



Editorial: Data and Information in Online Environments

Edgar Bisset-Alvarez¹

Published online: 20 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Editorial:

In recent years the world has faced one of its greatest challenges. A terrible disease has put researchers, scientists, professors, governments, and society on a career path to minimize the social and economic impacts caused by the COVID-2019 pandemic. In this scientific race, the importance of collaboration between the agents of the scientific community was verified, and the collection, tabulation, and manipulation of research data for the elaboration of didactic research materials. At the same time, all the content produced and disseminated posed great challenges to the scientific community, and the questioning of the veracity and reliability of scientific information became yet another problem to be solved. The volume of fake news around the pandemic grew every day, compromising human lives.

Several problems arose from this scenario: How can collaboration and data sharing between researchers, governments, and society help in the development of humanity? What are the technological, economic, political, and social challenges of facing the fake news phenomenon?

In order to discuss and suggest possible solutions to these problems, researchers from different areas of knowledge, such as data science, natural language processing, data engineering, big data, research evaluation, network science, sociology of science and communication, are working to address technological, economic, political and social challenges.

This special issue includes eight high-quality papers presented at the Second International Conference on Data and Information in Online Environments (DIONE 2021).

The first article proposes an automated technology surveillance method using a map-reduce model to deal with Big Data scenarios. Given the variety of information, sources to be considered and the considerable volume of data to be analyzed, organizations will be able to monitor internal and external technological environments and anticipate changes

with potentially positive or negative impacts on their business. The research delimited a series of activities necessary to enable the automation process and was divided into five processes: planning, collection, organization, intelligence, and communication. A software prototype was implemented to monitor some portals in the furniture and wood sector. A total of 2,918 publications, collected in four specialized portals, were pre-processed and analyzed properly, and the results generated later were presented to stakeholders. In the end, industry experts evaluated the proposal and preliminary results as very promising.

The second article proposes a technique for the expansion and generation of weakly supervised data, which facilitates the retrieval of information from specialists in the scientific area with neural models, aiming to reduce the difficulty of finding candidates with adequate profiles for a given activity. In the proposed technique, relevance judgments were created with heuristic techniques, enabling the use of models that require large volumes of data. The search for experts in the scientific area with neural models resulted in an even greater challenge due to the complexity of these models and the need for large volumes of data with relevant judgments or labels for their training. It was proposed, a technique using deep autoencoder to select negative documents or irrelevance judgments and, finally, a classification model based on recurrent networks called Dual Embedding LSTM that was able to outperform all compared baselines.

The third article aims to analyze the applicability of research productivity indices to evaluate research groups. These indices are useful to support decisions in the distribution of financial resources for the development of scientific research. Some of these indices aim to evaluate an individual researcher and/or research groups applying the same formula for both situations; in many cases, such a strategy is not simple. For this, several studies that applied the h index to research groups were evaluated, identifying the different ways of using this procedure. Through a literature review on these indices, we proposed an approach to select and aggregate scientific articles for evaluation by research groups. In addition, ways were presented to expand these metrics to

✉ Edgar Bisset-Alvarez
edgar.bisset@ufsc.br

¹ Federal University of Santa Catarina, Florianópolis, Brazil

be applied more fairly to assess quantitative and qualitative aspects in research groups.

The fourth article compares two machine learning approaches for the development of a metamodel capable of estimating the thermal load in single-family buildings. The metamodels evaluated were the Artificial Neural Networks and the Gradient Boosting Machine. Some of the equitable approaches proposed to reduce world energy consumption in a scenario of a growing global climate crisis were analyzed. One of the most advanced methods for this estimation uses computer simulations, which require a high level of technical knowledge. The results obtained during the research allowed us to observe a better performance in the indicators of the Gradient Boosting Machine approach in relation to the Artificial Neural Networks. The downside is that the Gradient Boosting Machine requires a relatively long training time, making its use in routine projects less feasible.

The fifth article proposes a framework for evaluating predictive models of academic failure based on machine learning, to facilitate early pedagogical interventions. We take as a case study a Brazilian undergraduate course in the distance learning modality. Seven classification models were run on normalized datasets, which included grades from three weeks of classes for a total of six weeks. Because it is a context of unbalanced data, the adoption of a single metric to identify the best predictive model of student failure would not be efficient. Therefore, the proposed framework considers 11 metrics generated by running the classifiers and applying exclusion and ordering criteria to produce a list of the best predictors. Finally, some possible applications were discussed and presented to minimize student failure..

The sixth article raises an investigation about the prediction and the importance of the resource to estimate the infection by COVID-19, using the Machine Learning approach. The work analyzed the inclusion of climatic characteristics, mobility, governmental actions, and a number of cases per health sub territory from an existing model. The Random Forest with Permutation Importance method was used to assess the importance and list the thirty most relevant that represents the probability of disease infection. Among all the characteristics, the following stand out: i) the health variables by region, ii) the period between the date of notification and the onset of symptoms, iii) characteristics of symptoms such as fever, cough, and sore throat, iv) variables of traffic flow and mobility, and also v) climate characteristics. The model was validated and reached an average accuracy of 81.82%, while sensitivity and specificity reached 87.52% and 78.67% respectively in the infection estimate. Therefore, the proposed investigation represents an alternative to guide authorities in understanding aspects related to the disease.

