Semantic relatedness measure using object properties in an ontology

Laurent Mazuel, Nicolas Sabouret

Laboratoire Informatique de Paris 6 (LIP6), France
{laurent.mazuel, nicolas.sabouret}@lip6.fr

7th International Semantic Web Conference, October 30th 2008
Outline

1. Problem
2. Theoretical measure
3. Evaluation
4. Conclusion
Outline

1. Problem
2. Theoretical measure
3. Evaluation
4. Conclusion
Semantic measure definition

Semantic measure:

*Computes a score of semantic similarity/relatedness/distance between two concepts defined in the same knowledge representation*
Semantic measure definition

**Semantic measure:**

*Computes a score of semantic similarity/relatedness/distance between two concepts defined in the same knowledge representation*

1. **Similarity:** only use attributes in common (e.g. moto-car)
2. **Relatedness:** use non-subsumption relation (e.g. gasoline-car)
**Semantic measure definition**

**Semantic measure:**

*Computes a score of semantic similarity/relatedness/distance between two concepts defined in the same knowledge representation*

1. **Similarity:** only use attributes in common (e.g. moto-car)
2. **Relatedness:** use non-subsomption relation (e.g. gasoline-car)

**In literature**

1. **Similarity:** well-studied on all KR
2. **Relatedness:** studied only in Gloss-based [Strube06] or Google [Cilibrasi06]
3. **Human/machine interaction system cannot use** Gloss-based or Google [Eliasson07]
Semantic measure definition

Semantic measure:

Computes a score of semantic similarity/relatedness/distance between two concepts defined in the same knowledge representation

1. Similarity: only use attributes in common (e.g. moto-car)
2. Relatedness: use non-subsumption relation (e.g. gasoline-car)

In literature

1. Similarity: well-studied on all KR
2. Relatedness: studied only in Gloss-based [Strube06] or Google [Cilibrasi06]
3. Human/machine interaction system cannot use Gloss-based or Google [Eliasson07]

Need for efficient relatedness on graph-based KR
Hypothesis

1. Graph-based knowledge representation (e.g. semantic networks, W3C SKOS):
   - Based upon hierarchical structure
   - With heterogeneous relations (part-of, etc.)

2. Extension of previous work on semantic similarity measure
Semantically correct path

- Introduced by [Hirst & St-Onge 98]
- Notion still used: [Aleksovski 06], [Hollink 06]
- Using all relations, must filter the set of all possible graph paths
  \[ \Rightarrow \text{set of patterns to recognize a semantically correct path, based on the combination of relation type in a path} \]
Semantically correct path

- Introduced by [Hirst & St-Onge 98]
- Notion still used: [Aleksovski 06], [Hollink 06]
- Using all relations, must filter the set of all possible graph paths
  \( \Rightarrow \) set of patterns to recognize a semantically correct path, based on the combination of relation type in a path

Examples

- \([is-a]^+ [part-of]^+ [includes]^+\): correct pattern
- \([is-a]^+ [part-of]^+ [includes]^+ [part-of]^+\): incorrect pattern
Semantically correct path

- Introduced by [Hirst&St-Onge98]
- Notion still used: [Aleksovski06], [Hollink06]
- Using all relations, must filter the set of all possible graph paths

\[ \Rightarrow \text{set of patterns to recognize a semantically correct path, based on the combination of relation type in a path} \]

Examples

- \([is\text{-}a]^+ [part\text{-}of]^+ [includes]^+\): correct pattern
- \([is\text{-}a]^+ [part\text{-}of]^+ [includes]^+ [part\text{-}of]^+\): incorrect pattern

We will only consider paths which are semantically correct
Single-relation path: hierarchical path

Single-relation path
- Path with only one type of relation
Single-relation path: hierarchical path

**Single-relation path**
- Path with only one type of relation

**Hierarchical single-relation path**
- Information theoretic approach introduced by [Resnik95]
- Each node has a weight:
  - ⇒ the *Information Content* function: $IC(x)$ [Resnik95, Seco04]
- Converted to edge weight by [Jiang&Conrath97]:
  \[
  W(path_{x \in \{isa,include\}}(x, y)) = |IC(x) - IC(y)|
  \]
Single-relation path: non-hierarchical path

