

Editorial

Sensor Data Analytics: Challenges and Methods for Data-Intensive Applications

Felipe Ortega * and Emilio L. Cano *

Data Science Laboratory, Research Centre for Intelligent Information Technologies, Rey Juan Carlos University, 28933 Madrid, Spain

* Correspondence: felipe.ortega@urjc.es (F.O.); emilio.lopez@urjc.es (E.L.C.)

Sensors have become a key element for the development of the Information Society. An ever-increasing number of improved sensor devices capture information for decision-making tools, either to be interpreted by humans or to be plugged back into the system for autonomous operation, self-diagnostics and resilience. It is possible to find sensor applications spanning almost any area, including healthcare and medicine, retail and logistics, smart agriculture and animal farming, industry digitalisation, smart cities, energy grids, transport or security, among many others [1]. Analytics is a term connected to the practice of data science that refers to the analysis of data using statistical tools and techniques, machine learning, information theory, pattern recognition and other methods. Outcomes stemming from this task constitute essential inputs for data-driven decision-making [2,3].

The current overabundance of data, generated in many cases by sensors, together with the refinement of standard methodologies for data science and engineering [4] has led to the rise of a fourth scientific paradigm, the so-called *data-intensive scientific discovery* [5]. Indeed, one of the most challenging aspects for the development of data-intensive applications has been how to cope with massive and complex datasets effectively, especially in situations in which real-time requirements arise [6,7]. In this regard, sensors provide an unrivalled data source to match these needs, as they can provide timestamped information with enough level of detail to characterise observed phenomena adequately.

Information theory [8] plays a central role for knowledge extraction in sensor data analytics, such as the analysis of data in the frequency domain [9], the essential concept of *entropy* [10] and efficient data representation and compression [11]. As a result, many new methods based on information theory have been developed in modern data science [12]. This Special Issue presents nine original contributions encompassing a wide variety of sensor data analytics applications, in which information theory is used to obtain knowledge from data in different domains.

Gajowniczek et al. [13] develop a novel method for data streams clustering, applicable to complete time series representing customer electricity consumption. This method leverages new Fast Fourier Transform (FFT) [9] features to improve its performance, showing the importance of information theory principles in this type of analysis. Wearable sensors tracking human activity and behaviour are at the core of several works, including applications in rehabilitation of visually impaired people [14], automated human activity recognition [15] and walking behaviour detection for elderly people [16]. The last two works attach importance to the application of information gain and neural networks to detect activity profiles accurately. Alfaro et al. [17] propose a new method to distribute the training process using the SVM algorithm, which can be applicable to Wireless Sensor Networks (WSN), aggregating the local contributions from individual sensors using Voronoi regions. Once again, this demonstrates the critical role of information aggregation in this kind of energy and location-aware sensor application. Sensor placement optimisation is the topic of another work [18], using Gaussian priors and the Fisher Information Matrix (FIM) to show important properties that can enhance recommendations on the best possible



Citation: Ortega, F.; Cano, E.L. Sensor Data Analytics: Challenges and Methods for Data-Intensive Applications. *Entropy* **2022**, *24*, 850. <https://doi.org/10.3390/e24070850>

Received: 2 June 2022

Accepted: 17 June 2022

Published: 21 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

location for a given device set. Sun et al. present an interesting application of sensor data analytics to estimate vehicles accident risk [19]. This is an emerging topic that raises significant interest among insurance companies, taking advantage of the more precise tracking capabilities enabled by built-in sensors installed in vehicles. In turn, Esteban-Escañó et al. present an interesting application of sensor data analysis to predict acidemia in electronic fetal monitoring [20], using machine learning algorithms, stressing the use of cross-entropy optimisation function along this process, to adjust the best possible predictive model. Finally, the last work [21] presents a novel methodology for cattle behaviour profiling and classification that, again, uses both time-domain and frequency-domain features to improve the accuracy of this classification task.

