Received: 22 May 2020 | Accepted: 12 June 2020
DOI: 10.1111/exsy.12607

Expert Systems WILEY

EDITORIAL

Special issue on "advances in visual analytics and mining visual data"

Visual and multimedia analytics provides an emerging field of research combining strengths from information analytics, geospatial analytics, scientific analytics, statistical analytics, knowledge discovery, data management and knowledge representation, presentation, production and dissemination, cognition, perception, and interaction (Chen, Chiang and Storey, 2012). The aim is to gain insight into homogeneous, contradictory, and incomplete data through the combination of automatic analysis methods with human background knowledge and intuition.

While the scope of visual analytics is broad, one principle that has emerged over the years is the need for visual analytics systems to leverage computational methods in data mining, knowledge discovery, and machine learning for large-scale data analysis. In these systems, the human operator works alongside the computational processes in an integrated fashion. Therefore, computing systems or services can sift through large amounts of data and identify the relevant information, while the human interactively explores the reduced data space to discover trends and patterns and make informed decisions. These two components operate in coordination, allowing for a continuous and cooperative analytical loop (Cybulski et al., 2015; Valdez et al., 2016). Top papers from Data 2018, Madrid, Spain and the best paper from FEMIB 2019 in Crete, Greece, have been invited. Through a robust and competitive review process, six papers have been selected. The summary of their contributions is as follows.

Arhipova et al. (2020) propose a data aggregation approach for mobile phones, which was conducted at 15-min intervals in the area of each cellular base station. The case study examines all of Latvia's municipalities, analysing the economic activity level in each municipality in comparison to the mobile phone activity in three periods: 2015–2016, 2017, and 2018. The authors concluded that economic activity in municipalities could be estimated, and the positive dynamics of regional development have been detected. Such data and the data analytics method, which provides an understanding of how economic activities evolve in real-time in particular to locations and economic activity centres, can improve regional development planning and plan implementation. In order to assess which are the centres of economic activity in each municipality and its sphere of influence, the patterns of human commuting and fluctuations of internal activity on workdays and weekends/holidays in 2017–2018 were determined. In general, there is a shortage of reliable data on human commuting within Latvia and its specific regions; therefore, the method described here provides a practical tool for regional governments to keep track of strategy implementation and for strategic gap analysis.

Khamayseh et al. (2019) investigate the issue of friends' management in the Social Internet of Things, and proposes a framework to manage friends' requests. The proposed framework consists of friend selection, friendship removal, and update-modules. It develops a weight based and Naïve Bayes Classifier based algorithms for the selection component. Moreover, a random service allocation model is proposed to construct a service-specific network model. This model is then used in the simulation setup to examine the performance of different friends' management algorithms. The performance of the proposed framework is evaluated using simulation under different scenarios. The obtained simulation results show improvement over other strategies in terms of the average degree of connections, average path length, local cluster coefficients, and Throughput.

Lara et al. (2019) propose an outlier detection method based on a clustering process. Their aim is to overcome the specificity of many existing outlier detection techniques that fail to take into account the inherent dispersion of domain objects. The outlier detection method is based on four criteria designed to represent how human beings (experts in each domain) visually identify outliers within a set of objects after analysing the clusters. This has an advantage over other clustering-based outlier detection techniques that are founded on purely numerical analysis of clusters. To validate the proposed method, they studied method outlier detection and efficiency in terms of runtime. The results of regression analyses confirm that their proposal is useful for detecting outlier data in different domains, with a false positive rate of less than 2% and reliability greater than 99%.

Park et al. (2019) demonstrate an online principal component analysis methodology based on online eigenvector transformation with the moving average of the data stream that can reflect concept drift. They compared the network intrusion detection performance based on an online transformation of eigenvectors with that of offline methods by applying three machine learning algorithms. Both online and offline methods demonstrated excellent performance in terms of precision. However, in terms of the recall ratio, the performance of the proposed methodology with integrated online eigenvector transformation was better; thus, the F1-measure also indicated better performance. The visualization of the principal component score shows the effectiveness of their method.

