Bo Yuan and Wenhuan Liu</i></li> <li>• Training Radial Basis Function Network with the Honey Bee Mating Optimization <i>Ming-Huwi Horng, Yung-Nien Sun, and Yun-Xiang Lee</i></li> <li>• Weighted Association Rule Mining using Particle Swarm Optimization <i>Russel Pears and Yun Sing Koh</i></li> <li>• Employing Nearest Neighbor classifier and Genetic Algorithm for Churn Prediction in Telecom <i>Adnan Idris, Muhammad Rizwan and Asifullah Khan</i></li> </ul>
<b>Lunch ( 12:15 -- 13:30 )</b>	
13:30 – 14:15	<b>Panel discussion</b>
14:15 – 15:15	<ul style="list-style-type: none"> <li>• An SVM-based Approach to Discover MicroRNA Precursors in Plant Genomes <i>Yi Wang, Cheqing Jin, Minqi Zhou and Aoying Zhou</i></li> <li>• Towards Recommender System Using Particle Swarm Optimization Based Clustering <i>Shafiq Alam, Gillian Dobbie, Patricia Riddle</i></li> </ul>
<b>Coffee Break ( 15:15 – 15:30 )</b>	
15:30 – 16:20	<ul style="list-style-type: none"> <li>• An Unsupervised Feature Selection Framework Based on Clustering <i>Sheng-yi Jiang , Lian-xi Wang</i></li> </ul>
16:20 – 16:30	<b>Closing</b>

**PAKDD-2011 Pre-Conference Program (Workshops & Tutorials)**  
**May 24, 2011**

**Workshop 3: Behavior Informatics (BI)**

**Venue: Madrid 5**

8:25 -- 8:30      ***Opening Welcome***

8:30 -- 9:20      ***Invited Talk***  
 Speaker: *Prof. Jeffery Xu Yu* (The Chinese University of Hong Kong)  
 Topic: *Graph Behaviors: Discovering Patterns When a Graph Evolves*

**Session 1: Social Network Mining**

9:30 -- 9:50      • Evaluating the Regularity of Human Behavior from Mobile Phone Usage Logs  
*Hyoungnyoung Kim and Ji-Hyung Park*

9:50 -- 10:10    • Explicit and Implicit User Preferences in Online Dating  
*Joshua Akehurst, Irena Koprinska, Kalina Yacef, Luiz Augusto Pizzato, Judy Kay and Tomasz Rej*

10:10 -- 10:30   • Simultaneously Modeling Reply Networks and Contents to Generate User's Profiles on Web Forum  
*Zhao Zhang, Weining Qian and Aoying Zhou*

**Coffee Break ( 10:30 --10:45 )**

10:45 -- 11:15   • Blogger-Link-Topic Model for Blog Mining  
*Flora Tsai*

11:15 -- 11:35   • A Random Indexing Approach for Web User Clustering and Web Prefetching  
*Miao Wan and Arne Jonsson*

11:35 -- 11:55   • Emotional Reactions to Real-World Events in Social Networks  
*Thin Nguyen, Dinh Phung, Brett Adams and Svetha Venkatesh*

11:55 -- 12:15   • Constructing Personal Knowledge Base: Automatic Key-phrase Extraction from Multiple-domain Web Pages  
*Yin-Fu Huang and Cin-Siang Ciou*

**Lunch ( 12:15 -- 13:00 )**

13:00 -- 13:50   ***Invited Talk***  
 Speaker: *Dr Jie Tang* (Tsinghua University)  
 Topics: *Social Prediction: Can We Predict Users' Action and Emotions?*

**Session 2: Behavior Mining**

14:00 -- 14:20   • GroupBuyer: A Personalized Group Buying Event Recommender System Using Social Information Filtering  
*Yun-Hui Hung, Jen-Wei Huang and Ming-Syan Chen*

