Ontology Alignment Evaluation in the Context of Multi-Agent Interactions

Paula Chocron and Marco Schorlemmer

Artificial Intelligence Research Institute, IIIA-CSIC
Bellaterra (Barcelona), Catalonia, Spain **

Abstract The most prominent way to assess the quality of an ontology alignment is to compute its precision and recall with respect to another alignment taken as reference. These measures determine, respectively, the proportion of found mappings that belong to the reference alignment and the proportion of the reference alignment that was found. The use of these values has been criticised arguing that they fail to reflect important semantic aspects. In addition, they rely on the existence of a reference alignment. In this work we discuss the evaluation of alignments when they are used to facilitate communication between heterogeneous agents. We introduce the notion of pragmatic alignment to refer to the mappings that let agents understand each other, and we propose new versions of precision and recall that measure how useful mappings are for a particular interaction. We then discuss practical applications of these new measures and how they can be estimated dynamically by interacting agents.

1 Introduction

Communication between heterogeneous agents has been identified as one important application for ontology alignments [9]. In dynamic and open environments such as multi-agent systems, agents with multiple backgrounds may not share their vocabularies or representations of meaning. Even when a common vocabulary is established, maintaining it over time can be a difficult task, particularly in dynamic domains [4]. To achieve meaningful communication it is therefore necessary to develop techniques that align the vocabularies that agents use, obtaining a translation that allows them to interpret the messages they receive correctly. If agents organise their vocabularies in some kind of taxonomy or ontology, a very reasonable approach is to take advantage of the diverse ontology alignment tools that were developed in the last decades [9]. However, language used in agent communication has its own particularities that should be taken into account when using alignments for this purpose; mainly, language is contextualised in the concrete interaction agents are performing. General purpose

** `pchocron,marco@iiia.csic.es`. This research has been funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 607062 /ESSENCE: Evolution of Shared Semantics in Computational Environments/.

ontology matchers do not take this into account, and despite being an important application, there is little research on the creation and use of alignments for agent interaction.

In this paper we focus on the problem of using ontology alignments as translators for agent communication, and particularly on their evaluation for that application. We are interested in developing measures to decide whether an alignment is useful for a particular interaction, that is, if using it will help agents communicate. Traditionally, ontology alignments are evaluated with respect to a human-crafted reference alignment, and accuracy measures count the elements in the intersection between the evaluated alignment and the reference. In this way, the *precision* of an alignment is defined as the proportion of found mappings that belong to the reference alignment, while the *recall* is the proportion of the reference alignment that was found. We propose an application-dependent evaluation technique that does not require the (possibly arbitrary) construction of a gold standard. In this way, we make a step towards considering the problem of "in situ evaluation", based on the idea that *"the relative quality or usefulness of a generated alignment also depends on its intended use"* [8].

Our approach considers agents taking part in task-oriented interactions, and defines a mapping as correct if it allows agents to finish the joint task successfully. This leads to the notion of *useful* and *misleading* mappings, which are, respectively, those that lead to the success or failure of an interaction. This new classification allows us to redefine the traditional precision and recall measures that are used for alignment evaluation, comparing an alignment against the specification of an interaction, thus providing a method for evaluating alignments that does not rely on a human-built alignment. We then show how these newly defined measures can be used by agents to improve the quality of their understanding, and sketch a method in which agents can estimate them online using their experience from interaction, making evaluation automatic.

2 Related Work

The use of the standard precision and recall notions from information retrieval for the evaluation of ontology alignments has been criticised by different authors, all of whom argue that these measures ignore important aspects of the problem that should also be taken into account to decide how good a solution is. The main approach to creating measures that are more appropriated for the nature of semantic mappings is the one of *semantic* precision and recall [6]. Here, Euzenat tackles the problem of the binary nature of traditional precision and recall (if a mapping is not found by the alignment, it is missing), by considering the relation between the consequences of the alignments instead of between the alignments themselves. In [11], Holling et al. propose new evaluation measures that take into account the frequency of use of the mappings found, as well as the semantic distance to an alignment. In [13], the authors introduce the notion of *relevance* of a mapping, that measures how often the mapped words appear in a particular context.

