Transfer Learning of Link Specifications

August 21, 2013

1. With the growth of of Linked Data Set, Link Discovery becomes one of the crucial issues in Semantic Web.
2. Specifying link specifications represents the main player in link discovery and the better specifications the better linking results.
3. Our work is motivated by the question:
   Can we reuse existing knowledge of links specifications to detect new link specifications and enhancing Link Discovery accuracy?.

Introduction

Link Discovery consists of tow basic steps:

1. Specifying the Link specifications.
2. Carrying out linking using a link discovery framework.

Many frameworks were developed to address quadratic a-priori runtime of Link Discovery

- LIMES
- SILK
- RDF-AI

Link Discovery

- Our work is related to tow research areas:
  - Link Discovery
  - Dectect link specifications
- Link Discovery main aim is finding links between tow datasets. Link Discovery is formalized as:
  - For source S and target T and relation ρ, compute the set M of pairs of instances $(s, t) \in S \times T$ such that $\forall (s, t) \in M : \rho(s, t)$.
  - $\rho(s, t)$ represents the projection of s and t into similarity space $\mathfrak{S}$ such that $\rho(s, t)$ is set iff $\sigma(s, t) \geq \tau$ is satisfied,
    where $\sigma : S \times T \rightarrow [0, 1]$ is a similarity function and $\tau \in [0, 1]$.

Link Specification

- Link Specification is a main step in Link Discovery
- Link Specifications has three compnents:
  - Two sets of restrictions $\mathcal{R}_1^S$ ... $\mathcal{R}_m^S$ resp. $\mathcal{R}_1^T$ ... $\mathcal{R}_k^T$ that specify the sets S resp. T,
  - A specification of a complex similarity metric σ via the combination of several atomic similarity measures σ_1, ..., σ_n and
  - A set of thresholds τ_1, ..., τ_n such that τ_i is the threshold for σ_i.

Transfer Learning

- Transfer Learning is a Machine Learning approach.
- Machine Learning goal is:
  Here the
  mathcal error happens
- In our approach we use *Transductive Transfer Learning*

Transfer Learning Framework I

Transfer Learning of link specifications is tackled through three problems:

- Restrictions similarity
  It is reduced to be Classes similarity as
  s rdf:type someClass
  The similarity function: $\zeta : 2^C \times 2^C \mapsto [0, 1]$
- Properties similarity
  The similarity function: $\pi : P \times P' \mapsto [0, 1]$,
  where P and P' are properties sets for C and C', the set all such property similarity functions is denoted as Π.
- Determining accuracy of link specifications
  link specification assessment function: $\alpha : Q \mapsto [0, 1]$.

Transfer Learning Framework II

- The overall similarity measure for Transfer Learning is represented as: $\omega(t, t') = \alpha(q') \cdot \zeta(\psi(q'), \mathcal{C}) \cdot \zeta(\psi'(q'), \mathcal{C}') \cdot r'(r(q', P_L, \pi), P'_L, \pi)$ Each function in similarit measure can be implemented in manifold approaches
- Class similarity function ζ is implemented in Framework using tow approaches:
  - label-based similarity
  - name-based similarity (URI similarity)
  - data-centric similarity

Experiemental setup I

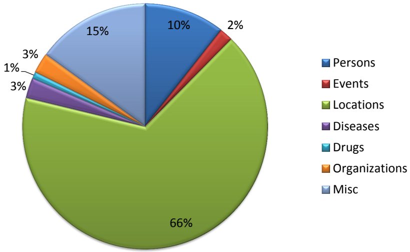
The goal of evaluation is tow-folded:

- Evaluating the accuracy of function f' to be based for predicting f'.
- Discover whether the functions f' for other domains could be used directly.

113 specifications were retrieved from LATC, each has manual links evaluation.

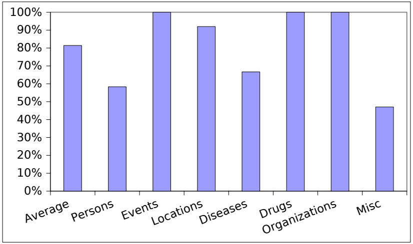
Experiemental setup II

- URI similarity is used where the source and target endpoints in a specification are alive
- The experiement was applied on 12 specifications out of specifications retrieved from LATC.
- The distributions of the specfications across different domains are:



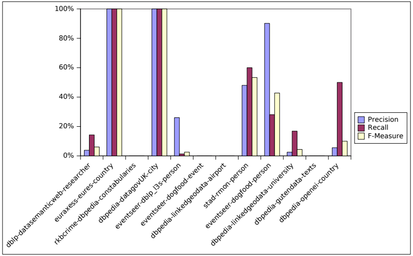
First experiements set Results

- Detecting right specfication average is 81%
- Detecting right specfications in geo-spatial domain is 92%
- Detecting right specfications in persons domain is 58.3%



Second experiements set Results

In the second experiements series, both source and target endpoints considered to be alive.



Conclusion

- Detecting best similar specification with mean reciprocal rank larger than or equal 0.81
- Transfer learning can not replace the learning of link specification in itself

Future Work

- Combining other learning approaches of link specification to transfer learning
- Using more sophisticated class and property similarity approaches

Questions

?