Inductive Venn-Abers Predictive Distributions: New Applications & Evaluation

Ilia Nouretdinov, James Gammerman
Centre for Reliable Machine Learning, Royal Holloway
University of London

September 7, 2024

Introduction

In this work, we revisit the Inductive Venn-Abers Predictive Distribution (IVAPD) framework for regression problems, first introduced in [Nouretdinov et al., 2018].

Contributions:

- Application of the IVAPD approach to real-world predictive maintenance and energy consumption forecasting tasks.
- Extension of the algorithm to online (real-time) learning settings.
- Examination of evaluation metrics for IVAPD.

Inductive Venn-Abers Predictive Distributions (IVAPD)

- We assume that the audience is familiar with Venn and Venn-ABERS prediction
- IVAPD is an extension of Inductive Venn-Abers that generates predictive distributions for continuous outcomes, offering calibrated uncertainty estimates for each prediction.
- It constructs a Cumulative Distribution Function (CDF) for each test instance, providing the probability that the outcome will fall below various thresholds.
- In our implementation, IVAPD updates predictions dynamically with each new instance, making it suitable for online scenarios.

IVAPD Algorithm: Initialization and Data Splitting

Step 1: Initialization

- Input dataset $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ where $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$
- Underlying predictor s for nearest neighbors

Step 2: Dynamic Data Splitting

- For each instance $j = 1, \ldots, n$, split data into:
  - Training set $T_P = \{(x_1, y_1), \ldots, (x_r, y_r)\}$
  - Calibration set $T_C = \{(x_{r+1}, y_{r+1}), \ldots, (x_h, y_h)\}$
- Apply feature selection (optional)
- Calculate scores $s_i = s(x_i, T_P \setminus (x_i, y_i))$ for training instances

IVAPD Algorithm: Calibration and Prediction

Step 3: Calibration using Isotonic Regression

- Apply isotonic regression on T_P to calibrate scores:

$$\sum_{i=1}^{r}(g(s_i) - y_i)^2 \to \min$$

Step 4: Scoring and Calibration for Test Example

- For each $x_j \in T_C$, calculate score $s_i = s(x_i, T_P)$
- Find s_k closest to s_i and assign $g_i := g_k$

Step 5: Construct Predictive Distribution

- Construct predictive set $\hat{Y} = \{y_i \in A : g_i = g_{h+1}\}$
- Calculate probabilities:

$$\hat{P}_0\{y_j \leq t\} = \frac{|\{\hat{y} \in \hat{Y} : \hat{y} \leq t\}|}{|A| + 1}, \quad \hat{P}_1\{y_j \leq t\} = \frac{|\{\hat{y} \in \hat{Y} : \hat{y} \leq t\}| + 1}{|A| + 1}$$

Datasets and Tasks

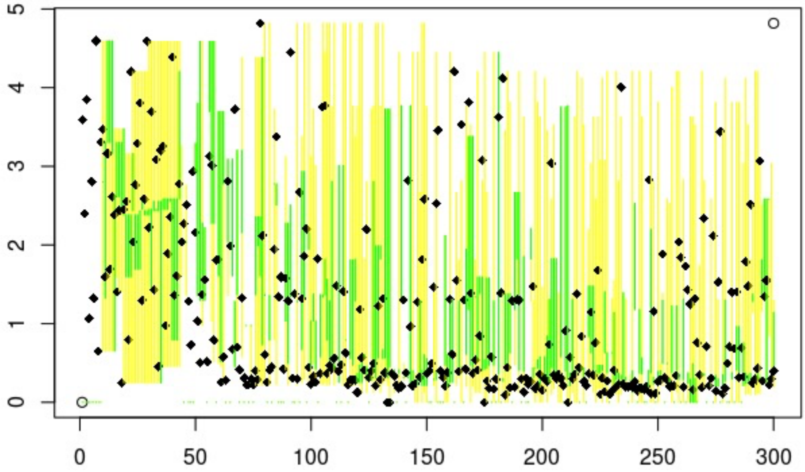
Naval Propulsion Plants Dataset ("NPP")

- 16 features related to Gas Turbine measurements.
- Labels: Compressor and Turbine degradation coefficients.

UCI Household Electric Power Consumption ("ECP")

- Predicts evening power consumption (at 18:00) based on data from the morning (00:00-12:00).

