Taylor & Francis
Taylor & Francis Group

Arguing from experience using multiple groups of agents

Maya Wardeh*, Trevor Bench-Capon and Frans Coenen

Department of Computer Science, The University of Liverpool, Liverpool L69 3BX, UK

A framework to support "*Arguing from Experience*" using groups of collaborating agents (termed participant agents/players) is described. The framework is an extension of the *PISA multi-party arguing from experience framework*. The original version of PISA allowed *n* participants to promote *n* goals (one each) for a given example. The described extension of PISA allows individuals with the same goals to pool their resources by forming "*groups*". The framework is fully described and its effectiveness illustrated using a number of classification scenarios. The main finding is that by using groups more accurate results can be obtained than when agents operate in isolation.

Keywords: arguing from experience; multi-party argumentation; group argument; association rules, argumentation strategies

1. Introduction

Arguing from Experience was introduced in Wardeh, Bench-Capon, and Coenen (2008a) to provide a computational model of argument based on inductive reasoning from past experience. The model allows participants to draw directly from past examples to find reasons for coming to a *view* on some current example (case), without the need to analyse this experience into rules and rule priorities. Such reasoning can be found in informal everyday arguments, where we often make use of our (personal) experience when conversing with other people, by observing certain *regularities* in this experience, and then employing these *regularities* to back up what is being said. To model this, instead of drawing rules from a knowledge base, we construct our arguments *on the fly* using association rule mining (ARM) techniques (Coenen, Leng, and Ahmed 2004; Wardeh, Bench-Capon, and Coenen 2007).

The setting in which we explore this approach is a debate about how to classify an example, the identified associations then provide reasons for and against particular classifications. This form of argument differs from the more common style of arguments (Prakken 2006) generated from a knowledge base. In those arguments, the existence of a theory in the form of a collection of rules is presupposed. In systems based on knowledge bases, persuasion occurs through one participant telling the other(s) something previously unknown, either a fact or a rule: in arguing from experience, opinions on rules differ in respect of confidence, and the participants try to find a rule they both can accept on the basis of their different experience.

Argument from experience has several advantages:

(1) it is a common form of natural argument;
(2) it avoids the knowledge acquisition bottleneck, since there is no requirement to rationalise experience into rules;

---

*Corresponding author. Email: maya@csc.liv.ac.uk, maroshcka@gmail.com

(3) the arguments are not constrained by a predetermined theory, but can be adapted to the current context and opponents;

(4) it can allow the different experiences of different agents to be considered separately, whereas merging this experience would mask significant considerations.

In the context of previous work, the *PADUA Protocol* (Wardeh et al. 2008a) allowed two agents to dispute a binary classification. This was extended in Wardeh, Bench-Capon, and Coenen (2009a) to the *PISA Framework* to allow more than two participants to engage in the dialogue, and so support argument about problems with a range of classifications, with each participant championing a different possibility. In this paper, we present a more flexible version of the protocol, implemented in an extension to PISA, to allow *groups* of agents to argue for a classification. It is conjectured that groups of agents collaborating to produce some result will be more effective than if the agents were operating in isolation. Agents advocating the same or a similar thesis can confer and jointly select which arguments to put forward. This extension raises a number of issues:

(1) the process by which such agents can collaborate;
(2) how to form such groups and what roles to assign to their members;
(3) how to facilitate the discussion within these groups in order to produce a single argument to present to opponents.

For the first question, we suggest that such agents should form what we term a *group of participants* and make a single move rather than acting independently for the same position; our answers to the other two questions will be given later in the paper.

The rest of this paper is organised as follows. In Section 2, a summary of the PISA framework is given. The notion of strategy in PISA is introduced in Section 3. In Section 4, we explain the process by which agents can join forces and argue as a group (team) in PISA. Section 5 provides a detailed example of the process described in the previous section. In Section 6, we present the results of a series of experiments which we have performed to explore the operation of PISA with groups to determine whether groups improve the quality of the classifications. Section 7 provides some comparison with related previous work. Finally we conclude with a summary of the main points and some future directions.

2. PISA framework for arguing from experience

As described in Wardeh et al. (2009a), PISA concerns dialogues where there is a range of options (classes) for classification, and each of the participants is the advocate of one of these options. In order to accommodate situations where there are more agents than opinions, the extensions presented here will enable the agents to act in groups (Section 4), one group for each opinion. PISA dialogues are open: participants may enter or leave when they wish. For turn taking, a structure with rounds is adopted, rather than a linear structure where a given agent is selected as the next speaker (e.g. the turn taking protocol in Bel-Enguix and Lopez (2006) where the current speaker chooses who will speak next). In each round, any participant who can make a legal move may do so. There is no limitation on the number of parties that can participate in any round but, to simplify the game, each participant is limited to one move per round. This turn taking policy allows participants to place their attacks/counter attacks as soon as they become appropriate, without the need to wait for a turn to contribute. This is not perhaps the most usual structure for human meetings, but it can be found in some board games such as Diplomacy, which also has the feature of a many player game in which every player is the opponent of every other player.

It is suggested that the structure is particularly appropriate in order to achieve fairness in the situations where every participant is playing for themselves, and has to regard every other participant as an opponent. The game terminates when no participant makes a contribution for two rounds (to ensure that they have really finished and not withheld a move for tactical reasons) or after some limiting number of rounds have been played: thus the termination of the game is guaranteed. The model is essentially that of a facilitated discussion, with the chairperson acting as the facilitator. The realisation of this model and the choices summarised above are considered in the following sub-sections.

2.1. *The components of the framework*

The framework comprises three key components:

(1) a number of *participant agents* (players), each advocating one opinion (goal/view/possible classification).
(2) A software agent playing the role of mediator, *the chairperson*, which does not advocate any position, but rather manages and facilitates communication between the clashing participant agents.
(3) A central argument structure, termed the *argumentation tree*, which stores the arguments exchanged in the course of PISA dialogues.

In the following subsections, each of these three components will be briefly discussed.

2.1.1. *The participant agents*

Each participant agent has its own individual repository of experience, in the form of tabular data set. *Tabular data set* is defined as follows: let $I = \{i_1, i_2, \ldots, i_n\}$ be a set of attributes, such that each attribute $i_k \in I$ has a set of possible values $V_i = \{v_{i_1}, v_{i_2}, \ldots, v_{i_m}\}$. Such that i_1 has a set of possible values $V_{i_1} = \{1, \ldots, j\}$, i_2 has a set of possible values $V_{i_2} = \{j + 1, \ldots, k\}$ and so on. Then a tabular data set D is a set of database records, each record $R \in D$ is a set of pairs $\langle i, v \rangle$: $i \in I$ and $v \in Vi$.

Using their distinct databases, PISA participant agents produce reasons for and against classifications by mining association rules (ARs) from their data sets using ARM techniques (the application of ARM to arguing from experience can be found in Wardeh et al. (2007)). Additionally, the agents are equipped with a strategy model (Section 3) to help them in generating their arguments. ARs (Agrawal, Imielinski, and Swami 1993) are probabilistic relationships expressed as rules of the form $A \rightarrow W$ which should be read as follows: *if A is true then W is likely to be true*, or *A is a reason to think W is true* where A and W are disjoint subsets of some global set of attributes. In the context of "*arguing from experience*", ARs represent a means to draw arguments from individual experiences. Such arguments can be read as follows: *In my experience, typically things with features A are Ws; this case has those features, and so it is likely to be a W.*

In the context of ARs, likelihood is usually represented in terms of a confidence value expressed as a percentage. For an AR $A \rightarrow W$ this is calculated as: support$(XW) \times 100/$support(X). Where support(X), where X is a set of pairs $\langle i, v \rangle$: $i \in I$ and $v \in V_i$, is the number of records in D that contains X (i.e. $|R_X|$ where $R_X \subseteq D$ and $X \subseteq R_X$).To limit the number of rules generated, only attribute sets with support above a user-specified support threshold are used to generate associations. To further limit the number of associations, only those rules whose confidence exceeds a user-specified confidence threshold are accepted.

Each participant agent can use one of six speech acts, which collectively form the basic building blocks for constructing dialogues in the proposed model. These speech acts fall under three basic types: (i) *stating* a position, (ii) *attacking* a position and (iii) *refining* a position, as follows:

- *Stating a position*: *Propose rule* (*SA1*). Allows generalisations of experience to be cited, by proposing a new AR with a confidence higher than a certain threshold (SA here stands for Speech Act).
- *Attacking a position*: We provide three ways of attacking a proposal, two of which reduce the confidence of a rule already proposed, while the third proposes a rule with higher confidence than an existing rule: (i) *Distinguish* (SA2) whereby the agents attempt to mine a new rule, by adding some new premise(s) (from the case under discussion) to a previously proposed rule, so that the confidence of the new rule is lower than the confidence of the original rule. If such a rule exists, then the agent can put forward SA2. (ii) *Unwanted consequences* (SA3) allows the inclusion of additional features in the consequences (conclusions) of the rule under discussion that do not match the case under consideration; and (iii) *Counter rule* (SA4) is used in a very similar manner to propose rule (SA1) to cite reasons for supporting a different classification.
- *Refining a position*: These speech acts enable the modification of a previously played rule to meet objections: (i) *Increase confidence* (SA5), whereby the agent attempts to mine a new rule, by adding additional features (from the case) to the premises of the a rule previously played to increase the confidence of the previous rule. If such a rule exists, then the agent can put forward SA5; and (ii) *Withdraw unwanted consequences* (SA6) excludes the unwanted consequences of a rule previously proposed, while maintaining a certain level of confidence.

Note that the proposed speech acts are very different from those found in persuasion dialogues based on belief bases, a summary of which can be found in Prakken (2006). Rather, these speech acts, especially SA1, SA2 and SA4, have a strong resemblance to the speech acts used in arguing on the basis of precedent examples in common law, especially (Ashley 1990; Aleven 1997). The PISA speech acts differ from case-based reasoning in law, however, as they reflect the whole of an individual agent's experience, represented by a data set of previous examples used *collectively*, rather than a single case taken to be the closest *precedent*. Unlike legal decisions, the authority of PISA's arguments comes from the frequency of occurrence in the set of examples rather than endorsement of a particular decision by a court with the appropriate status.

