

The 10th IEEE International Conference on Data Mining (ICDM'10)

**International Workshop on**  
**Topic Feature Discovery and**  
**Opinion Mining (TFDOM 2010)**

December 14, 2010 Sydney, Australia

**Workshop Co-Organizers**

- **Yuefeng Li**  
Queensland University of Technology, Australia  
Email: [y2.li@qut.edu.au](mailto:y2.li@qut.edu.au)
  
- **Ning Zhong**  
Maebashi Institute of Technology, Japan  
Email: [zhong@maebashi-it.ac.jp](mailto:zhong@maebashi-it.ac.jp)
  
- **Raymond Y. K. Lau**  
City University of Hong Kong, Hong Kong  
Email: [raylau@cityu.edu.hk](mailto:raylau@cityu.edu.hk)

**Workshop Publicity Chair**

- **Xiaohui (Daniel) Tao**  
Queensland University of Technology, Australia  
Email: [x.tao@qut.edu.au](mailto:x.tao@qut.edu.au)

## **Welcome Message from Workshop Co-Organizers**

Topic feature discovery and opinion mining are extremely challenging topics in modern information analysis, from both an empirical and a theoretical perspective. They are also of central interest and the critical steps for many Web personalized applications and recommender systems. The problems in topic feature discovery and opinion mining have charged continuously increasing attentions from researchers in data mining, Web intelligence, text mining, machine learning, natural language processing, and information retrieval communities. Strongly focusing on these two challenging topics and their surrounding areas, this workshop aims to enhance the current text mining and opinion mining techniques, and explores a new methodology to discover and interpret useful and interesting knowledge in text documents.

The workshop received 22 submissions in total, and out of which 9 papers were accepted. The acceptance rate for published papers is approximately 41%. Each paper submitted to the workshop was reviewed by at least two program committee members and, using these reviews, the workshop organizers selected papers for inclusion in the workshop proceedings and presentation at the workshop.

We appreciate the guidance and suggestions we received from the workshops co-chairs. We are also grateful to Xiaohui Tao (Daniel) for his excellent technical assistance in ensuring smooth functioning of the paper management functions. Finally, we thank all authors for their submissions and all attendees for their participation.

Yuefeng Li, Raymond Lau and

Ning Zhong

**TFDOM 2010 Workshop Co-Organizers**

## Workshop Description

Text information in the world can be roughly categorized into two main types: facts and opinions. Much effort has been invested in the past decades in the fact-based text processing, and many techniques have been developed for retrieving information or text mining. The opinion-based text processing, however, has received just limited attentions from researchers. Listening to other people's opinions is important, especially when taking account of valuable customers' opinions into an organization's decision making process. However, identifying relevant sources, extracting topic features, and summarizing opinions are all difficult tasks. These are the problems targeted by topic feature discovery and opinion mining.

Topic feature discovery aims to identify on-topic sources and extract relevance features for a given topic (e.g., a person, an event, or a government policy). Many empirical experiments illustrated that the performance of text mining was hindered because of the measures shipped from data mining. Such performance measures, for example, support and confidence, turn out to be unsuitable in the leveraging stage. By way of illustration, given a specified topic, usually a highly frequent pattern (normally a short pattern) is general and a specific pattern is lowly frequent. The objective of topic feature discovery is to find a suitable subset of features available in text documents to describe the requested topic.

Opinion mining, also known as sentiment analysis, aims to summarize and classify opinions. Compared with traditional topic-based or fact-based analysis, opinion mining tends to address the new problems raised by the applications concerning the subjective or opinionated expression. Its aim is to determine the inclination of a reporter with respect to some topics, or to extract the opinions from a large variety of digital texts containing opinionated content. Difficult problems in opinion mining include extracting opinionated information; classifying sentiments and subjectivity; analysing feature-based sentiments; identifying opinion spam; and applying opinions to problem solving or decision making. Therefore, there are many opportunities and feasibility of extensive research activities in the field of opinion mining.

Topic feature discovery and opinion mining are extremely challenging topics in modern information analysis, from both an empirical and a theoretical perspective. They are also of central interest and the critical steps for many Web personalized applications and recommender systems. The problems in topic feature discovery and opinion mining have charged continuously increasing attentions from researchers in data mining, Web intelligence, text mining, machine learning, natural language processing, and information retrieval communities. Strongly focusing on these two challenging topics and their surrounding areas, this workshop aims to enhance the current text mining and opinion mining techniques, and explores a new methodology to discover and interpret useful and interesting knowledge in text documents.