Also relevant are approaches that consider the use and evolution of alignments in a multi-agent environment. Both [10] and [2] propose methods to create alignments from scratch that are learned from the agent's interaction experience. In [7] and in [5] the authors propose techniques to repair alignments with information that is learned directly from observations made while interacting. A similar idea is proposed in [12], but in this case agents repair their ontologies instead of alignments.

3 A Pragmatic Approach to Alignment

We consider the problem of achieving meaningful communication between two agents a_1 and a_2 that need to interact to perform some task, but use potentially different vocabularies V_1 and V_2 respectively. Each agent can organise its vocabulary in its own way, using structures that go from simple lists of words to fully fledged ontologies. We only suppose that they can be matched with one of the existing tools to obtain an alignment between them.

Definition 1. *An* alignment $\mathcal{A}$ *between two vocabularies* V_1 *and* V_2 *is a finite set of* mappings *between words in* V_1 *and* V_2. *A mapping is defined as a quadruple* $\langle v_1, v_2, n, r\rangle$, *where* $v_1 \in V_1$, $v_2 \in V_2$, $n \in (0, 1]$ *is the degree of confidence on the mapping, and* r *is the kind of relation that holds between words. An alignment must contain at most one tuple for each pair* $v_1 \in V_1$, $v_2 \in V_2$. *[3]*

When working with an alignment $\mathcal{A}$, if a mapping $\langle v_1, v_2, n, r\rangle$ belongs to $\mathcal{A}$ we will write $v_1 \, r \, v_2$ (for example, $v_1 \equiv v_2$).

In general, techniques to build alignments between different vocabularies make use of the structure or additional information in the ontologies in which such vocabularies are organised. Other techniques use external resources, such as text corpora or the web. Still others have a completely syntactic approach. Extending the ideas in [2], we propose a different kind of alignments, that we call *pragmatic*. This kind of alignments are produced by only taking into account the interactions in which agents use their vocabularies. Let us first define the specifications of interactions, and then move to formalise the alignments.

3.1 Interaction Specifications

We specify interactions performed jointly by agents by means of *interaction protocols* that define all possible sequences of message exchanges. The multi-agent systems community has extensively discussed possible formalisms to describe these kind of protocols; in this work we stick to a generic approach that uses Finite State Automata. Since we focus on agents that communicate to perform a task together (for example, *ordering drinks*), the interaction can end successfully (if the task is completed) or can fail (if it is not). To decide this outcome, we introduce the notion of *state properties*, which are Boolean predicates assigned to final states to represent observations. Interactions are successful only if agents reach together final states with the same properties.

Definition 2. *Given two agents a_1 and a_2, a vocabulary V, and a set of state properties SP, an* interaction model IM *is defined as a tuple $\langle Q, q_0, \delta, F, \rho, speaks \rangle$ where Q is a finite set of states, $q_0 \in Q$ is the initial state, $F \subseteq Q$ is the set of final states, $\rho : F \to \mathcal{P}(SP)$ assigns a subset of state properties to each final state, $speaks : Q \to \{a_1, a_2\}$ assigns to each state its sender agent, and $\delta : Q \times V \to Q$ is a partial function called the* transition function.

Note that while we do not specify any particular turn-taking pattern, we do require that, for each state, all messages labelling transitions from this state share the same sender agent, who is determined with the *speaks* function. For simplicity reasons, we will consider that δ is undefined for the final states F.

In the rest of this paper, including all the definitions, we consider interactions between two agents a_1 and a_2 with interaction models $IM_i = \langle Q_i, q_i^0, F_i, \delta_i, \rho_i, speaks_i \rangle,\ i = 1, 2$. While IM_1 and IM_2 have the same set of agents ($\{a_1, a_2\}$), their vocabularies and state properties can differ; we will call them V_1, V_2 and SP_1, SP_2 respectively.

3.2 Pragmatic Alignments

Alignments between the vocabularies of two interaction models, that we will call *pragmatic alignments*, capture relations between the ways in which words are used in a conversation. In this way, a word v_1 from IM_1 matches with a word v_2 from IM_2 if an agent can interpret v_1 as v_2 in an interaction and finish the task successfully.

