Arguing From Experience to Classifying Noisy Data

Maya Wardeh, Frans Coenen and Trevor Bench-Capon

Department of Computer Science
University of Liverpool, L60 3BX, UK
maya@csc.liv.ac.uk, frans@csc.liv.ac.uk, tbc@csc.iv.ac.uk

Abstract. A process, based on argumentation theory, is described for classifying very noisy data. More specifically a process founded on a concept called "arguing from experience" is described where by several software agents "argue" about the classification of a new example given individual "case bases" containing previously classified examples. Two "arguing from experience" protocols are described: PADUA which has been applied to binary classification problems and PISA which has been applied to multi-class problems. Evaluation of both PADUA and PISA indicates that they operate with equal effectiveness to other classification systems in the absence of noise. However, the systems out-perform comparable systems given very noisy data. **Keywords:** Classification, Argumentation, Noisy data

1. Introduction

Argumentation is concerned with the logical reasoning processes required to arrive at a conclusion given two or more alternative view points. The process of argumentation can be conceptualised as a discussion about some issue that requires a solution, between a group of individuals with different points of view; where each member of the group attempts to persuade the others that his/her point of view, and the consequent solution, is the correct one. The discussion is conducted using a set of logical reasoning rules linking antecedents to consequents. Computer automation and modelling of the argumentation process has applications in legal reasoning, online auctions and so on. There is much reported work on automated argumentation, especially in the "two player" setting.

In automated argumentation (persuasion dialogue games) each "player" typically has access to their own Knowledge Base (KB) which is used to propose arguments founded on the rules and facts contained in the KB [20]. Arguments can be advanced to either promote a player's own desired outcome or to attack arguments advanced by other players. However, the use of a KB to support argumentation has several disadvantages. Firstly the construction of the KB requires domain experts and entails the well established knowledge acquisition bottle neck reported in the Knowledge Based System and Expert System literature. Secondly the KB is never up to date.

An alternative to the KB approach to argumentation, and that promoted in this paper, is for each player to use data mining techniques to "mine" the desired rules from a live database. The authors refer to this process as "arguing from experience" in the sense that each player's database can be considered to encapsulate that player's "experience". The arguing from experience idea was first explored by the authors in [26], where the PADUA two player argumentation system was introduced; and further developed in [27] where the PISA multi-player argumentation system was proposed. Both systems provided a mechanism for two (PADUA) or more (PISA) software agents to conduct dialogues to resolve disputes concerning the correct categorisation of particular examples. Both operate using an Association Rule Mining (ARM) technique to extract rules from their database repository of experience. The evidence presented in [26] and [27] indicated that the arguing from experience approach provides a natural representation of the participant's experience as a set of records, and the arguments as Association Rules (ARs).

In this paper the authors explore the application of the "arguing from experience" paradigm, advocated by both PADUA and PISA, to resolve classification (categorisation) problems, especially with respect to noisy data. The ability to handle noisy data is seen as important because it must be recognised that classification data will often contain wrongly classified examples, representing misconceptions and mistakes. In certain domains, such as welfare benefits, it is estimated that 30% or more of previous examples may have been wrongly classified [19]. Any classifier relying on such data must therefore be robust in the face of quite high levels of noise. Conceptually example cases are presented for classification, to either PADUA or PISA, and in each case the (PADUA or PISA) agents will argue for a particular classification through a persuasion process.

The investigation, reported here, establishes that arguing from experience in this manner provides a classification mechanism that can produce similar accuracies to those produced by other classification systems in the absence of noise, but can cope more readily given noisy input data (noise levels of up to 50%).

The rest of the paper is organised as follows. Section 2 provides some background information about the problem of classifying noisy data. Then in section 3 we give a summary of the argumentation from experience process and an overview of both PADUA and PISA. In section 4 an evaluation of PADUA, in the context of the binary classification of noisy data is presented. This is followed up in Section 5 by an evaluation of PADUA in a multi-class classification setting. Some final conclusions are presented in Section 6.

