Concept Convergence in Empirical Domains

Santiago Ontañón and Enric Plaza

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

Abstract. How to achieve shared meaning is a significant issue when more than one intelligent agent is involved in the same domain. We define the task of *concept convergence*, by which intelligent agents can achieve a shared, agreed-upon meaning of a concept (restricted to empirical domains). For this purpose we present a framework that, integrating computational argumentation and inductive concept learning, allows a pair of agents to (1) learn a concept in an empirical domain, (2) argue about the concept's meaning, and (3) reach a shared agreed-upon concept definition. We apply this framework to marine sponges, a biological domain where the actual definitions of concepts such as orders, families and species are currently open to discussion. An experimental evaluation on marine sponges shows that concept convergence is achieved, within a reasonable number of interchanged arguments, and reaching short and accurate definitions (with respect to precision and recall).

1 Introduction

How to achieve shared meaning is a significant issue when more than one intelligent agent is involved in the same domain. In this paper we focus on empirical domains, where intelligent agents are able to learn, in an individual way, the concepts that are relevant to describe that domain from examples. In this scenario, two or more agents will require some process for sharing, comparing, critiquing and (eventually) agreeing on the meaning of the concepts of a domain. Our proposal is that an agent communication process based on argumentation supports the required aspects to find a shared, agreed-upon meaning of concepts.

For instance, in zoology, the definition of "manta ray" (the largest species of ray) has been a subject of debate; another example is in the domain of astronomy, where the definition of "planet" has been subject of recent debate. If more than one expert is to collaborate in these domains, they need to reach a shared definition of these concepts. Notice that these examples do not deal with the issue of *ontology alignment* (where different names or terms for the same concept are aligned); rather, the debate is about the meaning and scope (with respect to an empirical domain) of a particular concept. In this article we propose a framework intended to model a particular kind of process to reach this shared meaning we call concept convergence.

We will define the task of *concept convergence* 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. Two agents achieve concept convergence when (a) they share a concept C within some shared terminology, (b) their individual meanings for C are equivalent in a field of application, and (c) each agent individually accepts this agreed-upon meaning. Notice that concept convergence is less general than the complex discussion on how many species of manta ray should be recognized or how should be defined the concept of planet; however, it is more clearly specified and we will show it can be automated for empirical domains[1].

The task of concept convergence can be performed by the integration of computational argumentation and inductive concept learning. We have developed A-MAIL, a framework allows the agents to argue about the concept they learn using induction [7]. A-MAIL is a unified framework where autonomous agents learn from experience, solve problems with their learnt hypotheses, autonomously generate arguments from experience, communicate their inductive inferences, and argue about them in order to reach agreements with other agents.

The remainder of this paper is organized as follows. First we formally define concept convergence. Then our empirical argumentation framework A-MAIL is described. Then we motivate the usefulness of concept convergence in the biological domain of marine sponges, including an experimental evaluation of two inductive agents arguing about definitions of several concepts. The paper closes with related work, conclusions and future work.

2 Concept Convergence

Our approach integrates notions and techniques from two distinct fields of study —namely inductive learning and computational argumentation— to develop a new approach to achieve concept convergence. We will define the meaning and definition of concepts in the framework of inductive concept learning, which is the process by which given an *extensional definition* of a concept C then an *intensional* definition of a concept C expressed in an ontology $\mathcal{O}$ is found.

Let $\mathcal{E} = \{e_1...e_M\}$ be a field of application composed of M individuals described in an ontology $\mathcal{O}$ and let $C \in \mathcal{O}$ be a concept: an *extensional description* of C is a subset of individuals $E^+ \subset \mathcal{E}$ that are instances of C. E^+ are called (positive) examples of C, while the rest of the examples $E^- = \mathcal{E} - E^+$ are called counterexamples (or negative examples).

[1] Notice that ontology alignment (or matching) is a related topic but it focuses on determining correspondences between concepts [3]. As such, alignment's main goal is to establish a "concept name correspondence" relationship such that a semantic interoperability is achieved by being capable of substituting a concept name by a corresponding name. Concept convergence is different, we assume that the individual members of a multiagent system have a common concept vocabulary, but they still do not share a precise shared *definition* of some concept(s).

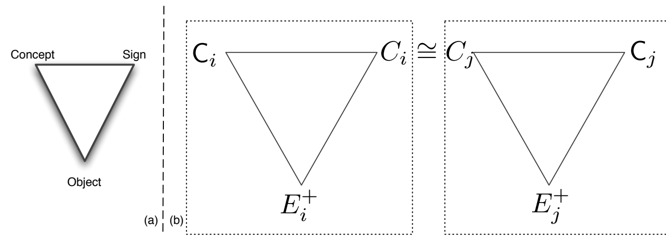


Fig. 1. (a) Semiotic triangle; (b) schema for two agents where a concept sign (C) is shared ($C_i \cong C_j$) while concept descriptions may be divergent ($\mathsf{C}_i \ncong \mathsf{C}_i$).

