Patterns

Inferring global-scale temporal latent topics from news reports to predict public health interventions for COVID-19

Highlights

- Automated prediction of public health intervention from COVID-19 news reports

- Inferred 42 country-specific temporal topic trends to monitor interventions

- Learned interpretable topics that predict interventions from news reports

- Transfer learning to predict interventions for each country on weekly basis

Authors

Zhi Wen, Guido Powell, Imane Chafi, David L. Buckeridge, Yue Li

Correspondence

david.buckeridge@mcgill.ca (D.L.B.), yueli@cs.mcgill.ca (Y.L.)

In brief

We developed a machine learning model called EpiTopics toward automatic detection of the changes in the status of non-pharmacological interventions (NPI) for COVID-19 from news reports. EpiTopics learns country-dependent topics from large numbers of COVID-19 news reports that do not have NPI labels. Subsequently, EpiTopics learns accurate connections between these topics and changes in NPI status from a set of labeled news reports, which enables accurate detection of temporal NPI status for each country referred to in the news reports.

CellPress

Article

Inferring global-scale temporal latent topics from news reports to predict public health interventions for COVID-19

Zhi Wen,[1,3] Guido Powell,[2,3] Imane Chafi,[1] David L. Buckeridge,[2,*] and Yue Li[1,4,*]

[1]School of Computer Science, McGill University, Montreal, QC, Canada
[2]School of Population and Global Health, McGill University, Montreal, QC, Canada
[3]These authors contributed equally
[4]Lead contact
*Correspondence: david.buckeridge@mcgill.ca (D.L.B.), yueli@cs.mcgill.ca (Y.L.)
https://doi.org/10.1016/j.patter.2022.100435

THE BIGGER PICTURE   Accurate, scalable detection of the timing of changes to public health interventions for COVID-19 is an important step toward automating evaluation of the effectiveness of interventions. We show that it is possible to train an interpretable deep-learning model called EpiTopics on media news data to predict (1) the interventions mentioned in individual news articles and (2) the temporal change of intervention status at the country level. We addressed a main challenge of label scarcity among the news reports. Using EpiTopics, we modeled the latent semantics from 1.2 million unlabeled news reports on COVID-19 over 42 countries recorded from November 1, 2019 to July 31, 2020, identifying 25 interpretable topics under 4 COVID-related themes. Using the learned topic model, we inferred topic mixture membership for each labeled article, which allowed us to learn an accurate connection between the topics and the public health interventions at both the document level and country level.

1 2 3 4 5 **Development/Pre-production:** Data science output has been rolled out/validated across multiple domains/problems

---

SUMMARY

The COVID-19 pandemic has highlighted the importance of non-pharmacological interventions (NPIs) for controlling epidemics of emerging infectious diseases. Despite their importance, NPIs have been monitored mainly through the manual efforts of volunteers. This approach hinders measurement of the NPI effectiveness and development of evidence to guide their use to control the global pandemic. We present EpiTopics, a machine learning approach to support automation of NPI prediction and monitoring at both the document level and country level by mining the vast amount of unlabeled news reports on COVID-19. EpiTopics uses a 3-stage, transfer-learning algorithm to classify documents according to NPI categories, relying on topic modeling to support result interpretation. We identified 25 interpretable topics under 4 distinct and coherent COVID-related themes. Importantly, the use of these topics resulted in significant improvements over alternative automated methods in predicting the NPIs in labeled documents and in predicting country-level NPIs for 42 countries.

INTRODUCTION

It has long been understood that organized community efforts are required to control the spread of infectious diseases.[1] These efforts, called public health interventions, include social or non-pharmaceutical measures to limit the mobility and contacts of citizens and pharmaceutical interventions to prevent and limit the severity of infections. Non-pharmaceutical interventions (NPIs) are not easily evaluated through experimental studies,[2] so evidence regarding the effectiveness of these public health interventions is usually obtained through observational studies. This is also partly due to the lack of an efficient method to monitor NPIs, which is challenging because their use is not recorded consistently within or across countries.

In the early stages of an emerging infectious disease, such as COVID-19, NPIs tend to be particularly important due to the lack

A One of the 2000 WHO documents with NPI labels

Country: Canada, Time: March 25, 2020 (WHO)

Today the Honourable Patty Hajdu, Minister of Health, announced an Emergency Order under the Quarantine Act that requires any person entering Canada by air, sea, or land to self-isolate for 14 days whether or not they have symptoms of COVID-19.
......
Individuals displaying symptoms of COVID-19 after arriving in Canada may not use public transportation to travel to their place of isolation. They also may not isolate in a place where they will be in contact with vulnerable people such as seniors and individuals with underlying health conditions.

Manually curated NPI labels for the WHO document:
- Detecting and isolating cases
- International travel measures

B One of 1.2 M AYLIEN news reports with no NPI label

Country: Canada, Time: March 16, 2020 (AYLIEN)

An office building in Downtown Winnipeg is in the process of being cleaned, and its occupants are being asked to work from home, after owners say they were notified someone who's recently been in the building has tested positive for COVID-19 and someone else has a possible case of the virus.
......
"If there was an at-risk event that we thought, of course we'll be trying to contact those people and if we needed to make that public, we'd do it."

C Country-level NPI labels at each CDC week recorded by WHO

NPIs in Canada on the week of March 16, 2020
- Communications and engagement
- Detecting and isolating cases
- Domestic Travel
- Financial packages
- Gatherings, businesses and services
- International travel measures
- Offices, businesses, institutions and operations
- Other measures
- School measures

Figure 1. Examples of news reports and NPI labels
(A) One of the 2,000 WHO documents with 2 manually curated NPI labels.
(B) One of the 1.2 million AYLIEN news reports with no NPI label observed. Nonetheless, the sentences related to 2 NPIs are highlighted.
(C) Country-level NPI labels recorded at each CDC week. The observed NPIs in the same CDC week in Canada are highlighted, which are implicated in the AYLIEN news reports in (B).

of specific pharmaceutical interventions. At the outset of the COVID-19 pandemic, recognizing the absence of systems for recording NPIs, multiple groups initiated projects to track the use of interventions for COVID-19 around the world. These projects have relied on the manual efforts of volunteers to review digital documents accessible online with minimal coordination across projects.[3] This approach to monitoring interventions is not sustainable and is difficult to be implemented or extended to other infectious diseases.[4]

Given that changes in the status of NPIs are usually described in digital online media (e.g., government announcements and news media) (Figure 1), machine learning methods have the potential to support the monitoring of NPIs. However, the application of machine learning methods to this task is complicated by the need to generate interpretable results and also by the limitations in the amount and quality of labeled data.

In this paper, we present a machine learning framework that uses online media reports to monitor changes in the status of NPIs that are used to contain the COVID-19 spread in 42 countries. Our method makes use of three types of data: around 2,000 documents for which a group supported by WHO provided labels for 15 NPIs, 1.2 million unlabeled news reports on COVID-19 made publicly available from an AI-powered news mining company AYLIEN, and country-level NPI labels at each CDC week (i.e., the epidemiological weeks defined by the Centers for Disease Control and Prevention) for 42 countries (Figure 1). This type of event-based surveillance can inform situational awareness, support policy evaluation, and guide evidence-based decisions about the use of public health interventions to control the COVID-19 pandemic.

Our overarching goal in this work is to develop a topic modeling framework[5] to predict NPI labels at both the document level and country level to support the monitoring of public health interventions. To this end, we developed a 3-stage framework called EpiTopics (Figure 2). At the center of our approach is the use of probabilistic topic modeling, which facilitates the interpretation of our results and is well suited to the noisy nature of online media. To address the limitation of the small number of NPI-labeled documents, we use transfer learning to infer the topic mixture from the labeled data using a topic model trained on a much larger amount of unlabeled COVID-19 news data. By clas-

sifying online media news into NPI categories based on interpretable, model-inferred topic distributions, we aim to automate many of the labor-intensive aspects of NPI tracking and enable systematic monitoring of interventions across larger geographical and conceptual scopes than are possible with currently used manual methods.

RESULTS

EpiTopics model overview
EpiTopics is a machine learning framework for predicting NPIs used to control the COVID-19 pandemic from the news data. The intended application of the framework is to predict each week, based on unlabeled news documents about COVID-19, any change in the status of multiple NPIs at the country level. To this end, EpiTopics implements a 3-stage machine learning strategy (Figure 2), which is described in general terms below and in greater detail in the experimental procedures.