2. Background

The data classification (categorization) problem is well established in the Knowledge Discovery in Data (KDD) and data mining community. A substantial number of mechanisms have been developed to generate classifiers, including Neural Networks and Support Vector Machine, decision tree algorithms, rule induction approaches, various mechanisms influenced by ideas take from genetic programming and biocomputation, and Classification Association Rule Mining (CARM). Both PISA and PADUA operate using CARM [18]. The basic idea of CARM is to generate a set of Classification Association Rules (AR) (a subset of the complete set of ARs) using ARM technology [1]. CARM offers a number of advantages including computational efficiency and, unlike many other classifier generators, easy understandability of the resulting classifier.

One of the challenges of the classification problem is how to deal with very noisy data (and data with many missing values). Of course in an ideal domain the training data will contain no noise, no errors and no missing attributes; but unfortunately, in most real world domains, this is not the case. Tolerating noise is particularly important when designing classifiers, as the accuracy of classification depends on the quality of the input dataset. Noise can also be artificially introduced to the datasets for different purposes such as preserving privacy [2,5].

Coping with noise can be addressed in different ways. One approach is to develop robust systems that allow for noise by avoiding over-fitting the model to the data ([2, 7]). Another approach is to pre-processing the input data before learning [6, 25] so as to eliminate tainted records [6, 25], but entails some major drawbacks:

- Eliminating whole records of "bad data" eradicates "potentially" useful information such as the associations between uncorrupted attributes.
- When there is a large amount of noise in the dataset, the amount of information in the remaining clean dataset may not be sufficient for building the classifier.
- In some cases eliminating "bad data" records is not possible because identifying these records can be an exhausting task, and may even require consulting expert opinion. This can be the case in datasets representing legal scenarios where the legislation can be misinterpreted.

A number of preprocessing techniques have been developed to correct corrupted (noisy) data such as:

- Deleting the corrupted fields and using the remaining, non-corrupted, fields for subsequent modeling and analysis [16].
- Cleaning of the dataset to remove noise (for example using Bayesian methods to clean corrupted data that have dependencies among features as described in [24])
- Correcting the misclassified data to improve classification accuracy based on the other predicted feature values as well as the corrected feature values [14].

– However, such preprocessing is not always feasible as it often requires expert consultation (for example to provide the model for the Bayesian network in [24]), or because the noise level is so high that the correction of the corrupted data is neither easy nor effective.

– The following sections provide an overview of how the proposed arguing from experience framework can cope with noisy data, without any need for (i) pre-processing or (ii) initial removal of corrupted data by providing a moderation mechanism; whereby several agents engage in an argumentation dialogue, each using their own database of cases (representing their former experience). The idea is that this will allow the agents to correct each others "misconceptions".

3. Classification through Argumentation Using PISA and PADUA

The objective of both PADUA and PISA is to allow a number of agents, each with their own "private" database of examples, to debate the correct classification of a new case. The classification can be binary (PADUA) or non-binary (PISA). In PADUA (Persuasive Argumentation Using Association Rules), a protocol to enable two agents to argue about the classification of a case was established. PISA (Pooling Information from Several Agents) extended the PADUA protocol to allow any number of software agents to engage in a dialogue. This was found to be particularly useful for multi-class classification (i.e. non-binary classification), since each possible classification can then have its own champion. As noted above the distinguishing feature of both PADUA and PISA was that the arguments used by the agents were derived directly from a database of previous examples using ARM [26]. In PADUA the background dataset of each agent was represented by the means of a T-tree (Total tree) data structure, a reverse set enumeration tree structure with fast look up properties [9].

Both PADUA and PISA operate using a basic set of speech acts for argument from experience dialogues between two or n parties respectively. These speech acts are supported by three different forms of dynamic ARM request:

1. Find a subset of the possible set of ARs that conform to a given set of constraints.
2. Distinguishing a given AR by adding additional attributes.
3. Generalising a given AR by removing attributes.

Using their distinct databases PADUA and PISA agents produce reasons for and against classifications.