Definition 1. *An* intensional definition C *of a concept* C *is a well formed formula built using the concepts in* O *such that it subsumes* ($\sqsubseteq$) *all positive examples of* C *and no counterexample of* C*:*

$$\forall e_i \in E^+ : \mathsf{C} \sqsubseteq e_i \;\wedge\; \forall e_j \in E^- : \mathsf{C} \not\sqsubseteq e_j$$

For simplicity, we will shorten the previous expression as follows: $\mathsf{C} \sqsubseteq E^+ \wedge \mathsf{C} \not\sqsubseteq E^-$. In this framework, we will define the task of concept convergence between 2 agents based on the notion of *semiotic triangle*. The well-known semiotic triangle in Fig. 1(a) expresses meaning as the relationship between sign, concept, and object. Specifically:

1. A *sign* is a designation of the concept in some ontology (in our framework the name of the concept $C \in \mathcal{O}$);
2. A *concept* is "A unit of thought constituted through abstraction on the basis of properties common to a set of objects" [ISO 5963:1985] (in our framework the intensional description C)
3. An *object* is a material or immaterial part of the perceived world (in our framework, the objects in $\mathcal{E}$)

Now, concept convergence between 2 agents means that each one has its own semiotic triangle concerning a particular concept, as shown in Fig. 1(b). We assume that both agents share the designation of the concept C in an ontology, which in Fig. 1(b) is expressed by the equivalence $C_i \cong C_j$. The agents do not share their intensional definitions of the concept —which we'll assume are consistent with their extensional representations of concepts E_i^+ and E_j^+. Moreover, the agents do not share their individual collections of examples E_i and E_j.

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

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 ($\mathsf{C}_i \ncong \mathsf{C}_i$) and extensional ($E_i^+ \neq E_j^+$) definitions of that concept,*

Find *a convergent, shared and agreed-upon intensional description ($\mathsf{C}_i' \cong \mathsf{C}_j'$) for C that is consistent for each individual with their extensional descriptions.*

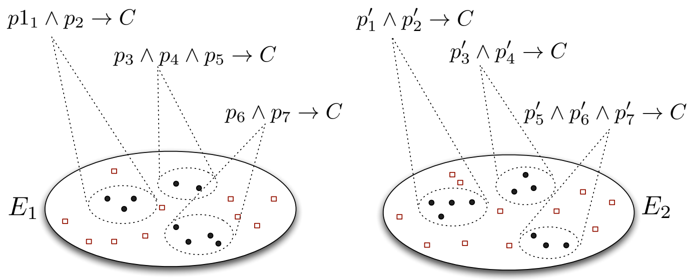


Fig. 2. The relationship of concept definitions for two inductive agents.

For example, in this paper we used the domain of marine sponge identification. The two agents need to agree on the definition of the target concept $C =$ *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: $\mathsf{C} =$ "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".

2.1 Empirical Argumentation for Concept Convergence

Concept convergence in empirical domains is modeled by agents that perform induction to achieve intensional definition of one or more concepts. Figure 2 shows the relationship of concept definitions for two inductive agents concerning a concept C. Each agent has a sample of examples of C and examples that are not C. The task of concept convergence is to find a shared and mutually acceptable definition for C that is consistent with the examples each agent has. The information exchanged during argumentation about how C should be defined is the information that will enact a process of belief revision in each individual agent until an agreed-upon definition is achieved. This paper focuses on 2-agent argumentation, leaving concept convergence among more agents for future work.

In the A-MAIL framework, an intensional definition of a concept C is represented as a *disjunctive description* $\mathsf{C} = r_1 \vee ... \vee r_n$, where each of the conjuncts r_i will be called a *generalization*, such that each positive example of C is subsumed by at least one of the generalizations, and no generalizations subsume any counterexample of C. When an example is subsumed by a generalization in C, we will say that the example is *covered*. Each one of these generalizations is a well formed formula representing a *generalization* of a set of examples. We assume that a *more-general-than* relation (*subsumption*) exists among generalizations, and when a generalization r_1 is more general than another generalization r_2 we write $r_1 \sqsubseteq r_2$. Additionally, if a generalization r is a generalization of an example e, we will also say that r is more general than e, or that r subsumes or covers e, noting it as $r \sqsubseteq e$. Moreover, for practical purposes the intensional definitions are allowed to subsume less than 100% of positive examples.

