Arguing from Experience To Classifying Noisy Data

Frans Coenen

Maya Wardeh **Trevor Bench-Capon**

University Of Liverpool
Department of Computer Science

Presentation Plan

Introduction : Arguing from Experience

Classifying Through Argumentation

Example 1

Example 2

More Experiments

Discussion

Arguing from Experience

Experience May Differ

- Geographically
- Exception may be only rarely encountered
- Samples may be abnormal

Advantages

- No previous analysis
  - No knowledge engineering bottleneck
  - Existing Databases can be used
  - Can deal with gaps and conflicts

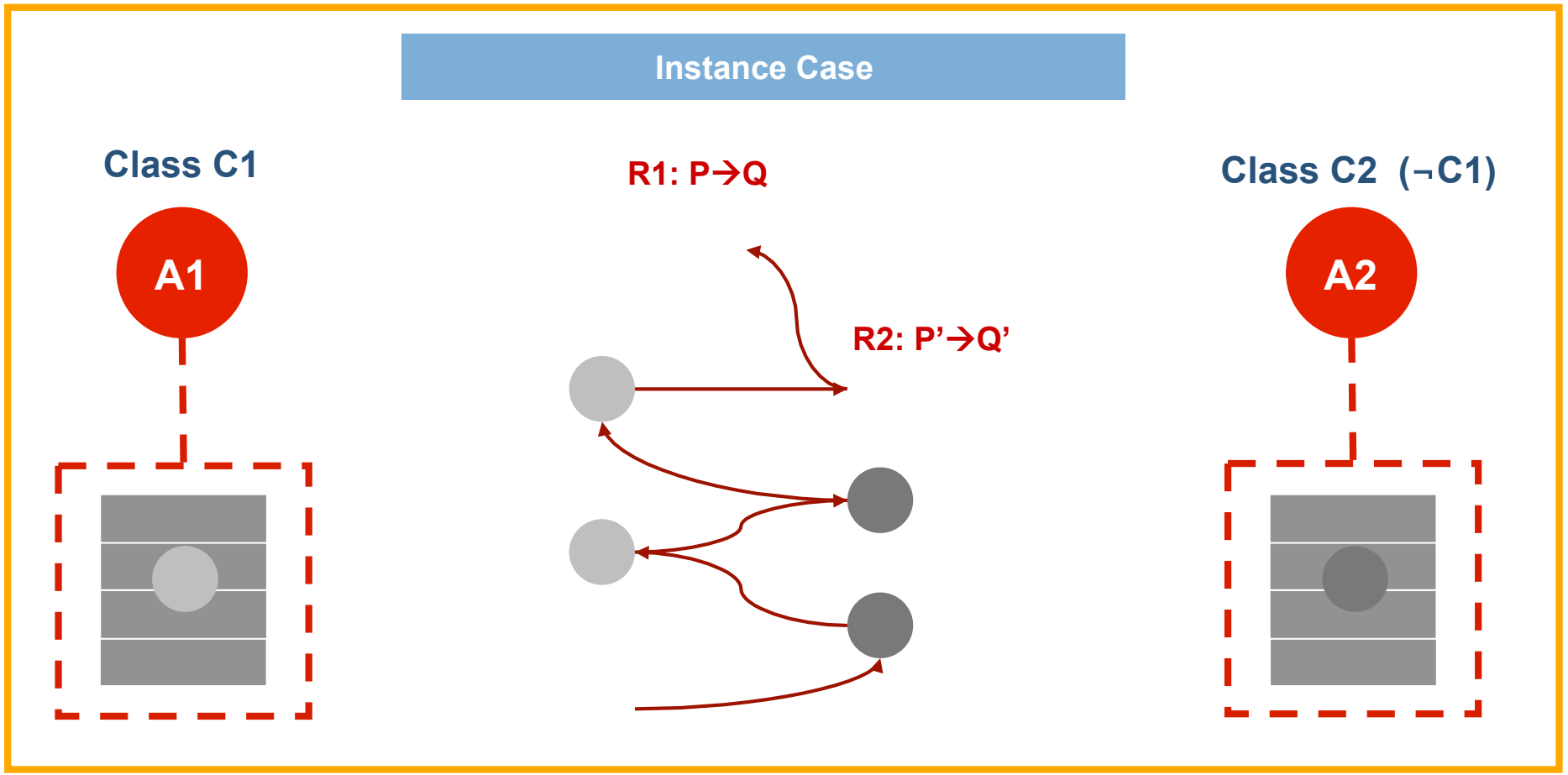
→ Previous Experience

- Common law
- Every day argument

What is PADUA?

