Vocabulary Alignment in Open and Heterogeneous Interactions: is it Possible?

Work in Progress

Paula Chocron
IIIA-CSIC
Universitat Autònoma de Barcelona
Bellaterra, Catalonia, Spain
pchocron@iiia.csic.es

ABSTRACT

Addressing the problem of vocabulary heterogeneity is necessary for the common understanding of agents that use different languages, and therefore crucial for the success of multi-agent systems that act jointly by communicating. In recent work, we have studied this problem from a new perspective, that does not require external knowledge or any previously shared meta-language. Instead, we assume that agents share the knowledge of how to perform the tasks for which they need to collaborate, and we show how they can learn alignments from repeated interaction. Importantly, in that work we require agents to share the *complete* knowledge of the task. In this extended abstract we present a sketch of an extension that would allow to consider, in a meaningful way, differences between the agents' specifications. To this aim, we propose a new kind of protocols with constraints that have weights to represent a punishment received when they are violated.

1. VOCABULARY ALIGNMENT FROM THE EXPERIENCE OF INTERACTION

The problem of aligning the vocabularies of heterogeneous agents to guarantee mutual understanding has been tackled several times in the past two decades, in general from one of two different perspectives. Some approaches [7, 4] consider the existence of external *contextual* elements, such as physical objects, that all agents perceive in common, and explore how those can be used to explain the meaning of words. A second group of techniques [5, 6] consider the situation, reasonable for agents that communicate remotely, in which this kind of context is not available. They do so by providing explicit ways of learning or agreeing on a common vocabulary (or alignment between vocabularies), that can include argumentation techniques, explanations, or definitions. These techniques always require agents to share a common meta-language. The question of how to communicate with heterogeneous interlocutors when neither a physical context nor a meta-language are available remains practically unexplored.

Appears in: *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

In recent work [1, 2, 3] we proposed a different approach, where the alignment is performed considering only the context given by the interactions in which agents engage. Agents are assumed to share the *knowledge of how to perform a task*, or, more concretely, the specification of an interaction. As an example, consider an *ordering drinks* interaction between an English speaking customer and an Italian waiter. We assume that both agents know the dynamics of the conversation (for example, that the customer can order wine and/or beer, only if they are asked the question *"what would you like to drink?"*, and that the waiter will ask for the color if wine is ordered). However, the words that are used are different (*vino* and *birra* instead of *wine* and *beer*). In the cited work, we show how agents can progressively learn which mappings lead to successful interactions from the experience of performing the task. After several interactions, agents converge to an alignment that they can use to always succeed at ordering and delivering drinks with that particular interlocutor.

In [3] the interactions are specified with *open protocols* that define linear temporal logic (LTL) constraints about what can be said. In this way, the ordering drinks scenario could be specified with the following two protocols, where W is the waiter, C is the customer (and $\mathfrak{P}_W, \mathfrak{P}_C$ their respective protocols), $say\colon A \times V$ (with A a set of agent names and V a vocabulary) is a predicate such that $say(a, v)$ is true if a says v at a given time, and $\Diamond, \Box, \bigcirc$ are the LTL operators that mean *eventually*, *globally* and *next* respectively.

$$\begin{gathered}\mathfrak{P}_W = \{\Diamond\, say(W, \textsf{da bere}),\\ \Box\,(\bigcirc\, say(C, \textsf{birra}) \to say(W, \textsf{da bere})),\\ \Box\,(\bigcirc\, say(C, \textsf{vino}) \to say(W, \textsf{da bere})),\\ \Box\,(say(C, \textsf{vino}) \to \Diamond\, say(W, \textsf{colore}))\}\end{gathered}$$

$$\begin{gathered}\mathfrak{P}_C = \{\Diamond\, say(W, \textsf{to drink}),\\ \Box\,(\bigcirc\, say(C, \textsf{beer}) \to say(W, \textsf{to drink})),\\ \Box\,(\bigcirc\, say(C, \textsf{wine}) \to say(W, \textsf{to drink})),\\ \Box\,(say(C, \textsf{wine}) \to \Diamond\, say(W, \textsf{color}))\}\end{gathered}$$

The approach for learning alignments from interactions is simple. Agents maintain a confidence distribution that assigns a value to each mapping between a foreign word and a word in their vocabulary. These values are updated according to what agents observe in interactions. Briefly, when an agent receives a word, it punishes all interpretations that are not possible because they violate some constraint. For example, if the customer receives *colore* right after saying *wine*, it infers that it can not mean *to drink*. By interacting repeatedly with different protocols, agents gradually learn an alignment between their vocabularies.

Until now, we required agents to share the entire structure of the interactions they perform. We do so by defining a notion of *compatibility* between protocols: two protocols are compatible if they accept exactly the same interactions as correct, modulo an alignment. Then, we require our agents to have only pairs of protocols that are compatible under one alignment. Of course, this raises an immediate question: what can agents learn if they do not share the protocol specifications? The short answer is that, if the protocols differ significantly, they have nothing to learn, since there is no alignment that is useful to perform the tasks together. If only some protocols differ, and in details, they can still infer an alignment with the same technique (although more slowly), since the learning method can automatically fix things that were wrongly learned.

2. WEIGHTED PROTOCOLS

We now propose an approach that considers more carefully the question of whether agents can align their vocabularies when their protocols are different. To this aim, we introduce a new version of these protocols, in which each constraint has a weight that represents a punishment received when that constraint is violated. This punishment can be interpreted, for example, as a way of expressing preferences (heavier constraints are those that agents prefer not to violate), or degrees of confidence on a constraint, when there is uncertainty about the interaction context.

