Declutter Your Justifications

Determining Similarity Between OWL Explanations

*Samantha Bail*, Bijan Parsia, Uli Sattler
The University of Manchester

WoDOOM, 8th October 2012
Background: OWL & Justifications

Definition

\[ J \text{ is a justification for } O \models \eta \text{ if } J \subseteq O, J \models \eta \]

and for all \( J' \subset J \) it holds that \( J' \not\models \eta \).

- Justifications *pinpoint the causes* for an entailment
  - Restrict attention to the relevant axioms
  - Focus on a potentially smaller set of axioms
- Best understood and most promising explanation type
  - for ontology *debugging* (understanding & fixing errors)
  - for ontology *comprehension*
Motivation: Multiple Justifications

- Entailments can have multiple justifications
  - potentially exponential in the size of the ontology
- We may also want to repair *multiple entailments*
- Large numbers of justifications are an unordered, unmanageable mess
- But: Justifications are often *similar* (if we look closely)
• Pizza ontology ($SHOIN$): 100 classes, 672 axioms
• >100 justifications for an entailment
• Pizza ontology ($SHOIN$): 100 classes, 672 axioms
• >100 justifications for an entailment
- Pizza ontology ($SHOIN$): 100 classes, 672 axioms
- >100 justifications for an entailment
Motivation: Multiple Justifications

- Entailments can have multiple justifications
  - potentially exponential in the size of the ontology
- We may also want to repair *multiple entailments*
- Large numbers of justifications are an unordered, unmanageable mess
- But: Justifications are often *similar* (if we look closely)
Motivation: Multiple Justifications

- Entailments can have multiple justifications
  - potentially exponential in the size of the ontology
- We may also want to repair *multiple entailments*
- Large numbers of justifications are an unordered, unmanageable mess
- But: Justifications are often *similar* (if we look closely)
- The logical diversity is not as great as it may seem
- ... how do we *determine similarity* between justifications?
Structural Equivalence

- **Structural equivalence** [1] of OWL axioms is well defined:
  - We have an equivalence relation!
  - ... but only a boring one

**Example**

1) \( \text{InterestingPizza} \equiv \text{Pizza} \sqcap \geq 3 \text{hasTopping} \)

2) \( \geq 3 \text{hasTopping} \sqcap \text{Pizza} \equiv \text{InterestingPizza} \)

Justification Isomorphism

- **Isomorphism** [2] between justifications is well defined
- It describes an equivalence relation

**Example**

\[ \mathcal{J}_1 = \{ A \subseteq B \cap \exists r.C, B \cap \exists r.C \subseteq D \} \models A \subseteq D \]
\[ \mathcal{J}_2 = \{ E \subseteq B \cap \exists s.F, B \cap \exists s.F \subseteq D \} \models E \subseteq D \]
\[ \phi = \{ A \mapsto E, C \mapsto F, r \mapsto s \} \]

We Want More!

... but these do not cover all possible “similarities”:

Example: Subexpressions

\[ J_1 = \left\{ A \subseteq B \cap C, B \cap C \subseteq D \right\} \models A \subseteq D \]

\[ J_2 = \left\{ A \subseteq \exists r.C, \exists r.C \subseteq D \right\} \models A \subseteq D \]
We Want More!

- ... but these do not cover all possible “similarities”:

**Example: Subexpressions**

\[
\mathcal{J}_1 = \{ A \sqsubseteq X1, B \cap C \sqsubseteq D \} \models A \sqsubseteq D
\]

\[
\mathcal{J}_2 = \{ A \sqsubseteq \exists^1 r.C, \exists^1 r.C \sqsubseteq D \} \models A \sqsubseteq D
\]
We Want More!

• ... but these do not cover all possible “similarities”:

Example: Subexpressions

\[ \mathcal{J}_1 = \{ A \sqsubseteq X_1, \ X_1 \sqsubseteq D \} \models A \sqsubseteq D \]
\[ \mathcal{J}_2 = \{ A \sqsubseteq \exists r.C, \ \exists r.C \sqsubseteq D \} \models A \sqsubseteq D \]
We Want More!

... but these do not cover all possible “similarities”:

Example: Subexpressions

\[ J_1 = \{ A \sqsubseteq \textcolor{blue}{X1}, \textcolor{blue}{X1} \sqsubseteq D \} \models A \sqsubseteq D \]
\[ J_2 = \{ A \sqsubseteq \textcolor{purple}{X2}, \exists r. C \sqsubseteq D \} \models A \sqsubseteq D \]
We Want More!

- ... but these do not cover all possible “similarities”:

Example: Subexpressions

\[ \mathcal{J}_1 = \{ A \subseteq X_1, X_1 \subseteq D \} \models A \subseteq D \]
\[ \mathcal{J}_2 = \{ A \subseteq X_2, X_2 \subseteq D \} \models A \subseteq D \]
We Want More!

