

Empirical Argumentation: Integrating Induction and Argumentation in MAS

Santiago Ontañón and Enric Plaza

IIIA, Artificial Intelligence Research Institute
CSIC, Spanish Council for Scientific Research
Campus UAB, 08193 Bellaterra, Catalonia (Spain).
`santi|enric@iia.csic.es`

Abstract. This paper presents an approach that integrates notions and techniques from two distinct fields of study —namely inductive learning and argumentation in multiagent systems (MAS). We will first discuss inductive learning and the role argumentation may play in multiagent inductive learning, and then how inductive learning can be used to realize an argumentation in MAS based on empirical grounds. We present a MAS framework for empirical argumentation and then we show how this is applied to a particular task where two agents argue in order to reach agreement on a particular topic. Finally, an experimental evaluation of the approach is presented evaluating the quality of the agreements achieved by argumentation.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence — Multiagent systems, Intelligent Agents. I.2.6 [**Artificial Intelligence**]: Learning.

General Terms

Algorithms, Experimentation, Theory.

Keywords

Argumentation, Learning.

1 Introduction

This paper presents an approach that integrates notions and techniques from two distinct fields of study —namely inductive learning and argumentation in multiagent systems (MAS). We will discuss first inductive learning and the role argumentation may play in multiagent inductive learning, and later how inductive learning can be used to realize an argumentation in MAS based on empirical grounds.

Multiagent inductive learning (MAIL) is the study of multiagent systems where individual agents have the ability to perform inductive learning, i.e. an agent is able to learn general descriptions from particular examples. Therefore, induction is a form of empirical-based inference, where what is true (or what is believed by the agent) is derived from the experience of that agent in a particular domain (such experience is usually represented with “cases” or “examples”). Notice that inductive inference is not deductive, and specifically it is not truth-preserving¹, and therefore it captures a form of empirical knowledge that can be called into question by new empirical data and thus needs to be revised.

The challenge of multiagent inductive learning is that several agents will inductively infer empirical knowledge that in principle is not the same, since that knowledge is dependent on each individual in two ways: the concrete empirical data an agent has encountered and the specific inductive method an agent employs.

Therefore, empirical knowledge will be different in principle for each individual agent, and based on a part of the empirical data (the one observed by each individual agent). Communication among agents is necessary in order to reach a shared and agreed-upon empirical-based generalization that is based on, and consistent with, all the empirical data available to a collection of agents. Instead of simply using a communication process that redistributes all data to all agents, we propose an argumentation-based communication process where agents can propose, compare and challenge the empirical knowledge of other agents, with the goal of achieving a more accurate, shared, and agreed-upon body of empirical knowledge.

From the point of view of argumentation in MAS, inductive learning provides a basis for automating, in empirical domains, a collection of activities necessary for implementing artificial agents that support argumentation: how to generate arguments, how to attack and defend arguments, and how to change an agent’s beliefs as a result of the arguments exchanged. Logic-based approaches to argumentation like DeLP [8] amend classical deductive logic to support defeasible reasoning. Our approach takes a different path, assuming agents that *learn their knowledge* (by using induction over empirical data) instead of assuming agents have been *programmed* (by giving them a rule-based knowledge base). Therefore, we need to specify empirical methods that are able to perform the required activities of argumentation (generating arguments and attacks, comparing arguments and revising an agent’s beliefs): this is the approach we will call *empirical argumentation* for MAS.

This paper presents a MAS framework for empirical argumentation that proposes a way to implement those activities on the basis of the inductive inference techniques developed in the field of Machine Learning. The structure of the paper starts by introducing the notions of inductive learning needed (Section 2). Then Section 3 presents the MAS framework for empirical argumentation called A-MAIL, while Section 4 shows the utility of the framework in the task of concept

¹ Inductive inference is not truth-preserving since new experiences may contradict past generalizations, albeit it is falsity-preserving.

convergence (in which two agents argue with the goal of achieving an agreement on a particular topic); an experimental evaluation of the approach is presented evaluating the quality of the agreements achieved by argumentation. The paper closes with sections on related work and conclusions.

2 Concept Induction

Inductive learning, and in particular concept learning, is the process by which given an *extensional definition* of a concept C (a collection of examples of C and a collection of examples that are not C) then an *intensional definition* (or generalization) of a concept C can be found. Formally, an *induction domain* is characterized as pair $\langle \mathcal{E}, \mathcal{G} \rangle$ where \mathcal{E} is the language describing examples or instances and \mathcal{G} is the language for describing generalizations; usually $\mathcal{E} \subset \mathcal{G}$ is assumed, but this is not necessary. A language is understood as the set of well formed formulas built from a domain vocabulary or ontology \mathcal{O} . The relation between languages \mathcal{E} and \mathcal{G} is established by the subsumption relation (\sqsubseteq); we say a generalization $g \in \mathcal{G}$ subsumes (or covers) an example $e \in \mathcal{E}$, $g \sqsubseteq e$, whenever e satisfies the properties described by g [6]. Different approaches to induction work with different languages, from propositional languages (attribute value vectors) to subsets of predicate logic (like Inductive Logic Programming that uses a sublanguage of Horn logic).

Given a collection of examples $E = \{e_1, \dots, e_M\}$ described in a language \mathcal{E} , an extensional definition of a concept C is a function $C : E \rightarrow \{+, -\}$, i.e. determines the subset E^+ of (positive) examples of C , and the subset E^- of counterexamples (or negative examples) for C . Induction is a function $I : \mathcal{P}(E) \times C \rightarrow \mathcal{G}$, which given a collection of examples and a target concept yields an intensional definition $C \in \mathcal{G}$; generally one single formula in \mathcal{G} is not sufficient to describe an intensional definition so it is usually described as a disjunction of generalizations $C = h_1 \vee \dots \vee h_n$.

