Data Mining and Knowledge Discovery (2019) 33:695–696
https://doi.org/10.1007/s10618-019-00623-5

Guest Editorial

Special Issue on Data Mining for Geosciences

Alípio Jorge[1] · Rui L. Lopes[2] · German Larrazabal[3] · Hamed Nikhalat-Jahromi[2]

Published online: 27 March 2019

Modern Geosciences have to deal with huge quantities and a wide variety of data. Examples include 2-D, 3-D and 4-D seismic surveys, well logs generated by sensors, detailed lithological records, satellite images, audio signals, meteorological records, human activity records and fossil data. These data serve important industries, such as the exploration of mineral deposits and the production of energy (Oil and Gas, Geothermal, Wind, Photovoltaic, Hydroelectric), are important in the study of the earth crust to reduce the impact of earthquakes, in land use planning, enable the study of climate evolution and have a fundamental role in sustainability. Geosciences thus offer a rich set of applications in very related but different problems. The variety of data formats and their volume require specific attention from Data Mining researchers, solution providers and Geoscientists. However, the opportunities for using and developing data mining techniques in the various scenarios of geosciences remain largely unexplored or uncharted and constitute an avenue of research challenges.

This special issue gathers a selection of papers with strong contributions in the area of Data Mining to the field of Geosciences in the large. We are particularly happy with the variety of domains and offered Data Mining solutions. The contributions are impactful for Geoscientists as well as for Data Miners working in other fields. There were twenty-three submissions to the special issue that were subject to rigorous

✉ Alípio Jorge
amjorge@fc.up.pt

Rui L. Lopes
rmlc@inesctec.pt

German Larrazabal
german.larrazabal@repsol.com

Hamed Nikhalat-Jahromi
hamed.nikhalat@gmail.com

[1] Department of Computer Science - FCUP, University of Porto/LIAAD - INESC TEC, Rua do Campo Alegre, 1021/1055, 4169-007 Porto, Portugal

[2] INESC TEC, Campus da FEUP, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

[3] Repsol USA, 2455 Technology Forest Boulevard, The Woodlands, TX 77381, USA

reviewing by both Geoscience and Data Mining experts. The five selected articles cover various areas within the proposed theme, from satellite image time series to acoustic remote sensing and data streams analysis.

- Méger et al. propose an automatic unsupervised technique for summarizing satellite image time-series using an entropy-based ranking and a swap randomization technique. This work fosters the identification and reuse of satellite image time-series with relevant phenomena information and provides tools and principles that enable urban expansion assessment, glacier dynamic analysis, snow cover mapping, forest mapping, earthquake monitoring, coastline detection, and soil erosion monitoring.
- Ceci et al. address several issues of remote sensing data in a holistic approach by performing online adaptive training of neural networks and having into account spatial autocorrelation in entropy measures. They show the effectiveness of this approach on renewable energy forecasting using photovoltaic power production data. The proposed methods enable to forecast both the consumption and the production in smart grid systems thus enabling dynamic pricing models, as well as proactive control of the macrogrid network.
- Airola et al. introduce a spatial cross-validation method to correct bias in AUC (Area Under the Curve) estimation in machine learning based classification, applying it to different classification methods on mineral prospectivity mapping data. The work shows how to avoid some misleading evaluation results on spatial data.
- Wilson et al. address issues of distributed acoustic sensing data streams in the Oil and Gas domain. They propose an online classification technique to identify anomalous regions within multivariate time series, based on the Multivariate Locally Stationary Wavelet model. As a result of their proposal the quality of the data is improved as well as the overall data mining predictions.
- Žliobaitė focuses on concept drift in transfer learning targeting the reconstruction of past environmental conditions and climate changes tracking over millions of years from the fossil record data. The aim is to infer population characteristics and distribution of species that existed in a distant past.

The accepted contributions illustrate the diverse range of topics currently being investigated, and of the applications of Data Mining in Geosciences. The diversity of the applications and methods addressed should be appealing to both the experts in the field and those who want a snapshot of the current breadth of Data Mining for Geosciences research.

We are deeply grateful to all the reviewers who helped us with the laborious task of evaluating the papers and who contributed with insightful comments that enabled the high quality of the selected works. We also would like to thank the support of Springer staff and of the responsible editor Johannes Fürnkranz. We especially show our appreciation to all the authors who submitted papers and who stood with us during this very tough selection and reviewing period.