

## Editorial

This third issue of the twelfth volume of JIDM brings six extended versions of papers published as full papers on the 2020 Brazilian Symposium on Databases (SBBD2020). SBBD is the largest venue in Latin America for presentation and discussion of research results and applications in the database domain. Twelve Full papers were accepted for presentation at the SBBD2020 and invited to submit an extended version to this JIDM special issue, incorporating at least 30% of new content with respect to the original paper. From the twelve invited papers six finally submitted an extended version to this special issue. Each submitted paper went through multiple review rounds guided by suggestions of a committee composed of senior researchers of the database community. The six accepted papers were presented in the following technical sessions of the SBBD2020: Query Language and Processing; Data Integration, Extraction and Cleaning; Database Management Systems; Data Mining and Machine Learning.

The first paper in this special issue, *Efficient Set Similarity Join on Multi-Attribute Data Using Lightweight Filters*, by Leonardo Andrade Ribeiro and co-authors, discusses the problem raised when integrating data from different data sources. A common problem in this setting is to find near-duplicate entries among data extracted from the different sources. The authors consider a multi-attribute join problem and propose an algorithm framework that incorporates the *FSSJoin* algorithm. Experimental results compare the proposed approach against competing solutions, showing competitive results in both runtime efficiency and memory consumption.

The paper *Evaluating Temporal Bias in Time Series Event Detection Methods* by Luciana Escobar and co-authors explores the detection of events in time-series and the implication of temporal bias on models' analysis metrics. The authors use the Harbinger framework to instantiate different event detection methods and to support classification metrics that take into account either prior or posterior detection with respect to labeled events marked in the time-series.

Isis Fogaça and Renato Bueno contribute with the paper *Analyzing Temporal Evolution of Complex Data Using Similaroty Queries*. The authors discuss the problem of computing similarity distance among complex data (eg. images) in a estimated point in time. The work interestingly maps objects in a metric space to a multidimensional space with temporal dimension, where distances of objects in an estimated time position can be calculated. The authors evaluate algorithms to compute K-NN and range queries.

In *SmarT: Machine Learning Approach for Efficient Filtering and Retrieval of Spatial and temporal Data in Big Data*, Sávio S.T. de Oliveira and co-authors investigate the problem of choosing a Big Data platform to run spatio-temporal queries. The platforms considered in the study are Apache Spark, Elasticsearch and SciDB. A machine learning model based on the XGBoost learner receives information about the query and the computing environment to select the target system.

The paper by Camila R. Lopes and co-authors, *An empirical assessment of quality metrics for diversified similarity searching*, discusses similarity search queries exploring diversity on the result set. The authors focus on improving quality metrics for the evaluation of query results with diversity and present the Diversity Features Model (DFM), a new approach which combines previous metrics to asses the quality of similarity search with diversity.

Finally, the work by Vanessa Souza and Karin Becker, *A Deep Learning Ensemble to Classify Anxiety, Depression and their Comorbidity from Texts of Social Networks*, which received the Prof. Jose Mauro de Castilho award as the Best paper of the SBBD2020, discusses the use of Deep Learning techniques for the classification of mental diseases, such as Anxiety and Depression, from text in social networks, such as Reddit.

These papers are illustrative of the high level work produced by the database community as presented in the Simpósio Brasileiro de Banco de Dados (SBBD2020).

We would like to thank all the authors and reviewers that contributed in putting together this special issue.

Enjoy your reading.

Fabio Porto  
*SBBD2020 Guest Editor*