Definition 1. *An intensional definition C of a concept C is a disjunct $C = h_1 \vee \dots \vee h_n$ (where $h_i \in \mathcal{G}$) such that its generalizations subsume (\sqsubseteq) all positive examples of C and no counterexample of C :*

$$\forall e_j \in E^+ : \exists h_i : h_i \sqsubseteq e_j \quad \wedge \quad \forall e_j \in E^- \wedge \forall h_i : h_i \not\sqsubseteq e_j$$

For simplicity, we will shorten the previous expression as follows: $C \sqsubseteq E^+ \wedge C \not\sqsubseteq E^-$.

As an exemplification of these notions, consider the case where C is the concept *Chair*; in this scenario, the set E may consists of chairs, benches, stools, tables and other furniture, and each specific positive example is a description of one concrete chair and each specific negative example is a description of one furniture item that is not a chair. Finally, and intensional description *Chair* of concept *Chair* might be “Chair is an object with a seat, four legs and a back.”

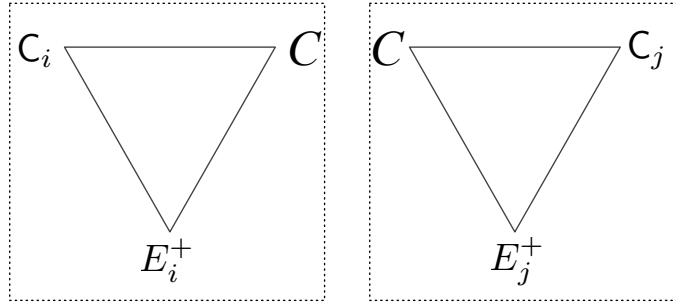


Fig. 1. Schema for two agents where a concept name (C) is shared while intensional descriptions are, in general, not equivalent ($C_i \not\equiv C_i$).

2.1 Inductive agents with empirical beliefs

Since we will focus on argumentation between two agents (say A_1 and A_2), and each agent will have certain beliefs according to what they have learnt, we will now explore how differences between these two agents relate to induction and argumentation. First, we will assume each agent has its own set of examples from which they may learn by induction (say E_1 and E_2) and they are both in principle unrelated although expressed in the same language \mathcal{E} . Furthermore, each agent may use, in principle, different induction techniques but they obtain generalizations in the same language \mathcal{G} . Thus, for any particular concept C two agents will have intensional descriptions C_1 and C_2 that are, in general, not equal or equivalent. Fig. 1 depicts these relationships between two agents beliefs about what C is based on their empirical data E_1 and E_2 .

Finally, since Def. 1 is too restrictive for practical purposes, Machine Learning approaches allow the intensional definitions to subsume less than 100% of positive examples by defining a confidence measure. The goal of induction is, given as a target the function $C : E \rightarrow \{+, -\}$, to find a new function C , which is a good approximation of C , in the sense of yielding a small error in determining when an example is a positive or negative example of C .

We will use the following confidence measure:

Definition 2. *The individual confidence of a hypothesis h for an agent A_i :*

$$B_i(h) = \frac{|\{e \in E_i^+ | h \sqsubseteq e\}| + 1}{|\{e \in E_i | h \sqsubseteq e\}| + 2}$$

$B_i(h)$ is the ratio of examples correctly covered by h over the total number examples covered by h ; moreover, we add 1 to the numerator and 2 to the denominator following the Laplace probability estimation procedure (which prevents estimations too close to 0 or 1 when very few examples are covered). Other confidence measures could be used, our framework only requires that the confidence measure reflects how much the set of examples known to an agent endorses

the hypothesis. Finally, a threshold τ is established and only hypotheses with confidence $B_i(h) > \tau$ are accepted as valid for the inductive process.

Definition 3. *A hypothesis h is τ -admissible for an agent A_i if $B_i(h) \geq \tau$, where $0 \leq \tau \leq 1$.*

3 An Empirical Approach to MAS Argumentation

This section will focus on how to integrate argumentation with inductive agents in scenarios where the goal is to achieve an agreement between two agents on the basis of their empirical knowledge. Here the *empirical* adjective refers to the observations of the real world that each agent has had access to and that is embodied in the set of examples E_1 and E_2 represented using a language \mathcal{E} .

Argumentation in Multiagent Inductive Learning (A-MAIL) is a framework where argumentation is used as a communication mechanism for agents that want to perform collaborative inductive tasks such as concept convergence. We do not claim, however, that A-MAIL is a new “argumentation framework” in the sense of Dung [4], it is intended as a framework to integrate argumentation processes and inductive processes in MAS. According to Dung, an argumentation framework $AF = \langle A, R \rangle$ is composed by a set of arguments A and an attack relation R among the arguments. A-MAIL is not a general logic framework and, although certainly we will define what we mean as arguments and attack relations, we take an empirical approach to argumentation. Thus, the main difference from Dung’s framework is that, since arguments are generated from examples, our approach necessarily defines a specific relation between arguments and examples, which is not part of the usual interpretations of Dung’s framework².

3.1 The A-MAIL Approach

Let us define both the kinds of arguments considered by A-MAIL and how arguments attack each other.

There are two kinds of arguments in A-MAIL:

Example Argument: $\alpha = \langle e, \overline{C} \rangle$ is a pair where an example $\alpha.e \in \mathcal{E}$ is related to a concept $\alpha.\overline{C} \in \{C, \neg C\}$, either endorsing C if $\alpha.e$ is a positive example or endorsing $\neg C$ if $\alpha.e$ is a counter-example of C .

Hypothesis Argument: $\alpha = \langle h, \overline{C} \rangle$ is a pair where $\alpha.h$ is a τ -admissible hypothesis and $\alpha.\overline{C} \in \{C, \neg C\}$. An argument $\langle h, C \rangle$ states that $\alpha.h$ is a hypothesis of C , while $\langle h, \neg C \rangle$ states that h is a hypothesis of $\neg C$, i.e. that examples covered by $\alpha.h$ do not belong to C .

