On fixing semantic alignment evaluation measures

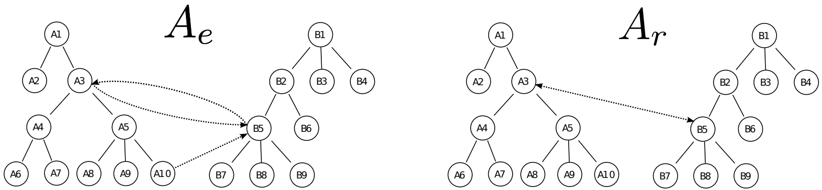
Jérôme David, Jérôme Euzenat

Workshop Ontology Matching 2008

Problems of precision and recall

These two alignments are equivalent :

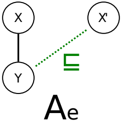
- $A_3 \sqsubseteq B_5$ and $A_3 \sqsupseteq B_5 \Leftrightarrow A_3 \equiv B_5$
- $A_3 \equiv B_5 \models A_{10} \equiv B_5$

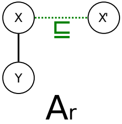


But with the classical model: *Precision* $= 0$ and *Recall* $= 0$!

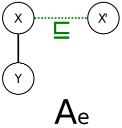
Semantic properties

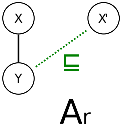
A solution : proposing measures respecting semantic properties [Euzenat, 2007]



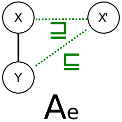


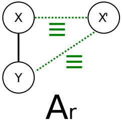
$A_r \models A_e \Rightarrow Precision = 1$
max-correctness





$A_e \models A_r \Rightarrow Recall = 1$
max-completeness





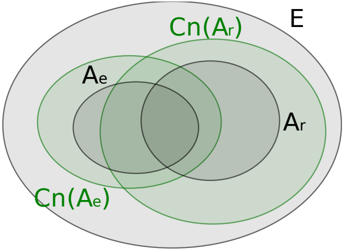
$A_r \equiv A_e \Leftrightarrow Precision = 1$ and $Recall = 1$
definiteness

1 - Ideal precision and recall

Replace A_e and A_r by their semantic closure $Cn(A_e)$ and $Cn(A_r)$
Semantic closure $Cn(\ldots)$ = set of correspondences deduced from alignment and ontologies

$$P_i = \frac{|Cn(A_e) \cap Cn(A_r)|}{|Cn(A_e)|}$$

$$R_i = \frac{|Cn(A_e) \cap Cn(A_r)|}{|Cn(A_r)|}$$



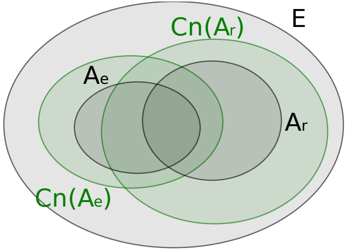
+ The three properties are satisfied
− Not always defined : $Cn(\ldots)$ could be infinite

2 - Semantic precision and recall

Use both alignments and their semantic closure

$$P_s = \frac{|A_e \cap Cn(A_r)|}{|A_e|}$$

$$R_s = \frac{|Cn(A_e) \cap A_r|}{|A_r|}$$



+ The three properties are satisfied
+ Always defined (contrarily to ideal precision and recall)
− But they still have some drawbacks...

Limitations of semantic precision and recall

Semantic precision and recall have **two drawbacks**:

1. Two semantically equivalent alignments could have different precision values
2. An alignment can have null precision and recall even if its semantic closure intersects those of the reference alignment

Limitations of semantic precision and recall

Semantic precision and recall have **two drawbacks**:

1. Two semantically equivalent alignments could have different precision values
2. An alignment can have null precision and recall even if its semantic closure intersects those of the reference alignment

Two other properties that a perfect semantic model must satisfies :

1. the **semantic-equality** property :

$$Cn(A_{e_1}) = Cn(A_{e_2}) \Rightarrow \begin{cases} P(A_{e_1}, A_r) = P(A_{e_2}, A_r) \\ R(A_{e_1}, A_r) = R(A_{e_2}, A_r) \end{cases}$$

2. the **overlapping-positiveness** property:

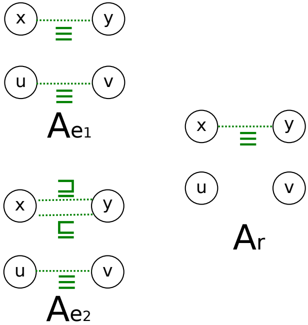
$P(A_e, A_r) = 0$ and $R(A_e, A_r) = 0$ iff $Cn(A_e) \cap Cn(A_r) = Cn(\emptyset)$

Limitations of semantic precision and recall

1st problem: Two semantically equivalent alignments could have different precision values

Case 1: problem occuring at alignment level:

a correspondence could be split into several correspondances



$A_{e_1} \equiv A_{e_2}$ but:

$P_s(A_{e_1}, A_r) = 1/2$

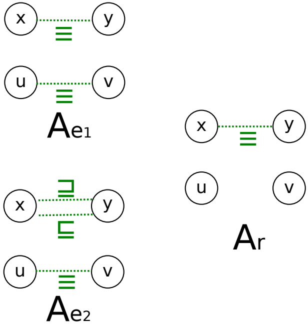
$P_s(A_{e_2}, A_r) = 2/3$

Limitations of semantic precision and recall

1^{st} problem: Two semantically equivalent alignments could have different precision values

Case 1: problem occuring at alignment level:

a correspondence could be split into several correspondances



$A_{e_1} \equiv A_{e_2}$ but:

$P_s(A_{e_1}, A_r) = 1/2$

$P_s(A_{e_2}, A_r) = 2/3$

A solution: normalize alignments

Syntactic normalisation of alignments

Goal of normalization: allows measures to satisfy the semantic-equality property when reasoning only at alignment level.

1. Use **alignment relation algebra**, i.e., write each alignment relation as a disjunction of elementary relations [Euzenat, 2008]
   - Elementary relations: $\Gamma = \{\sqsubset, \sqsupset, \equiv, \between, \perp\}$
   - Operators: meet ($\cup$), join ($\cap$), compose($.$), inverse($^{-1}$)
2. A pair of entities or formulas appear at most once in each alignment

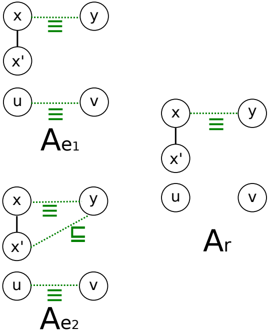
Examples :

- $x \sqsubseteq y$ becomes $x\{\sqsubset, \equiv\}y$
- $x \sqsubseteq y$ and $x \sqsupseteq y$ become $x\{\sqsubset, \equiv\} \cap \{\sqsupset, \equiv\}y$, i.e., $x\{\equiv\}y$

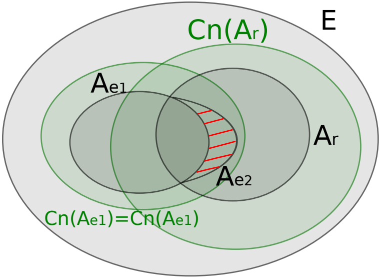
Limitations of semantic precision and recall

1st problem: Two semantically equivalent alignments could have different precision values

Case 2: problem occuring at ontological level (redundancy)



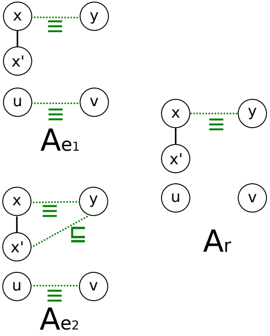
$$P_s(A_{e_1}, A_r) = 1/2$$
$$P_s(A_{e_2}, A_r) = 2/3$$



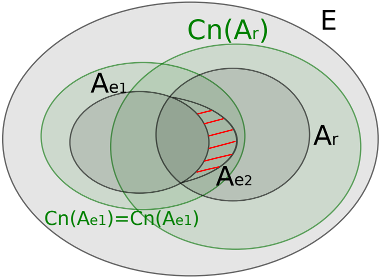
Limitations of semantic precision and recall

1st problem: Two semantically equivalent alignments could have different precision values

Case 2: problem occuring at ontological level (redundancy)



$P_s(A_{e_1}, A_r) = 1/2$
$P_s(A_{e_2}, A_r) = 2/3$

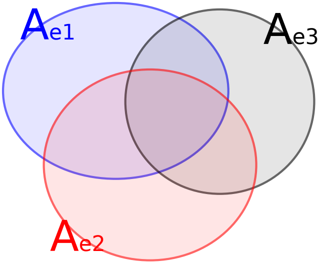


A solution: Λ-bounded precision and recall

Λ-bounded precision and recall

Idea: Restricting semantic closures to a set of alignments for enabling ideal precision and recall measures

Classical evaluation model:

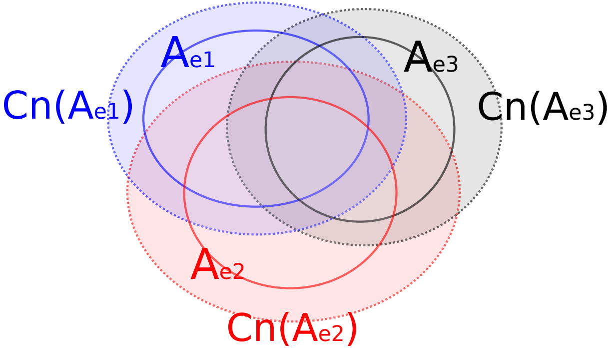


$$P = \frac{|A_e \cap A_r|}{|A_e|} \quad R = \frac{|A_e \cap A_r|}{|A_r|}$$

Λ-bounded precision and recall

Idea: Restricting semantic closures to a set of alignments for enabling ideal precision and recall measures

