Uncertainty in Computer Science 21/22

Handling uncertainty in machine learning

DAVIDE CIUCCI

Data sets
in tutorials

Data sets in
the wild

Uncertainty representation and Machine Learning

HOW MOST ALGORITHMS NEED DATA

Patient	Mitral rigurgitation	Acute dyspnea	Bicuspid aortic valve	EKG stress test
P1	No	Heart failure	Yes	Positive
P2	No	COPD	No	Negative
P3	Yes	Pneumonia	No	Positive
P4	Yes	Heart failure	Yes	Negative
P5	yes	Heart failure	No	Negative
P6	yes	COPD	No	Positive
P7	yes	pneumonia	No	Positive

Uncertainty representation and Machine Learning

HOW DATA ARE

Patient	Mitral rigurgitation	Acute dyspnea	Bicuspid aortic valve	EKG stress test
P1	No	Heart failure (highly confident)	Yes (sure)	Positive (highly confident)
P2	Mild	COPD (highly confident)	No (confident)	Negative (highly confident)
P3	Moderate	Pneumonia (highly confident)	No (confident)	Not performed
P4	Severe	Heart failure (highly confident) or pneumonia (low confident)	Yes (highly confident)	Negative (confident)
P5	Moderate or Severe	Heart failure (highly confident) or pneumonia (highly confident)	Undetermined	Undetermined
P6	Mild or Moderate	COPD or pulmonary embolism or pneumonia	Not applicable	Not available
P7	Undetermined	Heart failure and pneumonia	Not applicable	Positive (confident)

Dataset issues

- Lack of knowledge
  - Previous examples
- Lack of trust
  - Example: Circulating Tumor Cells (CTC) dataset, 617 instances, 50x50 RGB images, labeled by 11 raters, identification of *Circulating Tumor Cells* from fluorescence microscopy;
  - Are the raters reliable?
  - How to aggregate raters opinion?
  - Is it necessary to aggregate? https://pdai.info/

Not only the data: allowing uncertainty in the output

Patient	Temperature	Headache	Disease
P1	38	yes	Flu
P2	37	yes	None
P3	37	yes	Pneumonia

- A new patient with temperature = 37 and headache = yes
  - Force a decision? Abstain from decision? Return a set of possibilities?

Three-way decision (3WD)

- A recent paradigm, introduced by Y.Y. Yao (2012)
- Based on representing computational objects in threes (steps/dimensions/groups/parts/…)
- Three *literal* or Three *figurative*: “a few” (<7)
- A cognitive basis: we can process only a small number of units of information
- Examples: true/unsure/false; low/middle/high; yes/maybe/no,… —> orthopair
- 3WD narrow sense —> rough sets
- 3WD wide sense —> TAO model

TAO model

- TAO (trisecting-acting-outcome) model (2018)
  - Trisecting: how can we tri-partition the universe of studied objects?
  - Acting: what computational strategy can we apply to the three parts identified?
  - Outcome: how can we evaluate the adopted strategies?

TAO-based framework

Dealing with uncertainty in Machine Learning

	Trisecting	Acting	Outcome
Input	Split the dataset in certain/ uncertain instances	Take into account the dataset uncertainty and handle it	Define measures to quantify the dataset uncertainty, use them in the algorithm evaluation
Output	The output can contain instances with no decision	The ML algorithm abstains on uncertain instances	New measures to evaluate ML algorithms with abstention

Handling uncertainty in the input

Weak Supervision

The target or decision variable is imprecise or partially specified

Patient	Temperature	Headache	Disease
P1	38	yes	Flu
P2	37	yes	?

- Partially labelled data: uncertain or missing classification in the input
- Solution: extend existing ML algorithm to handle (without deleting or imputing) the uncertainty by means of orthopartitions
- Result: among the tested algorithms, TW-random forest has better result

Weak Supervision

The target or decision variable is imprecise or partially specified

Patient	Temperature	Headache	Disease
P1	38	yes	Flu
P2	37	yes	{None, Pneumonia}

- Superset Learning: instances are assumed to be labeled with a set of possible annotations containing the correct one
- Solution: feature reduction with rough sets + evidence theory to *figuring out the most plausible precise instantiation of the imprecise training data*
- Result: *statistically significant improvement* in predictive accuracy wrt to the state of the art

Multi-rater annotation

Patient	Temperature	Headache	Expert 1	Expert 2	Expert 3
P1	38	yes	Flu	Flu	None
P2	37	yes	None	Pneumonia	Flu

- Multiple experts are involved in the annotation process and may produce mutually inconsistent categorizations
- Solution: (binary decision) convert multi-rater to a unique (uncertain) value by means of different strategies, namely three-way decision, possibility and probability theory
- Results
  - The proposed algorithms are more effective than the traditional majority voting based approach
  - The three-way strategy and “surprising popular algorithm” outperform the other