14:20 -- 14:40	<ul style="list-style-type: none"> <li>Discovering Valuable User Behavior Patterns in Mobile Commerce Environments <i>Bai-En Shie, Hui-Fang Hsiao, Philip S. Yu and Vincent S. Tseng</i></li> </ul>
14:40 -- 15:00	<ul style="list-style-type: none"> <li>A novel method for community detection in complex network using new representation for communities <i>Yiwen Wang and Min Yao</i></li> </ul>
15:00 -- 15:20	<ul style="list-style-type: none"> <li>Link Prediction on Evolving Data using Tensor Factorization <i>Stephan Spiegel, Jérôme Kunegis, Jan Hendrik Clausen and Sahin Albayrak</i></li> </ul>
15:20 -- 15:40	<ul style="list-style-type: none"> <li>Permutation Anonymization: Improving Anatomy for Privacy Preservation in Data Publication <i>Xianmang He, Yanghua Xiao, Yujia Li, Wei Wang and Bole Shi</i></li> </ul>
<b>Coffee Break ( 15:40 --15:55 )</b>	
<b>Session 3: Pattern Discovery</b>	
15:55 -- 16:15	<ul style="list-style-type: none"> <li>Mining Causality from Non-categorical Numerical Data <i>Joaquim Silva, José Lopes and António Falcão</i></li> </ul>
16:15 -- 16:35	<ul style="list-style-type: none"> <li>Efficient mining Top-k Regular-frequent itemset using Compressed Tidsets <i>Komate Amphawan, Philippe Lenca and Athasit Surarerks</i></li> </ul>
16:35 -- 16:55	<ul style="list-style-type: none"> <li>A Fast Algorithm for Mining High Utility Itemsets <i>Show-Jane Yen, Chia Ching Chen and Yue-Shi Lee</i></li> </ul>
16:55 -- 17:15	<ul style="list-style-type: none"> <li>A Method of Similarity Measure and Visualization for Long Time Series Using Binary Patterns <i>Hailin Li, Chonghui Guo and Libin Yang</i></li> </ul>
<b>Session 4: Clustering Methods</b>	
17:15 -- 17:35	<ul style="list-style-type: none"> <li>A BIRCH-Based Clustering Method for Large Time Series Databases <i>Duong Tuan Anh and Vo Le Quy Nhon</i></li> </ul>
17:35 -- 17:55	<ul style="list-style-type: none"> <li>Visualizing Cluster Structures and Their Changes over Time by Two-Step Application of Self-Organizing Maps <i>Masahiro Ishikawa</i></li> </ul>
17:55 -- 18:15	<ul style="list-style-type: none"> <li>Analysis of Cluster Migrations using Self-Organizing Maps <i>Denny, Peter Christen and Graham Williams</i></li> </ul>
18:15 -- 18:35	<ul style="list-style-type: none"> <li>Clustering Fragmented Trajectories for Mining Movement Behaviors <i>Chih-Chieh Hung and Wen-Chih Peng</i></li> </ul>
18:35 -- 18:40	<b>Closing</b>

PAKDD-2011 Pre-Conference Program (Workshops & Tutorials)  
May 24, 2011

Doctoral Symposium on Data Mining ( DSDM )

Venue: Madrid 6 ( 9:00 – 11:00)  
Madrid 8 (13:30 – 16:50)

9:00 -- 9:05	<b>Opening Welcome</b>
9:05 -- 10:00	<b>Invited Talk</b> Speaker: <b>Dr. Jie Tang</b> Topic: Colorful Social Networks: Inferring Social Ties in Large-scale Networks
10:00 -- 11:00	<ul style="list-style-type: none"><li>• Age Estimation using Bayesian Process. <i>Yu Zhang</i></li><li>• Significant Node Identification in Social Networks <i>Chi-Yao Tseng and Ming-Syan Chen</i></li><li>• Mining Tourist Preferences with Twice-Learning <i>Chen Zhang and Jie Zhang</i></li></ul>
<b>Lunch ( 12:00 -- 13:30 )</b>	
13:30 -- 14:30	Invited talk Speaker: <b>Dr. Ping Luo</b> Topic: Cross-domain Classification: Generative vs. Discriminative Models
14:30 -- 15:30	<ul style="list-style-type: none"><li>• Improving Bagging Performance through Multi-Algorithm Ensembles <i>Kuo-Wei Hsu and Jaideep Srivastava</i></li><li>• Towards Cost-Sensitive Learning for Real-World Applications <i>Xu-Ying Liu and Zhi-Hua Zhou</i></li></ul>
<b>Coffee Break ( 15:30 – 15:45 )</b>	
15:45 -- 16:45	Panel Session
16:45 -- 16:50	<b>Closing</b>

PAKDD-2011 Pre-Conference Program (Workshops & Tutorials)  
May 24, 2011

**Workshop 4: Advances and Issues in Traditional Chinese Medicine Clinical Data Mining (AI-TCM)**

Venue: Madrid 7

9:00 -- 9:05

**Opening Welcome**

9:05 -- 10:00

**Keynote Speech**

*by Prof. Baoyan Liu*

*Vice President of China Academy of Chinese Medical Sciences*

**Session: Paper Presentation Session**

10:00-10:25

- Discovery of Regularities in the Use of Herbs in Traditional Chinese Medicine Prescriptions