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

$$K_i(\mathsf{C}_i, \mathsf{C}_j) = \frac{|\{e \in E_i | \mathsf{C}_i \sqsubseteq e \wedge \mathsf{C}_j \sqsubseteq e\}|}{|\{e \in E_i | \mathsf{C}_i \sqsubseteq e \vee \mathsf{C}_j \sqsubseteq e\}|}$$

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

$$K(\mathsf{C}_i, \mathsf{C}_j) = min(K_i(\mathsf{C}_i, \mathsf{C}_j), K_j(\mathsf{C}_j, \mathsf{C}_i))$$

Concept convergence is defined as follows:

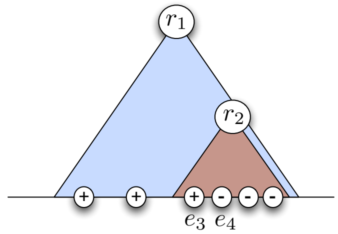
Definition 5. *Two intensional definitions are* convergent ($\mathsf{C}_i \cong \mathsf{C}_j$) *if* $K(\mathsf{C}_i, \mathsf{C}_j) \geq 1 - \epsilon$*, where* $0 \leq \epsilon \leq 1$ *is a the degree of divergence allowed.*

3 Empirical Argumentation

An argumentation framework $AF = \langle A, R \rangle$ is composed by a set of arguments A and an attack relation R among the arguments. In our approach we will adopt the semantics based on dialogical trees [1]. For a wider explanation the formal model underlying our framework see [5].

There are two kinds of arguments in A-MAIL:

- A *rule argument* $\alpha = \langle r, \overline{C} \rangle$ is a pair where r is a generalization and $\overline{C} \in \{C, \neg C\}$. An argument $\langle r, C \rangle$ states that induction has found a rule such that $r \to C$ (i.e. that examples covered by r belong to C), while $\langle r, \neg C \rangle$ states that induction has found a rule such that $r \to \neg C$ (i.e. that examples covered by r do not belong to C).
- An *example argument* $\alpha = \langle e, \overline{C} \rangle$ consists of an example $e \in \mathcal{E}$, which can be a positive or a negative example of C, i.e. $\overline{C} \in \{C, \neg C\}$.



$$\left.\begin{array}{l}\alpha_1 = \langle r_1, C\rangle \\ \\ \alpha_2 = \langle r_2, \neg C\rangle\end{array}\right\} \begin{array}{l}\alpha_2 \twoheadrightarrow \alpha_1 \\ \langle e_3, C\rangle \twoheadrightarrow \alpha_2 \\ \langle e_4, \neg C\rangle \twoheadrightarrow \alpha_1\end{array}$$

$$B_i(\alpha_1) = \frac{3+1}{6+2} = 0.5 \qquad B_i(\alpha_2) = \frac{3+1}{4+2} = 0.66$$

Fig. 3. Exemplification of several arguments, their confidences, and attack relations.

Moreover, we allow rules to cover some negative examples, while defining a confidence measure as follows:

Definition 6. *The* confidence $B_i(\alpha)$ *of a rule argument* α *for an agent* A_i *is:*

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

$B_i(\alpha)$ is the ratio of examples correctly covered by α over the total number examples covered by α. Moreover, we add 1 to the numerator and 2 to the denominator following the Laplace probability estimation procedure. Other confidence measures could be used, our framework only requires some confidence measure that reflects how much a set of examples endorses the argument.

Definition 7. *A rule argument* α *is* τ-acceptable *for an agent* A_i *if* $B_i(\alpha) \geq \tau$, *where* $0 \leq \tau \leq 1$.

In our framework, only τ-acceptable generalizations are allowed for a predetermined threshold τ. To ensure only highly quality rules are considered. Next, we will define attacks between arguments.

Definition 8. *An* attack *relation* ($\alpha \twoheadrightarrow \beta$) *between arguments* α, β *holds when:*
1. $\langle r_1, \widehat{C}\rangle \twoheadrightarrow \langle r_2, \overline{C}\rangle \iff \widehat{C} = \neg\overline{C} \wedge r_2 \sqsubset r_1$, *or*
2. $\langle e, \widehat{C}\rangle \twoheadrightarrow \langle r, \overline{C}\rangle \iff \widehat{C} = \neg\overline{C} \wedge r \sqsubseteq e$
(where $\overline{C}, \widehat{C} \in \{C, \neg C\}$*)*

Notice that a rule argument α only attacks another argument β if $\beta.r \sqsubset \alpha.r$, i.e. when β is a strictly more general argument than α. This is required since it implies that all the examples covered by α are also covered by β, and thus if they support opposing concepts, they must be in conflict.