Definition 3. *Consider IM_1 and IM_2 such that $speaks(q_1^0) = speaks(q_2^0)$. Extending [1], the* communication product *of IM_1 and IM_2 ($IM_1 \otimes IM_2$) is an interaction model $\langle Q, q^0, F, \delta, \rho, speaks \rangle$ over a language V that is the Cartesian product between V_1 and V_2, a set of agents $\{a_1, a_2\}$, and $SP = \{success, failure\}$, and such that:*

- *Q is a subset of the Cartesian product of Q_1 and Q_2 in which both states have the same senders, in other words, the states in Q are all possible ordered pairs $\langle q_1, q_2 \rangle$ with $q_1 \in Q_1,\ q_2 \in Q_2$, and $speaks_1(q_1) = speaks_2(q_2)$*
- *speaks is the speaker in q_1 or q_2: $speaks(\langle q_1, q_2 \rangle) = speaks_1(q_1) (= speaks_2(q_2))$*
- *the initial state q^0 is the pair $\langle q_1^0, q_2^0 \rangle$*
- *δ is defined as follows: $\langle q_1', q_2' \rangle = \delta(\langle q_1, q_2 \rangle, \langle v_1, v_2 \rangle)$ if $\delta_i(q_i, v_i) = q_i'$ for $i \in \{1, 2\}$*
- *F are all states in Q for which δ is not defined*
- *For $\langle q_1, q_2 \rangle \in F$, $\rho(\langle q_1, q_2 \rangle) = \{success\}$ if $q_1 \in F_1, q_2 \in F_2$, and $\rho_1(q_1) = \rho_2(q_2)$. It is $\{failure\}$ otherwise.*

With this construction, we can easily obtain all possible interactions between agents with two interaction models.

Definition 4. *An* interaction *between two interaction models IM_1, IM_2 is an accepted string in the communication product IM between IM_1 and IM_2. An interaction is* successful *if it ends in a state q such that $\rho(q) = \{success\}$, it is* unsuccessful *if $\rho(q) = \{failure\}$.*

These interactions can be seen as all possible combinations of uttered messages and their interpretations; our objective is to use them to define pragmatic alignments. An immediate approach consists in considering two words as equivalent if they belong to a successful interaction. In an alignment of this kind, one word in V_1 could be mapped to many words in V_2 if they have different interpretations in different states. Instead, agents will be interested in knowing which mapping is correct for each state. This information can be obtained from successful interactions if we consider deterministic FSAs in which any accepted string can be assigned to an unique sequence of states. In the following definition, mappings are parametrised by states in the communication product.

Definition 5. *A* pragmatic alignment *between interaction models IM_1, IM_2 is a set of tuples $\langle q, v_1, v_2, r\rangle$, where $q \in Q, v_1 \in V_1, v_2 \in V_2$, and $r \in \{\equiv, \between\}$.*

The relation between two words ($\equiv$ or $\between$) depends on whether finishing the interaction successfully is always possible after mapping them. To define formally their semantics, we will refer to each state in one of these accepted strings as $\langle q, v\rangle$, representing the state and the message.

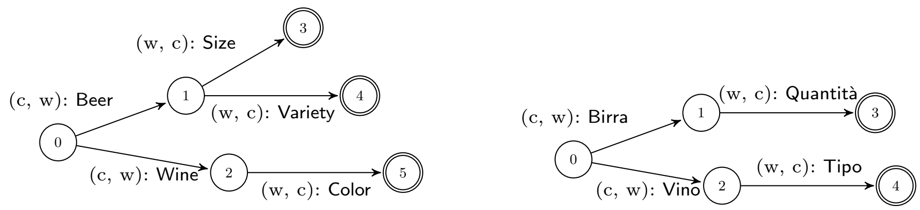
- $IM_1, IM_2 \models \langle\langle q_1, q_2\rangle, v_1, v_2, \equiv\rangle$ if there are interactions between IM_1 and IM_2 that include $\langle\langle q_1, q_2\rangle, \langle v_1, v_2\rangle\rangle$, and all strings accepted by IM_1 or IM_2 that include $\langle q_1, v_1\rangle$ or $\langle q_2, v_2\rangle$ are the projection of one of these interactions (the interaction can always end successfully after mapping v_1 with v_2).
- $IM_1, IM_2 \models \langle\langle q_1, q_2\rangle, v_1, v_2, \between\rangle$ if there exists at least one successful interaction between IM_1 and IM_2 that includes $\langle\langle q_1, q_2\rangle, \langle v_1, v_2\rangle\rangle$ (the interaction can end successfully at least for some cases after mapping v_1 with v_2).