2.1.2.   *The chairperson*

The chairperson is a neutral agent which administers a variety of tasks to facilitate multiparty arguing from experience dialogues. This mediator agent resembles the mediator artefact suggested in Oliva, Viroli, Omicini, and McBurney (2008a) and it has the following responsibilities:

- Starting a dialogue involving a set of participants to classify a given case.
- Accepting or rejecting proposed moves.
- Maintaining the argumentation tree to reflect the moves made by the participants and consequent changes in the status of the arguments.
- Making decisions regarding agents requesting to join or to withdraw.
- Monitoring the dialogue. This involves registering, for each round played, which agents have taken part and which have not.
- Terminating the dialogue, once a termination condition is satisfied.
- Announcing the game's winner through consultation of the argumentation tree.

- Excluding (removing) participants from the game if they fail to contribute in the game for a predetermined number of rounds.

2.1.3. *The argumentation tree*

The notion of an *argumentation tree* is used in PISA to describe the central data structure representing the arguments exchanged in a dialogue, and the attack relations between those arguments. This tree acts as a mediating artefact for the dialogue of the sort advocated in Oliva, McBurney, and Omicini (2008b). The tree structure comprises the arguments put forward by the participants, organised so that children are the attackers of their parents. It uses four colours to mark the status of the arguments played so far, and two types of links: *explicit* links representing direct attacks, and *implicit* links representing indirect attacks. The issue of addressing that arises in multiparty dialogues is resolved via the direct links. A move is addressed to the participant that played the argument attacked by this move (except for the first move in the game which is addressed to all the other participants).

The nodes of the argumentation tree represent the speech acts (moves) exchanged in the dialogue, and the links between them represent the attack relations between these moves. Each node has one of four colours: green, blue, red or purple. Nodes are either green or blue when introduced: green if they propose a new AR (SA1,SA4,SA5,SA6); or blue if they only attempt to undermine an existing association (SA2, SA3). Red nodes are those directly under attack and purple nodes are those indirectly attacked. Nodes change their colour according to Table 1. The tree also has a global value: the *Green Confidence,* which represents the highest confidence of the undefeated green node(s).

When a participant plays a move (*m*), it must satisfy a number of conditions in order to be added as a node to the argumentation tree, otherwise it will be rejected. These conditions are as follows:

- *m* is added to the tree if and only if it changes the colouring of the tree. In consequence, participants are not allowed to attack, for instance, red nodes (defeated moves), as these attacks will not change the colouring of the red node, nor that of the branch of the argumentation tree in which the red node is located. Note, however, that participants may attack purple nodes, as direct attacks against purple nodes will change their colouring to red.
- *m* must explicitly attack the move it is associated with (parent node).

Table 1. The colours used in the argumentation tree.

Colour	Meaning	Shifts to
Green	SA1, SA4, SA5 or SA6 node, undefeated in the given round	Red: If attacked by at least one undefeated node
		Purple If indirectly attacked by an undefeated green node with higher confidence
Red	The node is defeated in the given round	Green: If all attacks against it are successfully defeated and the original node colour was green
		Blue: If all attacks against it are successfully defeated and the original node colour was blue
Blue	SA2 or SA3 move node undefeated in the given round	Red: If attacked by at least one undefeated node
Purple	SA1, SA4, SA5 or SA6 move node indirectly attacked by a higher confidence green node, played by a different participant	Green: If all attacks against it are successfully defeated, and if the move(s) indirectly attacking this node was defeated
		Red: If attacked by at least one undefeated node

- A participant can put forward only one move per round (deciding which rule to play is strategy issue).
- Participants cannot play moves that weaken their position, such that another participant would take the lead. This condition holds when a participant tries to attack a blue node that was originally made to attack an argument proposed by other participants, unless this move changes the colouring of that argument to purple.

Additionally there are conditions to guard against repetition of the same or similar moves, to avoid needless repetition and the possibility of endless loops. A similar move is one where the content of the move is an AR with the same premises and conclusions, differing only in the degree of confidence.

(1) One participant cannot repeat the same attacking move (with the same AR) against different opponents if:
  (i) This attack is either a distinguishing or unwanted consequences attack.
  (ii) If all of the other previously played moves using this attack are still green (undefeated) on the argumentation tree.
(2) Participants cannot attack an opponent using moves that have already been played against and defeated by this opponent.
(3) Suppose two or more participants have coincidentally attacked the same opponent in the same round, using similar attacks. If the confidence is equal in all of these attacks, then the participant under these attacks is required to defend its proposal against them once only. Otherwise the chairperson chooses the attack with the highest confidence (lowest confidence in case of distinguishing) and discards the rest.

Moves attack one another as follows:

- Moves with speech acts SA1, SA4, SA5, and SA6 implicitly attack all other $SA1$, $SA4$, $SA5$, and $SA6$ moves played by other participants, which have content with lower confidence.
- Moves $SA2$ and $SA3$ affect only the nodes they directly attack.

Once a game has terminated, the chairperson consults the argumentation tree to determine the winner. The winner has to satisfy one of the following rules:

- *Rule* 1: If all the green nodes belong to the same participant, then the classification (goal/view) supported by that participant is undefeated. The case under discussion is classified accordingly. This condition is realised when no other participant has played an undefeated move with higher or similar confidence.
- *Rule* 2: If there are no green nodes, and all the blue nodes were played by the same participant, that participant wins, and the case is classified according to this participant-supported classification (goal/view) by the virtue of the fact that it defeats all other possible classifications.

It is not always the case that the dialogue games conducted within PISA result in a clear winner. There are two scenarios:

(1) Upon the termination of the game, there may be two or more green nodes with the same confidence, each belonging to a different participant. This situation may occur if the confidence value of these nodes are the highest (indirectly attacking all the other potentially green nodes), or if all the other nodes with higher confidence values are defeated.

(2) The argumentation tree may not contain any green moves at the end of the game, because, for example, all the green moves have been defeated in the course of the game. Alternatively, the undefeated blue nodes may have been played by a number of different participants.

The first case is considered a *strong tie situation*, as the participants have actually proposed an opinion within the game. One possible solution is to initiate a new game involving only the tying parties and see how this game develops. However, there is no guarantee that this game will not also lead to another tie. In this case, the chairperson will be forced to announce a tie (after the second game or after predefined number of games with the tying parties from previous ones). The second case is seen as a *weak tie situation*, as the tying participants did not actually have any positive support for their classifications at the end of the game. In such cases, enforcing a second game may be of great benefit, but with the requirement that the participants should propose as many reasons for their classifications as they can this time, by adopting a appropriate and possibly different strategy.

In the next section, we will formally define the notions introduced in this section.

2.2. *PISA formal framework*

PISA is formally defined as follows:

$$\text{PISA} = \langle \text{chairperson, argumentation tree, participants} \rangle \tag{1}$$

Participants is the set of participant agents taking part in PISA dialogue and is defined as follows.

Whereby $\forall P \in$ participants then:

$$P = \langle D_i, A_i \rangle, \tag{2}$$

where:

(1) $D_i = \{R_1, R_2, \ldots, R_m\}$ is a set of tabular data set records. Each record $R \in D$ is a subset of attributes in $I = \{i_1, i_2, \ldots, i_k\}$ the set of attributes. Each attribute $i \in I$ has a set of possible values $V_i = \{v_{i_1}, v_{i_2}, \ldots, v_{i_m}\}$. An AR $ar(P \rightarrow Q, c)$ can be identified in D_i as follows:
  (i) P: the premises of the rule – $P = \{(i_{p_1}, v_{x_1}), (i_{p_2}, v_{x_2}) \cdots (i_{p_k}, v_{x_k})\}$ – such that for each tuple (i_{p_h}, v_{x_j}): $i_{p_h} \in I$ and $v_{x_j} \in V_{i_q}$.
  (ii) Q: the conclusion of the rule – $Q = \{(i_{q_1}, v_{y_1}), (i_{q_2}, v_{y_2}) \ldots (i_{q_h}, v_{y_h})\}$ – such that for each tuple (i_{q_h}, v_{y_j}): $i_{q_h} \in I$ and $v_{y_j} \in V_{i_q}$.
  (iii) $P \cap Q = \emptyset$
  (iv) c: rule confidence, which means that $c\%$ of the records in D_i that contain P also contain Q (the conditional probability of Q given P (Agrawal et al. 1993).
(2) $A = \{A_1, \ldots, A_n\}$ the set of participant agents:

$$\forall A_i \in A : A_i = \langle \text{name}_{A_i}, C_{A_i}, \Sigma_{A_i}, V_{A_i}, \text{play}_{A_i} \rangle, \tag{3}$$

where:
  (i) name_{A_i}: the participant agent name.
  (ii) $CA_i = \{c_1, c_2, \ldots, c_n\}$ the set of classifications supported by this agent (i.e. the agent will try to prove that the case under discussion should be classified as one of these classes). In this application of PISA, CA_i has one element only $CA_i = \{c_{A_i}\}$.
  (iii) Σ_{A_i} is a representation of the underlying data set D_i. This representation enables participants to mine for the suitable ARs as needed. For instance, Σ_{A_i} might be

represented as the following tuple: $\Sigma_{A_i} = \langle T_{D_i}, \text{Conf}, \text{Genr}_{A_i} \rangle$, where T_{D_i} is a summarising tree of the data set D_i; Conf is the *Confidence Threshold*, representing the lowest acceptable confidence: rules with confidence lower than this threshold are considered invalid arguments; Genr_{A_i} is a function that mines a suitable rule, from the summarising data structure according to some set of conditions Cond: $\text{Genr}_{A_i} : T_{D_i} \times \text{Cond} \rightarrow ra$.

(iv) V_{A_i} represents the agent's view of the underlying argumentation tree structure. The implementation of this view depended on the agents' strategies: it could take into consideration only the moves that have been added to this structure in the last round; or a more sophisticated view covering the entire tree, or anything in between.

(v) play_{A_i} is the strategy function for agents taking part in PISA games. The specification of this function is given in Section 3.

