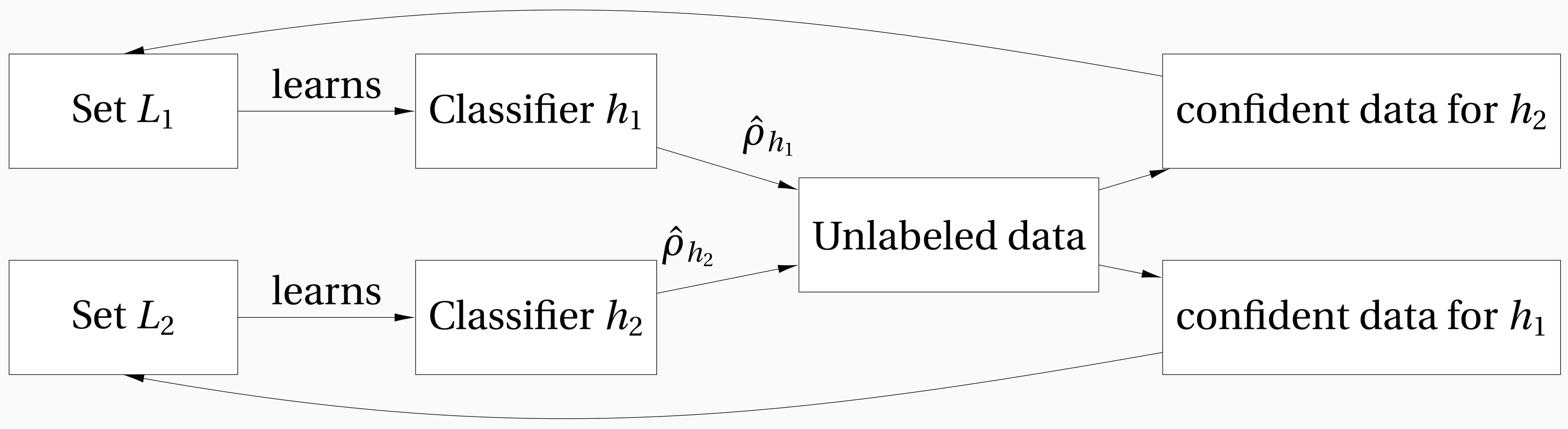
Learning calibrated belief functions from conformal predictions

Vitor Martin Bordini, Sebastien Destercke and Benjamin Quost

CNRS, UMR 7253 Heudiasyc, UTC

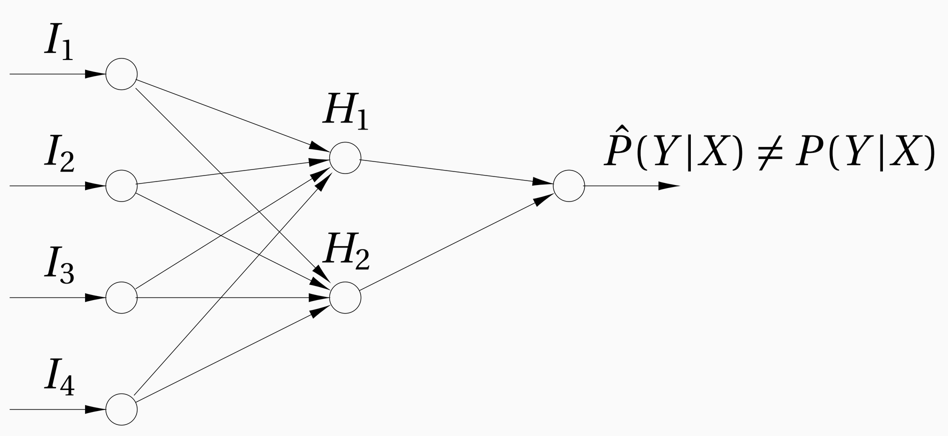
Objectives

- Estimate calibrated belief functions in a co-learning context.
- Use a minimum amount of calibration data.



Problem

- Common problem on machine learning predictions: poor calibration.
- Calibration definition: The level of confidence actually reflects the chance that the associated output turns out to be true.
- Inductive Conformal Prediction (ICP) [1] is a possible solution to this problem.
- What is the relation (if any) between ICPs and Belief Functions?