As an example, consider the interaction models in Figure 1, which represent fragments of interactions between a waiter (w) and a customer (c) to order drinks in English and Italian (state transitions should be read as $(sender, receiver)$: *message*.). Let IM_1 have $SP : \{size\_beer, kind\_beer, kind\_wine\}$, and IM_2 have $SP : \{kind\_beer, kind\_wine\}$, and $\rho_1(3) = size\_beer$,$\rho_1(4) = \rho_2(3) = kind\_beer$, $\rho_1(5) = \rho_2(4) = kind\_wine$. The mapping $\mathsf{Wine} \equiv \mathsf{Vino}$ in $\langle 0, 0\rangle$ is satisfied by IM_1, IM_2, because the interaction $(a_1 : \langle \mathsf{Wine}, \mathsf{Vino}\rangle, a_2 : \langle \mathsf{Color}, \mathsf{Tipo}\rangle)$ is successful in the communication product, and all accepted strings in IM_1 and IM_2 that include $mathsfWine$ and $mathsfVino$ respectively are projections of it. The mapping $\mathsf{Beer} \equiv \mathsf{Birra}$ in $\langle 0, 0\rangle$ is not, because there is no interaction that projects $(\mathsf{Beer}, \mathsf{Size})$. However, $\mathsf{Beer} \between \mathsf{Birra}$ is satisfied, because $(\langle \mathsf{Beer}, \mathsf{Birra}\rangle, \langle \mathsf{Variety}, \mathsf{Tipo}\rangle)$ is successful.

Pragmatic alignments are everything agents need to communicate successfully, but they are only useful in a particular context. Notice, for example, that mapping Tipo with Color is not correct in a general English-Italian translation; however in the context of ordering drinks it yields to common understanding.

4 Pragmatic Evaluation of Alignments

The quality of a vocabulary alignment is typically measured in comparison with a *reference alignment*, for which values of *precision* and *recall* are computed. As



(a) $IM_1 = \langle Q_1, q_1^0, \delta_1, F_1, \rho_1 \rangle$ (b) $IM_2 = \langle Q_2, q_2^0, \delta_2, F_2, \rho_2 \rangle$

Figure 1: Fragments of interaction models for ordering drinks

it is commonly done, we do not take into account the confidence degrees in these measures.

Definition 6. *Given an alignment $\mathcal{A}$, let $\mathcal{A}'$ denote the set of mappings of $\mathcal{A}$ for which we have removed the confidence degree, i.e., $\mathcal{A}' = \{\langle v_1, v_2, r\rangle \mid \langle v_1, v_2, n, r\rangle \in \mathcal{A}$ for some $n\}$. The* precision *of an alignment $\mathcal{A}$ with respect to a reference alignment $\mathcal{B}$ is the fraction of the mappings in $\mathcal{A}'$ that are also in $\mathcal{B}'$:*

$$precision(\mathcal{A}, \mathcal{B}) = \frac{\mid \mathcal{A}' \cap \mathcal{B}' \mid}{\mid \mathcal{A}' \mid}$$

while its recall *is the fraction of the mappings in $\mathcal{B}$ that were found by $\mathcal{A}$:*

$$recall(\mathcal{A}, \mathcal{B}) = \frac{\mid \mathcal{A}' \cap \mathcal{B}' \mid}{\mid \mathcal{B}' \mid}$$

Two problems arise when using these measures to assess the quality of an alignment $\mathcal{A}$ used for agent interaction. First, a reference alignment between the vocabularies may not be available. Second, even if it is, the measures do not take into account the way in which terms are used in an interaction. To show this, we performed a small experiment, based on the ones in [5], and let agents with heterogeneous vocabularies interact using alignments of different qualities. In Figure 2, we can see that recall is more relevant than precision; this is because the alignment counts as correct many mappings that are not actually necessary for interacting.

