ARGOS TITAN TOSCA meeting - 06/06/2024

IRAKLEION, CRETE

Quantifying Uncertainty with Conformal Prediction

Klea Panayidou

European University Cyprus

Joint work with

Yannis Konidakis & Greg Tsagkatakis

University of Crete & FORTH

TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS

FORTH INSTITUTE OF COMPUTER SCIENCE

Signal Processing Laboratory

CosmoStat

cea

Funded by the European Union

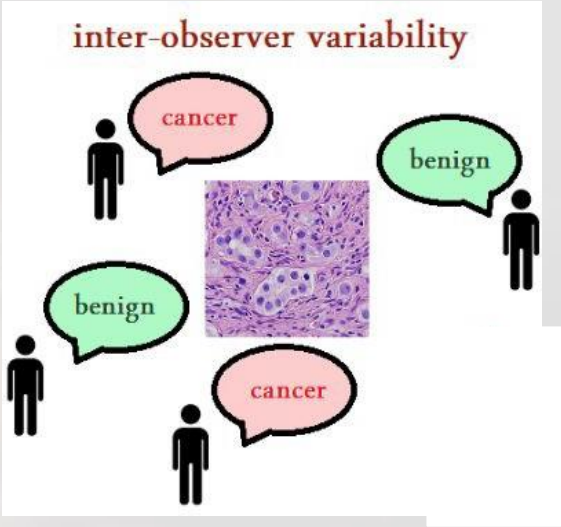
OUTLINE

- Motivation
- Exploratory results
- Conformal Training
- Remarks

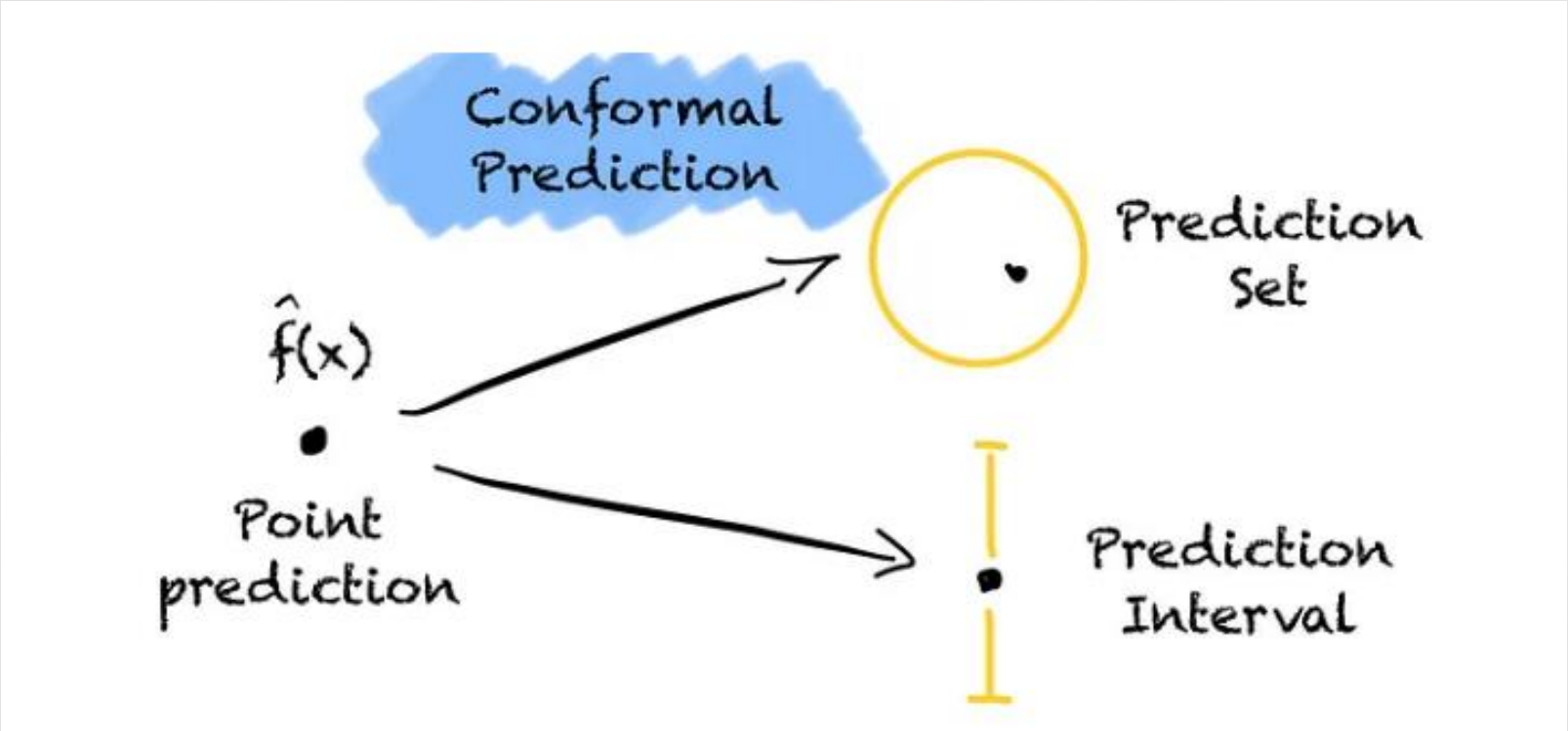
Motivation

Uncertainty

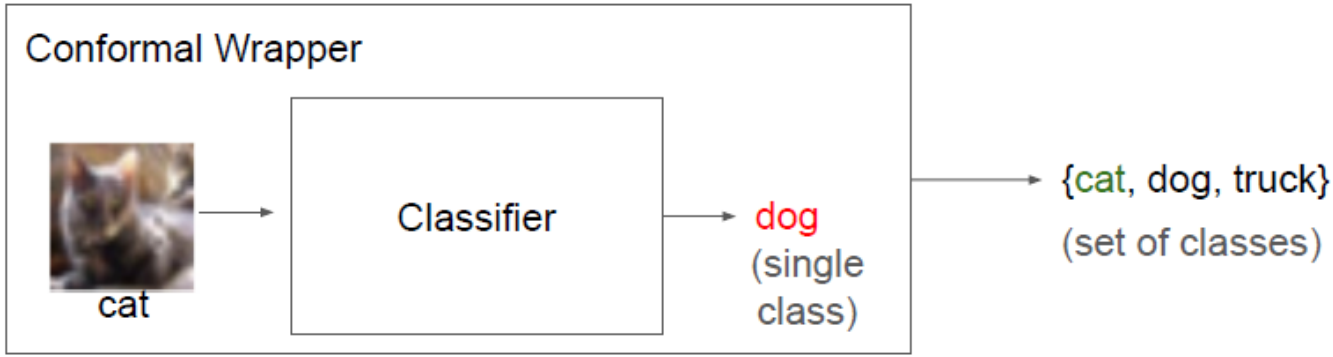
- *True* ground truth unknown / label errors
- Rare classes or long-tailed class distribution
- High-stakes and security-critical applications



Idea


Conformal Prediction
Prediction Set
$\hat{f}(x)$
Point prediction
Prediction Interval

Capturing uncertainty with Conformal Prediction



Number of predicted classes = ambiguity

True class is in the predicted confidence set
with user-specified probability!

Conformal Prediction sets

Given classifier $\pi_{\theta,y} \approx p(y|x)$, construct confidence sets $C_\theta(x) \subseteq [K] = \{1, \ldots, K\}$

- Assumes held-out calibration set, i.i.d. from the test distribution.

$C_\theta(x)$ guarantees *coverage*:

$$p(y \in C_\theta(x)) \geq 1 - \alpha$$

- confidence level α user-specified
- *inefficiency* = confidence set size $|C_\theta(x)|$ optimized

Conformal Prediction sets

Given classifier $\pi_{\theta,y} \approx p(y|x)$, construct confidence sets $C_\theta(x) \subseteq [K] = \{1, \ldots, K\}$

- Assumes held-out calibration set, i.i.d. from the test distribution.

$C_\theta(x)$ guarantees *coverage*:

$$p(y \in C_\theta(x)) \geq 1 - \alpha$$

- confidence level α user-specified
- *inefficiency* = confidence set size $|C_\theta(x)|$ optimized

Metrics

- Empirical and Marginal **Coverage**

$$\text{Cover} := \frac{1}{|I_{test}|} \sum_{i \in I_{test}} \delta\left[y_i \in C(x_i)\right], \quad \delta = 1 \text{ when arg is true, 0 else}$$

- **Inefficiency**

$$\text{Ineff} := \frac{1}{|I_{test}|} \sum_{i \in I_{test}} |C(x_i)|$$

Exploratory Results

Data – Model -Procedure

Data used:

- CIFAR10 (with size 10000)

Model used: resnet-50-finetuned-eurosat (Resnet50 fine tuned on CIFAR10)

- Accuracy: 0.852

Pipeline:

- Load pre-trained model
- Split test set on calibration and validation set
- Get Conformity Score as Softmax of network's output
- Calculate $\hat{q}$ threshold
- Calculate Prediction Sets on validation set
- Compute Inefficiency and Coverage

