

Guest editorial: New trends in data pre-processing methods for signal and image classification

Kemal Polat¹ · Hariharan Muthusamy³ · Rajendra Acharya² · Yanhui Guo⁴

Published online: 10 September 2017
© The Natural Computing Applications Forum 2017

A special issue of the Neural Computing and Applications (NCAA) is dedicated to “New trends in data pre-processing methods for signal and image classification.”

Data pre-processing is crucial for effective data mining. Low-quality data usually produce inaccurate and unpredictable outcomes. Today’s real-world data are greatly vulnerable to noise and getting lost due to either large data size or the sources of origin. Real-world data are often inconsistent and incomplete, and are possible to have several errors. These poor-quality data will result in poor-quality mining outcomes. Data pre-processing enhances the data standard and subsequently aids to refine the value of data mining outcomes. Data pre-processing performs certain processing on raw original data to prepare it for further processing or analysis. In short, data pre-processing

prepares original raw data for further processing. Data pre-processing converts the data into a form acceptable easily for further processing by the user.

Data pre-processing methods have various applications in signal and image processing such as signal and image pre-processing, feature extraction, feature dimension reduction, classification and idea or information extraction. Data pre-processing can be applied to (1) remove noise part from signal or image, otherwise increase the consequence of reaching incorrect conclusions using raw signal or image data, (2) extract features, reduce the dimensionality of the signal or image and maintain as much significant information as possible, and (3) develop concept formation from the signal or image data. The main focus of this issue is the application of soft computing on signal and image classification using different and new data pre-processing methods on the following problems: signal detection, image detection, personal identification systems, iris recognition, face recognition, biomedical signal and image classification.

Totally 57 papers were submitted and sent out for peer review. The peer review process was conducted according to the standing editorial policy of Neural Computing and Applications which results in the final versions of the 20 papers accepted and included in this special issue. A summary of the papers appears below.

R. Alejo et al. presented an improved dynamic sampling approach (ISDSA) for facing the multi-class imbalance problem. ISDSA uses the mean square error (MSE) and a Gaussian function to identify the best samples to train the neural network. Results show that ISDSA makes better exploitation of the training dataset, improves the MLP classification performance, and deals the multi-class imbalance problem successfully. In addition, results indicate that the proposed method is very competitive in terms

✉ Kemal Polat
kemal_polat2003@yahoo.com

Hariharan Muthusamy
wavelet.hari@gmail.com; hari@unimap.edu.my

Rajendra Acharya
aru@np.edu.sg

Yanhui Guo
yguo56@uis.edu; guoyanhui@gmail.com

¹ Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Abant Izzet Baysal University, 14280 Bolu, Turkey

² Department of ECE, Ngee Ann Polytechnic, 535, Clementi Road, Singapore 599489, Singapore

³ Department of Biomedical Engineering, School of Bio-Engineering, SRM University (Under Section 3 of UGC Act 1956), SRM Nagar, Kattankulathur, Kancheepuram District 603 203, Tamil Nadu, India

⁴ Department of Computer Science, University of Illinois at Springfield, One University Plaza, Springfield, IL, USA

of classification performance with respect to classical over-sampling methods (also, combined with well-known features selection methods) and other dynamic sampling approaches; even in training time and size, it is better than the over-sampling methods. Akhan Akbulut *et al.* proposed a cloud-based book recommendation service that uses a Principal Component Analysis-Scale Invariant Feature Transform (PCA-SIFT) feature detector algorithm to recommend book(s) based on a user-uploaded image of a book or collection of books. The high dimensionality of the image is reduced with the help of a principal component analysis (PCA) pre-processing technique. When the mobile application user takes a picture of a book or a collection of books, the system recognizes the image(s) and recommends similar books. The computational task is performed via the cloud infrastructure. Experimental results show that the PCA-SIFT-based cloud recommendation service is promising; additionally, the application responds faster when the pre-processing technique is integrated. The proposed generic cloud-based recommendation system is flexible and highly adaptable to new environments. U. Raghavendra *et al.* proposed an automated screening method for classifying normal and congestive heart failure (CHF) echocardiographic images affected due to dilated cardiomyopathy (DCM) using variational mode decomposition (VMD) technique. The texture features are extracted from variational mode decomposed image. These features are selected using particle swarm optimization (PSO) and classified using support vector machine (SVM) classifier with different kernel functions. We have validated our experiment using 300 four-chamber echocardiography images (150 normal and 150 CHF) obtained from 50 normal and 50 CHF patients. The proposed approach yielded maximum average accuracy, sensitivity, and specificity of 99.33, 98.66, and 100%, respectively, using *ten* features. Thus, the developed diagnosis system can effectively detect CHF in its early stage using ultrasound (US) images and aid the clinicians in their diagnosis. S. Senthil Kumar *et al.* applied uncertainty (i.e., rough set)-based pattern classification techniques for UCI healthcare data for the diagnosis of diseases from different patients. In this study, covering rough set model (CRS)-based classification (i.e., proposed pattern classification approach) applied for UCI healthcare data. Proposed CRS gives effective results than delicate pattern classifier model. A. Janani and M. Sasikala investigated different signal pre-processing approaches for enhancing the quality of functional near-infrared spectroscopy (fNIRS). Various signal pre-processing approaches such as band-pass filtering, correlation-based signal improvement (CBSI), median filtering, Savitzky–Golay filtering, wavelet denoising, and independent component analysis (ICA) have been investigated. The results show that the application of such filtering algorithms for

