

---

## Preface

---

### Shuigeng Zhou

Fudan University  
Shanghai, China  
E-mail: sgzhou@fudan.edu.cn

**Biographical notes:** Shuigeng Zhou is a Professor at the School of Computer Science, Fudan University, China. He received a PhD in Computer Science from Fudan University in 2000, a Master's degree from China University of Electronic Science and Technology (CUEST) in 1991 and a Bachelor's degree from Huazhong University of Science and Technology (HUST), China in 1988. His research interests include databases, data mining, web searching and information retrieval, bioinformatics and complex networks. He has published more than 100 technical articles in international/domestic journals and conferences. Recently, he has served as organiser or Program Committee (PC) member in a number of international conferences, including SIGMOD, ICDE, ER, DASFAA, MDM, PAKDD, ADBIS, PRICAI, FSKD, WAIM and APWeb. Currently, he is a member of ACM, IEEE CS and IEICE.

---

Data mining has been an active research area in the academic community in the past two decades. This special issue includes six selected papers from the Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'07).

FSKD'07 was held jointly with the Third International Conference on Natural Computation (ICNC'07) from 24–27 August 2007 in Haikou, Hainan, China. It received 1305 submissions from 35 countries/regions, in which 586 papers were included in the proceedings of FSKD'07 after a regular peer reviewing process. Through a careful selection, extended versions of 14 papers from the FSKD'07 proceedings were submitted for possible publication in this special issue. After a round of peer reviewing and a round of revision, six papers were finally accepted for publication in the special issue. The six papers cover various aspects of data mining, including clustering, co-location mining, feature selection, knowledge reduction, semantic analysis and hybrid wavelet model construction.

Yan *et al.* ('Multifractal-based cluster hierarchy optimisation algorithm') propose the Multifractal-based Cluster Hierarchy Optimisation (MFCHO) algorithm to construct the cluster hierarchy tree from disjoint initial clusters based on multifractal theory, which integrates the cluster similarity with the cluster shape and cluster distribution. Wan *et al.* ('KNFCOM-T: a  $k$ -nearest features-based co-location pattern mining algorithm for large spatial data sets by using T-trees') present the notion of  $k$ -nearest features (simply  $k$ -NF)-based co-location pattern. The  $k$ -NF set of a spatial feature's instances is used to evaluate the spatial relationship between this feature and any other feature. A  $k$ -NF-based co-location pattern mining algorithm by using T-tree (KNFCOM-T in short) is further developed to identify the co-location patterns in large spatial data sets. Manzour *et al.*

(‘Genome-wide efficient attribute selection for purely epistatic models via Shannon entropy’) propose and discuss a convenient framework for modelling epistasis using the information-theoretic concepts and algorithms inspired by such an approach. A new visualisation approach is also introduced for the purpose of better illustrating epistasy for cases where the number of loci is more than two or three. Xu and Zhang (‘Knowledge reduction and matrix computation in inconsistent ordered information systems’) propose the assignment reduction and approximation reduction for inconsistent ordered information systems. The properties and relationships between assignment reduction and approximation reduction is discussed. The dominance matrix and decision assignment matrix are also proposed for information systems based on dominance relations and the algorithm of assignment reduction is introduced. Hu and Du (‘A semantic analysis of Chinese radicals’) present a quantitative analysis of Chinese radicals’ semantics by mapping radicals and the corresponding characters’ semantics to the WordNet synsets hierarchy, explore the semantics strength of Chinese radicals and discover the entropy of Chinese radicals set. Zhang *et al.* (‘Hybrid wavelet model construction using orthogonal forward selection with boosting search’) investigate sparse regression modelling by using a generalised kernel model where each kernel regressor has its individually tuned centre vector and diagonal covariance matrix. An orthogonal least squares forward selection procedure is employed to select the regressors one by one using a guided random search algorithm. In order to prevent the possible overfitting, a practical method to select the termination threshold is used. A novel hybrid wavelet is constructed to make the model sparser.

Editing a special issue requires a lot of effort from many people. I am grateful to all the authors contributing to this special issue. I appreciate the reviewers who helped in reviewing the submissions and selecting quality papers. I thank Professor David Taniar, the Editor-in-Chief of IJBIDM, for supporting this special issue and especially for his excellent coordination and cooperation in preparing the special issue. I sincerely hope that this special issue is interesting and helpful to the readers.