The *argumentation tree* is defined as follows:

$$\text{ArgT} : \langle \text{Nodes}, \text{Attack} \rangle, \tag{4}$$

where $\text{Nodes} = \{n_0, \ldots, n_m\}$ is the nodes of the argumentation tree such that $\max(|\text{Nodes}|) < \text{predefinedthreshold}$. Attack is the attack relation between the tree's nodes, as represented by the direct and indirect attacks identified previously.

The *Chairperson* entity is defined in the form of a tuple as follows:

$$\text{Chairperson} = \langle P, \varphi, \text{Diag}_{\text{initial}}, \text{Diag}_{\text{tie}}, \text{start}, O \rangle, \tag{5}$$

where:

(1) P is a protocol specifying the legal moves at each stage of a dialogue. P is formally defined as the function: $P : M \rightarrow 2^M$, where M is the set of possible moves. Thus P identifies the combination rules for the dialogue games taking place under this framework. Table 2 summarises these rules and indicates where a new set of reasons is introduced to the discussion. Each move $m \in M$ is a tuple $m = \langle \text{sa}, \text{content} \rangle$ such that $\text{sa} \in SA$ is the move's speech act. SA represents the Speech Acts discussed in Section 2.1.1; content is the content of this move: $\text{content} = \text{ar}(p \rightarrow Q, c)$.

(2) φ: The instance under discussion.

(3) $\text{diag}_{\text{initial}}$: PISA initial dialogue and is identified as follows $\text{diag}_{\text{initial}} = dm_1, \ldots, dm_n$, where $dm_i \in \text{DM}$ is a dialogue move: $dm = \langle A_i, H, m, t \rangle$, where

  (i) $A_i \in A$ is the agent that utters the move.

 (ii) $H \subseteq A$ denotes the set of agents to which the move is addressed.

(iii) $m \in M$ is the move.

Table 2. The legal next speech acts in PISA.

Speech act	Label	Next speech act
SA1	Propose rule	SA3, SA 2, SA4
SA2	Distinguish	SA3, SA5, SA1
SA3	Unwanted cons	SA6, SA1
SA4	Counter rule	SA3, SA2, SA1
SA5	Increase conf	SA3, SA2, SA4
SA6	Withdraw unwanted cons	SA3, SA2, SA4

(iv) $t \in DM$ is the target of the move (the move which it replies to). dm_1 (the initial move) does not reply to any other move.

(4) Diag_{tie}: set of PISA dialogues to resolve ties. Such that $|\text{Diag}_{\text{Tie}}| \leq TS$. Where TS is a predefined threshold.

(5) start: a function that begins a certain PISA dialogue, e.g. start($\text{diag}_{\text{tie}} \in \text{Diag}_{\text{tie}}$) begins a tie resolution dialogue, start($\text{diag}_{\text{initial}}$) begins the initial one.

(6) O denotes the outcome rules of the dialogues. These rules define for each dialogue $d \in (\text{Diag}_{\text{tie}} + \text{diag}_{\text{initial}})$ and instance φ the winners and losers of d with respect to φ. PISA applies the following rules:

    (a) Winners: $w_\varphi(d, A_w \subseteq A)$ is true if $\forall A_i \in A_w$, then $C_{A_i} \subseteq O(d, \varphi)$.

    (b) Losers: $l_\varphi(d, A_l \subseteq A)$ is true if $\forall A_i \in A_l$, then $\forall c \in C_{A_i} then c \notin O(d, \varphi)$.

    (c) $O(d, \varphi) = \{o1, \ldots, o_m\}$ such that $\forall o \in O$, then $o \in \bigcup_{j=1}^{j=|A|}(C_{Aj})$, and:

- $o \in \text{consequences}(AR(N_G))$: $N_G \in \text{Node}(ArgT)$ and the colour of each node from N_G is *Green*, or
- $o \in C_{\text{dominantBlue}}$ such that $\text{dominantBlue} \in A$ and it has played all the blue nodes on the argumentation tree.

    (d) If $|w_\varphi| > 1$ then there is a tie between the agents in A_w.

2.3. *PISA dialogue protocol*

Assume that we have an instance that requires classification and a number of PISA participant agents, each promoting one of the possible classifications in the domain from which this instance was drawn; the PISA dialogue protocol works as follows:

(1) Before the start of the dialogue, the chairperson selects one participant ($P_1 = (A_1, D_1) \in$ Participants) to start the dialogue. At the first round, P_1 proposes a new argument $ARG1$ (the content of dm_1) such that the confidence of this argument is higher than the user-defined threshold. The chairperson instantiates a new argumentation tree, whose root represents ARG1. If P_1 fails to play an opening move, the chairperson selects another participant to commence the dialogue. If all the participants fail to propose an opening argument, the dialogue terminates with failure.

(2) In the second round, the other participants attempt to attack ARG1 using any of the attacking speech acts discussed previously. If all the participants fail to attack ARG1, the dialogue terminates, and the case is classified according to the rule advanced by P_1. Otherwise, the argumentation tree data structure is updated with the submitted attacks.

(3) Before the beginning of each of the subsequent rounds, the chairperson removes the participant agents that have not taken part in the last m rounds from the dialogue and updates the argumentation tree accordingly. If only one agent remains in the game after all the other participants have withdrawn or were removed by the chairperson, then the dialogue is terminated, and the case is classified according to the class promoted by this participant. Otherwise, any participant who can play a legal move, according to the protocol presented in Table 1, can do so; and the argumentation tree data structure is updated with the submitted attacks

(4) If one round passes without any participant making a legal move, the chairperson informs all the involved agents that the dialogue will terminate in the following round if they do not make any move.

If two consecutive rounds pass without any new moves being submitted to the argumentation tree, or if n rounds have passed without reaching an agreement, the chairperson terminates the dialogue. Then, it consults the argumentation tree to identify the winner(s). If no winner

can be identified, a tie-break dialogue is initiated between the tied parties, and the dialogue restarts from step 1. Otherwise, the case under discussion is classified according to the classification proposed by the winner. The suggested facilitated discussion scenario enjoys the following advantages:

- *It increases the flexibility of the overall operation of PISA*: By assigning the majority of protocol surveillance to the chairperson, the system gains significant flexibility with regard to the participating agents. For instance, the system can be switched between "closed" and "open" by applying a few limited changes to the chairperson definition, while the rest of the participants remain unaffected.
- *It is a very simple structure*: There is no complicated turn taking procedure involving a choice of the next participant, allowing the internal implementation of the participants to be kept as simple as possible.
- *It provides a fair dialogue environment*: The organisational configuration of the dialogue is neutralised by restricting the control tasks to the chairperson, which is not allowed to take sides in the dialogue.

3. Strategy model for PISA

The interaction of arguments in PISA is viewed as a form of argumentation dialogue game McBurney and Parsons 2002. This section is concerned with the strategy[1] problem in PISA: choosing the best move among the set of possible moves available to each participant agent at each round of the PISA dialogues. The outline of *PISA strategy problem* is as follows: each agent must select (i) the kind of move to be presented in the dialogue; (ii) the particular content of the move; and, most importantly, (iii) an opponent to direct this move against. We think that the notion of speech acts and content selection in PISA is best captured at different levels, as suggested in Moore (1993). A six-level strategy model was therefore designed for individual player agents taking part in PISA dialogues. The proposed six levels are divided into two tiers. The lower tier encapsulates the basic strategy model, while the upper tier determines how PISA players use the argumentation tree in selecting their moves. The proposed strategy model works as explained below.

The lower tier comprises a four-level layered strategy model:

- *Level* 0 – *Game mode*: This level distinguishes PISA games into two basic classes: (i) *win* mode, in which participants attempt to win using as few steps as possible, so exposing the least amount of information to the opponent. (ii) *dialogue* mode, in which games are played to fully explore the characteristics of the underlying argumentation system and the dialogue game.
- *Level* 1 – *Agent profiles* Amgoud and Maudet (2002): PISA has two agent profiles: (i) *Agreeable* agent accepts whenever possible (attempts to agree with arguments proposed by other participant agents, and only promotes own argument when agreement is not possible), or (ii) *Disagreeable* agent only accepts when no reason not to. Additionally, PISA players applying an agreeable profile may try to agree either with all the other participants or with a pre-specified group of participants.
- *Level* 2 – *Strategy mode*: *Build* or *destroy* mode. In the first mode, participants aim to win the game by proposing new rules, thus *building* a strong argument. In the second, players try to win by *destroying* the adversary's argument by undermining them either by distinguishing (SA2) these arguments or by pointing to their unwanted consequences (SA3).
- *Level* 3 – *Tactics*: Concerns choosing some appropriate content for the arguments depending on the tactics and heuristics suggested.

The upper tier identifies the manner by which PISA participant agents infer what move to play next from the current status of the argumentation tree. This tier comprises two strategy levels:

- *Level U2*: defines whether an agent *has to* participate in the current round of dialogue or not.
- *Level U1*: defines the process by which agents choose their next moves with respect to the argumentation tree. By consulting this tree, participants can base their decision on a number of issues with relation to their next moves, such as which opponent to attack next and which speech act to use.

Level U1 of the layered strategy has three different modes for deducing the next moves from the current status of the argumentation trees:

(1) *Full tree inference mode*: Enables the derivation of strategies using a full view of the argumentation tree.
(2) *Leaf nodes inference mode*: Leads to strategies with a limited view of the argumentation tree, considering only the leaf nodes of the tree (un-attacked moves), when making decision about what moves to play next.
(3) *One leaf node only inference mode*: By which the attention of participants is focused on one and only one un-attacked move.

Recall from Section 2 that a strategy function Play_a was identified for each PISA participant agent ($a \in A$). The details of this function are as follows:

$$\text{Play}_a : M_{\text{poss}} \times R_{\text{poss}} \times D_{\text{current}} \times S_a \times \text{Tactics}_a \rightarrow M_{\text{poss}}. \tag{6}$$

S_a is the *strategy matrix* for the given agent such that:

$$S_a = [\text{havePart}_a, \text{tm}_a, \text{gm}_a, \text{profile}_a, \text{sm}_a], \tag{7}$$

where havePart_a identifies whether a has to take part in the next round or not, havePart_a is either true or false; tm_a is the tree inference mode used by a, and is one of {full, leaves, oneleaf}; $\text{gm}_a \in \text{GM}$ is the game mode, where $\text{GM} = \{\text{win, dialogue}\}$; $\text{profile}_a \in \text{Profile}$ is the agent profile, where $\text{Profile} = \{\text{agreeable, disagreeable}\}$; $\text{sm}_a \in \text{SM}$ is the strategy mode, where $\text{SM} = \{\text{build, destroy}\}$. Tactics_a is the tactics matrix including the move preference and the best move content tactics.

