Visualizing the Pace of COVID-19 Research: An Experimental Study of All India Institute of Medical Sciences (AIIMS), New Delhi

Manika Lamba[1][0000−0002−2022−3098]

University of Delhi, India
lambamanika07@gmail.com
https://manika-lamba.github.io

Abstract. With the emergence of a new variant of coronavirus in 2019, it became important to devote a significant amount of time to reading and identifying relevant studies to better understand COVID-19. It is urgent to sort out important, effective, and meaningful information from large databases to guide scientific research and promote the proper control, prevention, diagnosis, and treatment of COVID-19. Thus, this paper presents an experimental study that shows the application of topic modeling, and data visualization for COVID-19 research from 2010 to 2021 for the All India Institute of Medical Sciences (AIIMS), New Delhi, India. The results unveil the focus of scientific research, thereby giving deep insights into how the Indian medical society contributes to combating the COVID-19 pandemic.

Keywords: COVID-19 · Topic Modeling· Structural Topic Model (STM)

1 Introduction

Scientists have been studying the human coronavirus since its discovery in the 1960s (Deng et al., 2020). With the emergence of a new variant of coronavirus in 2019, it became important to devote a significant amount of time to reading and identifying relevant studies to better understand COVID-19. It is urgent to sort out important, effective, and meaningful information from large databases to guide scientific research and promote the proper control, prevention, diagnosis, and treatment of COVID-19. Topic modeling can be used as a way to identify and visualize the pace of COVID-19 from scientific documents.

Thus, this paper presents an experimental study that shows the application of topic modeling, and data visualization for COVID-19 research from 2010 to 2021 for the All India Institute of Medical Sciences (AIIMS), New Delhi, India. This study will help to (i) determine the core topics published by which AIIMS, New Delhi researchers related to COVID-19; and (ii) visualize the results using a dashboard and storyboard.

2   Literature Review

2.1   All India Institute of Medical Sciences (AIIMS), New Delhi

AIIMS is a public hospital and medical research university based in New Delhi, India. The institute is governed by the AIIMS Act, 1956, and operates autonomously under the Ministry of Health and Family Welfare of India. AIIMS, New Delhi was ranked 1st in South Asia in 2020 in the category of Life Sciences and Medicine by QS WUR (AIIMS, New Delhi - Top Medical University in India, 2021) and was ranked 1st among all medical colleges in India by the National Institutional Ranking Framework in 2021 (NIRF Rankings 2021, 2021). Therefore, it was found fit to be used to conduct this experimental study to understand the landscape of COVID-19 research in India.

2.2   COVID-19 Research

Topic modeling has been applied to COVID-19 research to reflect people's emotional responses to the pandemic and academic research hotspots. For example, Jang et al. (2021) analyzed COVID-19-related tweets using topic modeling and aspect-based sentiment analysis to investigate people's reactions and concerns about COVID-19 in North America and Canada. Similarly, Ordun et al. (2020) also traced the distinctiveness of topics, key terms and features, speed of information dissemination, and network behaviors for COVID-19 tweets by using Latent Dirichlet Allocation (LDA) to generate topics.

From the perspective of scientific article analysis, Haghani et al. (2020) used bibliometric analysis to describe COVID-19 research areas and their relative importance. Sonbhadra et al. (2020) extracted the activity and trends of coronavirus-related research articles using machine learning approaches from the paper content level. Dong et al. (2020) used LDA to track semantic relationships between topics and compare the topic distribution between COVID-19 and other CoV infections. Lui et al. (2021) used LDA to allocate COVID-19 research papers into 50 topics and found relevant studies that were the cornerstone of COVID-19 research. Bras et al. (2020) developed a novel automated theme-based visualization method for quick discovery of COVID-19 topics and research resources.

Therefore, it can be concluded that many studies have been conducted that used topic modeling to study the scientific literature on COVID-19 using Latent Dirichlet Allocation (LDA) algorithm on tweets, newspapers, and research articles but the present study retrieved data from three different databases for the top Indian Medical Institute and uses a different topic modeling algorithm (Structural Topic Model) than the reviewed literature.