Protocol for Argumentation Dialogue Using Association rule mining

A Dialogue Game to Argue about CLASSIFICATION



What is PADUA?

PADUA Speech acts

1: Propose Rule

PADUA Protocol

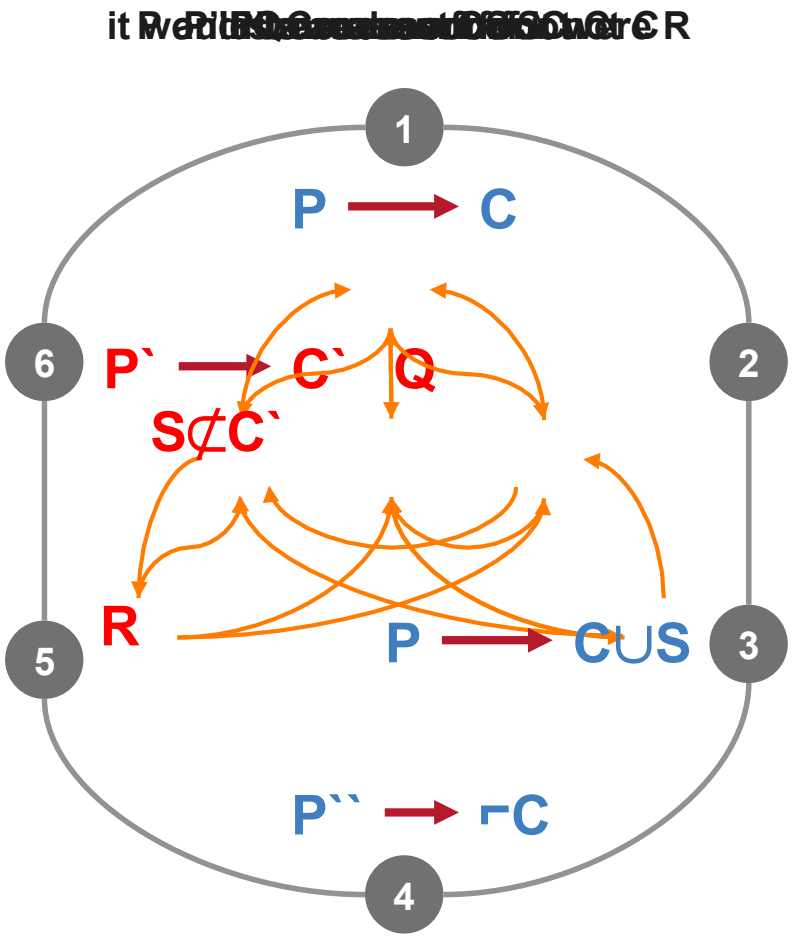
2: Distinguish

3: unwanted consequences

4: Counter Rule

5: Increase Confidence

6: Withdraw unwanted consequences



What is PISA?