Investigation

- Different Calibration Set Sizes
- Added Noise
- Different Image Sizes

Different Calibration Sizes

- Different Calibration Sizes:

	n=500	n=1000	n=5000
Inefficiency	2.0662	2.6121	2.5328
Coverage	0.9779	0.9893	0.9872

Mean test confidence sets (1 → class in set, 0 → not in set):

	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10
n=500	0.2221	0.1532	0.2476	0.2631	0.2106	0.2248	0.1909	0.1913	0.1871	0.1756
n=1000	0.2686	0.1928	0.3304	0.3310	0.2909	0.2584	0.2429	0.2579	0.2261	0.2131
n=5000	0.2616	0.1880	0.3254	0.3206	0.2798	0.2498	0.2384	0.2432	0.2154	0.2106

Different Resolution

- Different Image Sizes (n=1000):

	s=32	s=64	s=128	s=224
Inefficiency	9.7883	9.6240	6.1646	2.6121
Coverage	0.9864	0.9920	0.9890	0.9893

Mean test confidence sets (1 → class in set, 0 → not in set):

	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10
s=32	0.9920	0.9998	0.9629	0.9914	0.9612	0.9760	0.9153	0.9897	1.0000	1.0000
s=64	0.9991	0.9838	0.9938	0.9590	0.9459	0.9164	0.9151	0.9422	0.9907	0.9780
s=128	0.8252	0.5917	0.6729	0.6016	0.6247	0.5158	0.4419	0.6031	0.7257	0.5621
s=224	0.2686	0.1928	0.3304	0.3310	0.2909	0.2584	0.2429	0.2579	0.2261	0.2131

Added Noise

"Out of" Distribution

<Description>

- Added Noise $\sim \mathcal{N}(0,1)$ (n=1000):

	+noise	no noise
Inefficiency	6.9836	2.6121
Coverage	**0.6832**	0.9893

Mean test confidence sets (1 → class in set, 0 → not in set):

	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10
+noise	0.9987	1.0000	0.8449	**0.0939**	0.9523	**0.0104**	**0.0833**	1.0000	1.0000	1.0000
no noise	0.2686	0.1928	0.3304	0.3310	0.2909	0.2584	0.2429	0.2579	0.2261	0.2131

Noise levels

- Different Levels of noise on calibration and validation data:

	σ=0.1	σ=0.2	σ=0.3	σ=0.4	σ=0.5
Inefficiency	8.9636	9.8929	9.8771	7.5928	5.1481
Coverage	0.9869	0.9849	0.9857	0.7339	0.5264

Mean test confidence sets (1 → class in set, 0 → not in set):

	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10
$\sigma = 0.1$	0.9967	0.9609	0.9716	0.9162	0.8613	0.7219	0.7921	0.8212	0.9954	0.9262
$\sigma = 0.2$	1.0000	1.0000	1.0000	0.9953	0.9997	0.9309	0.9670	1.0000	1.0000	1.0000
$\sigma = 0.3$	1.0000	1.0000	1.0000	0.9988	1.0000	0.8866	0.9918	1.0000	1.0000	1.0000
$\sigma = 0.4$	0.9998	1.0000	0.9792	0.3229	0.9931	0.0018	0.2960	1.0000	1.0000	1.0000
$\sigma = 0.5$	6.9e-01	1.0e+00	1.49e-01	2.22e-04	3.4e-01	0.0	3.3e-04	9.58e-01	9.9e-01	1.0

Exchangeability ?

- Different Levels of noise only on validation data:

	σ=0.1	σ=0.2	σ=0.3	σ=0.4	σ=0.5
Inefficiency	7.1151	1.1772	0.9846	0.9844	0.9844
Coverage	0.9023	0.1148	0.1011	0.1005	0.0995

