

Guest editorial: Special issue on advances in intelligent data processing and analysis

Chee Peng Lim · Margarita N. Favorskaya ·
Lakhmi C. Jain

Received: 24 May 2014 / Accepted: 5 July 2014 / Published online: 26 July 2014
© The Natural Computing Applications Forum 2014

In today's information era, data (especially digital data) are generated at an unprecedented rate in terms of volume and velocity. Data gathered from the real world are often multifaceted, e.g. spatiotemporal data, multivariate data, and multimodal data, and exist in a variety of formats, e.g. structured/unstructured text, time-/frequency-domain signals, static images, and dynamic video streams. The available data samples also need to be processed carefully, e.g. cleansing, filtering, and transforming, before useful information can be extracted for inferring conclusions and supporting decision-making.

Intelligent techniques stemming from the domain of soft computing, which include neural, fuzzy, and evolutionary computing methodologies and other related techniques, provide a viable approach to either automatically or semi-automatically process and analyse a large volume of data. Neural computing techniques are useful for processing low-level data samples and extracting information out from data samples, while fuzzy computing techniques are beneficial for analysing high-level human linguistic variables and performing inference/reasoning with if-then rules elicited from human experts. On the other hand, evolutionary computing techniques are capable of searching in a high-dimensional space and finding solutions for optimization problems.

In this special issue, a total of nine of articles reporting recent research findings in intelligent data processing and analysis from either theoretical or practical perspectives are presented. It should be noted that these articles represent only a small fraction of researches in this fast-moving domain. We aspire that this special issue could provide useful insight and stir further research in the area of intelligent data processing and analysis. A summary of each article is as follows.

In Knauer et al., an ensemble of Radial Basis Function (RBF) neural networks with γ -divergence-based similarity measures is first formulated. Fusion of the RBF outputs with decision trees is then performed. In addition, a selection scheme for subsets of RBF networks based on their relevance in the fusion process is proposed. A number of hyperspectral imaging data sets are deployed for evaluation. Different tree-based learners and combination strategies, which include AdaBoost with decision trees, random forests, and pruned decision trees, are experimented. The results demonstrate the usefulness of the proposed fusion tree approach to combining multiple RBF outputs in deriving an accurate classification system.

The use of Genetic Algorithm (GA) to design an appropriate structure of the Fuzzy ARTMAP (FAM) neural network is described in Loo et al. The network parameters and the order of training data are first determined by a GA. Then, an ensemble of FAM networks is optimized with another GA to improve its performance in undertaking data classification problems. Probabilistic voting is employed to combine the predictions from multiple FAM networks. The usefulness of the proposed model is demonstrated using benchmark problems from the UCI Machine Learning repository.

To reduce patient waiting time and to improve quality of healthcare, Wang et al. adopt the Fuzzy Min-Max (FMM)

C. P. Lim
Deakin University, Waurn Ponds, Victoria, Australia

M. N. Favorskaya
Siberian State Aerospace University, Krasnoyarsk, Russia

L. C. Jain (✉)
University of South Australia, Mawson Lakes, SA, Australia
e-mail: Lakhmi.Jain@unisa.edu.au

neural network to undertake patient admission prediction problem in the emergency department of a hospital. In addition to providing accurate prediction, the FMM network is endowed with a rule extraction facility utilizing a GA. The GA is used to maximize prediction accuracy and minimize the number of FMM hyperboxes, therefore, providing an optimal configuration for rule extraction. The results with a large real patient database reveal the effectiveness of the proposed model in producing accurate prediction with the capability of explaining its predictions with if-then rules.

A co-evolutionary genetic watermarking scheme based on wavelet packet transform is proposed by Chen and Huang. By treating the task of embedding watermarks into images as an optimization problem, a cooperative co-evolutionary GA is used to select the appropriate basis of wavelet packet transform and determine the sub-bands for watermark embedding. The experimental results show that the proposed model is able to select the best wavelet packet transform basis and sub-bands to increase watermark robustness in conjunction with some specific image processing methods. In addition, image fidelity and watermark robustness can be adjusted through the relevant parameters of the proposed model.

A hybrid model combining genetic programming and simulated annealing to undertake problems in biochemical network modelling and optimization is proposed by Rausanu et al. The hybrid model is used to optimize both the network topology and reaction rates of biochemical networks. The focus is on automatic identification of network structures and their corresponding kinetic constants. Genetic programming is used for generation of network topologies while simulated annealing is employed for optimization. Promising results from a series of simulation studies are reported.

In Girsang et al., the Ant Colony Optimization (ACO) algorithm is used together with the Analytical Hierarchy Process (AHP) in handling multi-criteria decision-making problems. The ACO algorithm is employed to solve issues pertaining to inconsistency of comparison matrices in AHP. The matrix elements are used as the path for the colony of ants to construct their tour such that optimal matrices that satisfy the consistency requirements can be found. A series of experimental studies reveal that the

resulting ACO-based model is useful to overcome the inconsistent pairwise weight matrices problem in AHP.

To tackle feature selection problems in data classification, a combined wrapper framework with a two-step methodology is proposed by Stanczyk. A pre-processing step that uses a simple wrapper to rank the characteristic features through the sequential backward elimination method is first conducted. The next step uses a predictor to exploit the resulting ordering of features and to reduce the number of features. Applicability of the proposed model to author attribution in the field of stylometry, a branch of science that involves the analysis of writing styles, is demonstrated.

A patent time series processing component with trend identification functionality is proposed by Chen et al. A technology intelligence framework is constructed. The piecewise linear representation method is used in the framework to generate and quantify the trend of patent publication activities. As such, the trend turning points can be identified, which in turn provides the trend tags to the existing text mining component. The proposed framework allows both text-based and time-based knowledge to be combined so that useful technical insight can be provided to aid decision-making in the domain of technology strategy-making support.

An approach to modelling customer behaviours using a Hidden Markov Model coupled with radio frequency identification and point of sales data is proposed by Sano and Yada. Information on sales areas and the number of bargain products are used as covariates of the proposed customer model. The attractiveness of each sales area and the effects of bargain sales pertaining to each customer's behaviour are then quantified. The proposed approach enables re-configuration of the sales areas in order to attract customers. In addition, effective bargain sales strategies can be formulated based on information pertaining to the optimal quantity of bargain products.

The guest editors are grateful to the authors for contributing their articles and the reviewers for reviewing the articles. Thanks are due to Professor John MacIntyre for the opportunity in organizing this special issue and the editorial team for their assistance, especially Preethi Vijayakumar's excellent assistance in keeping the review process on track. We appreciate it very much.