2.3   Structural Topic Model (STM)

The Structural Topic Model (STM) was first developed by Roberts et al. in 2013. It has primarily been made for social scientists who wants to measure

latent variables using document-level covariates to identify the variation in either topic prevalence (i.e., the proportion of document devoted to a given topic) and/or topic content (i.e., the rate of word uses within a given topic). STM included corpus structure to reflect the interest in making inferences about observed covariates rather than predicting covariate values in the unseen text (Fig. 1).
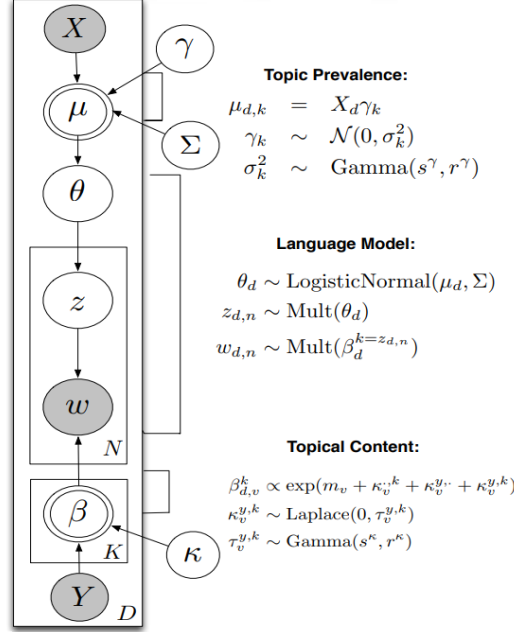

Topic Prevalence:

$$\mu_{d,k} = X_d \gamma_k$$
$$\gamma_k \sim \mathcal{N}(0, \sigma_k^2)$$
$$\sigma_k^2 \sim \mathrm{Gamma}(s^\gamma, r^\gamma)$$

Language Model:

$$\theta_d \sim \mathrm{LogisticNormal}(\mu_d, \Sigma)$$
$$z_{d,n} \sim \mathrm{Mult}(\theta_d)$$
$$w_{d,n} \sim \mathrm{Mult}(\beta_d^{k=z_{d,n}})$$

Topical Content:

$$\beta_{d,v}^k \propto \exp(m_v + \kappa_v^{\cdot,k} + \kappa_v^{y,\cdot} + \kappa_v^{y,k})$$
$$\kappa_v^{y,k} \sim \mathrm{Laplace}(0, \tau_v^{y,k})$$
$$\tau_v^{y,k} \sim \mathrm{Gamma}(s^\kappa, r^\kappa)$$

Fig. 1. Plate Diagram for Structural Topic Model (Roberts et al., 2013)

3   Methodology

The data was retrieved from Web of Science, Scopus, and PubMed databases for the query, "corona" OR "covid-19" OR "coronavirus" restricted to AIIMS, New Delhi. The search was performed in April 2021. Metadata for 853 studies were identified from all three databases which were then merged and removed for duplicates using the *bibliometrix* package (Aria and Cuccurullo, 2017) in R. Additionally, the document type viz., *comment, letter, correction, editorial, editorial material, editorial material-early access, erratum, published erratum, meeting abstract, note, and retracted*, were also removed. After the preliminary cleaning, a total of 388 publications were finalized for the period 2010-2021 from 199 journals for the study. The structural topic model was performed using stm package (Roberts et al., 2013) in R.