Hawashin et al. (2019) propose a new efficient hybrid similarity measure for recommender systems based on a combination of the user interest-user interest similarity measure and the user interest-item similarity measure. This hybrid similarity measure improves the existing work in three aspects. First, it improves the current recommender systems by using actual user interests. Second, it provides a comprehensive

evaluation of an efficient solution to the cold start problem. Third, this method works well even when no co-rated items exist between two users. They demonstrate that their proposal is efficient in terms of accuracy, execution time, and applicability. Their proposed similarity measure achieves a mean absolute error (MAE) as low as 0.42, with 64% applicability and execution time as low as 0.03 seconds, while the existing similarity measures from the literature achieve an MAE of 0.88 at their best. These results demonstrate the superiority of their proposed similarity measure in terms of accuracy, as well as having a high applicability percentage and a very short execution time.

Vangipuram et al. (2020) propose (a) a novel imputation technique for imputation of missing data values; (b) a classifier based on feature transformation to perform classification and (c) imputation measure for similarity computation between any two instances that can also be used as the similarity measure. The performance of the proposed classifier is studied by using imputed datasets obtained through applying Kmeans, F-Kmeans and proposed imputation methods. Experiments are also conducted by applying existing and proposed classifiers on the imputed dataset obtained using the proposed imputation technique. For experiments, authors have used an open-source dataset named distributed smart space orchestration system publicly available from Kaggle. Experimental results are also validated using the Wilcoxon non-parametric statistical test. The performance of their proposed approach is better when compared to existing classifiers when the imputation process is performed using F-Kmeans and K-Means imputation techniques. It is also observed that accuracies for attack classes scan, malicious operation, denial of service, spying, data type probing, wrong setup are 100% while it is 99% for malicious control attack class when both the proposed imputation and classification technique are applied.

We thank the Expert Systems Office and Editor-in-Chief for their support and kind assistance in the completion of this special issue.

Victor Chang[1]
Shadi A. Aljawarneh[2]
Chung-Sheng Li[3]

[1]*School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough, UK*
[2]*Jordan University of Science and Technology, Irbid, Jordan*
[3]*PwC Labs, San Jose*

Correspondence
Victor Chang, School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough, UK.
Email: ic.victor.chang@gmail.com

ORCID

Victor Chang https://orcid.org/0000-0002-8012-5852
Shadi A. Aljawarneh https://orcid.org/0000-0001-5748-4921
Chung-Sheng Li https://orcid.org/0000-0001-9072-1140

REFERENCES

Arhipova, I., Berzins, G., Brekis, E., Binde, J., Opmanis, M., Erglis, A., & Ansonska, E. (2020). Mobile phone data statistics as a dynamic proxy indicator in assessing regional economic activity and human commuting patterns. *Expert Systems*, e12530. https://doi.org/10.1111/exsy.12530

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*, 1165–1188.

Cybulski, J. L., Keller, S., Nguyen, L., & Saundage, D. (2015). Creative problem solving in digital space using visual analytics. *Computers in Human Behavior*, *42*, 20–35.

Hawashin, B., Lafi, M., Kanan, T., & Mansour, A. (2019). An efficient hybrid similarity measure based on user interests for recommender systems. *Expert Systems*, e12471. https://doi.org/10.1111/exsy.12471

Khamayseh, Y., Mardini, W., Atwood, J. W., & Aldwairi, M. (2019). Dynamic framework to mining internet of things for multimedia services. *Expert Systems*, e12404. https://doi.org/10.1111/exsy.12404

Lara, J. A., Lizcano, D., Rampérez, V., & Soriano, J. (2019). A method for outlier detection based on cluster analysis and visual expert criteria. *Expert Systems*, e12473. https://doi.org/10.1111/exsy.12473

Park, S., Seo, S., Jeong, C., & Kim, J. (2019). Online eigenvector transformation reflecting concept drift for improving network intrusion detection. *Expert Systems*, e12477. https://doi.org/10.1111/exsy.12477

Valdez, A. C., Ziefle, M., Verbert, K., Felfernig, A., & Holzinger, A. (2016). Recommender systems for health informatics: State-of-the-art and future perspectives. In H. Andreas (Ed.) *Machine learning for health informatics* (pp. 391–414). Cham: Springer.

Vangipuram, R., Gunupudi, R. K., Puligadda, V. K., & Vinjamuri, J. (2020). A machine learning approach for imputation and anomaly detection in IoT environment. *Expert Systems*, e12556. https://doi.org/10.1111/exsy.12556