A number of different strategies can be derived from the two-tier strategy model discussed above according to the values given to each of the parameters of S_a using the above strategy function. Three basic strategies were derived from the given model in relation to level U2 from the upper tier, from which a number of sub-strategies can be derived. Note that strategies are numbered from S1 to SN, and where appropriate sub-strategies of a strategy are indicated using SK-M, for example S1-1, and so on. Table 3 provides a summary of these strategies and sub-strategies. The table gives the strategy identifier (column 1) and where appropriate the root strategy (column 2); columns 3 and 4 provide the tree inference (Level U1) and strategy modes (respectively). The last column presents the ranking of each strategy: strategies are ranked from one downwards, with one being considered the *smartest* strategy. The significance of the rankings will be clarified in Section 4. The strategies are discussed in more detail below.

(*S1*) *Attack whenever possible strategy:* PISA participant agents may adopt this strategy to enhance their chances of winning the game by being as aggressive as possible: they attack whenever

Table 3. Summary and suggested ranking of PISA strategies.

Name	Strategy (S1, S2, S3)	Tree inference mode	Strategy mode	Rank
S3	–	–	–	1
S2-3-2	S2	Tree dependent – Full	–	2
S1-3-2	S1	Tree dependent – Full	–	3
S2-3-1	S2	Tree dependent – Leaf nodes	–	4
S1-3-1	S1	Tree dependent – Leaf nodes	–	5
S2-2-1	S2	Focused	Build	6
S2-2-2	S2	Focused	Destroy	7
S2-1-1	S2	Blind	Build	8
S2-1-2	S2	Blind	Destroy	9
S1-2-1	S1	Focused	Build	10
S1-2-2	S1	Focused	Destroy	11
S1-1-1	S1	Blind	Build	12
S1-1-2	S1	Blind	Destroy	13

they can do so with a legal move. Three sub-strategies are derived from the attack whenever possible strategy according the modes of U1 identified above. Each mode will use a different process to identify its opponents:

(1) *Blind attack whenever possible sub-strategy* (S1-1): Here, PISA participants attempt to arbitrarily attack any of the undefeated previous moves; thus the opponent identification process is random (blind). This sub-strategy applies a one leaf node only inference mode. Here two sub-sub-strategies are distinguished (build S1-1-1 and destroy S1-1-2).

(2) *Focused attack whenever possible sub-strategy* (S1-2): Directs the participant's attacks according to some ordering of the identified undefeated previous moves. Thus, it promotes a leaf nodes inference mode. The proposed strategy uses the following ordering: the agents will attempt to attack any undefeated nodes representing the most threatening direct attacks against moves they have previously played. Such attacks are identified as the leaf nodes directly attacking moves placed by a particular agent. Here two sub-sub-strategies are also distinguished (build S1-2-1 and destroy S1-2-2).

(3) *Flexible attack whenever possible sub-strategy* (S1-3): Represents the most sophisticated of the three sub-strategies derived from S1. Here individual participant agents adopt a focused strategy similar to the one discussed above. However, instead of being restricted to a build or destroy strategy mode, players may choose whether to undermine an existing leaf node (undefeated move) or to propose a counter attack against this node. The switch from build to destroy mode, and the other way around, depends on the inference mode (Level U1) of the applied strategy. Two sub-sub-strategies are derived:

  (i) *Leaf nodes inference flexible attack whenever possible –* (S1-3-1): Deduces next moves from the set of the previous un-attacked moves (leaf nodes). Here, besides distinguishing between the most threatening direct attacks and other leaf nodes, the agents apply some ordering on the most threatening direct attacks, and then try to defeat these attacks. Here we assume a simple order of direct threats: Green attacks are ordered in a descending order according to their confidence followed by the blue attacks ordered in an ascending order according to their confidence.

  (ii) *Full tree inference flexible attack whenever possible –* (S1-3-2): Promotes a full tree perception with the intention of determining an order in which the player should attempt to attack its opponents.

(*S2*) *Attack only when needed strategy*: Here, participants attack their opponents only when all their proposed arguments so far have been successfully attacked, or when their attempts to undermine all the other participants have failed. S2 is further divided into six sub-strategies (S2-1-1, S2-1-2, S2-2-1, S2-2-2, S2-3-1, and S2-3-2), corresponding to the variants proposed for S1 (see above).

(*S3*) *Attack to prevent forecasted threat strategy*: S3 anticipates forthcoming attacks against the participant's existing proposals; thus it is the most sophisticated strategy type in PISA. Here agents deduce their best next moves based on the entire argumentation tree and use their own heuristics trying to calculate which of their previous moves may be the weakest link in their argument, and then either propose new rules to strengthen their position, or attack the positions of other participants before they have the chance to attack them.

4. Arguing in groups

Having introduced the PISA framework for multiparty arguing from experience, the process whereby PISA agents can collaborate together to jointly produce their arguments can be considered. In the extended variant of PISA described here, where there are more agents than opinions, individual participant agents advocating the same thesis (e.g. the same possible classification) are required to *join forces* and act as a single group of participants. Each group then acts as a single participant. This notion of groups prevents individual players sharing the same objective from arguing without consulting each other and consequently causing contradictions among themselves or attacking each other. Group formation in PISA is a straightforward process: before the start of a new dialogue, all participants advocating the same thesis (classification) are required to form a group of participants. This also applies to new agents joining an ongoing dialogue. In each group, one member is selected to be the leader of the group who is usually the smartest and most experienced member of the group. A player's *smartness* relates to its strategy. Hence, the smartest member is the one with the most sophisticated strategy among the group's members, where strategies are ranked according to their level of understanding of the history and the process of the dialogue. The leader guides the inter-group dialogue, and selects which of the moves suggested by the group's members to play in the next round. This inter-group dialogue is a variation of targeted broadcasting where only group members can *listen* to what is being discussed. The leader can also redirect other members' moves against different opponents, or advise them to follow its own strategy. This allows the group to benefit from the different moves suggested by its members according to their different strategies and based on their different experiences.

4.1. *Group types*

Two factors define the nature of groups: the strategy to be adopted and the experience of its members. Groups in PISA are divided into two types according to the strategy factor:

(1) *Homogeneous groups*: Groups of participant agents which apply the same strategy. In such groups, the most experienced player (the one with the largest background data set) is the group's leader. If two or more of the group members share the same level of experience, then one of them is selected at random to represent the group.

(2) *Heterogeneous groups*: Groups of participant agents with different strategies. Therefore a strategy ranking is applied to determine who is the smartest player among the group and thus best suited to be its leader. If two or more players have the smartest strategy, then the most experienced one is selected. If they also have the same experience then one of them is selected at random. PISA applies the strategy ranking described in Table 3 to determine the smartest possible strategy.

4.2. *The role of the group leader*

The group leader (GL) has authority over the other members of the group that allows the leader to perform the following tasks:

- Selecting the best move at every round of the dialogue, from the candidate moves suggested by the group's members. Here, the differences in the member's experiences will greatly influence the leader's decision; members with different experience will often promote different content for their chosen moves, even where all the members apply similar strategies.
- The leader can compel the more experienced members (if any) to act according to the leader's strategy. If a more experienced member suggests one move, in a given round, and if the leader judges that a better move could be produced by this member by following the leader's strategy, then the leader can instruct this agent to attempt to generate the leader's choice. Additional conditions are applied to ensure that the leader exerts the above authority only when needed, that is if the experienced members of the group apply weak strategies, and where other members have failed to produce adequate moves.
- The leader can redirect moves suggested by the other members against opponents other than the ones they have chosen. For instance, if one member suggests an *increase confidence* move against one opponent, then the leader may change this move to a *propose new rule* directed at another opponent (possibly because this opponent threatens the group more). The leader is allowed to redirect moves only when redirection is more rewarding than the original move.

The role of the leader is not fixed, and it may change during the course of the dialogue when a new member joins the group, or when the current leader leaves the dialogue. Therefore PISA uses a token-based technique to identify the group's leader, similar to that of Ambroszkiewicz, Matyja, and Penczek (1998) where the token is used as a sign of decision power among a team of software agents.

4.3. *Intergroup dialogue model*

At the start of a new round, if the leader applies a wait-and-see strategy (S2), then the leader assesses the argumentation tree and makes a decision as to whether the group has to participate or not. If there is no need to take part in the round, the leader passes the round. Otherwise it requests the other members to suggest moves, and the following dialogue process occurs:

(1) Once all the group members have suggested moves according to their strategies and experience, the leader compares the member's moves against its own strategy and selects the best move according to its own strategy (e.g. if applying a build strategy, then the move with the highest confidence), and submits it.

(2) Otherwise if the leader cannot find a move that follows its own strategy and cannot generate this move itself then it can re-direct the submitted moves (if possible) to match its own strategy (e.g. transforms an increase confidence move to counter attack).

(3) Otherwise if there exists at least one group member such that the member's experience exceeds a predefined limit, then the leader can ask this member to generate a move following the leader's strategy. If a successful move is generated, the leader submits it to the chairperson. Otherwise the leader re-submits the request to the other group members.

(4) Otherwise if no desirable moves were generated, then the leader can select a move from the initially submitted moves to play. If all the group members fail to suggest any move, the leader passes the round and submits no moves.