*Nevin L. Zhang, Runsun Zhang and Tao Chen*

10:25-10:50

- COW: A Co-evolving Mimetic Wrapper for Herb-Herb Interaction Analysis in TCM Informatics

*Dion Detterer and Paul Kwan*

**Coffee Break ( 10: 50 – 11:05 )**

11:05 -- 11:30

- Selecting Appropriate Interestingness Measure to Evaluate the Correlation of Syndrome Elements and Symptoms

*Lei Zhang, Qi-Ming Zhang, Yi-Guo Wang and Dong-Lin Yu*

11:30 -- 11:55

- The Impact of Feature Representation to the Biclustering of Symptoms-Herbs in TCM

*Simon K. Poon, Roger Luo and Runshun Zhang*

11:55 -- 12:40

**Panel Discussion: Future Directions and Issues with TCM Clinical Data Mining**

12:40 -- 12:45

**Closing**

PAKDD-2011 Pre-Conference Program (Workshops & Tutorials)  
May 24, 2011

**Workshop 5: Data Mining for Healthcare Management (DMHM)**

**Venue: Madrid 8**

9:00 -- 9:05	<b>Opening Welcome</b>
9:05 -- 9:55	<b>Invited Talk</b> Speaker: <i>Dr. Vipin Gopal (Director, Clinical Analytics, Humana)</i> Topic: <i>Predictive Modeling for Care Management: Impact on Outcomes, Costs and Engagement</i>
9:55 -- 10:20	<ul style="list-style-type: none"><li>• Usage of Mobile phones for Personalized Healthcare Solutions. <i>M Saravanan, S Shanthi and S Shalini</i></li></ul>
10:20 -- 10:45	<ul style="list-style-type: none"><li>• Robust Learning of Mixture Models and Its Application on Trial Pruning for EEG Signal Analysis. <i>Boyu Wang and Feng Wan</i></li></ul>
<b>Coffee Break ( 10: 45 – 11:00 )</b>	
11:00 -- 11:25	<ul style="list-style-type: none"><li>• An Integrated Approach to Multi-Criteria-based Health Care Facility Location Planning. <i>Wei Gu, Baijie Wang and Xin Wang</i></li></ul>
11:25 -- 11:50	<ul style="list-style-type: none"><li>• Medicinal Property knowledge Extraction from Herbal Documents for Supporting Question Answering System. <i>Chaveevan Pechsiri, Sumran Painual and Uraiwan Janviriyasopak</i></li></ul>
11:50 -- 11:55	<b>Closing</b>

PAKDD-2011 Tutorials  
May 24, 2011

**Amazon Tutorial: Scientific Computing in the Cloud**

**Venue: Madrid VIP (马德里贵宾厅)**

9:00 -- 12:00

**Presenter:**

- **Simone Brunozzi**, Amazon Web Services Technology Evangelist for Asia Pacific (APAC)

**Abstract:**

Data Mining, High Performance Computing, Data Processing, and many other scientific use cases are finding new opportunities in the Cloud. Amazon Web Services offers Infrastructure Web Services in the Cloud, providing flexible, on demand computing capacity. Many companies around the world are already successfully running their infrastructure on AWS.

What about the Scientific world?

Is Cloud Computing still a good fit?

Simone Brunozzi, APAC Technology Evangelist for Amazon Web Services, will explain how Cloud Computing can be successfully adopted, offering details on some successful use cases. A Q&A session will conclude the talk.

You can follow Simone on Twitter, @simon, or on his blog, [www.brunozzi.com](http://www.brunozzi.com).

