Towards Rule Learning Approaches to Instance-based Ontology Matching

Frederik Janssen[1], Faraz Fallahi[2], Jan Noessner[3], Heiko Paulheim[1]

[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

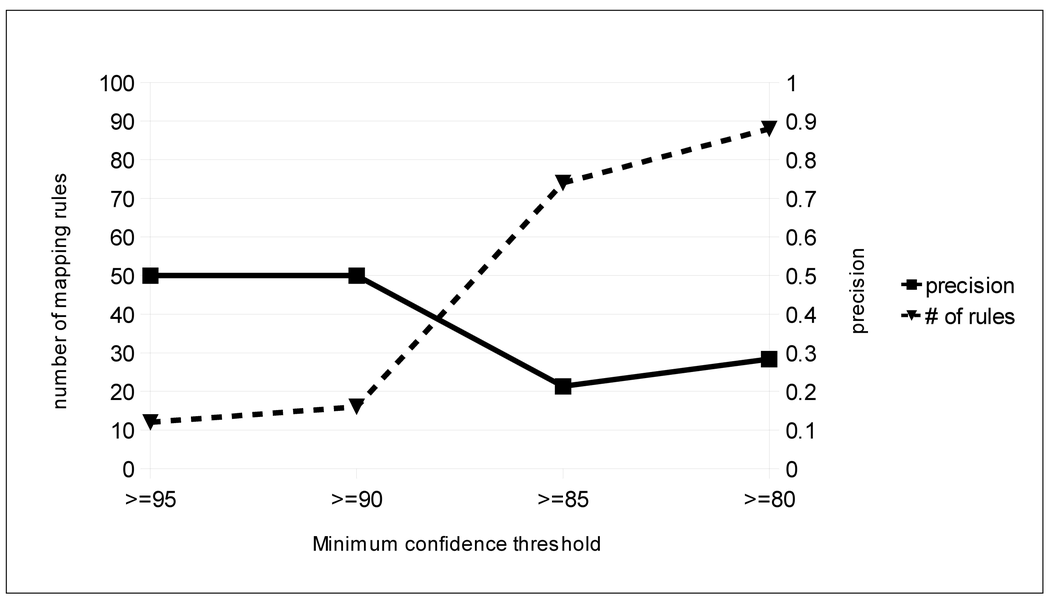
Case Study 1 – Creating mappings by association rule mining

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

$$\texttt{DBpedia-owl:ProtectedArea} \leftarrow \texttt{yago:Park}$$
$$\texttt{yago:Park} \leftarrow \texttt{DBpedia-owl:ProtectedArea}$$
$$\Rightarrow \quad \texttt{DBpedia-owl:ProtectedArea} \equiv \texttt{yago:Park}$$

Preliminary Results

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

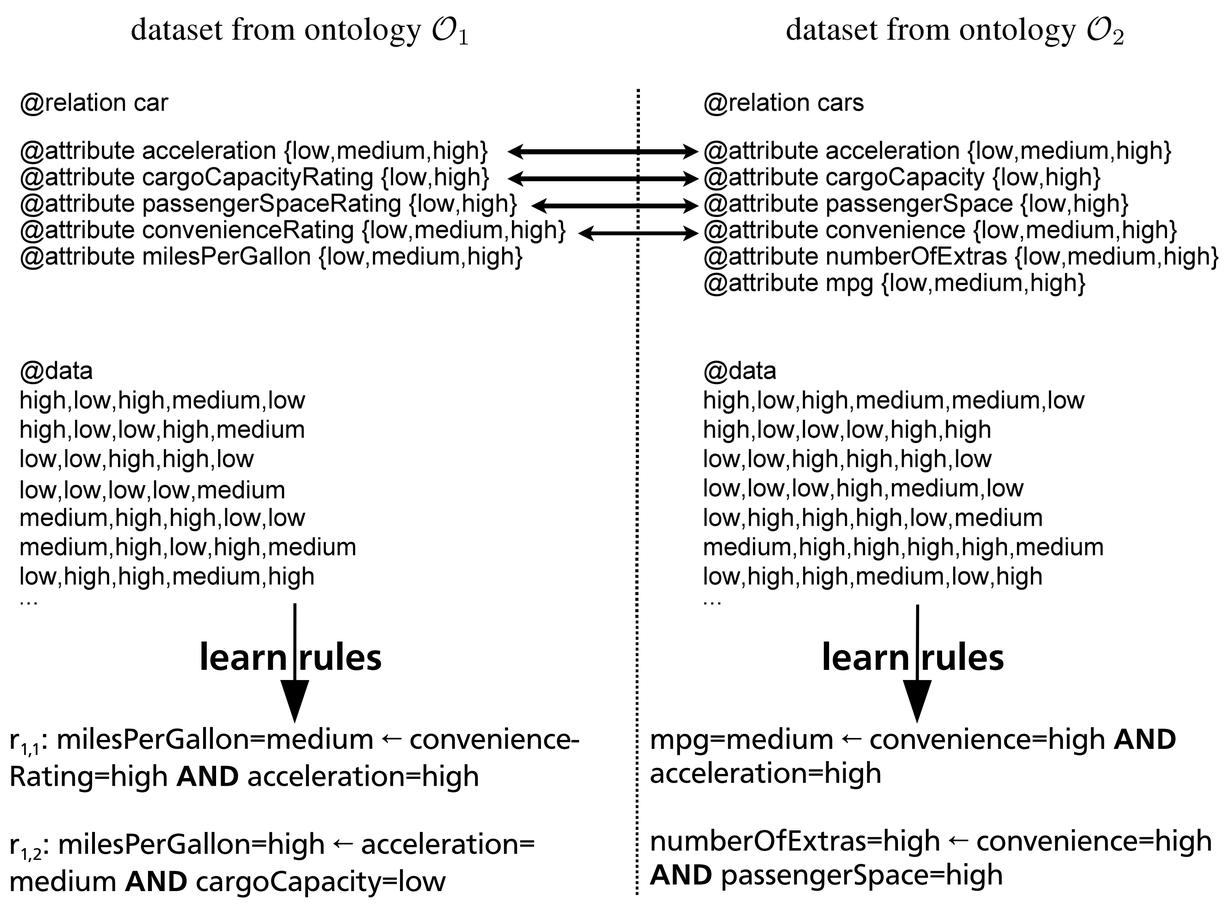


- approach is also able to find complex mappings, such as

$$\geq 1\texttt{DBpedia-owl:name} \sqsubseteq \texttt{yago:Person}$$

Case Study 2 – Refining mappings by separate-and-conquer rule learning

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



- *Idea:* similar rule sets → mapping candidate

$$sim_R(R, R') = \frac{\sum_{sim_r(r_{1,i}, r_{2,j}) \geq \theta} tp(r_{1,i}) + tp(r_{2,j})}{|D_1| + |D_2|}$$

e.g., with $sim_r(r, r') = \begin{cases} 1 \text{ if } r \text{ matches } r' \text{ exactly} \\ 0 \text{ otherwise} \end{cases}$

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