Conformal Multistep-Ahead Multivariate Time-Series Forecasting

Filip Schlembach   F.SCHLEMBACH@STUDENT.MAASTRICHTUNIVERSITY.NL
Evgueni Smirnov   SMIRNOV@MAASTRICHTUNIVERSITY.NL
Maastricht University, Maastricht, The Netherlands

Irena Koprinska   IRENA.KOPRINSKA@SYDNEY.EDU.AU
The University of Sydney, Sydney, Australia

Abstract

This paper proposes a method for conformal multistep-ahead multivariate time-series forecasting. The method minimizes the coverage loss when the data exchangeability assumption does not properly hold. This is done by weighting residual quantiles while computing prediction intervals. Preliminary experiments on real data demonstrate the method's utility.
Keywords: Conformal Prediction, Time Series Forecasting

1. Introduction

Background   Time series forecasting in critical domain applications is required to provide predictions together with uncertainty quantification (to improve decision making). Recently, conformal prediction (CP) has been proposed for this task since it allows prediction intervals with a theoretic coverage guarantee when the data is finite and no assumption on data distribution is imposed. However, the main obstacle in this context is the data exchangeability requirement needed for the coverage guarantee (Vovk et al., 2005). In this paper we consider two recent adaptations of conformal prediction that deal with this obstacle and can be used for time series forecasting (Barber et al., 2022; Stankevičiūtė et al., 2021). They are considered in the context of inductive conformal prediction (Papadopoulos et al., 2002) and then combined in a new method proposed in this paper.

Inductive Conformal Prediction   Consider a set D of examples $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$ consisting of objects x_i and their associated labels y_i with $i = 1, \ldots, n$. The inductive conformal predictor (ICP) first splits D into a proper training set D_{train} of examples (x_i, y_i) with $i = 1, \ldots, l$ and a calibration set D_{cal} of examples (x_i, y_i) with $i = l+1, \ldots, n$. Then it fits an underlying model $\hat{\mu}$ to D_{train} and uses the empiric quantiles Q of the residuals $r_i = |y_i - \hat{\mu}(x_i)|$ of the calibration examples $(x_i, y_i) \in D_{cal}$ to generate prediction interval $\hat{\mathbf{y}}_{n+1} = \hat{\mu}(x_{n+1}) \pm \mathsf{Q}_{1-\alpha}(\{r_i, i = l+1, \ldots, n\})$ for any new object x_{n+1} on confidence level $1 - \alpha$. If D is exchangeable, Vovk et al. (2005) show that ICP has a coverage guarantee.

Beyond Exchangeability   Recently, multiple attempts have been made to lift the exchangeability requirement. Barber et al. (2022) introduce the *coverage gap* that measures by how much the prediction intervals generated by ICP undercover the real labels compared to the targeted coverage rate. It is proven that this coverage gap is theoretically bound even if the data is not exchangeable. In addition, Barber et al. (2022) introduce a weighted quantile function $\mathsf{Q}_{1-\alpha}(\{(r_i, w_i), i = l+1, \ldots, n\})$ in the ICP setting. By reducing

the weights w_i of nonconformity scores that contribute to an increase in the coverage gap, the quality of the predicted intervals can be improved. If, for instance, the weights decrease more for older training examples, the method is able to adapt to distribution shifts in an online setting. However, using the residuals as nonconformity measure limits the method to single target and one-step-ahead predictions.

Multistep-Ahead Time Series In this task, we have a series of ordered observations $o_j \in \mathbb{R}, j = 1, \ldots, t$ and need to generate predictions $\hat{o}_j \in \mathbb{R}, j = t+1, \ldots, t+h$ for the next h time steps. Stankevičiūtė et al. (2021) assume exchangeable set D of examples (x_i, y_i) with $i = 1, \ldots, n$ where $x_i = [o_1^{(i)}, \ldots, o_t^{(i)}]^T$ and $y_i = [o_{t+1}^{(i)}, \ldots, o_{t+h}^{(i)}]^T$ which is split into a proper training set D_{train} and a calibration set D_{cal} as for ICP. A model $\hat{\mu} : \mathbb{R}^t \to \mathbb{R}^h$ is fitted to D_{train} and used to generate predictions $\hat{y}_i, i = l+1, \ldots, n$ for D_{cal}. This results in h-dimensional residuals $r_i = [|y_{i,j} - \hat{\mu}(x_i)_j|, j = 1, \ldots, h]^T$ for each calibration example (x_i, y_i). Given a new object x_{n+1}, the residuals are used to construct a prediction region $\hat{y}_{n+1} = [\hat{\mu}(x_{n+1}) \pm \mathsf{Q}_{1-\alpha/h}(\{r_{i,j}, i = l+1, \ldots, n\}), j = 1, \ldots, h]^T$ from h conformal predictors using the Bonferroni correction to reach the desired coverage rate.