More information on AWS can be found at: <http://aws.amazon.com>

**Tutorial 4: Data Stream Mining: Challenges and Techniques**

**Venue: Madrid VIP (马德里贵宾厅)**

13:30 -- 17:00

**Presenter:**

- **Latifur Khan**, University of Texas at Dallas, USA
- **Wei Fan**, IBM T.J.Watson Research, USA
- **Jiawei Han**, University of Illinois at Urbana-Champaign, USA
- **Jing Gao**, University of Illinois at Urbana-Champaign, USA
- **Mohammad Mehedy Masud**, University of Texas at Dallas (UTD), USA

**Abstract:**

Data streams are continuous flows of data. Examples of data streams include network traffic, sensor data, call center records and so on. Their sheer volume and speed pose a great challenge for the data mining community to mine them. Data streams demonstrate several unique properties: infinite length, concept-drift, concept-evolution, feature-evolution and limited labeled data. Concept-drift occurs in data streams when the underlying concept of data changes over time. Concept-evolution occurs when new classes evolve in streams. Feature-evolution occurs when feature set varies with time in data streams. Data streams also suffer from scarcity of labeled data since it is not possible to manually label all the data points in the stream. Each of these properties adds a challenge to data stream mining. This tutorial presents an organized picture on how to handle various data mining techniques in data streams: in particular, how to handle classification and clustering in evolving data streams by addressing these challenges.

PAKDD-2011 Tutorials  
May 27, 2011

**Tutorial 1: Handling Concept Drift: Importance, Challenges and Solutions**

**Venue:** Madrid 2

14:00 -- 17:30

**Presenter:**

- **Albert Bifet**, University of Waikato, New Zealand
- **João Gama**, University of Porto, Portugal
- **Mykola Pechenizkiy**, Eindhoven University of Technology, the Netherlands
- **Indrė Žliobaitė**, Bournemouth University, UK

**Abstract:**

In the real world data often arrives in streams and is evolving over time. Concept drift in supervised learning means that the underlying distribution of the data is changing. As a result the predictions might become less accurate as the time passes, or opportunities to improve the accuracy might be missed. Therefore, the learning models need to adapt to changes quickly and accurately. The proposed tutorial aims to provide a unifying view on the basic and applied concept drift research in data mining and related areas. In the first part we will introduce the problem of concept drift, discuss why changes appear in supervised learning and motivation to handle them. We will overview what types of application tasks are available. In the second part we will present available approaches and techniques to handle concept drift, discuss evaluation issues and open source software. In the third part we will reflect on the past, present and future of concept drift research and outline future research directions. We will focus on the link between research scenarios and application needs.

**Tutorial 2: Privacy-Preservation of Graphs and Social Networks**

**Venue:** Madrid 3

14:00 -- 17:30

**Presenter:**

- **Xintao Wu**, University of North Carolina at Charlotte, USA
- **Xiaowei Ying**, University of North Carolina at Charlotte, USA

**Abstract:**

Social networks often contain some private attribute information about individuals as well as their sensitive relationships. The privacy concerns associated with data analysis over social networks have incurred the recent research on privacy-preserving social network analysis, particularly on privacy-preserving publishing and querying social network data. Compared with well studied privacy-preservation techniques for tabular data, it is more challenging to design effective anonymization and perturbation techniques for publishing and mining social network data because of the difficulties in modeling link structures in social networks. In this tutorial, we survey the very recent research development on privacy-preserving publishing and analyzing social network data. In the first part of the

tutorial, we focus on privacy preserving publishing social network data and categorize the state-of-the-art anonymization methods on graphs in three main categories: K-anonymity based privacy preservation via edge modification, probabilistic privacy preservation via edge randomization, and privacy preservation via generalization. In the second part of the tutorial, we focus on privacy preserving querying and mining social network data and present the very recent private query answering techniques based on differential privacy. We finally discuss challenges and propose new research directions in this area.

### Tutorial 3: Data Mining for Healthcare Management (DMHM)

Venue: Madrid 6

14:00 -- 17:30

**Presenter:**

- **Prasanna Desikan**, Boston Scientific, USA
- **Jaideep Srivastava**, University of Minnesota, Portugal
- **Kuo-Wei (David) Hsu**, National Chengchi University, Taiwan

**Abstract:**

Data Mining for Healthcare Management (DMHM) has been instrumental in detecting patterns of diagnosis, decisions and treatments in healthcare. Data mining has aided in several aspects of healthcare management including disease diagnosis, decision-making for treatments, medical fraud prevention and detection, fault detection of medical devices, healthcare quality improvement strategies. Data mining was initially a success in the healthcare industry as it was used to detect fraudulent claims processing. However, since then large collections of transactional data and also data due to mergers and acquisitions has provided businesses enough opportunities to analyze and extract informative hidden patterns to reduce costs. Researchers from both academia and industry have recognized the potential of data mining and its impact on improved healthcare by discovering patterns and trends in large amounts of complex data generated by healthcare transactions. Data mining essentially helps discover interesting business insights to help

This tutorial will provide an up-to-date introduction to the increasingly important field of data mining in healthcare management, as well as an overview of research directions in this field. It will cover the most representative research activities and directions in data mining based healthcare management techniques. In this tutorial, we first provide an introduction to healthcare management and then survey the research in this field. Next, an overview of emerging research in data mining for healthcare is presented. This tutorial will help researchers by providing a survey on the research till date, enable the understanding of how data mining can be useful for healthcare management and motivate them to pursue new research in this field. It will also be useful for practitioners from industrial organizations to understand how data mining techniques can help them leverage the potential of large amounts of healthcare data that has been collected.

