Argumentation on Meaning: a Semiotic Model for Contrast Set Alignment

Kemo ADRIAN [a,b] and Enric PLAZA [a]
[a] IIIA-CSIC, Campus UAB, Carrer de Can Planes, 08193 Bellaterra, Catalonia, Spain
[b] Universitat Autonoma de Barcelona, 08193 Bellaterra, Catalonia, Spain

Abstract. Argumentation theory has been investigated as a possible way to link ontologies and machine learning, in multi-agent systems facing situations of heterogeneous knowledge. We are investigating scenarios where this argumentation takes place over one contrast set, a segment of two agents' ontology. Addressing a whole contrast set is challenging, as argumentation theory is normally used to align one pair of concepts from a contrast set. Our approach to use argumentation theory to align a whole contrast set goes in three parts. First, we clarify and define what is an acceptable alignment between two contrast sets with the notion of agreement. Then, we formalize as disagreements the pairing relations between concepts that prevent this agreement. Finally, we propose a model that identifies and resolves disagreements. This article focuses on the presentation of our model together with a preliminary experimental evaluation.

Keywords. Ontologies, Machine Learning, Inductive Learning, Argumentation Theory, Multi-Agent System.

1. Introduction

In machine learning, agents can learn to classify elements – called object or *examples* – from their environment. This learning gives to the agents a structured knowledge about a domain, a knowledge which is individual to each learning and therefore to each agent. A classification can be seen as a *contrast set* of an ontology, a particular subset of the ontology regrouping a set of classes or *concepts* partitioning a specific domain. For instance, the colors partition the visible spectrum domain of any ontology including it. Therefore, aligning the classes of a same domain from different ontologies becomes similar to aligning two classifiers. The classifiers, however, have the support of their learning data-sets to complement their ontological knowledge. An approach that takes advantage of these two aspects in the task of classifier alignment is the formal argumentation theory, where both the examples and the generalizations learned over these examples can be used as arguments in order to create new generalizations which allow consistent classifications over larger contexts.

Ontology alignment and machine learning both address the question of *meaning*. There approaches to that question, however, are different. In previous work [2], we already discussed that machine learning tends to have an *externally grounded* approach of meaning [5][11], while ontology alignment can either consider meaning as a *conceptual web* [6] or externally grounded. Argumentation has often been used in a conceptual web approach of ontology alignment [12][7], but recent work on interaction based on-

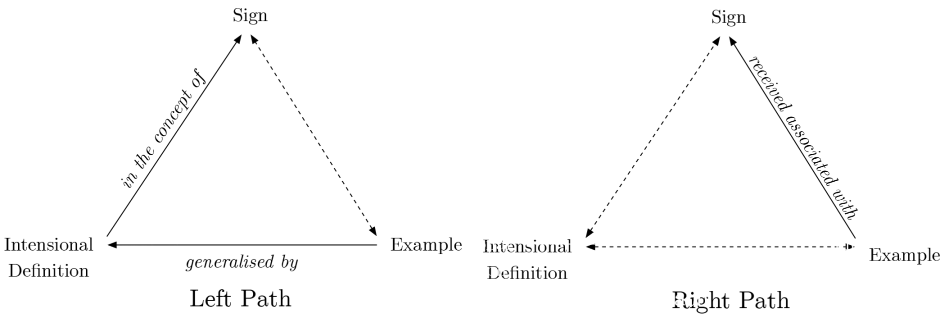

Figure 1. Association Paths as Semiotic Triangles.

tology alignment also uses grounded concepts [4]. We aim to provide an argumentation framework in the latter paradigm, which focuses on limiting the information exchanged between the agents.

A classifier alignment can be conceptualized as a naming game between two agents, where they both need to name the examples from a set U called their *context*, achieving the highest number of examples named similarly. In order to do so, the agents need to be able to *associate* each example from the context with a sign from a *lexicon*. We distinguish two methods for an agent to map a context U to a lexicon L: the *right*-path and the *left*-path associations. The right-path associations of a classifier K, noted $U \overset{r}{\underset{K}{\mapsto}} L$, are received from an external source by the agent, for instance from the experimenter as a training data-set. They consist of an array of example-sign pairs. The left-path associations of a classifier K, noted $U \overset{l}{\underset{K}{\mapsto}} L$, however, use the knowledge of the classifier to associate a sign with an example. The example is fed to the classifier as an input, and the same classifier outputs its sign. Two right-path associations can map a same example to different signs, creating inconsistencies in the knowledge of the agents. Moreover, only examples from the context known by an agent can be named through right-path associations. A same left-path association, however, can name new examples through the process of classification. Moreover, the result of a classification being reproducible, two agents classifying with the same generalizations are guaranteed to classify in a consistent way and score high in their naming game.