### Interpreting COVID-19 topics learned from the AYLIEN dataset
At the first stage, we sought to learn a set of unbiased latent topic distributions by mining a large corpus of news articles about COVID-19, but without the NPI labels. We used an open-access dataset of articles publicly released by an AI news company AYLIEN, which included 1.2 million news articles about COVID-19 from 42 countries recorded from November 1, 2019, to July 31, 2020 (data and code availability). All articles were related to COVID-19, but not necessarily related to particular NPI implementations. Although NPI labels were not available for these articles, in contrast to the smaller datasets with NPI labels (described next), this corpus was suitable for learning semantically diverse topics related to COVID-19 due to its size and coverage. To capture country-dependent topic dynamics, we developed an unsupervised model called MixMedia,[6] which was adapted from the dynamic embedding topic model (DETM)[7] (experimental procedures, discussion; Figure 2A).

We experimented with different numbers of topics and chose 25 topics based on a common metric called topic quality[7] (experimental procedures; Figure S1). We annotated the 25 topics learned from the AYLIEN data at stage 1 based on the

Stage 1. Learning COVID-19 topic and word embeddings from 1.2 million AYLIEN dataset over 42 countries
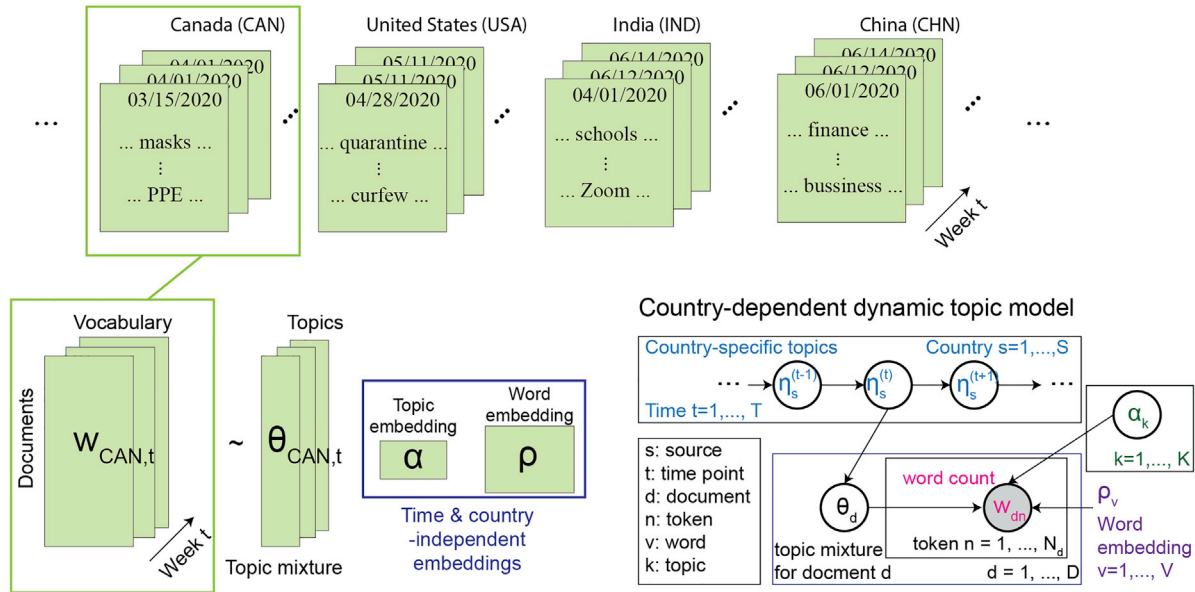
Stage 2. Supervised transfer-learning for document-level NPI prediction
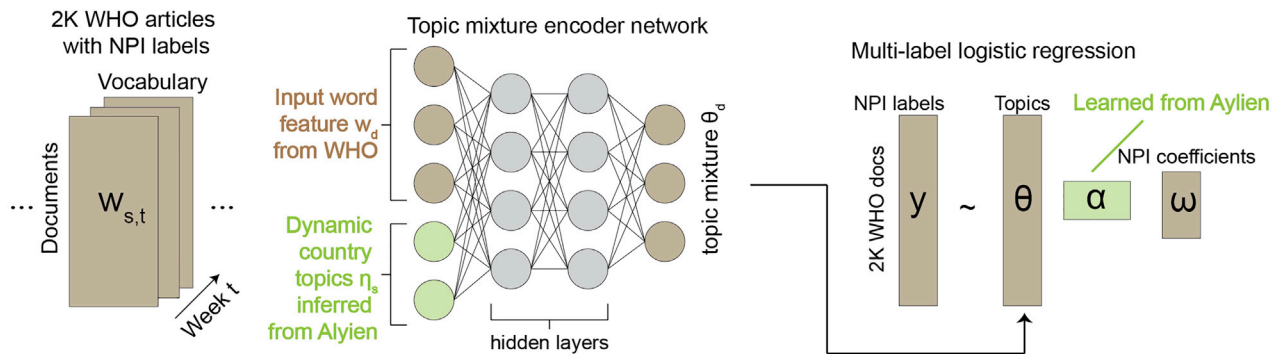
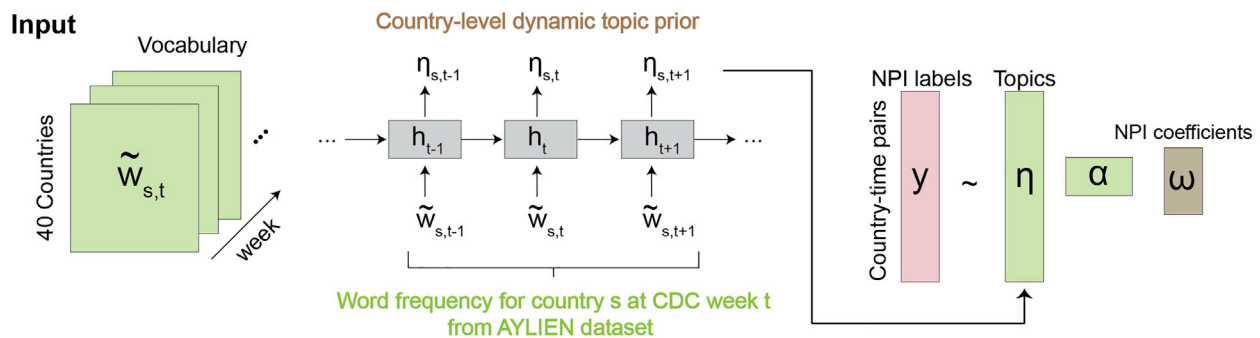## Stage 3. Transfer learning from document-level to country-level NPI retrospective prediction

Figure 2. EpiTopics model overview

(A) Unsupervised learning of COVID-19 topic and word embeddings from 1.2 million AYLIEN dataset over 42 countries. For each country, we extracted a set of news articles related to COVID-19 observed from November 1, 2019, to July 31, 2020, from AYLIEN. For each document (e.g., $w_{CAN,t,d}$ for document d published at time t from Canada), we inferred their topic mixture (θ_d). Meanwhile, we learned the global topic embedding (α) and word embedding (ρ) from the corpus.

(legend continued on next page)

most probable words under each topic (Figure 3) and then grouped these topics into four themes: (1) central health-related issues, (2) broader social impacts, (3) specific locations, and (4) indirectly related issues.

Under theme I, several topics addressed *central health-related issues* of the pandemic. For instance, the most frequent words in topic "M3 – COVID burden" included "deaths" and "confirmed cases"; topic "M12 – healthcare/hospitals" included "health," "staff," "doctors," "beds"; topic "M22 – testing/tracing" included "tested," "positive," "symptoms"; topic "M25 – vaccine/research" included "vaccine," "trial," "drug"; topic "M17 – masks" included personal protective measures, such as "wearing," "masks," (washing) "hands," and "social distancing"; and topic "M9 – global pandemic" included "health," "pandemic," and "world."

Under theme 2, topics were associated with *broader social impacts*. For example, topic "M20 – schools" reflected changes in education including words, such as "school," "online," "classes," "exams"; topic "M14 – international travel" focused on changes in international travel, with top words, such as "flight" and "passenger"; topic "M10 – government restrictions" included "lockdowns" and "work" (from) "home"; topic "M23 – time/home" was related to spending time at home and family; topic "M8 – food & delivery" discussed "restaurant," "delivery," "grocery," and "stores"; topic "M21 – sports leagues" was associated with sports teams and events; and topic "M6 – entertainment" was related to the impacts on entertainments, such as "film" and "music." There were also topics related to financial impacts, such as topic "M1 – business/work" on "businesses" and "companies"; topic "M5 – economy/markets" on "stock markets" and topic "M7 – finance/economy" on impacts on "unemployment," "income," and "government financial packages."

