Ann Math Artif Intell (2015) 74:1–7
DOI 10.1007/s10472-014-9429-3

Guest editors' preface to the special issue on conformal prediction and its applications

Harris Papadopoulos · Vladimir Vovk · Alexander Gammerman

Published online: 12 September 2014

Quantifying the uncertainty of predictions produced by classification and regression techniques is an important problem in the field of Machine Learning. This special issue is dedicated to Conformal Prediction (CP), which is a recently developed framework for producing provably valid measures of confidence in predictions. It can be used for extending conventional Machine Learning algorithms and thus developing methods, called Conformal Predictors (CPs), whose predictions are guaranteed to satisfy a given level of confidence without requiring anything more than that the data are generated independently by the same probability distribution (i.i.d.). More specifically, CPs produce as their predictions *prediction regions*, which are sets of labels sufficient to satisfy the required level of confidence.

The idea behind Conformal Prediction originated in a series of discussions at Royal Holloway, University of London, in the summer of 1996, between Alexander Gammerman, Vladimir N. Vapnik and Vladimir Vovk. These discussions, which were mainly concerned with Vapnik's work on Support Vector Machines (SVM), led to the realization that the number of support vectors used by an SVM could serve as basis for the production of confidence measures for individual predictions (in fact, similar ideas may be traced back to the joint work of Chervonenkis and Vapnik in the 1960s: see [5], footnote 4). This initial idea was described in [8] and was later improved to its current form in [37] and [33]. The first Conformal Predictor proposed in [37] and [33], which was then called Transductive Confidence

H. Papadopoulos (✉)
Computer Science and Engineering Department, Frederick University, 7 Y. Frederickou St., Palouriotisa, Nicosia 1036, Cyprus
e-mail: h.papadopoulos@frederick.ac.cy

V. Vovk · A. Gammerman
Computer Learning Research Centre, Department of Computer Science
Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK

V. Vovk
e-mail: v.vovk@rhul.ac.uk

A. Gammerman
e-mail: a.gammerman@cs.rhul.ac.uk

Machine, was based on SVMs. Soon the framework started being applied to other popular machine learning algorithms, such as k-nearest neighbours for both classification [32] and regression [31], Ridge Regression [20], Random Forests [4] and Genetic Algorithms [13]. At the same time comparisons of the CP framework with the Bayesian framework and the the theory of Probably Approximately Correct learning (PAC theory) were performed in [18] and [21].

In 2002 a modified version of the framework called Inductive Conformal Prediction (ICP) was proposed in [29] and [30] for the batch setting and in [36] for the on-line setting. ICP was developed in order to addresses the relative computational inefficiency of CP by replacing, as its name suggests, the transductive inference used in the latter with inductive inference. ICP has also been combined with a number of conventional machine learning algorithms such as Ridge Regression [29], k-nearest neighbours for both classification [30] and regression [31] and Artificial Neural Networks for both classification [22] and regression [27]. In these studies ICP has been shown to have an important computational efficiency advantage over CP without suffering a significant accuracy loss when dealing with large or even moderately large data sets.

Another modification of CP called Venn Prediction (VP) was proposed in 2004 by Vovk et al. [39]. VP produces well-calibrated *multiprobabilistic predictions*, which are sets of probability distributions for the true classification of each new example. Such a set can be summarized by lower and upper bounds for the conditional probability of the new example belonging to each one of the possible classes. The resulting bounds are guaranteed to contain well-calibrated probabilities (up to statistical fluctuations). Again, as in the case of CP, the only assumption made for obtaining this guarantee is that the data are i.i.d. The VP framework has been combined with the Nearest Neighbour classifier in [39] and [7], with Support Vector Machines in [16, 41], with Logistic Regression in [19] and with Artificial Neural Networks (ANN) in [23, 24].

In 2005 a book titled "Algorithmic Learning in a Random World" authored by Vovk, Gammerman and Glenn Shafer was published by Springer [38]. This book describes in detail the theoretical background of CP and its modifications together with examples of their combination with particular conventional algorithms. It also introduces a generalization of CP and VP called *On-line Compression Modeling*, which generalizes the i.i.d. assumption to any data generating mechanism that can be summarized by what is called an *on-line compression model*. Conformal prediction was also the subject of the second *Computer Journal* lecture of the British Computer Society by Gammerman and Vovk titled "Hedging Predictions in Machine Learning" [9].