2. New Method

Our new method is essentially the method of conformal multistep-ahead time series of (Stankevičiūtė et al., 2021) that relaxes the exchangeability assumption using the method of (Barber et al., 2022). It produces set D of examples (x_i, y_i) from the time-series data following (Stankevičiūtė et al., 2021), however, without assuming that the set is exchangeable. Thus, prediction region $\hat{y}_{n+1}$ for any new example x_{n+1} is computed as $[\hat{\mu}(x_{n+1}) \pm \mathsf{Q}_{1-\alpha/h}(\{(r_{i,j}, w_{i,j}), i = l+1, \ldots, n\}), j = 1, \ldots, h]^T$ by ICP that employs the weighted quantile function from (Barber et al., 2022). We further extend our new method to the multivariate time-series setting assuming that $o_j \in \mathbb{R}^d$ which implies that the prediction region $\hat{y}_{n+1}$ is in $\mathbb{R}^{2 \times d \times h}$. This is done by using the past residuals of a particular time step and dimension to estimate the prediction region.

3. Preliminary Results

We predict features *Nswdemand*, *Vicdemand* and *Transfer* using the first 20 000 entries of ELEC2 dataset (Harries, 1999) with $t = 192$, $h = 12$, and a stride of 12 between consecutive windows, resulting in 1650 examples. These are split into a proper training set, a calibration set, and a test set containing 660, 660 and 330 examples, respectively, along the time axis. We note that *features Vicdemand and Transfer display a strong distribution shift in the test set*, shown in Figures 1(*b*) and 1(*c*).

We choose a recurrent neural network as underlying model and fit it to the training set. We compute the residuals using the calibration set and evaluate the results on the test set. Whenever a test instance is processed, its residual is added to the residuals of the calibration instances in an online manner. The experiment is conducted five times for different random initializations of the neural network and the results are averaged. Figure 2 shows the coverage rate on the test set for different confidence levels $1 - \alpha$ using the standard quantile function (no weights), the weighted quantile function with linearly decreasing weights and the weighted quantile function with exponentially decreasing weights. Figure 3 displays the

average prediction interval width for different confidence levels $1 - \alpha$. For confidence values of $1 - \alpha > 0.5$, the use of the weighted quantile function produces valid intervals, even as the distribution of the test set shifts, which confirms the utility of the proposed method.
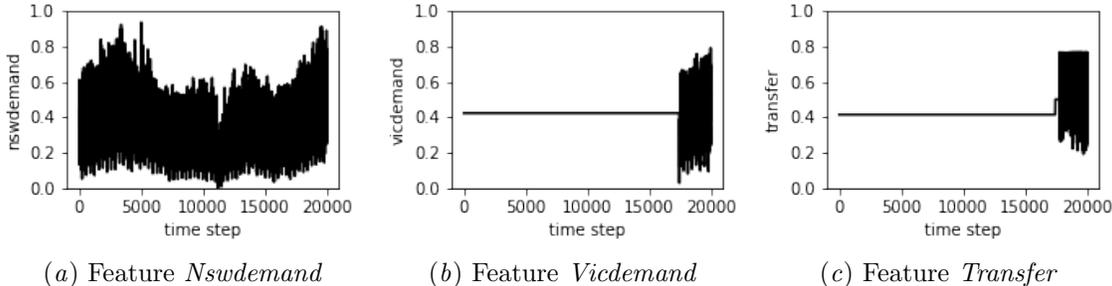

(a) Feature *Nswdemand* (b) Feature *Vicdemand* (c) Feature *Transfer*

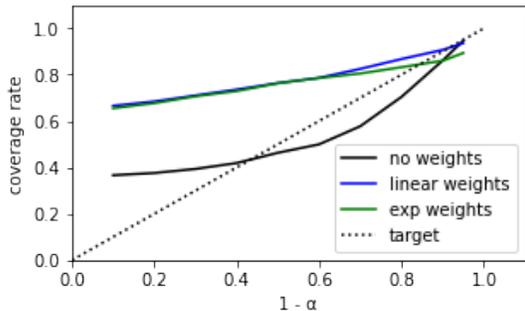Figure 1: Time distributions of features
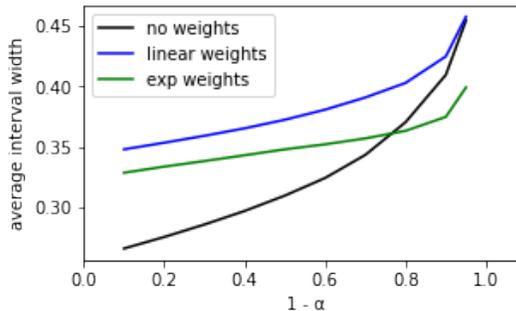

Figure 2: Empirical coverage rate.

Figure 3: Average interval width.

References

R.F. Barber, E.J. Candes, A. Ramdas, and R.J. Tibshirani. Conformal prediction beyond exchangeability. *arXiv:2202.13415 [stat]*, March 2022.

M. Harries. Splice-2 comparative evaluation: Electricity pricing. Technical report, School of Computer Science and Engineering, University of New South Wales, 1999.

H. Papadopoulos, K. Proedrou, V. Vovk, and A. Gammerman. Inductive Confidence Machines for Regression. In *Proceedings of 13th European Conference on Machine Learning (ECML 2002)*, volume 2430, pages 345–356. Springer, 2002.

K. Stankevičiūtė, A.M. Alaa, and M. van der Schaar. Conformal Time-Series Forecasting. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, 2021.

V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic learning in a random world.* Springer, New York, 2005.