Under theme 3, some topics focus on impacts of the virus on *specific countries*. Topic "M2 – COVID China" focuses on the pandemic in China and particularly in Wuhan city; topic "M4 – COVID USA/local" focused on the United States and its specific states and cities; and topic "M13 – COVID India" focused on the pandemic in India.

Under theme 4, a few topics were not directly related to COVID-19, although they were sometimes discussed in the context of COVID-19 as they reflected events occurring during the pandemic. For example, topic "M18 – outdoors" discusses "water," "space," and "parks," which are activities with reduced risk of COVID-19; topic "M11 – government/state" was associated with states or governments; topic "M15 – internet/social media" focused on social media and the internet, which are crucial for distributing COVID-19 information; topic "M16 – legal" involves "court," "law," and "order"; topic "M19 – police brutality/protests" was associated with police brutality, which sparked large-scale protests during the pandemic; and topic "M24 – US election" was associated with US presidential election, for which the pandemic was a central issue.

Taken together, these observations indicate that the 25 topics inferred by our EpiTopics model from the AYLIEN news dataset represent diverse aspects of the COVID-19 pandemic. These inferred topics provide a rich foundation against which we can characterize online media reports about COVID-19. Notably, although subjective by nature, the above manual topic labeling is an important aspect of topic analysis, as the labels aid in the interpretation of the subsequent country-level analysis and NPI prediction tasks.

Country-topic dynamics

Continuing to explore the results generated by EpiTopics at stage 1, we examined temporal patterns at the country level in the posterior distributions of topic popularity (η) inferred by our EpiTopics model. These patterns reflect changes over time in the COVID-19-related media news as the pandemic unfolded in each country (Figure 4). As expected, media news in January and February from all of the 42 countries are predominantly about topic "M2 – China COVID." In March, the topic popularity began to diverge for different countries. Interestingly, the change in topic popularity for many countries meaningfully reflected the progression of the pandemic within those countries. For instance, topic "M13 – India COVID" started to rise at the end of March and stayed high for India media news. Asian countries, South Korea, China, and Japan tended to focus more on topic "M12 – China COVID" even after February, whereas the same topic dropped more sharply after February for countries in other global regions. Spain was the only country that exhibited high probability for topic "M21 – sports leagues," perhaps due to the strong impacts of COVID on these activities in Spain or possibly reflecting sampling bias from the data, which was focused on English-language media. In contrast to other countries, we observed more dramatic changes in topic dynamics for the US with the focus shifting from topic "M2 – China COVID" in February to topic M4 for COVID in local regions of the US, then to multiple topics around June, including topic M9 for the worldwide epidemic, topic "M24 – US election," and the more general topic M23 for staying home. We caution that this theme may reflect the greater amount of articles collected for the US compared with other countries. More generally, given that we considered only articles published in English, our results may be biased toward media content from English-speaking countries.

Predicting document-level interventions

At the second stage, we trained a linear classifier to predict the 15 NPI labels mentioned in the 2,049 news articles from the WHO dataset (data and code availability) using the topic embeddings that were generated from the network encoder of the Mix-Media model trained at stage 1 (EpiTopics-stage 2: Predicting document-level interventions; Figure 2B). We compared the prediction performance of EpiTopics and 2 baseline methods using the macro area under the precision-recall curve (AUPRC) and the

---

(B) Supervised transfer learning for document-level NPI prediction. We used the trained topic mixture encoder from stage 1 to infer the topic mixture of the 2,000 WHO news reports. The resulting topic mixture were then used as the input features to predict NPI labels in a logistic regression model.

(C) Supervised transfer learning from document-level to country-level NPI prediction. Here, we used country-specific topic trajectories η inferred at stage 1 as the input features to predict country-level NPIs at each time point in a pre-trained logistic regression model, where the linear coefficients were already fit at stage 2 for the document-level NPI prediction task.

Figure 3. Learned topics and the top words under each topic
The sizes of the words are proportional to their topic probabilities. The background colors indicate the themes we gave to the topics.

weighted AUPRC (Table 1). The baseline methods predicted NPIs from the bag-of-words (BOW) feature vector of a document (i.e., the raw word count) using either a linear classifier or a two-layered feedforward network. Overall, EpiTopics outperformed the 2 baseline methods by a large margin, highlighting the benefit of incorporating topic mixtures (Figure 5A). While the prediction accuracy greatly varied across individual NPIs, we achieved the best prediction performance for "financial packages," "school measures," and "international travel measures" (AUPRC > 0.6). In addition, the use of topic modeling makes
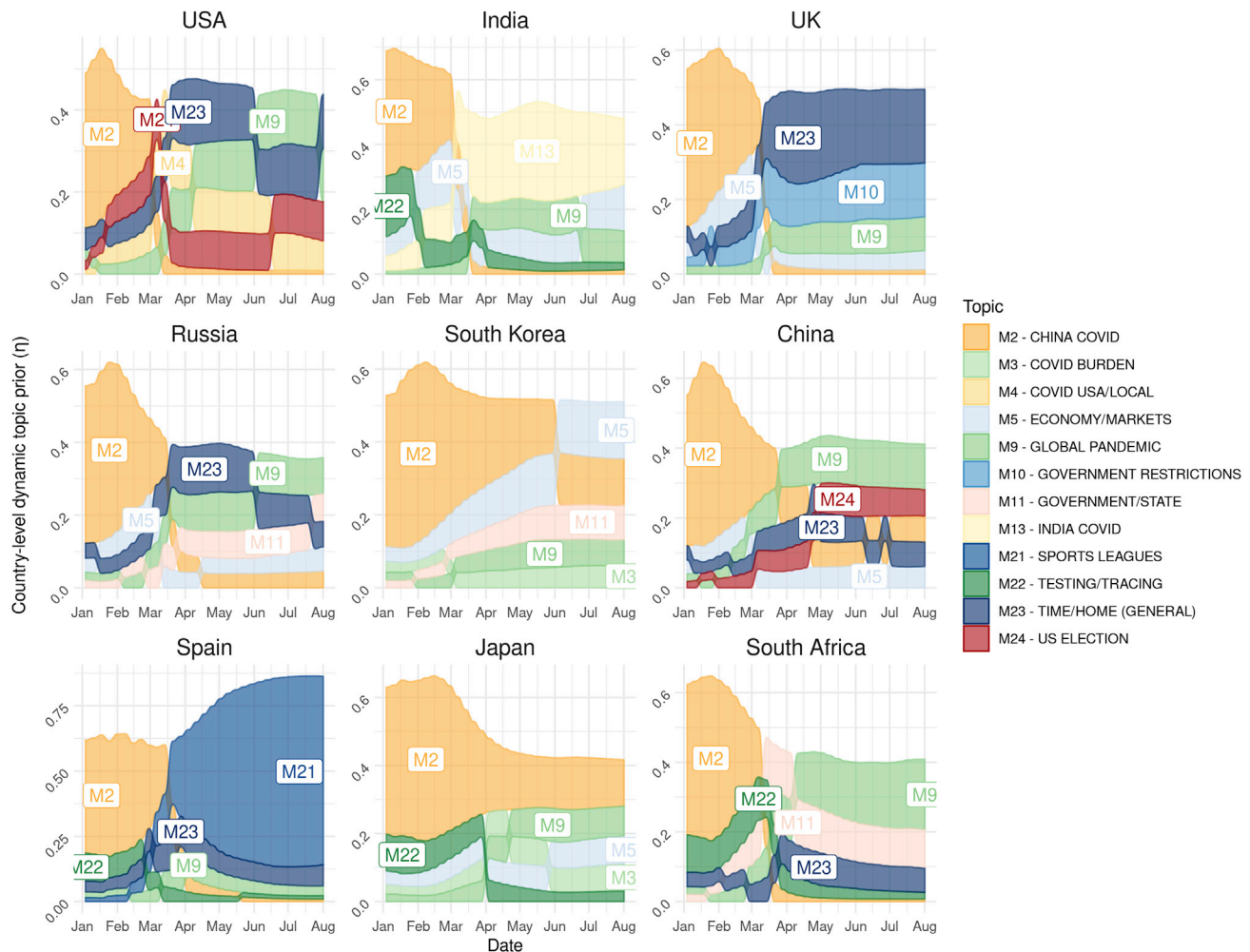
Figure 4. Temporal topics progression in example countries from January 2020 to July 2020
Different topics are represented in different colors. The height of each color block is proportional to the topic probability, in that country at that specific time. The topics were ranked in descending order vertically. To avoid cluttering the plot, only the top 5 topics are displayed for each country.

the results from our method highly interpretable via the learned classifier weights, which quantify the associations between the 25 topics and the 15 NPIs (Figure 5B). For example, we observed that topic "M20 – schools" was strongly associated with the NPIs school measures, topic "M5 – economy/markets" was associated with financial packages, and topic "M21 – sports league" was associated with "gatherings, businesses and services." Notably, although some topics, such as M21 alone did not have conceptual association with any NPI directly, their change correlated with changes in NPIs. They therefore could contribute to the prediction of NPIs. This result highlights the benefits of pre-training EpiTopics on a large unlabeled corpus to capture topics correlated with NPIs.