Argumentation frameworks are used in artificial intelligence to manage contentious information among multi-agent systems and draw conclusions from it. Among them, the framework AMAIL [8] has been shown to be a promising way to align two classifiers which are using a feature structures formalism to represent their knowledge. In order to align two classifiers, the AMAIL framework uses argumentation over each pair of concepts that belong to different agents but are supposed to be equivalent in order to find a new common set of generalizations for the two agents to have similar left-path associations. Unfortunately, this means that the AMAIL algorithm needs to know which concepts are expected to be equivalents in order to align them. Moreover, the AMAIL algorithm focuses on an approach based on the prevalence of right-path associations, which make the framework unable to work on scenarios with inconsistent data, where there are unique examples associated to different signs through the right-path.

Figure 1 shows that these two paths, which are linking signs, generalizations and examples draw a triangle that evokes a *semiotic triangle* [10], which is why we are describing our model to align classifiers as a semiotic model. The elements of our model are named from their semiotic equivalents, and they are used to formalize the notion of classifier alignment into the partitioning, pairing and adjustment of fairly-equivalent concepts from different contrast sets. The notion of alignment becomes an *agreement* over the meaning, as the agents use argumentation on each of these pairs of concepts in order to align them. By introducing the notion of pairing relations, we allow the agents to use an argumentation framework inspired from AMAIL that can be used without a pre-existing mapping of the expected equivalence relations between the concepts of the different agents. Moreover, we adopt a *constructivist* approach on argumentation, where the left-path associations prevails over the right-path associations, which allows the model to work on scenarios with inconsistent data. The full formalism of our approach has been formalized in a synthetic article [2]. Section 2 presents a short summary of this article.

2. Approach

In our approach, objects – called examples – are represented as feature terms [3]. Feature terms, also called feature structures or ψ-terms, are a generalization of first-order terms that have been introduced in theoretical computer science in order to formalize object-oriented capabilities of declarative languages. Feature terms correspond to a different subset of first-order logic than description logics, but have a similar expressive power. Generalizations of these examples can also be represented as feature terms. The basic operation between feature terms is *subsumption*: we will use $\psi_1 \sqsubseteq \psi_2$ to express that a term ψ_1 subsumes another term ψ_2 – that is to say ψ_1 is more general (or equal) than ψ_2. Our approach is broadly inspired by semiotics, and and we will now introduce some basic notions of that approach.

2.1. Semiotic Elements

The semiotic elements are the building blocks of our knowledge representation. They help us to formalize the notions of example-sign associations, consistency and alignment. The first three semiotic elements are the *extensional* definition E, the *intensional* definition I and the *sign* s. An extensional definition is a set of examples, an intensional definition is a set of generalizations and a sign is just a symbol (a string of characters in our case). These three semiotic elements can be combined into a *concept* C, which is the fourth and last type of semiotic elements. A concept is composed of an extensional, an intensional definition, and a sign, such that the intensional definition subsumes the examples of the extensional definition. The semiotic elements of the concept C are $s(C), I(C)$ and $E(C)$ and we note $C = \langle s(C), I(C), E(C) \rangle$. In a concept, the association between $s(C)$ and $E(C)$ represents the right-path associations, while the associations between $s(C)$ and an example subsumed by a generalization $g \in I(C)$ are the left-path associations.

2.2. Contrast Set

A single concept only partitions a subset of a context U through its associations. In order to partition a whole context, the agents need a set of concepts. A *contrast set* $K = \langle U, S = \{C_1, \ldots, C_n\} \rangle$ is a relation between a context U and the set of concepts $\{C_1, \ldots, C_n\}$ such that these concepts partition U, meaning that each example in U is subsumed by one and only one concept from S; this means any intensional definition $I(C)$

does not subsume any example in another $E(C')$ whenever $C, C' \in S$. A variation of the contrast set is the *hypothesis*. A hypothesis H is a contrast set without the constraint on the relation between the context and the set of concepts. A hypothesis is usually used by one agent A_1 to project the left-path associations of another agent A_2 on its own context U_1. Therefore, the set of concepts of the other agent may or may not partition U_1, and therefore may or may not be a contrast set.