**Topics of interest include, but are not limited to:**

- Relevance feature discovery
- Opinion mining and sentiment analysis
- Information filtering and retrieval
- Text mining
- Text categorizations
- Ontology mining
- Information extraction
- Sentiment and subjectivity classification
- Feature-based sentiment analysis
- Recommender systems
- Web personalization
- Evaluation methodologies in topic feature discovery and opinion mining

## Keynote Speaker

Professor Bing Liu, "Opinion Mining: What is there for Data Miners"

**Biography:** Prof. Liu is a professor of Computer Science at the University of Illinois at Chicago (UIC). He received his PhD in Artificial Intelligence from the University of Edinburgh. Before joining UIC, he was with the National University of Singapore. His current research interests include opinion mining (or sentiment analysis), text and Web mining, data mining, and machine learning. He has published extensively in these fields. He has also written a textbook titled "Web Data Mining: Exploring Hyperlinks, Contents and Usage Data" published by Springer. On professional services, Liu has served as associate editors of IEEE Transactions on Knowledge and Data Engineering, and SIGKDD Explorations, and is on the editorial boards of several other journals. He has also served or serves as program chairs of IEEE International Conference on Data Mining (ICDM-2010), ACM Conference on Web Search and Data Mining (WSDM-2010), ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2008), SIAM Conference on Data Mining (SDM-2007), ACM Conference on Information and Knowledge Management (CIKM-2006), and Pacific Asia Conference on Data Mining (PAKDD-2002). In addition, he has served extensively as area chairs and program committee members of leading conferences on data mining, Web mining, natural language processing, and machine learning. Further information about him can be found at <http://www.cs.uic.edu/~liub>.

**Abstract:** Opinion mining (or sentiment analysis) is the computational study of people's opinions, appraisals, attitudes and emotions toward entities, individuals, issues, events, topics and their attributes. It has become a very active research area in the past few years due to challenging research problems and a wide arrange of applications. There are now at least 40 companies in USA alone that provide some kinds of opinion mining services. Opinions are important because they are key influences on our behaviors. It is well known that our beliefs and perceptions of reality are to a considerable degree conditioned on how others see the world. For this reason, when we need to make a decision we often seek out the opinions of others. This is true not only for individuals but also for organizations. With the explosive growth of social media (i.e., reviews, forum discussions, blogs and social networks) in the past 10 years, individuals and organizations are increasingly using these media for their decision making. Nowadays, if one wants to buy a consumer product, one is no longer limited to asking one's friends and families for opinions because there are many user reviews of products on the Web. For an organization, it may no longer be necessary to conduct opinion polls, surveys, and focus groups in order to gather public opinions about its products and services because there is an abundance of such information publicly available. However, finding and monitoring opinion sites on the Web and distilling the information contained in them remains a formidable task because of the proliferation of diverse sites. Each site typically contains a huge volume of opinionated

text that is not always easily deciphered in forum posts and blogs. An average human reader has difficulty identifying the relevant sites and accurately summarizing opinions contained in them. In this talk, I will first discuss the key technical problems involved from both the research and practical application points of view. Interestingly but not surprisingly, they are almost all data mining problems with some NLP flavor. As an example, I will then focus on solving one or two problems and go into details.

## Accepted Papers

- ◆ Infrequent Purchased Product Recommendation Making Based on User Behaviour and Opinions in E-commerce Sites  
*Noraswaliza Abdullah, Yue Xu, Shlomo Geva, and Jinghong Chen*
- ◆ The PARIS Algorithm for Determining Latent Topics  
*Michal Aharon, Ira Cohen, Arik Itskovitch, Inbal Marhaim, and Ron Banner*
- ◆ A Block Mixture Model for Pattern Discovery in Preference Data  
*Nicola Barbieri, Massimo Guarascio, and Giuseppe Manco*
- ◆ Mining Arabic Business Reviews  
*Mohamed Elhawary and Mohamed Elfeky*
- ◆ Sentence-Level and Document-Level Sentiment Mining for Arabic Texts  
*Noura Farra, Elie Challita, Rawad Abou Assi, and Hazem Hajj*
- ◆ High-Order Concept Associations Mining and Inferential Language Modeling for Online Review Spam Detection  
*C.L. Lai, K.Q. Xu, Raymond Y.K. Lau, Yuefeng Li, and Dawei Song*
- ◆ Mining Users' Opinions Based on Item Folksonomy and Taxonomy for Personalized Recommender Systems  
*Huizhi Liang, Yue Xu, and Yuefeng Li*
- ◆ A Framework for Emotion Mining from Text in Online Social Networks  
*Mohamed Yassine and Hazem Hajj*
- ◆ Augmenting Chinese Online Video Recommendations by Using Virtual Ratings Predicted by Review Sentiment Classification  
*Weishi Zhang, Guiguang Ding, Li Chen, and Chunping Li*