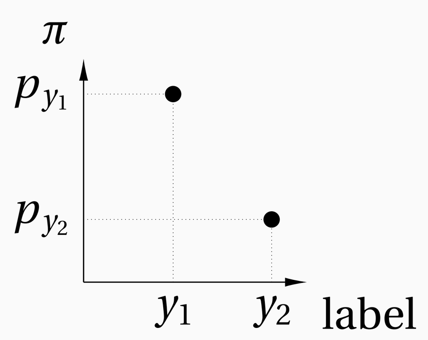


Inductive Conformal Prediction

- Dataset $Z = \{(x_i, w_i), w_i \in \Omega | i = 1, \ldots, n\}$ is exchangeable.
- Compute non-conformity scores α_i.
- Computes p-values, ICP output, by comparing the non-conformity scores of a single exemple and the ones of the calibration set.
- P-values property: $P(\{p(w_i) \leq \delta\}) \leq \delta, w_i \in \Omega$.
- In the exemple below, y_1 is a better prediction than y_2 because $p_{y_1} > p_{y_2}$.
- Advantages: Simple to implement/understand and with a rigorous theory behind it.
- Drawbacks: Calibration set (needs more data) and significantly slower.

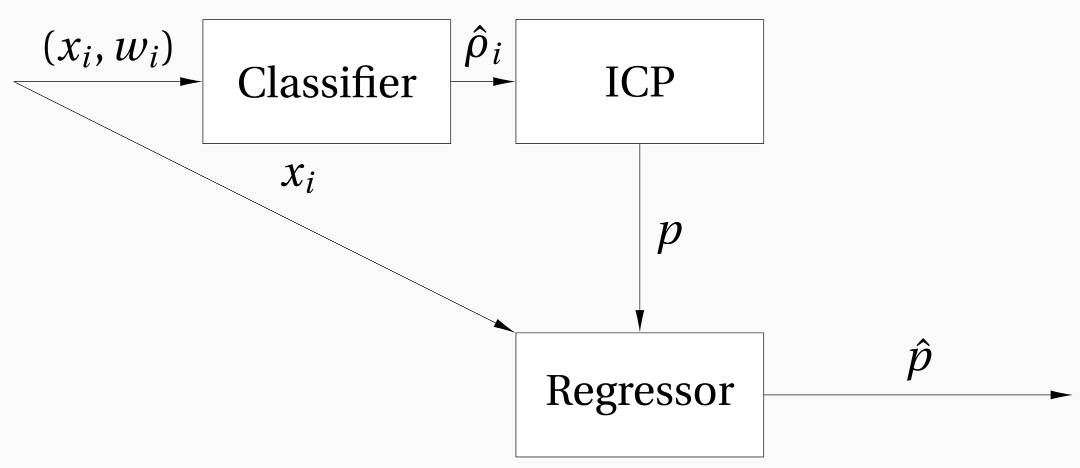
Possibility theory

- Possibility distribution: $\pi : \Omega \mapsto [0, 1]$
- Necessity Measure N and Possibility Measure Π (equivalent to Belief and Plausibilty Functions, respectively): $\Pi(B) = 1 - N(B^C) = \max_{x \in B} \pi(x), B \subseteq \Omega$.
- Limitation: we can only extract imprecise probabilities from a possibility distribution.
- α-cut: $\pi_\alpha = \{x \in \mathbb{R} | \pi(x) > \alpha\}$[2].
- Property: $P(\pi_\alpha) \geq 1 - \alpha$.
- In the example below, we can compute the mass functions as $m(\{y_1\}) = p_{y_1} - p_{y_2}$ and $m(\{y_1, y_2\}) = p_{y_2}$.



Our solution

- Our hypothesis: ICP outputs can be learned directly from a machine learning model.
- We need p-values as labels, which doesn't exist in any public datasets.
- We train a model that estimates probability distributions and then we apply the ICP on this model output to compute p-values.
- This p-values are the labels to train a regressor.
- P-value vector p can be interpreted as a possibility distribution π.
- Estimation of calibrated belief functions via possibility distribution.

