

Reliable Predictive Intervals for the Critical Frequency of the F2 Ionospheric Layer

Harris Papadopoulos and Haris Haralambous¹

Abstract. This paper addresses the problem of reliably predicting an important HF communication systems parameter, the critical frequency of the F2 ionospheric layer, with the use of a new machine learning technique, called Conformal Prediction (CP). CP accompanies the predictions of traditional machine learning algorithms with measures of confidence. The proposed approach is based on the well-known Ridge Regression technique, but instead of the point predictions produced by the original method, it produces predictive intervals that satisfy a given confidence level. Our experimental results on an extended critical frequency dataset show that the obtained intervals are well-calibrated and narrow enough to be useful in practice.

1 INTRODUCTION

HF communications is an effective communication mode which facilitates long and short range links by taking advantage of the ability of the ionosphere to reflect radiowaves in the frequency range up to 30 MHz. The most decisive parameter in achieving viable links is the choice of the transmission frequency. The optimal selection of this parameter is related to the monthly median critical frequency of the F2 ionospheric layer which is a function of the electron density profile at a given location and is subject to diurnal and long-term ionospheric variability characteristics. In addition to these considerations one must also take into account the frequency allocations defined by the International Telecommunication Union (ITU) which impose frequency ranges within which potential users must confine their transmissions based on their type of service (e.g. amateur, broadcast, aeromobile). These criteria pose a significant challenge to frequency prediction procedures which must be in place to provide adequate long-term frequency management. In the context of these procedures it is very beneficial to specify certain predictive intervals for the selection of the critical frequency as a function of season, time-of-day and solar activity which are the influential factors on its variability. Predictive intervals provide a quantitative measure for the range within which a value may lie so that the chance of selecting a near optimum operating frequency is high.

In this work we propose the use of a recently developed machine learning framework, called Conformal Prediction (CP) [3], for producing predictive intervals that satisfy a required level of confidence. The most attractive feature of Conformal Predictors (CPs) is that they are well-calibrated, meaning that in the long run their predictive intervals for some confidence level $1 - \delta$ will not contain the true label of an example with a relative frequency of at most δ [3].

Due to the relatively large size of the dataset used in this study, we follow a modified version of the original CP approach, called

Inductive Conformal Prediction (ICP). ICP was first introduced in [1] as a solution to the computational inefficiency problem of CPs, which rendered them unsuitable for large datasets. The particular ICP we use is based on the dual variables Ridge Regression algorithm [2].

2 RIDGE REGRESSION INDUCTIVE CONFORMAL PREDICTION

We are interested in examples $z_i \in Z$ where each z_i is a pair (x_i, y_i) consisting of an object $x_i \in \mathbb{R}^d$ and its label $y_i \in \mathbb{R}$, and the only assumption we make is that all z_i are independent and identically distributed (i.i.d.). The main idea behind the CP framework is to assign every possible label $\tilde{y}$ to the new unlabeled example x_n and check how likely it is that the pair $(x_n, \tilde{y})$ was generated independently by the same probability distribution as a set of known examples. The ICP version of the approach requires two distinct sets of training examples, called the *proper training set* and the *calibration set*. The proper training set, which is typically much larger, is used for generating a decision rule D using a conventional machine learning algorithm, in this case dual form Ridge Regression (RR). The resulting decision rule is then used for assigning a value α_i to each example $z_i : i = 1, \dots, q$ in the calibration set that indicates how strange, or non-conforming, it is. This value, called the *non-conformity score* of the example $z_i = (x_i, y_i)$, is measured as the degree of disagreement between the prediction $\hat{y}_i = D(x_i)$ and the actual label y_i . As an example the typical non-conformity measure for regression is

$$\alpha_i = |y_i - \hat{y}_i|. \quad (1)$$

The same process is used for calculating the non-conformity score $\alpha_n^{\tilde{y}}$ of $(x_n, \tilde{y})$. This non-conformity score is then compared with those of the calibration examples to evaluate how unusual $\tilde{y}$ is. This is achieved with the function

$$p(\tilde{y}) = \frac{\#\{i = 1, \dots, q, n : \alpha_i \geq \alpha_n^{\tilde{y}}\}}{q + 1}. \quad (2)$$

The output of this function, which lies between $\frac{1}{q+1}$ and 1 is called the p-value of $\tilde{y}$.

An important property of (2) is that if $z_1, \dots, z_q, z_n$ are i.i.d., then $\forall \delta \in [0, 1]$ and for all probability distributions P on Z ,

$$P^{q+1}\{(z_1, \dots, z_q, z_n) : p(y_n) \leq \delta\} \leq \delta, \quad (3)$$

where y_n denotes the true label of x_n ; a proof can be found in [3]. According to this property, we could exclude all labels with a p-value under some very low threshold, or *significance level*, δ and have at most δ chance of being wrong. Consequently, given a confidence level $1 - \delta$ a regression ICP outputs the predictive interval

$$\{\tilde{y} : p(\tilde{y}) > \delta\}. \quad (4)$$

¹ Computer Science and Engineering Department, Frederick University, Cyprus, email: {h.papadopoulos, h.haralambous}@frederick.ac.cy

As it is impossible to explicitly consider every possible label $\tilde{y} \in \mathbb{R}$, the RR-ICP follows a different process for computing the interval (4) efficiently. It sorts the non-conformity scores of the calibration examples in descending order obtaining $\alpha_{(1)}, \dots, \alpha_{(q)}$ and outputs

$$(\hat{y}_n - \alpha_{(s)}, \hat{y}_n + \alpha_{(s)}), \quad (5)$$

where $s = \lfloor \delta(q+1) \rfloor$. This interval is equivalent to (4) provided the non-conformity scores are computed according to (1).