• ... but these do not cover all possible “similarities”:

Example: Subexpressions

\[ J_1 = \{ A \sqsubseteq X_1, X_1 \sqsubseteq D \} \models A \sqsubseteq D \]
\[ J_2 = \{ A \sqsubseteq X_2, X_2 \sqsubseteq D \} \models A \sqsubseteq D \]

Example: Different number of axioms

\[ J_1 = \{ A \sqsubseteq B, B \sqsubseteq C \} \models A \sqsubseteq C \]
\[ J_2 = \{ A \sqsubseteq B, B \sqsubseteq C, C \sqsubseteq D \} \models A \sqsubseteq D \]
We Want More!

• ... but these do not cover all possible “similarities”:

Example: Subexpressions

\[ J_1 = \{ A \sqsubseteq X_1, X_1 \sqsubseteq D \} \models A \sqsubseteq D \]
\[ J_2 = \{ A \sqsubseteq X_2, X_2 \sqsubseteq D \} \models A \sqsubseteq D \]

Example: Different number of axioms

\[ J_1 = \{ A \sqsubseteq B, B \sqsubseteq C \} \models A \sqsubseteq C \]
\[ J_2 = \{ A \sqsubseteq B, B \sqsubseteq \ldots \sqsubseteq D \} \models A \sqsubseteq D \]
Subexpression-Isomorphism

Definition: S-Isomorphism

Two justifications \((\mathcal{J}_1, \eta_1), (\mathcal{J}_2, \eta_2)\) are \(s\)-isomorphic \(((\mathcal{J}_1, \eta_1) \cong_s (\mathcal{J}_2, \eta_2))\) if there exists a justification \((\mathcal{J}, \eta)\) and two injective substitutions \(\phi_1, \phi_2\), such that \(\phi_1(\mathcal{J}) = \mathcal{J}_1, \phi_2(\mathcal{J}) = \mathcal{J}_2, \phi_1(\eta) = \eta_1, \text{ and } \phi_2(\eta) = \eta_2\).

- S-isomorphism is reflexive, symmetric, and transitive
  - It is an equivalence relation
  - It *partitions* a set of justifications
Lemma-Isomorphism

- **Lemma**: “intermediate proof step”
- Entailment of a subset of a justification [3]

Lemma-Isomorphism

- **Lemma**: “intermediate proof step”
- Entailment of a subset of a justification [3]

Example: Lemmatisation

\[
\{ A \sqsubseteq \exists r. B, \\
B \sqsubseteq C \\
C \sqsubseteq D \\
\exists r. D \sqsubseteq E \}\]

Justification

Lemma-Isomorphism

- *Lemma*: “intermediate proof step”
- Entailment of a subset of a justification [3]

**Example: Lemmatisation**

\[
\{ A \subseteq \exists r. B, \\
B \subseteq C \\
C \subseteq D \\
\exists r. D \subseteq E \} \quad \rightarrow \quad B \subseteq D
\]

Justification Lemma

Lemma-Isomorphism

- **Lemma**: “intermediate proof step”
- Entailment of a subset of a justification [3]

**Example: Lemmatisation**

\[
\begin{align*}
\{ A \subseteq \exists r.B, \\
B \subseteq C \\
C \subseteq D \\
\exists r.D \subseteq E \}
\end{align*}
\]

\[
\Rightarrow
\begin{align*}
B \subseteq D
\end{align*}
\]

\[
\Rightarrow
\begin{align*}
\{ A \subseteq \exists r.B, \\
B \subseteq D \\
\exists r.D \subseteq E \}
\end{align*}
\]

Lemma-Isomorphism

Definition: L-Isomorphism

Two justifications \((\mathcal{J}_1, \eta_1), (\mathcal{J}_2, \eta_2)\) are \(\ell\)-isomorphic \(((\mathcal{J}_1, \eta) \simeq_\ell (\mathcal{J}_2, \eta))\) if there exist lemmatisations \(\mathcal{J}_1^{\Lambda_1}, \mathcal{J}_2^{\Lambda_2}\) which are \(s\)-isomorphic: \(\mathcal{J}_1^{\Lambda_1} \simeq_s \mathcal{J}_2^{\Lambda_2}\).