Predicting country-level interventions

At the third stage, we performed transfer learning to predict the 15 NPI labels at the country level for 42 countries by transferring model parameters learned from stages 1 and 2 to stage 3 (Figure S3). The goal of the final stage is to predict whether there is any change in the status of a specific NPI at the country level

from online media reports (Figure 2C). Here, a change in NPI status includes implementation, modification, and phase-out. We do not distinguish the types of change in our current model due to the limitations of the labeled data.

For input features, we used the -projected country-dependent topic posterior ($\eta\alpha$) inferred from the AYLIEN dataset at stage 1. The weights (ω) learned at stage 2 for the document-level NPI prediction were reused as the NPI-topic classifier weights. These weights were first used directly to predict NPIs at the country level without further training, an approach which we call zero-shot transfer. To improve upon this model, we also fine-tuned the topic-NPI coefficients ω to tailor the model toward country-level NPI prediction. We call this approach the fine-tuned model. Finally, as a baseline model, we also trained a classifier by randomly initializing and fitting the topic-NPI coefficients ω while fixing the same α-projected country-dependent topic posterior inferred from stage 1. We call this approach the from-scratch model. Details are described in experimental procedures.

Overall, the "zero-shot-transfer" model performed significantly better than "random" model, and fine-tuning the

Table 1. AUPRC scores for document-level NPI prediction

	Weighted AUPRC	Macro AUPRC
BOW + linear	0.165 (0.001)	0.108 (0.001)
BOW + feedforward	0.149 (0.022)	0.099 (0.007)
Document topic mixtures + linear	0.408 (0.001)	0.316 (0.001)

The AUPRC scores are computed on individual NPIs, and then averaged without weighting (macro AUPRC) or weighted by NPIs' prevalence (weighted AUPRC). Both BOW + linear and BOW + feedforward use the normalized word vector (i.e., bag-of-words [BOW]) for each document to predict NPI label. All methods are each repeated 100 times with different random seeds. Values in the brackets are standard deviations over the 100 experiments.

transferred topic-NPI linear weights performed significantly better than training from scratch (Figure 6C; Table 2). The results were mostly consistent for individual NPI prediction (Figures 6A and S2) and across countries (Figure S5). Furthermore, the predicted probabilities of NPIs were explained by the dominant country-level topic priors and their associations with each NPI (Figure 7). For instance, the high probabilities for financial packages in the Philippines starting from March (Figure 7), matched well with the true labels, and the high probabilities of this NPI were explained by the increasing probabilities of topics "M5 – economy/markets" and "M7 – finance/economy," and their strong association with NPI financial packages learned in the classifier (Figures 5B and 7).

We observed that the prevalence of NPIs was highly correlated with individual NPIs' AUPRC scores, especially for the random baseline (Figures 6A and S4B). This result was expected as the precision of completely random predictions is approximately equal to the proportion of the positive labels.[8] For non-random predictions, on a small dataset such as ours, it is also reasonable that the prevalence plays an important role. In addition to the prevalence of an NPI, the nature of an NPI and the way it is defined and grouped can impact performance. In some groups, the NPIs can be diverse (e.g., "other measures"), making these groups less conceptually coherent than other groups and harder to predict. Also, it is possible that some NPI status changes were not recorded by the trackers and therefore were missing in our dataset.

We also explored the feasibility of predicting country-level NPIs from document-level topic mixtures θ_d inferred from the AYLIEN dataset (EpiTopics-stage 3: Predicting country-level interventions). Compared with predicting from country-level topic priors (Table 3), we observed minor differences in performance of zero-shot transfer and fine-tuning. In addition, we trained another linear model from scratch using the document-level topic mixture to predict country-level labels associated with each document. While this approach has the advantage of using more documents to train a classifier, we observed similar performance compared with the fine-tuning approach. Notably, directly training on country-level topic mixture as input was much more efficient than training on the document-level topic mixture as there were one million documents and only a few hundred country-week pairs. Therefore, compared with the document-level alternative, the results highlight the advantage of directly operating on the inferred country-topic mixture in terms of both the interpretability and the computational efficiency.

DISCUSSION

In this paper, we present EpiTopics, a framework for surveillance of public health interventions using online news reports. This framework makes use of transfer learning to address the limited amount of labeled data and topic modeling to guide the model interpretation. Specifically, we used transfer learning in 2 places. First, the topic mixture encoder network trained on the 1.2 million unlabeled news reports related to COVID-19 at stage 1 was transferred to inferring the topic mixture of the 2,000 NPI-labeled WHO documents at stage 2 (Figure 2B). Second, we used temporal country-specific topics generated by the recurrent neural network encoder from stage 1 as input features and the already trained document-level NPI linear classifier to predict the country-level NPI at stage 3 (Figure 2C). In terms of topic modeling, our approach learned 25 interpretable topics, over 4 distinct and coherent COVID-related themes, whose country-specific dynamics appear to reflect evolving events in 42 countries. These topics contributed to significant improvements over alternative automated methods in predicting NPIs from documents. While country-level predictions of some interventions were less accurate, the topic-NPI associations drawn from the previous step improved performance over the baseline models.

In the broader context, our work contributes to both public health and machine learning. From a public health perspective, the EpiTopics framework demonstrates the feasibility of using machine learning methods to meaningfully support automation of the surveillance of public health interventions. In the context of the COVID-19 pandemic, our methods can be used to enhance and facilitate manual efforts by research teams tracking interventions. While automated methods for screening have been used by existing COVID-19 trackers,[3,9] their limited accuracy and interpretability suggests that a hybrid or semi-automated approach is needed. With interpretable country-level topic dynamics, our model offers high-level insights along with classification guideposts for analysts attempting to monitor tens of thousands of NPI events around the world. More generally, expanding our approach to new corpora could further the development of rigorous methods for more general surveillance of public health interventions, which are not usually subject to routine monitoring. Therefore, EpiTopics complements the current, mainly manual approaches to monitoring NPIs and has the potential to be expanded to the surveillance of public health interventions in general.