Figure 3 exemplifies some arguments and with their corresponding attacks. Positive examples of the concept C are marked with a positive sign, whereas negative examples are marked with a negative sign. Rule 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 generalization r_1 supporting C, which covers 3 positive examples and 3 negative examples, and thus has confidence 0.5, while argument α_2 has a generalization r_2 supporting $\neg C$ with confidence 0.66, since it covers 3 negative examples and only one positive example. Two example arguments are shown: $\langle e_3, C\rangle$ and $\langle e_4, \neg C\rangle$. Now, $\alpha_2 \twoheadrightarrow \alpha_1$ because α_2 supports $\neg C$, α_1 supports C and $r_1 \sqsubset r_2$. Additionally $\langle e_3, C\rangle \twoheadrightarrow \alpha_2$, since e_3 is a positive example of C, α_2 supports $\neg C$ and $r_2 \sqsubseteq e_3$.

Next we will summarily define when arguments *defeat* other arguments, based on the idea of argumentation lines [1]. An *Argumentation Line* $\alpha_n \twoheadrightarrow \alpha_{n-1} \twoheadrightarrow ... \twoheadrightarrow \alpha_1$ is a sequence of arguments where α_i attacks α_{i-1} and α_1 is called the *root*. Notice that odd arguments are generated by the agent whose generalization is under attack (the *proponent*) and the even arguments are generated by the agent attacking that generalization (the *opponent*).

Moreover, 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. 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 may appear only in the leaves of an argumentation tree. The example-free argumentation tree T^f corresponding to T is a tree rooted in α that contains the same rule arguments of T but no example arguments.

In order to determine whether the root argument α is warranted (undefeated) or defeated the nodes of the α-rooted tree are marked U (undefeated) or D (defeated) according to the following (cautious) rules: (1) every leaf node is marked U; (2) each inner node is marked U iff all of its children are marked D, otherwise it is marked D.

Finally we will define the status 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:

- $R_i^t = \{\langle r, C\rangle | r \in \{r_1, ..., r_n\}\}$, the set of rule arguments representing the current intensional definition $\mathsf{C}_i^t = r_1 \vee ... \vee r_n$ for agent A_i.
- G^t contains the collection of arguments generated before t by either agent, and belonging to a tree rooted in an argument in $R_i^{t'}$, where $t' \leq t$.

R_j^t is the same for agent A_j. Now we can turn to integrate inductive learning with computational argumentation.

3.1 Argument Generation Through Induction

Agents need two kinds of argument generation capabilities: generating an initial intensional definition from examples, and generating attacks to arguments.

When an agent A_i that wants to generate an argument β that attacks another argument α, β has to satisfy four conditions: a) support the opposite concept than α, b) have a high confidence $B_i(\beta)$ (at least being τ-acceptable), c) satisfy

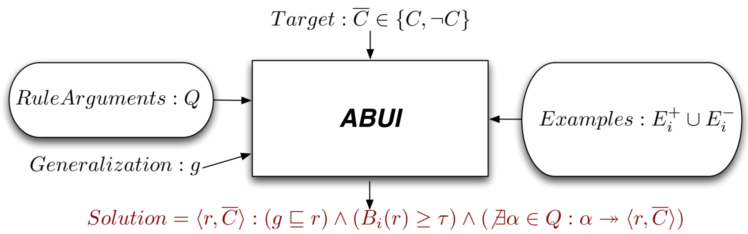


Fig. 4. ABUI is an inductive concept learning algorithm which can take additional background knowledge, in the form of arguments, into account.

$\beta \twoheadrightarrow \alpha$, and d) β should not be defeated by any argument previously generated by any of the agents. Existing inductive learning techniques cannot be applied out of the box for this process, because of the additional restrictions imposed. For this purpose, we developed the Argumentation-based Bottom-up Induction (ABUI) algorithm, capable of performing such task [7]. However, any algorithm which can search the space of rules, looking for one which satisfies the four conditions stated before would work in our framework.

ABUI is an inductive method for concept learning which, in addition to training examples, can take into account additional background knowledge in the form of arguments (see Fig. 4). ABUI is a bottom-up inductive learning method, which tries to generate rules that cover positive examples by starting from a positive example and generalizing it as much as possible in order to cover the maximum number of positive examples and while covering the minimum number of negative examples possible. During this generalization process, ABUI only considers those generalization which will lead to arguments not being defeated by any rule in the background knowledge. Specifically, ABUI takes 4 input parameters: a target concept $\overline{C} \in \{C, \neg C\}$, a set of examples $E_i^+ \cup E_i^-$, a generalization g, and a set of arguments Q which both agents have agreed to be true. ABUI finds (if it exists) an argument $\beta = \langle r, \overline{C}\rangle$ such that: $(g \sqsubseteq r) \wedge (B_i(r) \geq \tau) \wedge (\nexists \alpha \in Q : \alpha \twoheadrightarrow \langle r, \overline{C}\rangle)$.