14:00 -- 15:40

**Presenter:**

- **Gerald Schaefer**, Loughborough University, UK

**Abstract:**

In this tutorial we will look at a variety of techniques and methods for effective and intuitive image database visualization and browsing. Image browsing systems attempt to provide the user with a more intuitive interface, displaying more images at once in order to harness the cognitive power of the human mind in order to recognize and comprehend an image in seconds. Interaction with a traditional QBE system can often lead to a confusing and frustrating user experience. Formulating queries from images can prove difficult for the user, and the 'black-box' state of such approaches means that users typically cannot derive how the system is retrieving these results, and are thus unable to modify the query in order to improve the results returned by the system. Providing an overview of the database to the user allows for intuitive navigation throughout the system. An overview of the collection will give the user a good indication whether or not an image or image class they have in mind might actually be present in the database. In some cases, the entire database will be displayed to the user on a single display. The user can then focus on regions of the visualizations that they are attracted to or believe will harbor a particular concept they have in mind. Browsing such visualizations when arranged according to image similarity, can increase the rate of retrieval. These visualizations are usually achieved through dimensionality reduction, whereby the relationships between images in a high-dimensional feature space are maintained as best possible in a reduced 2D (or 3D) space which is more comprehensible to the user. In case image collections are too large to fit to a single display, images can be grouped according to similarity through the application of a clustering procedure. The user is then able to navigate through these clustered groups of images in order to browse the collection. An overview of the database is provided by initially presenting the user with a representative image for each cluster. Clustering can also be performed in a hierarchical manner which in turn allows for visualization of very large datasets. Another way in which image databases can be displayed is through graph-based visualizations. In these approaches, links are formed between images that are deemed similar or that share a common concept, while the images themselves form the nodes of the graph. The whole connected graph, or part thereof, is then displayed to the user for visualization and navigation. While a visualization of an image collection is useful for providing an overview of the contained images, it provides only part of a useable image database navigation system. Once a collection is visualized, users should have the ability to interact with it in order to arrive at the image(s) they are looking for. Typical operations here include panning and zooming which allow the user to focus on images in a different part of the visualization space, respectively on images which were previously hidden. Finally, it is evident that the usefulness of image browsing systems needs to be evaluated. While there is relatively little work on this, we will look at various measures to attempt to measure how well image browsers are suited for various tasks.

16:00 -- 17:30

**Presenter:**

- **Aixin Sun**, Nanyang Technological University (NTU), Singapore
- **Sourav S Bhowmick**, Nanyang Technological University (NTU), Singapore

**Abstract:**

With the advances in digital photography (e.g., digital cameras and mobile phones) and social media sharing web sites, a huge number of multimedia content is now available online. Most of these sites enable users to annotate web objects including images with free tags. A key consequence of the availability of such tags as meta-data is that it has created a framework that can be effectively exploited to significantly enhance our ability to understand social images. Such understanding paves way to the creation of novel and superior techniques and applications for searching and browsing social images contributed by common users. The objective of this tutorial is to provide a comprehensive background on state-of-the-art techniques for analysis and mining of tags associated with social images.

The tutorial is structured as follows. In the first part, we provide a comprehensive understanding of social image tags. We present a brief survey on studies related to motivation behind tagging and impact of various tagging systems that are used by users to create tags. We shall use Flickr as an example tagging system to illustrate various concepts. In the second part, we shall describe state-of-the-art techniques for mining tags that are effective in describing its annotated resources (social images). Specifically, we shall describe techniques that enable us to quantitatively measure a tag's ability to describe the image content of social images. Note that this issue is one of the most fundamental problems in multimedia analysis and retrieval. The third part of the tutorial is devoted to describing state-of-the-art data mining techniques for discovering relationships between tags and how such knowledge is useful for various tag-based applications such as tag recommendation, tag disambiguation and tag-based browsing systems. We conclude by identifying potential research directions in this area.