Non-hierarchical path

\[ W(path_X(x, y)) = TC_X \times \left( \frac{|path_X(c_1, c_2)|}{|path_X(c_1, c_2)| + 1} \right) \]

- With \( TC_X \) the weight of an infinite-length path of type \( X \)

Motivation

- \( TC_X \): bound the value in \([0, TC_X]\)
- \( \frac{n}{n+1} \): approximate the IC function shape [Seco04]
Final distance

Weight of a mixed-path

- The function \( T(path(x,y)) \) computes the minimal set of single-relation paths

\[
W(path(x,y)) = \sum_{p \in T(path(x,y))} W(p)
\]
Final distance

Weight of a mixed-path

- The function $T(path(x, y))$ computes the minimal set of single-relation paths

$$W(path(x, y)) = \sum_{p \in T(path(x, y))} W(p)$$

Final distance

- Function $HSO(p)$ is true iff $p$ is a valid path w.r.t. HSO rules.

$$dist(c_1, c_2) = \min_{\{p \in \pi(c_1, c_2) | HSO(p) = true\}} W(p)$$
Outline

1. Problem
2. Theoretical measure
3. Evaluation
4. Conclusion
Evaluation

Protocol

- KR: WordNet 3.0, IC [Seco04], using part-of only
- Test: [Miller&Charles91], [Finkelstein01] for WordSimilarity-353
  - M&C: 30 couples, test similarity (e.g. magician-wizard)
  - WS-353: 353 couples, test relatedness (e.g. computer-keyboard)
Evaluation

Protocol

- KR: WordNet 3.0, lC [Seco04], using part-of only
- Test: [Miller&Charles91], [Finkelstein01] for WordSimilarity-353
  - M&C: 30 couples, test similarity (e.g. magician-wizard)
  - WS-353: 353 couples, test relatedness (e.g. computer-keyboard)

<table>
<thead>
<tr>
<th>Measures</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M&amp;C</td>
</tr>
<tr>
<td>Rada</td>
<td>0.638</td>
</tr>
<tr>
<td>Resnik</td>
<td>0.804</td>
</tr>
<tr>
<td>Lin</td>
<td>0.836</td>
</tr>
<tr>
<td>Jiang &amp; Conrath</td>
<td>0.880</td>
</tr>
<tr>
<td>Hirst &amp; St-Onge</td>
<td>0.847</td>
</tr>
<tr>
<td>Our measure, $TC_{part-of} = 0.4$</td>
<td>0.902</td>
</tr>
</tbody>
</table>
$TC_X$ study with [M&C91]

Miller & Charles

![Graph showing correlation vs. TCX]

- **Ours**
- **J&C**
- **H&SO**
TCX study with WS353

WS-353

Correlation

TCX

- Ours
- J&C
- H&SO

L. Mazuel, N. Sabouret (LIP6)
Outline

1. Problem
2. Theoretical measure
3. Evaluation
4. Conclusion
## Conclusion & future work

### Conclusion

- A new relatedness measure on graph-based knowledge model
  - With information theoretic approach
  - With semantic path patterns
  - *With a new formula for non-hierarchical path*
- Evaluated on classical benchmark & gives good result

### Future work

- Test with: Others KR model (e.g., SNOMED v3.5 Fr, 105,000 concepts)
- Integrated in a human/machine interaction system
- Extension to OWL Lite?
Conclusion & future work

Conclusion

- A new relatedness measure on graph-based knowledge model
  - With information theoretic approach
  - With semantic path patterns
  - *With a new formula for non-hierarchical path*
- Evaluated on classical benchmark & gives good result

Future work

- Test with:
  - Others KR model (*e.g.* SNOMED v3.5 Fr, 105.000 concepts)
  - Integrated in a human/machine interaction system
- Extension to OWL Lite?
Thank you for your attention!
Have you any question?