Pattern-Guided Association Rule Mining for Complex Ontology Alignment

Beatriz Lima, Daniel Faria, and Catia Pesquita

LASIGE, Dep. Informática, Faculdade de Ciências, Universidade de Lisboa, Portugal

Abstract. Aligning real-world ontology pairs often requires establishing complex correspondences, as conceptual differences between them may be too profound to bridge with simple equivalence correspondences. Yet, most ontology alignment algorithms are restricted to finding simple equivalences between ontology entities.

This work presents a suite of novel algorithms for Complex Ontology Alignment (COA) that rely on a targeted application of Association Rule Mining (ARM) to known complex alignment patterns. This approach reduces the ARM search space, and enables the application of tailored semantic filtering algorithms for refining the mappings.

We evaluated our approach using a pattern-oriented manual method, which yielded a global weighted precision of 75%, but revealed our approach was unable to find mappings for some of the patterns present in the reference. On the other hand, our approach found several mappings for patterns not present in the reference with high weighted precision, highlighting the importance of establishing evaluation metrics that consider varying degrees of correctness while being fully automated.

Keywords: Ontology Matching · Ontology Alignment · Complex Ontology Matching · Association Rule Mining

1 Introduction

Ontology alignment is critical to address the semantic heterogeneity problem, as it finds correspondences that enable integrating data across the Semantic Web. One of its biggest challenges is that ontology schemas often differ conceptually, making it necessary to establish complex correspondences. A complex correspondence is an ontology mapping where at least one of the mapped entities is an expression, rather than a simple ontology entity. The expressions used in complex mappings include restrictions (e.g. $\forall x, y, o_1 : MarriedPerson(x) \equiv o_2 : hasSpouse(x, y) \wedge o_2 : Person(y)$) and constructions using logical operatores (e.g. $\forall x o_1 : Mother(x) \equiv o_2 : Parent(x) \wedge o_2 : Woman(x)$).

The relevance of the Complex Ontology Alignment (COA) sub-field has been acknowledged by the Ontology Alignment Evaluation Initiative (OAEI) [1] who

---

[1] `http://oaei.ontologymatching.org`

introduced a Complex track in 2018 [6]. As of 2020, only three out of twelve participating systems were able to produce complex mappings (AMLC [2], AROA [9] and CANARD [8]) and their performance was very modest in comparison with the results of simple matching tracks [4]. The strategies employed by these systems can be divided into two categories: lexical and instance-based approaches. AMLC employs a lexical approach, which is inherently limited to finding the subset of complex mappings where there is lexical similarity between all entities mapped. CANARD, AROA and this work employ instance-based approaches, which use statistical and pattern mining techniques over a dataset of individuals shared (or mapped) between the two ontologies. AROA uses an Association Rule Mining (ARM) algorithm, FP-Growth [3], over a transaction database derived from the instance-level triples shared by two ontologies, thus demonstrating how a complex ontology alignment dataset can be transformed into a traditional ARM problem. Predefined complex alignment patterns [5] are then used to filter the generated association rules and produce complex ontology mappings.

ARM exploring a shared set of instances is a promising approach. However, the fact that we have prior knowledge of the complex alignment patterns we want to find makes it inefficient to use a "catch-all" ARM algorithm to perform an exhaustive search for frequent itemsets, and only use the knowledge of the patterns *a posteriori* to filter the rules. Therefore, we propose to invert this paradigm, by using predefined complex alignment patterns to guide ARM. This effectively reduces the search space and allows the application of semantic-based filtering algorithms tailored to each kind of pattern, to select and refine the most relevant mappings.

2 Algorithms

Our ontology alignment approach consists of the following steps:

1. An initial ontology loading step retrieves the set of shared individuals between the two ontologies and organises the ontology information (types, relations and property values of each individual, ranges and domains of the properties and hierarchical relations between classes) in hash-tables.
2. For each complex alignment pattern, an individual pattern matching algorithm iterates through the set of shared individuals, and, for each individual, it searches the hash-table data structures containing the relevant data for the targeted alignment pattern. For each mapping candidate found, we increment the support (i.e., the frequency) of the source and target entities in the mapping and the support of the mapping itself (i.e., the fraction of shared individuals that have both the source and target entities).
3. A common ARM matching algorithm is then invoked by each pattern matching algorithm to filter mapping candidates by support and confidence, therefore extracting association rules.
4. Filtering algorithms select which of the candidate mappings to include in the final alignment, excluding redundant mappings and conflicting mappings with lower confidence. An aggregator algorithm combines mappings for the

same entity into a single mapping using logical operators, such as "AND" and "OR".

Our algorithms cover eight distinct complex patterns, from which seven were found in the $cmt - conference$ dataset (see Table 1). Additionally, they can produce combinations of these patterns through disjunction and conjunction.