² Some approaches may consider “counter-examples” as a kind of arguments. This is certainly true, but in our approach there is a constitutive relation between examples and arguments (the “empirical” approach) that is different from merely accepting counter-examples as arguments.

Since hypotheses in arguments are generated by induction, they have an associated degree of confidence for an individual agent:

Definition 4. *The confidence of a hypothesis argument α for an agent A_i is:*

$$B_i(\alpha) = \begin{cases} \frac{|\{e \in E_i^+ | \alpha.h \sqsubseteq e\}| + 1}{|\{e \in E_i | \alpha.h \sqsubseteq e\}| + 2} & \text{if } \alpha.\bar{C} = C \\ \frac{|\{e \in E_i^- | \alpha.h \sqsubseteq e\}| + 1}{|\{e \in E_i | \alpha.h \sqsubseteq e\}| + 2} & \text{if } \alpha.\bar{C} = \neg C \end{cases}$$

Consequently, we can use the threshold τ to impose that only arguments with a strong confidence are admissible in the argumentation process; however, we require only a strong confidence on the part of the individual agent uttering such an argument.

Definition 5. *An argument α generated by an agent A_i is τ -admissible iff it is a hypothesis argument and $B_i(\alpha) > \tau$, or if it is an example argument.*

From now on, only τ -admissible arguments will be considered within the A-MAIL framework.

Now we can define the attack relation between arguments; we will use the notation $\bar{C} \in \{C, \neg C\}$ to specify the boolean valuation of a concept as true or false, while $\hat{C} = \neg \bar{C}$ is the dual boolean valuation.

Definition 6. *The attack relation ($\alpha \rightarrow \beta$) between two τ -admissible arguments α, β holds when:*

1. $\langle h_1, \hat{C} \rangle \rightarrow \langle h_2, \bar{C} \rangle \iff \hat{C} = \neg \bar{C} \wedge h_2 \sqsubseteq h_1$, or
2. $\langle e, \bar{C} \rangle \rightarrow \langle h, \hat{C} \rangle \iff e.\bar{C} = \neg h.\hat{C} \wedge h \sqsubseteq e$

where $\bar{C} \in \{C, \neg C\}$.

Notice that a hypothesis argument α only attacks another argument β if $\beta.h \sqsubseteq \alpha.h$, i.e. when α is (strictly) more specific than β . This is required since it implies that all the examples covered by α are also covered by β , and thus if one supports C and the other $\neg C$, they must be in conflict.

Figure 2 shows some examples of arguments and attacks. Positive examples of the concept C are marked with a positive sign, whereas negative examples are marked with a negative sign. Hypothesis arguments are represented as triangles covering examples; when an argument α_1 subsumes another argument α_2 , we draw α_2 inside of the triangle representing α_1 . Argument α_1 has a hypothesis supporting C , which covers 3 positive examples and 2 negative examples, and thus has confidence 0.57, while argument α_2 has a hypothesis supporting $\neg C$ with confidence 0.60, since it covers 2 negative examples and only one positive example. Now, $\alpha_2 \rightarrow \alpha_1$ because α_2 supports $\neg C$, α_1 supports C and $\alpha_1.h \sqsubseteq \alpha_2.h$. Moreover, $\varepsilon_3 \rightarrow \alpha_2$, since e_3 is a positive example of C while α_2 supports $\neg C$ and covers this example ($\alpha_2.h \sqsubseteq \varepsilon_3.e$).

Notice that the viewpoint on the (empirical) admissibility of an argument or of an attack depends on each individual agent, as shown in Fig 3 where

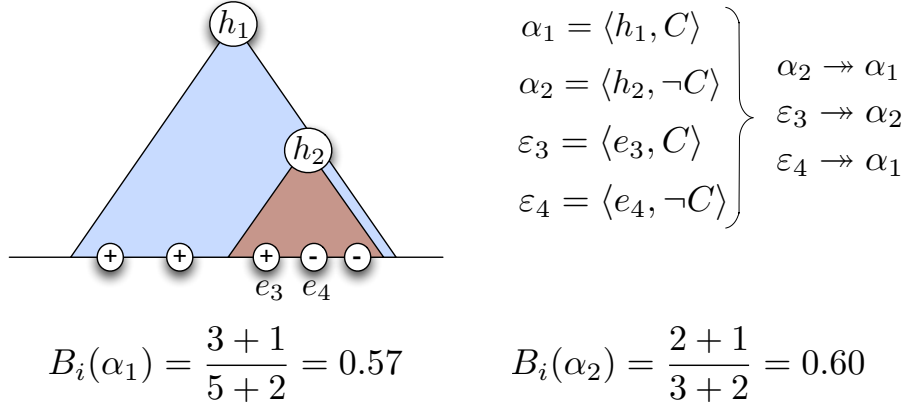


Fig. 2. An illustration of the different argument types, their confidences and attacks.

two agents A_i and A_j compare arguments α_1 and α_2 for hypotheses h_1 and h_2 , assuming $\tau = 0.75$. From the point of view of agent A_i (the Opponent), proposing argument α_2 as an attack against argument α_1 of agent A_j (the Proponent) is a sound decision, since for A_i , α_1 is not τ -admissible, while α_2 is. However, from the point of view of the Proponent of α_1 , α_2 is not τ -admissible. Thus, A_j does not accept α_2 and will proceed by attacking it; on the other side, had it found that confidence $B_j(\alpha_2)$ was higher than $B_j(\alpha_1)$ agent A_j would have revised its beliefs by dismissing α_1 and accepting α_2 , as we will explain later.

Next we will define when arguments *defeat* other arguments, based on the idea of argumentation lines [8].