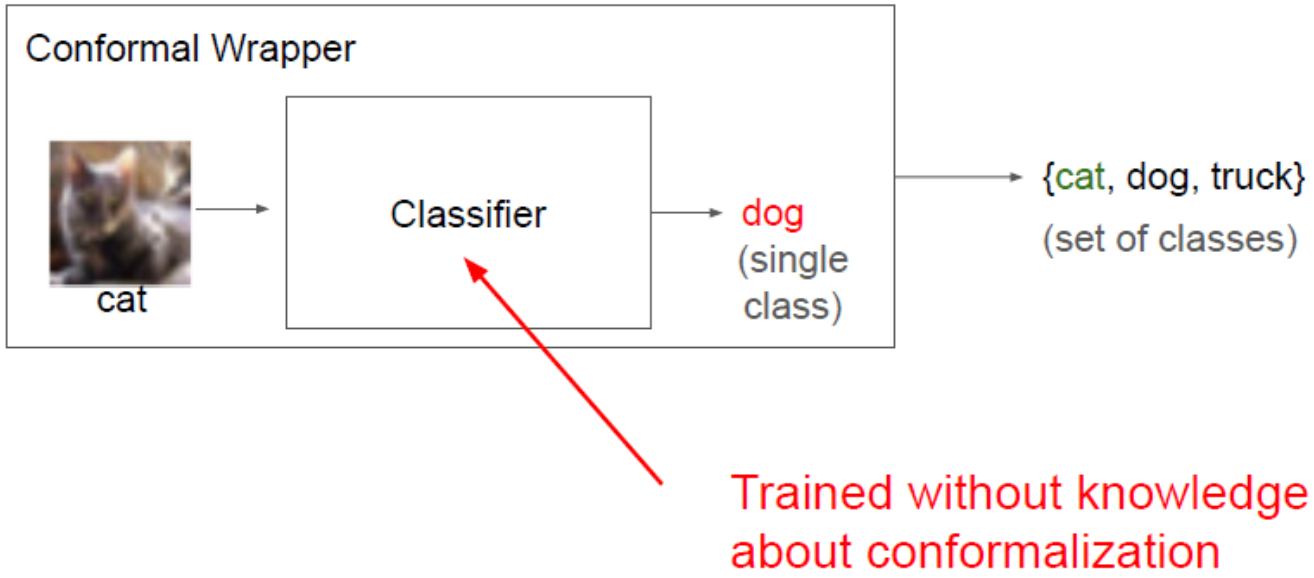
Mean test confidence sets (1 → class in set, 0 → not in set):

	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10
$\sigma = 0.1$	0.9915	0.8856	0.8964	0.7918	0.6036	0.3196	0.4666	0.4881	0.9813	0.8030
$\sigma = 0.2$	3.27e-03	0.187	1.01e-03	0.0	1.1e-04	0.0	0.0	0.0	4.4e-03	0.99e
$\sigma = 0.3$	0.0	1.1e-04	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
$\sigma = 0.4$	0.	0.	0.	0.	0.	0.	0.	0.	0.	1.
$\sigma = 0.5$	0.	0.	0.	0.	0.	0.	0.	0.	0.	1.

Conformal Training

Conformal Training

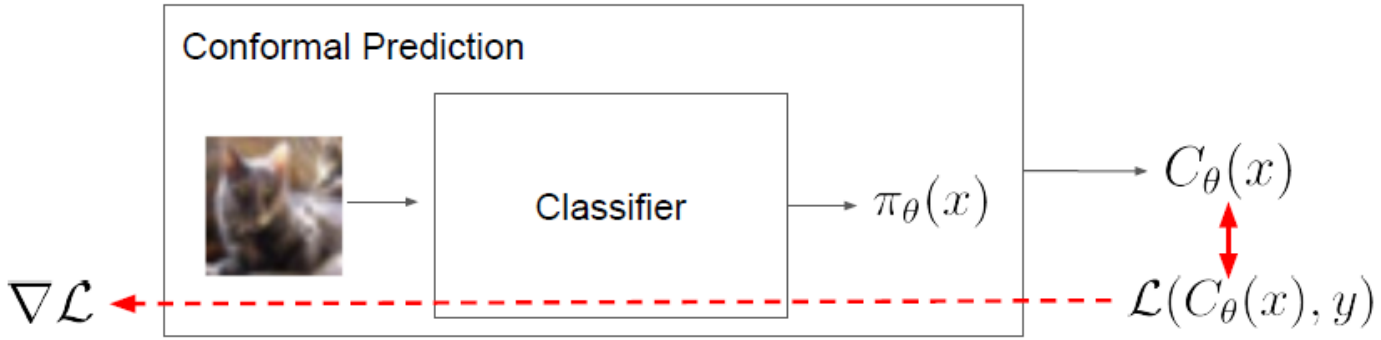
Predict confidence sets with *coverage guarantee*:



Stutz, D.; Dvijotham, K. D.; Cemgil, A. T.; and Doucet, A. 2022. Learning Optimal Conformal Classifers.
In International Conference on Learning Representations

Conformal Training

High-level idea: allow to train classifier “through” the conformal wrapper.



Stutz, D.; Dvijotham, K. D.; Cemgil, A. T.; and Doucet, A. 2022. Learning Optimal Conformal Classifers. In International Conference on Learning Representations

Remarks

- More examples
- Explore when and how CP can be useful – ground truth unknown, hallucinations etc.
- Conformal Training – incorporate in the modeling phase