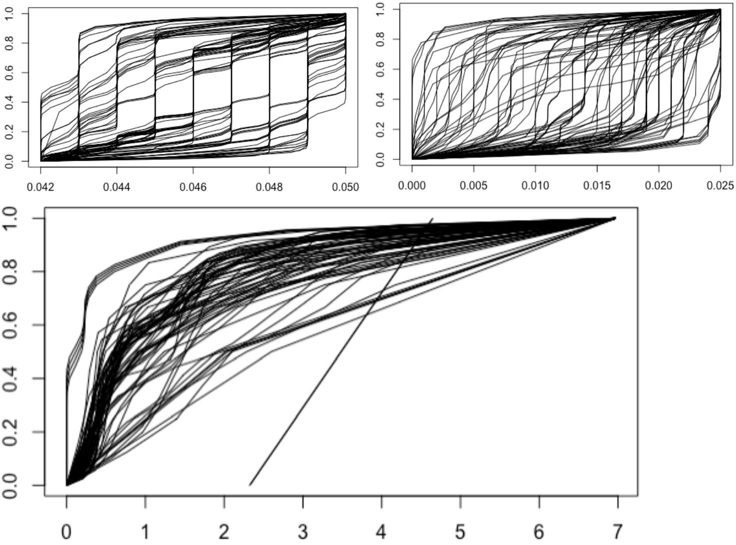
Visualising Predictive Distributions (ECP data)



- x-axis: Instance index ordered by time.
- y-axis: Energy consumption at 18:00.
- Black points: labels
- Yellow bars: full predictive distribution (green = IQR)

Cumulative Distribution Functions

Above: NPP Gas turbine dataset (2 different labels)
Below: ECP power consumption dataset



Evaluation Metrics

- **Scoring Rules**: Continuous Ranked Probability Score (CRPS) measures the mean squared difference between predicted and observed cumulative distributions.
- **Sharpness Measures**: Assess the precision of probabilistic predictions without requiring labels.
  - **Interval Width (W)**: the narrowest interval within which the label is expected to fall, with confidence level at least $(1 - \varepsilon)$.
  - **Variance (V)**: The average variance within the distribution; lower variance indicates tighter predictions.
  - **Probability Distribution Spread (P)**: Evaluates the average difference between upper and lower CDF estimates P_0 and P_1

Evaluating The Predictive Distributions

parameters feat.	nei.	C	V	W $\varepsilon = 0.25$	W $\varepsilon = 0.5$	W $\varepsilon = 0.75$	P
NPP(1) data set							
5	5	0.00117	0.00225	0.00424	0.00213	0.000582	0.0493
5	20	0.00130	0.00237	0.00484	0.00263	0.000809	0.0379
5	100	0.00143	0.00250	0.00577	0.00338	0.00110	0.0189
all	5	0.000964	0.00207	0.00339	0.00160	0.000369	0.0559
all	20	0.00124	0.00232	0.00454	0.00239	0.000744	0.0442
all	100	0.00143	0.00249	0.00579	0.00335	0.00107	0.0172
best param.		(all,5)	(all,5)	(all,5)	(all,5)	(all,5)	(all,100)
NPP(2) data set							
5	5	0.00183	0.00636	0.00773	0.00321	0.00116	0.108
5	20	0.00237	0.00635	0.00964	0.00450	0.00166	0.0880
5	100	0.00379	0.00729	0.0153	0.00841	0.00329	0.0507
all	5	0.00179	0.00642	0.00757	0.00283	0.000954	0.112
all	20	0.00239	0.00648	0.00993	0.00449	0.00158	0.0925
all	100	0.00415	0.00747	0.170	0.00989	0.00399	0.0294
best param.		(all,5)	(5,20)	(all,5)	(all,5)	(all,5)	(all,100)
ECP data set							
5	5	0.607	1.780	2.418	0.973	0.425	0.166
5	20	0.624	1.751	2.365	0.923	0.398	0.150
5	100	0.591	1.657	2.211	0.872	0.380	0.123
20	5	0.622	1.797	2.496	1.009	0.437	0.168
20	20	0.608	1.764	2.380	0.969	0.422	0.157
20	100	0.591	1.675	2.287	0.934	0.436	0.137
all	5	0.613	1.825	2.485	0.995	0.450	0.178
all	20	0.604	1.752	2.343	0.942	0.402	0.153
all	100	0.586	1.692	2.276	0.917	0.419	0.136
best param.		(all,100)	(5,100)	(5,100)	(5,100)	(5,100)	(5,100)

Conclusions & Future Work

- Explored IVAPD for regression, generating reliable predictive distributions.
- Demonstrated online application in energy consumption and predictive maintenance.
- Found interval width (W-criterion) to be a useful metric for evaluation when the true labels aren't available.

Future Work:

- Extend analysis to more datasets and underlying metrics.
- Compare with other probabilistic methods such as conformal predictive distributions and Bayesian approaches
- Explore additional metrics, not just accuracy and sharpness of individual predictions.

References

I. Nouretdinov, D. Volkhonskiy, P. Lim, P. Toccaceli, and A. Gammerman. Inductive venn-abers predictive distribution. *Proceedings of Conformal Prediction with Applications*, 91: 1–22, 2018.