Since their development CP and ICP have been applied successfully to a variety of challenging real world problems, such as the early detection of ovarian cancer [10], the classification of leukaemia subtypes [3], the diagnosis of acute abdominal pain [26], the assessment of stroke risk [15], the recognition of hypoxia in electroencephalograms (EEGs) [40], the recognition of gestures [35], the prediction of plant promoters [34], the assessment of the risk of complications following a coronary drug eluting stent procedure [1], the prediction of network traffic demand [6] and the estimation of effort for software projects [28]. VP has also been applied to important real world problems, such as the classification of internet traffic [7], the detection of Vesicoureteral Reflux [25] and the assessment of osteoporosis risk [14]. The framework has also been extended to additional problem settings such as semi-supervised learning, anomaly detection [17], feature selection [2], outlier detection, change detection in streams [12] and active learning [11].

In 2012 the first Workshop on Conformal Prediction and its Applications (CoPA 2012) was organized as part of the eighth IFIP International Conference on Artificial Intelligence

Applications and Innovations, which was held in September 27–30, 2012 at the Sithonia peninsula of Halkidiki, Greece. The success of this first Workshop, at which ten quality papers were presented, lead to this Special Issue and the organization of the second CoPA Workshop in 2013, in Paphos, Cyprus. The third CoPA Workshop will take place this year (2014) in Rhodes, Greece. This special issue consists of extended versions of the best papers presented at CoPA 2012 together with independent submissions. In total eleven high quality manuscripts are included in this issue, which were accepted after a rigorous two-round review process. Next we give a short description of the papers presented in this issue.

In the first paper, Vladimir Vovk proposes a new CP method called Cross-conformal Prediction (CCP), which is a hybrid of Inductive Conformal Prediction and Cross-validation. As a result of this combination, CCP has comparable computational efficiency with that of the ICP while its empirical validity and informational efficiency are studied using the well-known Spambase and USPS data sets. Additionally the paper develops a conditional version of CCP and empirically demonstrates its conditional validity.

Jing Lei et al. consider the problem of exploratory analysis and visualization of functional data. The authors extend conformal prediction to construct simultaneous prediction bands based on finite dimensional projections. In particular the constructed prediction bands are guaranteed to have a required level of coverage of a random curve drawn from the underlying process. Additionally the authors propose a new approach based on CP for the construction of clustering trees with finite sample interpretation for functional data, in which case the standard notion of density is no longer applicable and has to be replaced by that of pseudo-density.

Vineeth N. Balasubramanian et al. propose a methodology for information fusion under the CP framework in both the classification and regression settings. Their methodology applies CP to each data source independently and then combines the p-values obtained from each source to produce the fused prediction region for the required confidence level. The authors examine the use of six different methods for combining the p-values obtained from the individual classifiers and regressors. Specifically they experiment with three quantile combination methods, two order statistic methods and one learning-based method.

Rikard Laxhammar and Göran Falkman consider the problem of detecting anomalous sub-trajectories on-line. They propose an algorithm called *Inductive Conformal Anomaly Detector* (ICAD), which is a general anomaly detection method with a well-calibrated alarm rate based on the concept of ICP. They also propose a non-conformity measure for ICAD, called *Sub-Sequence Local Outlier* (SSLO), which is designed for examples represented as sequences of data points of varying length, such as trajectories. The authors examine the performance of the proposed technique, ICAD with SSLO, on an unlabeled data set of real vessel trajectories.

Frank-Michael Schleif et al. examine the application of CP in problems in which the data are expressed in terms of domain specific similarities or dissimilarities, also called proximity data. They extend a recent prototype based classifier for dissimilarity data using the CP framework. The use of CP enables the proposed approach to both produce reliable confidence measures with each prediction and to automatically adjust the number of prototypes to be used in the model.

Martin Eklund et al. investigate the application of CP to the drug discovery process. Specifically they apply ICP for the estimation of confidence in the predictions of Quantitative Structure-Activity Relationships (QSAR) models for properties of chemical compounds. The estimation of confidence in QSAR predictions is very important for

making informed decisions in the particular field. The authors experiment on four large data sets, two of which are binary classification and two are regression, of historical data collected from four different experiment assays.

Khuong An Nguyen and Zhiyuan Luo consider the problem of indoor localisation based on the signal strength of wireless Bluetooth. The authors employ CP to achieve better accuracy in location predictions and to provide a reliability estimate with each prediction. They consider both a classification and a regression approach. In the case of classification they propose a new underlying algorithm for indoor localisation based on the Weighted k Nearest Neighbour method. In the case of Regression they use the Ridge Regression Conformal Predictor proposed in [20]. The proposed approaches are evaluated on data sets collected in two different environments.

