Editorial: Deep Mining Big Social Data

Xiaofeng Zhu^{1,2} · Gerard Sanroma³ · Jilian Zhang⁴ · Brent C. Munsell⁵

Published online: 13 September 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

The internet revolution has made information acquisition easy and cheap so that it has been producing massive Web/social data in our real life. The emergence of big social media has lead researchers to study the possibility of their exploitation in order to identify hidden knowledge. However, a huge number of issues appear in obtained big social data [23, 24, 26, 28]. First, there are incomplete social data due to all kinds of reasons, such as security and private information. Second, the structure of social data is different, including structured data (e.g., social Web data), semi-structured data (e.g., XML data) and unstructured data (e.g., social networks). Third, the Web data are often high-dimensional. However, current computer techniques can only deal with structured, complete and moderate-sized-dimensional data. Moreover, current computer technologies can only mine the basic structure and are not capable of mining their natural complex structure (or deep structure). Hence, there is a huge gap between existing technologies and the real requirements of actual big social data. In this case, deep mining of big social data (such as data preprocessing, deep pattern discovery, pattern fusion, and outlier/noise detection) stands as an interesting promise to relief such a gap [4, 8, 22, 25, 27].

In [7], Komarasamy et al. proposed a multi-phase scheduling method to deal with parallel jobs in hierarchical model. The proposed method includes job preprocessing, prioritization,

✉ Xiaofeng Zhu
seanzhuxf@gmail.com

Gerard Sanroma
gerard.sanroma-gueell@dzne.de

Jilian Zhang
Jilian.z.2007@smu.edu.sg

Brent C. Munsell
munsellb@cofc.edu

¹ Guangxi Normal University, Guangxi, China

² Massey University Auckland Campus, Auckland, New Zealand

³ German Center for Neurodegenerative Diseases, Bonn, Germany

⁴ Jinan University, Guangzhou, China

⁵ Department of Computer Science at the College of Charleston, Charleston, USA

and scheduling among nodes in cloud storage. Moreover, it uses intermediate idle nodes in preprocessing and batch processing to avoid starvation and to mitigate unwanted delay.

In [5], Huang et al. proposed an intelligent trading method based on Naive Bayes algorithm and AdaBoost algorithm. This method first employs dual clustering to detect transaction patterns, and then uses the discovered trading patterns to predict market trends based on the Naive Bayes algorithm. Finally, the Adaboost algorithm is used to promote the naive Bayes classifier as a robust classifier and compared with the other four existing algorithms.

In [6], Huang et al. proposed a CAD system to overcome opacity in new CAD systems. The experimental results show that the proposed system has good performance in the diagnosis of mammary glands and can effectively identify whether it is a benign breast tumor in breast diseases.

In [13], Rao et al. proposed an active learning scheme to utilize the labeled and unlabeled images to build the initial Support Vector Machine (SVM) classifier for image retrieving.

In [15], Wan et al. proposed the development of a suitable coding framework and provided developers with the right coding skills. In this process, the paper designed a probabilistic expert ranking model, added the regularity of the project as a regularization of the graph to the expert ranking, and added the correlation propagation graph.

In [21], Zhang et al. proposed a new feature selection method for data classification that efficiently combines the discriminative capability of features with the ridge regression model. It first uses linear judgement analysis to establish the global structure of training data to help identify discriminative features. And then, the ridge regression model is used to evaluate the feature assurance and discriminant information so as to obtain a representative sparse matrix. Finally a new subset of the selected features is applied to the linear support vector machine for data classification.

In [18], Wen et al. proposed a spectral clustering method especially to process the high-dimensional data, via first using an affinity matrix learning method to learn a high-quality affinity matrix from the intrinsic feature space, and then utilizing the local PCA to clip the affinity matrix for solving the intersection problem. Finally, this paper employed a robust clustering method to conduct the clustering tasks directly on the affinity matrix, so that it can overcome the cluster-specification problem and the initialization sensitivity problem.

In [16], Wang et al. proposed a type of node similarity, based on the frequency of connections between nodes to describe the changeable relationships between entities over a period.