Definition 7. An Argumentation Line $\alpha_n \rightarrow \alpha_{n-1} \rightarrow \dots \rightarrow \alpha_1$ is a sequence of τ -admissible arguments where α_i attacks α_{i-1} and α_1 is called the root.

Notice that odd-numbered arguments are generated by the agent whose hypothesis is under attack (the Proponent of the root argument α_1) and the even-numbered arguments are generated by the Opponent agent attacking α_1 . Moreover, since hypothesis arguments can only attack other hypothesis arguments, and example arguments can only attack hypothesis arguments, example arguments can only appear as the left-most argument (e.g. α_n) in an argumentation line.

Definition 8. An α -rooted argumentation tree T is a tree where each path from the root node α to one of the leaves constitutes an argumentation line rooted on α . The example-free argumentation tree T^f corresponding to T is a tree rooted in α that contains the same hypothesis arguments of T and no example argument.

Therefore, a set of argumentation lines rooted in the same argument α_1 can be represented as an argumentation tree, and vice versa. Notice that example arguments can only appear as leafs in any argumentation tree.

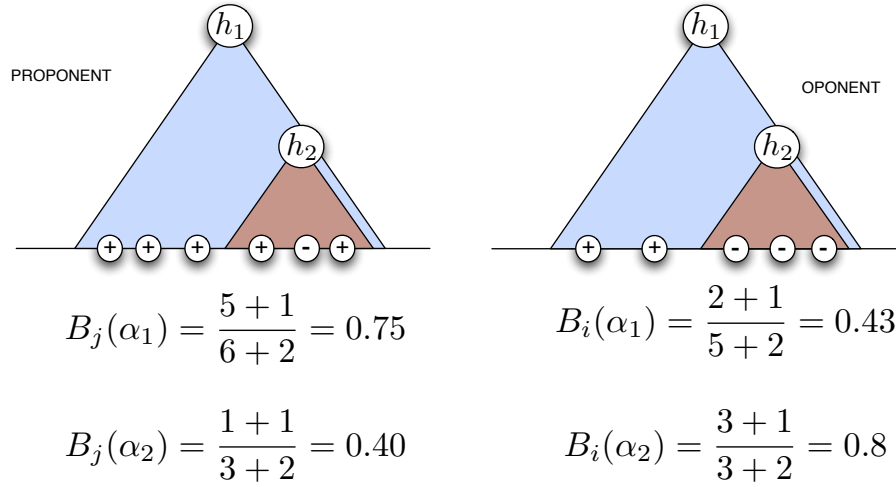


Fig. 3. An comparison of two individual viewpoints on arguments, attacks, and acceptability.

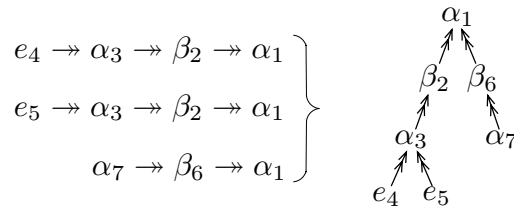


Fig. 4. Multiple argumentation lines rooted in the same argument α_1 can be composed into an argumentation tree.

Figure 4 illustrates this idea, where three different argumentation lines rooted in the same α_1 are shown with their corresponding argumentation tree. The α_i arguments are provided by the Proponent agent (the one proposing the root argument) while β_i arguments are provided by the Opponent trying to attack the Proponent's arguments.

Definition 9. Let an T α -rooted argumentation tree, where argument α belongs to an agent A_i , and let T^f be the example-free argumentation tree corresponding to T . Then the root argument α is warranted (or undefeated) iff all the leaves of T^f are arguments belonging to A_i .

Now we are able to define the state of the argumentation among two agents A_i and A_j at an instant t as the tuple $\langle R_i^t, R_j^t, G^t \rangle$, consisting of:

- $R_i^t = \{ \langle h, C \rangle \mid h \in \{h_1, \dots, h_n\} \}$, the set of arguments defending the current intensional definition $C_i^t = h_1 \vee \dots \vee h_n$ of agent A_i ;

- R_j^t is the same for A_j .
- G^t contains the collection of arguments generated before t by either agent, and belonging to one argumentation tree rooted in an argument in $R_i^t \cup R_j^t$.

3.2 Argument Generation Through Induction

Agents need two kinds of argument generation capabilities: generating *empirical arguments* based on the individual examples known to an agent, and generating *attack arguments* that attack arguments provided by other agents; notice that a defense argument is simply $\alpha' \rightarrow \beta \rightarrow \alpha$, i.e. an attack on the argument attacking a previous argument. Concerning empirical arguments, they are generated by using induction to find an initial intensional definition C from examples; for this reason, an agent A_i can generate an intensional definition of C by using any inductive learning algorithm capable of learning concepts as a disjunction of hypothesis, e.g. learning algorithms such as CN2[2] or FOIL[12].

Attack arguments, however, require a more sophisticated form of induction. When an agent A_i wants to generate an argument β such that $\beta \rightarrow \alpha$, A_i has to find an inductive hypothesis h for β that satisfies four conditions:

1. $\beta.h$ should support the opposite concept than α : namely $\beta.C = \neg\alpha.C$,
2. $\beta.h$ should have a high confidence $B_i(\beta)$ (at least being τ -admissible),
3. $\beta.h$ should satisfy $\alpha.h \sqsubset \beta.h$, and
4. β should not be attacked by any undefeated argument in G^t .

Currently existing inductive learning techniques cannot be applied out of the box, because they do not satisfy these conditions (mainly the last two conditions).

In previous work, we developed the Argumentation-based Bottom-up Induction (ABUI) algorithm, capable of performing such task [9]; this is the inductive algorithm used in our experiments by the agents. However, any algorithm which can search the space of hypotheses looking for a hypothesis which satisfies the four conditions stated before would work in our framework: e.g. CN2 could be modified in a way that the search of rules is restricted to the subspace of rules satisfying $\alpha.h \sqsubset \beta.h$.