Handling uncertainty in the output

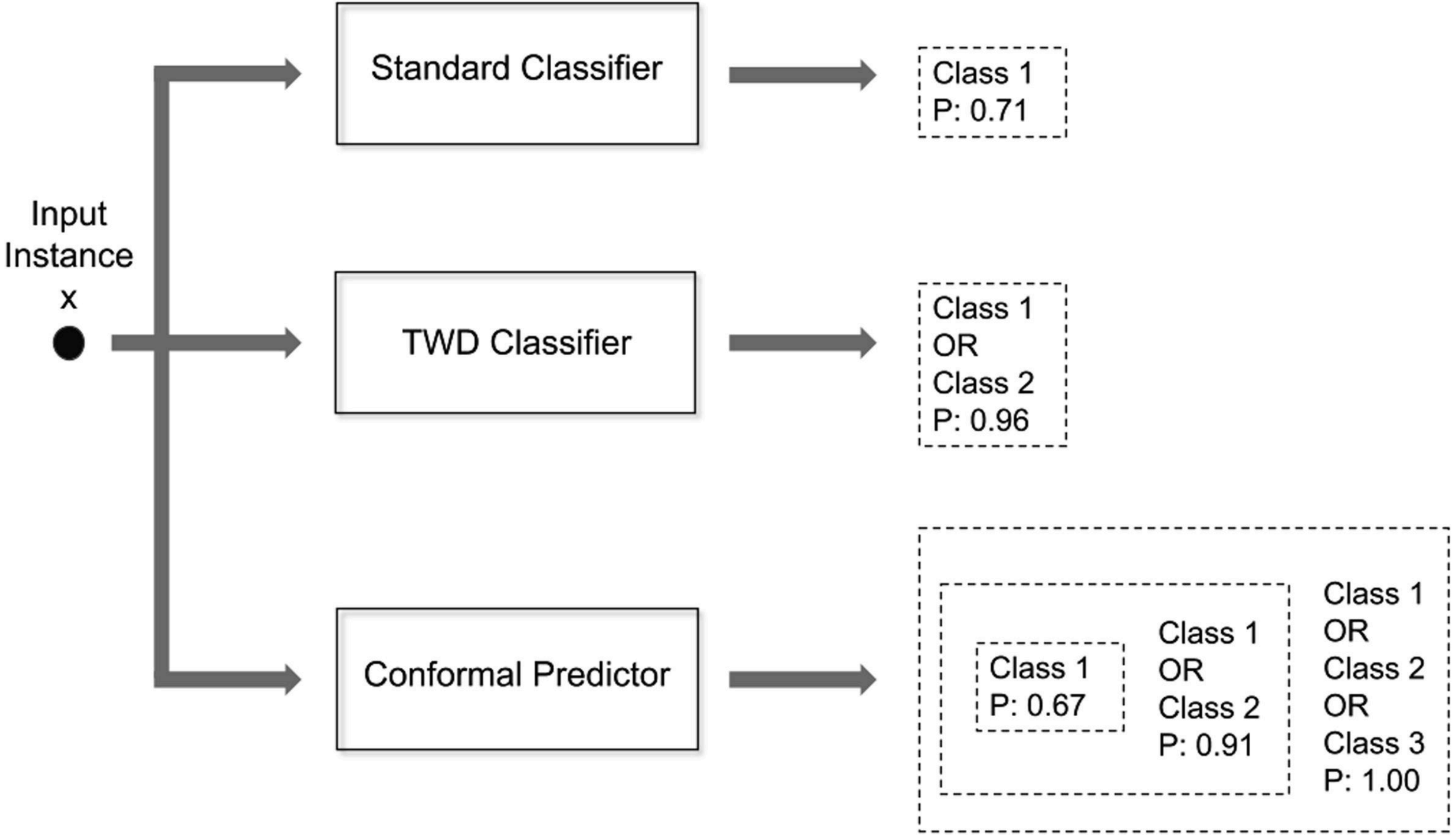
Cautious Learning

- Not enough evidence to take a decision
- a generalization of supervised learning in which the Machine Learning (ML) models are allowed to express set-valued predictions
- trade-off between different quality dimensions
  - Cost-sentitiveness: the model properly takes into account information about the utilities and costs of the different alternative decisions
  - Validity: the performance of the model can be reliably bounded, usually through a theoretical analysis
  - Efficiency: the set-valued predictions provided by the model are as informative as possible

Cautious Learning

- Solution: define algorithms that can abstain (three-way strategy)
  - a general method based on *cost of abstention* vs *cost of error*
  - ad hoc methods: TW-decision tree, TW-random forest based on orthopartition
- Result: three-way algorithms offer a trade-off among accuracy and coverage (the points that are classified)

- Compared with conformal prediction
  - Bound the error of three-way classifiers
  - Under which conditions the two approaches are equivalent


Input
Instance
x
Standard Classifier
Class 1
P: 0.71
TWD Classifier
Class 1
OR
Class 2
P: 0.96
Conformal Predictor
Class 1
P: 0.67
Class 1
OR
Class 2
P: 0.91
Class 1
OR
Class 2
OR
Class 3
P: 1.00

Clustering

- Soft Clustering: There are several known *soft clustering* algorithms
- Problem: Lack of evaluation methods
- Solution 1: we defined new external measures for rough/three-way clustering
- Result: we can now compare different three-way clustering algorithms
  - Compared several rough/three-way algorithms

Soft Clustering

- Solution 2: a general framework for external soft evaluation
- Two dimensions
  - ambiguity: the inability to uniquely assign an object to a single cluster (typical of rough clustering)
  - partial assignment: the assignment of objects to multiple clusters, each with a given degree of membership (typical of fuzzy clustering)
- Interpretation of soft clustering as representing a distribution over hard clusterings
- Computational complexity problems —> approximate solutions

Conclusion

- Advantages for Machine Learning
  - Handling (not hiding) of uncertainty with no pre-processing
  - Allows for robust and human-in-the-loop ML
    - The final decision should be from a human
  - Can be applied also to standard algorithms with little modification
- Future Directions
  - Interactive Uncertainty Management based on active learning and human-in-the-loop 3way ML
  - Deal with different forms of uncertainty in an integrated way