From a machine learning perspective, in contrast to the existing approaches that require large amounts of labeled data, our approach exploits the rich information from the vast amount of unlabeled news reports. We accomplish this by adapting the DETM framework.[7] We made two important modifications to the original DETM so that it would be more suitable for predicting NPI changes. First, we inferred temporal, country-dependent topic probabilities to represent the evolving pandemic situation within each country. Second, in contrast to the original DTEM, we inferred a set of static topic embedding to improve interpretation. This allowed us to analyze each topic separately without keeping track of the evolution of each topic over time within small time intervals (i.e., CDC weeks). Taken together, these modifications to the DETM allowed us to analyze the country-dependent topic trajectories over time by leveraging a large set of highly interpretable global topic distributions.
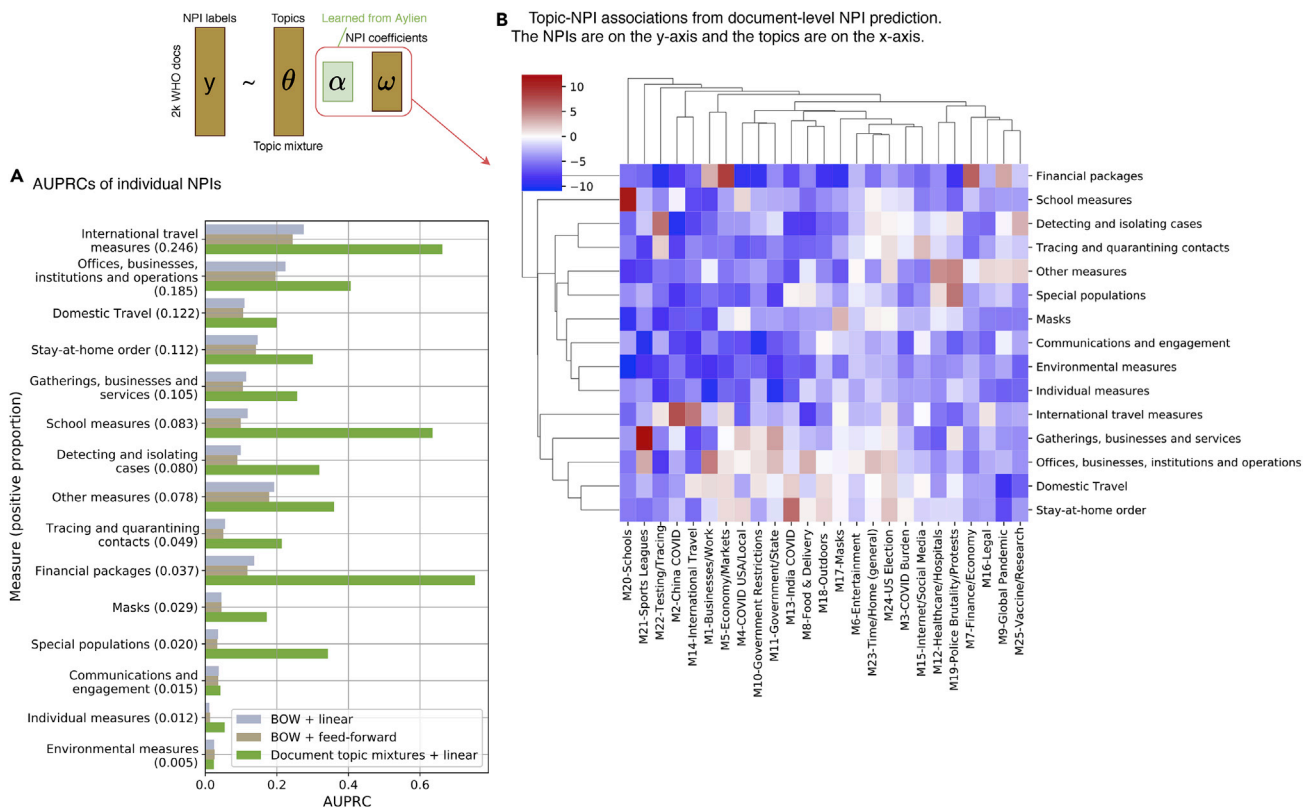
Figure 5. Document-level NPI prediction results
(A) AUPRC scores on individual NPI predictions. We compared 3 methods: BOW + linear used bag-of-word (BOW) features (i.e., the word count vectors) to predict NPIs with a linear classifier, BOW + feedforward used a feedforward neural network to predict NPIs using the BOW features as inputs, and document topic mixture + linear: EpiTopics, which uses the inferred topic mixture from the trained encoder at stage 1 to predict NPIs by fitting a linear classification model. (B) Topic-NPI associations learned by the linear classifier. The heatmap displays the linear coefficients with red, white, and blue indicating positive association, no association, and negative association, respectively.

In conclusion, our current work lays the methodological foundation for providing automated support for global-scale surveillance of public health interventions. Our work will greatly facilitate the manual NPI labeling process by automatically prioritizing relevant documents from a large-scale corpus. We envision that our work will also inspire further research to transform the way that public health interventions are monitored with advanced machine learning approaches.

Limitations of the study

Our work can be extended in several future directions. First, our study is focused on the time frame from the beginning of COVID-19 to July 2020. Application of the EpiTopics framework to data covering a wider time frame is possible, for example, through the fall of 2021 when many countries experienced a fourth wave. Since the beginning of the second wave, many interventions were implemented, and vaccines were approved and administered across the globe, which would likely affect media coverage and NPI prediction.

Second, this study was conducted using 2 datasets consisting of COVID-19-related news reports, each of which has their own limitations. The AYLIEN dataset lacks clear documentation on its sampling process and only includes records from English-language sources. These records may not be representative of media

discussion of interventions across all countries and in some cases English-language reporting of one country may not be representative of its media as a whole. While the WHO data have a very broad geographical coverage, issues with the validity of URLs in this dataset limited the records available for labeled data. Admittedly, these limitations of data sources restrict the generalizability of EpiTopics, in that countries not included in the training data are considered out-of-distribution and the EpiTopics model is unable to predict and monitor NPIs for these countries. We intend to improve the generalizability of EpiTopics by exploring other sources of data such as CoronaNet[3] and data from Global Public Health Intelligence Network,[10] which may cover a broader range of countries, languages, NPIs, tasks, and documents.[9,11]

Third, the documents are represented as a BOW for topic modeling. This representation omits rare words and stop words that are infrequent and too frequent, respectively. The order of words in sentences is also ignored. Therefore, we may lose some semantic meaning of the documents, especially when a sentence contains negation (e.g., deaths are not high). To this end, we intend to explore more complex neural language models such as ELECTRA[12] and BERT[13] at the expense of computations and sophisticated techniques, such as attention mechanisms for model interpretability.[14] Also, these neural language models operate at the sentence level, whereas our model operates at the

A AUPRCs of individual NPIs. Values in brackets are proportion of positive labels.
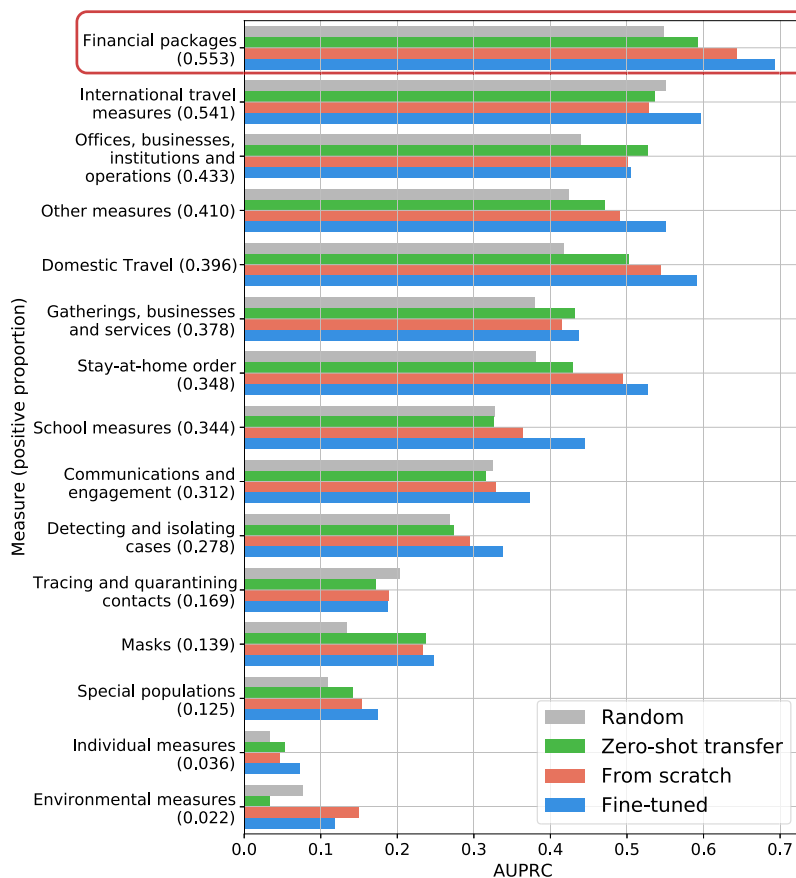
B Precision-recall curve of "Financial packages"
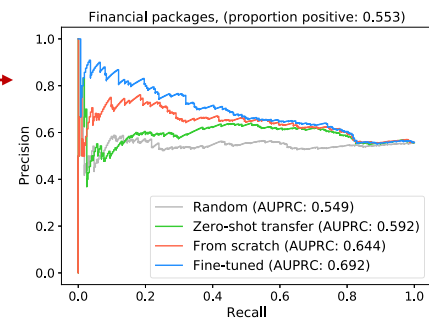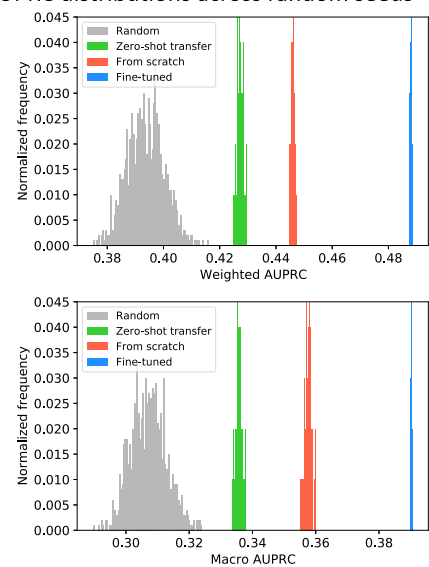

C AUPRC distributions across random seeds

Figure 6. Country-level NPI predictions

(A) AUPRC scores on individual NPI predictions at the country level. "random," "zero-shot transfer," "from-scratch," and "fine-tuned" are methods that predict NPIs using random features, pre-trained linear coefficients from stage 2 at the document-level NPI predictions, training the linear coefficients from random initialization, and fine-tuning the pre-trained linear coefficients from stage 2, respectively. The numbers in the brackets indicate the fraction of the positive labels for that NPI, which are positively correlated with the corresponding AUPRC.