2.3. Pairing Relations

A pairing relation qualifies the relation between two concepts in a given context. A pairing relation is independent of the extensional definitions of the concepts, as the left-path associations are used to compute the pairing relations between concepts. The process of computing a pairing relation is presented with more details in our previous paper [2]. The pairing relation between two concepts C_1 and C_2 in a context U is computed by finding these two concepts' adjunct sets $Adj(C_1, U)$ and $Adj(C_2, U)$ such that adjunct set is $Adj(C_k, U) = \{e \in U | I(C_k) \sqsubseteq e\}$. From these adjunct sets, three pairing partial sets are obtained: $U_{1,\bar{2}}$, $U_{\bar{1},2}$ and $U_{1,2}$. The two firsts sets are the two set differences of $Adj(C_1, U)$ and $Adj(C_2, U)$, while the third set is the intersection of $Adj(C_1, U)$ and $Adj(C_2, U)$. The three pairing partial sets are defining a r-triplet, a triplet of Boolean values $r(C_1, C_2, U)$ that represents whether or not each pairing partial set is empty. Each of the 2^3 possible r-triplets is mapped to a particular pairing relation: namely, *overlap*, *inclusion*, *equivalence* and *disjunction* are the four main types of pairing relations.

The main advantage of this model of computation is the fact that we can obtain the overall r-triplet $r(C_1, C_2, U_O)$ of two concepts C_1 and C_2 in the overall context of our agents, by applying a composition law to the agents' two local r-triplets $r(C_1, C_2, U_1)$ and $r(C_1, C_2, U_2)$. This composition law allows the agents to understand how their concepts are semantically related at the general level without having to exchange their examples. In the first version of our model, we were only using Boolean values in our r-triplets to represent the emptiness of each pairing partial set. This did not allowed us to consider cases where the intensional definition learned for a concept was not *exactly* subsuming the concept's extensional definition. In order to manage a certain degree of error when the agents are using inductive learning, the r-triplets are now containing the cardinal number of each pairing partial set. This modification has consequences on the composition law, but inferring overall triplets from local triplets remains useful to limit the number of examples exchanged by the agents.

2.4. Agreement on Meaning

The aim of an argumentation is to have the two agents involved in the argumentation reach a state of *mutual intelligibility*. This state is reached when both agents are classifying any given example of their overall context in two concepts that share a same sign. When an agent is convinced that this state has been reached, we say that this agent is in *synchronic* agreement. Besides the mutual intelligibility, the agents also want to still be able to discriminate, after the argumentation, the examples that they originally were classified in different concepts. If the new contrast set of an agent is, indeed, a refinement of the old one, we say that this agent is in *diachronic* agreement. Before an argumentation, an agent is always in diachronic agreement but not in synchronic agreement.

The argumentation on meaning helps the agents to reach a synchronic agreement without breaking the diachronic agreement.

The synchronic agreement can be expressed as a set of conditions on the overall pairing relations of the two agents. A violation of these conditions is called a synchronic *disagreement*. We list seven types of synchronic disagreements, divided in four families. The overlap disagreement and the hypo/hypernymy disagreement are semantic disagreement, caused by concepts that have nor disjoint neither equivalent concepts in the other contrast set. The synonymy and the homonymy disagreements are lexical disagreements, respectively caused by equivalent concepts that are not sharing their sign and disjoint concepts that are sharing the same sign. An indiscernible disagreement is caused by a relation between two concepts that is left with too few proper examples on one concept. A self disagreement is caused by two overlapping concepts from the same contrast set. Finally, the untranslatable disagreement is caused by a concept not having an equivalent concept in another contrast set.

3. Model

Our two agent model of argumentation will tackle the disagreements between the agents sequentially, reducing the number of disagreements while keeping a diachronic agreement for both agents. Before the argumentation, each agent has an initial contrast set that is learned over examples of a domain (a ‘data set’), including a set of right-path associations (supervised learning). The argumentation takes place turn by turn, with one agent receiving a token at the beginning of the argumentation, taking a set of actions, then passing the token to the other agent. An agent can take actions only when it has the token, and the last action that it takes is always to pass the token to the other agent. The token is exchanged until termination is detected. The actions taken by an agent while it has the token are decided by the agent’s inputs and its current state. The two variables that impact the behaviour of one agent at a given turn are the agent’s state (a qualitative variable) and the messages that this agent has received from the other agent. Each agent has the same set of possible states, making our argumentation model *symmetric*.