The MEETING ROOM

Turn Taking Policy

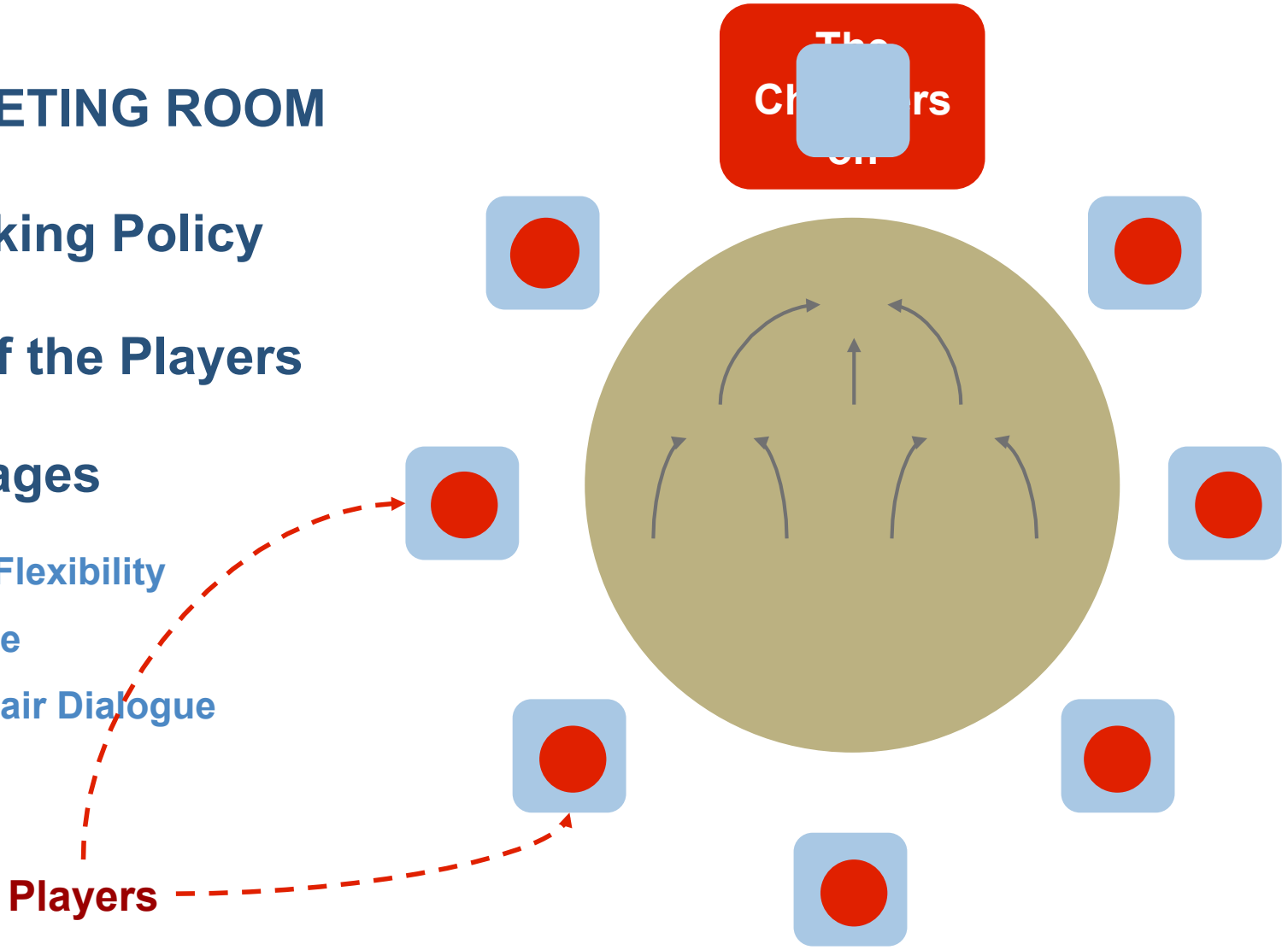
Roles of the Players

Advantages

Increases Flexibility

Very Simple

Provides Fair Dialogue



Presentation Plan

Introduction : Arguing from Experience

Classifying Through Argumentation

Example 1

Example 2

More Experiments

Discussion

Classifying through Argumentation

Arguing From Experience for Classification

Provides means to debate the possible classifications of a given case between a number of participants, each defending one possibility.

Each agent attempts to persuade the others that their point of view is the correct one

Each agent formulates arguments for or against a classification from a background dataset of past examples which is mined as required.