4   Results

4.1   Topic Modeling

"Topic modeling makes an excellent tool for discovery and helps to uncover evidence already present in the text. A topic can be defined as the main idea discussed in a text, i.e., the theme or subject of different granularities. In contrast, topic modeling acts as a text mining approach to understand, organize, process, extract, manage, and summarize knowledge. It is based on statistical and machine learning techniques to mine meaningful information from a vast corpus of unstructured data and is used to mine document content. There are no machine-readable annotations that can tell the topic modeling programs about the semantic meaning of the words in the text. Thus, it infers abstract topics based on similar patterns of word usage in each document" (Lamba and Madhusudhan, 2022).
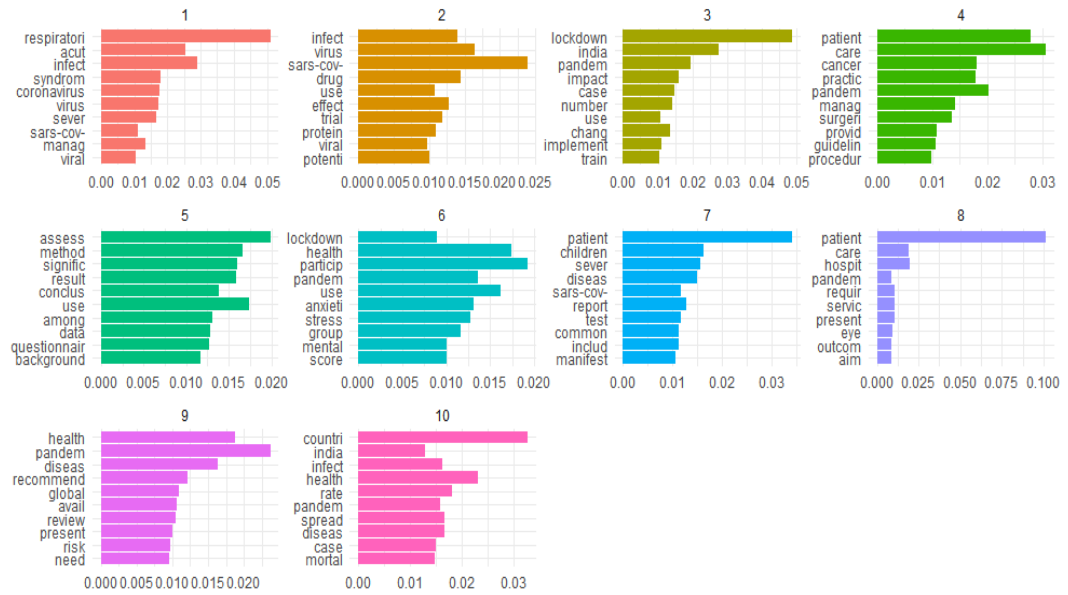

Fig. 2. Core Topics in COVID-19 Research Published by AIIMS, New Delhi

The Fig. 2 demonstrates the structural topic model results for COVID-19 research published by AIIMS, New Delhi. It presents the top 10 words for 10 core topics identified for the study. The topics were labeled after consulting the top words and the top representative articles together for the particular topic. The 10 sub-domains of COVID-19 research where the AIIMS, New Delhi researchers are publishing the most are (Fig. 2):

   (i) Respiratory Infection (Topic 1)
  (ii) Drugs Trial and Effectiveness (Topic 2)
 (iii) Lockdown in India (Topic 3)
 (iv) Patient Care, Practice, and Management (Topic 4)
  (v) Behavioral Studies (Topic 5)
 (vi) Mental Health (Topic 6)
(vii) Systematic Reviews and Metanalysis (Topic 7)
(viii) Guidelines to Perform Surgeries in COVID-19 Patients (Topic 8)
 (ix) Statements and Recommendations (Topic 9)
  (x) Mortality (Topic 10)

The Fig. 3 presents a network correlation plot to identify the interrelated topics. It can be observed from the figure that Topics 4 (patient care, practice, and management), 9 (statements and recommendations), and 10 (mortality) are related to each other. Topics 1 (respiratory infection), and 7 (systematic reviews and metanalysis) are related to each other. Topics 5 (behavioral studies), and 6 (mental health) are related to each other for the study. Topics 2 (drugs trial and effectiveness), 3 (lockdown in India), and 8 (guidelines to perform surgeries in covid-19 patients) are isolated and are not related to any of the other topics.
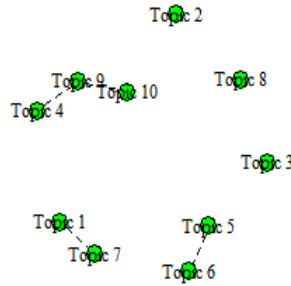

Fig. 3. Topic Correlation

"The output of topic modeling is not entirely human-readable, and one way to understand the results is through visualization" (Lamba and Madhusudhan, 2022). Therefore, we constructed a storyboard and dashboard that is human-readable and easy to comprehend as well as interpret. The Fig. 4 demonstrates the screenshot of the dashboard prepared to visualize and search the topic modeling results for the study. It is published at https://manika-lamba.github.io/stm/. The Fig. 5 presents the screenshot of the storyboard that summarizes the study results and is published at https://manika-lamba.shinyapps.io/covid_aiims/.