In [14], Sood proposed a new concept of free space fog to collect available free resources from all assigned jobs to eliminate Deadlock.

In [10], Liu et al. proposed a new data clustering algorithm based on potentiality model. The algorithm merges sub-clusters using clustering merging criteria to automatically terminate the clustering process.

In [9], Li et al. proposed a new problem in dynamic traffic networks. It first considers two dynamic transportation networks, which are the traffic software spatial network and the dynamic public transportation network, and then uses uncertain trajectory data to establish a spatial network of traffic areas.

In [19], Xie et al. proposed a method based on information theory to optimize the tag interaction with high efficiency. In order to generate a recommendation list, this paper applies probabilistic matrix decomposition techniques to predict user preferences and overcome level sparsity. It enhances level sparsity by embedding similar user and resource information.

In [29], Zhu et al. proposed a unsupervised feature selection method by embedding subspace learning regularization (PCA) into a feature selection framework.

In [3], Hu et al. proposed a popular route construction method named GRID based on collective knowledge. The experimental results on two real data sets show that this method is superior to the most advanced methods in terms of efficiency and efficacy.

In [2], Gu et al. proposed a strategy to resolve the ambiguity problems in short text categorization. By using Bi-directional Recurrent Neural Networks (Bi-RNN) and linear discriminant analysis (LDA), the proposed method can catch more contextual and latent semantic information for categorization. Apart from that, it uses topic model to enhance the neural network which represents short texts.

In [12], Pan et al. uses a multi-scale fully convolutional neural networks for regression based on density maps. It applies regression on the structured proximity space for patches to get larger value, and then uses convolutional regression networks to detect different kinds of cells based on features maps.

In [17], Wang et al. tried to apply self-representation of each feature to make the data set sparse, then used Frebenius norm and Locality Preserving Projection (LPP) as regularization term to avoid over-fitting and preserve local relations.

In [1], Gao et al. proposed a bin-based attack model to re-identify social individuals in social networks. Besides that, it also proposes the k-anonymity scheme to protect social individuals. Experiments showed that the proposed method is effective.

In [11], Menasria et al. concentrated on private information protection of Accelerometer-based activity recognition by leveraging the connection between irrelevant private information and relevant private information. By doing this, it can reduce the usage of the irrelevant information to protect our private information. Experiments illustrate that the proposed method can reduce the leakage of privacy.

In [20], Zhang tried to proposes a target-source framework to minimize the total cost by controlling another cost while minimizing one kind of cost scale. Besides, it also proposes a cost-sensitive learning model to help analyze the complex information. Experimental results showed that the proposed method works well on real medical data.

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

References

1. Gao, J., Ping, Q., Wang, J.: Resisting Re-Identification Mining on Social Graph Data. *World Wide Web Journal*, this issue
2. Gu, Y., Gu, M., Long, Y., Xu, M.G., Yang, Z., Zhou, J., Qu, W.: An Enhanced Short Text Categories Model with Deep Abundant Representation. *World Wide Web Journal*, this issue
3. Hu, G., Shao, J., Ni, Z., Zhang, D.: A graph based method for constructing popular routes with check-ins. *World Wide Web Journal*, this issue
4. Hu, R., Zhu, X., Cheng, D., He, W., Yan, Y., Song, J., Zhang, S.: Graph
5. Huang, Q., Kong, Z., Li, Y., Yang, J., Li, X.: Discovery of Trading Points Based on Bayesian Modeling of Trading Rules. *World Wide Web Journal*, this issue
6. Huang, Q., Zhang, F., Li, X.: A Novel Breast Tumor Ultrasonography CAD System Based on Decision Tree and BI-RAD Features. *World Wide Web Journal*, this issue
7. Komarasamy, D., Muthuswamy, V.: Priority Scheduling with Consolidation based BackFilling algorithm in Cloud. *World Wide Web Journal*, this issue
8. Lei, C., Zhu, X.: Unsupervised feature selection via local structure learning and sparse learning. <https://doi.org/10.1007/s11042-017-5381-7> (2017)