The process of classification – 3 ARM queries

Find a subset of rules that conform to a given set of constraints

Distinguish a given rule by adding additional attributes

Generalise a given rule by removing attributes

Presentation Plan

Introduction : Arguing from Experience

Classifying Through Argumentation

Example 1

Example 2

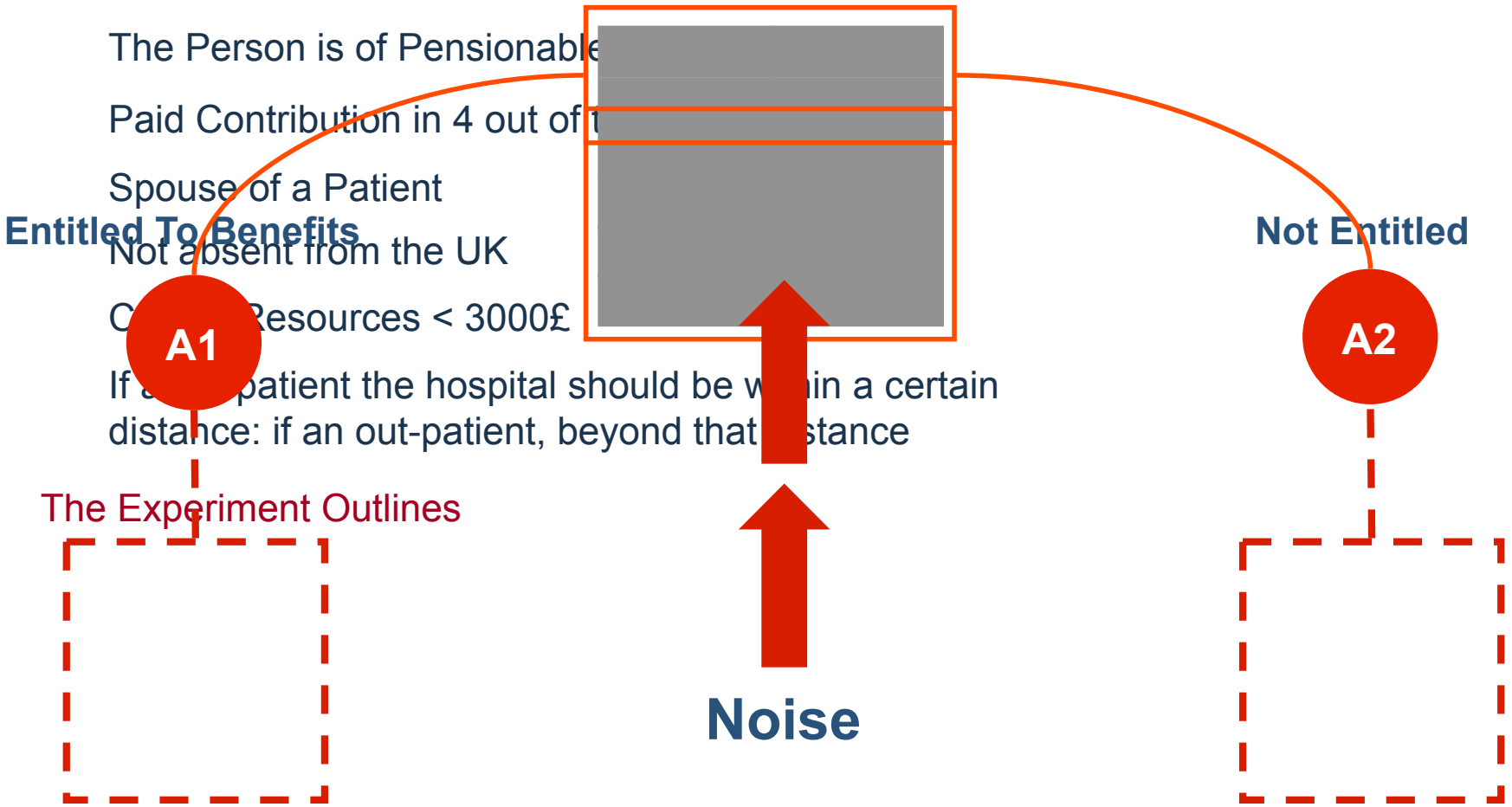
More Experiments

Discussion

Example 1 – Welfare Benefits

The Welfare Benefits

The benefit is payable if 6 conditions, covering wide range of possibilities.