In this section we propose adaptations of the traditional precision and recall measures that evaluate an alignment taking as reference, not a human-crafted standard, but a pragmatic alignment obtained from two interaction models. We introduce the notions of *useful* and *misleading* mappings for those that belong to successful and unsuccessful interactions respectively. In this first approach we will only consider alignments with $\equiv$ relations, the problem of analysing other relations is left for future work.

Definition 7. *Consider an alignment $\mathcal{A}$ between vocabularies V_1 and V_2 and the already defined interaction models IM_1 and IM_2. A mapping $\langle v_1, v_2, n, \equiv \rangle \in \mathcal{A}$ is* useful *with respect to IM_1, IM_2 if $\langle v_1, v_2 \rangle$ appears in a successful*

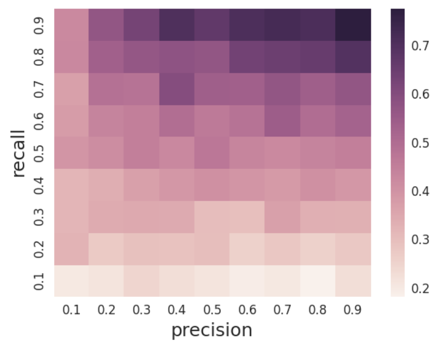


Figure 2: Success rates for different values of precision and recall

interaction between IM_1 *and* IM_2*. It is* misleading *if the same pair appears in an unsuccessful interaction between* IM_1 *and* IM_2*.*

Notice that there can be mappings in $\mathcal{A}$ that are neither useful or misleading. We will call *relevant* to the mappings that can be classified in useful or misleading, or equivalently, those between pairs that belong to an interaction between the models. More surprisingly, a mapping can be both useful and misleading at the same time, if the relation in the pragmatic alignment is $\between$. This allows for different possibilities when computing precision and recall. In this paper we consider as correct all useful alignments.

To define precision and recall for $\mathcal{A}$ with respect to IM_1, IM_2, let $useful$ and $relevant$ be, respectively, the sets of useful and relevant mappings of $\mathcal{A}$ with respect to the interaction models. Let $\mathcal{A}_p$ be the pragmatic alignment between IM_1 and IM_2, and let us define $pragmatic = \{\langle v_1, v_2, r\rangle$ if $\langle q, v_1, v_2, r\rangle \in \mathcal{A}_p$ for some $q \in Q\}$. Pragmatic precision and recall are defined as follows:

$$recall = \frac{\mid useful \mid}{\mid pragmatic \mid}$$

$$precision = \frac{\mid useful \mid}{\mid relevant \mid}$$

As argued in [11], we may want to take into account not only how many, but also which of the mappings are found by the alignment. Finding a correct mapping for a very common word should have more impact in the precision than finding a mapping for a rarely used one. This can be taken into account in the pragmatic precision and recall measures we just defined, by simply considering *useful* and *relevant* as multi-sets:

- *useful:* for each state $q \in Q$, all mappings in $\mathcal{A}_p$ that are useful in q
- *relevant:* for each state $q \in Q$, all mappings in $\mathcal{A}_p$ that are relevant in q

Precision is defined in the same way, and recall as:

$$recall = \frac{\mid useful \mid}{\mid \mathcal{A}_p \mid}$$

It is worth noting that, with these definitions, possible values for pragmatic precision and recall are determined by the structure of interaction models. For example, consider a linear interaction model in which each state has only one outgoing arrow. There are no possible misleading matches with this protocol; therefore the minimum level of precision for alignments is necessarily 1.