A weighted protocol over a vocabulary V and a set of agents A is a set $\mathfrak{P}$ of pairs $\langle c, \rho \rangle$, where c is a LTL constraint over instantiations of $say\colon A \times V$, and $\rho \in [0,1]$. As an example, consider again the waiter and the customer. Assume they have the same constraints as before with high weight, but now the waiter also believes that the customer should not order two different alcoholic beverages in one interaction. This constraint, however, is less strict than the others, since the waiter is willing to accept that behaviour some times. The protocols would look as follows.

$$\begin{gathered}\mathfrak{P}_W = \{\langle \Diamond\, say(W, \textsf{da bere}), 1\rangle,\\ \langle \Box\, (\bigcirc\, say(C, \textsf{birra}) \rightarrow say(W, \textsf{da bere})), 1\rangle,\\ \langle \Box\, (\bigcirc\, say(C, \textsf{vino}) \rightarrow say(W, \textsf{da bere})), 1\rangle,\\ \langle \Box\, (say(C, \textsf{vino}) \rightarrow \Diamond\, say(W, \textsf{colore})), 1\rangle,\\ \langle \Diamond\, say(C, \textsf{birra}) \rightarrow \neg \Diamond\, say(C, \textsf{vino}), 0.5\rangle\}\end{gathered}$$

$$\begin{gathered}\mathfrak{P}_C = \{\Diamond\, say(W, \textsf{to drink}), 1\rangle,\\ \langle \Box\, (\bigcirc\, say(C, \textsf{beer}) \rightarrow say(W, \textsf{to drink})), 1\rangle,\\ \langle \Box\, (\bigcirc\, say(C, \textsf{wine}) \rightarrow say(W, \textsf{to drink})), 1\rangle,\\ \langle \Box\, (say(C, \textsf{wine}) \rightarrow \Diamond\, say(W, \textsf{color})), 1\rangle\}\end{gathered}$$

Since we do not require protocols to be compatible in any way, there is no single *correct* alignment that agents need to find. Instead, we can define a measure of adequacy of an alignment to a pair of protocols, or to a set of pairs of protocols. We propose a first approach to define this measure. Given $\mathfrak{P}_1$ and $\mathfrak{P}_2$ over V_1 and V_2 respectively, the adequacy of an alignment $\mathcal{A}$ between V_1 and V_2 can be measured for $\mathfrak{P}_1$ as follows. Consider an interaction i given by a sequence of messages that agents have sent to each other (formally, a sequence of pairs $\langle agent, word \rangle$). Let $\rho(\mathfrak{P}_2, i)$ be the punishment for i in $\mathfrak{P}_2$, this is, the added weight of all violated constraints. Then let $Poss(\mathfrak{P}_2)$ be all the interactions for which $\rho(\mathfrak{P}_2, i) = 0$. Then the adequacy of the alignment $\mathcal{A}$ for $\mathfrak{P}_1$ is

$$\frac{\sum_{i \in Poss(\mathfrak{P}_2)} \rho(\mathfrak{P}_1, \mathcal{A}(i))}{|Poss(\mathfrak{P}_2)|}$$

where $\mathcal{A}(i)$ is the translation of i via $\mathcal{A}$. Notice that this measure is unilateral, and the adequacy of $\mathcal{A}$ for $\mathfrak{P}_2$ is not necessarily the same in the other direction.

The second aspect to take into account is how agents update their confidences in mappings between words from the experience of interacting. A first simple approach uses the punishment that would correspond to each interpretation. Let $\omega(v', v)$ be the previous confidence value an agent has for the mapping between v and v'. Suppose a_2 receives a word v_1 from a_1 after interaction i. Then, for all $v \in V_2$,

$$\omega(v_1, v) := \omega(v_1, v) - \rho(\mathfrak{P}_1, i.\langle a_1, v\rangle)$$

where $i.\langle a_1, v\rangle$ is the interaction obtained by appending the message v sent by a_1 to i.

Research Questions.

Although we still need to work towards a stable framework, we think there are interesting questions that could be explored considering this kind of protocols, such as:

- What kind of updating strategies lead agents to their most adequate alignment?
- How does the distribution of different constraints in protocols affect the convergence to an alignment? And the distribution of weights, or the frequency with which each task is performed?

REFERENCES

[1] M. Atencia and M. Schorlemmer. An interaction-based approach to semantic alignment. *Journal of Web Semantics*, 12-13:131–147, 2012.

[2] P. Chocron and M. Schorlemmer. Attuning ontology alignments to semantically heterogeneous multi-agent interactions. In *ECAI 2016 - 22nd European Conference on Artificial Intelligence, The Hague, The Netherlands*, pages 871–879, 2016.

[3] P. Chocron and M. Schorlemmer. Vocabulary alignment in openly specified interactions. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017) (To appear)*, 2017.

[4] J. V. Diggelen, R. Beun, F. Dignum, R. M. V. Eijk, and J. J. Meyer. Ontology negotiation: Goals, requirements and implementation. *Int. J. Agent-Oriented Softw. Eng.*, 1(1):63–90, Apr. 2007.

[5] G. Santos, V. Tamma, T. R. Payne, and F. Grasso. A dialogue protocol to support meaning negotiation. (extended abstract). In *Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '16, pages 1367–1368, Richland, SC, 2016.

[6] N. Silva, G. I. Ipp, and P. Maio. An approach to ontology mapping negotiation. In *Proceedings of the Workshop on Integrating Ontologies*, pages 54–60, 2005.

[7] L. Steels. The origins of ontologies and communication conventions in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 1(2):169–194, Oct. 1998.