ARs [1] are probabilistic relationships which can be viewed as rules of the form X → Y (read as if X is true then Y is likely to be true, or X is a reason to think Y is true) where X and Y are disjoint subsets of some global set of attributes. Likelihood is usually represented in terms of a *confidence* value expressed as a percentage. This is calculated as support(XY)×100/support(X) where the *support* of an itemset is the number of records in the data set in which the itemset occurs. To limit the number of ARs generated only itemsets whose support is above a user specified support threshold, referred to as frequent itemsets, are used to generate associations. To further limit the number of ARs only those rules whose confidence exceeds a user specified confidence threshold are accepted. In the context of this paper the antecedent of an AR represents a set of reasons for believing the example should be classified as expressed in the consequent. Neither PADUA nor PISA use a specialized CARM algorithm, instead they are found on the Apriori T ARM algorithm described in [9] and then classify the test cases by the means of the dialogue.

There are six speech acts (moves) used in PADUA [26] and PISA [27] dialogues which form three categories of "move" as follows:

1. *Propose Rule*: Move that allows generalizations of experience to be cited, by which a rule (AR) with a confidence higher than a certain threshold is proposed.

2. *Attacking Moves*: These moves argue that the reasons given in a rule proposed by another agent are not decisive in this case. This can be achieved using one of the following three speech acts:
   - *Distinguish*: Add one or more premises (antecedent items) to a previously proposed rule, so that the confidence of the new rule is decreased.
   - *Counter Rule*: Similar to the "propose rule" move, but used to cite a generalization leading to a different classification.
   - *Unwanted Consequence*: Move to suggest that a certain consequent (conclusion) of the proposed rule does not match the case under consideration.
3. *Refining Moves*: Moves that enable a rule to be refined to meet objections. This can be achieved using either of the following two speech acts:
   - *Increase Confidence*: Replace one or more premises (antecedent items) in a previously proposed AR so as to increase the confidence of the rule.
   - *Withdraw unwanted consequences*: Exclude unwanted consequences of a rule that has been previously proposed (while maintaining a certain level of confidence). In other words, by trying to withdraw unwanted consequences, the player aims to refine a rule it previously proposed (instead of proposing a new rule).

For each of the above six moves a set of legal next moves (i.e. moves that can possibly follow each move) is defined. Table 1 summarizes the rules for "next moves", and indicates where a new set of reasons is introduced to the discussion.

Table 1. Speech acts (moves) in PADUA-PISA

Move	Label	Next Move	New AR
1	Propose Rule	3, 2, 4	Yes
2	Distinguish	3, 5, 1	No
3	Unwanted Cons	6, 1	No
4	Counter Rule	3, 2, 1	Yes
5	Increase Conf	3, 2, 4	Yes
6	Withdraw Unwanted Cons	3, 2, 4	Yes

4. Evaluation Using Welfare Benefits Dataset

In this section we assess the effectiveness and robustness of PADUA as a classifier with respect to noise using a Welfare Benefits dataset. The model used to introduce noise was the same as that reported in [19]; for an N% noise level in a dataset of I instance, (N*D) instances were randomly selected and the class label changed to some other randomly selected value (with equal probability) from the set of available classes. The noise levels used in this study are: 2%, 5%, 10%, 20% and 40. The noise was introduced to training sets only and not to the test sets.

The rest of this section is organised as follows. The Welfare benefits dataset, used for the evaluation, is discussed in Sub-section 4.1. The various classifiers with which PADUA was compared are presented in Sub-Section 4.2. The ensuing results are discussed in Sub-Section 4.3.

4.1 The Welfare Benefits Dataset

The Welfare Benefits dataset was originally developed by Bench-Capon [3] and has been used in several experiments [19, 4, 15]. The data in this dataset concerns a fictional welfare benefit paid to pensioners to defray expenses for visiting a spouse in hospital. The benefit is payable if six conditions are satisfied:

1. The person is of pensionable age (60 for a woman, 65 for a man):
2. The person has paid contributions in four out of the last five relevant contribution years;
3. The person is a spouse of the patient;
4. The person is not absent from the UK;
5. The person does not have capital resources amounting to more than 3000;
6. If the patient is an in-patient the hospital should be within a certain distance: if an out-patient, beyond that distance.

Conditions 3 and 4 are Boolean necessary conditions, one which must be true and one which must be false. Condition 5 is a threshold on a continuous variable representing a necessary condition. Condition 2 relates five Boolean variables, only four of which need be true. Conditions 1 and 6 relate the relevance of one variable to the value of another: in 1 gender is relevant only for ages between 60 and 65, and in 6 the effect of the distance variable depends on the Boolean saying whether the patient is an in-patient or an outpatient. The wide range of conditions covered by this dataset, is one of the reasons why the dataset was selected to evaluate PADUA, as it demonstrates how well PADUA can cope with noise and how well it can cope with correlated conditions (as well as the other types of conditions used in this dataset).