Example 1 – Welfare Benefits

Noise	**PADUA**	**RDT**	IGDT	TFPC	CBA	CMAR	FOIL	CPAR	PRM	CN2	ABCN2
0	**99.9**	**100**	92.5	98.5	99.2	96.8	99.7	67.1	66.7	99.5	99.8
2	**99.9**	**98.6**	88.2	98.3	100	98.8	100	65.4	65.4	97.8	98.4
5	**99.3**	**99.6**	93.3	99.86	98.8	98.1	94.2	65.4	65.4	96.4	96.9
10	**98.5**	**98.3**	92.8	97.08	91.9	97.2	93.2	64.4	64.4	93.5	94.7
20	**97.8**	**97.3**	90.9	98.8	86.9	97.3	88.9	61.7	63.6	88.7	92.0
40	**97.1**	**96.4**	90.4	96.3	94.0	96.8	89.4	58.1	57.9	83.3	85.0

Presentation Plan

Introduction : Arguing from Experience

Classifying Through Argumentation

Example 1

Example 2

More Experiments

Discussion

Example 2 – Housing Benefits

The Housing Benefits

Housing benefit is payable if 6 conditions, covering wide range of possibilities.

The Person is of Pensionable Age

Paid Contribution in 3 out of the last 5 years

Same data generation process

Similar Two Players Games

Not Absent From The UK or Armed Forces

Income < 20%

Capital Resources < 3000£

The Experiment Outlines

Example 1 – Housing Benefits

Noise	**PADUA**	**RDT**	IGDT	TFPC	CBA	CMAR	FOIL	CPAR	PRM
0%	**99.86**	**99.72**	77.00	98.33	97.36	99.31	100.00	64.03	66.81
2%	**99.72**	**97.78**	76.25	98.61	99.86	98.01	96.67	63.75	64.72
5%	**99.58**	**98.89**	64.31	96.53	97.50	98.61	94.44	65.28	65.14
10%	**98.61**	**98.75**	73.61	93.61	91.11	95.69	87.08	63.61	64.92
20%	**96.81**	**98.19**	73.06	93.89	96.25	96.50	86.39	62.28	64.58
40%	**96.11**	**92.22**	64.44	83.06	92.08	92.92	86.11	60.97	61.25
50%	**94.03**	**88.75**	62.22	54.72	84.17	85.31	78.19	59.58	61.81

Housing Benefits – 4 Classes

Housing Benefits for PISA

Housing Benefit is either Payable, Payable with Priority, Partial Benefit or Not Payable.

One dataset is generated and noise is added in the same manner.

The training set is split into four equal datasets, each is given to one participant.

PISA is applied to classify the cases in the test set.

Housing Benefits – 4 Classes

Noise	**PISA**	**RDT**	IGDT	TFPC	CBA	CMAR	FOIL	CPAR	PRM
0%	**98.47**	**94.44**	68.19	92.56	90.28	86.75	92.25	75.83	75.83
2%	**97.64**	**90.56**	67.75	91.81	90.14	86.25	92.22	75.42	68.06
5%	**97.36**	**93.47**	62.92	89.72	90.69	85.00	91.39	73.33	73.89
10%	**96.53**	**92.92**	60.97	86.81	89.17	84.25	92.36	70.83	72.64
20%	**95.69**	**91.94**	60.56	80.83	88.89	83.75	89.31	70.78	70.61
40%	**94.44**	**90.31**	56.35	69.86	86.81	81.75	80.56	63.06	63.06
50%	**93.75**	**88.36**	61.81	45.83	62.71	80.50	70.42	63.06	65.83

Presentation Plan

Introduction : Arguing from Experience

Classifying Through Argumentation

Example 1

Example 2

More Experiments

Discussion

Real – World Datasets

Applying PISA and PADUA with Real Datasets

PISA and PADUA were applied using 7 real world datasets.