Let L be the inductive algorithm used by an agent A_i ; then to attack an argument $\alpha = \langle h, \overline{C} \rangle$ for C has to generate an argument $\beta = \langle h', \widehat{C} \rangle$ such that $\beta \rightarrow \alpha$. The agent calls L to generate such hypothesis h' , then:

- If L returns an individually τ -admissible h' , then β is the attacking argument to be used.
- If L fails to find an argument, then A_i looks for examples attacking α in E_i . If any exist, then one such example is randomly chosen to be used as an attacking argument.

Otherwise, A_i is unable to generate any argument attacking α . If an argument or example is not enough to defeat another argument, additional arguments or examples can be sent in subsequent rounds of the A-MAIL protocol (see below).

3.3 Empirical Belief Revision

During argumentation, agents exchange arguments which contain new hypotheses and examples. These exchanges contain empirical knowledge that agents will integrate with their previous empirical beliefs. Consequently, their beliefs will change in such a way that their hypothesis are consistent with the accrued empirical evidence: we call this process empirical belief revision. The belief revision process of an agent A_i at an instant t , with an argumentation state $\langle R_i^t, R_j^t, G^t \rangle$ starts whenever A_i receives an argument from another agent:

1. If it is an example argument ε then $\varepsilon.e$ is added as a new example into E_i , i.e. A_i expands its extensional definition of C .
2. Whether the received argument is an example or an hypothesis, the agent re-evaluates the confidence of any argument in R_i^t or G^t : if any of these arguments becomes no longer τ -admissible for A_i they removed from R_i^{t+1} and G^{t+1} .
3. If any argument α in R_i^t became defeated, and A_i is not able to expand the argumentation tree rooted in α to defend it, then the hypothesis $\alpha.h$ will be removed from C_i . This means that some positive examples in E_i will not be covered by C_i any longer. The inductive learning algorithm is called again to generate new hypotheses that cover the now uncovered examples.

Thus, we have presented the A-MAIL approach to empirical argumentation for MAS. Notice that, as shown in Fig. 5 all aspects of the argumentation process (generating arguments and attacks, accepting arguments, determining defeat, and revising beliefs) are supported on an empirical basis and, from the point of view of MAS, implemented by autonomous decision making of artificial agents. The activities in Fig. 5 permit the MAS to be self-sufficient in a domain of empirical enquiry, since individual agents are autonomous and every decision is based on the empirical knowledge available to them.

The next section presents an application of this MAS framework to reach agreements in MAS.

4 Concept Convergence

We have developed A-MAIL as part of our research line on deliberative agreement³, in which 2 or more artificial agents use argumentation to reach different forms of agreement. In this section we will present a particular task of deliberative agreement called concept convergence. The task of *Concept Convergence* is defined as follows: Given two or more individuals which have individually learned non-equivalent meanings of a concept C from their individual experience, find a shared, equivalent, agreed-upon meaning of C .

Definition 10. *Concept Convergence (between 2 agents) is the task defined as follows:*

³ This is part the project Agreement Technologies: <http://www.agreement-technologies.org/>

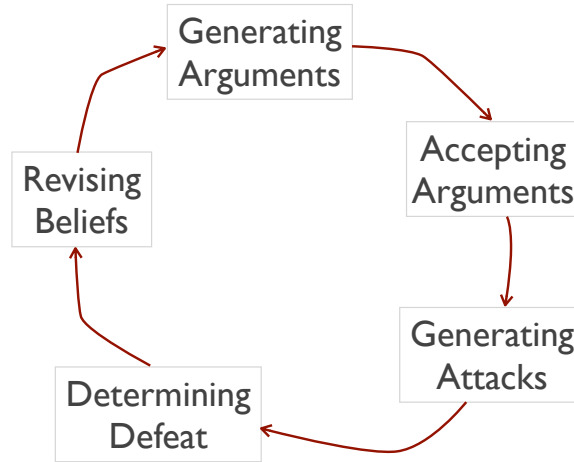


Fig. 5. The closed loop of empirically based activities used in argumentation.

Given two agents (A_i and A_j) that agree on the sign C denoting a concept ($C_i \cong C_j$) and with individually different intensional ($C_i \not\cong C_i$) and extensional ($E_i^+ \neq E_j^+$) definitions of that concept,

Find a convergent, shared and agreed-upon intensional description ($C_i \cong C_j$) for C that is consistent for each individual with their extensional descriptions.

For example, in the experiments reported in this paper, we used the domain of marine sponge identification. The two agents need to agree on the definition of the target concept $C = \text{hadromerida}$, among others. While in ontology alignment the focus is on establishing a mapping between the ontologies of the two agents, here we assume that the ontology is shared, i.e. both agents share the concept name *hadromerida*. Each agent has experience in a different area (one in the Atlantic, and the other in the Mediterranean), so they have collected different samples of *hadromerida* sponges, those samples constitute their extensional definitions (which are different, since each agent has collected sponges on their own). Now they want to agree on an intensional definition C , which describes such sponges. In our experiments, one such intensional definition reached by one of the agents is: $C =$ “all those sponges which do not have gemmules in their external features, whose megascleres had a tylostyle smooth form and that do not have a uniform length in their spikulate skeleton”. In the remainder of this paper we will present how agents can combine argumentation and inductive learning to argue about such definitions.

Concept convergence is assessed individually by an agent A_i by computing the *individual degree of convergence* among two definitions C_i and C_j as:

Definition 11. *The individual degree of convergence among two intensional definitions C_i and C_j for an agent A_i is:*

$$K_i(C_i, C_j) = \frac{|\{e \in E_i | C_i \sqsubseteq e \wedge C_j \sqsubseteq e\}|}{|\{e \in E_i | C_i \sqsubseteq e \vee C_j \sqsubseteq e\}|}$$

where K_i is 0 if the two definitions are totally divergent, and 1 when the two definitions are totally convergent. The degree of convergence corresponds to the ratio between the number examples covered by both definitions (intersection) and the number of examples covered by at least one definition (union). The closer the intersection is to the union, the more similar the definitions are.