To generate a β such that $\beta \twoheadrightarrow \alpha$, the agent calls ABUI with $g = \alpha.r$ and with the set of agreed upon arguments Q (the subset of arguments in G^t which are undefeated).

- If ABUI returns an individually τ-acceptable β, then β is the attacking argument to be used.
- If ABUI 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.

3.2 Belief Revision

During argumentation, agents exchange arguments which contain new rules and examples. The Belief Revision process of an agent A_i triggered at an instant t, with an argumentation state $\langle R_i^t, R_j^t, G^t\rangle$ works as follows:

1. Each example argument in G_i^t is added to E_i, i.e. A_i expands its extensional definition of C.
2. Since E_i might have changed, the confidence in any argument in R_i^t or G^t might have changed. If any of these arguments becomes not individually τ-acceptable 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 rule $\alpha.r$ will be removed from $\mathsf{C_i}$. As a consequence, some positive examples in E_i will not be covered by $\mathsf{C_i}$ any longer. Then ABUI is called with the now uncovered examples to find new rules that cover them and that will be added to $\mathsf{C_i}$.

3.3 Concept Convergence Argumentation Protocol

The concept convergence argumentation process follows an iterative protocol composed of a series of rounds, during which two agents argue about the individual rules that compose their intensional definitions of a concept C. At every 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 in the current round. At the end of each round the token is passed on to the other agent. This cycle continues until $\mathsf{C}_i \cong \mathsf{C}_j$.

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

1. Each agent A_i communicates their current intensional definition by sharing R_i^0. The token goes to one agent at random, and the protocol moves to 2.
2. The agents share $K_i(\mathsf{C}_i, \mathsf{C}_j)$ and $K_j(\mathsf{C}_j, \mathsf{C}_i)$, their individual convergence degrees. If $\mathsf{C}_i \cong \mathsf{C}_j$ 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 with the token, A_i, communicates 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, α' is sent to A_j. Also, if any of the undefeated arguments $\beta \in R_j^t$ is not individually τ-acceptable for A_i, and A_i can find an argument β' to extend any β-rooted argumentation line, in order to attack β, then β' is sent to A_j. If any 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 the protocol moves to 5.
5. If there is any example $e \in E_i^+$ such that $\mathsf{C}_j^t \not\sqsubseteq e$, A_i sends e to A_j (since 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 agent is allowed to send twice the same argument. 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 rules 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-acceptable rule might be negative.

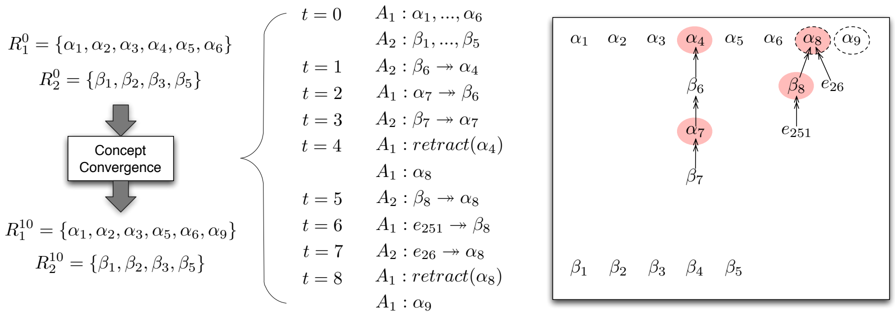


Fig. 5. An example concept convergence argumentation. Left hand side shows starting point and result. The middle shows the list of messages exchanged during the protocol. Right hand side shows the resulting argumentation trees.

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 rule τ-acceptable to one agent that could not be τ-acceptable to the other agent when both know the same collection of examples.

An example process of concept convergence is shown in Fig. 5. On the left hand side are the arguments (concept definition) of each agent before and after. In the middle, Fig. 5 shows the messages exchanged during the protocol, and on the right hand side the argumentation trees used. We can see that in round $t = 0$ the agents just exchange the arguments that compose their concept definitions. Then, in rounds 1, 2 and 3, the agents are arguing about the argument α_4, when ends up being defeated (shaded node). As a consequence, agent A_1 retracts α_4 and proposes a new one, α_8 (dashed node). The agents argue about α_8 in rounds 5 to 7, and eventually α_8 is defeated. Finally, agent A_1 retracts α_8, and proposes a new argument α_9, which is accepted (not attacked) by A_2. In this example, A_1 does not attack any argument in the definition of agent A_2.

