Measuring Semantic Similarity within Reference Ontologies to Improve Ontology Alignment

Valerie Cross and Pramit Silwal

Computer Science and Software Engineering Department,
Miami University, Oxford, OH 45056
crossv@muohio.edu

1 Mediating Matcher with Semantic Similarity

Some ontology alignment (OA) systems find an identical bridge concept in a reference ontology to which both the source and target concept can be mapped. Then a mapping between the two is produced. Using semantic similarity within a reference ontology can find more mappings than with only identical bridge concepts. A wide variety of semantic similarity measures were implemented within AgreementMaker [1] to use semantic similarity to evaluate OA mappings [2]. Initial results of enhancing AgreementMaker with a new matcher, the mediating matcher with semantic similarity (MMSS) in place of its mediating matcher (MM) are in [3]. Briefly, the MMSS uses the MM to first produce a set of mappings M_{ST} between source and target concepts with an exact match on the bridge concepts, i.e., $b_S = b_T$ as

$$M_{ST} = \{(s, t, mapSim_{SR} * mapSim_{TR}) \mid s \in O_S, b_S, b_T \in O_R, t \in O_T : \exists (s, b_S, mapSim_{SR},) \in M_{SR} \wedge \exists (t, b_T, mapSim_{TR},) \in M_{TR} \wedge b_S = b_T\}$$

M_{SR} contains mapping from the source O_S to the reference O_R using BSM^{lex} matcher and similarly for M_{TR} with O_T. U_S contains source concepts s from M_{SR}, which are not selected by the original MM and similarly U_T for the target concepts t.

$$U_S = \{s \mid s \in O_S : \exists (s, b_S, mapSim_{SI}) \in M_{SI} \wedge \nexists\, t \in O_T : (s, t, sim_{ST}) \in M_{ST}\}$$

For each (s, t) in U_S x U_T, semantic similarity between all their b_S and b_T are calculated, and the maximum is used in determining the enhanced mapping set as

$$E_{ST} = \{(s, t, agg(mapSim_{SR}, mapSim_{TR}, bridgeSim)) \mid s \in U_S, b_S, b_T \in O_R, t \in U_T : \exists (s, b_S, mapSim_{SR}) \in M_{SR} \wedge \exists (t, b_T, mapSim_{TR},) \in M_{TR} : bridgeSim = \max_{b_S, b_T \in O_R} (semSim(b_S, b_T))\}.$$

Different *agg* operators are possible, but here minimum is used with the rationale that the final mapping between *s* and *t* is not any stronger than the weakest similarity between the pairs of concepts, (s,b_S), (t,b_T), and (b_S,b_T). Different semantic similarity measures can be used for *semSim*. The standard Lin semantic similarity measure is used with information content described as in [3]. An additional threshold value may be set to eliminate mappings in E_{ST} whose aggregated similarity falls below the threshold. $M_{ST} \cup E_{ST}$ is input to the linear weighted combination (LWC) operation

2 Experimental Results using OAEI 2011 Anatomy Track

To compare the MMSS to the MM in OAEI 2011 AgreementMaker configuration using its matchers and hierarchical LWCs, experiments are performed with the OAEI 2011 anatomy track and Uberon as the reference ontology. The results are shown in Table 1. At the 0.9 threshold level, the OAEI 2011 AgreementMaker configuration with MMSS (OAEI-MMSS) produced 2 more mappings than with MM (OAEI-MM), but no more correct mappings. Examining the mappings showed OAEI-MMSS found 3 new correct ones but lost 3 correct ones found by OAEI-MM. Further analysis suggests that the interaction among AgreementMaker's matchers, its local quality measures (LQM) used as weighting for its LWCs, and the hierarchical organization of its LWCs have subtle effects on the mappings eventually selected for the final result.

Table 1. OAEI 2011 AgreementMaker MM vs. its MMSS version

	Produced	Correct	Precision	Recall	F-measure
OAEI-MM	1439	1348	93.7	88.9	91.2
OAEI-MMSS -0.9	1441	1348	93.5	88.9	91.2
OAEI-MMSS-0.95	1441	1350	93.7	89.1	91.3
OAEI-MMSS-0.95, PSM kept	1443	1353	93.8	89.2	91.4

The OAEI 2011 AgreementMaker configuration hierarchically combines its Parametric String-based Matcher (PSM), Vector-based Multi-word Matcher (VMM) and Lexical Similarity Matcher (LSM) represented as LWC3(LWC1(LSM+MM) + LWC2(PSM + VMM)) using LQMs to weight each component. MMSS is substituted for MM. Other hierarchical combinations of matchers in experiments did not perform better than OAEI-MM. However, three different hierarchical combinations produced new mappings not found by either the OAEI-MM or OAEI-MMSS for a total of 9 new correct mappings. More work is needed to determine possible heuristics to be able to keep the lost 3 mappings and also retain the 9 new mappings. Examining LQM weighting showed LQMs for MMSS are usually higher than for the PSM or VMM; therefore, the MMSS dominates in the final results. The third table row shows going from 0.9 to 0.95 eliminates incorrect mappings to retain the lost mappings. The last row shows by keeping identical source-target mappings from the PSM, a higher F-measure is achieved, better than AgreementMaker's OAEI 2011 result.

References

1. Cruz, I. F., Stroe, C., Caimi, F., Fabiani, A., Pesquita, C., Couto, F. M., Palmonari, M.: Using AgreementMaker to Align Ontologies for OAEI 2011. Ontology Matching Workshop, International Semantic Web Conference (2011)
2. Cross, V. and Hu, X: Using Semantic Similarity in Ontology Alignment. OM Workshop, 10th Int. Semantic Web Conference (ISWC 2011), Bonn Germany (2011)
3. Cross, V., Silwal, P., and Morell, D: Using a Reference Ontology with Semantic Similarity in Ontology Alignment. International Conference on Biomedical Ontologies (ICBO), July 22 – 25, Graz, Austria (2012)