The dataset comprises of 2400 records such that half are classified as “entitled” (to benefit) and the other half to “not entitled”. 70% of these rows were used as the training set and the rest (30%) as the test set. Noise was then applied to the training set (as defined above). The training set used for each of the noise levels, was split into two equal subsets, one given to the proponent and the other to the opponent in PADUA. The two players argued to classify the 720 cases in the testing set.

4.2 Comparator Classifiers

The operation of PADUA was compared against a variety of standard classifiers, covering a range of classification paradigms, as follows:

Decision Trees: Classification using *decision trees* was one of the earliest reported classification approaches. Quinlan’s C4.5 is arguably the most referenced decision tree algorithm [23]. One of the most significant issues in decision tree generation is deciding on the *splitting criteria*. Of the approaches have been proposed in the literature, two have been used in the evaluation described here:

- Select most frequently occurring item; the Random Decision Tree (RDT) algorithm.
- Select according to highest information gain; the Information Gain Decision Tree (IGDT) algorithm.

Information gain [21] is one of the standard measures used in decision tree construction.

TFPC (Total From Partial Classification) ([10], [11]), is a Classification Association Rule Mining (CARM) algorithm founded on the TFP (Total From Partial) ARM algorithm ([12], [13]); which, in turn, is an extension of the Apriori-T (Apriori Total) ARM algorithm. TFPC is designed to produce Classification Association Rules (CARs) whereas Apriori-T and TFP are designed to generate Association Rules (ARs). In its simplest form TFPC determines a classifier according to the support and confidence framework.

CBA (Classification Based on Associations) is another CARM algorithm developed by Liu et al [17]. CBA operates using a two stage approach to generating a classifier: (i) generate a complete set of CARs, (ii) prune the set of CARs, using the cover principle, to produce a classifier.

CMAR (Classification based on Multiple Association Rules) is a further CARM algorithm developed by Li et al [18]. CMAR also operates using a two stage approach to generating a classifier: (i) generate the complete set of CARs according to a user supplied support threshold to determine frequent (large) item sets, and a confidence threshold to confirm CRs, (ii) prune this set to produce a classifier.

FOIL – CPAR – PRM: FOIL (First Order Inductive Learner) is an inductive learning algorithm for generating Classification Association Rules (CARs) developed by Quinlan and Cameron-Jones [22]. This algorithm was later further developed by Yin and Han to produce the PRM (Predictive Rule Mining) CAR generation algorithm PRM was then further developed, by Yin and Han, to produce CPAR (Classification based on Predictive Association Rules) [28].

CN2 The CN2 algorithm [7,8] consists of a “covering” algorithm and a search procedure that finds individual rules by performing a beam search. The covering algorithm induces a list of rules that cover all the examples in the learning set. Roughly, the covering algorithm starts by finding a rule, and then it removes from the set of learning examples those examples that are covered by this rule, and adds the rule to the set of rules. This process is repeated until all the examples are removed. There are two versions of CN2: one induces ordered list of rules, and the other unordered list of rules.

ABCN2 (Argument Based CN2) [19] is an extension of CN2 (see above). ABCN2 augmented the original CN2 algorithm to take into account arguments that explain misclassified examples: another pass uses these arguments to constrain the rules generated.

4.3. The results

For the experiments the support threshold value was fixed to 1% and the confidence threshold value to 70% for all the relevant classifiers (including PADUA). Table2 shows the affect of adding noise to the Welfare dataset on the accuracy of each classifier. As expected the accuracy of all the classifiers drops as the noise level increases. When using clean data (no noise) RDT out performs all the other classifiers, with PADUA producing acceptable results. However, as the noise level increases it can be observed that PADUA is more tolerant to noise: the PADUA accuracy drops only 2.78% even when the noise level is increased to 40%, while the accuracy of RDT drops 3.61%. The other classifiers suffer more severe drops in their accuracy levels, for example the FOIL accuracy drops 10.28% between the noise levels. The results therefore indicate that PADUA is more tolerant to noise than all the other classifiers. The results for CN2 and ABCN are taken from [19], while the others were produced as part of the experiment.