5   Conclusion

We are dealing with the massive explosion of COVID-19 research in the last 2.5 years, thus, being able to benefit from such an unprecedented amount of data
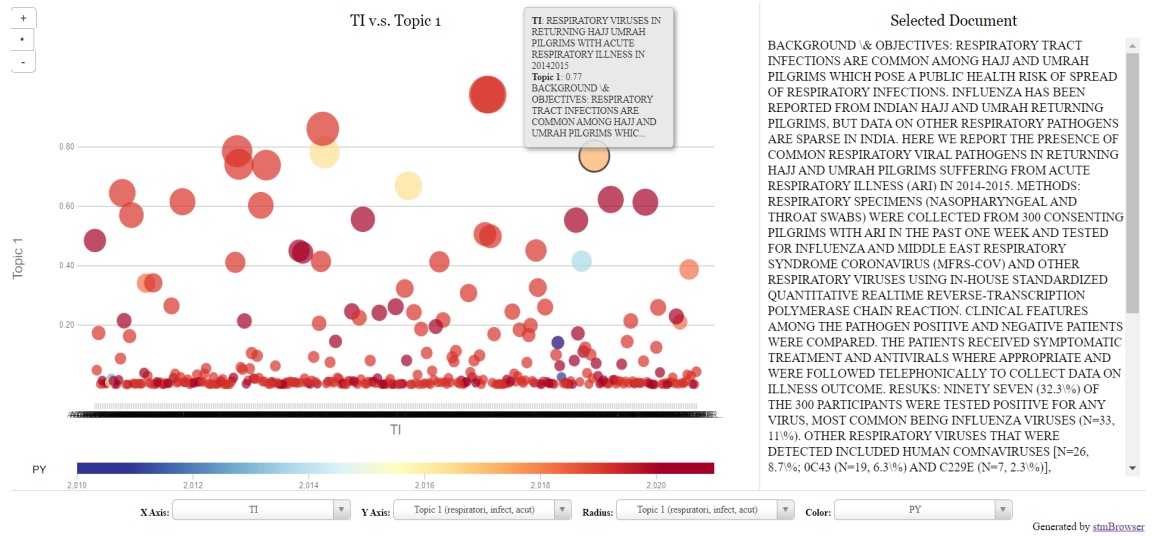
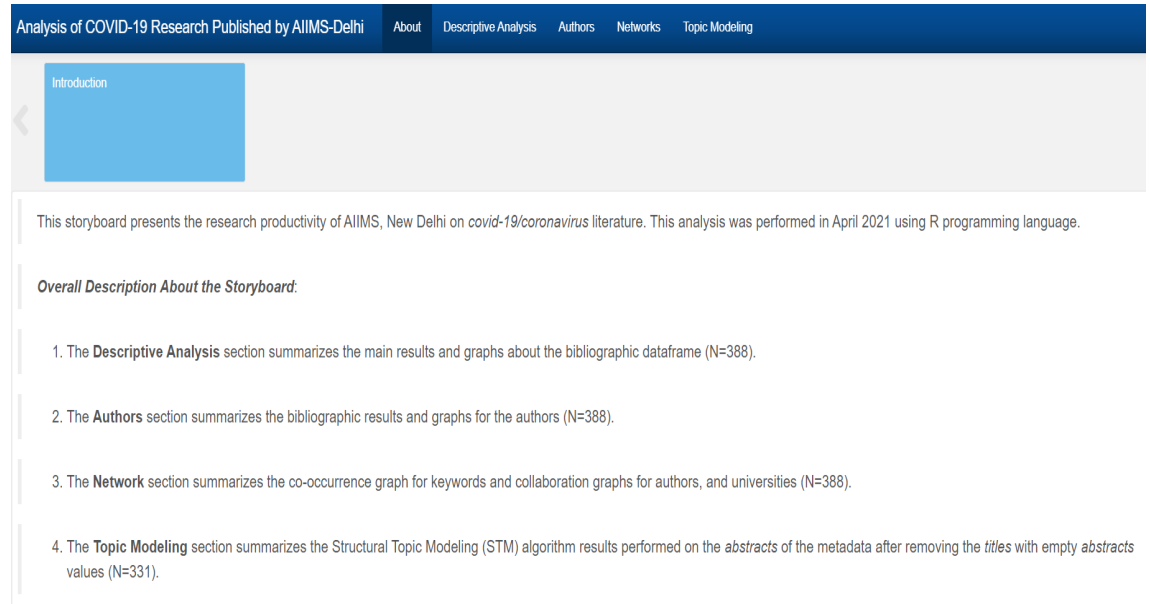**Fig. 4.** Screenshot of the Dashboard showing Topic Modeling Results