4 Concept Convergence for Marine Sponges

Marine sponge classification poses a challenge to benthologists because of the incomplete knowledge of many of their biological and cytological features, and due to the morphological plasticity of the species. 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. Due to these problems, the classification or sponges into different classes is a challenging problem which is still under discussion among specialists.

The problem that we use as our test bed is that of learning which are the features that distinguish the different orders of sponges among each other, i.e. finding their intensional definition. We will focus on the scenario where two different experts have collected sponges in different locations and that these

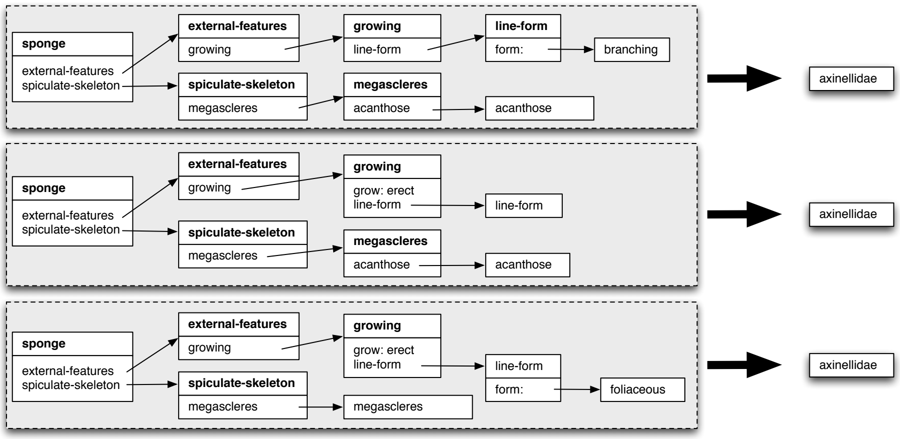


Fig. 6. Original concept definition learnt by an agent for the Axinellidae class, and composed of three rules.

sponges are properly classified into their respective orders. Now, the two experts are interested in having a specific agreed definition of each of the different order of sponges, so that their classification is clear in the future.

We have designed an experimental suite with a collection of 280 marine sponges pertaining to three different orders of the Demospongiae class (Astrophorida, Hadromerida and Axinellidae), taken from the Demospongiae dataset from the UCI repository. For our evaluation, we divide this collection of sponges in two disjoint sets, and give each set to one agent, which corresponds to an expert. Given a target order, say Axinellidae, each agent learns by induction a definition which characterizes all the sponges belonging to that order, and does not cover any sponge from any other order. After that, both agents argue about those definitions to reach an agreement using A-MAIL. The expected result is that the definition they reach after argumentation is better than the definitions they found individually (it is in agreement with the data known to both agents), and that it is achieved without exchanging large amounts of information.

Figure 6 shows an example definition of Axinellidae found by one agent in our experiments. The definition is composed of three rules. The first one, for instance states that "all the sponges which have a branching line-form growing and acanthose in the megascleres in the spikulate-skeleton" are Axinellidae.

Figure 7 shows two arguments (α_3 and β_4) as generated in one of our experiments by 2 agents while arguing about the definition of the Axinellidae order. An agent A_1 had proposed α_3, stating that "all the sponges which have a branching line-form growing and megascleres in the spikulate skeleton" are Axinellidae. This was so, since this rule was consistent with A_1 knowledge, i.e. with the set of sponges A_1 knew. However, this rule turned out to be too general, since it covered some sponges known to the other agent, A_2, which were not Axinellidae. In order to attack this rule, agent A_2 generated the argument β_4, which states

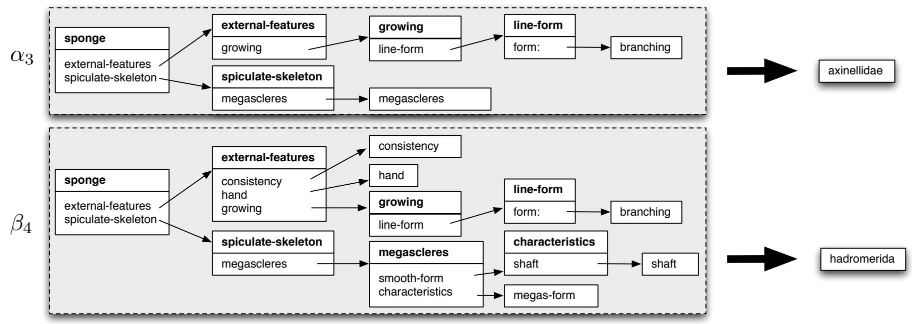


Fig. 7. Examples of arguments α_3 and β_4, where β_4 is attacking argument α_3.

