A Conversation with Dr. Yong Shi

Yong Shi
Chinese Academy of Sciences
Beijing 100080, China
yshi@gucas.ac.cn

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

Chinese KDD or Data Mining research started from the early 1990's when the first group of international KDD communities was formed. Since the demand of using information technology (IT), including computing tools, and Internet communications, was growing in Chinese economic reform, many scholars in universities and institutes have paid attention on KDD related research. A number of books in the field, such as Artificial Intelligence by Ruqian Lu (1996 in Chinese) and Knowledge Discovery by Zhongzhi Shi (2011 in Chinese), have recorded the progress of KDD in China. These works have also influenced young generations who were working on KDD at home and aboard.

In recent years, Chinese governmental branches and industries have built their own databases. Some of them, especially China's commercial banks and mobile communication industries, have been migrating the databases into data warehouse and applying techniques of KDD to solve their business decision making problems. Chinese ICT (information and communications technology) market becomes the largest one in the world. For example, there are 457 Chinese million using Internet service in 2011. It was forecasted that by 2015, the Internet population of Chinese will reach 1.2 billion. As a significant number of researchers from both research institutes and universities are showing their increasing interest in doing various KDD problems, the National Natural Science Foundation of China (NSFC) has sponsored a large number of KDD proposals since 1990's. According to China's National Science and Technology Development Mid-Long Term Planning (2006-2020), "theories and methods of large-scale information processing and knowledge mining" have been identified as one of the key supporting technologies in fundamental scientific research for the national prioritized strategic needs. This has demonstrated that Chinese government's strong commitment on KDD related research and applications

2. Please describe your expertise and contribution to KDD.

The Chinese Academy of Science (CAS) has initiated the CAS Research Center on Data Mining & Knowledge Economy (CASDTKE) in October 2004 as an integration research body based on the Graduate University of Chinese Academy of Science. This center then has been formally renamed as the CAS Research Center on Fictitious Economy & Data Science (CASFEDS), where Yong Shi (the author) serves as its Executive Director, in January 2007. With the multidisciplinary nature of KDD, financial markets, environmental sciences and public management, CASFEDS has three major functions: fundamental and theoretical development, application-oriented research, and thank tank of Chinese government. For the last seven years, CASFEDS has been granted more than 30 million RMB by NSFC, the Ministry of Chinese Science and Technology, CAS and National Audit Office of China for its KDD related research projects. It has published more than 300 research papers in the international journals and conferences, including a number of major KDD related journals.

In 2011, the International Journal of Information Technology and Decision Making (ITDM), founded by Yong Shi (the author) in 2002, has a new ISI Web of Knowledge Impact Factor of 3.139, with ranking in the category Computer Science, Artificial Intelligence, 7th out of 108; Computer Science, Information Systems, 5th out of 126; Computer Science, Interdisciplinary Application, 7th out of 97; and Operations Research & Management Science, 3rd out of 74. A number of KDD papers have been published in ITDM. With the support of its publisher World Scientific Co., the editorial board of ITDM has built **Herbert Simon Award** for Outstanding Contribution in Information Technology and Decision Making.

In KDD applications, CASFEDS has been actively working with the People's Bank of China (PBC, or China's Central Bank), Industrial and Commercial Bank of China, China Investment Co., China Reinsurance Co., Shenzhen Development Bank, Netease Co., BHP Billion. Co., Australia, and other corporations to solve various KDD problems. Among them, CASFEDS and PBC formed a joint research team under my leadership in developing the National Credit Scoring System via KDD technologies, called "China Score", from the world largest personal credit database of 650 million accounts hosted by PBC. It will be eventually serving all 1.3 billion population of China for their daily financial activities, such as bank accounts, credit card application, mortgage, personal loans, etc. This can become the most influential event of KDD techniques to human kind so far. In addition, CASFEDS has provided more than two dozen of policy recommendations based on its KDD research to Chinese government, which have already produced fundamental impact on the strategic decision making on Chinese economic developments.

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

The future of KDD can be further developed in two important directions. The first one is to focus on the basic scientific problems, such as how the phenomenon, structure and interpretation of data lead a theory of data, how the rules developed from different human experts can be incorporated to perform a effective data preprocessing, and how various human

knowledge can be combined with the hidden patterns from KDD processes to derive useful knowledge or intelligent knowledge as an efficient solution to the target problems, etc. This line of research requires the creation of a new subject, can be called "Data Science", based on the preliminary areas of KDD or data mining, knowledge management, and wisdom discovery. The second direction of the KDD future will be the capability development of KDD tools to deal with real-life applications. Traditional KDD has emphasized on using different algorithms for various testing datasets for finding a high efficiency of the proposed methodology. A challenging question now open to all of scholars and practitioners in KDD: can we demonstrate more significant KDD related solutions that affect human daily life, global economic development, and environmental changes so that more and more people will believe the power of KDD and widely use KDD tools?

About the author:

Yong Shi is professor at Chinese Academy of Sciences and Research Center on Fictitious Economy and Data Sciences, Beijing 100080, PR China.

http://www.mscas.ac.cn/teachers/shiyong.htm