The seventh article presents and discusses the factors that can influence the state of anomie of researchers in data sharing. A table was presented with the factors that influence

the perception and attitude of researchers in sharing data. The low adherence between the perception and practice of established norms and the actions of researchers in the data sharing process results in a state of anomie that constitutes an important barrier to data access. The data revealed that indicators of the Cognitive, Normative, Career, Resources, and Social pillars influence the perception and attitude of sharing or retaining data. It is hoped that the results will provide insight into the factors that influence researchers to share their data.

The eighth article analyzes the object of communication and the media, focusing on the sports information service platform on the subject of communication, from the perspective of statistics in sports as an example to analyze the characteristics of information and channel sink. Using the analytical hierarchy process, from the content of networked sports information, audience experience, organization of networked sports information, and dissemination of sports information and network environment in four aspects, build the evaluation index system of the sports model. dissemination of sports information and determine the weight of each index. The results showed that: the premise of the sports information dissemination model is based on a network technology to serve the public, the right to guide the public to build a harmonious sports information network for the purposes of the business model, for the development of the sports site, expand business partners, reposition the sports network station business operation mode, break the old ideas, improve the connotation of the sports site of the operation mode.

The ninth article develops a new optimization-oriented classifier to classify sentiment degrees. Focusing on opinions and ratings posted by users on social media. Here, reviews are considered where features are extracted using reviews. In addition, significant resources such as SentiWordNet-based resources, statistical resources, context-based resources, and term frequency-inverse document frequency (TF-IDF) based resources are taken from the review. These features are adapted into the Hierarchical Attention Network (HAN) to categorize the degree of sentiment. HAN training is performed using the proposed Competitive Swarm Water Wave Optimization (CSWWO) algorithm. The developed CSWWO algorithm is newly designed integrating the Competitive Swarm Optimizer (CSO) and Water Wave Optimization (WWO) techniques. Thus, the proposed CSWWO-based HAN model classifies sentiment into five classes, such as bad, better, good, very good, and excellent. Finally, the handling of the data flow is performed by concept deviation detection and prototype-based adaptation.

The tenth article analyzes cooperative wideband spectrum sensing (CWSS) for cognitive radio under imperfect feedback channels and proposes an algorithm to overcome the effects of imperfections in feedback channels. A

probabilistic model is proposed to model the behavior of feedback channels. An algorithm based on repetition channel code for cooperative wideband spectrum sensing (R-CWSS) is proposed to overcome the effects of imperfect feedback channels. First, the modified theoretical analysis for the existing algorithm called partial band Nyquist sampling-based CWSS (PBNS-CWSS) is given, and then the same is extended to include the imperfect feedback channels. Then, a complete theoretical analysis is provided for the proposed R-CWSS algorithm. All the theoretical analysis provided in this paper is verified using simulations. The analysis demonstrates that the performance of CWSS is greatly affected by the imperfect feedback channels. Also, the proposed R-CWSS shows superior performance compared to recently proposed state-of-the-art algorithms. Finally, the effects of different parameters on the performance of R-CWSS are studied.

The eleventh article presents a remote patient monitoring system (RPMS). The Internet of Things (IoT) and integrated cloud computing technologies are used for the implementation. The system can continuously measure different physiological parameters with the appropriate degree of accuracy required by medical standards. A Personal Service Application (PSA) was developed for Android devices, which acts as a gateway between the RPMS and the Cloud. The developed PSA offers visualization and storage of physiological parameters locally as well as in the cloud, along with real-time data transmission for remote monitoring and further analysis. RPMS has been implemented and validated in the state-of-the-art patient monitoring system.

Acknowledgements The guest editor appreciates the guidance received by the Editor-in-Chief, Prof. Imrich Chlamtac for his supportive guidance during the entire editorial process. Special thanks also go to our reviewers for their effort in reviewing the manuscripts submitted to the special issue. Finally, it is appreciated the support received by the Post-graduate Program of Information Science, and the Federal University of Santa Catarina (PGCIN/UFSC).

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



Edgar Bisset-Alvarez is a graduate with an M.Sc in Library and Information Science from the University of Havana, Cuba, and a Ph.D. degree in Information Science from the São Paulo State University, Brazil. He is currently a professor and post-graduate program in the Information Science coordinator with the Federal University of Santa Catarina, Brazil, and serves as editor-chief of *Encontros Bibli Journal*. Associated with the Working Group GT-8 Information and Technology of the

National Association for Research and Graduate Studies in Information Science (ANCIB), Brazil. He participates as a researcher in the research group ITI-RG (Intelligence, Technology, and Information—Research Group) at the Federal University of Santa Catarina.