Definition 12. *The joint degree of convergence of two intensional definitions C_i and C_j is:*

$$K(C_i, C_j) = \min(K_i(C_i, C_j), K_j(C_j, C_i))$$

Concept convergence is defined as follows:

Definition 13. *Two intensional definitions are convergent ($C_i \cong C_j$) if $K(C_i, C_j) \geq 1 - \epsilon$, where $0 \leq \epsilon \leq 1$ is a the degree of divergence allowed.*

The next section describes the protocol to achieve concept convergence.

4.1 Argumentation Protocol

The CC argumentation process follows an iteration protocol composed of a series of rounds, during which two agents will argue about the individual hypotheses that compose their intensional definitions of a concept C . At each round t of the protocol, each agent A_i holds a particular intensional definition C_i^t , and only one agent will hold a *token*. The holder of the token can assert new arguments and then the token will be passed on to the other agent. This cycle will continue until $C_i \cong C_j$.

The protocol starts at round $t = 0$ and works as follows:

1. Each agent A_i communicates to the other their current intensional definition by sharing R_i^0 . The token is given to one agent at random, and the protocol moves to 2.
2. The agents share their individual convergence degrees ($K_i(C_i, C_j)$ and $K_j(C_j, C_i)$). If $C_i \cong C_j$ the protocol ends with success; if no agent has produced a new attack in the last two rounds then the protocol ends with failure; otherwise it moves to 3.
3. If modified by belief revision, the agent A_i with the token, sends a message communicating its current intensional definition R_i^t . Then, the protocol moves to 4.
4. If any argument $\alpha \in R_i^t$ is defeated, and A_i can generate an argument α' to defend α , the argument will be sent to the other agent. Also, if any of the undefeated arguments $\beta \in R_j^t$ is not individually τ -admissible for A_i , and

A_i can find an argument β' to extend any argumentation line rotted in β , in order to attack it, then β' is sent to the other agent. If at least one of these arguments was sent, a new round $t + 1$ starts; the token is given to the other agent, and the protocol moves back to 2. Otherwise, if none of these arguments could be found, the protocol moves to 5.

5. If there is any example $e \in E_i^+$ such that $C_j^t \not\sqsubseteq e$ (i.e. a positive example not covered by the definition of A_j), A_i will send e to the other agent, stating that the intentional definition of A_j does not cover e . A new round $t + 1$ starts, the token is given to the other agent, and the protocol moves to 2.

Moreover, in order to ensure termination, no argument is allowed to be sent twice by the same agent. A-MAIL ensures that the convergence of the resulting concepts is at least τ if (1) the number of examples is finite, (2) the number of hypotheses that can be generated is finite. Convergence higher than τ cannot be ensured, since $100 \times (1 - \tau)\%$ of the examples covered by a τ -admissible hypothesis might be negative, causing divergence. Even when both agents use different inductive algorithms, convergence is assured since by assumption they are using the same finite generalization space, and there is no hypothesis τ -admissible by one agent that could not be τ -admissible by the other agent when both use the same acceptability condition over the same collection of examples.

4.2 Experimental Evaluation

In order to empirically evaluate A-MAIL with the purpose of concept convergence we used the marine sponge identification problem. Sponge classification is interesting because the difficulties arise from the morphological plasticity of the species, and from the incomplete knowledge of many of their biological and cytological features. Moreover, benthology specialists are distributed around the world and they have experience in different benthos that spawn species with different characteristics due to the local habitat conditions. The specific problem we target in these experiments is that of agreeing upon a shared description of the features that distinguish one order of sponges from the others.

To have an idea of the complexity of this problem, Figure 6 shows a description of one of the sponges collected from the Mediterranean sea used in our experiments. As Figure 6 shows, a sponge is defined by five groups of attributes: ecological features, external features, anatomy, features of its spikulate skeleton, and features of its tracts skeleton. Specifically, we used a collection of 280 sponges belonging to three different orders of the demospongiae family: axinellida, hadromerida and astrophorida. Such sponges were collected from both the Mediterranean sea and Atlantic ocean. In order to evaluate A-MAIL, we used each of the three orders as target concepts for concept convergence. In an experimental run, we split the 280 sponges randomly among the two agents and, given a target concept, the goal of the agents was to reach a convergent definition of such concept. The experiments model the process that two human experts undertake when they get together to discuss over which features determine whether a sponge belongs to a particular order.

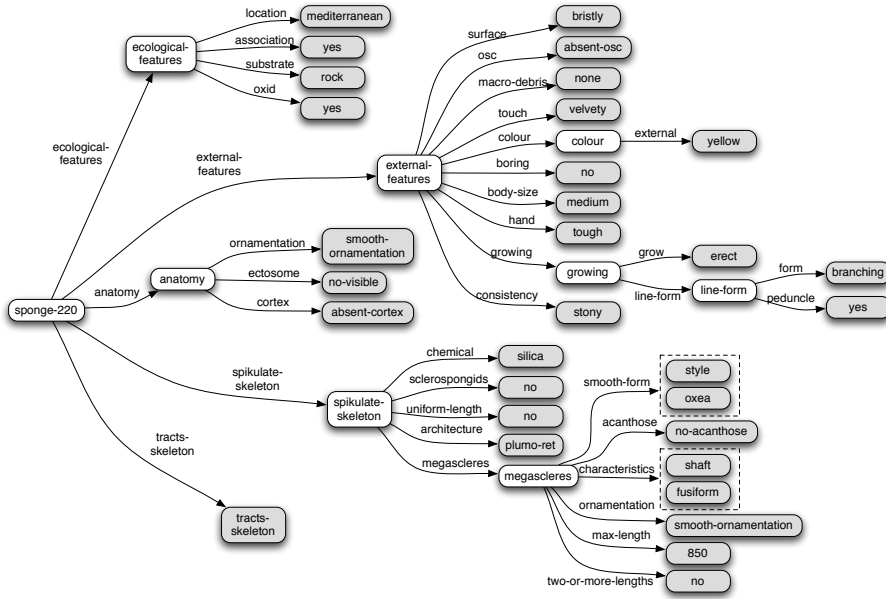


Fig. 6. A description of one of the sponges of the Axinellida order used in our experiments.