functional fNIRS signal could effectively classify motor tasks to develop BCI applications. Fatih Kayaalp *et al.* designed a wireless sensor networks (WSN)-based real-time monitoring system to detect and locate the leaks on multiple positions on water pipelines by using pressure data. Three multi-label classification methods such as the random *k*-label sets (RAkELd), binary relevance *k*-nearest neighbors (BRkNN) and BR with SVM have been used. From the results, they recommended that multi-label classification methods can be used for the detection and localization of the leaks in the pipeline systems successfully. Manpreet Kaur *et al.* proposed an approach for feature selection using local searching (sequential backward selection and mutual information maximization algorithm) and global optimization techniques (genetic algorithm, differential evolution or particle swarm optimization). The results show that the proposed approach provides an improvement in terms of both the classification accuracy and the computation time. Muhammed Kursad Ucar *et al.* presented a method for detection of respiratory arrests in obstructive sleep apnea (OSA) patients using features extracted from the photoplethysmography (PPG) and machine learning techniques. Different machine learning techniques such as *k*-nearest neighbors, radial basis function neural network, probabilistic neural network, multi-layer feedforward neural network, and ensemble classification method and their performances are compared. R. Sindhu *et al.* proposed sine–cosine algorithm (SCA) for feature selection with elitism strategy and new updating mechanism. The potency of improved SCA is compared with its basic SCA, genetic algorithm, and particle swarm optimization. Rajeev Sharma *et al.* proposed a new technique for automated classification of sleep stages based on iterative filtering of electroencephalogram (EEG) signals. Poincare plot descriptors and statistical measures are applied as input features for different classifiers in order to classify sleep stages. The classifiers such as naive Bayes, nearest neighbor, multilayer perceptron, C4.5 decision tree, and random forest are applied. The results show that the proposed method has provided better tenfold cross-validation classification accuracy than other existing methods. E. C Orosco and F. D Sciascio presented high-order statistics (HOS) cumulants (auto-, cross-, and full-joint third-order cumulants)-based techniques to classify myoelectric signals. A myoelectric control scheme and its experimental application are executed with normal and disabled subjects, reaching a classification rate of 90%, in average. Jothi Ganesan *et al.* presented tolerance rough set firefly-based quick reduct-based feature selection to choose the conspicuous features of medicinal information. The results show that the proposed method outperforms the current supervised feature selection techniques. Yanhu Guo *et al.* proposed a novel image segmentation algorithm

based on neutrosophic c-means clustering and indeterminacy filtering method. Both artificial and natural images were utilized to evaluate the performance of the proposed method. The experimental results show that the proposed algorithm has better performances quantitatively and qualitatively. Aysun Sezer et al. proposed a computer-based diagnosis (CBD) system to recognize normal and edematous humeral head images by using texture features derived from Hermite transform. The results show that the proposed system is a promising tool for classification of edematous and normal bone from PD-weighted MR images. Ismail Kirbas and Musa Peker proposed a new method that enables the determination of P and S wave arrival time in noisy recordings based on the hybrid usage of empirical mode decomposition (EMD) and Teager–Kaiser energy operator (TKEO) algorithms. The results show that the proposed system gives effective results in the automatic detection of P and S wave arrival time. Jinrong He et al. proposed an unsupervised feature selection based on decision graph. From statistical tests on the averaged classification accuracies over 16 real-life datasets, it is observed that the proposed method obtains better or comparable ability of discriminant feature selection compared with the state-of-the-art methods. Erkan Deniz proposed artificial neural network-based maximum power point tracking (MPPT) algorithm for a solar permanent magnet synchronous motor (PMSM) drive system used without a

boost converter and batteries. The use of three-phase PMSM presents more efficient solutions to the trading solar systems with DC motor or induction motor. Thus, an effective solar system is achieved. U. Rajendra Acharya et al. proposed empirical mode decomposition (EMD) for an automated identification and classification of normal and CHF using heart rate variability (HRV) signals. In this work, HRV signals are subjected to EMD to obtain intrinsic mode functions (IMFs). From these IMFs, thirteen nonlinear features are extracted. The proposed automated technique is able to identify the person having CHF, and this method may act as a valuable tool for increasing the survival rate of many cardiac patients. Mehmet Dursun proposed a new approach-based correlation and discrete wavelet transform-based rule to eliminate EOG artifacts from the sleep EEG signals. An improvement about 8.03% in classification accuracy with regard to the un-cleaned EEG signals is achieved. Wu Ziheng et al. proposed an improved fuzzy c-means clustering algorithm (FCM) in which an adaptive weight vector and an adaptive exponent are introduced and the optimal values of the fuzziness parameter and adaptive exponent are determined by simulated annealing (SA) and particle swarm optimization (PSO). The results demonstrate that the proposed algorithm can avoid local optima and significantly improve the clustering performance.