4.1 An Example: ordering drinks

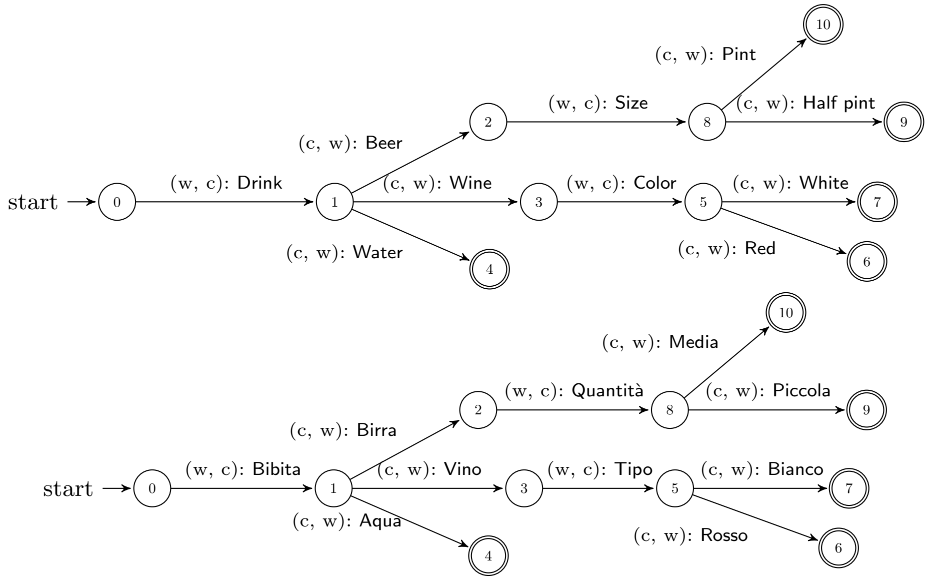


Figure 3: English and Italian interaction models for ordering drinks

Consider the alignments presented in Table 1 applied to the ordering drinks scenario represented by the protocols in Figure 3. According to an English-Italian dictionary, they would both have precision 0.5 (Wine $\equiv$ Vino and Red $\equiv$ Rosso are correct). Depending on the way of using the dictionary, Media $\equiv$ Half Pint could also be considered correct, giving the second alignment a precision of 0.75. However, they are clearly not equally useful when used by agents interacting, because the second alignment has a misleading mapping Media $\equiv$ Half Pint. Using our values, both alignments have a recall of 0.2 (Wine $\equiv$ Vino, Red $\equiv$ Rosso are the useful alignments found), but the first one has a precision of 1 and the second one of 0.66.

Alignment 1	
$v_1 \in V_1$	$v_2 \in V_2$
Bibita	Water
Vino	Wine
Rosso	Red
Quantità	Pint

Alignment 2	
$v_1 \in V_1$	$v_2 \in V_2$
Bibita	Water
Vino	Wine
Rosso	Red
Media	Half Pint

Table1: Two alignments for the Ordering Drinks example

5 Pragmatic Precision and Recall in Practice

In their pragmatic version, precision and recall are not only indicators of how useful an alignment is for a particular interaction, but can also be actively used by semantically heterogeneous agents to improve their mutual understanding. Methods to learn pragmatic alignments and to transform traditional alignments into pragmatic ones can be obtained by adapting the techniques developed in [2] and [5] respectively. In this section we focus on the practical application of the evaluation of pragmatic alignments. We first analyse how pragmatic precision can be used to improve automatic matching techniques, and then sketch a method in which agents can estimate them from the experience of interaction.

5.1 Using Pragmatic Precision and Recall

Consider an agent that interacts with another one using an alignment that it does not trust completely. If the agent translates the messages it receives by always following the alignment, it would very frequently fail to communicate when the alignment has any misleading mapping. To avoid this situation, the following heuristic can be used to decide when to follow the alignment and when to explore.