<i>C</i>	<i>Centralized</i>		<i>Individual</i>			<i>A-MAIL</i>		
	P	R	P	R	K	P	R	K
Axinellida	0.98	1.00	0.97	0.95	0.80	0.97	0.95	0.89
Hadromerida	0.85	0.98	0.89	0.91	0.78	0.92	0.96	0.97
Astrophorida	0.98	1.00	0.97	0.97	0.93	0.98	0.99	0.97

Table 1. Precision (P), Recall (R) and degree of convergence (K) for the intensional definitions obtained using A-MAIL versus those obtained using .

We compared the results of A-MAIL with respect to agents which do not perform argumentation (*Individual*), and to the result of centralizing all the examples and performing centralized induction (*Centralized*). Thus, the difference between the results of *individual* agents and agents using A-MAIL should provide a measure of the benefits of A-MAIL for concept convergence, whereas comparing with *Centralized* gives a measure of the quality of the outcome. All the results are the average of 10 executions, $\epsilon = 0.05$ and $\tau = 0.75$.

Table 1 shows one row for each of the 3 concepts we used in our evaluation: Axinellida, Hadromerida and Astrophorida. For each setting we show three values: precision, measuring how many of the examples covered by the definition are actually positive examples; recall, measuring how many of the total number of positive examples in the data set are covered by the definition; and convergence, as defined in Definition 12. The first thing we see in Table 1 is that A-MAIL is

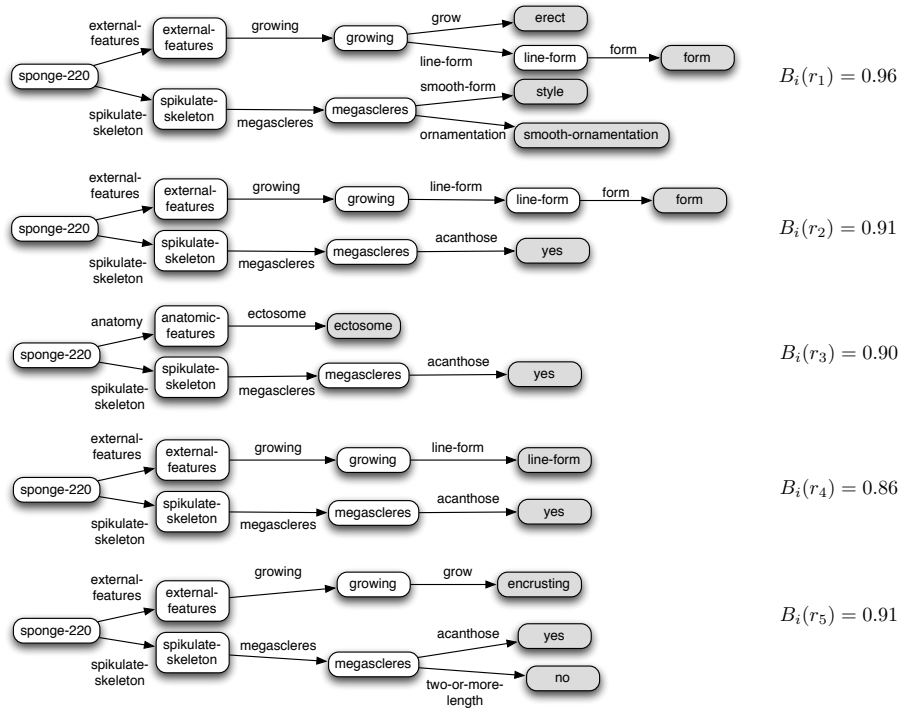


Fig. 7. Set of rules forming the definition of Axinellida and obtained by one of the agents using A-MAIL in our experiments.

able to increase convergence from the initial value appearing in the Individual setting. For all concepts except for Axinellida the convergence was higher than 0.95 (i.e. $1 - \epsilon$). Total convergence was not reached for that concepts because in our experiments $\tau = 0.75$, allowing hypotheses to cover some negative examples and preventing overfitting. This means that acceptable hypotheses can cover some negative examples, and thus generate some divergence. Increasing τ could improve convergence but if would make finding hypotheses by induction more difficult, and thus recall might suffer. Moreover, even precision and recall improve thanks to argumentation, reaching values close to the ones achieved by a Centralized setting.

Figure 7 shows the set of rules that one of the agents in our experiments using A-MAIL obtained as the definition of the concept Axinellida. For instance, the first rule states that “all the sponges with an erect and line-form growing, and with megascleres in the spikulate skeleton which had style smooth form and smooth ornamentation belong to the Axinellida order”. By looking at those rules, we can clearly see that both the growing external features and the characteristics of the megascleres are the distinctive features of the Axinellida order.

In summary, we can conclude that A-MAIL successfully achieves concept convergence by integrating argumentation and inductive learning, in addition to improve the quality of the intensional definition (precision and recall). This is achieved by exchanging only a small percentage of the examples the agents know (as opposed to the Centralized setting where all the examples are given to a single agent, which might not be feasible in some applications). Additionally, in average, the execution time of A-MAIL is lower than that of a centralized strategy.

