



Advances in applying soft computing techniques for big data and cloud computing

B. B. Gupta¹ · Dharma P. Agrawal² · Shingo Yamaguchi³ · Michael Sheng⁴

Published online: 16 October 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Soft computing is a set of methods which are highly consistent at the target level and distinctive at the method level. Soft computing mainly includes rough set, fuzzy logic, neural network, genetic algorithm, particle swarm algorithm, chaos and other parts, while compared with traditional hard computing, soft computing is closer to the objective things themselves, and human thinking more closely with more intelligent to complete the solution to the problem. Particularly, soft computing for big data and cloud computing is profoundly affecting the global business and world economy, becoming one of the major driving forces for industrial prediction, classification and better human life (Plageras et al. 2018; Psannis et al. 2018; Cho 2002; Stergiou et al. 2016). With the rapid development and widespread application of big data and artificial intelligence, large, complex and diverse example datasets widely appear in model building and data classification for meeting various requirements. At the same time, the massive datasets could contain sensitive information that the data owners are unwilling to contribute their resources, which creates a certain barrier on the data collection. In the data outsourcing and multi-party computation, the encryption and other technologies are used to protect the privacy of sensitive information. However, traditional machine learning and soft computing techniques cannot be used to classify encrypted data directly, and there are still many challenges that need to be addressed such as in machine learning and soft computing algorithms, learn ability, multi-party computation, privacy preserving, outsourcing computation and computing security. Moreover, as in every new technology, some challenges face the vision of the big data and cloud computing, which are the administrative policies and

security concerns (i.e., secure data storage, secure computation, network security, data privacy, usage privacy, location privacy; Gupta et al. 2016, 2017a, 2018; Jain and Gupta 2016).

These considerations have led to this special issue a venue for big data and cloud computing researchers applying soft computing techniques around the world to share their state-of-the-art research and development, particularly on advances computing technologies and related areas that could be beneficial to handle various research challenges in this domain. This has helped us to collect high-quality articles that reported recent research advances in applying soft computing techniques for big data and cloud computing, covering various topics of interest. This special issue contains 20 papers dealing with different aspects of big data and cloud computing research issues and other related areas (Zhang et al. 2016; Lee et al. 2017; Gupta et al. 2017b; Negi et al. 2013). The first article entitled, “An efficient and secure searchable public key encryption scheme with privacy protection for cloud storage” authored by Debiao He et al., presents an efficient and secure searchable public key encryption with privacy protection (SPE-PP). Authors then provide a concrete construction of SPE-PP that uses the Diffie–Hellman shared secret key, and authors prove that it can resist these attacks. Both the theoretical analysis and the experimental results show that proposed scheme achieves strong security along with high efficiency. The second article entitled, “Dimension reduction in radio maps based on the supervised kernel principal component analysis” authored by Bing Jia et al., presents a supervised approach to take the full advantage of the information available for building radio maps, i.e., location labels attached to fingerprints, to compress original radio maps. Specifically, in the offline phase, the supervised kernel PCA (SKPCA) method is employed to derive a nonlinear and optimal embedding in a low-dimensional subspace; in the online phase, any sample vector containing received signal strengths can be projected onto the optimal subspace in real time for further localization processing. Experiments are carried out not only in a real environment but also

✉ B. B. Gupta
bbgupta@nitkr.ac.in

¹ National Institute of Technology Kurukshetra, Kurukshetra, India

² University of Cincinnati, Cincinnati, USA

³ Yamaguchi University, Yamaguchi, Japan

⁴ Macquarie University, Sydney, Australia

using an open dataset. It is shown that the compressed radio maps based on SKPCA have much smaller sizes than their original radio maps, but achieve similar localization performance and significantly outperform the other two popular PCA-based unsupervised dimension reduction methods, i.e., PCA and PCA-MLE. The third article entitled, “Concurrent workflow budget-and deadline-constrained scheduling in heterogeneous distributed environments” authored by Naqin Zhou et al., presents a new dynamic scheduling algorithm for concurrent workflows. This algorithm proposes a uniform ranking that considers the time and costs for both workflows and workgroups to assign priorities for tasks. In the resource selection phase, it controls the resource selection range for each task based on an optimistic budget for the current task and selects resources for the current task according to a defined bi-factor. The experimental results show that the proposed algorithm outperforms the existing algorithms in both consistent and inconsistent environments.