Matching Criterion.
Consider an agent a_1 with interaction model IM_1 and an alignment $\mathcal{A}$. When receiving v_2 in state $q_1 \in Q_1$, a_1 needs to decide how to interpret it, or which outgoing arrow from q_1 to follow. Let $U(q_1)$ be the set of all these possible interpretations. For each $v_1 \in U(q_1)$, a_1 computes the value of the mapping as:

$$\mathcal{V}(v_1, v_2) = \begin{cases} n & \text{if } \langle v_1, v_2, n, \equiv \rangle \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases}$$

let $\hat{\mathcal{V}}(v_1, v_2)$ be the normalized values for $v_1 \in U(q_1)$, and consider an exploration parameter $\alpha \in [0, 1]$. The criterion consists in choosing $v_1 \in U(q)$ with probability:

$$p(v_1) = \alpha\, \hat{\mathcal{V}}(v_1, v_2) + (1 - \alpha) \frac{1}{\mid U(q) \mid}$$

A reasonable question is how to choose a good value of α. It is easy to see that the values that give better results in terms of rate of successful interactions depend on the pragmatic precision of $\mathcal{A}$ with respect to IM_1 and the protocol IM_2 of the agent a_1 interacts with. If precision is high, agents should trust more on the alignment, if it is low they should rely more on the random exploration.

To show this, we performed a short experiment, in which we analyse the rate of success of interactions between agents that use different values of α and have alignments of different qualities. We used the customer and waiter agents from the example in Section 4.1 and let them interact for 150 times, measuring in how many cases they succeeded. As a simplification, we used only alignments that had the same values of precision and recall; this should be extended in future work to consider more realistic values. We defined three alignment quality levels: low (precision and recall 0.2), medium (precision and recall 0.5) and a high (precision and recall 0.8) quality. Figure 4 shows the results. As expected, when the alignment is good with respect to the interaction, best results are obtained with a high α, while for bad alignments it is better to make random choices. For medium quality, there is almost no difference, since the probability of a mapping being correct is similar to the one of choosing randomly the right option.

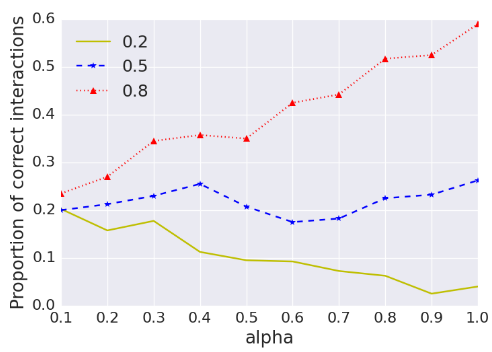


Figure 4: Success rates for different values of α

5.2 Estimating Pragmatic Precision and Recall

Although pragmatic precision and recall can be useful in practice, in most applications it is not realistic to expect agents to know them beforehand. In what follows we discuss how agents can use the experience of interaction to automatically estimate the values of precision and recall of an alignment. This would be useful not only to improve their behaviour as explained before, but also to evaluate alignments in a dynamic, distributed way.

Let us first focus on estimating recall. In this case, agents can simply use the proportion of the mappings they made in successful interactions that were already in $\mathcal{A}$.

$$recall_{est} = \frac{|\text{ mappings in successful interactions} \cap \mathcal{A} \text{ }|}{|\text{ mappings in successful interactions }|}$$

Estimating precision is more complicated. A first attempt could be to consider:

$$precision_{est} = \frac{|\text{ mappings in successful interactions} \cap \mathcal{A} \text{ }|}{|\text{ relevant mappings seen }|}$$

However, this considers as incorrect all the relevant mappings that were not part of successful interactions. This can sub-estimate the precision, particularly in the first steps, when an estimation is needed most.

Alternatively, we propose to use a learning strategy that estimates gradually the precision of $\mathcal{A}$ by analysing which of the mappings that were made are likely to be correct and which ones are not. A possibility is to use a technique proposed in [5], where all mappings start with a confidence equal to the one in $\mathcal{A}$ (or 0 if it is not a mapping in $\mathcal{A}$), and after an interaction they are updated as follows:

- After a successful interaction, the confidence in all mappings that were made is set to 1. These mappings are not updated in following interactions.
- After an unsuccessful interaction, a negative punishment is applied to the mappings made. At the same time, mappings are updated according to the quality of the aligning possibilities found later; if mappings with large confidence appeared as options after making one match, that match will increase its value.