3 Evaluation

We integrated our algorithms in the ontology matching system AMLC [1,2], and assessed their performance in the $cmt - conference$ alignment [2] by manually classifying the mappings according to a rating scale consisting of the following five categories with associated scores:

- Correct [1.0]: The mapping is formally correct (regardless of whether it is present in the reference alignment).
- Nearly correct [0.75]: Only minor corrections necessary (e.g., alter the mapping relation type or substitute a class for its sub- or super-class).
- Plausible [0.5]: The mapping seems sensible and no information in the ontologies or reference alignment contradicts it.
- Implausible [0.25]: The mapping seems incorrect and is likely derived from biases in the dataset, but no information in the ontologies or reference alignment contradicts it.
- False [0.0]: The mapping is contradictory to the reference alignment and/or ontologies.

Our approach allows for fine tuning of matchers and filters, specific to each pattern, yielding precise results (Table 1). While it was unable to find mappings for some of the patterns present in the reference, it found several mappings for patterns not present in the reference with high weighted precision.

4 Conclusions

We developed a novel complex ontology matching method based on pattern-guided ARM, which represents a paradigm shift by making used of the alignment patterns to steer, rather than filter, the ARM process. Our manual evaluation revealed that the majority of mappings we found are correct or nearly correct, even if not present in the reference alignment. These results highlight the importance of establishing evaluation metrics that consider varying degrees of correctness while being fully automated. Going forward we will investigate the computational performance of our approach versus classical ARM, and extend the types of patterns it captures.

---

[2] Available at: `http://oaei.ontologymatching.org/2020/complex/index.html#popconf`; Reference alignments provided by Thiéblin *et al.* [7]

Table 1. Pattern-oriented analysis of the results obtained in the $cmt - conference$ alignment using the filtered approach. N: number of mappings; Ref: reference alignment; W: weighted. The total alignment size does not correspond to the sum of pattern occurrences as the same mapping may contain multiple patterns.

Pattern	**Ref. N**	**Result N**	**W.Precision (%)**
Class - Class	16	10	77.5
Class - *cardinality* restriction on Object Property	5	22	73.9
Class - *someValues* restriction on Object Property	4	6	58.3
Class - *hasValue* restriction on Data Property	-	-	-
Class - *someValues* restriction on Data Property	-	1	100
Object Property - Object Property	10	3	75.0
Data Property - Data Property	1	-	-
Object Property - Data Property	-	-	-
Object Property - *InverseOf* Object Property	2	-	-
Object Property - Object Property+range restriction	-	9	80.6
Object Property - Object Property+domain restriction	-	8	78.1
Data Property - Data Property+domain restriction	-	-	-
Total alignment	35	51	75.0

Acknowledgments This work was supported by FCT through the LASIGE Research Unit (UIDB/00408/2020 and UIDP/00408/2020). It was also partially supported by the KATY project which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101017453.

References

1. Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I.F., Couto, F.M.: The AgreementMakerLight ontology matching system. In: On the Move to Meaningful Internet Systems: OTM 2013 Conferences. pp. 527–541. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
2. Faria, D., Pesquita, C., Tervo, T., Couto, F., Cruz, I.: AML and AMLC results for OAEI 2019. vol. 2536, pp. 101–106 (2019)
3. Han, J., Pei, J., Yin, Y.: Mining frequent patterns without candidate generation. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. p. 1–12. SIGMOD '00, Association for Computing Machinery (2000)
4. Pour, N., Algergawy, A., Amini, R., Faria, D., Fundulaki, I., Harrow, I., Hertling, S., Jimenez-Ruiz, E., Jonquet, C., Karam, N., et al.: Results of the Ontology Alignment Evaluation Initiative 2020. vol. 2788, pp. 92–138. CEUR-WS (2020)
5. Ritze, D., Meilicke, C., Sváb-Zamazal, O., Stuckenschmidt, H.: A pattern-based ontology matching approach for detecting complex correspondences. vol. 551 (2009)
6. Thiéblin, É., Cheatham, M., Santos, C., Sváb-Zamazal, O., Zhou, L.: The First Version of the OAEI Complex Alignment Benchmark. In: International Semantic Web Conference. vol. 2180. Springer (2018)

7. Thiéblin, É., Haemmerlé, O., Hernandez, N., Trojahn, C.: Task-oriented complex ontology alignment: Two alignment evaluation sets. In: Gangemi, A., Navigli, R., Vidal, M.E., Hitzler, P., Troncy, R., Hollink, L., Tordai, A., Alam, M. (eds.) The Semantic Web. pp. 655–670. Springer International Publishing, Cham (2018)
8. Thiéblin, E., Haemmerlé, O., Trojahn, C.: CANARD complex matching system: results of the 2019 OAEI evaluation campaign. vol. 2536, pp. 114–122 (2019)
9. Zhou, L., Hitzler, P.: AROA Results for 2020 OAEI. vol. 2788, pp. 161–167 (2020)