Summarising Literature on Mental Health in Academia: Machine Learning Methods and Human Expertise

Olga Lezhnina and Gábor Kismihók
Learning and Skill Analytics Research Group,
TIB -Leibniz Information Centre for Science and Technology, Hannover, Germany

Purpose

Studying mental health in academic environment is a complicated topic, which was for a long time underrepresented in literature (Gurtie et al., 2017). In recent years, research findings have been accumulated in the area (Mattijssen et al., 2021), and their number is growing. The findings were summarized in systematic reviews, such as Sabagh et al. (2018) on faculty burnout, Hazell et al. (2020) on mental health of doctoral researchers, or Salimyedeh et al. (2021) on coping with stress in academia. However, this traditional approach deals with rather limited samples of publications: for instance, 36 papers in Sabagh et al. (2018), 22 papers in Hazell et al. (2020), and 52 papers in Salimyedeh et al. (2021). In this study, we rely upon advanced machine learning techniques that make it possible to obtain a wider picture of the area on larger samples of literature. Our task is to illustrate how machine learning methods and human expertise complement each other in summarizing literature on mental health in academia.

Design

Web scrapping and data analysis was conducted with Python on two samples of publications. The first was gathered from PubMed[1] as a reliable health-related resource, and the second was collected by ReMo experts[2]. We used abstracts of the papers for our analysis, as they are concise, informative, and comparable with each other (Daenekindt & Huisman, 2020). For scrapping the PubMed database, we specified a query with such terms as “mental health” and “student”, “postdoc” or “graduate” in the title or abstract. The query had to be amended iteratively; for instance, the term “student” had to be supplemented with “…NOT t-test” after discovering that the query yielded papers unrelated to academia but containing “Student’s t-test” among the methods listed in the abstract.

[1] https://pubmed.ncbi.nlm.nih.gov/
[2] See Researcher Mental Health Library (Zotero).

In both samples, rows with missing abstracts were removed. The resulting PubMed sample consisted of 1289 papers from 454 sources, and the ReMo sample of 245 papers from 102 sources. The publication years were in the range from 1970s to nowadays, with most papers in both samples published after 2015, as shown in Figure 1.

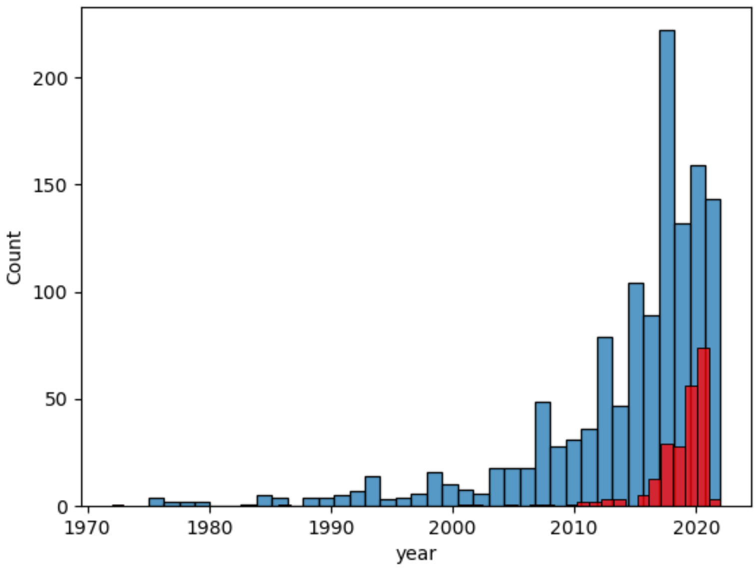

Figure 1. Publication year histograms for the PubMed sample (blue) and the ReMo sample (dark red).

To summarize the abstracts in each sample, topic modelling – an unsupervised machine learning method using patterns of word co-occurrence to reveal latent topics in the texts – was applied (Chaney & Blei, 2012). Topic modeling was conducted with Latent Dirichlet Allocation (LDA, see Blei et al., 2003). Relationships between most relevant words in each topic were visualized with graphs (Miranda-Jiménez et al., 2014).

Results

To determine the number of topics in each sample, the coherence score (Röder et al., 2015) was used. The results, as presented in Figure 2, indicated that three topics were optimal for the PubMed sample, and two topics for the ReMo sample.