Fig. 5. Screenshot of the Storyboard

without being overwhelmed poses a significant obstacle for researchers, especially when facing the emergency of finding a cure, and more about the different variants of the coronavirus. Topic modeling and data visualization are critical for exploring and communicating information effectively and helping researchers to continue to progress. For this reason, the author applied topic modeling, and visualization methods to analyze 388 articles published on COVID-19 from AIIMS, New Delhi retrieved from three different databases. The study presents an overview of early studies of the COVID-19 crisis published from AIIMS, New Delhi at different scales including topic variation, and their inner interactions. It also identifies papers that are regarded as the cornerstones for different topics in the development of COVID-19 research published by AIIMS, New Delhi. The results unveil the focus of scientific research, thereby giving deep insights into how the top Indian medical society contributes to combating the COVID-19 pandemic. The methodology from this study can be applied to trace the literature of any field on any topic of interest. This study can be used to analyze large datasets of COVID-19 literature to investigate and visualize the ongoing advancements of early scientific research on COVID-19 from the perspective of Artificial Intelligence.

References

1. AIIMS, New Delhi - Top Medical University in India, https://www.whatincity.com/listing/aiims-new-delhi/. Last accessed 10 May 2022
2. Aria, M. and Cuccurullo, C.:bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics **11**(4), 959–975 (2017)
3. Bras, P.L., Gharavi, A., Robb, D.A., Vidal, A.F., Padilla, S. and Chantler, M.J.:Visualising COVID-19 Research, ArXiv:2005.06380 [Cs], http://arxiv.org/abs/2005.06380. Last accessed 11 May 2022
4. Dong, M., Cao, X., Liang, M., Li, L., Liu, G. and Liang, H.:Understand research hotspots surrounding COVID-19 and other coronavirus infections using topic modeling, medRxiv, https://doi.org/10.1101/2020.03.26.20044164. Last accessed 11 May 2022
5. Haghani, M., Bliemer, M.C.J., Goerlandt, F. and Li, J.:The scientific literature on Coronaviruses, COVID-19 and its associated safety-related research dimensions: A scientometric analysis and scoping review, Safety Science **129**, 104806 (2020)
6. Jang, H., Rempel, E., Roth, D., Carenini, G. and Janjua, N.Z.:Tracking COVID-19 Discourse on Twitter in North America: Infodemiology Study Using Topic Modeling and Aspect-Based Sentiment Analysis, Journal of Medical Internet Research **3**(2), e25431 (2021)
7. Lamba, M. and Madhusudhan, M.:Topic Modeling", Text Mining for Information Professionals: An Uncharted Territory, Springer International Publishing, Cham, 105—137 (2022)
8. Liu, J., Nie, H., Li, S., Chen, X., Cao, H., Ren, J., Lee, I., et al.:Tracing the Pace of COVID-19 Research: Topic Modeling and Evolution, Big Data Research **25**, 100236 (2021)

9. NIRF Rankings 2021: AIIMS Delhi declared best medical college, IIT Madras tops overall category, Firstpost, https://www.firstpost.com/india/nirf-rankings-2021-aiims-delhi-declared-best-medical-college-again-iit-madras-tops-overall-category-9950561.html. Last accessed 10 May 2022

10. Ordun, C., Purushotham, S. and Raff, E.:Exploratory Analysis of Covid-19 Tweets using Topic Modeling, UMAP, and DiGraphs, ArXiv:2005.03082 [Cs], http://arxiv.org/abs/2005.03082. Last accessed 11 May 2022

11. Roberts, M.E., Tingley, D., Stewart, B.M. and Airoldi, E.M.:The Structural Topic Model and Applied Social Science, In: NIPS 2013 Workshop on Topic Models: Computation, Application, and Evaluation, https://scholar.princeton.edu/files/bstewart/files/stmnips2013.pdf

12. Sonbhadra, S.K., Agarwal, S. and Nagabhushan, P.:Target specific mining of COVID-19 scholarly articles using one-class approach, Chaos, Solitons & Fractals **140**, 110155 (2020)