Andrea Murari et al. investigate the development of CPs based on Fuzzy Logic Classifiers both in the supervised and unsupervised setting. The authors define a number of CPs based on fuzzy membership function outputs. An advantage of the proposed methods is that they can be used in conjunction with any fuzzy membership function and parameters according to the problem at hand. The behavior of the proposed Fuzzy Logic CPs is examined on one-dimensional and two-dimensional artificial data.

Antonis Lambrou et al. propose an inductive version of the Venn Prediction framework. The motivation of this work, as in the case of the ICP, is the relative computational inefficiency of the original transductive version of the framework. The authors perform an extensive experimental evaluation of the proposed version of the framework combined with a SVM as underlying algorithm in terms of computational efficiency, accuracy and quality of its probabilistic outputs in both the online and the batch settings.

Ilia Nouretdinov et al. examine the application of Venn Prediction to proteomic mass spectrometry data. Specifically the authors apply a Venn Predictor based on logistic regression to three real-world MALDI-TOF data sets concerning the diagnosis of heart disease and the early detection of ovarian and breast cancer. Their experimental results on the three data sets confirm the empirical validity of the proposed approach. Moreover the produced probabilistic intervals are shown to be very narrow and therefore almost as precise as single probabilities. In terms of accuracy both the VP and its underlying algorithm provide highly accurate predictions well in advance of the actual diagnosis.

Finally, Jesús Vega et al. apply two concepts from the CP framework to nuclear fusion diagnostics. Specifically they first examine the use of VP for predicting the number of perturbations that can appear at any time in the plasma emissivity. To this end they apply VP combined with the Nearest Centroid algorithm to nine simulated data sets and compare the results to those of the Naive Bayes classifier. The authors then consider the case in which labeled training examples do not exist and apply exchangeability testing for detecting the plasma transition from an unperturbed to a perturbed state.

In concluding this preface we would like to thank all the authors for submitting their work to this special issue and all the reviewers for dedicating their time and effort to the reviewing process. We are also grateful to the AMAI Editor-in-Chief Professor Martin Golumbic for approving and supporting this special issue.