Table 2. Accuracy versus Noise (PADUA – Welfare Dataset). The CN2 and ABCN2 results are those given in [19].

Noise	PADUA	Rand DT	Info Gain DT	TFPC	CBA	CMAR	FOIL	CPAR	PRM	CN2	ABCN2
0	99.86	100	92.50	98.47	99.17	96.81	99.72	67.08	66.67	99.47	99.76
2	99.86	98.6	88.19	98.33	100	98.75	100	65.36	65.36	97.78	98.42
5	99.31	99.6	93.33	99.86	98.75	98.1	94.17	65.36	65.36	96.36	96.96
10	98.47	98.3	92.78	97.08	91.94	97.19	93.19	64.44	64.44	93.51	94.69
20	97.78	97.3	90.97	98.75	86.94	97.33	88.89	61.67	63.61	88.69	92.00
40	97.08	96.4	90.44	96.25	94.03	96.80	89.44	58.06	57.92	83.26	85.03

5. Experimenting with Housing Benefit dataset

In the above section PADUA, a two player argumentation protocol, was evaluated in the context of binary valued classification using an artificial welfare benefits dataset. In this section multi-class classification problems are considered using a second artificial housing benefits data set where benefits are again payable if certain conditions are satisfied. This dataset, although also originally a two class set, was selected because it is easy to modify from two classes to 3,4, or five classes so as to evaluate the operation of PISA. For completeness PADUA was also applied to the dataset.

The scenario that the housing benefits dataset is intended to reflect is a fictional benefit Retired Persons Housing Allowance (RPHA), which is payable to a person who is of an age appropriate to retirement, whose housing costs exceed one fifth of their available income, and whose capital is inadequate to meet their housing costs. Such persons should also be resident in the UK, or absent only by virtue of "service to the nation", and should have an established connection with the UK labour force. These conditions need to be interpreted and applied [26]. For this data set we used an interpretation very similar to the previous example, the only difference was that here we employed more flexible contribution and residency conditions. We also removed the patient-distance correlated condition. This simplified the dataset, and made modification, for the purpose of PISA, an easier task.

5.1. Evaluation using PADUA

In this sub-section PADUA is further evaluated by applying it to the above housing benefits set configured in terms of two classes: entitled and not entitled. For the evaluation 2400 records were again generated distributed evenly over the two classes. The not entitled cases were generated such that they fail to meet one and only one condition of the five conditions listed above. Noise was then applied to this dataset in the same manner as in the previous evaluation. However, in this case an extra noise level of 50% was added to this experiment. The dataset was randomly split into a 70% training set and a 30% test set. Noise was then applied to the training set in the same manner as reported above. Again the training dataset used for each of the noise levels was split equally between two PADUA players and the two players allowed to "argue" to classify the 720 cases in the test set (using the same support\confidence level as in the previous test). This experiment was not applied to CN2 or ABCN2, which were not available to us.

Table3 shows the affect of adding noise to the housing benefit dataset on the accuracy of each classifier. Here it can be notice that FOIL is the best classifier when using correct data (unlike the previous experiment), but again it can be observed that as the accuracy of all the classifiers drops with the increase in noise level in the data, PADUA is again more tolerant of noise that the other classifiers. The accuracy of PADUA drops 5.83% as the noise level is increased from 0% to 50% whereas the accuracy of FOIL (which worked well with clean data) drops 21.81% and the accuracy of Random Decision Trees drops 10.97%.

Table 3. Accuracy versus Noise (PADUA – Hosuing Benefit Dataset)

Noise	PADUA	Random DT	Info Gain DT	TFPC	CBA	CMAR	FOIL	CPA R	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

1.2. Evaluation using PISA

In order to use the Housing Benefits datasets to test PISA, the conditions mentioned in the previous sections were interpreted such that the final output would be increased from just two classes (entitled or not entitled). For the purpose of the example presented here a fourfold classification was used: fully entitled, entitled with priority, partially entitled and not entitled. The requirements for each class were defined as follows:

1. *Fully Entitled*: Candidates will be entitled to full housing benefit allowance if they satisfy all the above five conditions.
2. *Entitled with Priority*: candidates will entitle to housing benefit allowance with priority if they satisfy the entitling conditions and also satisfy the following:
   - Paid Contribution in four out of the last five years and
     - Have less capital than the original limit (this is interpreted as 1000£ less than the original limit).
     - Have has less income (5%) than the original limit.
   - They are member of the armed forces and have paid the contribution fees in five out of the last five years.
3. *Partially Entitled*: Candidates will be entitled to a lower rate of benefit if they satisfy the age condition but they either:
   - Have slightly more capital than the original limit (i.e. +1000£ more than the original limit), but have paid contributions in 4 (or 5) years out of the last five.
   - Or they have slightly more available income (i.e. +5%) than the original limit, but have paid contributions in 4 (or 5) years out of the last five.
   - Merchant navy members are also partially entitled if they satisfy all the other conditions and have paid the contribution in five out of the last five years.
4. *Not Entitled*: If the candidate fails to satisfy the conditions for full or partial entitlement.

In the same manner as reported above 2400 records were generated equally distributed over the classifications (entitled, priority entitled, partially entitled and not entitled). The same noise levels used for PADUA (0%, 2%, 5%, 10%, 20%, 40%, and 50%) were applied to the dataset. The training dataset used for each of the noise levels, was split into four equal subsets, each subset was given to one PISA player, and the four players in each subtest argued to classify the 720 cases in the test set. The support value was again fixed to 1% and confidence to 50% for all the classifiers (including PISA).

Table 4 shows the affect of adding noise to the housing benefit dataset on the accuracy of each classifier. From the table it can be seen that the overall accuracy level is lower than that recorded for the binary classification. The best overall classifier is PISA with an accuracy level starting with 98.47% for clean data and dropping to 93.75% when a 50% noise level is introduced indicating that the PISA protocol copes extremely well with noisy data compared to the other classifiers used in the evaluation.

Table 4. Accuracy versus Noise (PISA)

Noise	PISA	Random DT	Info Gain DT	TFPC	CBA	CMAR	FOIL	CPA R	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