(B) Precision-recall curve of different methods on predicting financial packages.

(C) Distributions of the weighted AUPRC scores and macro AUPRC scores over the 15 NPI predictions for the 4 methods across 100 repeated experiments with random initializations.

document and country levels. Future work can explore ways of exploiting manually labeled sentences[9] and investigate their performance on the task of NPI prediction.

Fourth, we can predict future country-level NPIs at the next time point based on the previous time points using the country-level topic trend and previous NPI states. This can be done in an auto-regressive framework. Furthermore, incorporating active case counts (i.e., the number of newly infected people) may also potentially improve NPI prediction. The challenge here is the scarce observations of positive NPI changes at the country level over time. Having more granular sub-region-level NPI labels (e.g., cities instead of countries) spanning over longer chronology of the pandemic may help training these more sophisticated models.

Finally, geolocation extraction is a challenging task.[15] We expect that some proportion of countries were misclassified in the construction of the AYLIEN media dataset. While the large

size of the dataset prohibits exhaustive validation, a non-exhaustive manual inspection of a subset of documents confirmed that a significant majority of country assignments were correct. To reduce the impact of this misclassification on our results, we limited our analyses to the country level rather than to the sub-region level, where misclassification is more likely. We intend to expand geolocation extraction in a future study to develop regional predictions.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Materials availability

This study did not generate physical materials.

Table 2. AUPRC scores for country-level NPI prediction

	Weighted AUPRC	Macro AUPRC
Random eta, alpha, and omega	0.394 (0.007)	0.307 (0.006)
Random eta and omega	0.394 (0.007)	0.307 (0.006)
Random omega	0.398 (0.014)	0.312 (0.011)
Zero-shot transfer	0.427 (0.001)	0.336 (0.001)
Trained from scratch	0.446 (0.001)	0.358 (0.001)
Fine-tuned	0.488 (0.000)	0.390 (0.000)

Random baselines are each repeated 1,000 times with different random seeds, and the rest are each repeated 100 times with different random seeds. Values in the brackets are standard deviations over the repeated experiments.

Data and code availability

WHO. Supported by the World Health Organization and led by a team at the London School of Hygiene and Tropical Medicine, the WHO Public Health and Social Measures (WHO-PHSM) dataset (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/phsm) represents merged and harmonized data on COVID-19 public health and social measures combining databases from seven international trackers. These individual trackers used mostly similar methods of manually collecting media articles and official government reports on COVID-19 interventions. The database has de-duplicated, verified, and standardized data into a unified taxonomy, geographic system, and data format. After validating entries across multiple trackers, the WHO team resolved conflicting information, such as different dates or disagreements in enacted measures, and carried forward only information from the best ranked sources (e.g., government documents over media reports and social media). We selected English-language records annotated with URL links to static content unchanged since its release. These pages were scraped using the Beautiful Soup and Newspaper3k Python libraries. Of over 30,000 records available at the time of writing, 8,432 records had valid URLs for scraping, of which 2,049 represented unique documents (articles or reports) that aligned with the period and countries in the larger AYLIEN dataset.

To balance the need for fine-grained NPI classes and the need for sufficient data points for model development, the WHO taxonomy covering 44 different categories of NPIs has been regrouped into 15 interventions (Table S1). The selection of categories to group together was determined by their conceptual similarity as well as their frequency. The prevalence of the 15 NPIs at the document and country levels are shown in Figure S4. We split the data into a training set and a testing set with an 80%–20% ratio and tuned the hyperparameters on the training set.

AYLIEN. While the WHO dataset contains expert-curated documents with NPI labels, the documents available were too few and not representative of general media content on which a surveillance system may operate. Therefore, in addition to the WHO dataset, we also use another dataset created by a News Intelligence Platform called AYLIEN (https://aylien.com/resources/datasets/coronavirus-dataset). Although this dataset has no NPI label, it has a lot more COVID-19-related news reports. In particular, the AYLIEN dataset has 1.2 million news articles related to COVID-19, spanning from November 2019 to July 2020. The data were compiled and generated using a proprietary NLP platform established by the AYLIEN company. The data were provided in the form of a json file that can be directly imported into the Python environment as a Panda DataFrame. We split 50% of data points (country-time pairs) into a training set and the remaining half into a testing set. The data processing is discussed next.

EpiTopics software. The EpiTopics code and processing scripts are available at GitHub: https://github.com/li-lab-mcgill/covid-npi. Additional supplemental items are available from Mendeley Data at https://doi.org/10.17632/nwwsvwrj93.1.

Data processing

We adopted the same data processing pipeline for both datasets. Specifically, we removed white spaces, special characters, non-English words, single-letter words, infrequent words (i.e., words appearing in fewer than ten documents). We also removed the same stop words as in Dieng et al.[16] as well as common non-English stop words "la," "se," "el," "na," "en," "de." For WHO documents, we obtained country information from the "SOURCE" field associated with each WHO document. For AYLIEN documents, we took the "country" sub-field under the "source" field for the country origin of the document. In addition, the content of the news reports are in the "body" field, and we use the "published at" for the publication date of the reports. After removing documents without empty source fields, we have 1.2 million documents in the AYLIEN dataset. To minimize the effect of noise in time stamps and also to reduce sparsity of the data, we reduced the temporal resolution to weeks by grouping the documents and the NPIs within the same CDC weeks (epidemiological weeks defined by the Centers for Disease Control and Prevention; https://wwwn.cdc.gov/nndss/downloads.html).

In total, there are 1,176,916 documents in the AYLIEN dataset, over 42 countries and spanning 38 CDC weeks from November 2019 to July 2020. After applying the same pre-processing procedure to WHO and discarding the documents whose country or source is not seen in AYLIEN, we have 2,049 unique documents in the WHO dataset with NPI labels. There are 109,377 and 8,012 unique words in AYLIEN and WHO, respectively. Among them, 366 unique words in WHO are not present in AYLIEN. Manual inspection confirms these 366 words are non-lexical or non-English and were discarded.

EpiTopics-stage 1: Unsupervised dynamic embedded topic

The first stage of EpiTopics (Figure 2A) is built upon our previous model called MixMedia,[6] which was adapted from the embedded topic model (ETM)[16] and the DETM.[7] The data generative process of EpiTopics is as follows:

1. Draw a topic proportion θ_d for a document d from logistic normal $\theta_d \sim \mathcal{LN}(\eta_{s_d}^{(t_d)}, \delta^2 \mathbf{I})$:

$$\delta_d \sim \mathcal{N}(0, I); \theta_d = \text{softmax}(\delta_d) = \frac{\exp(\delta_d)}{\sum_k \exp(\delta_{kd})},$$

where s_d and t_d are indices for the country of the document d and the time at which document d was published, respectively;

2. For each token n in the document, draw word:

$$w_{dn} \sim \text{Cat}(\theta_d \beta) = \prod_v (\theta_d \beta_v)^{[w_{dn} = v]},$$

where β is a normalized $K \times V$ probability matrix with each row endowing a topic distribution over the vocabulary of size V:

$$\beta_k = \frac{\exp(\beta_k^*)}{\sum_v \exp(\beta_{kv}^*)},$$

we further factorize the unnormalized topic distribution β^* into a $K \times L$ topic embedding matrix α and a $L \times V$ word embedding matrix ρ:

$$\beta^* = \alpha\rho.$$

The above K-dimensional time-varying topic prior $\eta_s^{(t)}$ is a dynamic Gaussian variable, which depends on the topic prior at the previous time point of the same source:

$$p(\eta_s^{(0)}) = \mathcal{N}(0, \mathbf{I}_K) \quad \text{if } t = 0$$
$$p(\eta_s^{(t)} | \eta_s^{(t-1)}) = \mathcal{N}(\eta_s^{(t-1)}, \zeta^2 \mathbf{I}_K) \quad \text{otherwise.}$$