Harris Papadopoulos
Vladimir Vovk
Alexander Gammerman

References

1. Balasubramanian, V.N., Gouripeddi, R., Panchanathan, S., Vermillion, J., Bhaskaran, A., Siegel, R.M.: Support vector machine based conformal predictors for risk of complications following a coronary drug eluting stent procedure. In: Proceedings of the IEEE Conference on Computers in Cardiology 2009, pp. 5–8 (2009)
2. Bellotti, T., Luo, Z., Gammerman, A.: Strangeness minimisation feature selection with confidence machines. In: Intelligent Data Engineering and Automated Learning (IDEAL 2006), LNCS, vol. 4224, pp. 978–985. Springer (2006)
3. Bellotti, T., Luo, Z., Gammerman, A., Delft, F.W.V., Saha, V.: Qualified predictions for microarray and proteomics pattern diagnostics with confidence machines. Int. J. Neural Syst. **15**(4), 247–258 (2005)
4. Bhattacharyya, S.: Confidence in predictions from random tree ensembles. In: Proceedings of the 11th IEEE International Conference on Data Mining (ICDM 2011), pp. 71–80. Springer (2011)
5. Chervonenkis, A.: Chervonenkis's recollections. In: Vovk, V., Papadopoulos, H., Gammerman, A. (eds.) Measures of Complexity: Festschrift for Alexey Chervonenkis, chap. 1. Springer, Berlin (2014). To appear
6. Dashevskiy, M., Luo, Z.: Network traffic demand prediction with confidence. In: Proceedings of the IEEE Global Telecommunications Conference 2008 (GLOBECOM 2008), pp. 1453–1457. IEEE (2008)
7. Dashevskiy, M., Luo, Z.: Reliable probabilistic classification and its application to internet traffic. In: Proceedings of the 4th international conference on Intelligent Computing (ICIC 2008): Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues, LNCS, vol. 5226, pp. 380–388. Springer (2008)
8. Gammerman, A., Vapnik, V., Vovk, V.: Learning by transduction. In: Proceedings of the fourteenth conference on uncertainty in artificial intelligence, pp. 148–156. Morgan Kaufmann, San Francisco (1998)
9. Gammerman, A., Vovk, V.: Hedging predictions in machine learning: The second computer journal lecture. Comput. J. **50**(2), 151–163 (2007)
10. Gammerman, A., Vovk, V., Burford, B., Nouretdinov, I., Luo, Z., Chervonenkis, A., Waterfield, M., Cramer, R., Tempst, P., Villanueva, J., Kabir, M., Camuzeaux, S., Timms, J., Menon, U., Jacobs, I.: Serum proteomic abnormality predating screen detection of ovarian cancer. Comput. J. **52**(3), 326–333 (2009)
11. Ho, S.S., Wechsler, H.: Transductive confidence machine for active learning. In: Proceedings of the International Joint Conference on Neural Networks 2003, vol. 2, pp. 1435–1440 (2003)
12. Ho, S.S., Wechsler, H.: A martingale framework for detecting changes in data streams by testing exchangeability. IEEE Trans. Pattern Anal. Mach. Intell. **32**(12), 2113–2127 (2010)
13. Lambrou, A., Papadopoulos, H., Gammerman, A.: Reliable confidence measures for medical diagnosis with evolutionary algorithms. IEEE Trans. Inf. Technol. Biomed. **15**(1), 93–99 (2011)
14. Lambrou, A., Papadopoulos, H., Gammerman, A.: Osteoporosis risk assessment with well-calibrated probabilistic outputs. In: Artificial Intelligence Appications and Innovations (AIAI 2013), IFIP AICT, vol. 412, pp. 432–441. Springer (2013)
15. Lambrou, A., Papadopoulos, H., Kyriacou, E., Pattichis, C.S., Pattichis, M.S., Gammerman, A., Nicolaides, A.: Evaluation of the risk of stroke with confidence predictions based on ultrasound carotid image analysis. Int. J. Artif. Intell. Tools **21**(4), 1240,016 (2021). doi:10.1142/S0218213012400167
16. Lambrou, A., Papadopoulos, H., Nouretdinov, I., Gammerman, A.: Reliable probability estimates based on support vector machines for large multiclass datasets. In: Artificial Intelligence Appications and Innovations (AIAI 2012), IFIP AICT, vol. 382, pp. 182–191. Springer (2012)
17. Laxhammar, R., Falkman, G.: Online detection of anomalous sub-trajectories: A sliding window approach based on conformal anomaly detection and local outlier factor. In: Artificial Intelligence Appications and Innovations (AIAI 2012), IFIP AICT, vol. 382, pp. 192–202. Springer (2012)
18. Melluish, T., Saunders, C., Nouretdinov, I., Vovk, V.: Comparing the Bayes and Typicalness frameworks. In: Proceedings of the 12th European Conference on Machine Learning (ECML'01), Lecture Notes in Computer Science, vol. 2167, pp. 360–371. Springer (2001)
19. Nouretdinov, I., Devetyarov, D., Burford, B., Camuzeaux, S., Gentry-Maharaj, A., Tiss, A., Smith, C., Luo, Z., Chervonenkis, A., Hallett, R., Vovk, V., Waterfield, M., Cramer, R., Timms, J.F., Jacobs, I., Menon, U., Gammerman, A.: Multiprobabilistic venn predictors with logistic regression. In: Artificial Intelligence Appications and Innovations (AIAI 2012), IFIP AICT, vol. 382, pp. 224–233. Springer (2012)