3.1. General functions of an agent

Our agents have 8 principal functions. Their first and basic function is their ability to send elements of their knowledge (semiotic elements, r-triplets, pairing relations, etc.) to the other agent through messages. The seven remaining functions are listed below.

Deleting Concepts. The agents can delete concepts from their contrast sets at anytime. Since the agents are building hypotheses as an image of each others contrast sets, when an agent A_1 deletes a concept C from its contrast set it notifies the other agent A_2 with a message, asking A_2 to delete C from its hypothesis.

Creating Concepts. The agents can create new concepts. The easiest way is to use a set of right-path associations, and use them as a set of inputs for inductive learning. The inductive learning generates a set of generalizations that become the intensional definition of the new concept(s), while the lexicon of the set of associations becomes the sign(s) of the new concepts and the examples become the extensional definition(s). Instead of creating an intensional definition from an extensional definition, the agents can also create an extensional definition from a set of generalizations. If an agent A_i with a context U_i receives an intensional definition I associated to a sign s, it can create a new concept $C = \langle s, I, \{e \in U_i | I \sqsubseteq e\} \rangle$. Finally, the agents can create concepts through argumentation. The creation of a new concept C through argumentation triggers a secondary argumentation within the argumentation on meaning, where the agents first define the properties of the adjunct set $Adj(C, U_O)$ of the new concept in the overall context. Then, an agent A_1 uses induction to learn an intensional definition I of the new concept from the adjunct set $Adj(C, U_1)$ over its local context U_1. I might be inaccurate, as the examples from the other agent's context are not taken into account during its creation, and in the argumentation process the agents will modify I until both are satisfied. We do not detail this argumentation process in this article, but it is fairly similar to the AMAIL algorithm. Finally, the agents decide on a sign for the new concept.

Reassigning Overlaps. When a set of examples is left-path associated to two labels, the agents can choose to reassign them to one of these two labels. Any method can be used to reassign them, as long as this method assigns each example to the same label independently of the agent using it. In our implementation, we are using the anti-unification similarity measure [9] between each example that needs to be reassigned and the intensional definitions of the two concepts that have the two involved labels for signs. Each example is reassigned to the label linked to the intensional definition that has the highest similarity with them.

Determining Overall Pairing Relations. In order to compute the overall pairing relation between two concepts, the agents first compute their local adjunct sets. With them, they compute the local pairing partial sets, then the local r-triplets. After exchanging their local r-triplets, the agents can compute the overall pairing relation using composition laws presented in our previous publication [2].

Finding disagreements. Once an agent knows about the overall pairing relations between its concepts and the other agent's concepts, this agent can find the disagreements between them as these disagreements are directly expressed as configurations of pairing relations and signs. Using the overall pairing relations allow the agents to resolve their disagreements over the whole set of examples that they collectively have access to, while guarantying that the pairing relations used to qualify the disagreements are the same for both agents —unlike local relations that are individual to each agent.

Looking for transitivity issues. When our model admits a degree of error, the properties of the pairing relations change. The most problematic change is the lost of transitivity for the relation of equivalence. Given three concepts A, B and C, if we had $A \equiv_U B$ and $B \equiv_U C$, then we had automatically $A \equiv_U C$ with Boolean r-triplets. However, once we introduce an error threshold, the relation between B and C is not necessarily an equivalence anymore. To address this issue, the agents need to look for such configurations with their overall pairing relations and delete either A or C. The concept that has the most examples shared with B is the one deleted.

Reassigning lexicon items. The lexicon of an agent is the set of signs used in its current contrast set. Once the agents have aligned their contrast sets in a configuration that allows a one-to-one mapping of their concepts, they can reassign their lexicon items in order to get rid of their lexical disagreements. Moreover, when concepts are created they have new signs added into the lexicon. In order to reuse the old signs that became unassigned following their concepts deletion, the agents can reassign them to the concepts that are using new signs after reaching mutual intelligibility.

3.2. Resolving disagreements

Since disagreements are always related to pairs of concepts, removing one concept of the two resolves the disagreement. However, the deletion can cause new disagreements to appear. New disagreements caused by the resolution of explored disagreements are addressed in the same way as any disagreement.