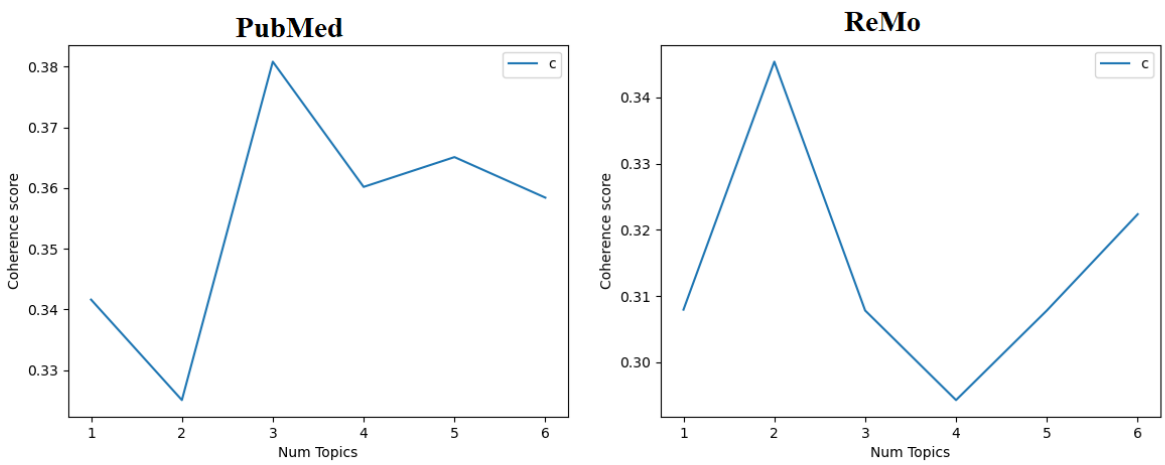

Figure 2. The coherence score graphs for the PubMed sample (on the left) and the ReMo sample (on the right).

LDA for each sample was explored with interactive visualisations. A screenshot of Jupiter notebook visualisation for the PubMed sample is presented in Figure 3. The left panel shows the general view of the model: the size of each circle indicates how prevalent the topic is, and the distance between the circles indicates how the topics relate to each other. The right panel shows the relevant words (tokens) for varying values of relevance metric λ, which can be interactively adjusted with a slider.

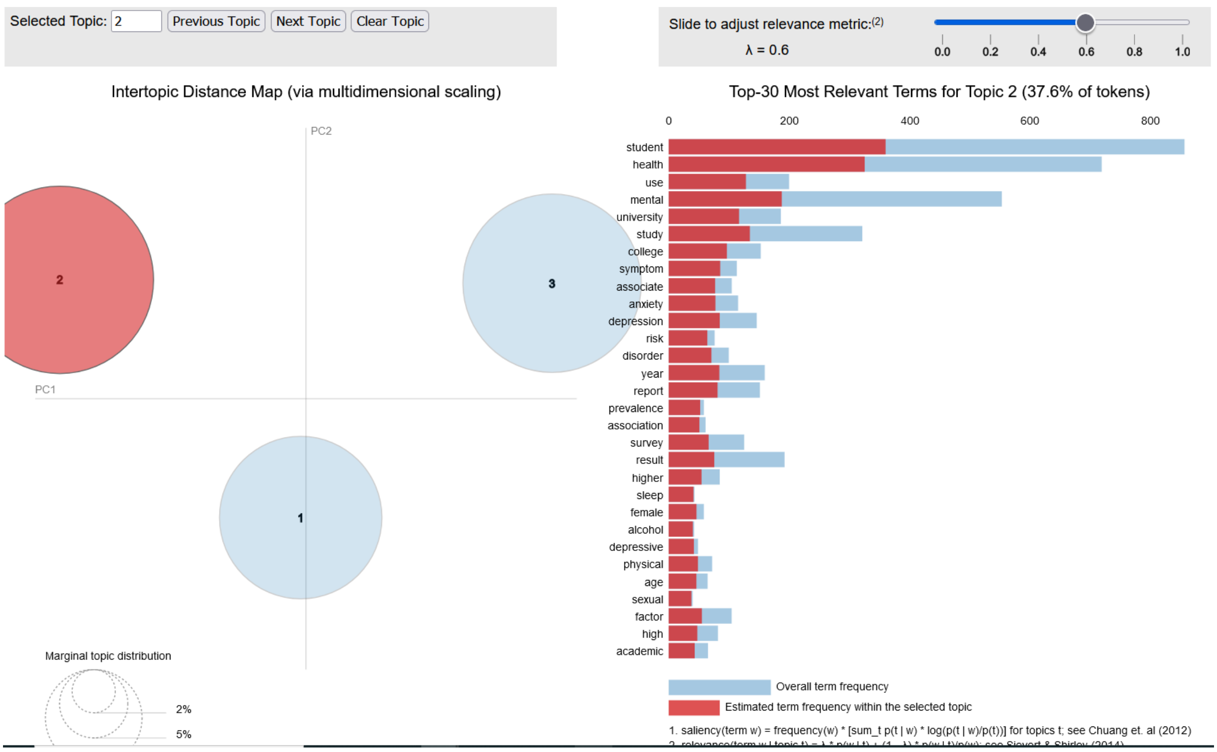

Figure 3. The screenshot of the LDA visualisation for the PubMed sample, with the largest topic selected.