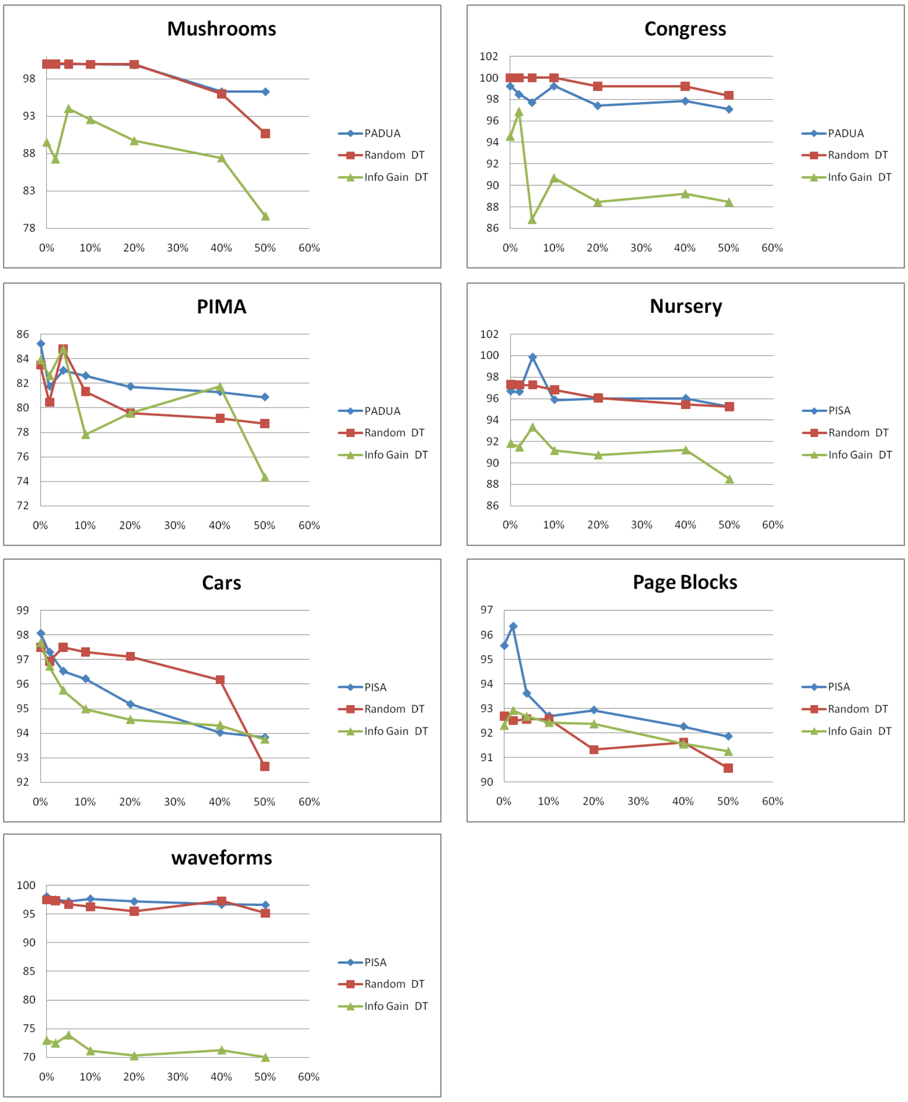
6. Further evaluation

The tests described above, use artificial datasets, mainly because we have full understanding of these datasets. But relying on just artificial datasets is not enough to demonstrate the tolerance to noise of PISA and PADUA. In this section we list some of the results obtained when testing PISA and PADUA using 7 real datasets. PADUA was used with the datasets containing 2 classes only (Mushrooms, Congress and PIMA) while PISA was applied to datasets with 3 classes (Wave Forms), 4 classes (Nursery and Car Evaluation) and 5 classes (Page Blocks).

This test compared the operation of both PADUA and PISA with the same classifiers as used before, but in this section we will only report on the comparison with decision trees classifiers, because decision trees were found to be the closest “competitors” to PADUA and PISA.

The results of this evaluation (figure 1) show a similar pattern to the benefits experiments: the accuracy of almost all the classes dropped when the noise percentage was increased. The only case in which PADUA or PISA performed worse than random trees with high level of noise is when the congress dataset was used. The reason is that this dataset is very small (435 rows), which means that each player has only 152 cases from which they should mine their arguments (association rules). This is rather a small size when a high level of confidence is used.

Fig. 1. Real datasets study (*in these graphs the horizontal axe represents the noise level and the vertical represents the accuracy*).



7. Conclusions

In this paper we have presented an overview of PADUA and PISA, two argumentation from experience systems applicable to two and multiplayer argumentation respectively. We have described how both systems can be applied to the classification problem and illustrated this by detailed experiments using two artificial welfare scenarios/data sets, and summary results for seven real datasets. Of note, other than the operation of the two systems, is that the argumentation from experience concept can successfully be applied to address classification. The results obtained indicate that the systems' performance is comparable to, or better than, other classification approaches. The particular advantage that the approach offers is that it operates very successfully in noisy environments, outperforming competitor classification systems. Ability to handle noisy data sets is of significant importance in many domains where sufficient data can only be obtained at the cost of including misclassified records. The authors are greatly encouraged by the reported results and are currently undertaking further investigation to evaluate the systems performance on a wider range of datasets.

References

[1] R. Agrawal, T. Imielinski, and A. N. Swami. Mining association rules between sets of items in large databases. In P. Buneman and S. Jajodia, editors, *SIGMOD Conference*, pages 207–216. ACM Press, 1993.

[2] R. Agrawal and R. Srikant. Privacy-preserving data mining. In *Proc. of the ACM SIGMOD Conference on Management of Data 2000 (SIGMOD'00)*, pages 439{450. ACM Press, May 2000.

[3] Bench-Capon, T. (1993). Neural Nets and Open Texture. In Fourth International Conference on AI and Law, 292–297. ACM Press: Amsterdam.

[4] Bench-Capon, T. and Coenen, F. (2000). An Experiment in Discovering Association Rules in the Legal Domain. In 11th International Workshop on Database and Expert Systems Applications, 1056–1060. IEEE Computer Society: Los Alamitos.