4.4. *Discussion*

Thus far, the processes by which groups of individual participant agents with a common objective (goal, classification) are formed, and a leader for each group is selected, have been described. Note that the proposed notion of groups prevents individual participant agents, sharing the same objective, from arguing without consulting each other and consequently causing contradictions among themselves or attacking each other. The intuition behind group formation in PISA is that each of the group members will generate the best possible argument according to their experience/strategy, but only by the inter-group dialogue process (Section 4.3) can the group produce the best possible overall argument. Thus, groups in PISA represent a means for agents to form coalitions. Another interesting issue is *temporary coalitions* in which a number of participants may attempt to *temporarily agree* with each other for strategic reasons, in order to overcome stronger opponents. In this case, a number of participants could form a *temporary coalition* by which they join forces and cease attacking each other for a limited number of rounds for the purposes of defeating the stronger opponent(s). Once this goal is achieved, say when the stronger opponent(s) drops out of the dialogue game, then the participants in the "temporary coalition" can break up and resume attacking each other as they would have done prior to forming the coalition. Equipping PISA participant agents with a mechanism to form temporary coalitions is seen as a substantial further extension of the PISA Framework, and a number of issues remain to be addressed in future work, if a successful implementation of such coalitions is to be realised.

5. Example

One motivation for group formation in PISA is that agents advocating the same opinion should not attack each other. Another motivation is that the overall performance of PISA could benefit from dividing the same collection of data available to one agent between a reasonable number of agents forming a group arguing for the same opinion as the original agent. The idea is that the operation of the group would allow its members to jointly produce better arguments than would be the case if one agent used all the group's experience. This section illustrates this notion of groups by the means of an example using an artificial data set representing a fictional housing benefit scenario. This scenario concerns a fictional benefit, *Retired Persons Housing Allowance* (RPHA) (as previously used in Bench-Capon (1993) and Mozina, Zabkar, Bench-Capon, and Bratko (2005)), 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 (established as less than £3000). Such persons should also be resident in the country, or absent only by virtue of *service to the nation*, and should have an established connection with the UK labour force. The scenario was interpreted in such a way that four classes could be identified:

(1) (*Fully*) *entitled*: The candidate satisfies all of the above conditions (as in Bench-Capon (1993)).
(2) *Entitled with priority*: The candidate satisfies all of the above and one of the following: (i) has paid contributions in four out of the last 5 years and either has no more than £2000 capital, or the housing costs are substantially more than is needed for entitlement; or (ii) is a member of the armed forces and has paid contributions in each of the last 5 years.
(3) *Partially Entitled*: The candidate satisfies the age condition, and also satisfies one of the following: (i) has paid contributions in four out of the last 5 years and either has no more than £1000 above the original capital limit (£3000), or the housing costs are only slightly

below what is required for entitlement; or (ii) is employed in the Merchant Navy and has
paid contributions in all of the last 5 years.

(4) *Not entitled*: The candidate fails to satisfy any of the above.

In our experiment, we assume four different benefit offices providing RPHA in four different
regions, each with a data set of 1500 benefit records. Each data set is assigned to a PISA participant
agent. Thus a total of four participant agents engage in dialogues regarding the classification
of RPHA claims, each agent defending one of the above four classifications. The support and
confidence thresholds were set to 1% and 50%, respectively (the default thresholds in the data
mining community). Given the specific new case of a female applicant, aged 78 years, who is a
UK resident whose capital and income fall in the right range ($\leq$ £2000 and $\leq$15%, respectively),
and who has paid contributions in 4 out of the last 5 years them, according to the above listed
conditions listed, the case should be classified as entitled with priority.

In one of the executions, the dialogue proceeds as illustrated by the argument tree presented
in Figure 1(a). The diagram features four agents labelled PR (priority entitled), EN (entitled), PE
(partially entitled) and NE (not entitled) who all apply the same *focused build attack when possible*
strategy (S1-1-1). The diagram should be viewed in conjunction with Table 4 which presents the
dialogue, and gives the premises and conclusions of the various rules put forward. The dialogue
commences when the chairperson invites EN to propose the opening argument, EN proposes $R1$.
$R1$ is then attacked by the other three agents in the second round ($R2$, $R3$, and $R4$). At the end
of round 2, PE is ahead as it has the best un-attacked rule. In round three, PE proposes a counter

(a)   (b)
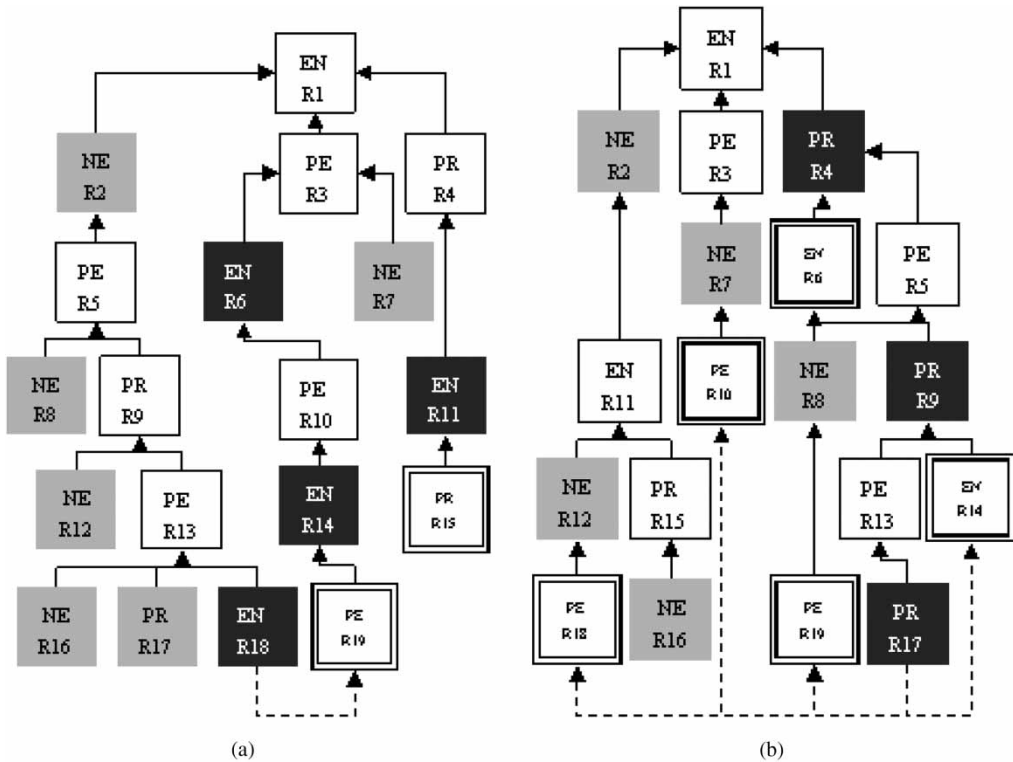
Figure 1. The argumentation tree for the example presented in Section 5: (a) four individual participant
agents; (b) three individual agents and one group (PR).

Table 4. The result- dialogue of the first example (c = confidence).

1   R1: EN – Proposes a new rule: Residency = UK and Contrib Y1 = paid → entitled. c = 50.61%.
2   R2: NE – Distinguishes a previous move. The case has the additional feature: Contrib Y5 = not paid → entitled with c 20.0% only.

       R3: PE – Proposes a counter rule against a move previously played by EN: Contrib Y1 = paid and Contrib Y2 = paid → priority entitled. c = 54.79%.

       R4: PR – Distinguishes a previous move. The case has the additional feature: Capital ≤£2000 → entitled with c = 35.66% only.

3   R5: PE – Proposes a counter rule against a move previously played by NE: Contrib Y1 = paid, Contrib Y2 = paid and Contrib Y3 = paid → partially entitled. c = 62.29%.

       R6: EN – Proposes a new rule against a move previously played by PE: Income ≤15% and Contrib Y2 = paid → entitled. c = 54.83%.

       R7: NE distinguishes PE's argument by pointing out that capital ≤ £2000 and residence = UK → partial benefits with c = 19.24% only.

4   R8: NE – Distinguishes PE's previous move (R5). The case has the additional feature: Residence = UK → partially entitled with c 27.60% only.

       R9: PR – Proposes a new rule against a move previously played by PE: Gender = female. 75 ≤ Age ≤ 80, Income ≤ 15%, Capital ≤ £2000 and Contrib Y2 = paid → priority entitled. c = 74.83%.

       R10: PE – Increases the confidence of a previous rule by stating that the case has additional feature: Gender = female → partially entitled. c = 70.05%.

       R11: EN – Proposes a new rule against a move previously played by PE: Residency = UK, Contrib Y3 = paid and Contrib Y4 = paid → entitled. c = 69.12%.

5   R8: NE – Distinguishes PE's previous move (round 4). The case has the additional features: Residence = UK and Contrib Y5 = not paid → partially entitled with c 23.71% only.

       R13: PE – Proposes a new rule against a move previously played by PR. The case has the following features: Gender = female, Contrib Y1 = paid, Contrib Y2 = paid and Contrib Y3 = paid → partially entitled. c = 76.83%.

       R14: EN – Proposes a new rule against a move previously played by PR. The case has the following features: Gender = female, 75 ≤ Age ≤ 80, Residency = UK and Contrib Y2 = paid → entitled. c = 74.24%.

       R15: PR – Increases the confidence of a previous rule by stating that the case has additional features: Contrib Y3 = paid, Contrib Y4 = paid → entitled. c = 75.58%.

6   R16: NE – Distinguishes PE's argument (R13) by pointing out that the additional features: Residency = UK and Capital ≤£2000 → partial benefits with c = 25.36% only.

       R17: PR – Distinguishes PE's argument (R13) by pointing out that the additional features: Income ≤15% and Capital ≤£2000 → with c = 14.22% only.

       R18: EN – Proposes a new rule against a move previously played by PE: 75 ≤ Age ≤ 80, Residency = UK and Contrib Y5 = Not paid → entitled. c = 83.54%.

       R19: PE – Proposes a new rule against a move previously played by EN: Gender = female, 3 75 ≤ Age ≤ 80, Contrib Y1 = paid, Contrib Y2 = paid, Contrib Y3 = paid, Contrib Y4 = paid → partially entitled. c = 82.69%.

rule against NE's previous rule (*R*5), EN proposes a new rule (*R*6) to attack the current best rule (*R*3) and NE distinguishes PE's argument (*R*7). Thus PE maintains its lead. Note that PR has not contributed to round three.

