**To do list for IAA vs F-measure: Due April 13th**

|  |  |  |
| --- | --- | --- |
| X | Kevin | 1. Add more introduction about the comparison between IAA and F-measure in the intro - Kevin |
|  | Mayla | 1. Put reference numbers in the order that they appear |
| X | Mayla | 1. Use subscript 1 in F1 measure - F­1 measure - DONE |
| X | Mayla | 1. Remove redundancies in the 2 sentences in intro: Typically we compare the IAA score to the precision, recall, or F1 measure of the computer's performance. To evaluate computer annotation performance we use precision, recall, or F1 measure. (removed this sentence in next paragraph) - DONE |
| X | Kevin | 1. Rephrase precision and recall to better explain what it is here: Precision (P) is the proportion of actual correct answers to computer output correct answers (true positives). Recall (R) is the proportion of actual correct answers to computer output correct answers (true positives) plus answers that the computer deemed incorrect but are actually correct. - Kevin |
|  | Kevin | 1. Fix spelling errors: |
|  | Kevin | * 1. "We can also weight". I think "weigh" is better here. – weight! Missing e in calculatd  IAA verse F1: rather "versus" Figures: should be called "figure", not "graph" Rephrase "when we look at ..." Generally, critically look at the use of "look at", and "we see", and similar constructions - Kevin |
| X | Kevin | 1. Rephrase sentence to fix prioritization: "In order to make funding decisions and provide accurate information to the public about technology" |
|  | Kevin | 1. The higher the IAA only tells you that ... Rephrase - Kevin |
|  | Mayla | 1. References: 5 should include a url to the github repo. 18 should include url or similar reference. |
| X | Kevin | 1. Describe better how the literature search was performed – Kevin |
| X | Kevin | * 1. On which repositories was the search performed? |
| X | Kevin | * 1. What key words were used? |
| X | Kevin | * 1. How many papers resulted from the search? |
| X | Kevin | * 1. How many discarded to find the 12 papers? |
| X | Kevin | * 1. Reasons: robustness of approach and better motivate why such few articles |
|  | Kevin/Mayla | Add Carpenter citations on Bayesian approach |
|  | Kevin | Check comments from Patrick |
|  | Kevin | Replace “Jacob Cohen” |
|  | Kevin | Re-think the adj/ann solution |
|  | Kevin | Problems with determining the estimated chance agreement leading to under-stated kappa |
|  | Kevin | Point out that alternative to correlation is noise, and noise is not what we see |
|  |  | Redo figures |