Semantic Disagreements. Semantic disagreements are resolved by deleting one of the concepts that causes them and creating new concepts for the examples of the overall context that become unassigned. *A hypo/hypernymy* is resolved by removing the hypernym from its contrast set and replacing it by two co-hyponyms, one of them being the hyponym of the disagreement. The co-hyponym is created through argumentation. An *overlap* is resolved by creating a concept that covers the examples at the intersection of the two overlapping concepts. This new concept causes hypo/hypernymy disagreements where the hypernyms are the two overlapping concepts. When these two hypo/hypernymies are resolved, the hypernyms are deleted and the overlap is resolved as well. An *indistinguishable* disagreement is resolved by deleting one of the two concepts. Unlike in the other types of disagreement, the examples remain unassigned.

Lexical Disagreements. Lexical disagreements are resolved by changing the signs of the concepts that cause them. A *synonymy* disagreement is resolved by changing the signs of the two synonyms for a new sign. A *homonymy* disagreement is resolved by changing the sign of the two homonyms for two different signs.

Untranslatable Disagreements. There is only one concept involved in an untranslatable disagreement. The sign-intensional definition association of this concept is sent to the agent that does not have the concept in its contrast set, and this agent will create a new concept for this association.

Self-Disagreements. A self disagreement can only be caused by an overlap between two concepts of a same contrast set. Unlike overlap disagreements, however, the self-disagreements are not solved by creating a new concept but by reassigning some of their examples. Doing so in self-disagreements does not cause diachronic disagreements, as it would be the case for overlap disagreements. The agents reassign the examples from the intersection of these two concepts to these concepts and create new intensional definitions for them through argumentation, creating two new concepts with the same signs as the old ones.

3.3. Organization of an argumentation

An argumentation over concept meaning goes as follow: first the agents exchange the sign-intensional definition associations for each of their concepts. Knowing these associations, each agent can compute the overall pairing relation between each of the concepts involved. Once done, the agents list their disagreements and proceed to resolve them one by one. The agents prioritize the resolution of their disagreements in the following order: first self-disagreements, then semantic disagreements, then untranslatable disagreements, and finally lexical disagreements. Each time that a new concept has been created and added to a contrast set, the agents exchange their new local pairing relations and compute the missing overall relations. Each time that a new concept has been added or removed to a contrast set, the agents are updating their list of disagreements. Once there are no overall pairing relations causing disagreements anymore, the agents update their lexicon a last time. The termination condition being met, the token is confiscated from the agents and the argumentation stops.

4. Experiments on Two Agent Disagreements

Our model has multiple properties that can be tested. The first and most important is its *generality*, that is to say that our model succeeds to reach mutual intelligibility between two agents through a monotonic refinement of their contrast sets for all disagreement types and any combination thereof.

4.1. Experimental Variables

Parameters. There are three parameters in our model: the error threshold τ_E, the argument acceptability – a parameter of our inductive learning algorithm ABUI – and the redundancy of the examples between the two agents' initial contexts. The error threshold is set as 5, a value that has been experimentally found challenging in preliminary studies. The argument acceptability is set at 0.75, the value used in AMAIL tests. The redundancy is set to *O%*, meaning that the initial contexts of the two agents are always completely disjoint in our setup. Since the information shared by the two agents is minimal, this setup is the most challenging for an argumentation.

Set-up Disagreements. The only independent variables of our experiment is the number and types of disagreement that we set up. Only four types of disagreements are set up: overlaps, hypo/hypernymies, synonymies and homonymies. A set up disagreement is not a disagreement as we defined it. A Set-up Disagreement (SD) is a set of inconsistent right-path associations distributed as training sets among two agents in order to obtain a specific type of disagreement between the agents once the agents learned their initial contrast sets as a result of their initial training. SDs are always set-up in the overall context.

4.1.1. Dependent Variables

In our experiments we measure five dependent variables:

Synchronic Agreement Ratio. The Synchronic Agreement Ratio (SAR) is the ratio between the of examples from the overall context that are named through a left path association with the same unique sign by both agents, over the total number of examples in the overall context. SAR measures how well the agents have reached mutual intelligibility:

Definition 1 (SAR) *Let A_1 and A_2 be two agents, with contrast sets $K1$ and $K2$. The Synchronic Agreement Ratio of A_1 and A_2 is:*

$$SAR(A_1,A_2) = \frac{|\{e \in U_O | e_{K1} \overset{l}{\mapsto} s \wedge e_{K2} \overset{l}{\mapsto} s\}|}{|U_O|}$$

Diachronic Agreement Ratio. The Diachronic Agreement Ratio (DAR) is the inverse of the diachronic disagreement ratio, a measure which is the ratio between (1) the number of pairs of examples from an agent's initial context that were not in a same concept in the initial contrast set *and* are in a same concepts the final contrast set, and (2) the number of pairs of examples that were in two different concepts in the initial contrast set. The DAR measures how well the principle of concept refinement is maintained throughout the argumentation process.