The screenshot depicts the model with three well-separated topics and relevant words for the largest topic with λ = .60. The relevant words in the topic (with λ = 1) can be visualised as nodes in a graph, in which edges are understood as co-occurrences of these words. Such a graph for the largest topic in the PubMed sample is shown in Figure 4.

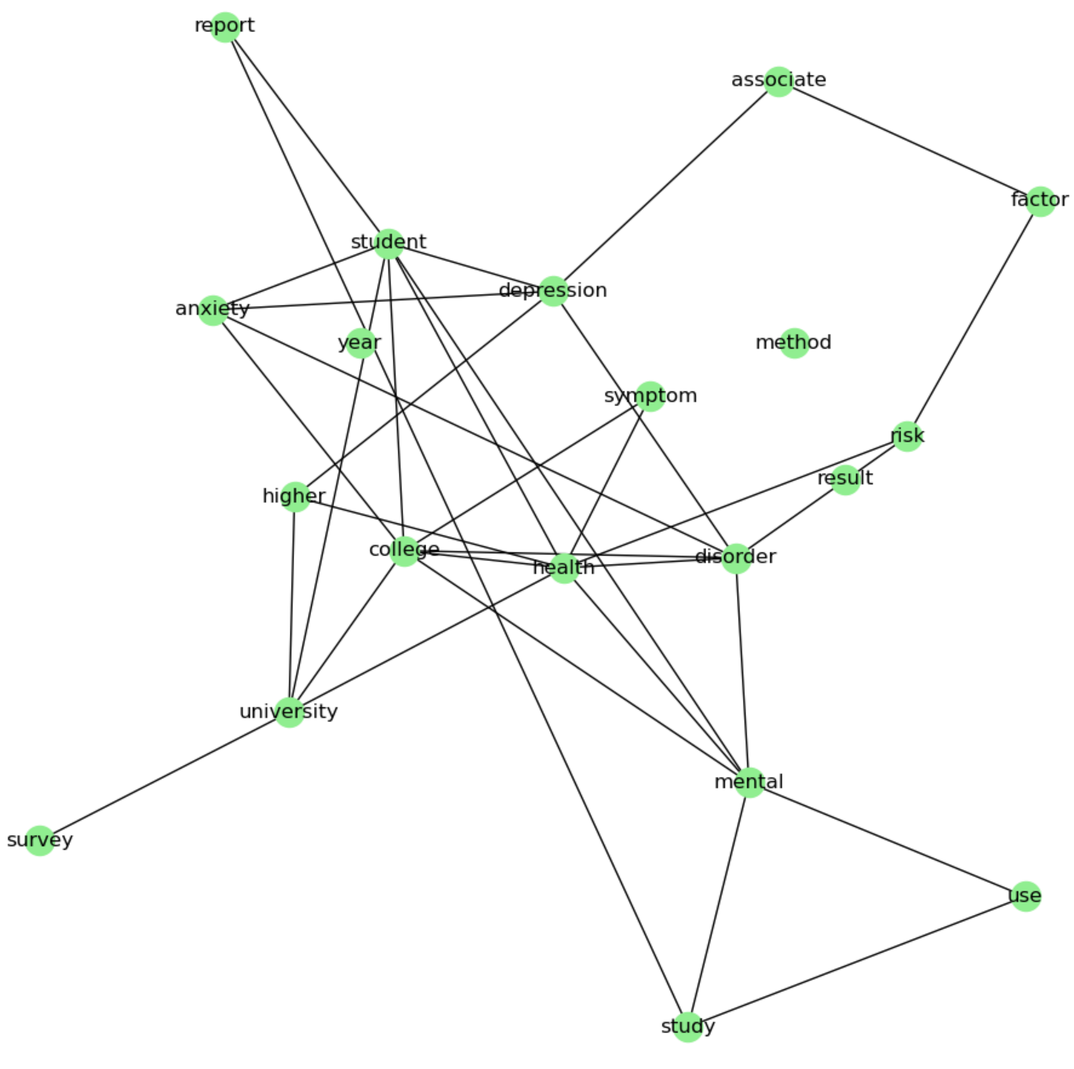

Figure 4. A graph visualising the relations between most relevant words in the largest topic of the PubMed sample.

