



Recent advances of statistics in computational intelligence (RASCI)

Muhammed J. A. Patwary¹ · James N. K. Liu² · Honghua Dai³

Published online: 18 December 2017

© Springer-Verlag GmbH Germany, part of Springer Nature 2017

Computational intelligence is usually considered to be the capability of a computer to learn a specific task from data or experimental observation. Computational intelligence plays a pivotal role in solving real-world complex problems. One of the main objectives of computational intelligence models is to discover the statistical disciplines hidden in the data sets. The most significant theoretical fundamental in developing different computational intelligence models is considered to be statistical analysis. There is a continuous scientific attempt to get better approaching into the cryptic information underneath the huge stack of statistical model and data that we come across to solve a specific problem. As a result, there has been a major change towards quantitative analysis of statistical methods as well as data through various computational approaches.

Computational approaches that have been extremely popular and found significant application include Neural Network (NN) which provides with a representation framework for statistical construct, regression models that are based on pure statistical theory, Support Vector Machine (SVM) is based on framework of statistical learning, Bayesian method is also regarded as part of statistics, etc. On the whole, all the major tasks of computational intelligence can be represented as statistical methods and these methods can be explained from a statistical point of view.

This special issue is an effort to bring together appealing works that use advanced statistical theory and methodology for handling the computational intelligence. After a strict blind peer review, we selected 12 papers for this issue. The following is a brief introduction to the 12 selected papers.

The paper “Experimental study on generalization capability of extended naïve Bayesian classifier” by Si-si Chen, Juan-juan Cao, Li-li Gan, Qing-ge Song and Di Han, mainly investigates the Extended naïve Bayesian classifier (ENBC’s) generalization capability based on the real classification datasets. The analysis shows that ENBC is not stable and its aggregation weight is sensitive to the order of training samples and ENBC indeed has higher generalization capability than the existing NBC.

The paper titled “NBWELM: naïve Bayesian-based weighted extreme learning machine” authored by Jing Wang, Lin Zhang, Juan-juan Cao and Di Han discusses the generalization performance of imbalance classification problems. This paper reported that WELM strengthens the influence of outliers to some extent because of weight assignment strategy. The authors propose a Naïve Bayesian-based.

Weighted Extreme Learning Machines (NBWELM), in which the weight is determined with the flexible naïve Bayesian (FNB) classifier. While comparing with the existing WELMs, NBWELM achieves the better classification accuracy as a result of effective discrimination and handling to the outliers in training samples. The experimental results reported in this paper demonstrate that the proposed method is able to improve the generalization capability of classification problems.

The paper titled “A novel online ensemble approach to handle concept drifting data streams: diversified dynamic weighted majority” authored by Parneeta Sidhu and M. P. S. Bhatia, discusses an online ensemble approach, a diversified dynamic weighted majority (DDWM) to classify new query data instances which have varying conceptual distributions. The authors keep two different sets of weighted ensembles which differ in their level of variety. Based on the classification accuracy, an expert is updated or removed in either of the ensembles. Depending on the final global prediction

✉ Muhammed J. A. Patwary
jamshed@szu.edu.cn

James N. K. Liu
james.liu@polyu.edu.hk

Honghua Dai
hdai@deakin.edu.au

¹ Big Data Institute, College of Computer Science and Software Engineering, ShenZhen University, Shenzhen, China

² Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

³ School of Information Technology, Deakin University, Melbourne Campus, 21 Burwood Highway, Burwood, VIC 3125, Australia

of the algorithm and the global prediction of the ensemble for any data instance a new expert is added. No matter what the size of the dataset, type of drift or presence of noise, the experimental study conducted in this paper shows that DDWM provides very high accuracy to classify new data instances.

The paper “The effect of online reviews on e-tailers’ pricing in a dual-channel market with competition” authored by Kaiqiang Guo, Hongwei Wang, Yuan Song and Zhanqi Du, investigates the problem of fixing the price for maximal profits on the basis of online reviews of a product. The authors build a review-involved pricing model taking utility theory and game theory into account. To construct the model the authors first grouped their consumers into three categories based on the channel of information they refer for purchasing decisions. The categories are as follows: (1) online information only (2) online and offline information and (3) offline information only. To show the association between optimized revenue and other features such as the additional cost of online shopping, a number of reviews and applause rate of reviews a numerical simulation is conducted in this study. The experimental results show that the rate of applause is positively correlated with the pricing and profit in e-tailers. Besides, the paper reported, by reducing additional cost of online shopping e-tailers can improve their revenue.

The paper “Sentimental feature selection for sentiment analysis of Chinese online reviews” by Lijuan Zheng, Hongwei Wang, and Song Gao studies the feature selection process for sentiment analysis of Chinese online reviews. First of all, N-char-grams and N-POS-grams are chosen to be the possible sentimental features. Secondly, to select the feature subsets, improved documents frequency method is used. Then, Boolean weighting method is employed to calculate the feature weight. Finally, experiments are conducted based on online reviews of the mobile phone. To test the significance of experimental results, Chi square test is performed. The experimental results show that the sentiment analysis of Chinese online reviews finds higher accuracy when taking 4-POS-grams as features and it is also reported that low order N-char-grams can achieve better performance than high order N-char-grams features. Moreover, noteworthy improvement in sentiment analysis of Chinese online reviews can be achieved through improved document frequency.

