National Library of Medicine

Supplemental Q&A from NLM Office Hours

February 28, 2024

1. Did you or will you publish a paper with the technical details of MTIX regarding the Machine Learning algorithm applied?
    a. This paper describes the MTIX machine learning algorithm used for Descriptors: https://ceur-ws.org/Vol-2936/paper-22.pdf, and this paper describes the machine learning algorithm for Qualifiers https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8075546/

2. How does MTIX determine which of the MeSH terms are major concepts?
    a. The MTIX machine learning algorithm assigns a score to each predicted Descriptor. Descriptors with high scores (above a decision threshold) are designated as major topics/IM. However, there is a list of Descriptors (e.g., Check Tags) that can never be major topics/IM.

3. What do you suggest for adjusting searches for searching checktags with a lower recall?
    a. To find additional articles that the algorithm may have missed, we suggest casting a wider net by 1) using additional related MeSH terms, such as “Infant” rather than just “Infant, Newborn” and 2) using related keywords in addition to MeSH terms.

4. Will MTIX be used to develop or inform new MeSH terms?
    a. We have no plans to use MTIX to develop new MeSH terms.

5. I remember talk a number of years ago about a group of articles that were accidently indexed twice. Check tags were the most consistently used part of the indexing. I do not remember the results of the regular MeSH terms. Has NLM looked at this to see how MTIA/MTIX compares?
    a. Yes, checktags do appear to be the part of indexing applied with the greatest inter-indexer consistency, and both MTIA and MTIX are also at their strongest with checktags. The consistency with which different indexers used various terms in the training data has a significant impact on MTIX performance. That means MTIX performance is generally stronger with concepts that are more clearly defined, distinct entities, because indexers have applied these terms in more consistent ways, and weaker with concepts that are more abstract and overlapping, because indexers have made more variable choices in the vocabulary used to represent those ideas.

6. I've noticed that there are MeSH terms that dropped significantly in usage when the switch to auto-indexing happened (for example 'Sensitivity and Specificity' or 'United Kingdom'). I don't know whether that was over-indexing in the past or under-indexing now but I'm interested to know whether NLM have performed any analyses based on indexing changes over time?
    a. We have not yet performed a formal analysis, but it is something we are aware of and would like to track more closely in the future. Since MTIX relies on statistical probabilities to decide on indexing terms, it seems likely it will have a vocabulary distribution that more closely resembles recent manual indexing than MTIA.

7. What happens to abstracts which are longer than MTIX’s input limits? Are they cut?

a. The MTIX machine learning pipeline includes a convolutional neural network (CNN) model and two fine-tuned BiomedBERT models. The CNN model has maximum input lengths of 64 and 768 tokens for the title and abstract, respectively. The BERT models have a maximum input length of 512 tokens, and the input to the BERT models is the concatenated journal name, title, abstract, and candidate MeSH term names. The maximum input lengths of the CNN model are rarely exceeded, but we do sometimes have to truncate the end of the abstract for the BERT models. We average the predictions of the three pipeline models and in most cases at least two out of three models will have seen the full title and abstract text.

8. Do you expect a noticeable increase in volume with saved searches and alerts as a result of the MTIX indexing changes?
    a. We do not expect there to be a noticeable increase in volume with saved searches and alerts when the shift from MTIA to MTIX indexing occurs later this spring. However, if you have any questions about the results of your saved searches, please write to the help desk.

9. I worry that indexing on title and abstract will mean there is no way to search for adverse events which are often secondary outcomes and only in the full-text. I just wonder if you have any suggestions?
    a. This could be a potential impact of indexing on title and abstract alone. We don't have any specific suggestions at this time, but this is certainly an area that may merit further study and evaluation.

10. How does NLM manage text word indexing? Particularly, device names.
    a. Typically, a single word or term is available to be searched in PubMed as soon as it appears on a record. If a device name is comprised of a single word, you can search for the device name in double quotes.

