

## Preface

This volume contains the proceedings of the 5th Symposium on Conformal and Probabilistic Prediction with Applications (COPA 2016), which was co-organized by Royal Holloway, University of London, UK, and Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain. The Symposium was held at CIEMAT during April 20–22, 2016.

Conformal prediction is a recently developed framework for complementing the predictions of machine learning algorithms with reliable measures of confidence. The framework produces well-calibrated confidence measures for individual examples without assuming anything more than that the data are generated independently from the same probability distribution.

Since its development the framework has been applied to many popular techniques, such as support vector machines, k-nearest neighbors, neural networks, ridge regression etc., and has been successfully applied to many challenging real world problems, such as the early detection of ovarian cancer, the classification of leukemia subtypes, the diagnosis of acute abdominal pain, the assessment of stroke risk, the recognition of hypoxia in electroencephalograms (EEGs), the prediction of plant promoters, the prediction of network traffic demand, the estimation of effort for software projects and the back-calculation of non-linear pavement layer moduli. The framework has also been extended to additional problem settings such as semi-supervised learning, anomaly detection, feature selection, outlier detection, change detection in streams, and active learning. Recent developments in collecting large volumes of data have also required its adjustment to handle “big data”.

The aim of this symposium is to serve as a forum for the presentation of new and ongoing work and the exchange of ideas between researchers on any aspect of conformal and probabilistic prediction and their applications.

While the previous four annual gatherings (COPA 2012 to COPA 2015) were devoted mainly to conformal predictors, they also included extensions of conformal predictors to Venn predictors. The title of this year's event reflects the expanded scope explicitly and covers all kinds of probabilistic prediction, not only Venn prediction.

The popularity of conformal prediction in the machine-learning community is growing. As evidence of this we can mention the following events that took place after COPA 2015. In June 2015, a special issue on “Conformal Prediction and its Applications” of the *Annals of Mathematics and Artificial Intelligence* (Volume 74, Issues 1–2) was published. In July 2015, Henrik Boström, Alexander Gammerrman, Ulf Johansson, Lars Carlsson, and Henrik Linusson presented the tutorial “Conformal Prediction: A Valid Approach to Confidence Predictions” at the 2015 International Joint Conference on Neural Networks (Killarney, Ireland). An EU Horizon 2020 project on drug design that started in September 2015 adopted conformal prediction as one of the main tools for selecting useful chemical compounds. In December 2015, an Indo-UK workshop on “Mathematical Foundations of Probabilistic Conformal

Prediction and Its Applications in Machine Learning” was held at the Indian Institute of Technology in Hyderabad, India. In January 2016, there was a session on “Data-Intensive Methods and Conformal Predictions” at the International Conference on Pharmaceutical Bioinformatics (ICPB 2016) in Pattaya, Thailand.

Overall, 14 papers were accepted for presentation at the symposium after being reviewed by at least two independent academic referees. The authors of these papers come from 11 different countries, namely: Austria, Cyprus, Italy, The Netherlands, Russia, Spain, Sweden, Switzerland, Ukraine, the UK and the USA.

The volume is divided into three parts. The first part presents the invited paper “Learning with Intelligent Teacher” by Vladimir Iapnik and Rauf Izmailov, devoted to learning with privileged information and emphasizing the role of the teacher in the learning process.

The second part is devoted to the theory of conformal prediction. The two papers in this part investigate various criteria of efficiency used in conformal prediction (Vladimir Vovk, Valentina Fedorova, Ilia Nouretdinov, and Alexander Gammerman) and introduce a universal probability-free version of conformal predictors (Vladimir Vovk and Dusko Pavlovic).

The core of the book is formed by the third part, containing experimental papers describing various applications of conformal prediction. This part opens by “Conformal Predictors for Compound Activity Prediction” (Paolo Toccaceli, Ilia Nouretdinov and Alexander Gammerman), applying conformal prediction to big and imbalanced datasets in the field of drug discovery. The following paper, “Conformal Prediction of Disruptions from Scratch: Application to an ITER Scenario” by Raul Moreno, Jesús Vega, and Sebastian Dormido-Canto, demonstrates advantages of conformal prediction over the conventional methodology in the field of nuclear fusion. In “Evaluation of a Variance-Based Nonconformity Measure for Regression Forests” Henrik Boström, Henrik Linusson, Tuve Löfström and Ulf Johansson continue their empirical investigation of conformal prediction based on random forests; their new algorithms achieve impressive computational efficiency while retaining predictive efficiency. This part is concluded by four papers proposing valuable extensions of the framework of conformal prediction in various directions. First, Antonis Lambrou and Harris Papadopoulos (“Binary Relevance Multi-label Conformal Predictor”) extend the framework to multi-label classification. The second extension is proposed by Andrea Murari, Saeed Talebzadeh, Jesús Vega, Emmanuele Peluso, Michela Gelfusa, Michele Lungaroni, and Pasqualino Gaudio in “A Metric to Improve the Robustness of Conformal Predictors in the Presence of Error Bars”: now all data, including the attributes of the objects to be labelled, are not precise but are obtained using a noisy measurement procedure. The third paper, by Shuang Zhou, Evgeni Smirnov, Ralf Peeters, and Gijs Schoenmakers (“Decision Trees for Instance Transfer”), applies the ideas of conformal prediction to the case where the test data are generated from a distribution different from that generating the training data. Finally, Giovanni Cherubin and Ilia Nouretdinov (“Hidden Markov Models with Confidence”) extend the methodology of conformal prediction to the popular setting of hidden Markov models.

The third part contains theoretical and experimental papers in general machine learning. It opens by two theoretical papers, “Variable Fidelity Regression Using Low Fidelity Function Blackbox, and Sparsification” by Alexey Zaytsev and “Effective

Design for Sobol Indices Estimation Based on Polynomial Chaos Expansions” by Evgeny Burnaev, Ivan Panin, and Bruno Sudret. Apart from theoretical results, both papers provide convincing empirical validation. The next two papers are devoted to two different, both very important, applications: medicine (“Joint Prediction of Chronic Conditions Onset: Comparing Multivariate Probits with Multiclass Support Vector Machines” by Shima Ghassem Pour and Federico Girosi) and information security (“Method of Learning Malware Behavior Scripts by Sequential Pattern Mining” by Alexandra Moldavskaya, Victoria Ruvinskaya, and Evgeniy Berkovich). The final paper, “Extended Regression on Manifolds Estimation” by Alexander Kuleshov and Alexander Bernstein, solves several interrelated problems in the area of regression on manifolds.

We are very grateful to the Program and Organizing Committees; the success of the symposium would have been impossible without their hard work. We are also indebted to the sponsors: Royal Holloway, University of London, and CIEMAT. Our special thanks to Yandex for their help and support in organizing the symposium and the special Alexey Chervonenkis Memorial Lecture.

March 2016

Alexander Gammerman  
Zhiyuan Luo  
Jesús Vega  
Vladimir Vovk



Conformal and Probabilistic Prediction with Applications  
5th International Symposium, COPA 2016, Madrid,  
Spain, April 20-22, 2016, Proceedings  
Gammerman, A.; Luo, Z.; Vega, J.; Vovk, V. (Eds.)  
2016, XII, 229 p. 45 illus., Softcover  
ISBN: 978-3-319-33394-6