

Cost-Sensitive Learning: Preface

Luís Torgo

*Faculty of Computer Science, Dalhousie University
Halifax, Canada*

LTORGO@DAL.CA

Stan Matwin

*Faculty of Computer Science, Dalhousie University
Halifax, Canada*

STAN@CS.DAL.CA

Gary Weiss

*Department of Computer and Information Science, Fordham University
New York, USA*

GAWEISS@FORDHAM.EDU

Nuno Moniz

*LIAAD-INESC TEC DCC-FCUP, University of Porto
Porto, Portugal*

NMMONIZ@INESCTEC.PT

Paula Branco

*LIAAD-INESC TEC DCC-FCUP, University of Porto
Porto, Portugal*

PAULA.BRANCO@DCC.FC.UP.PT

This volume contains the Proceedings of the International Workshop on Cost-Sensitive Learning - COST'2018. This workshop was co-organized by the Laboratory of Artificial Intelligence and Decision Support (INESC TEC), the Department of Computer Science in Faculty of Sciences of Porto University (Portugal), the Faculty of Computer Science of Dalhousie University and the Department of Computer and Information Science in Fordham University. The workshop was co-located with the *SIAM: International Conference on Data Mining (SDM) 2018* and was held on the 5th of May 2018 in the San Diego Marriott Mission Valley, San Diego, California, USA.

Research on data mining and machine learning tasks is usually developed under assumptions of uniform preferences, where cases are equally important, and issues such as data acquisition costs are not considered. However, many real-world data mining applications involve complex settings where such assumptions do not hold. Frequently, predictive analytics involve settings where the consideration of costs is unavoidable. Such costs can appear at all stages of the data mining process, e.g. data acquisition, modelling or model deployment. The main goal of this workshop is to address these tasks involving the consideration of costs and/or benefits that may arise from different sources.

The most frequently studied setting involves binary classification tasks with costs considered at the evaluation level. In this case, different penalizations and/or benefits are assigned to different errors and/or accurate predictions, and a cost matrix is used to express this domain-dependent information used in assessing the performance of the models. However, other predictive tasks may also be cost dependent, such as regression and time series or data streams forecasting tasks. Moreover, there are other issues which, although relevant, are still unsolved or need improved solutions, such as performance evaluation, different cost

of acquisition and use of particular attributes and applications involving unsupervised or semi-supervised tasks.

Tackling the issues raised by cost-sensitive learning problems is crucial to both academia and industry, as it allows the development of more suitable and robust systems for complex settings. For industry partners, this presents the opportunity to develop frameworks targeting specific contexts, embedding in the solutions the necessary domain knowledge. Examples include dealing with budgeted resources, limited space or computational time, prediction of rare events and anomaly detection. Considering these factors is often crucial for the successful deployment of predictive models in these industrial environments.

This workshop received a diversity of high quality inter-disciplinary contributions discussing various aspects of cost-sensitive learning. Overall, there were 12 paper submissions for COST'2018, out of which 3 papers were accepted for inclusion in workshop proceedings and assigned a presentation slot, together with time for questions and answers. These papers cover different aspects of the cost-sensitive learning problem. [Miller et al. \(2018\)](#) address a common problem in operation contexts where obtaining ground truth labels is highly costly, and propose an evaluation framework for binary classifiers in such scenarios. Focusing on the problem of model selection, an extension of the ROC Convex Hull method is proposed by [Meekins et al. \(2018\)](#) to include additional cost information in the selection of optimal classifiers. Finally, [Kriege et al. \(2018\)](#) focus on an interesting problem related to the recognition of cuneiform script, and propose a graph model embedding tailored distance measures while controlling the computation cost of the prediction phase, along with efficient heuristics for the computation and demonstration of its effectiveness in classification tasks. The workshop included an invited talk by Professor Luis Torgo, from the Dalhousie University, entitled "Utility-based Regression".

We would like to thank all of the authors and the Program Committee members that enabled a successful workshop for their hard work and commitment. We also want to deeply thank the SDM 2018 Workshop and Tutorial Chairs for their support in the logistics of this workshop.

Organizing Committee

- Luís Torgo (Faculty of Computer Science, Dalhousie University; Department of Computer Science, Faculty of Sciences, University of Porto; LIAAD - INESC TEC)
- Stan Matwin (Faculty of Computer Science, Dalhousie University)
- Gary Weiss (Department of Computer and Information Science, Fordham University)
- Nuno Moniz (Department of Computer Science, Faculty of Sciences, University of Porto; LIAAD - INESC TEC)
- Paula Branco (Department of Computer Science, Faculty of Sciences, University of Porto; LIAAD - INESC TEC)

Program Committee

- Naoki Abe (IBM)

- Roberto Alejo (Tecnológico de Estudios Superiores de Jocotitlán)
- Colin Bellinger (University of Alberta)
- Seppe Vanden Broucke (Katholieke Universiteit Leuven)
- Nitesh Chawla (University of Notre Dame)
- Christopher Drummond (National Research Council Canada)
- Inês Dutra (DCC - FCUP)
- Tom Fawcett (Apple)
- Mikel Galar (Universidad Pública de Navarra)
- Nathalie Japkowicz (American University)
- Charles Ling (Western University)
- Dragos Margineantu (Boeing Research and Technology)
- Ronaldo Prati (Universidade Federal do ABC)
- Foster Provost (NYU Stern)
- Jose Hernandez-Orallo (Universitat Politècnica de València)
- Rita Ribeiro (LIAAD / INESC Tec)
- Shengli Victor Sheng (University of Central Arkansas)
- Marina Sokolova (University of Ottawa)

References

- Nils M. Kriege, Matthias Fey, Denis Fisseler, Petra Mutzel, and Frank Weichert. Recognizing cuneiform signs using graph based methods. In *Proceedings of the International Workshop on Cost-Sensitive Learning (COST 2018)*, volume 91 of *Proceedings of Machine Learning Research*, pages 31–44, SDM, San Diego, California, USA, 3–5 May 2018. PMLR.
- Ryan Meekins, Stephen Adams, Peter A. Beling, Kevin Farinholt, Nathan Hipwell, Ali Chaudhry, Sherwood Polter, and Qing Dong. Cost-sensitive classifier selection when there is additional cost information. In *Proceedings of the International Workshop on Cost-Sensitive Learning (COST 2018)*, volume 91 of *Proceedings of Machine Learning Research*, pages 17–30, SDM, San Diego, California, USA, 3–5 May 2018. PMLR.
- Benjamin A. Miller, Jeremy Vila, Malina Kirn, and Joseph R. Zipkin. Classifier performance estimation with unbalanced, partially labeled data. In *Proceedings of the International Workshop on Cost-Sensitive Learning (COST 2018)*, volume 91 of *Proceedings of Machine Learning Research*, pages 4–16, SDM, San Diego, California, USA, 3–5 May 2018. PMLR.