Definition 2 (DAR) *Let A be an agent that has an initial contrast set $K = (U, Q)$ and a final contrast set $K' = (U', Q')$. The Diachronic Agreement Ratio of A is:*

$$DAR(A) = 1 - \frac{|\{e_1, e_2 \in U | e_1 \,{}_{K}\!\overset{l}{\mapsto} s \wedge e_2 \,{}_{K}\!\overset{l}{\mapsto} s' \wedge e_1 \,{}_{K'}\!\overset{l}{\mapsto} s'' \wedge e_2 \,{}_{K'}\!\overset{l}{\mapsto} s'' \wedge s \neq s'\}|}{|\{e_1, e_2 \in U | e_1 \,{}_{K}\!\overset{l}{\mapsto} s \wedge e_2 \,{}_{K}\!\overset{l}{\mapsto} s' \wedge s \neq s'\}|}$$

Exchanged Examples Ratio. The exchanged examples ratio (EER) corresponds to the number of examples that have been sent from one agent to the other through messages, divided by the number of examples in the overall context.

Coverage Ratio. The coverage ratio (CR) of an agent corresponds to the number of examples from the overall context that can be associated with a sign by that agent through left-path associations, divided by the number of examples in the overall context.

Observed Disagreements. The observed disagreements (OD) are proper disagreements that can be observed between two pairs of agents' contrast sets at a given time; we measure their types and number. The OD are different from the SD (that are not proper disagreements). The OD can can be measure either before or after the argumentation.

4.2. *Experimental Setup*

The generality property of our model is tested over the Soybean data-set. In our experiment, we are only considering the classes that are supported by at least $2 \times \tau_E$ examples, in order to have enough examples for each agent's concepts. This leaves 290 examples distributed over 15 classes. These examples and the sign associated to their class are then split into two disjoint subsets as right-path associations. The proportion of each sign in each subset is kept at 1:1. Then, we set-up each type of disagreement by modifying the sign associated to each example differently in each subset.

Four types of disagreements are being setup in each experiment: overlap, hypo/ hypernymies, synonymies and homonymies. These four types are randomly ordered for each experiment, and a random number of each disagreement type is selected and then setup, one type after the other. For instance, if the overlap type has been ordered first, we select a random number between 0 and the number of classes of Soybean with more than τ_E examples divided by three (the number of classes required to setup an overlap disagreement) and we merge as many times three random classes in order to obtain as many overlap disagreements. For each of the 200 runs of the experiment, we setup a new random arrangement of disagreements.

Hypo/hypernyms are set up by replacing a pair of signs by a unique sign in one of the two subsets. Overlaps are set up by replacing a pair of signs in each subset by a unique sign, one of the replaced signs being shared by both pairs. Synonyms are set up by replacing a sign by a new sign in one of the two subsets. Homonyms are set up by replacing two signs from different subsets by one unique sign.

Once these disagreements have been set up, the two subsets are used by two agent as a training set to learn its initial contrast set through concept creation using right-path associations. The examples of a subset become the context of the agent, while the signs of the subset become its lexicon. After they both learned their initial contrast set, a token is randomly given to one agent and the argumentation starts. We tested 200 random setups, followed each time by one argumentation process.

4.3. Experimental Results

Figure 2 shows the results of our experiments measuring SAR, DAR, EER and CR before and after the argumentation. The left hand plot shows the average of each measure over the 200 random setups. In the case of DAR and CR, they are also averaged over the corresponding individual measure of both agents. We can see that the agents significantly moved toward a synchronic agreement after an argumentation, while diachronic agreement is not compromised. Only a portion of all the examples in the context have been exchanged, proving that the agents are really reaching mutual intelligibility by argument exchange, and not by exchanging all of their contexts.The coverage ratio (CR) average decreases after the argumentation, meaning that less examples can be classified. However, CR average remains over 0.8, meaning that the global classification of our agents is not over-fitted on their respective contexts.