Concept	*Centralized*		*Individual*			A-MAIL		
	P	R	P	R	K	P	R	K
Axinellidae	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 different methods.

that "all the sponges which have a branching line-form growing, a hand, and a shaft in the smooth form of the megascleres" are actually Hadromeridae. Since A_1 could not attack β_4, α_3 is defeated.

4.1 Experimental Evaluation

We perform concept convergence on each of the 3 orders in the marine sponges data set: Astrophorida, Hadromerida and Axinellidae. In an experimental run, we randomly split the data among the two agents and, given a target concept, the goal of the agents was to reach a convergent definition of such concept. We compare 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 concept learning (*Centralized*). Comparing the results of *Individual* agents and agents using A-MAIL provides a measure of the benefits of A-MAIL, 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 $\tau = 0.75$. We used the same induction algorithm, ABUI, for all the experiments.

Table 1 shows one row for each of the 3 concepts we used in our evaluation; for each one we show three values: precision, (P, how many examples covered that are actually positive examples); recall, (R, how many positive examples in the data set are covered by the definition); and convergence degree (K, as defined in Definition 4). The first thing we see is that indeed A-MAIL is able to increase convergence from the Individual setting. Moreover, for all concepts

Concept	*Centralized*		*Individual*		*A-MAIL*			
	time	R	time	R	time	R	NE	NR
Axinellidae	82.3s	7	40.8s	4.10	65.2s	6.65	10.7	15.6
Hadromerida	173.3s	11	75.6s	6.15	164.8s	9.2	18.5	32.6
Astrophorida	96.7s	6	47.7s	7.00	50.6s	4.1	4.1	9.7

Table 2. Comparison of the cost and quality of obtaining intensional definition from examples using different settings. Cost is measured in time (in seconds), and for A-MAIL, also the average number of example arguments (NE) and rule arguments (NR) exchanged. Quality is measured by the average number of rules (R) in intensional definitions.

except for Axinellidae the convergence degree is higher than 0.95 (i.e. $1 - \epsilon$). 100% convergence is not reached because $\tau = 0.75$ in our experiments. This means that acceptable rules can cover some negative examples, which allows for the appearance of some divergence. Increasing τ could improve convergence but makes finding rules by induction more difficult, and thus recall might suffer. Finally, notice that argumentation also improves precision and recall that reach values close to the ones achieved by Centralized.

Table 2 shows the average cost of each of the three settings. Column *time* shows the average CPU time used in each execution; when there are 2 agents (in the Individual and A-MAIL settings) individual time is obtained dividing 2. The Centralized setting uses more time on average than either Individual or A-MAIL settings. Table 2 also shows the average number of examples and of rule arguments exchanged among the agents, showing that A-MAIL only requires the exchange of a small amount of examples and arguments in order to converge.

Quality of solution is estimated by compactness of concept descriptions. The definitions found by A-MAIL are more compact (have less rules) than the definitions found by a Centralized approach. For instance, for the concept Astrophorida, the Centralized setting obtains a definition consisting of 6 rules, whereas A-MAIL generates only 4.1 rules on average.

In summary, we can conclude that A-MAIL successfully achieves concept convergence. 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 strategy where all the examples are given to a single agent, which might not be feasible in some applications). Moreover, the execution time of A-MAIL is on average lower than that of a centralized strategy. An interesting implication of this is that A-MAIL could be used for distributed induction, since it achieves similar results than a centralized approach, but at a lower cost, and in a distributed fashion.

5 Related Work

In our approach to concept convergence, we used our A-MAIL framewok [7]. A-MAIL is a framework which integrates inductive learning techniques with computational argumentation. In previous work, we applied A-MAIL to the task of distributed inductive learning, where agents are interested in benefitting from data known to other agents in order to improve performance. In this paper, we have used A-MAIL for a different task: concept convergence, where the goal is for two agents to *coordinate* their definitions of specific concepts. This process can be used, as we have shown, to model the process of argumentation between biology specialists about the definition of specific species. However, A-MAIL can be used for other tasks such as joint deliberation (when agents what to reach an agreement on a specific decision to a particular problem).

The integration of arguments into a machine learning framework is a recent idea, receiving increasing attention, as illustrated by the argument-based machine learning framework [4]. The main difference between this framework and A-MAIL is that in argument-based machine learning, arguments are given as the *input* of the learning process, while A-MAIL *generates* arguments by induction and uses them to reach agreements among agents.

