



## Agreement after alignment improvements?

**Patrice Lopez** <patrice.lopez@science-miner.com>  
To: James Howison <jhowison@ischool.utexas.edu>  
Cc: "cfdu@utexas.edu" <cfdu@utmail.utexas.edu>

Wed, Mar 11, 2020 at 8:00 PM

My answer below.

On 11 Mar 2020, at 19:31 , James Howison <jhowison@ischool.utexas.edu> wrote:

Hi Patrice,

Caifan and I reviewed these "one coder found many, the other found none" and we're confident that the coder that found none did review the article. That's because that boolean comes straight from a "coded\_no\_mentions" statement in the RDF file they produced. That was not generated by default so they would have had to explicitly add that. So we will have to count those as disagreement.

So I have updated my agreement counts to cover that. From 228 initially, I obtain now 260 documents which are multiple-annotated (versus 258 for you, I have not checked the two edge cases yet!). Here are the revised agreement measures - it's a bit lower as expected, but the difference is smaller than I was thinking:

\*\*\*\*\* Inter-Annotator Agreement (Percentage agreement) \*\*\*\*\* PMC SET \*\*\*\*\*

number of documents annotated by multiple annotators: 260

software      agreement measure:    0.7072  
                standard error:        0.0001  
                confidence interval: [0.7069-0.7074]  
                number of agreements: 2876  
                number of samples:    4067

version-number agreement measure:    0.8087  
                standard error:        0.0004  
                confidence interval: [0.808-0.8094]  
                number of agreements: 896  
                number of samples:    1108

version-date   agreement measure:    0.4267  
                standard error:        0.0013  
                confidence interval: [0.4241-0.4293]  
                number of agreements: 160  
                number of samples:    375

creator        agreement measure:    0.4217  
                standard error:        0.0002  
                confidence interval: [0.4213-0.4221]  
                number of agreements: 1093  
                number of samples:    2592

url            agreement measure:    0.6682  
                standard error:        0.0021  
                confidence interval: [0.664-0.6723]  
                number of agreements: 149  
                number of samples:    223

all fields     agreement measure:    0.6185  
                standard error: 0.0001  
                confidence interval: [0.6184-0.6186]  
                number of agreements: 5174

number of samples: 8365

\*\*\*\*\* Inter-Annotator Agreement (Percentage agreement) \*\*\*\*\* ECON SET \*\*\*\*\*

number of documents annotated by multiple annotators: 26

software agreement measure: 0.7206  
standard error: 0.0004  
confidence interval: [0.7198-0.7214]  
number of agreements: 797  
number of samples: 1106

version-number agreement measure: 0.7672  
standard error: 0.0008  
confidence interval: [0.7656-0.7688]  
number of agreements: 402  
number of samples: 524

version-date agreement measure: 0.4335  
standard error: 0.0021  
confidence interval: [0.4293-0.4376]  
number of agreements: 101  
number of samples: 233

creator agreement measure: 0.592  
standard error: 0.0005  
confidence interval: [0.5911-0.593]  
number of agreements: 595  
number of samples: 1005

url agreement measure: 0.713  
standard error: 0.0042  
confidence interval: [0.7048-0.7212]  
number of agreements: 77  
number of samples: 108

all fields agreement measure: 0.6626  
standard error: 0.0002  
confidence interval: [0.6623-0.6629]  
number of agreements: 1972  
number of samples: 2976

However we think this should be reported in (at least) two steps:

1. How many overlapping full\_quotes were identified? (straight percentage agreement)
2. Given two coders looking at an over-lapping full\_quote, what agreement was there on the other codes?

So these "zero found" examples should reduce the agreement for step 1, but not have impact on step 2 (since there wasn't a chance for agreement).

Thoughts?

This is an interesting distinction indeed! To realise such report in two steps, I would need to rewrite a bit my agreement count process. Do you plan to implement it ?

Patrice

[Quoted text hidden]