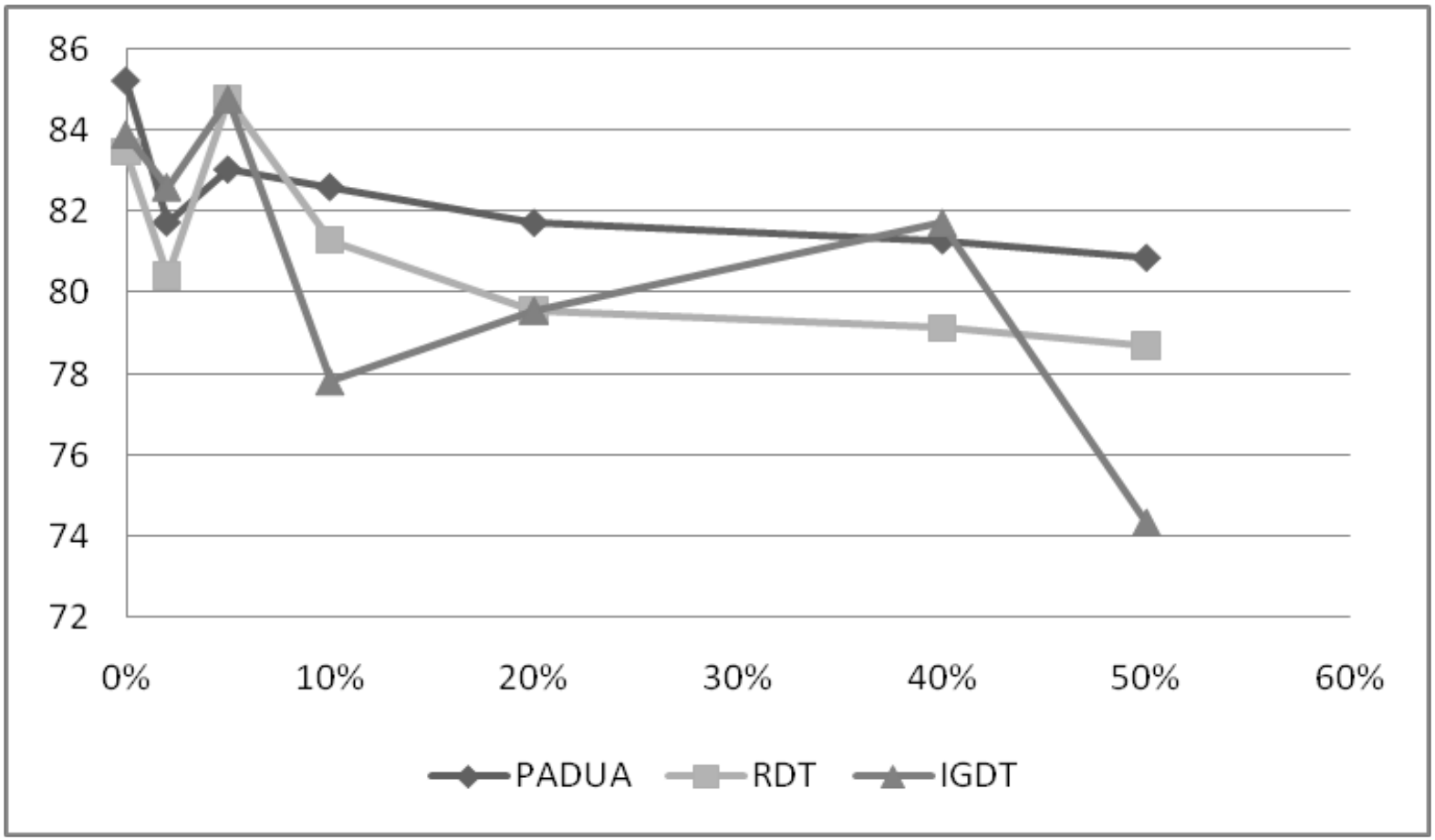
The tests compared PISA and PADUA to the same classifiers as before.

The tests used the same noise model as before.

The obtained results show a similar pattern to the benefits experiments: the accuracy of almost all the classes dropped when the noise percentage was increased.

Real – World Datasets

Example Wisconsin Breast Cancer Dataset



Presentation Plan

Introduction : Arguing from Experience

Classifying Through Argumentation

Example 1

Example 2

More Experiment

Discussion

Discussion and Conclusions

Arguing From Experience

Provides a means for an **unlimited number** of agents to engage in discussion about a classification on the basis of raw data, unmediated by knowledge representation effort to present this data in the form of rules.

Arguing from Experience Applications

In domains in which there are large volumes of data available.

several distributed datasets generated from different samples

With high level of noise in the training dataset that are expensive to clean.

Thanks For Listening