Our work is also related to multiagent inductive learning. One of the earliest in this area was MALE [9], in which a collection of agents tightly cooperated during learning, effectively operating as if there was a single algorithm working on all data. Similar to MALE, DRL [8] is a distributed rule learning algorithm based on finding rules locally and then sending them to the other agents for evaluation. The idea of merging theories for concept learning has been also studied in the framework of Version Spaces [2].

6 Conclusions

This paper has presented the task of concept convergence. Concept convergence is different from ontology alignment in that we are not trying to find correspondence between ontologies, but reach shared definitions to known concepts. Since concept convergence is a broad subject we have focused on empirical domains. We have proposed to use inductive learning techniques to represent concepts and computational argumentation to regulate the communication process. For this purpose we have summarized A-MAIL, a framework that integrates inductive learning and computational argumentation; this integration is achieved by (1) considering rules learned by inductive learning as arguments, and (2) developing inductive learning techniques that are able to find new generalizations that are consistent with or attack a given set of arguments.

We have motivated the approach in the biological domain of marine sponges, where definitions of taxonomic concepts are still under debate. Experiments in this domain show that computational argumentation integrated with induction is capable of solving the concept convergence task, and the process is efficient (in the sense of the number of arguments that need to be exchanged).

As part of our future work, we intend to investigate more complex settings of concept convergence, and other tasks than can be performed by integrating induction with argumentation. Concerning concept convergence, we have started by focusing on the 2-agent scenario, but we intend to investigate concept convergence for n agents. Since computational argumentation is traditionally modeled as a dialogue between 2 agents, moving to a n-agents scenario requires more complex interaction models, such as those of committees (following argumentation-based deliberation in committees as in [6]). Another avenue of research is convergence on more than one concept; when these concepts are interdependent we surmise our current approach would work when dependencies are not circular; circular dependencies would require a more sophisticated approach.

Moreover, integrating induction with argumentation allows other kinds of tasks, such are using argumentation among agents to improve the individual inductive model [7]; another task is deliberative agreement, where 2 or more agents disagree on whether a situation or object is an instance of a concept C and user argumentation to reach an agreement on that issue.

Acknowledgments. This research was partially supported by projects Next-CBR (TIN2009-13692-C03-01), Aneris (PIF08-015-02) and Agreement Technologies (CONSOLIDER CSD2007-0022).

References

[1] Carlos I. Chesñévar, Guillermo R. Simari, and Lluís Godo. Computing dialectical trees efficiently in possibilistic defeasible logic programming. In *Logic Programming and Nonmonotonic Reasoning*, Lecture Notes in Computer Science, pages 158–171. Springer, 2005.

[2] Haym Hirsh. *Incremental version-space merging: a general framework for concept learning*. PhD thesis, Stanford University, Stanford, CA, USA, 1989.

[3] Yannis Kalfoglou and Marco Schorlemmer. Ontology mapping: The state of the art. In Y. Kalfoglou, M. Schorlemmer, A. Sheth, S. Staab, and M. Uschold, editors, *Semantic Interoperability and Integration*, number 04391 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2005.

[4] Martin Mozina, Jure Zabkar, and Ivan Bratko. Argument based machine learning. *Artificial Intelligence*, 171(10-15):922–937, 2007.

[5] Santiago Ontañón, Pilar Dellunde, Lluis Godo, and Enric Plaza. Towards a logical model of induction from examples and communication. In *Proceedings of the 13th International Conference of the Catalan Association for Artificial Intelligence*, volume In Press of *Frontiers in Artificial Intelligence*. IOS Press, 2010.

[6] Santiago Ontañón and Enric Plaza. An argumentation-based framework for deliberation in multi-agent systems. In I. Rahwan, S. Parsons, and C. Reed, editors, *Argumentation in Multi-Agent Systems*, volume 4946 of *Lecture Notes in Artificial Intelligence*, pages 178–196. Springer Verlag, 2008.

[7] Santiago Ontañón and Enric Plaza. Multiagent inductive learning: an argumentation-based approach. In *ICML-2010*, page [http://www.icml2010.org/papers/284.pdf]. Omnipress, 2010.

[8] F. J. Provost and D. Hennessy. Scaling up: Distributed machine learning with cooperation. In *Proc. 13th AAAI Conference*, pages 74–79. AAAI Press, 1996.

[9] Sati S. Sian. Extending learning to multiple agents: Issues and a model for multi-agent machine learning (MA-ML). In *EWSL-91: Proc. European Working Session on Machine Learning*, pages 440–456, London, UK, 1991. Springer Verlag.