On Partitioning for Ontology Alignment

Sunny Pereira[1], Valerie Cross[1], Ernesto Jiménez-Ruiz[2]

[1] Miami University, Oxford, OH, United States, [2] University of Oslo, Norway

1 Methods

Ontology alignment (OA) for two very large ontologies becomes time consuming and memory intensive. A general approach to address these challenges is to partition each ontology into cohesive blocks. (i.e., partitions). Ontology partitioning brings new challenges: how best to partition each ontology into blocks and whether the partitioning process on each ontology should be independent of each other. In this paper, we present preliminary work to determine the suitability of partitioning strategies to improve the performance of OA systems, especially those unable to cope with the largest datasets.

The PBM (Partition Block Matching) [2,3], PAP (partition, anchor, partition) and APP (anchor, partition, partition) [1] partitioning methods have been implemented as independent methods from the alignment system. In the preliminary experiments included in this paper we report results for the systems LogMap [4] and FCA-Map [7]. In [1], [2], and [3] a path-based semantic [6] similarity measure is used to determine link strength between concepts within an ontology when creating blocks. In these experiments, the path-based Wu-Palmer [6] as well as information content based Lin [5] semantic similarity measures are considered. The ontology structure is used in determining the information content (IC) for a concept. The link strengths are calculated between concepts that only differ by one in their depth within the ontology. The authors of the PBM method use ISUB to find the anchors between concepts. In our experiments, anchors are found using an exact label match between two concepts in the two different ontologies. Each identified block pair represents a matching (sub)task, however, since blocks are only characterized by a set of concepts, they are first converted to (logical) ontology modules and then given to the ontology alignment system as input.

The initial experiments were performed on task 1 of the OAEI *largebio* track,[1] involving small fragments of FMA and NCI, using all three methods. The results using Wu-Palmer are shown below in Table 1 and those for Lin in Table 2. The parameters used are an η of 0.05 for PBM, an α of 0.75 for APP. A maximum block size of 500 and a depth difference of one for semantic similarity calculation is used for all three methods. Blocks with only one concept are considered isolated blocks. *Coverage* represents how many of the entities occurring in the OAEI reference alignments are present in the identified block pairs. The precision and recall are calculated over the combined alignment results for all the matching tasks (i.e., pair of modules extracted from the block pairs). FMA blocks (resp. NCI blocks) represents the number of total blocks produced after partitioning of the FMA ontology (resp. NCI ontology).

The results from task 1 suggest that the PBM method provides much higher recall values than the other two methods. The Wu-Palmer measure performed slightly better than Lin. The next experiments examined how the PBM with the Wu-Palmer performed on the OAEI *largebio* tasks that use the whole ontologies, that is, task 2, task 4 and task 6. The maximum block size is 3000. Table 3 presents these results.

[1] http://www.cs.ox.ac.uk/isg/projects/SEALS/oaei/

Table 1. Experiments in *largebio* task 1 suing Wu-Palmer. Matching with LogMap.

Method	FMA Blocks		NCI Blocks		Matching Tasks	Coverage	Precision	Recall	Time (s)	
	#	Isolated	#	Isolated					Partitioning	Matching
PBM	55	15	141	60	87	0.821	0.845	0.743	40.248	85.162
PAP	60	13	141	60	58	0.451	0.870	0.410	39.827	58.517
APP	50	15	143	60	48	0.518	0.870	0.472	41.644	53.157

Table 2. Experiments in *largebio* task 1 using Lin. Matching with LogMap.

Method	FMA Blocks		NCI Blocks		Matching Tasks	Coverage	Precision	Recall	Time (s)	
	#	Isolated	#	Isolated					Partitioning	Matching
PBM	46	6	180	53	83	0.801	0.833	0.728	52.454	81.689
PAP	37	5	180	53	37	0.348	0.861	0.321	56.508	39.423
APP	46	6	180	53	46	0.483	0.862	0.439	56.704	49.938

Table 3. Experiments with *largebio* whole ontologies using PBM with Wu-Palmer.

Task	System	Source Blocks		Target Blocks		Matching Tasks	Coverage	Precision	Recall	Time (s)	
		#	Isolated	#	Isolated					Partitioning	Matching
FMA-NCI	LogMap	151	2	256	91	69	0.763	0.468	0.675	649	76.7
	FCA-Map							0.506	0.698		$\approx$ 8 hrs
FMA-SNOMED	LogMap	388	9	3352	3273	154	0.594	0.571	0.423	4,807	385
SNOWMED-NCI	LogMap	3357	3160	693	427	443	0.666	0.725	0.491	6,623	937

2 Discussion and future work

In this paper we have presented a preliminary evaluation of state of the art partitioning algorithms for ontology alignment. The obtained results are not good as expected since, after the partitioning and identification of the (sub)matching tasks, the coverage of the entities in the reference alignments is rather low. For example, in the FMA-SNOMED case only 59% of the entities appearing in the reference alignment are covered by the modules in the identified matching tasks. In this case 41% of the entities were lost in either isolated blocks or blocks for which a suitable pair could not be found.

As expected, given the coverage of entities in the reference alignment, the results obtained by LogMap are very low as compared to the results reported for LogMap in last OAEI campaign. In addition the partitioning step represents a considerable overhead with respect LogMap's computation times. Nevertheless, FCA-Map was successfully run in task 2 of the *largebio* track using partitioning,[2] while the system could not cope with the task when given the whole FMA and NCI ontologies.

In the close future we aim at investigating new algorithms to provide a suitable partitioning for ontology alignment where the loss of coverage in the identified (sub)matching tasks, in terms of entities of the reference alignments, is minimized. We also intend to perform an extensive evaluation of the novel partitioning algorithms with all OAEI participating systems, especially those failing to cope with the largest tasks.

References

1. Hamdi, F., et al.: Alignment-based partitioning of large-scale ontologies. SCI, vol. 292 (2010)
2. Hu, W., Qu, Y.: Block matching for ontologies. In: Int'l Sem. Web Conf. (2006)
3. Hu, W., et al.: Matching large ontologies: A divide-and-conquer approach. DKE (2008)
4. Jiménez-Ruiz, E., Cuenca-Grau, B.: LogMap: Logic-based and scalable ontology matching. In: ISWC (2011)
5. Lin, D., et al.: An information-theoretic definition of similarity. In: ICML (1998)
6. Wu, Z., Palmer, M.: Verbs semantics and lexical selection. In: ACL (1994)
7. Zhao, M., Zhang, S.: FCA-Map results for OAEI 2016. In: Ontology Matching (2016)

[2] Not tested in tasks 4 and 6 due to limited experimental time