For the ease of interpretation, we assume that the L-dimensional topic embedding α_k for each topic k is time-invariant and follows a standard Gaussian distribution:[6]

$$p(\alpha_k) = \mathcal{N}(0, \mathbf{I}_L).$$
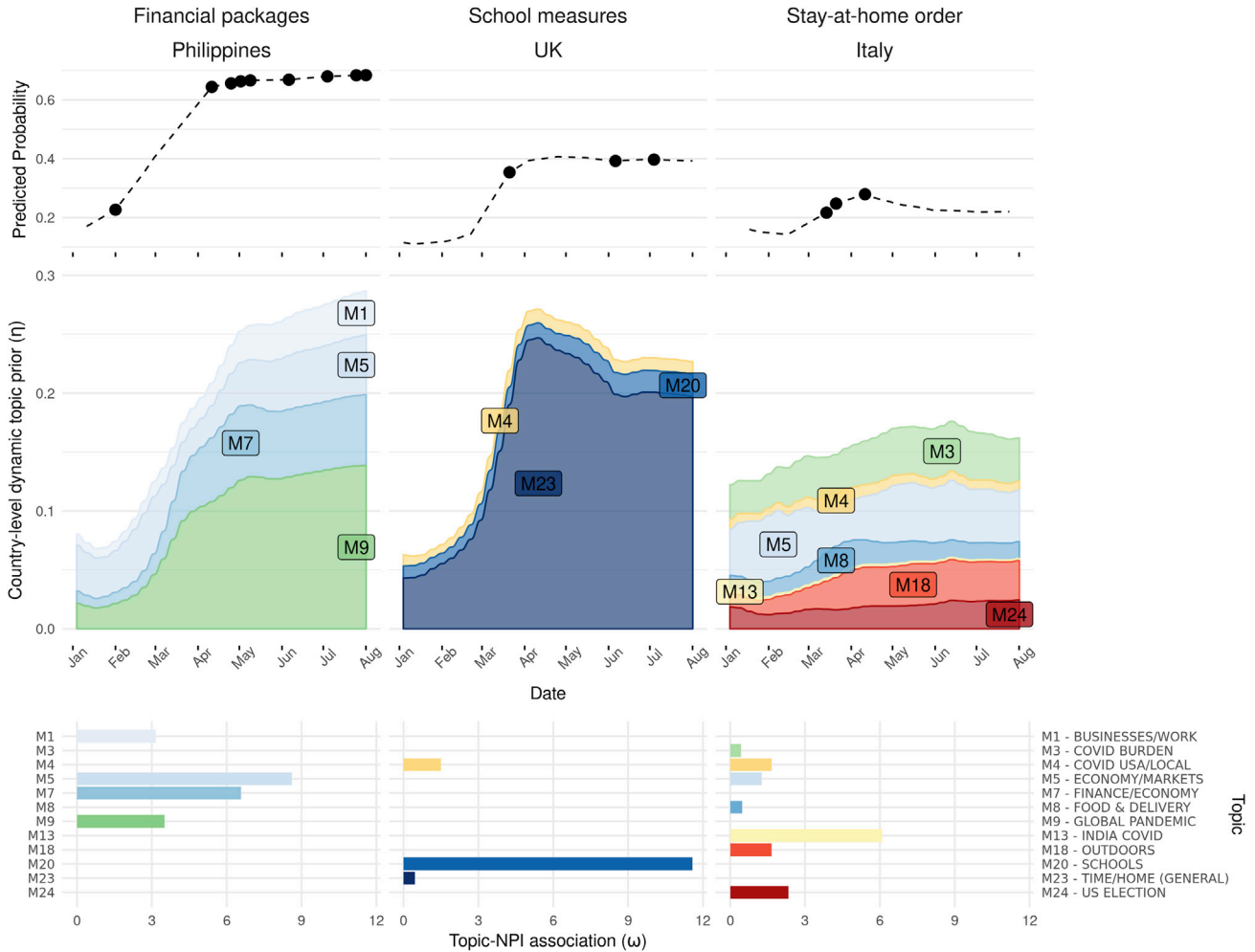
The word embedding ρ are fixed.

Figure 7. Example of topic dynamics and NPI predictions (using the fine-tuned method) for three select countries
In the top panel, the predicted probabilities over time of financial packages in the Philippines, "school measures" in the UK, and "stay-at-home orders" in Italy are represented by the dashed lines. The dots represent the date of the NPIs. These examples were selected to highlight a range of good to poor prediction accuracy (left to right). Predictions are informed by temporal topic dynamics within the AYLIEN dataset, represented in the middle panel. The bottom panel shows weights of each topic for a given NPI prediction, based on topic-NPI associations (ω) learned via predicting NPIs from the media articles in the WHO dataset. Topics with the predictive associations below 0 (i.e., $\omega_k < 0$) are not shown.

Inference

Our model has several latent variables including the dynamic topic prior η_s per source s, the topic mixture per document θ_d, the topic assignment per word per document token z_{dn}, and the topic embedding α_k per topic k. Word embedding ρ is treated as fixed point estimates and optimized via empirical Bayes. The data likelihood follows a categorical distribution:

$$p(\mathcal{D}|\theta,\beta) = \prod_d^D \prod_n^{N_d} N_d p(w_{dn}|\beta,\theta_d) = \prod_{d,n} \sum_k \theta_{dk}\beta_{w_{dn}k} = \prod_{d,n} \theta_d \beta_{w_{dn}} = \prod_{d,v}(\theta_d\beta_v)^{w_{dv}},$$

(Equation 1)

where $\mathcal{D} = \{w_1, w_2, ..., w_D\}$ denote the corpus, w_d denotes the frequency of each word in document d, D is the number of documents in the corpus, N_d is the number of tokens in document d, $\beta_{kv} = \text{softmax}(\alpha_k\rho_v)$ for word $w_{dn} = v$ and topic k, and w_{dv} is the total count of word index v in document d. Therefore, the log likelihood is: $\log p(\mathcal{D}|\theta,\beta) = \sum w_{dv}\log(\theta_d\beta_v)$

The posterior distribution of the other latent variables $p(\eta,\alpha,\theta|\mathcal{D})$ are intractable. To approximate them, we took an amortized variational inference approach using a family of proposed distributions $q(\eta,\alpha,\theta)$:[7]

$$q(\eta,\alpha,\theta) = \prod_d q\left(\theta_d \Big| \eta_{s_d}^{(t_d)}, w_d\right) \prod_s \prod_t q\left(\eta_s^{(t)}\Big|\eta_s^{(t-1)}, \bar{w}_s^{(t)}\right) \prod_k q(\alpha_k),$$

where

$$\eta_s^{(0)} = 0, \quad \left[\mu_s^{(t)}, \log\left(v_s^{(t)}\right)^2\right] = LSTM\left(\left[\eta_s^{(t-1)}, \bar{w}_s^{(t)}\right]; W_\eta\right)$$
$$\eta_s^{(t)} \sim \mu_s^{(t)} + \text{diag}(v_s^{(t)})\mathcal{N}(0,I_K),$$

(Equation 2)

$$= NNET\left(\left[\eta_{s_d}^{(t_d)}, \bar{w}_d\right]; W_\theta\right), \quad \delta_d \sim \mu_d + \text{diag}(\sigma_d)\mathcal{N}(0,I_K), \quad \theta_d = \text{softmax}(\delta_d),$$

(Equation 3)

$$\alpha_k \sim m_k + \text{diag}(\zeta_k)\mathcal{N}(0,I_L).$$

(Equation 4)

Here using the BOW representation, w_d denotes a $V \times 1$ vector of the word frequency of document d over the vocabulary of size V; $\bar{w}_d$ is the normalized word frequencies; $\bar{w}_s^{(t)}$ denotes average word frequency at time t for source s.

Using the variational autoencoder framework,[17] the function $NNET(x; W_\theta)$ is a feedforward neural network parametrized by W_θ; $LSTM(x; W_\eta)$ is a long

Table 3. AUPRC scores for country-level NPI prediction from topics at document and country level

	Topic level	Weighted AUPRC	Macro AUPRC
Random eta, alpha, and omega	Country	0.411 (0.007)	0.323 (0.006)
Random eta and omega	Country	0.411 (0.007)	0.323 (0.006)
Random omega	Country	0.413 (0.017)	0.326 (0.013)
Zero-shot	Country	0.451 (0.001)	0.352 (0.001)
	Document	0.448 (0.001)	0.350 (0.001)
From scratch	Country	0.455 (0.001)	0.367 (0.001)
	Document	0.506 (0.000)	0.398 (0.000)
Fine-tuned	Country	0.503 (0.000)	0.404 (0.000)
	Document	0.506 (0.000)	0.397 (0.000)

Values in the brackets are standard deviations. Random baselines are each repeated 1,000 times with different random seeds, and the rest are each repeated 100 times with random seeds.

short term memory (LSTM) network[18] parametrized by $\mathbf{W}_\eta$. Because of the Gaussian properties, we use the re-parametrization trick[17] to stochastically sample the latent variable θ_d, $\eta_s^{(t)}$, and α_k from their means with added Gaussian noise weighted by their variances, as shown in Equations 2, 3, and 4, respectively.