In round four: NE distinguishes PE's rule from round three (*R*8), PR proposes a counter rule against PE's rule of round three (*R*9), PE increases the confidence of its previous move (*R*3) by playing *R*10 and EN proposes a new rule *R*11. Now PR is winning, but in the fifth round its winning rule (*R*10) can be distinguished by NE (R12). Additionally, both EN and PE have moves in this round (*R*13 and *R*14, respectively). PR however can increase the confidence in its previous move (*R*4) by playing *R*15, but the resulting rule does not have enough confidence to maintain PR's lead. In the sixth round, NE distinguishes PE's argument (*R*13) using *R*16, PR distinguishes *R*13 using *R*17, and EN and PE play new rules (*R*18 and *R*19, respectively). No more arguments are now possible, and so the final classification is that the candidate should be entitled to normal rather than priority benefit, which is the wrong classification.

Let us now assume that the data available to the office arguing that the case should classify as *priority entitled* are divided between four PISA Participant Agents as follows: *P*1 (25%), *P*2 (15%), *P*3 (35%), and *P*4 (25%), and that these agents form a group with P3 being the most experienced agent. Let us also assume that *P*1 and *P*2 apply a build strategy (S1-1-1) and that *P*1

is selected as the GL. $P3$ and $P4$ apply a destroy strategy (S1-1-2). The new dialogue commences in the same manner as the previous one, and EN proposes the same rule $R1$ as before. In round two, the other three agents attack $R1$; EN and PE play the same moves they have used in the previous example ($R2$ and $R3$). PR however attacks using a different rule (Figure 1(b)). Both Leader and $P2$ propose a counter rule against EN's move from round one, while $P3$ and $P4$ suggest distinguishing this move. GL chooses its own move, as it has the highest confidence and consequently plays R4$'^2$:

- $R4'$: PR - Proposes a Counter Rule against a move previously played
  by EN. Gender = female and Capital $\leq$ £2000 → priority entitled.
  c = 53.89%.

In round three, PE, EN, and NE play the same moves as before ($R5'$, $R6$, and $R7$, respectively). However, here PE directs its move against PR's previous move ($R4'$). In round 4, NE and PE play the same moves ($R8$ and $R10$) as before; EN also has the same move ($R11'$), but here it is directed against NE's move from round two ($R2$). In the PR group, participants $P2$ and $P4$ have no moves, the leader suggests a counter move against PE but it cannot be played (has confidence of 61.83% only), and $P3$ suggests distinguishing PE's move from the last round ($R6$). GL, however requests the following from $P3$:

- RequestMove(propose counter rule, confidence >62.29%).

$P3$ succeeds in mining the requested move, and GL puts forward the resulting move ($R9'$):

- $R9'$: PR - proposes a new rule against a move previously played by
  PE. 75 $\leq$ Age $\leq$ 80, income $\leq$15%, capital $\leq$ £2000 and Contrib Y2 = Paid
  → priority entitled. c = 72.13%.

In round five, NE PE and EN play the same moves as before ($R12$, $R13$ and $R14$, respectively). In the group PR, a similar decision to the previous round is taken, and the leader puts forward the move requested from P3:

- $R15'$: PR - proposes a counter rule against a move previously played
  by PE: 75 $\leq$ Age $\leq$ 80, income $\leq$15%, Contrib Y1 = Paid, Contrib Y2 = Paid
  and Contrib Y3 = Paid → priority entitled. c = 77.34%.

Finally, in round six, all participants make moves:

- PE and EN play the same moves as the first example ($R18$ and $R19$, respectively).
- EN distinguishes PR's argument ($R15'$) by pointing out that residency = UK and contribution Y5 = not paid, only gives partial benefits with 33.69% confidence.
- Similarly to the previous rounds, GL(PR) directs its members to play the following move:
  - ($R17'$): PR - proposes a counter rule against a move previously
    played by PE: 75 $\leq$ Age $\leq$ 80, income $\leq$ 15%, Contrib Y1 = Paid, Contrib
    Y2 = Paid, Contrib Y3 = Paid and Contrib Y4 = Paid → priority entit-
    led. c = 84.55%.

No more arguments are possible at this stage, and so the final classification is that the candidate should classify as priority entitled. Note that the right class emerged as a result of PR being a group of four agents, and even though each agent individually had less data than the first example; collectively they were able to generate better arguments.

6. Experimental evaluation

The effectiveness of *Arguing from Experience* has been demonstrated through a number of experiments reported in previous papers (e.g. PADUA was evaluated in Wardeh, Bench-Capon, and Coenen (2008b)). PISA was first evaluated in Wardeh et al. (2009a), in which it was shown that PISA also performed better than, or as well as, its competitors. The particular benefits of Arguing from experience for classification were, however, shown most clearly in Wardeh, Bench-Capon, and Coenen (2009b), where PISA was shown to produce robust results even when the data sets of the included agents were infected with different levels of noise (up to 50%).

In this section, we empirically evaluate the proposed extension to the PISA framework for argumentation-based classification using groups. The evaluation comprises a number of experiments designed to test the hypothesis that argumentation-based classification using groups improves the quality of the resulting predictions compared with other methods such as decision trees or ensemble paradigms. Moreover, we also conjecture that the improved accuracy achieved will increase as the number of agents supporting each possible classification increases. Four categories of experiments were conducted to investigate both the application of the extended version of PISA (with groups) as a classifier and the operation of the group concept in general. The four categories of experiment may be itemised as follows:

- *Effectiveness of groups in argumentation*: Experiments to demonstrate that classification using PISA groups produces more accurate results than when using PISA agents in isolation.
- *Group size*: Experiments to determine the effect of varying the group size.
- *Data size*: Experiments to determine the effect on accuracy according to the amount of data available to individuals in a group.
- *Data balance*: Experiments to investigate the outcome where the data available to the individuals in a group are unbalanced.

These experiments are discussed in detail in the following four subsections.

6.1. *Effectiveness of groups in argumentation*

The first set of experiments applied PISA to a number of real-world data sets taken from the *UCI Repository of machine learning databases* Blake and Merz (1998), selected to cover a wide variety of domains. Two experimental setups were used. In the first, each data set was equally divided among a number of participant agents equal to the number of possible classes in each data set; while in the second, PISA was run using a number of groups (equal to the number of possible classes in each data set), each comprising four participant agents. Figure 2 show the classification accuracies (The percentage of correctly classified cases) of ten cross validation (TCV) tests, applied using PISA with and without groups. These results are compared against those of C4.5 (Quinlan 1993) and Bagging (Brieman 1996). The results indicate that PISA performs consistently well (producing high accuracy with most of the data sets), often outperforming the other two classification paradigms included in the experiment; and giving comparable results to the decision tree methods, which are particularly well suited to data sets such as the nursery data set. The results demonstrate that PISA tends to perform better using groups than individual agents: it seems that the greater the number of separate databases (group size) the greater the number of arguments found, enabling a more thorough exploration of the problem. Note that the only data set where the accuracy of the predictions generated using PISA with groups is worse than all the other paradigms is the relatively small Pima data set, where each participant could be assigned only 96 records.
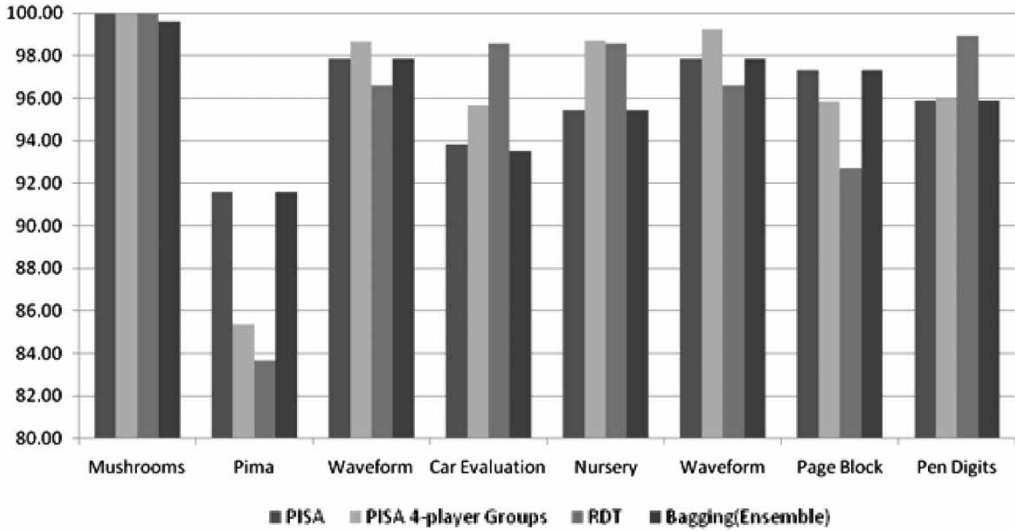
Figure 2. Results of TCV tests using the same number of players per group using UCI repository data sets.

6.2. *Group size*

In the second set of experiments, PISA was applied to a number of variation of the artificial housing benefits data sets (introduced in Section 5). The objective was to investigate the operation of groups within PISA with respect to group size. This experiment assumed that the amount of data available to each PISA group was fixed and equally divided among its members, similar to an ensemble-like approach to classification (Brieman 1996). Thus, if too many players were assigned to a single group, each will be assigned a very small data set and the members would not be able to mine adequate rules. Five sets of TCV tests were carried out using the housing benefits data set. PISA was run using four groups comprising (respectively) 2, 4, 6, 8, and 10 (participant) agents each, each group corresponding to one of the possible four classes in the data set. In each test, the original data set was divided equally among the players. Figure 3(a) presents the results of these tests, which indicate that, in general, PISA operates better when using groups of players.