       A phrase made up of multiple words must appear in the phrase index if you want to search for that exact phrase without Automatic Term Mapping. You can check if a device name is included in the phrase index by going the Advanced Search page, typing in your phrase, and then clicking Show Index which is located under the Add button. If the phrase is included in the phrase index, it will show in this list.

       If the device name you are searching for does not appear yet in the phrase index, you might consider using a proximity search with an N value of 0 to look for that term. For example, the phrase "cardiac and cancer care" does not yet appear in the phrase index. To search for this phrase, you might try the proximity search:

       "cardiac and cancer care"[tiab:~0]

       For more information about phrase searching and proximity search, please see the PubMed user guide:
       - Searching for a phrase: https://pubmed.ncbi.nlm.nih.gov/help/#searching-for-a-phrase

- Browsing the index of terms: https://pubmed.ncbi.nlm.nih.gov/help/#browsing-show-index
- Proximity searching: https://pubmed.ncbi.nlm.nih.gov/help/#proximity-searching

11. Have you analyzed how well MTIX performs on searches of varying complexity, e.g., complex concepts to search like health services topics versus simpler concepts like well described single clinical indications?
    a. The consistency with which different indexers used various terms in the training data has a significant impact on MTIX performance. That means MTIX performance is generally stronger with concepts that are more clearly defined, distinct entities, because indexers have applied these terms in more consistent ways, and weaker with concepts that are more abstract and overlapping, because indexers have made more variable choices in the vocabulary used to represent those ideas.
    Where concepts are inherently fuzzy and overlapping, searchers have typically needed to use multiple related MeSH terms and keywords for comprehensive retrieval. This will continue to be the case.

12. Would MTIX index graphic medicine differently than MTIA?
    a. Both MTIX and MTIA accept only text as inputs. Therefore, any differences in how they indexed graphic medicine would be limited to differences in how they treat the available text from that publication, much like any other abstract.

13. What are the search precision implications of automatic indexing, i.e., how reliably does automatic indexing assign the correct MeSH at the correct tree level?
    a. It's difficult to answer this precisely, for the reasons covered in the presentation. For MTIX, the overall raw precision is 76% and raw recall is 72%. If we extrapolate from the results of our small evaluation, we assume about 2/3 of precision errors and 1/5 of recall errors are actually as good or better as the original indexing (meaning, these "errors" are actually correct terms from the correct tree level). In that case, MTIX has an overall precision of about 90% and recall of about 77%.

14. Is NLM tracking known automatic indexing errors and if so, is that information publicly available? We will be able to identify therefore certain indexing elements (e.g., specific MeSH, publication types, check tags, etc.) that are more prone to automatic indexing errors so we can change how we search.
    a. This depends on the type of error and the algorithm. When we discover MTIA has applied a term inappropriately because of a bad trigger term (often an ambiguous acronym), we modify the MTIA rule or trigger list as needed, and we also find and correct any other similar occurrences in PubMed. Additionally, we maintain a list of terms that MTIA may use incorrectly when it encounters metaphorical language. Curators review all articles indexed with these terms. Because we correct these in curation for MTIA and because MTIX typically does not make this kind of error, you should not usually need to adjust your searches to avoid this type of mistake.

       In terms of more general errors, we have also looked at MTIX performance for the

different trees of the MeSH vocabulary. We typically see too much variability between test sets to have reliable statistics for individual MeSH terms unless they are very frequent, like checktags and certain publication types. However, to repeat an answer given above, MTIX performance is generally stronger with concepts that are more clearly defined, distinct entities, because indexers have applied these terms in more consistent ways, and weaker with concepts that are more abstract and overlapping, because indexers have made more variable choices in the vocabulary used to represent those ideas. So, it typically performs better in areas like organisms, chemicals, diseases, and anatomy, and lower in areas like psychology, sociology, policy, and methods. In general, recall takes a bigger hit than precision on those weaker areas, so MTIX is more likely to omit a term than to apply it incorrectly.