Exercise: Entity Linking Evaluation

November 12, 2016

In this exercise, you will evaluate the performance of entity linking systems. We provide you with two sample texts and the ground truth, and you need to evaluate the output of an entity linking system.

Note. For all the tasks we use partial matching of mentions.

Text 1

Ground truth:

Mention	Entity
mpla	<wikipedia:MPLA>
angola	<wikipedia:Angola>
1992 elections	<wikipedia:Angolan_presidential_election,_1992>
multiparty democracy	<wikipedia:Multi-party_system>

System output:

Mention	Entity
1992 elections	<wikipedia:Philippine_general_election,_1992>
angola	<wikipedia:Angola>
multiparty democracy	<wikipedia:multiparty_democracy>

Text 2

Ground truth:

Mention	Entity
sweet potato	<wikipedia:Sweet_potato>
bell pepper	<wikipedia:Bell_pepper>
tomato	<wikipedia:Tomato>

System output:

Mention	Entity
potato	<wikipedia:Sweet_potato>
perennials	<wikipedia:Perennial_plant>

Task 1

Given the above ground truth and system results, answer the following questions:

1. What is Precision for Text 1? $\frac{1}{3} = 0.33$

2. What is Precision for Text 2? $\frac{1}{2} = 0.5$

3. What is P_{macro}? $(\frac{1}{3} + \frac{1}{2})/2 = 0.41$

4. What is Recall for Text 1? $\frac{1}{4} = 0.25$

5. What is Recall for Text 2? $\frac{1}{3} = 0.33$

6. What is R_{macro}? $(\frac{1}{4} + \frac{1}{3})/2 = 0.29$

7. What is $F1$ score? $\frac{2*0.41*0.29}{0.41+0.29} = 0.17$

Macro-averaged evaluation metrics are computed as:

$$P_{macro} = \sum_{d \in D} \frac{|\hat{A} \cap A|}{|A|}/|D|, \qquad R_{macro} = \sum_{d \in D} \frac{|\hat{A} \cap A|}{|\hat{A}|}/|D|, \qquad F1 = \frac{2.P_{macro}.R_{macro}}{P_{macro} + R_{macro}}$$

where $\hat{A}$ denotes ground truth annotations, A is system generated annotations, and $|D|$ is the total number of documents (i.e. 2 in this exercise).

Task 2

Given the above ground truth and system results, answer the following questions:

1. What is P_{micro}? $\frac{1+1}{2+3} = 0.4$

2. What is R_{micro}? $\frac{1+1}{3+4} = 0.28$

3. What is $F1$ score? $\frac{2*0.4*0.28}{0.4+0.28} = 0.15$

Micro-averaged evaluation metrics are computed as:

$$P_{micro} = \frac{|\hat{A}_D \cap A_D|}{|A_D|}, \qquad R_{micro} = \frac{|\hat{A}_D \cap A_D|}{|\hat{A}_D|}, \qquad F1 = \frac{2.P_{micro}.R_{micro}}{P_{micro} + R_{micro}}$$

where $\hat{A}$ denotes ground truth annotations and A is system generated annotations.