The right hand plot in Figure 2 shows the average distance between the individual measures of SAR, DAR, EER and CR. SAR is calculated locally by using the local context of each agent instead of the overall context. EER uses the count of the examples *sent* by each agent, and CR uses the count of the examples covered by at least one of each agent's concepts. Recall that DAR is essentially an individual measure. The scale on which these distances range goes from 0 to 0.08, meaning that for each measured ratio, the difference between the two corresponding individual ratios is two orders of magnitude smaller.

Figure 3 shows the count of observed disagreements at the beginning (left side) and at the end (right side) of the argumentation, averaged over the 200 different setups. As expected, new disagreement types have appeared that are not limited to the four types included in the setups by us. The plots on the top row show the overall disagreements count, while the bottom row shows the local disagreements count of an individual agent. Since the argumentation protocol is symmetric and the disagreement setup is random, the other agent displays a similar profile. We count zero overall disagreements after the argumentation, showing the general effectiveness of our approach.

Local disagreements may still remain: as local contexts are different from the overall context, different pairing relations hold in local contexts than in the overall context. For instance, local synonymy disagreements are detected after the argumentation. They are due to a set of concepts that have different signs, but they all have empty adjunct sets in one local context and therefore are seen as equivalent to one another. Two equivalent concepts that have two different signs cause a synonymy disagreement. If this situation arises in one local context but not in the other, one agent perceives a local synonymy disagreement while the other agent does not. In the overall context, the two involved concepts each have proper examples in their adjunct sets and therefore they are not equivalent, so no overall synonymy disagreement arises.

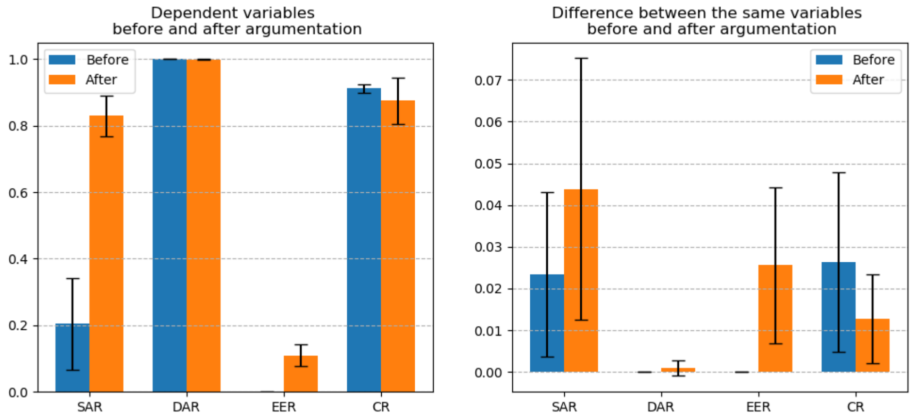

Figure 2. The left plot represents five different types of ratio measured on the overall context before and after argumentation: *synchronic agreement* ratio, *diachronic agreement* ratio, *example exchanges* ratio and *coverage* ratio. Each of these ratio can also be measured on the local context of one agent. The right plot represents, for each type of ratio presented in the left plot, the distance between the two ratios measured on the two local contexts of the agents.

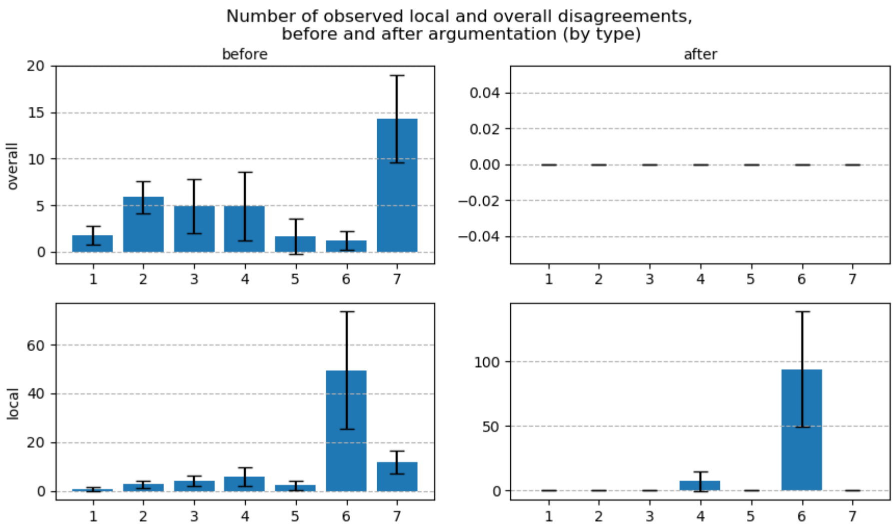

Figure 3. The plots show the number of disagreements counted before (left-hand) and after (right hand) the argumentation, and using (below) the local contexts and (above) the overall context. The x-axis shows the disagreement types labeled as follows: 1. Self-Disagreements, 2. Overlaps, 3. Hypo/hypernymies, 4. Synonymies, 5. Homonymies, 6. Indistinguishable disagreements, 7. Untranslatable disagreements