The fourth article entitled, “Verifiable privacy-preserving single-layer perceptron training scheme in cloud computing” authored by Xiaoyu Zhang et al., focuses on designing a novel privacy-preserving single-layer perceptron training scheme which supports batch patterns training and verification for the training results on the client side. In addition, adopting classical secure two-party computation method, authors design a novel lightweight privacy-preserving predictive algorithm. Both two participants learn nothing about other’s inputs, and the calculation result is only known by one party. Detailed security analysis shows that the proposed scheme can achieve the desired security properties. Authors also demonstrate the efficiency of the proposed scheme by providing the experimental evaluation on two different real datasets. The fifth article entitled, “Incentive-driven attacker for corrupting two-party protocols” authored by Yilei Wang et al., presents a model that can be considered as the extension of the seminal work presented by Groce and Katz (Annual international conference on the theory and applications of cryptographic techniques, Springer, Berlin, pp 81–98, 2012), while making significant contribution in addressing the corruption of two parties in two-party protocols. To the best of authors’ knowledge, this is the first time to model the corruption of both parties in two-party protocols. The sixth article entitled, “A new container scheduling algorithm based on multi-objective optimization” authored by Bo Liu et al., defines a metric method for every key factor and establishes a scoring function for each one and then combines them into a composite function. The experimental results show that compared with the other three well-known algorithms: Spread, Binpack and Random, Multiopt increases the maximum TPS by 7% and reduces the average response time per request by 7.5% while consuming roughly same allocation time.

The seventh article entitled, “Ciphertext retrieval via attribute-based FHE in cloud computing” authored by Yong Ding et al., presents an attribute-based fully homomorphism encryption scheme with LSSS matrix and a ciphertext-based retrieval scheme on this basis. Firstly, the retrieval scheme supports fine-grained-cum-flexible access control along with “Query-Response” mechanism to enable users to efficiently retrieve desired data from cloud servers. Secondly, the scheme supports considerable flexibility to revoke system privileges from users without updating the key client, which reduces the burden of the client greatly. Security analysis shows that the proposed scheme can resist collusion attack. A comparison of the performance with existing CP-ABE schemes indicates that proposed scheme reduces the computation cost greatly for users and guarantees only the owners of the evaluation; keys can perform ciphertext retrieval, which is also one of the main advantages of the proposed scheme. The eighth article entitled, “Efficient and secure big data storage system with leakage resilience in cloud computing” authored by Yinghui Zhang et al., presents an efficient and secure big data storage system in cloud computing, in which a leakage-resilient encryption scheme serves as the main ingredient. What’s more, proposed formal security proof analysis indicates that the proposed scheme can ensure users’ data privacy even if the partial key is leaked in cloud computing. Finally, the leakage resilience analysis indicates that the leakage ratio in the proposed scheme can reach roughly 1/3 and is higher than other schemes. Performance comparisons show the practicability of the proposed scheme for big data security in cloud computing. The ninth article entitled, “A self-feedback strategy differential evolution with fitness landscape analysis” authored by Ying Huang et al., presents a new self-feedback strategy differential evolution (SFSDE) algorithm based on fitness landscape analysis of single-objective optimization problem. In the SFSDE algorithm, in the analysis of the fitness landscape features of fitness distance correlation, a self-feedback operation is used to iteratively select and evaluate the mutation operators of the new SFSDE algorithm. Moreover, mixed strategies and self-feedback transfer are combined to design a more efficient DE algorithm and enhance the search range, convergence rate and solution accuracy. Finally, the proposed SFSDE algorithm is implemented to optimize soil water textures, and the experimental results show that the proposed SFSDE algorithm reduces the difficulty in estimating parameters, simplifies the solution process and provides a novel approach to calculate the parameters of the Van Genuchten equation. In addition, the proposed algorithm exhibits high accuracy and rapid convergence and can be widely used in the parameter estimation of such nonlinear optimization models.

The tenth article entitled, “A comparative performance analysis of evolutionary algorithms on k -median and facil-

