Seventeenth International Conference Zaragoza-Pau on Mathematics and its Applications
Jaca, September 4–6th 2024

Data Science and Location Theory in a nutshell

Justo Puerto,

SUMMARY

The design, management and use of any type of complex network requires a methodology to handle its parameters, detect deficiencies and coordinate their resources to solve the problems that arise. Developing methods to carry out such actions demands, among other things, the preliminary screening of large masses of data, quantitative analysis to design better information structures, often organized as networks, and the solution of optimization problems related to clustering, location, routes, allocation of flows and traffic of any kind, distribution of intelligent sensors, early detection of extreme observations, profiling user behavior and operations planning, often under an environment of risk or uncertainty, etc. All those operations involve large masses of data that must be integrated in all phases of the operational analysis. The standard approach of handling separately/sequentially data and design is defective and lacks the important gain of integration. Data integration, data reduction, feature selection, outliers detection, intelligent segmentation or separability are the driving challenges that relies on tools such as machine learning, statistical analysis, optimization, and mathematical programming.

One step forward to bring the gap of integration in data science is the application of techniques from optimization and statistics. In this talk, we focus on one important challenge: "integration of design, optimization and data". This approach sets a very ambitious objective: to improve the data science paradigm integrating techniques of location and networks analysis and vice versa. The two features of datafication and universalization of the available information establish a subtle difference with the standard methodologies of location science and network analysis: the representation of complex environments with large masses of data imposes the need to apply more advanced tools of mathematics and machine learning to allow the design, the effective treatment and use of data at all levels and the optimization of the problems that arise from them.

In general terms, the challenge that is currently posed in this field is, not only, to incorporate the methodology of data science into the analysis of large scale networks, in order to deal with problems that involve large masses of data ("Big Data"); but also, reciprocally, how to make use of the tools and models of optimization and network design in data science. This talk will surf over recent results of our team (see references) in this respect, showing how modern location analysis techniques improve several machine learning methodologies including regression, unsupervised and supervised classification and community detection.

Keywords: data science, location, mathematical programming,...

AMS Classification: 62R07, 90B80, 90B85

References

[1] Benati, Stefano and Puerto, Justo. 2023. A network model for multiple selection questions in opinion surveys. Quality & Quantity. ISSN 1573-7845.

[2] Benati, Stefano; Puerto; Justo; Rodríguez Chía, Antonio M. and Temprano Francisco 2023. Overlapping communities detection through weighted graph community games. PLOS ONE. Public Library of Science. 18-4, pp.1-35.

[3] Stefano Benati; Justo Puerto; Antonio M. Rodríguez Chía; Francisco Temprano García. 2022. A Mathematical Programming approach to Overlapping community detection. Physica A: Statistical Mechanics and its Applications. Elsevier. 602, pp.127628.

[4] Alfredo Marín Pérez; Luisa I. Martínez Merino; Justo Puerto; Antonio M. Rodríguez Chía. 2022. The soft-margin Support Vector Machine with ordered weighted average. Knowledge-Based Systems. Elsevier. 237, pp.107705.

[5] Víctor Blanco; Japón, Alberto; Puerto, Justo. 2022. A mathematical programming approach to SVM-based classification with label noise. Computers & Industrial Engineering. 172, pp.108611-108611. ISSN 0360-8352.

[6] Víctor Blanco; Alberto Japón Sáez; Justo Puerto, Antonio Rodríguez Chía. 2021. Robust optimal classification trees under noisy labels. Advances in Data Analysis and Classification volume. Springer. 16, pp.155-179.

[7] Stefano Benati; Diego Ponce López; Justo Puerto (AC); Antonio Rodríguez Chía. (3/4). 2021. A branch-and-price procedure for clustering data that are graph connected. European Journal of Operational Research. Elsevier. 297-3, pp.817-830. ISSN 0377-2217.

[8] Víctor Blanco; Diego Ponce; Justo Puerto; Antonio M. Rodríguez-Chía 2021. On the multisource hyperplanes location problem to fitting set of points. Computers and Operations Research. Elsevier. 128, pp.105124. ISSN 0305-0548.

[9] Víctor Blanco; Justo Puerto, Antonio M. Rodríguez-Chía;. 2020. On lp-support vector machines and multidimensional kernels. Journal of Machine Learning Research. Journal of Machine Learning Research. 21, pp.1-29.

[10] José J. Calviño; Miguel López Haro; Juan M. Muñoz Ocaña; Justo Puerto; Antonio M. Rodríguez Chía. 2022. Segmentation of Scanning-Transmission Electron Microscopy Images using the Ordered Median Problem. European Journal of Operational Research. Elsevier. 302-2, pp.671-687.

[1]IMUS
University of Seville, Spain
email: puerto@us.es