20. Nouretdinov, I., Melluish, T., Vovk, V.: Ridge regression confidence machine. In: Proceedings of the 18th International Conference on Machine Learning (ICML'01), pp. 385–392. Morgan Kaufmann, San Francisco (2001)
21. Nouretdinov, I., Vovk, V., Vyugin, M.V., Gammerman, A.: Pattern recognition and density estimation under the general i.i.d. assumption. In: Proceedings of the 14th Annual Conference on Computational Learning Theory and 5th European Conference on Computational Learning Theory, Lecture Notes in Computer Science, vol. 2111, pp. 337–353. Springer (2001)
22. Papadopoulos, H.: Inductive Conformal Prediction: Theory and application to neural networks. In: Fritzsche, P. (ed.) Tools in Artificial Intelligence, chap. 18, pp. 315330 (2008). InTech, Vienna, Austria. http://www.intechopen.com/download/pdf/pdfs_id/5294
23. Papadopoulos, H.: Reliable probabilistic prediction for medical decision support. In: Proceedings of the 7th IFIP International Conference on Artificial Intelligence Appications and Innovations (AIAI 2011), IFIP AICT, vol. 364, pp. 265–274. Springer (2011)
24. Papadopoulos, H.: Reliable probabilistic classification with neural networks. Neurocomputing **107**, 59–68 (2013)
25. Papadopoulos, H., Anastassopoulos, G.: Probabilistic prediction for the detection of vesicoureteral reflux. In: Proceedings of the 14th International Conference on Engineering Applications of Neural Networks (EANN 2013), CCIS, vol. 383, pp. 253–262. Springer (2013)
26. Papadopoulos, H., Gammerman, A., Vovk, V.: Reliable diagnosis of acute abdominal pain with conformal prediction. Eng. Intell. Syst. **17**(2-3), 115–126 (2009)
27. Papadopoulos, H., Haralambous, H.: Reliable prediction intervals with regression neural networks. Neural Netw. **24**(8), 842–851 (2011). doi:10.1016/j.neunet.2011.05.008
28. Papadopoulos, H., Papatheocharous, E., Andreou, A.S.: Reliable confidence intervals for software effort estimation. In: Proceedings of the 2nd Workshop on Artificial Intelligence Techniques in Software Engineering (AISEW 2009), CEUR Workshop Proceedings, vol. 475 (2009). CEUR-WS.org. ceur-ws.org/Vol-475/AISEW2009/22-pp-211-220-208.pdf
29. Papadopoulos, H., Proedrou, K., Vovk, V., Gammerman, A.: Inductive confidence machines for regression. In: Proceedings of the 13th European Conference on Machine Learning (ECML'02), LNCS, vol. 2430, pp. 345–356. Springer (2002)
30. Papadopoulos, H., Vovk, V., Gammerman, A.: Qualified predictions for large data sets in the case of pattern recognition. In: Proceedings of the 2002 International Conference on Machine Learning and Applications (ICMLA'02), pp. 159–163. CSREA Press (2002)
31. Papadopoulos, H., Vovk, V., Gammerman, A.: Regression conformal prediction with nearest neighbours. J. Artif. Intell. Res. **40**, 815–840 (2011). doi:10.1613/jair.3198
32. Proedrou, K., Nouretdinov, I., Vovk, V., Gammerman, A.: Transductive confidence machines for pattern recognition. In: Proceedings of the 13th European Conference on Machine Learning (ECML'02), LNCS, vol. 2430, pp. 381–390. Springer (2002)
33. Saunders, C., Gammerman, A., Vovk, V.: Transduction with confidence and credibility. In: Proceedings of the 16th International Joint Conference on Artificial Intelligence, vol. 2, pp. 722–726. Morgan Kaufmann, Los Altos (1999)
34. Shahmuradov, I.A., Solovyev, V.V., Gammerman, A.J.: Plant promoter prediction with confidence estimation. Nucleic Acids Res. **33**(3), 1069–1076 (2005)
35. Sprinkhuizen-Kuyper, I., Vuurpijl, L., van Pinxteren, Y.: Reliable gesture recognition with transductive confidence machines. In: Dai, H., Liu, J.N., Smirnov, E. (eds.) Discovery, ReliableKnowledge, pp. 183–200, Springer (2012)
36. Vovk, V.: On-line confidence machines are well-calibrated. In: Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science, pp. 187–196. IEEE Computer Society (2002)
37. Vovk, V., Gammerman, A., Saunders, C.: Machine-learning applications of algorithmic randomness. In: Proceedings of the 16th International Conference on Machine Learning (ICML'99), pp. 444–453. Morgan Kaufmann, San Francisco (1999)
38. Vovk, V., Gammerman, A., Shafer, G.: Algorithmic learning in a random world. Springer, New York (2005)
39. Vovk, V., Shafer, G., Nouretdinov, I.: Self-calibrating probability forecasting. In: Advances in Neural Information Processing Systems, vol. 16, pp. 1133–1140. MIT Press (2004)
40. Zhang, J., Li, G., Hu, M., Li, J., Luo, Z.: Recognition of hypoxia EEG with a preset confidence level based on EEG analysis. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN 2008), part of the IEEE World Congress on Computational Intelligence (WCCI 2008), pp. 3005–3008. IEEE (2008)

41. Zhou, C., Nouretdinov, I., Luo, Z., Adamskiy, D., Randell, L., Coldham, N., Gammerman, A.: A comparison of venn machine with platts method in probabilistic outputs. In: Proceedings of the 7th IFIP International Conference on Artificial Intelligence Appications and Innovations (AIAI 2011), IFIP AICT, vol. 364, pp. 483–490. Springer (2011)