ity location problems” authored by Xue Peng et al., presents an approach to investigate the approximation performance of two evolutionary algorithms for the k-median and facility location problems from a theoretical perspective. For k-median problem, authors show that the evolutionary algorithms with standard bit mutation (SBM) operator can obtain approximation ratios 5 and $3 + 2/p$ in expected polynomial time, while the evolutionary algorithm with somatic contiguous hypermutation (CHM) operator cannot obtain them. For facility location problem, authors show that the evolutionary algorithms with SBM operator can obtain approximation ratio 3 in expected polynomial time, while the evolutionary algorithm with CHM operator cannot obtain it. Further, authors prove that on a facility location instance the evolutionary algorithm with CHM operator greatly outperforms the evolutionary algorithm with SBM operator. The eleventh article entitled, “A virtual cluster embedding approach by coordinating virtual network and software-defined network” authored by Yusong Tan et al., presents a coordinated VCE approach, called CoVCE. The approach combines virtual network with SDN to form virtual SDN and weakens slightly the constraints of VM placement. The network centrality, correlation property and resource fragmentation are optimized with multiple objectives to receive more requests, increase the throughput and decrease network delay and runtime. The CoVCE method integrates not only the logical topology of virtual machines in the cluster but also the relationship with virtual switch, virtual link and control services. This does not only optimize the placement location of virtual components, but also allow the virtual resources to migrate upon virtual cluster requests. The method further improves the utilization of physical resources and reduces resource fragmentation. According to extensive simulation and emulation experiments and comparison with correlative algorithms, CoVCE effectively reduces network delays, offers a higher embedding efficiency, improves user experiences and, to some extent, also improves the revenue/cost ratio and throughput. The twelfth article entitled, “Online handwritten signature verification using feature weighting algorithm relief” authored by Li Yang et al., presents a novel writer-dependent online signature verification technique based on Relief. For each user, the proposed verification scheme falls into two phases: the training phase and the test phase. In the training phase, authors select a signature as the base signature from real signatures and construct two different kinds of signature pairs: the real signature pairs and the forged signature pairs. Each pair of real signatures consists of two real signatures, and each pair of forgeries is made up of a real signature and a forgery. A group of combined features are defined for each pair of signatures. Then, authors select more stable combined features by means of the Relief algorithm and compute a matrix used as a matching template, which contains the value of each combined feature for each pair of

signatures. In the test phase, the classification of the pair of signatures composed of the test signature and the base reference signature is determined by the K-nearest neighbor. Experiments are conducted on the authors own database and SVC2004, and the results indicate that the propose method shows lower FAR and FRR than do traditional methods on the same databases.

The thirteenth article entitled, “Face pose estimation based on kernelized maximum separability” authored by Xiao-Zhang Liu et al., presents a face pose estimation method based on maximum separability of pose templates and incorporate the kernel technique to help model the nonlinear distribution of face patterns under pose variations. The proposed method involves a kernel subspace projection phase and the nearest neighbor classification. Experimental results on the CMU PIE face database and the UMIST database show that the proposed method can achieve high pose estimation performance. The fourteenth article entitled, “External Archive Matching Strategy for MOEA/D” authored by Feng Wang et al., presents an external archive matching strategy which selects solutions’ most matching archive solutions as parent solutions. The offspring solutions generated by this strategy can maintain a good convergence ability. To balance convergence and diversity, a perturbed learning scheme is used to extend the search space of the solutions. The experimental results on three groups of test problems reveal that the solutions obtained by MOEA/D-EAM have better convergence and diversity than the other four state-of-the-art algorithms. The fifteenth article entitled, “Towards Biology-inspired Solutions for Routing Problems of Wireless Sensor Networks with Mobile Sink” authored by Nan Jiang et al., presents draws an inspiration from the single-celled organism—slime mold *physarum polycephalum* of foraging process and combined with the traditional clustering algorithm to design routing scheme. The application of traditional clustering algorithm and mobile sink technology in WSNs is on the basis of this scheme design. The simulation results show that compared with the existing routing protocol, the new scheme can effectively improve the self-adaptation of the network, reduce the energy consumption and improve the overall performance of the network under the complicated network environment. The sixteenth article entitled, “A generic optimization method of multivariate systems on graphic processing units” authored by Guohong Liao et al., presents a generic GPU acceleration framework for multivariate systems with various orders. The experiment results show that the proposed optimization method can effectively improve the performance of multivariate cryptosystems on GPUs.

The seventeenth article entitled, “Research on QoS Service Composition Based on Coevolutionary Genetic Algorithm” authored by Yuanzhang Li et al., presents a coevolu-