In summary, these contributions offer a diverse and representative portfolio of sensor data analytics applications in different scenarios, in which information theory and data science methods perform a central role in order to successfully accomplish the proposed challenges in each case.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Perros, H. *An Introduction to IoT Analytics*; Chapman & Hall/CRC Data Science Series; CRC Press: Boca Raton, FL, USA, 2021.
2. Anderson, C. *Creating a Data-Driven Organization*; O'Reilly Media Inc.: Sebastopol, CA, USA, 2015.
3. Provost, F.; Fawcett, T. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*; O'Reilly Media Inc.: Sebastopol, CA, USA, 2013.
4. Wirth, R.; Hipp, J. CRISP-DM: Towards a Standard Process Model for Data Mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, Manchester, UK, 11–13 April 2000; Volume 1, pp. 29–40.
5. Hey, A.J.; Tansley, S.; Tolle, K.M. *The Fourth Paradigm: Data-Intensive Scientific Discovery*; Microsoft Research Redmond: Redmond, WA, USA, 2009; Volume 1.
6. Kelleher, J.D.; Tierney, B. *Data Science*; MIT Press: Cambridge, MA, USA, 2018.
7. Bifet, A.; Gavaldá, R.; Holmes, G.; Pfahringer, B. *Machine Learning for Data Streams: With Practical Examples in MOA*; MIT Press: Cambridge, MA, USA, 2018.
8. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
9. Proakis, J.G.; Manolakis, D.G. *Digital Signal Processing: Principles, Algorithms and Applications*; Pearson: London, UK, 2006.
10. Bishop, C.M. *Pattern Recognition and Machine Learning*; Information Science and Statistics; Springer: Berlin/Heidelberg, Germany, 2006.
11. MacKay, D.J. *Information Theory, Inference and Learning Algorithms*; Cambridge University Press: Cambridge, UK, 2003.
12. Rodrigues, M.R.D.; Eldar, Y.C. (Eds.) *Information-Theoretic Methods in Data Science*; Cambridge University Press: Cambridge, UK, 2021. [[CrossRef](#)]
13. Gajowniczek, K.; Bator, M.; Ząbkowski, T. Whole Time Series Data Streams Clustering: Dynamic Profiling of the Electricity Consumption. *Entropy* **2020**, *22*, 1414. [[CrossRef](#)] [[PubMed](#)]
14. Reyes Leiva, K.M.; Jaén-Vargas, M.; Cuba, M.Á.; Sánchez Lara, S.; Serrano Olmedo, J.J. A Proposal of a Motion Measurement System to Support Visually Impaired People in Rehabilitation Using Low-Cost Inertial Sensors. *Entropy* **2021**, *23*, 848. [[CrossRef](#)] [[PubMed](#)]
15. Liu, L.; He, J.; Ren, K.; Lungu, J.; Hou, Y.; Dong, R. An Information Gain-Based Model and an Attention-Based RNN for Wearable Human Activity Recognition. *Entropy* **2021**, *23*, 1635. [[CrossRef](#)] [[PubMed](#)]
16. Aznar-Gimeno, R.; Labata-Lezaun, G.; Adell-Lamora, A.; Abadía-Gallego, D.; del Hoyo-Alonso, R.; González-Muñoz, C. Deep Learning for Walking Behaviour Detection in Elderly People Using Smart Footwear. *Entropy* **2021**, *23*, 777. [[CrossRef](#)] [[PubMed](#)]
17. Alfaro, C.; Gomez, J.; Moguerza, J.M.; Castillo, J.; Martinez, J.I. Toward Accelerated Training of Parallel Support Vector Machines Based on Voronoi Diagrams. *Entropy* **2021**, *23*, 1605. [[CrossRef](#)] [[PubMed](#)]
18. Zhou, R.; Chen, J.; Tan, W.; Yan, Q.; Cai, C. Optimal 3D Angle of Arrival Sensor Placement with Gaussian Priors. *Entropy* **2021**, *23*, 1379. [[CrossRef](#)] [[PubMed](#)]
19. Sun, S.; Bi, J.; Guillen, M.; Pérez-Marín, A.M. Driving Risk Assessment Using Near-Miss Events Based on Panel Poisson Regression and Panel Negative Binomial Regression. *Entropy* **2021**, *23*, 829. [[CrossRef](#)] [[PubMed](#)]
20. Esteban-Escañó, J.; Castán, B.; Castán, S.; Chóliz-Ezquerro, M.; Asensio, C.; Laliena, A.R.; Sanz-Enguita, G.; Sanz, G.; Esteban, L.M.; Savirón, R. Machine Learning Algorithm to Predict Acidemia Using Electronic Fetal Monitoring Recording Parameters. *Entropy* **2022**, *24*, 68. [[CrossRef](#)] [[PubMed](#)]
21. Cabezas, J.; Yubero, R.; Visitación, B.; Navarro-García, J.; Algar, M.J.; Cano, E.L.; Ortega, F. Analysis of Accelerometer and GPS Data for Cattle Behaviour Identification and Anomalous Events Detection. *Entropy* **2022**, *24*, 336. [[CrossRef](#)] [[PubMed](#)]