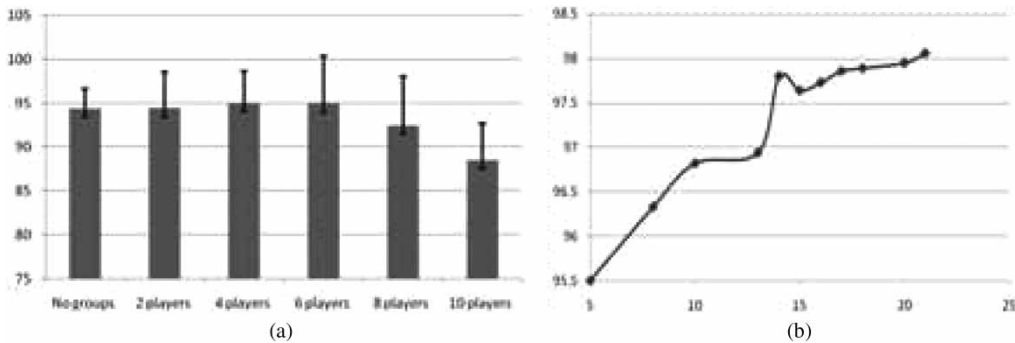(a)   (b)

Figure 3. Relation between number of players and PISA accuracy when players are given fixed amount of data. (a) Results of the TCV tests using the same number of agents per group. Error bars represent the standard deviation. (b) Results of the TCV tests using same number of players per group. Error bars represent the standard deviation.

More than one agent advocating the same classification, using the notion of groups, seems to have positive effects on the accuracy of the resulting dialogues, due to the fact that a range of arguments can be mined from the different data sets, each presenting different experience. Thus more options will be available to the group as a whole, from which the group's leader can select the best course of action. However, the increase in accuracy is relative to the amount of data given to each agent in the dialogues. If the data is not sufficient to mine adequate ARs, the operation of PISA will not benefit from dividing the data any further. Thus, for a fixed amount of data, performance begins to degrade if there are too many players in the group (i.e. the data available to each individual gets too small).

6.3. *Data size*

The Third experiment was conducted to investigate the relationship between the amount of data available to the group as a whole and the performance of the group. For this purpose, a number of Housing Benefits data sets, each comprising 1000 records, were generated. PISA was run with four groups, each comprising the same number of players. But here, each player was given one of the generated data sets. Thus, the more players that join a group, the more experience available to the group as a whole. A number of TCV tests were conducted. For each test, the number of players in each group was randomly generated. Figure 3(b) shows the result of these tests. Note that the overall accuracy benefits from the increase in the size of the groups. For instance, the average accuracy when each group comprised five players was 95.5%. However as more members were added, the accuracy increased to reach 98.06%.

6.4. *Unbalanced groups*

The fourth set of experiments (again using TCV) were directed at determining the effects of unbalanced groups on accuracy. For example, a big group might win when it should lose, while a small group might loose when it should win. For these purposes, the four groups in the previous set of experiments were assigned random numbers of players. Each player was given a similar-sized housing benefits data set, as above. Four random players' allocations were generated, and the TCV test was repeated for each allocation (Table 5). The results of these tests suggest that the two effects highlighted above hold. In order to clarify this point, supplementary information was generated with respect to the winning groups for each dialogue in the four TCV tests. The results demonstrated that in TCV1, group G4 (with only one member) has failed to win 30.16% of the cases that should classify according to its advocated class. Also, 94.25% of these cases were won by G1 (the group with the highest number of players). Similar results were reported with the other TCV tests. Table 6 illustrates, for each group, the percentage of cases that this group has failed to classify correctly. In every case, the smallest group misses the most opportunities and the largest group fewest.

Table 5. Four random allocations for random number of participants per group.

Participants	TCV1	TCV2	TCV3	TCV4
G1	8	5	1	3
G2	4	8	7	8
G3	4	1	4	2
G4	1	3	5	3
PISA accuracy	90.21%	91.73%	93.91%	94.38%

Table 6. Percentage of misclassified cases that should have been won by each group from Table 5.

Test	TCV1	TCV2	TCV3	TCV4
G1	0.00%	2.81%	21.05%	4.90%
G2	1.74%	0.00%	0.31%	0.00%
G3	3.26%	29.08%	1.90%	10.90%
G4	30.16%	1.19%	1.10%	6.68%

6.5. *Evaluation summary*

Our analysis suggests that PISA can be profitably exploited as a valuable means of classification. The advantageous features of PISA with groups, identified in the foregoing, may be itemised as follows:

- By using groups, better (more accurate) results are obtained than when agents are used in isolation.
- PISA can be used with data sets of any size. However, it is advised to use PISA with moderately large sets so that the individual players have a reasonably sized "experience (case) base" to argue from. Very large data sets should be split and groups used. Very small data sets perform less well, and do not afford the advantages that can come from groups.
- By dividing one data set among a reasonable number of agents, distributed over a number of groups, PISA can act as an ensemble technique whose performance is compatible with other well-known ensemble methods.
- PISA has a wide range of configuration parameters, such as the number of players in each group and the strategy setup for each player; by changing some of these parameters, the course of PISA dialogues can be modified to better fit the characteristics and size of the underlying data set (i.e. nature of the application).

A further, general advantage of PISA is that it does not require a training phase, unlike other similar systems. Furthermore, PISA produces a limited number of classification ARs sufficient to classify the given case without the need to generate all the possible rules.

7. Related work

Concerning the automatic generation of arguments from collections of previous examples, the work in Governatori and Stranieri (2001) investigates the feasibility of knowledge discovery and data mining in order to facilitate the discovery of defeasible rules for legal decision-making. Arguing from experience presents an approach to argumentation related to that of Governatori and Stranieri (2001) and bridges the gaps in their proposal (e.g. their technique can operate only on small data sets). More importantly, PISA offers a more efficient means to exploit databases for the production of arguments.

In Ontañon and Plaza (2008), an argumentation framework for learning agents is articulated. This framework has similarities with the one proposed here in that it takes the experience, in the form of past cases, of agents into consideration. However, the work in Ontañon and Plaza (2008) uses Case-Based reasoning techniques to generate arguments from past examples and provides a means to define the preference relation between the generated arguments. PISA, on the other hand, implements a more straightforward ARM techniques to produce arguments and applies a preference relation based upon the support/confidence measures. Ontañon and Plaza (2008) also presented a framework for multiparty argumentation to enable a committee of agents to jointly

deliberate about given cases. However, unlike PISA, the communication between the arguing agents is direct (there is no mediator agent).

An earlier example of un-mediated multi party argumentation can be found in Tambe and Jung (1999)), where turn taking is tokenised. The protocol consists of a series of rounds. In the initial round, the agents state their individual predictions of the discussed case, and broadcast these predictions to the involved agents. Then, in the subsequent rounds, a token passing mechanism enables agents (one at a time) to attack other agents (also one at a time) if they disagree with their prediction. When an agent receives an attack, it informs the attacker whether it accepts the counterargument (and changes its prediction) or not. Also, when an agent has the token, it can answer to such attacks by generating the counter attacks. When all the agents have had the token once, the token returns to the first agent, and so on. The communication between the agents continues in the same manner until they all agree on a prediction, or if a given number of rounds has passed and no agent has generated any counterargument. Moreover, if – at the end of the argumentation – the agents have not reached an agreement, then a voting mechanism that uses the confidence of each prediction as weights is used to decide the final solution. PISA differs in that it relies on an argumentation artefact (the argumentation tree), and mediator agent (the chairperson) to facilitate the argumentation process between a number of agents. Thus, PISA agents focus on generating the best arguments rather than on the communication among themselves. Additionally, instead of voting, PISA applies a tie-resolution mechanism where agreement cannot be reached in a given number of rounds.

Some work has also been done regarding *n*-person argumentation games. Pham, Thakur, and Governatori (2008) presents a defeasible logic approach to address situations requiring agents to settle on a common goal despite the fact that their agendas may contain conflicting goals. This group of agents applies the majority rule to identify the "*most common*" claim. This approach is argued to simplify the complexity of *n*-person argumentation games into two-group games: one supports the major claim, the other opposing it. PISA, on the other hand, addresses situations where there is no "*major claim*", but rather each participant has its own claim, and the situation requires consideration of each of these claims as legitimate stand-alone claims.

Teamwork has been the focus of much research in the fields of distributed AI and multi agent systems. Horling and Lesser (2005) identifies an agent team as consisting of a number of cooperative agents which have agreed to work together towards a common goal. Additionally, teams attempt to maximise the utility of the team (goal) itself, rather than that of the individual members. Thus, the notion of groups in PISA can be likened to teamwork: the group members share the common goal of establishing their mutual point of view. Within a team, each agent will take on one or more roles as needed to address the subtasks required by the team goal. The roles may change over time in response to planned or unplanned events, while the high-level goal itself usually remains relatively consistent. In PISA groups, there are only two roles: team leader and regular member. The leader role can shift to new members if they have a better strategy or more experience than the current leader. Thus it is a dynamic role. A similar notion of leadership can be found in literature (Klusch and Gerber 2001; Bai and Zhang 2005) in which the leading agent acts as a representative and intermediary for the group as a whole.

In Dignum, DuninKeplicz, and Verbrugge (2001), a four-stage formal model of teamwork is presented. The first stage is potential recognition in which the agent that takes the initiative tries to find out which agents are potential candidates for achieving a given overall goal and how these can be combined in a team. The second stage is team formation to construct the team that will try to achieve the goal. The third stage is plan formation. Here the team divides the goal into subtasks, associates these with actions and allocates these actions to team members. The last stage is plan execution in which the team members execute the allocated actions and monitor the appropriate colleagues. The first stage is straightforward in PISA, because the chairperson assign

participants to groups (teams) according to their goals. Thus, the agents need not to carry these tasks themselves. Additionally, group formation in PISA is a clear-cut process when compared with the work on team formation in the literature (a survey of which can be found in Horling and Lesser (2005)). This is mainly because PISA groups members aim at achieving one task: mining the best possible argument in the context of the ongoing dialogue, according to their strategy and experience. The in-group dialogue process presented in this paper falls under the last two stages. Again the task in PISA is relatively simple and homogeneous: the group members attempt to generate the best possible argument, from their individual sets of experience, and then apply this argument in the best way in the current dialogue, whereas teams more generally are seen as undertaking much more complex and varied tasks, and may even require different abilities from different team members.