5 Related Work

Concerning argumentation in MAS, previous work focuses on several issues like a) logics, protocols and languages that support argumentation, b) argument selection and c) argument interpretation, a recent overview can be found at [13].

The idea that argumentation might be useful for machine learning was discussed in [5], but no concrete proposal has followed, since the authors goal was propose that a defeasible logic approach to argumentation could provide a sound formalization for both expressing and reasoning with uncertain and incomplete information as appears in Machine Learning. Since the possible hypotheses can be induced from data could be considered an argument, and then by defining a proper attack and defeat relation, a sound hypotheses can be found. However, they did not develop the idea, or attempted the actual integration of an argumentation framework with any particular machine learning technique. Amgoud and Serrurier [1] elaborated on the same idea, proposing an argumentation framework for classification. Their focus is on classifying examples based on all the possible classification rules (in the form of arguments) rather than on a single one learned by a machine learning method.

A related idea is that of *argument-based machine learning* [7], where some examples are augmented with a justification or “supporting argument”. The idea is that those supporting arguments are then used to constrain the search in the hypotheses space: only those hypotheses which classify examples following the provided justification are considered. Notice that in this approach, arguments are used to augment the information contained in an example. A-MAIL uses arguments in a different way. A-MAIL does not require examples to be augmented with such supporting arguments; in A-MAIL the inductive process itself generates arguments. Notice, however, that both approaches could be merged, and that A-MAIL could also be designed to exploit extra information in the form of examples augmented with justifications. Moreover, A-MAIL is a model for multiagent induction, whereas argument-based machine learning is a framework for centralized induction which exploits additional annotations in the examples in the form of arguments.

The idea of using argumentation with case-based reasoning in multiagent systems has been explored by [11] in the AMAL framework. Compared to A-MAIL, AMAL focuses on lazy learning techniques where the goal is to argue about the classification of particular examples, whereas A-MAIL, although uses

cases and vase bases, allows agents to argue about rules generated through inductive learning techniques. Moreover, the AMAL framework explored a related idea to A-MAIL, namely learning from communication [10]. An approach similar to AMAL is PADUA [14], an argumentation framework that allows agents to use examples to argue about the classification of particular problems, but they generate association rules and do not perform concept learning.

6 Conclusions

The two main contributions of this paper are the definition of an argumentation framework for agents with inductive learning capabilities, and the introduction of the concept convergence task. Since our argumentation framework is based on reasoning from examples, we introduced the idea of *argument admissibility*, which measures how much empirical support an argument has, which is used to define an *attack* relation among arguments. A main contribution of the paper has been to show the feasibility of a completely automatic and autonomous approach to argumentation in empirical tasks. All necessary processes are autonomously performed by artificial agents: generating arguments from their experience, generating attacks to defeat or defend, changing their beliefs as a result of the argumentation process — they are all empirically based and autonomously undertaken by individual agents.

The A-MAIL framework has been applied in this paper to the concept convergence task. However, it can also be seen as a multi-agent induction technique to share inductive inferences [3]. As part of our future work, we want to extend our framework to deal with more complex inductive tasks, such achieving convergence on a collection of interrelated concepts, as well as scenarios with more than 2 agents. Our long term goal is to study the relation and integration of inductive inference and communication processes among groups of intelligent agents into a coherent unified MAS framework.

Acknowledgments

This research was partially supported by projects Next-CBR TIN2009-13692-C03-01 and Agreement Technologies CONSOLIDER CSD2007-0022.

References

- [1] Leila Amgoud and Mathieu Serrurier. Arguing and explaining classifications. In *Argumentation in Multi-Agent Systems*, volume 4946 of *LNCS*, pages 164–177, 2008.
- [2] Peter Clark and Tim Niblett. The CN2 induction algorithm. In *Machine Learning*, pages 261–283, 1989.
- [3] Winton Davies and Peter Edwards. The communication of inductive inferences. In *ECAI '96: Selected papers from the Workshop on Distributed Artificial Intelligence Meets Machine Learning, Learning in Multi-Agent Environments*, pages 223–241, London, UK, 1997. Springer-Verlag.

- [4] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.
- [5] Sergio Alejandro Gómez and Carlos Iván Chesñevar. Integrating defeasible argumentation and machine learning techniques. *CoRR*, cs.AI/0402057, 2004.
- [6] Tom Mitchell. *Machine Learning*. McGraw-Hill, 1997.
- [7] Martin Mozina, Jure Zabkar, and Ivan Bratko. Argument based machine learning. *Artificial Intelligence*, 171(10-15):922–937, 2007.
- [8] Carlos I. Chesnevar, Guillermo R. Simari, and Lluís Godo. Computing dialectical trees efficiently in possibilistic defeasible logic programming. In *LNAI/LNCS Series (Proc. 8th Intl. LPNMR Conf)*, pages 158–171. Springer, 2005.
- [9] Santi Ontañón and Enric Plaza. Multiagent inductive learning: an argumentation-based approach. In *Submitted*, 2010.
- [10] Santiago Ontañón and Enric Plaza. Case-based learning from proactive communication. In *IJCAI*, pages 999–1004, 2007.
- [11] Santiago Ontañón and Enric Plaza. Learning and joint deliberation through argumentation in multiagent systems. In *Proc. AAMAS’07*, pages 971–978, 2007.
- [12] J. R. Quinlan. Learning logical definitions from relations. *Machine Learning*, 5:239–266, 1990.
- [13] Iyad Rahwan and Guillermo R. Simari. *Argumentation in Artificial Intelligence*. Springer Publishing Company, Incorporated, 2009.
- [14] Maya Wardeh, Trevor J. M. Bench-Capon, and Frans Coenen. PADUA: a protocol for argumentation dialogue using association rules. *Artificial Intelligence in Law*, 17(3):183–215, 2009.