In the paper “Information-decomposition-model-based missing value estimation for not missing at random dataset” authored by Shigang Liu, Honghua Dai and Min Gan, an information decomposition imputation (IDIM) algorithm is suggested where fuzzy membership function is used to address the missing value problem of dataset under the condition of not missing at random (NMAR). With the help of an example, the proposed IDIM algorithm is described first. Then, it is compared with some existing algorithms

for evaluation. Normal root means square error (NRMSE) and TP + TN are calculated to compare the accuracy of the proposed algorithm with existing algorithms. The experimental results show that the proposed algorithm has higher accuracy than the existing imputation approaches.

In the paper “A novel hybrid model using teaching–learning-based optimization and a support vector machine for commodity futures index forecasting” by Shom Prasad Das and Sudarsan Padhy, Combining support vector machine (SVM) with teaching–learning-based optimization (TLBO), a hybrid method is developed, which is used to forecast a financial derivatives instrument (the commodity futures contract index). Unlike other optimization methods, the proposed SVM-TLBO model does not use user-specified control parameters. The experimental results demonstrate that the proposed model outperforms than the particle swarm optimization (PSO) + SVM hybrid and standard SVM models. When compared with standard SVM regression proposed model shows substantial improvements.

The paper “Application of fuzzy analytic hierarchy process in the risk assessment of dangerous small-sized reservoirs” by Ma Jing, Yang Jie, Li Shou-yi and Wang Lu, proposes a fuzzy AHP risk assessment method which is suitable for small-sized reservoirs. The method is based on statistical analysis of the wracked small-sized reservoirs. The proposed method is used to identify risk factors very effectively. Two engineering cases are used to verify the applicability and effectiveness of this method. The experimental results show that the proposed method clearly assesses the level of risk of dangerous small-sized reservoirs.

In the paper “Towards flexible management in enterprise network: an enhanced routing protocol” authored by Shu Yang, Yuanyuan He, Mingwei Xu and Yong Jiang, a new routing protocol called TwoD-IP routing is proposed to facilitate enterprise network management. With the help of TwoD-IP routing protocol, flexible management can be achieved by providing different views on the same network for different users. Authors design the routing protocol algorithm and implement it on a quagga-based router. Eventually, authors evaluate their proposed mechanism using simulations with BRITe generated topologies. The experimental results show that the proposed mechanism substantially reduces the network configuration complexities.

The paper “Universal knowledge discovery from big data using combined dual-cycle” authored by Bin Shen, proposes a dual-cycle methodology with three types of cycle-driven Universal Knowledge Discovery (UKD) process, i.e., big-data-cycle-driven, mechanism-cycle-driven and combined-dual-cycle-driven mining. To demonstrate the effectiveness of the proposed methodology, a case study is also conducted in this paper. This paper can be used for the further development of universal knowledge discovery from big data.

In the paper “Robust image watermarking scheme in lifting wavelet domain using GA-LSVR hybridization” authored by Rajesh Mehta, Navin Rajpal, Virendra P and Vishwakarma, an image watermarking scheme is proposed in lifting wavelet domain using the grouping of genetic algorithm and Lagrangian support vector regression. Nonoverlapping blocks of the approximate image are selected based on the fuzzy entropy. To train the LSVR, a dataset is formed from the important and stable feature of these selected blocks. The output of LSVR is used to insert the twisted image of the watermark. Each watermark bit is embedded into the selected blocks of an approximate image using optimized multiple scaling factors obtained by the genetic algorithm. The experimental results of the proposed scheme compared with the existing techniques for several image processing operations like JPEG compression, an addition of noise, sharpening, contrast enhancement, gamma correction, scaling and cropping show the high imperceptibility and robustness. The watermarking scheme developed in this paper shows resistance against common signal processing

attacks and geometric attacks such as scaling and cropping but lacks to show resistance against rotation and translation attacks.

The paper “Colour image segmentation with histogram and homogeneity histogram difference using evolutionary algorithms” authored by Sushil Kumar, Millie Pant, Manoj Kumar and Aditya Dutt proposes a fitness function based on pixel-by-pixel values for color image segmentation and optimizes these values through evolutionary algorithms such as differential evolution (DE), particle swarm optimization (PSO) and genetic algorithms (GA). Experimental results show that DE outperforms PSO and GA based on computational time and quality of the segmented image.

I would like to extend my sincere thanks to all the authors and reviewers for their enormous contributions to this special issue section. I would also like to thank, Professor Xizhao Wang, the Editor-in-Chief of International Journal of Machine Learning and Cybernetics, for his continuous support in organizing this category.