tionary genetic algorithm for web service composition based on quality of service (QoS), which fully considers the individual relationships among populations. The real coding method is adopted to solve the service selection problem based on QoS, so that the negative effect of the long length of chromosomes in the algorithm is avoided. Moreover, in view of the difficulty of determining the weight of each QoS attribute in web services, authors propose to use the entropy method to determine the weights of each one. Compared with the traditional genetic algorithm, the experimental results show that the proposed algorithm converges faster in the service composition, and the fitness of the optimal solution is higher. The eighteenth article entitled, “A Bibliometric Analysis of Text Mining in Medical Research” authored by Tianyong Hao et al., systematically reviews the existing academic research outputs of the field from Web of Science and PubMed by using techniques such as geographical visualization, collaboration degree, social network analysis and topic modeling analysis. Specifically, publication statistical characteristics, geographical distribution, collaboration relations and research topic are quantitatively analyzed. This study contributes to the text mining enhanced medical research field in a number of ways. First, it provides the latest research status for researchers who are interested in the field through literature analysis. Second, it helps scholars become more aware of the research subfields through hot topic identification. Third, it provides insights to researchers engaging in the field and motivates attention on the relevant research. The nineteenth article entitled, “An Efficient Index Structure for Distributed k Nearest Neighbours Query Processing” authored by Min Yang et al., presents an efficient distributed solution for k-NN query over moving objects to tackle the increasingly large scale of data. This approach includes a new grid-based index called block grid index (BGI) and a distributed k-NN query algorithm based on BGI. There are three advantages of this approach: (1) BGI can be easily constructed and maintained in a distributed setting; (2) the algorithm is able to return the results set in only two iterations; (3) the efficiency of k-NN query is improved. The efficiency of proposed solution is verified by extensive experiments with millions of nodes. The twentieth article entitled, “Information retrieval methodology for aiding scientific database search” authored by Samuel Marcos-Pablos et al., presents an iterative methodology that applies text mining and machine learning techniques to a downloaded corpus of abstracts from scientific databases, combining automatic processing algorithms with tools for supervised decision making in an iterative process sustained on the researchers’ judgement, so as to adapt, screen and tune the search output. The paper ends showing a working example that employs a set of developed scripts that implement the different stages of the proposed methodology.

We would like to express our special thanks to Prof. Antonio Di Nola, the Editor-in-Chief of *Soft Computing*, for their great support and efforts throughout the whole publication process of this special issue. Moreover, this special issue is due to encouragement of Prof. Vincenzo Loia, the Editor responsible for special issues who is instrumental in the organization process and to the Springer Journal Editorial Office for their continuous support to publish this special issue. Many individuals have contributed toward the success of this issue. Special thanks are due to dedicated reviewers who found time from their busy schedule to review the articles submitted in this special issue. In addition, we are also grateful to all the authors for submitting and improving their papers. This special issue presents some selected papers dealing with different aspects of big data and cloud computing research issues and other related areas and also emphasizes many open questions (Hossain et al. 2018; Wang et al. 2018).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Cho S-B (2002) Incorporating soft computing techniques into a probabilistic intrusion detection system. *IEEE Trans Syst Man Cybern C Appl Rev* 32(2):154–160
- Gupta BB, Agrawal DP, Yamaguchi S (2016) Handbook of research on modern cryptographic solutions for computer and cyber security. IGI Global Publisher, Hershey
- Gupta S et al (2017a) Detection, avoidance, and attack pattern mechanisms in modern web application vulnerabilities: present and future challenges. *Int J Cloud Appl Comput (IJCAC)* 7(3):1–43
- Gupta BB, Gupta S, Chaudhary P (2017b) Enhancing the browser-side context-aware sanitization of suspicious HTML5 code for halting the DOM-based XSS vulnerabilities in cloud. *Int J Cloud Appl Comput (IJCAC)* 7(1):1–31
- Gupta BB, Yamaguchi S, Agrawal DP (2018) Advances in security and privacy of multimedia big data in mobile and cloud computing. *Multimed Tools Appl* 77(7):9203–9208
- Hossain MS et al (2018) Cloud-assisted secure video transmission and sharing framework for smart cities. *Future Gener Comput Syst* 83:596–606
- Jain AK, Gupta BB (2016) A novel approach to protect against phishing attacks at client side using auto-updated white-list. *EURASIP J Inf Secur* 2016(1):9
- Lee K et al (2017) A comparative evaluation of atrial fibrillation detection methods in Koreans based on optical recordings using a smartphone. *IEEE Access* 5:11437
- Negi P, Mishra A et al (2013) Enhanced CBF packet filtering method to detect DDoS attack in cloud computing environment. *Int J Comput Sci Issues (IJCSI)* 10(2):142

- Plageras AP, Psannis KE et al (2018) Efficient IoT-based sensor BIG Data collection-processing and analysis in smart buildings. *Future Gener Comput Syst* 82:349–357
- Psannis K, Stergiou C, Gupta BB (2018) Advanced media-based smart big data on intelligent cloud systems. *IEEE Trans Sustain Comput* 3:1
- Stergiou C et al (2016) Secure integration of IoT and cloud computing. *Future Gener Comput Syst* 78:964–975
- Wang Y, Liu Q, Hou HD et al (2018) Big data driven outlier detection for soybean straw near infrared spectroscopy. *J Comput Sci* 26:178–189
- Zhang Z et al (2016) Social media security and trustworthiness: Overview and new direction. *Future Generation Computer Systems*. Elsevier, Amsterdam

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.