In this work an extended version of non-conformity measure (1) is used, which normalises (1) according to the predicted accuracy of D on the example in question. This measure is defined as

$$\alpha_i = \frac{|y_i - \hat{y}_i|}{e_i^\mu}, \quad (6)$$

where μ_i is the prediction of the value $\ln(|y_i - D(x_i)|)$ produced by a second RR decision rule generated using the training objects. This non-conformity measure was shown in [1] to give tighter predictive intervals than (1). In this case the predictive interval for x_n becomes

$$(\hat{y}_n - \alpha_{(s)} e_n^\mu, \hat{y}_n + \alpha_{(s)} e_n^\mu), \quad (7)$$

where again $s = \lfloor \delta(q+1) \rfloor$.

3 MEASUREMENT DATA

The data used in this study were recorded over two different measurement periods. The first consisted of measurements recorded from November 1987 to November 1996, while the second consisted of recent measurements from September 2008 to January 2010. The recorded measurements were used for creating two sets, consisting of 2046 and 384 values respectively, with the monthly median values of the F2 layer critical frequency for different hours. Its worth noting that these do not correspond to the median values of all 24 hours for each month, as there were many missing values in the recorded data.

The input features used for each monthly median value were: (a) the quadrature components of the hour-of-day defined as

$$\sinhour = \sin\left(2\pi \frac{hour}{24}\right) \quad (8)$$

and

$$\coshour = \cos\left(2\pi \frac{hour}{24}\right), \quad (9)$$

(b) the quadrature components of the month defined as

$$\sinmonth = \sin\left(2\pi \frac{month}{12}\right) \quad (10)$$

and

$$\cosmonth = \cos\left(2\pi \frac{month}{12}\right), \quad (11)$$

and (c) a monthly average index of solar activity, R12, to represent the long-term effect of the sun on the ionization of the F2 ionospheric layer, which directly affects its median critical frequency.

4 EXPERIMENTS AND RESULTS

Before conducting our experiments all features were normalised by setting their means to zero and their standard deviations to one. The first dataset of 2046 values was used as training set for generating the RR decision rule of the ICP, while the second was used for forming the calibration and test sets. In order to test the method on all 384

Table 1. Tightness and empirical reliability results.

	Confidence Level		
	90%	95%	99%
Median Interval Width (MHz)	2.00	2.44	2.89
Percentage of Label Range (%)	16.0	19.5	23.1
Percentage of Errors (%)	9.90	5.03	0.83

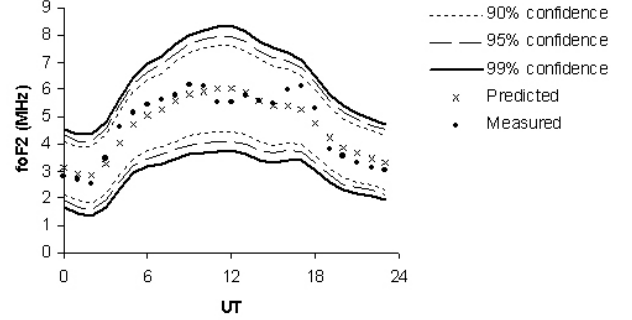


Figure 1. Actual and predicted values along with the 90%, 95% and 99% confidence intervals for April 2009.

values of the second dataset a 6-fold cross validation procedure was employed. The dataset was split into 6 equal parts and each part was used in turn as a test set with the remaining parts used for calibration. To ensure that the results were not dependent on the particular split of examples into parts this process was repeated 10 times with different splits. The kernel used by RR for the decision rule was RBF with $\sigma = 1$ and ridge parameter $a = 0.05$. These parameters were determined by trying out different values on 10-fold cross validation experiments with the training dataset.

Table 1 reports the results of the RR-ICP for the 90%, 95% and 99% confidence levels. The first row reports the median widths of the predictive intervals, while the second transforms these values into the percentage they represent over the range of label values in the data. The values reported here are very satisfactory considering that we can have confidence as high as 99% in a region that covers less than 1/4th (23.1%) of the range of possible values. The third row of the table checks the empirical reliability of the intervals by reporting the percentage of examples for which the true value was not included in the output interval. These values confirm that the predictive intervals of RR-ICP are well-calibrated. The plot in Figure 1 corresponds to the monthly median measured and predicted values of foF2 along with estimated confidence intervals for each hour corresponding to April 2009. The size of the predictive intervals appears to maximise around noon which is reasonable as the ionosphere is greatly ionized and highly variable around this time.

ACKNOWLEDGEMENTS

This work is funded by the Cyprus Research Promotion Foundation (project PROEM/0308/13: “Ionospheric Modeling and Prediction”).

REFERENCES

- [1] Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman, ‘Inductive confidence machines for regression’, in *Proceedings of the 13th European Conference on Machine Learning (ECML’02)*, volume 2430 of *LNCS*, pp. 345–356. Springer, (2002).
- [2] Craig Saunders, Alex Gammerman, and Volodya Vovk, ‘Ridge regression learning algorithm in dual variables’, in *Proceedings of the 15th International Conference on Machine Learning (ICML’98)*, pp. 515–521, San Francisco, CA, (1998). Morgan Kaufmann.
- [3] Vladimir Vovk, Alex Gammerman, and Glenn Shafer, *Algorithmic Learning in a Random World*, Springer, New York, 2005.