- L-isomorphism with arbitrary lemmas is **not transitive**!
- Arbitrary lemmatisations may be **entirely different** from the original justification
  - e.g. lemmatisation = entailment
Lemma Restriction

- We need to restrict the selection of lemmas:
  - Allow only *summarising* lemmas
  - Allow only *obvious steps* to be substituted
- Frequent pattern: *Atomic subsumption chains*
  - ... they seem like a good start!
  - We substitute atomic subsumption chains in a justification with their entailment
BioPortal Survey

- 83 ontologies from BioPortal
  - 85 to 70,015 axioms (median: 962)
  - 23 to 33,913 classes (median: 552)
  - Expressivity: $\mathcal{AL}$ to $\mathcal{SROIQ}(\mathcal{D})$
- Computed 6,744 justifications in total
BioPortal Survey

- 83 ontologies from BioPortal
  - 85 to 70,015 axioms (median: 962)
  - 23 to 33,913 classes (median: 552)
  - Expressivity: $\mathcal{AL}$ to $\mathcal{SROIQ}(\mathcal{D})$
- Computed 6,744 justifications in total
- Large reductions visible across corpus:
  - 7.7 -11 justifications per template
  - all to iso: 90.9%
  - iso to s-iso: 25.7%
  - s-iso to l-iso: 15.8%
A potential application...

<table>
<thead>
<tr>
<th><strong>Entailments</strong></th>
<th><strong>S-iso templates</strong></th>
<th><strong>Strict iso templates</strong></th>
<th><strong>L-iso templates</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>abnormal_connective_tissue SubClassOf event</td>
<td>template 1 (2 axioms)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aspirin SubClassOf event</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pulmonary_disorder SubClassOf event</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>relative SubClassOf event</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stroke SubClassOf event</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aspirin SubClassOf disease_or_disorder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avoidance SubClassOf event</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bleeding_disorder SubClassOf event</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>body_piercing SubClassOf disease_or_disorder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cutaneous_bleeding SubClassOf disease_or_disorder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>musculoskeletal_bleeding SubClassOf disease_or_disorder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shaving SubClassOf disease_or_disorder</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Selected Instantiations</strong></th>
<th><strong>Instantiations of selected templates</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>aspirin SubClassOf disease_or_disorder</td>
<td></td>
</tr>
<tr>
<td>aspirin SubClassOf disease_or_disorder</td>
<td></td>
</tr>
<tr>
<td>aspirin SubClassOf disease_or_disorder</td>
<td></td>
</tr>
<tr>
<td>aspirin SubClassOf disease_or_disorder</td>
<td></td>
</tr>
<tr>
<td>musculoskeletal_bleeding SubClassOf disease_or_disorder</td>
<td></td>
</tr>
<tr>
<td>musculoskeletal_bleeding SubClassOf disease_or_disorder</td>
<td></td>
</tr>
<tr>
<td>musculoskeletal_bleeding SubClassOf disease_or_disorder</td>
<td></td>
</tr>
<tr>
<td>musculoskeletal_bleeding SubClassOf disease_or_disorder</td>
<td></td>
</tr>
</tbody>
</table>

**Selected Instantiations**

**Explanations**

**Explanation 1**

- musculoskeletal_bleeding SubClassOf disease_or_disorder
  1) musculoskeletal_bleeding SubClassOf has_treatment some treatment
  2) has_treatment Domain disease_or_disorder

**Explanation 2**

- musculoskeletal_bleeding SubClassOf disease_or_disorder
  1) musculoskeletal_bleeding SubClassOf has_treatment some generic_bleeding_treatment
  2) has_treatment Domain disease_or_disorder

**Explanation 3**

- musculoskeletal_bleeding SubClassOf disease_or_disorder
  1) musculoskeletal_bleeding SubClassOf has_age_of_onset some age_of_onset
  2) has_age_of_onset Domain disease_or_disorder
Another potential application...

\[ J_1 = \{ A \subseteq X_1, X_1 \subseteq D \} \models A \subseteq D \]
\[ J_2 = \{ A \subseteq X_2, X_2 \subseteq D \} \models A \subseteq D \]

\[ J_1 = \{ A \subseteq B, B \subseteq C \} \models A \subseteq C \]
\[ J_2 = \{ A \subseteq B, B \subseteq \ldots \subseteq D \} \models A \subseteq D \]
Another potential application...

\[ \mathcal{J}_1 = \{ A \subseteq X_1, ~ X_1 \subseteq D \} \models A \subseteq D \]
\[ \mathcal{J}_2 = \{ A \subseteq X_2, ~ X_2 \subseteq D \} \models A \subseteq D \]

\[ \mathcal{J}_1 = \{ A \subseteq B, ~ B \subseteq C \} \models A \subseteq C \]
\[ \mathcal{J}_2 = \{ A \subseteq B, ~ B \subseteq \ldots \subseteq D \} \models A \subseteq D \]

OWL developer: “That would be tremendously helpful”
Summary and Future Work

- We have introduced new equivalence relations:
  - Subexpression-isomorphism
  - Lemma-isomorphism
- We implemented isomorphism detection
- We surveyed a set of BioPortal ontologies
- Future work:
  - Exploit equivalence relations in OWL applications
  - Explore further obvious steps for lemmatisations
Template Frequency