9. Li, J., Wang, Y., Zhong, Y., Guo, D., Zhu, S.: Aggregate Location Recommendation in Dynamic Transportation Networks. *World Wide Web Journal*, this issue
10. Liu, X., Liu, Y., Xie, Q., Li, L., Li, Z.: A potential-based clustering method with hierarchical optimization. *World Wide Web Journal*, this issue
11. Menasria, S., Wang, J., Lu, M.: The Purpose Driven privacy Preservation for Accelerometer-based Activity Recognition. *World Wide Web Journal*, this issue
12. Pan, X., Yang, D., Li, L., Liu, Z., Yang, H., Cao, Z., He, Y., Ma, Z., Chen, Y.: Cell detection in pathology and microscopy images with multi-scale fully convolutional neural networks. *World Wide Web Journal*, this issue
13. Rao, Y., Liu, W., Fan, B., Song, J., Yang, Y.: A Novel Relevance Feedback Method for CBIR. *World Wide Web Journal*, this issue
14. Sood, S.K.: SNA based QoS and Reliability in Fog and Cloud Framework. *World Wide Web Journal*, this issue
15. Wan, Y., Chen, L., Xu, G., Zhao, Z., Tang, J., Wu, J.: SCSMiner: Mining Social Coding Sites for Software Developer Recommendation with Relevance Propagation. *World Wide Web Journal*, this issue
16. Wang, R., Ji, W., Song, B.: Durable relationship prediction and description using a large dynamic graph. *World Wide Web Journal*, this issue
17. Wang, R., Zong, M.: Joint self-representation and subspace learning for unsupervised feature selection. *World Wide Web Journal*, this issue
18. Wen, G., Zhu, Y., Cai, Z., Zheng, W.: Self-tuning Clustering for High-dimensional Data. *World Wide Web Journal*, this issue
19. Xie, Q., Xiong, F., Han, T., Liu, Y., Li, L., Bao, Z.: Interactive Resource Recommendation Algorithm Based on Tag Information. *World Wide Web Journal*, this issue
20. Zhang, S.: Multiple-Scale Cost Sensitive Decision Tree Learning. *World Wide Web Journal*, this issue
21. Zhang, S., Cheng, D., Hu, R., Deng, Z.: Supervised Feature Selection Algorithm via Discriminative Ridge Regression. *World Wide Web Journal*, this issue
22. Zhang, Y., Zhou, G., Jin, J., Wang, X., Cichocki, A.: Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis. *International Journal of Neural Systems* **24**(4). <https://doi.org/10.1142/S0129065714500130> (2014)
23. Zhang, Y., Zhou, G., Jin, J., Zhao, Q., Wang, X., Cichocki, A.: Sparse Bayesian classification of EEG for brain-computer interface. *IEEE Trans. Neural Netw. Learn. Syst.* **27**(11), 2256–2267 (2016)
24. Zhang, S., Li, X., Zong, M., Zhu, X.: Efficient kNN classification with different numbers of nearest neighbors. *IEEE Trans. Neural Netw. Learn. Syst.* **29**(5), 1774–1785 (2018)
25. Zheng, W., Zhu, X., Zhu, Y., Hu, R., Lei, C.: Dynamic graph learning for spectral feature selection, *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-017-5272-y> (2017)
26. Zheng, W., Zhu, X., Wen, G., Zhu, Y., Yu, H., Gan, J.: Unsupervised feature selection by self-paced learning regularization, *Multimedia Tools and Applications*. <https://doi.org/10.1016/j.patrec.2018.06.029> (2018)
27. Zhu, X., Li, X., Zhang, S., Xu, Z., Yu, L., Wang, C.: Graph PCA hashing for similarity search. *IEEE Trans. Multimed.* **9**(9), 2033–2044 (2017)
28. Zhu, X., Zhang, S., Hu, R., Zhu, Y., et al.: Local and global structure preservation for robust unsupervised spectral feature selection. *IEEE Trans. Knowl. Data Eng.* **30**(3), 517–529 (2018)
29. Zhu, Y., Zhang, X., Wang, R., Zheng, W., Zhu, Y.: Self-Representation and PCA Embedding for Unsupervised Feature Selection. *World Wide Web Journal*, this issue