## Workshop Program

<b>0900-1000am</b>		<b>Session 1 (Chair: Yuefeng Li)</b>	
0900-0915am	Workshop Opening <i>Yuefeng Li</i>		
0915-1000am	Keynote: Opinion Mining: What is there for Data Miners <i>Bing Liu</i>		
<b>1000-1030am</b>		<b>Coffee Break</b>	
<b>1030-1200pm</b>		<b>Session 2 (Chair: Raymond Y.K. Lau)</b>	
1030-1050am	Sentence-Level and Document-Level Sentiment Mining for Arabic Texts <i>Noura Farra, Elie Challita, Rawad Abou Assi, and Hazem Hajj</i>		
1050-1110am	Mining Users' Opinions Based on Item Folksonomy and Taxonomy for Personalized Recommender Systems <i>Huizhi Liang, Yue Xu, and Yuefeng Li</i>		
1110-1130am	A Framework for Emotion Mining from Text in Online Social Networks <i>Mohamed Yassine and Hazem Hajj</i>		
1130-1150am	Augmenting Chinese Online Video Recommendations by Using Virtual Ratings Predicted by Review Sentiment Classification <i>Weishi Zhang, Guiguang Ding, Li Chen, and Chunping Li</i>		
<b>1200-1330pm</b>		<b>Lunch</b>	
<b>1330-1530pm</b>		<b>Session 3 (Chair: Yuefeng Li)</b>	
1330-1350pm	The PARIS Algorithm for Determining Latent Topics <i>Michal Aharon, Ira Cohen, Arik Itskovitch, Inbal Marhaim, and Ron Banner</i>		
1350-1410pm	Infrequent Purchased Product Recommendation Making Based on User Behaviour and Opinions in E-commerce Sites <i>Noraswaliza Abdullah, Yue Xu, Shlomo Geva, and Jinghong Chen</i>		
1410-1430pm	A Block Mixture Model for Pattern Discovery in Preference Data <i>Nicola Barbieri, Massimo Guarascio, and Giuseppe Manco</i>		
1430-1450pm	Mining Arabic Business Reviews <i>Mohamed Elhawary and Mohamed Elfeky</i>		
1450-1510pm	High-Order Concept Associations Mining and Inferential Language Modeling for Online Review Spam Detection <i>C.L. Lai, K.Q. Xu, Raymond Y.K. Lau, Yuefeng Li, and Dawei Song</i>		
<b>1530-1600pm</b>		<b>Coffee Break</b>	
<b>1600-1700pm</b>		<b>Session 4 (Chair: Yuefeng Li)</b>	
1600-1650pm	Panel Discussion Session		
1650-1700pm	Workshop Closing <i>Raymond Y.K. Lau</i>		

## Program Committee

- Albert Au yeung      NTT Communication Science Laboratories, Japan
- Ling Chen            University of Technology, Sydney, Australia
- Michael Gamon        Microsoft Research, USA
- Xiaoying Gao         Victoria University of Wellington, New Zealand
- Shlomo Geva          Queensland University of Technology, Australia
- Jimmy Huang         Youk University, Canada
- Qingliang Miao        Chinese Academy of Sciences, China
- Stuart E. Middleton    University of Southampton, UK
- Chunping Li         Qinghua University, China
- Qiudan Li            Chinese Academy of Sciences, China
- Tao Li                Florida International University, USA
- Wenjie Li            Hong Kong Polytechnic University, China
- Yang Liu             York University, Canada
- Yuee Liu             Northwest A&F University, China
- Yue Lu                University of Illinois at Urbana-Champaign, USA
- Manabu Okumura        Tokyo Institute of Technology, Japan
- Luiz Augusto Pizzato   University of Sydney, Australia
- Franco Salvetti       Microsoft Corp., USA
- Giovanni Semeraro     University of Bari, Italy
- Dian Tjondronegoro    Queensland University of Technology, Australia
- Hui Wang            Ulster University, UK
- Yue Xu                Queensland University of Technology, Australia
- Yiyu Yao             Regina University, Canada
- Bei Yu                Syracuse University, USA
- Yunqing Xia         Qinghua University, China
- Wei Xu                Renmin Univeristy, China
- Markus Zanker        University Klagenfurt, Austria
- Daniel Zeng         The University of Arizona, USA
- Songmao Zhang        Chinese Academy of Sciences, China
- Yanchang Zhao        Centrelink, Australia