

# A Conversation with Professors Deyi Li and Jie Tang

Deyi Li

Institute of Chinese Electrical System Engineering  
Beijing 100840, China  
leedeyi@nlsde.buaa.edu.cn

Jie Tang

Tsinghua University  
Beijing 100084, China  
jietang@tsinghua.edu.cn

## 1. Please share with us your view on the history and important milestones of the Chinese KDD research and application areas.

Roughly speaking, Chinese KDD research mainly underwent three stages. It was in 1993 when National Science Foundation of China (NSFC) started to sponsor research on knowledge discovery and data mining. This can be considered as the first stage. The major research around that time was focused on “Knowledge Discovery from Database”, including sub-topics such as frequent mining and association rule mining from databases. The research was mainly conducted in academic institutes. The second stage started from the end of 1990’s, with the emergence and the rapid proliferation of Web-based applications. People started to notice that the largest data source for mining is the information on the Web instead of traditional databases. At the same time the mining tasks became more diversified. In the second stage, the term “Web Mining” became popular in the field. Research labs on “knowledge engineering”, “web/internet mining” have been built in different research institutes and rapidly developed. Several web search companies also emerged in this stage such as Baidu and Sogou. The third stage began around 2005, when online social applications and media (such as, in China, Tencent, Sina Weibo, Renren) become a prevalent and complex force to influence our daily life. Indeed, Tencent, the largest social network in China, already has more than 700 million registered users, the same number of Facebook; Sina Weibo has attracted 250 million users in the past two years, a figure higher than Twitter. These online networks grow very fast and they provide a huge amount of user generated content, which presents great opportunities in understanding the science of these networks. Accordingly, the emphasis of the research started to switch to mining social networks. This is a more diverse research field, attracting researchers from a wide range of academic fields, including theory and algorithms, data mining and machine learning, computer systems and networks, statistical physics and complex systems, social psychology, economics and managerial science. Another important change in this stage is that Chinese companies are paying more and more attention to data mining research. Not only Chinese Internet companies (e.g., Tencent, Baidu, Sogou, Youdao, etc.) but also communication/hardware IT companies (e.g., China Mobile, Huawei, ZTE, Lenovo) started to build data mining research labs. There is little doubt that for now it is the best time for data mining in China.

## 2. Please describe your expertise and contribution to KDD.

Deyi Li is a member of the Chinese academy of engineering and the International Academy of Sciences for Europe and Asia. He is also an adjunct professor in Tsinghua University, National

University of Defense Technology, Beijing University of Posts and Telecommunications, and Beihang University. Li’s first paper on KDD was published in 1996 on a system for mining knowledge in large relational databases. Deyi Li was also the first to propose “Cloud Models” in China. At present, the research focus of the authors’ team is to use the Cloud Models to represent data’s randomness, fuzziness, and to mine associations from data. On the basis of the cloud models, our team extensively studied super entropies, high order statistics for Gaussian distribution, probabilistic factor graph models. We proposed a general and flexible framework, namely iMiner, for providing mining services for large-scale data. So far, we have developed several instances of the framework, i.e., iMiner for academic ([Arnetminer.org](http://Arnetminer.org)), iMiner for news ([Newsminer.net](http://Newsminer.net)), iMiner for pictures, and iMiner for patents ([Pminer.org](http://Pminer.org)). The Arnetminer is designed to search and perform data mining operations against academic publications on the Internet, using social network analysis to identify connections between researchers, conferences, and publications. It has attracted more than 210 million visits from 218 countries in the world, and also was widely used to provide academic services for publisher such as Elsevier and conferences such as KDD, ICDM, and WSDM.

## 3. Please share with us your view on the future of KDD both in China and the world.

In our view, there are two important research directions for data mining. The first is to investigate new mathematical models to explain the “noisy” information space. This is especially true when we study data mining on the social networking data. The user-generated data is not necessarily as clean as the standard text on the traditional Web (e.g., news articles). What is the fundamental difference between these two types of information? Moreover, the online social network presents a new network structure, which is no doubt very different from traditional webpage based network and our physical network, but what is the fundamental difference? Moreover, how does information spread in the social network and will information/network finally reach an equilibrium and how? Another interesting direction is to identify emerging data mining tasks. This is important for the data mining community.

## 4. REFERENCES

- [1] Jiawei Han, Yongjian Fu, Wei Wang, Jenny Chiang, Wan Gong, Krzysztof Koperski, Deyi Li, Yijun Lu, Aymnmohamed Rajan, Nebojsa Stefanovic, Betty Xia, Osmar R. Zaiane. DBMiner: A System for Mining Knowledge in Large Relational Databases. KDD’1996. pp.250~255

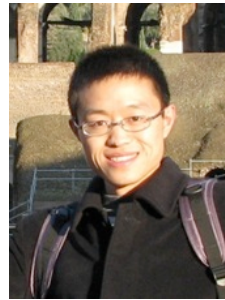
- [2] Deyi Li. iMiner: From Passive Searching to Active Pushing. APWeb'2011. pp.3~3
- [3] Deyi Li, Kaichang Di, Deren Li, Xuemei Shi. Mining Association Rules with Linguistic Cloud Models. PAKDD'1998. pp.392~393
- [4] Deyi Li, Jiawei Han, Xuemei Shi, Man-chung Chan. Knowledge representation and discovery based on linguistic atoms. Knowl.-Based Syst., 1998: 431~440
- [5] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, Zhong Su. ArnetMiner: extraction and mining of academic social networks. KDD'2008. pp.990~998
- [6] Jie Tang, Jimeng Sun, Chi Wang, Zi Yang. Social influence analysis in large-scale networks. KDD'2009. pp.807~816
- [7] John E. Hopcroft, Tiancheng Lou, Jie Tang, and Liaoruo Wang. Detecting Community Kernels in Large Social Networks. ICDM'2011.

---

**About the authors:**



**Deyi Li** is a member of the Chinese academy of engineering and the International Academy of Sciences for Europe and Asia. He is the Chairman of the Chinese Artificial Intelligence Society (<http://caai.cn/>). He is a professor of software engineering at Tsinghua University, China, and heads the Information Science Directorate of the Natural Science Foundation of China. His current research interests include networked data mining, artificial intelligence with uncertainty, and cloud computing. <http://www.arnetminer.org/viewperson.do?naid=499431&keyword=Deyi%20Li>



**Jie Tang** is an associate professor in Tsinghua University. His research interests are social network analysis, data mining, and machine learning. He has published over 100 research papers in major international journals and conferences. He is the initiator of the system Arnetminer.org, which has attracted millions of users from more than 200 countries. Jie is the local chair for KDD 2012. <http://keg.cs.tsinghua.edu.cn/jietang/>