5. Discussion and Future Work

The experimental results show that our model can significantly improve the mutual intelligibility of two agents having multiple disagreements, while keeping their new shared classification as a refinement of both old ones. Moreover, the agents do not need to exchange an important number of examples from their contexts in order to reach this mutual intelligibility. The mutual intelligibility is attained with low variance, meaning that the combinations of disagreements that our agents have to effectively resolve does not impair the ability of our system to reach the mutual intelligibility.

The model used in the experiments uses a systematic strategy to search for and resolve every overall disagreements between the agents. The results of another strategy, a lazy strategy that takes place in a naming game and that triggers an argumentation only when two agents encounter inconsistent classifications of an example during their naming game, will be presented in later publications. The hypothesis of generality is not the only hypothesis tested on our model. Later publications will include tests of: the *domain independence* hypothesis, testing our model over different data sets, the *scalability* hypothesis, that our model is able to achieve mutual intelligibility over the contexts of increasing sizes of a same domain without increasing exponentially the number of arguments exchanged, the *simplicity* hypothesis, that the number of concepts in the final contrast sets of our agents are not larger than the number that we could expect from a brute force alignment after a transfer of all the examples from one agent to another, and finally the *paradigm shift* hypothesis, that the SAR and the DAR achieved by an AMAIL argumentation after an argumentation using our model is better than the SAR and DAR achieved by an AMAIL argumentation alone, while the total computation time of our model plus an AMAIL run is lesser than an AMAIL run alone.

Future work will focus on applying our model to different types of classifiers than inductive learners, most likely deep neural networks. Possible approaches to achieve this goal have already been discussed in [1].

Acknowledgements

This research has been partially funded by the project ESSENCE: Evolution of Shared Semantics in Computational Environments (ITN 607062).

References

[1] Adrian, Kemo, From Learning Through Signal Processing to Argumentation on Ontological Representations, *Workshop on Semantic Deep Learning* (SemDeep). October 2017.

[2] Adrian, Kemo, and Plaza, Enric, An Approach to Interaction-Based Concept Convergence in Multi-Agent Systems, *The Joint Ontology Workshops Proceedings*. May 2017.

[3] Carpenter, B, Typed Feature Structures: An Extension of First-Order Terms. In V.Saraswat & K.Ueda (Eds.), *Logic programming: Proceedings of the 1991 international symposium*, 1991, p. 187–201.

[4] Euzenat, J, Interaction-based Ontology Alignment Repair with Expansion and Relaxation, *IJCAI*, 2017, p.185-191

[5] Field, H, Logic, Meaning and conceptual role, *The Journal of Philosophy*, 1977, vol. 74(7), p. 379-409.

[6] Harnard, S, The Symbol Grounding Problem, *Physica D: Nonlinear Phenomena*, 1990, vol. 42(1-3), p. 335-346.

[7] Laera, L, Blacoe, I, Tamma, V, Payne, R, Euzenat, J, Bench-Capon, T, Argumentation over Ontology Correspondences in MAS, *AAMAS*, 2007, p. 228

[8] Ontañón, Santiago, and Plaza, Enric, Coordinated Inductive Learning using Argumentation-based Communication, *Autonomous Agents and Multiagent Systems Journal*, March 2015, Vol 29, Issue 2 p. 266-304.

[9] Ontañón, Santiago, and Plaza, Enric, Similarity Measures over Refinement Graphs, *Machine Learning Journal*, 2012, 87(1): p.57-92.

[10] Charles Sanders Peirce. Collected Papers. *Harvard U.P*, 1931-1958.

[11] Rapaport, W, Holism, Conceptual-role Semantics, and Syntactic Semantics, *Minds and Machines*, 2002, vol. 12(1), p. 3-59.

[12] Trojahn dos Santos, C, Quaresma, P, Vieira, R, An Argumentation Framework Based on Confidence Degrees to Combine Ontology Mapping Approaches, *IJMSO*, 2008, vol. 3(2), p. 142-150