Finally, some work on the use of argumentation to facilitate classification can be found in the literature. Amgoud and Serrurier (2008) propose a formal argumentation framework for classification, in which arguments can be constructed for and against each possible classification of a given example. Similar to PISA, these arguments can then be evaluated and a plausible classification of the example suggested. However, the work in Amgoud and Serrurier (2008) assumes that a set of hypothesis is associated with the learning examples, whereas PISA has no such assumption. Additionally, the strength of each argument is identified such that the arguments coming from the set of training examples are stronger than arguments built from the set of hypotheses, whereas PISA applies *support/confidence* measures (Agrawal et al. 1993) to evaluate the strength of the generated arguments. Moreover, the work in Amgoud and Serrurier (2008) only focuses on the single agent situation.

Other work has tried to improve the performance of classification methods by combining them with argumentation techniques. In Mozina et al. (2005), the idea of augmented examples is introduced to improve the results of one machine learning technique CN2 (Clark and Niblett 1989). However, their approach requires consulting an expert, who decides on the right class labels for misclassified cases. The expert also generates the arguments associated with such cases, so as to drive a refinement process. Their approach has a single learning agent, and a human expert to explain misclassifications. Additionally, their argumentation process involves two options only, while PISA can handle any number of options. In Ontañon and Plaza (2010), the concept of learning from communication was introduced to show that the accuracy of the learning process can benefit from communication (via argumentation) among a number of agents.

8. Conclusions and future directions

We have provided a mechanism by which any number of agents can participate in a dialogue to decide how a case should be classified using arguments mined (*on the fly*) from their individual data sets. The agents form groups of agents supporting each possible classification and these groups collaborate to decide on the best next move. Using groups can improve the quality of the classification, even when the data are simply divided among different agents (i.e. no more additional data made available) over the situation where each classification was supported by a single agent. Our experiments suggest that for any given problem there is an optimal size for the agent's data sets, and allowing the available data to be divided among agents in this way ensures that the available data can be deployed to the best effect, whatever be its size. Further experiments show that more data improve the quality of classification provided that the data can be deployed in optimally sized chunks. In future work, it would be interesting to explore dynamic groups, whereby agents that become convinced of a different classification can join the appropriate group

rather than leaving the game. Also we intend to explore the effect of coalitions whereby two groups can temporarily join forces against a currently stronger adversary.

Notes

1. Note that we use the strategy not in the special sense peculiar to the game theory, but rather as a heuristic guide to move selection, as in ordinary talk of games where, for example, a chess player adopts an aggressive or defensive strategy.
2. Primed rules (e.g. $R4'$) represent rules (or moves) that are different in the second dialogue. For instance, $R4'$ is different from $R4$, whereas, $R1$, $R2$ and $R3$ are identical in both dialogues

References

Agrawal, R., Imielinski, T., and Swami, A. (1993), 'Mining Association Rules Between Sets of Items in Large Databases', in *Proc. ACM SIGMOD Conf. on Management of Data* (*SIGMOD'93*), ACM Press, DC, USA, pp. 207–216.

Aleven, V. (1997), 'Teaching Case Based Argumentation Through an Example and Models', PhD thesis, University of Pittsburgh, USAW, Pittsburgh, PA.

Ambroszkiewicz, S., Matyja, O., and Penczek, W. (1998), 'Team Formation by Self-interested Mobile Agents', in *Multi-agent Systems, Lecture Notes in Computer Science*, Springer, Brisbane, Queensland, Australia, pp. 1–15.

Amgoud, L., and Maudet, N. (2002), 'Strategical Considerations for Argumentative Agents (Preliminary Report)', in *Proc. 9th Int. Workshop on Non-Monotonic Reasoning* (*NMR'02*), France: Toulouse, pp. 409–417.

Amgoud, L., and Serrurier, M. (2008), 'Arguing and Explaining Classifications', in E. H. Durfee, M. Yokoo, M. N. Huhns and O. Shehory (Eds.), *Argumentation in Multi-agent Systems*, Springer LNCS (Vol. 4946), Heidelberg: Springer, pp. 164–177.

Ashley, K.D. (1990), *Modelling Legal Argument*, Cambridge, MA: MIT Press.

Bai, Q., and Zhang, M. (2005), 'Dynamic Team Forming in Self-interested Multi-agent Systems', in *AI 2005: Advances in Artificial Intelligence*, Springer, pp. 674–683.

Bel-Enguix, G., and López, D.J. (2006), 'Membranes as Multi-agent Systems: An Application to Dialogue Modelling', in *Professional Practice in AI*, Springer, pp. 31–40.

Bench-Capon, T.J.M. (1994), 'Neural Nets and Open Texture', in *Proc. 4th Int. Conf. on AI and Law* (*ICAIL'94*), Amsterdam: ACM Press, pp. 292–297.

Blake, C.L., and Merz., C.J., (1998), *UCI Repository of Machine Learning Databases*, Irvine, CA: University of California, Department of Information and Computer Science. *http://www.ics.uci.edu/mlearn/MLRepository.html*.

Brieman, L., (1996). 'Bagging Predictors', in *Machine Learning* (Vol. 24), Springer, pp. 123–140.

Clark, P., and Niblett, T. (1989), 'The CN2 Induction Algorithm', in *Machine Learning* (Vol.3, 4th ed.), Springer, pp. 261–283.

Coenen, F., Leng, P.H., and Ahmed, S. (2004), 'Data Structure for Association Rule Mining: T-trees and p-Trees', in *IEEE Journal of Transactions in Knowledge and Data Engineering* (Vol. 16, 6th ed.), pp. 774–778.

Dignum, F., DuninKeplicz, B., and Verbrugge, R. (2001), 'Agent Theory for Team Formation by Dialogue', in *ATAL '00 Proceedings of the 7th International Workshop on Intelligent Agents VII. Agent Theories Architectures and Languages*, Springer, pp. 141–156.

Governatori, G., and Stranieri, A. (2001), 'Towards the Application of Association Rules for Defeasible Rules Discovery', in *Proc. 14th Annual Conf. on Legal Knowledge and Information Systems* (*JURIX'01*), Amsterdam: IOS Press, pp. 63–75.

Horling, B., and Lesser, V. (2005), 'A Survey of Multi-agent Organizational Paradigms', in *The Knowledge Engineering Review* (Vol. 19, 4th ed.), Cambridge University Press, pp. 281–316.

Klusch, M., and Gerber., A. (2001), 'Dynamic Coalition Formation among Rational Agents', in *Intelligent Systems* (Vol. 17, 3rd ed.), IEEE, pp. 42–47.

McBurney, P., and Parsons, S. (2002), 'Games that Agents Play: A Formal Framework for Dialogues Between Autonomous Agents', in *Logic, Language and Information* (Vol. 11, 3rd ed.), pp. 315–334.

Moore, D. (1993), 'Dialogue Game Theory for Intelligent Tutoring Systems', PhD thesis, Leeds Metropolitan University.

Mozina, M., Zabkar, J., Bench-Capon, T., and Bratko, I. (2005), 'Argument Based Machine Learning Applied to Law', in *Artif. Intell.* (Vol. 12, 1st ed.), Springer, pp. 53–73.

Oliva, E., Viroli, M., Omicini, A., and McBurney, P. (2008), 'Argumentation and Artifact for Dialogue Support', in *Proc. 5th Int. Workshop on Argumentation in Multiagent Systems* (*ArgMAS'08*), Lisbon, Portugal, pp. 24–39.

Oliva, E., McBurney, P., and Omicini, A. (2008b), 'Co-argumentation Artifact for Agent Societies', in *Argumentation in Multi-Agent Systems*, Springer, pp. 31–46.

Ontañon, S., and Plaza, E. (2008), 'An Argumentation-based Framework for Deliberation in Multi-agent Systems', in *Argumentation in Multi-Agent Systems*, Springer, pp.178–196.

Ontañon, S., and Plaza, E. (2010), 'Learning, Information Exchange, and Joint-deliberation Through Argumentation in Multi-agent Systems', in *On the Move to Meaningful Internet Systems: OTM 2008 Workshops*, Springer, pp. 150–159.

Pham, D.H., Thakur, S., and Governatori, G. (2008), 'Settling on the Group's Goals: An *n*-Person Argumentation Game Approach', in *Proc. 11th Pacific Rim International Conference on Multi-Agents* (*PRIMA'08*), Springer, pp. 328–339.

Prakken, H. (2006), 'Formal Systems for Persuasion Dialogue', in *The Knowledge Engineering Review* (Vol. 21), Cambridge University Press, pp. 163–188.

Quinlan, J.R. (1993), 'Combining Instance-based and Model-based Learning', in *Proc. 10th Int. Conf. on Machine Learning*, Amherst, MA: Morgan Kaufmann, pp. 236–243.

Tambe, M., and Jung, H. (1999), 'The Benefits of Arguing in a Team', in *AI Magazine* (Vol. 20, 4th ed.), AAAI, pp. 85–92.

Wardeh, M., Bench-Capon, T., and Coenen, F.P. (2007), 'Dynamic Rule Mining for Argumentation Based Systems', in *Proc. 27th SGAI Int. Conf. on AI* (*AI'07*), Cambridge: Springer, pp. 65–78.

Wardeh, M., Bench-Capon, T., and Coenen, F.P. (2008a), 'Arguments from Experience: The PADUA Protocol', in *Proc. 2nd Conf. on Computational Models of Argument* (*COMMA'08*), Toulouse, France: IOS press, pp. 405–416.

Wardeh, M., Bench-Capon, T., and Coenen, F.P. (2008b), 'Argument Based Moderation of Benefit Assessment', in *Legal Knowledge and Information Systems. Frontiers in Artificial Intelligence and Applications*, eds. E. Francesconi, G. Sartor, and D. Tiscornia (Vol. 189), Amsterdam, The Netherlands: IOS Press, pp. 128–137.

Wardeh, M., Bench-Capon, T., and Coenen, F.P. (2009a), 'Multi-party Argument from Experience', in *Proc. 6th Int. Workshop on Argumentation in Multiagent Systems* (*ArgMAS'09*), Hungary: Budapest.

Wardeh, M., Bench-Capon, T., and Coenen, F.P. (2009b), 'An Arguing from Experience Approach to Classifying Noisy Data', *Proc. 11th Int. Conf. on Data Warehousing and Knowledge Discovery* (*DaWaK'09*), Springer LNCS 5691, ISSN 0302-9743, pp. 354–365.