Three topics in the PubMed sample were:

1) "Medical treatment-related", with 28.3% of tokens and most relevant words including 'medical', 'stress', 'service', 'help', 'treatment', 'patient'.
2) "Mental symptom-related", with 37.6% of tokens and most relevant words including 'disorder', 'risk', 'symptom', 'factor', 'depression', 'anxiety'.
3) "Nursing training-related", with 34.1% of tokens and most relevant words including 'training', 'intervention', 'nurse', 'practice', 'nursing', 'learning'.

Two topics in the ReMo sample were:

1) "Intervention-related", with 50.4.6% of tokens and most relevant words including 'support', 'intervention', 'stress', 'social', 'provide', 'depression'.
2) "Career-related", with 49.6% of tokens and most relevant words including 'researcher', 'doctoral', 'group', 'supervisor', 'working', 'career'.

For the three topics in the PubMed sample, as well as for the two topics in the ReMo sample, interactive LDA visualisations and graphs were constructed (see Supplemental Materials and the code on the GitHub). These results of automated text analysis can be further used to explore the topics in detail, to select papers assigned to a specific topic, to find out which concepts prevail in the literature and which are underrepresented, etc.

Implications

As scientific research is based on accumulation and processing of previous findings, it is vital to develop effective methods of summarising the literature in the area. Machine learning methods are useful tools of inductive data-driven approach, which gives an overview of the area based on a large sample of papers; in our study, more than a thousand abstracts from PubMed papers were analyzed. It is crucial, however, to employ human experts on certain stages of analysis. In our study, the query for the PubMed sample had to be amended by the authors, while the sample collected by the ReMo experts was used without additional adjustments. Interpreting the results of machine learning methods (such as topics and graph relations) also requires domain-specific knowledge of human experts. Therefore, integration of machine learning and human expertise might be considered the most effective method of summarizing literature on mental health in academia and needs to be more widely applied in the area.

Resources

Blei, D., Ng, A. & Jordan, M. (2003). Latent Dirichlet Allocation. *The Journal of Machine Learning Research*. 3. 601-608. Retrieved from https://dl.acm.org/doi/10.5555/944919.944937

Chaney, A., & Blei, D. (2012). Visualizing Topic Models. *Proceedings of the International AAAI Conference on Web and Social Media*, 6(1), 419-422. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/14321

Daenekindt, S., & Huisman, J. (2020). Mapping the scattered field of research on higher education. A correlated topic model of 17,000 articles, 1991–2018. *High Educ* 80**,** 571–587. https://doi.org/10.1007/s10734-020-00500-x

Guthrie, S., Lichten, C.A., van Belle, J., Ball, S., Knack, A., & Hofman, J. (2017). Understanding mental health in the research environment: A Rapid Evidence Assessment. *Santa Monica, CA: RAND Corporation.* https://www.rand.org/pubs/research_reports/RR2022.html

Hazell, C.M., Chapman, L., Valeix, S.F. et al. (2020). Understanding the mental health of doctoral researchers: a mixed methods systematic review with meta-analysis and meta-synthesis. *Syst Rev* 9, 197. https://doi.org/10.1186/s13643-020-01443-1

Mattijssen, L., Bergmans, J. E., van der Weijden, I., & Teelken, J. C. (2021). In the eye of the storm: the mental health situation of PhD candidates. *Perspectives on medical education*, 10(2), 71–72. https://doi.org/10.1007/s40037-020-00639-4

Miranda-Jiménez, S., Gelbukh, A. & Sidorov, G. (2014). Conceptual Graphs as Framework for Summarizing Short Texts. *International Journal of Conceptual Structures and Smart Applications*. 2. 55-75. 10.4018/IJCSSA.2014070104.

Röder, M., Both, A. & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. *Proceedings of the eight International Conference on Web Search and Data Mining, Shanghai, February 2-6.*

Sabagh, Z., Hall, N.C., & Saroyan, A. (2018) Antecedents, correlates and consequences of faculty burnout, *Educational Research*, 60:2, 131-156, DOI: 10.1080/00131881.2018.1461573

Salimzadeh R., Hall N.C., & Saroyan, A. (2021) Examining Academics' Strategies for Coping With Stress and Emotions: A Review of Research. *Front. Educ*. 6:660676. doi: 10.3389/feduc.2021.660676