Intelligent Techniques for Data Science

Tianrui Li[1, *], Jie Lu[2, †], Luis Martínez López[3, ‡]

[1]School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, People's Republic of China

[2]Decision System and e-Service Intelligence Lab, Centre for Quantum Computation and Intelligence Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

[3]Computer Science Department, University of Jaén, Spain

With the extraordinary spread of computers and sensors, enormous amounts of data are generated every day in a range of areas—search engines, social media, healthcare organizations, insurance companies, financial industry, retail, and many others. Data science refers to the theories, methods, and applications for extracting previously unavailable and potentially highly useful information from data. This field has evolved as a hybrid of research in data mining, machine learning, computational intelligence, databases, algorithms, statistics, operations research, visualization, privacy and security. It is helping us make sense out of vast quantities of information. However, how to use these data by an effective and ethical way is a significant challenge to science and to society as a whole.

Intelligent techniques, including artificial intelligence, neural networks, fuzzy logic, granular computing, rough sets, expert systems, case-based reasoning, evolutionary algorithms and swarm computing, have been successfully applied in many fields including data science. This special issue is devoted to the use of intelligent techniques for data science that reflects their current development obtained from selected papers submitted to the 8th International Conference on Intelligent Systems and Knowledge Engineering (ISKE2013) held in Shenzhen, China, during November 20-23, 2013. This issue encompasses seven papers that present the application of different intelligent techniques to different data science problems ranging from recommender systems to recognition processes passing by others like activity simulation, fuzzy trading systems, deep learning and incremental learning.

The paper coauthored by Wei Wang, Guangquan Zhang, and Jie Lu investigates the collaborative filtering with an entropy-driven user similarity in recommender systems. It aims at improving recommendation performance, by means of a novel collaborative filtering approach that uses an information entropy-driven user similarity model to measure the relative difference between ratings and a Manhattan distance-based model to address the fat tail problem through estimating the alternative active user average rating. Experimental results on public and private data sets validate the effectiveness of the proposed recommendation approach.

* Author to whom all correspondence should be addressed: e-mail: trli@swjtu.edu.cn.

† e-mail: jielu@it.uts.edu.au

‡ e-mail: martin@ujaen.es

The paper coauthored by Huisi Wu, Lei Wang, Feng Zhang, and Zhenkun Wen presents a method for leaf recognition from a big hierarchical image database. The textural gradient histogram with the shape context is combined to form a more distinctive feature for leaf recognition. To achieve efficient leaf image retrieval, the big database is divided into a set of subsets based on Mean-shift clustering on the extracted features, and hierarchical k-dimensional trees is built to index each cluster in parallel. The parallel indexing and searching schemes are implemented in the MapReduce framework. Both visual results and statistical results from extensive experiments on different databases demonstrate the validation of the proposed method.

The paper coauthored by Jorge Castro, Francisco J. Quesada, Ivan Palomares, and Luis Martinez presents a consensus-driven group recommender system. It presents a novel technique for group recommendations based on consensus reaching processes to provide the added value to Group Recommender Systems of improving group members' satisfaction regarding the recommended items. A Case study illustrates that the proposed approach may improve the results by applying consensus reaching processes in group recommendation compared with classical group recommending techniques.

The paper coauthored by Qihe Liu, Xiaonan Hu, Mao Ye, Xianqiong Cheng, and Fan Li applies deep learning in gas recognition under sensor drift. Unsupervised features are automatically extracted by using the deep learning technique and a gas classifier based on these features is then presented. Simulated results show that the proposed method is robust to gas sensor drift.

The paper coauthored by Jie Hu, Tianrui Li, Hongmei Chen, and Anping Zeng addresses an incremental learning approach for updating approximations in rough set model over dual universes. Principles for updating approximations when the object sets of two universes vary with time are discussed and the four corresponding incremental algorithms to compute approximations are then designed. An illustrative example shows the feasibility of the proposed method. Experimental results show that it can effectively shorten the running time of computing approximations compared with the non-incremental method.

The paper coauthored by Qiong Bao, Bruno Kochan, Tom Bellemans, Lieve Creemers, Yongjun Shen, Davy Janssens, and Geert Wets investigates the extension of the activity-based micro-simulation modeling framework FEATHERS from the Subzone zoning system to a more disaggregated zoning system, i.e., Building block. The data processing procedure is elaborated to implement the FEATHERS framework under the Building block zoning system. The observed as well as the predicted travel demand in Flanders based on the Subzone and Building block zoning systems are compared. The extended modeling system is further applied to investigate the potential impact of light rail initiatives on travel demand at a local network in Flanders.

Finally the paper coauthored by Rodrigo Naranjo, Albert Meco, Javier Arroyo, and Matilde Santos proposes an intelligent trading system with fuzzy rules and fuzzy capital management. In this system, fuzzy logic is applied not only for defining the trading rules, but also for managing the capital to invest. Two fuzzy decision support systems are developed, in which the first one

employs fuzzy logic to design the trading rules and to apply the stock market technical indicators and the other one improves this fuzzy trading system by adding a fuzzy strategy to manage the capital to trade. A new technical market indicator to produce short and long entry signals is then introduced based on the moving average convergence divergence indicator whose parameters are optimized by genetic algorithms. A comparison of the proposed trading system with its classical non-fuzzy version indicates that the use of fuzzy logic for capital management is effective.

Last but not least, as Guest Editors of this special issue on intelligent techniques for data science, we would like to express our gratitude to all the authors for their contribution and thank all the referees for their careful and constructive reviews that led to further improvement of the articles. We also would like to give our sincere gratitude to the Editor-in-Chief of the Journal, Prof. Ronald R. Yager, for accepting to publish this special issue and for his guidance throughout the publication process.

Acknowledgments