TECHNISCHE UNIVERSITÄT DARMSTADT

Towards Rule Learning Approaches to Instance-based Ontology Matching



Frederik Janssen¹, Faraz Fallahi², Jan Noessner³, Heiko Paulheim¹

¹ Knowledge Engineering Group, Technische Universität Darmstadt, Germany

² ontoprise GmbH, Karlsruhe, Germany

³ KR & KM Research Group, University of Mannheim, Germany

Main Problems of lexical distance measures or pattern recognition for Ontology Matching: complex mappings cannot be found

- e.g., in multi-lingual schemas there is no lexical similarity at all *Remedy:* use of machine learning techniques, focus on symbolic representations (such as rules) Advantages:
- interpretability: enhanced methods for comparison and combination of rules and rule sets • capability of finding complex mappings

<u>Case Study 1 – Creating mappings by</u> association rule mining

- *Idea:* use association rule learning to find mappings
- using binary features for classes
- conclude mappings for symmetrical rules, e.g.

<u>Case Study 2 – Refining mappings by</u> <u>separate-and-conquer rule learning</u>

- *Given:* two ontologies and some existing mappings (e.g., found by a lexical matcher)
- Goal: find additional mappings

dataset from ontology \mathcal{O}_1

dataset from ontology \mathcal{O}_2

DBpedia-owl:ProtectedArea ← yago:Park yaqo:Park ← DBpedia-owl:ProtectedArea

DBpedia-owl:ProtectedArea \equiv yago:Park \Rightarrow

Preliminary Results

• *Data set:* manual partial mapping between DBpedia and YAGO



@relation car

@attribute acceleration {low,medium,high} @attribute cargoCapacityRating {low,high} @attribute passengerSpaceRating {low,high} @attribute convenienceRating {low,medium,high} @attribute milesPerGallon {low,medium,high}

@data high,low,high,medium,low high,low,low,high,medium low,low,high,high,low low,low,low,low,medium medium,high,high,low,low medium, high, low, high, medium low,high,high,medium,high

learn | rules

 $r_{1,1}$: milesPerGallon=medium \leftarrow convenience-Rating=high **AND** acceleration=high

 $r_{1,2}$: milesPerGallon=high \leftarrow acceleration= medium **AND** cargoCapacity=low

@relation cars

@attribute acceleration {low,medium,high} @attribute cargoCapacity {low,high} @attribute passengerSpace {low,high} @attribute convenience {low,medium,high} @attribute numberOfExtras {low,medium,high} @attribute mpg {low,medium,high}

@data

high,low,high,medium,medium,low high,low,low,low,high,high low,low,high,high,high,low low,low,low,high,medium,low low,high,high,high,low,medium medium,high,high,high,high,medium low,high,high,medium,low,high

learn|rules

mpg=medium ← convenience=high **AND** acceleration=high

numberOfExtras=high ← convenience=high **AND** passengerSpace=high

• *Idea:* similar rule sets \rightarrow mapping candidate

Minimum confidence threshold

approach is also able to find complex mappings, such as

 ≥ 1 DBpedia-owl:name \sqsubseteq yago:Person

Conclusions

 reformulation of ontology matching as problems of (association) rule learning first experiments show that both approaches work



Challenges:

 create suitable benchmark data sets for complex mappings similarity measures for rules and rule sets • parameter tuning of rule learning algorithms