[5] M. Bendou and P. Munteanu. Learning bayesian networks from noisy data. In *Proc. of the 5th International Conference on Enterprise Information Systems (ICEIS 2003)*, pages 26{33, April 22-26 2003.

[6] C. E. Brodley and M. A. Friedl. Identifying and eliminating mislabeled training instances. *AAAI/IAAI*, 1, 1996.

[7] P. Clark and T. Niblett. The cn2 induction algorithm. *Machine Learning*, 3(4): 261-283, 1989.

[8] Clark, P. and Boswell, R. (1991). Rule Induction with CN2: Some Recent Improvements. In Machine Learning – Proceeding of the Fifth Europen Conference (EWSL-91), 51–163. Berlin.

[9] F. Coenen, P. H. Leng, and S. Ahmed. Data structure for association rule mining: T-trees and p-trees. *IEEE Trans. Knowl. Data Eng.*, 16(6):774–778, 2004.

[10] Coenen, F. and Leng, P. (2005). *Obtaining Best Parameter Values for Accurate Classification*. Proc. ICDM'2005, IEEE, pp597-600.

[11] Coenen, F., Leng, P. and Zhang, L. (2005). *Threshold Tuning for Improved Classification Association Rule Mining*. Proceeding PAKDD 2005, LNAI3158, Springer, pp216-225.

[12] Coenen, F., Leng, P. and Ahmed, S. (2004a). *Data Structures for association Rule Mining: T-trees and P-trees*. IEEE Transactions on Data and Knowledge Engineering, Vol 16, No 6, pp774-778.

[13] Coenen, F.P. Leng, P. and Goulbourne, G. (2004b). *Tree Structures for Mining Association Rules*. Journal of Data Mining and Knowledge Discovery, Vol 8, No 1, pp25-51.

[14] G. H. John. Robust decision trees: Removing outliers from databases. In *Proc. of the First International Conference on Knowledge Discovery and Data Mining (KDD'95)*, pages 174 - 179. AIII Press, 1995.

[15] Johnston, B. and Governatori, G. (2003). Induction of Defeasible Logic Theories in the Legal Domain. In Ninth International Conference on AI and Law, 204–213. ACM Press: Edinburgh.

[16] J. Kubica and A. Moore. Probabilistic noise identification and data cleaning. Technical Report CMU-RI-TR-02-26, CMU, 2002.

[17] Liu, B. Hsu, W. and Ma, Y (1998). *Integrating Classification and Assocoiation Rule Mining.* Proceedings KDD-98, New York, 27-31 August. AAAI. pp80-86.

[18] Li W., Han, J. and Pei, J. (2001). CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules. Proc ICDM 2001, pp369-376.

[19] M. Mozina, J. Zabkar, T. Bench-Capon and I. Bratko, Argument based machine learning applied to law, *Artificial Intelligence* **13** (1) (2005), pp. 53–73.

[20] H. Prakken. Formal systems for persuasion dialogue. *Knowledge Eng. Review*, 21(2):163–188, 2006.

[21] J. R. Quinlan. Simplifying decision trees. *International Journal of Man-Machine Studies*, 27(3): 221-234, 1987.

[22] Quinlan, J. R. and Cameron-Jones, R. M. (1993). *FOIL: A Midterm Report.* Proc. ECML, Vienna, Austria, pp3-20.

[23] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1998.

[24] S. Schwarm and S. Wolfman. Cleaning data with bayesian methods, 2000.

[25] C. M. Teng. *Correcting Noisy Data*. Machine Learning, 1999.

[26] Wardeh, M., Bench-Capon, T. and Coenen, F.P. Arguments from experience: The padua protocol. In P. Besnard, S. Doutre, and A. Hunter, editors, *COMMA*, volume 172 of *Frontiers in Artificial Intelligence and Applications*, pages 405–416. IOS Press, 2008.

[27] Wardeh, M., Bench-Capon, T. and Coenen, F.P. *PISA - Pooling Information from Several Agents: Multiplayer Argumentation From Experience*. In Proc. AI'2008, Springer, London, pp133-146.

[28] Yin, X. and Han, J. (2003). CPAR: Classification based on Predictive Association Rules. Proc. SIAM Int. Conf. on Data Mining (SDM'03), San Francisco, CA, pp. 331-335.