Ideal evaluation model (not always defined):

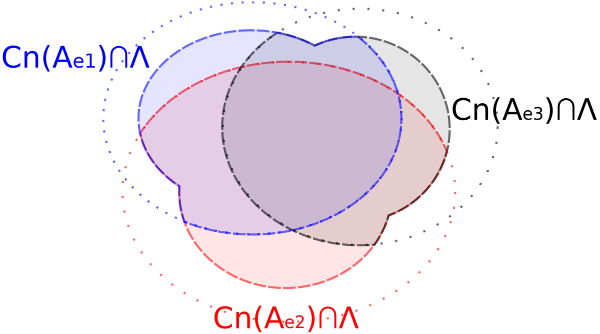


$$P_i = \frac{|Cn(A_e) \cap Cn(A_r)|}{|Cn(A_e)|} \quad R_i = \frac{|Cn(A_e) \cap Cn(A_r)|}{|Cn(A_r)|}$$

Λ-bounded precision and recall

Idea: Restricting semantic closures to a set of alignments for enabling ideal precision and recall measures

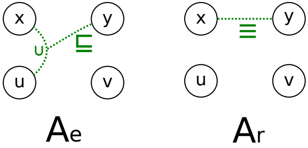
Bounded evaluation model (always defined):



$$P_i = \frac{|Cn(A_e) \cap Cn(A_r) \cap \Lambda|}{|Cn(A_e) \cap \Lambda|} \quad R_i = \frac{|Cn(A_e) \cap Cn(A_r) \cap \Lambda|}{|Cn(A_r) \cap \Lambda|}$$

Limitations of semantic precision and recall

2^{nd} problem: the semantic closures of A_e and A_r intersects but A_e has null semantic precision and recall values.



$$\left.\begin{array}{l} A_e \models x \sqsubseteq y \\ A_r \models x \sqsubseteq y \end{array}\right\} Cn(A_e) \cap Cn(A_r) = \{x \sqsubseteq y\}$$

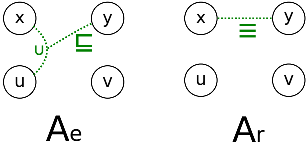
$$\text{but } P_s(A_e, A_r) = 0 \text{ and } R_s(A_e, A_r) = 0$$

Semantic relaxed precision and recall

Idea: introducing semantics in relaxed precision and recall [Ehrig and Euzenat, 2005]

- Relaxed measures are function of proximity functions σ between individual correspondences.
- New σ measures based on relation algebra

Example on σ precision: $\sigma_{prec}(x \cup u\{\sqsubset, \equiv\}y, x\{\equiv\}y)$?



A_e states $x \cup u\{\sqsubset, \equiv\}y$

O_1, O_2, and $x\{\equiv\}y \models x \cup u\{\sqsupset, \equiv\}y$

$$\sigma_{prec} = \frac{|\{\sqsubset, \equiv\} \cap \{\sqsupset, \equiv\}|}{|\{\sqsupset, \equiv\}|} = 0.5$$

Conclusion

- Identified specific problems remaining with semantic precision and recall
- Expressed them as properties
  - semantic-equality
  - overlapping-positiveness
- Defined two specific measures for countering them
  - Λ-bounded measures: do not provide absolute values
  - Relaxed semantic measures: properties are respected only at correspondence level
- Work to integrate them in a common framework

Conclusion

- Identified specific problems remaining with semantic precision and recall
- Expressed them as properties
  - semantic-equality
  - overlapping-positiveness
- Defined two specific measures for countering them
  - Λ-bounded measures: do not provide absolute values
  - Relaxed semantic measures: properties are respected only at correspondence level
- Work to integrate them in a common framework

Conclusion

- Identified specific problems remaining with semantic precision and recall
- Expressed them as properties
  - semantic-equality
  - overlapping-positiveness
- Defined two specific measures for countering them
  - Λ-bounded measures: do not provide absolute values
  - Relaxed semantic measures: properties are respected only at correspondence level
- Work to integrate them in a common framework

Marc Ehrig and Jérôme Euzenat.
Relaxed precision and recall for ontology matching.
In Benjamin Ashpole, Marc Ehrig, Jérôme Euzenat, and Heiner Stuckenschmidt, editors, *Proceedings of the Workshop on Integrating Ontologies*, volume 156, pages 25–32. CEUR-WS.org, 2005.

Jérôme Euzenat.
Semantic precision and recall for ontology alignment evaluation.
In *Proceedings of 20th International Joint Conference on Artificial Intelligence (IJCAI 07)*, pages 248–253, Hyderabad (IN), 2007.

Jérôme Euzenat.
Algebras of ontology alignment relations.
In *Proceedings of the 7th International Semantic Web Conference (ISWC 08)*, pages 387–402. Springer, 2008.