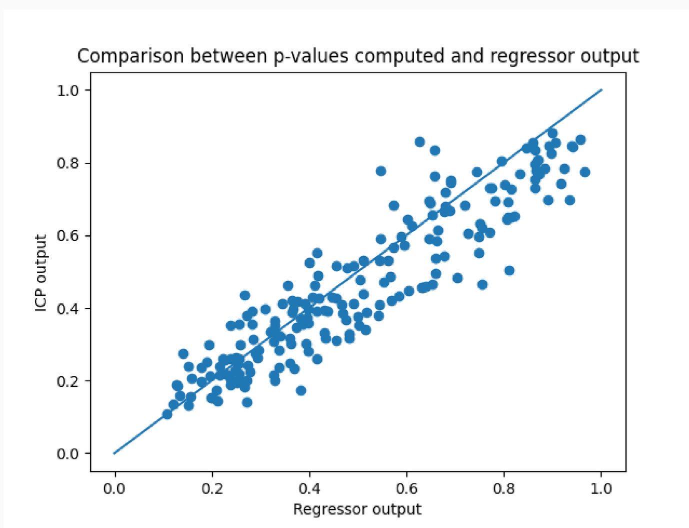


Experiments

- CIFAR-10 dataset.
- Classifier: Fitnet backbone [3] + softmax layer.
- Regressor: Fitnet backbone + linear layer with activation function $\phi(x) = \frac{e^{x_i}}{\max_{x_i \in x} e^{x_i}}$.
- $\pi(\emptyset) = 0$.
- Training parameters: batch size 25, learning rate 0.001, momentum 0.9 and an Adam optimizer.

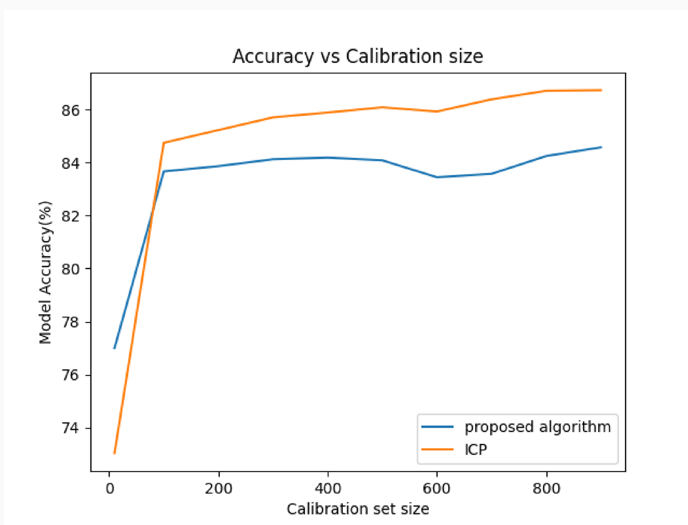
P-values comparison

- Comparison between ICP and the regressor outputs.
- Calibration dataset = 10% of the test dataset.
- The Mean Square Root(MSR) and the R2 coefficient are 0.02 and 0.8, respectively.
- Output averaged by the number of classes.
- Results for 200 outputs. Ideal result = Blue line.



Calibration size influence

- Goal = Change the calibration set size from 10 to 1000 instances and check how the accuracy of ICP and our algorithm grows.
- Regressor performs slightly worse after 100 instances, at most 2.5% below than the ICP, but has a better performance with less data.



Conclusion

- Calibration techniques make model predictions statistically valid.
- The ICP is a popular calibration technique but it is slower and requires more data.
- Our algorithm decrease the dependence of ICP on the calibration dataset while also being less computationally expensive and having similar performance.
- However, it still requires a minimum amount of data and takes more time to learn.
- Future works may solve this problem using co-learning techniques[4][5].

References

[1] Harris Papadopoulos. Inductive conformal prediction: Theory and application to neural networks. In Paula Fritzsche, editor, *Tools in Artificial Intelligence*, chapter 18. IntechOpen, Rijeka, 2008.

[2] Sebastien Destercke and Ines Couso. Ranking of fuzzy intervals seen through the imprecise probabilistic lens. *Fuzzy Sets and Systems*, 278:20–39, 2015. Special Issue on uncertainty and imprecision modelling in decision making (EUROFUSE 2013).

[3] Dmytro Mishkin and Jiri Matas. All you need is a good init, 11 2015.

[4] Yingda Xia, Dong Yang, Zhiding Yu, Fengze Liu, Jinzheng Cai, Lequan Yu, Zhuotun Zhu, Daguang Xu, Alan Yuille, and Holger Roth. Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. *Medical Image Analysis*, 65, 10 2020.

[5] Yann Soullard, Sebastien Destercke, and Indira Thouvenin. Co-training with credal models. volume 9896, pages 92–104, 09 2016.