To estimate precision, let *increased* be the set of all the mappings made that are in $\mathcal{A}$ and for which the calculated confidence is greater or equal to the one in $\mathcal{A}$. Precision can then be estimated as:

$$precision_{est} = \frac{|\text{ increased} \cap \mathcal{A} \text{ }|}{|\text{ relevant mappings seen }|}$$

This can improve the precision estimation in early stages, since mappings that are likely to be correct (because many good mappings were found after them) would still increase their value. These are preliminary ideas, that we plan to further develop and evaluate experimentally in future work.

6 Conclusions

We consider the ideas presented in this paper to be a first step towards the development of ontology alignment tools that are particularly designed for agent interaction. These tools would require novel reasoning techniques that take into account contextual information about the tasks that are being performed to build mappings of high pragmatic precision and recall. To this aim, a first technical requirement is the formalisation of a language that allows to express properties of the domain together with information about the interaction. To apply the ideas we propose here, it may be necessary to adapt them to more complex descriptions of interactions, or to incomplete ones.

References

1. Manuel Atencia and Marco Schorlemmer. Formalising interaction-situated semantic alignment: The communication product. In *Tenth International Symposium on Artificial Intelligence and Mathematics (ISAIM'08)*, Fort Lauderdale, Florida, USA, jan 2008.
2. Manuel Atencia and W. Marco Schorlemmer. An interaction-based approach to semantic alignment. *Journal of Web Semantics*, 12:131–147, 2012.
3. Paolo Bouquet, Jérôme Euzenat, Enrico Franconi, Luciano Serafini, Giorgos Stamou, and Sergio Tessaris. Specification of a common framework for characterizing alignment. Deliverable D2.2.1, Knowledge Web, 2004.
4. Alan Bundy and Fiona McNeill. Representation as a fluent: An ai challenge for the next half century. *IEEE Intelligent Systems*, 21(3):85–87, 2006.
5. Paula Chocron and Marco Schorlemmer. Attuning ontology alignments to semantically heterogeneous multi-agent interactions. In *Proceedings of the 22nd European Conference on Artificial Intellignce (ECAI)*, 2016 (to appear).
6. Jérôme Euzenat. Semantic precision and recall for ontology alignment evaluation. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, pages 348–353, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.
7. Jérôme Euzenat. First Experiments in Cultural Alignment Repair. In *Semantic Web: ESWC 2014 Satellite Events*, volume 8798, pages 115–130, 2014.
8. Jérôme Euzenat, Christian Meilicke, Heiner Stuckenschmidt, Pavel Shvaiko, and Cássia Trojahn. Ontology alignment evaluation initiative: Six years of experience. *Journal on Data Semantics XV*, pages 158–192, 2011.
9. Jérôme Euzenat and Pavel Shvaiko. *Ontology Matching*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.
10. Claudia V. Goldman, Martin Allen, and Shlomo Zilberstein. Learning to communicate in a decentralized environment. *Autonomous Agents and Multi-Agent Systems*, 15(1):47–90, August 2007.
11. Laura Hollink, Mark Van Assem, Shenghui Wang, Antoine Isaac, and Guus Schreiber. Two variations on ontology alignment evaluation: Methodological issues. In *Proceedings of the 5th European Semantic Web Conference on The Semantic Web: Research and Applications*, ESWC'08, pages 388–401, Berlin, Heidelberg, 2008. Springer-Verlag.
12. Fiona McNeill and Alan Bundy. Dynamic, automatic, first-order ontology repair by diagnosis of failed plan execution. *International Journal on Semantic Web and Information Systems*, 3(3), 2007.
13. Willem Robert Van Hage, Hap Kolb, and Guus Schreiber. Relevance-based evaluation of alignment approaches: The oaei 2007 food task revisited. In *Proceedings of the 3rd International Conference on Ontology Matching - Volume 431*, OM'08, pages 234–238, Aachen, Germany, Germany, 2008. CEUR-WS.org.