

Dear Editor, Reviewers,

First we would like to thank both reviewers for the time invested in reviewing our manuscript. We have made revisions to the manuscript according to the detailed responses posted in the interactive discussion (see below). We rephrased objective 2. We trimmed the abstract and removed the NLP jargon to make it more accessible. We also updated the topic modelling section, performed the analysis again and discussed it more. Following the recommendation of RC2, we also move one figure and one table to the appendix. We also made some minor modifications throughout the manuscript.

RC1

Interesting study regarding the use of natural language processing methods to extract information from the growing volume of scientific literature. The authors not only illustrate the use of different algorithms but also try to evaluate them numerically. In general, a well written manuscript. However, I think there is a lack of discussion and some of their objectives/aims are weakly met. The "relationship extraction" section is interesting and well written and the authors might want to put the same effort in the rest of the sections.

We appreciate that you find the study interesting and we thank you for your useful comments on the content that helped to improve the manuscript. We would like to state that the primary aim of the study was to demonstrate a practical workflow of several NLP techniques for summarising a large body of scientific literature. This was not properly reflected in the aims of our study. We have modified the aims accordingly in the revised version of the manuscript.

We acknowledge that the "topic analysis" part is less developed and weakly matched the objective 2 of addressing if a paper was relevant or not to a topic. In this regard, we restructured the content around topic classification in the manuscript. Instead of classifying "new papers" in different topics, we now demonstrate how to identify groups of manuscripts (in our case, groups around different types of "agricultural practices") and observe which groups are less represented (or absent). In this way, we can show practices less studied and identify possible knowledge gaps. This also serves as a first classification to identify on which topic would a meta-analysis be well suited for instance.

Comments

- Abstract: The beginning abstract seems a bit disconnected with the rest of the manuscript. Climate change is a hot topic but the paper itself is not related to that. I would suggest re-framing the abstract to match the content of the manuscript.

We have rephrased the abstract such that the main focus is on the NLP techniques to summarise a large body of scientific environmental literature and then present the OTIM en Meta corpus as a case study on which we applied these techniques.

- Assessing the ability of an algorithm such as regex: I find this evaluation a bit strange. The algorithms itself is infallible in the sense that it always finds what you tell it to find if it is present in the text. The algorithm is only restricted by the capacity of the user to generate valid regular expressions.

We agree that the regex algorithm is infallible but indeed, in this case, we want to estimate how well user-defined regexes are able to recover specific information. We have made this clear in the manuscript that we do not assess the ability of the regex algorithm itself but rather the ability of the user-generated regular expressions to match relevant content considering the trade-off between generality and their specificity.

- Topic modelling: There is no discussion.

Further discussion has been added, especially on how topic classification can be used as one of the first steps of the presented semi-automated NLP workflow for information summary and identifying groups of abundant literature where a meta-analysis can be useful.

- How did you achieve your second aim (to illustrate the ability of topic classification to classify a new paper as relevant to a given topic)?

(see general comment)

- You mention that topic modelling "can help identify knowledge gaps". How? Did you find any? If your aim is to present a practical workflow, perhaps you should guide the user to achieve that.

We agree that a practical interpretation is a useful addition to the manuscript. We now give an example in the manuscript (around irrigation).

- Why did you select 6 topics instead of 9. You only mention that you are trying to maximise the coherence, which is higher for 9 topics.

That is a fair point and has been corrected in the revised version of the manuscript.

- How does the number of topics might affect your workflow? Is selecting the highest coherence score infallible?

It is not infallible and we found that choosing a number of topics between 6 and 9 topics tends to lead to the same groups. The variability in coherence for each number of topics can be large, especially for a relatively small corpus as we have. This is now discussed in the revised version of the manuscript.

- Could you elaborate on how excluding monograms increased the coherence? From the term frequencies (Fig 7) I do not see many soil-related terms, which seems strange. Perhaps they were ignored since they appeared as monograms? I do agree that bi- and even trigrams are important but I have usually seen them added to a selection of monograms.

In our case, the inclusion of monograms led to words like 'soil', 'treatment', 'water', 'crop' or 'tillage' to appear prominently in the different topics. This did not allow us to differentiate the topic so well and the average topic coherence in this case was $C_v = 0.4$. With only bi-grams, some of these words carried more meaning: "conventional tillage", "soil water", "cover crop" and hence enabled better to see what the topic is about. This is the reason why, in this case, we preferred to only use bi-grams. This remark is a good point and we recognize that the addition of monograms as seen in other work can sometimes help. This is now discussed in the revised manuscript.

General:

Overall this manuscript fits well with SOIL, and the methodology as well as the results will be of interest to readers. The nature of the study, involving "natural language processing for metadata extraction from environmental {soil} scientific publications" is inherently multidisciplinary, and complex! The necessary methods are well discussed and well referenced, and the appendix of the NLP software will be a big help to researchers in this field. The results relating agricultural practices and soil and site properties are novel and important.

We appreciate that you find this manuscript well suited for the journal SOIL and more specifically to a multi-disciplinary topic related to agricultural practices. We are also glad to hear that our effort towards a reproducible workflow (by the means of notebooks, github repository) is acknowledged.

Specific:

Most SOIL readers are probably substantially unfamiliar with NLP and would benefit from more focused guidance by the authors, which can be accomplished perhaps mostly easily by a trimmed revision. For example the Abstract is overly complex; the Introduction states the objectives of the study on just four lines 96-100, and a trimmed Abstract could focus simply on the achieving of the objectives.

Agree. As mentioned in reply to RC1, we refocused the abstract around "NLP techniques" and the objectives we want to address in this work. Additionally, we made sure that the NLP specific language is explained and simplified to make the abstract accessible to most.

The Material and Methods section is appropriately long, given the emphasis on methods, but could be edited to be more uniformly coherent. Perhaps part of that could be fixed by reformatting the variety of figures, and relegating some of them to just the appendix.

Figure 3 and Table 2 are now part of the appendix to ease the flow through the Material and Methods section.

Most of the figures in the Results section are important, but much of the other discussions in Results are really recommendations and can be eliminated or partly moved to Conclusions.

Thank you for the feedback. We have made some small edits but found that it was sometimes more useful to leave the recommendation/discussion close to the results.

Technical:

I see Reviewer #1 listed some technical issues, most of which I believe can be handled by trimming as suggested.

See reply to RC1.