To learn the above variational parameters $\Theta_q = \{\mathbf{W}_\theta, \mathbf{W}_\eta, \mathbf{m}_\alpha, \zeta_\alpha\}$, we optimize the evidence lower bound (ELBO), which is equivalent to minimizing the Kullback-Leibler (KL) divergence between the true posterior and the proposed distribution $KL(q(\Theta)|p(\Theta|\mathcal{D}))$:

$$ELBO = \mathbb{E}_q[\log p(\mathcal{D}|\eta, \alpha, \theta)] - KL[q(\eta, \alpha, \theta)||p(\eta, \alpha, \theta)]. \quad \text{(Equation 5)}$$

We optimize ELBO with respect to the variational parameters using amortized variational inference.[7,17,19,20] Specifically, we sample the latent variable η using Equation 2, θ using Equation 3, α using Equation 4 based on a mini-batch of data. We then use those samples as the noisy estimates of their variational expectations for the ELBO (5). The ELBO is optimized with backpropagation using Adam.[21]

Selecting the topic number based on topic quality
We use topic quality to select the best number of topics. Topic quality is calculated as the product of topic diversity and topic coherence. Topic diversity is defined as the percentage of unique words in the top 25 words from each topic across all topics, and topic coherence is defined as the average point-wise mutual information of the top 10 most likely words under each topic:

$$TC = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{45}\sum_{i=1}^{10}\sum_{j=i+1}^{10} f(w_i, w_j). \quad \text{(Equation 6)}$$

Implementation
EpiTopics-stage 1 or unsupervised MixMedia[6] is implemented and trained with PyTorch 1.5.0. We used 300-dimensional word and topic embeddings. We implemented the neural network in Equation 3 with a two-layer feedforward network, with a hidden size of 800 and ReLU activation. We implemented the LSTM in Equation 2 with a 3-layer LSTM with a hidden size of 200 and a dropout rate of 0.1. We set the initial learning rate to 10^{-3} and reduced it by

1/4 up to the minimal value of 10^{-7} if the validation perplexity is not reduced within the last 10 epochs. The model was trained for 400 epochs with a batch size of 128. The number of topics, 25, was chosen based on the best topic quality, as shown in Figure S1.

EpiTopics-stage 2: Predicting document-level interventions
At this stage, we use the unsupervised MixMedia trained at stage 1 to generate the expected topic mixtures of each article in the WHO dataset to predict the change in the status of an NPI reported in the article. The inputs to the classifier are the topic mixture of the WHO article produced by the MixMedia model trained on the AYLIEN data in the first stage. Specifically, the topic mixtures θ_d for WHO document d from country s_d observed at the t_d^{th} CDC week were produced from a feedforward encoder trained at stage 1:

$$\mu_d, \log \sigma_d^2 = NNET\left(\left[\widehat{\eta}_{s_d}^{(t_d)}, \bar{\mathbf{w}}_d\right]\right), \quad \text{(Equation 7)}$$

$$\theta_d = softmax(\mu_d), \quad \text{(Equation 8)}$$

where $NNET$ is the trained encoder of MixMedia in Equation 3, $\bar{\mathbf{w}}_d$ is the normalized word frequencies of document d, and $\widehat{\eta}_{s_d}^{(t_d)} \equiv \mu_{s_d}^{(t_d)}$ is the expectation of the country-level topic mixture for country s_d at time t_d, which is produced by the trained LSTM of MixMedia in Equation 2.

We then fit a linear classifier to predict the 15 NPIs for document d using its topic mixture θ_d:

$$\widehat{\mathbf{y}}_d = \sigma(\theta_d \alpha \omega + \mathbf{b}), \quad \text{(Equation 9)}$$

where $\sigma(x) = 1/(1 + e^{-x})$ is the sigmoid function, $\widehat{\mathbf{y}}_d$ is the vector of predicted probabilities of the 15 NPIs, ω is the linear weights, and $\mathbf{b}$ is the bias vector. Notably, here we projected the topic mixture (θ) by the topic embedding (α, also learned from stage 1) to yield a document embedding input matrix. This α embedding allowed for higher dimensional representation of concepts relevant to NPIs for better prediction performance than using θ alone. To predict the NPI mentioned in an article, we fit a logistic regression model. We chose a linear classifier here for the ease of interpretability and also due to small training sample size. The associations between topics and NPIs can be obtained by inspecting the inner product of the topic embedding and regression coefficients: $\alpha\omega$, which gives a $K \times 15$ matrix indicating the associations between the K topics and 15 NPIs (Figure 5).

The loss function here is cross-entropy plus the L2 regularization:

$$L = \sum_d - \mathbf{y}_d \log\widehat{\mathbf{y}}_d - (1 - \mathbf{y}_d)\log(1 - \widehat{\mathbf{y}}_d) + \lambda\sum_{c=1}^{C=15}\left\|\omega_l\right\|^2. \quad \text{(Equation 10)}$$

Implementation
The linear classifier was implemented with PyTorch 1.5.0. We used Adam optimizer with a learning rate of 6×10^{-3} and L2-penalty λ of 10^{-5}. We trained the classifier for 400 epochs with a batch size of 512. To mitigate class imbalance over the 15 NPIs, we adjusted the weights for the per-NPI loss by up-weighting the loss of minority NPI class labels and down-weighting the loss of majority NPI class labels:

$$\text{weight of class } c = \frac{\# \text{ of positive labels in total}/\# \text{ of positive labels in class } c}{\sum_c (\# \text{ of positive labels in total}/\# \text{ of positive labels in class } c)}. \quad \text{(Equation 11)}$$

EpiTopics-stage 3: Predicting country-level interventions
Prediction from country-level topic priors
After the linear classifier for document-level NPI prediction is trained, its weights are transferred to the prediction task for country-level NPI labels:

$$\widehat{\mathbf{y}}_s^{(t)} = \sigma\left(\widehat{\eta}_s^{(t)}\widehat{\alpha}\widehat{\omega} + \widehat{\mathbf{b}}\right). \quad \text{(Equation 12)}$$

We experimented three models as follows.

For the zero-shot-transfer model, we directly use the country-level topic mixture η and topic embedding α from learned stage 1 and the trained linear classifier coefficients $\widehat{\omega}$ and biases $\widehat{\mathbf{b}}$ from stage 2.

For the fine-tune model, the classifier weights are further updated for predicting country-level interventions. Specifically, we initialize the classifier weights with the learned parameters $\widehat{\omega}$ and further adjust it to predict country-level NPIs.

For the from-scratch model, we initialize the classifier weights by random values sampled from standard normal and learn the weights from scratch to predict the 15 NPIs from the training country-time pairs. More details are described in "experiments."

Prediction from document-level topic mixtures

We also experimented with an alternative approach by predicting country-level NPIs using documents' topic mixtures. Specifically, the country-level NPI probabilities are calculated by taking the average of the predictions of the document-level NPIs for the corresponding country and time:

$$\widehat{\mathbf{y}}_s^{(t)} = \frac{1}{\left|\mathcal{D}_s^{(t)}\right|} \sum_{d \in \mathcal{D}_s^{(t)}} \sigma(\theta_d \alpha \omega + \mathbf{b}), \qquad \text{(Equation 13)}$$

where $\mathcal{D}_s^t \subset \mathcal{D}$ is the set of documents associated with country s and time point t. Therefore, this method can only be trained on predicting NPIs for the country and time pairs, where there is at least one document observed. Similarly to the method above, predictions can be made in a zero-shot transfer way, or the classifier weights are updated from the learned weights for the fine-tuned model or from random weights for the from-scratch models.

Experiments
Baselines

We omitted the comparison with other models at EpiTopics-stage 1 and refer the readers to our MixMedia work for comparing the topic quality with baseline topic models.[6] For document NPI predictions, we compared the EpiTopics-stage 2 model (i.e., using the inferred document-level topic mixtures of each WHO document to predict their NPI labels) against the baseline models that are trained to directly use the input word frequencies as the BOW to predict WHO labels. We experimented with two baseline methods namely linear (BOW + linear) and feedforward network (BOW + feedforward).

The BOW + linear baseline is trained with learning rate 8×10^{-5} and weight decay 0.002 for 100 epochs. The BOW + feedforward baseline is trained with learning rate 10^{-4} and weight decay 0.002 for 100 epochs. The feedforward network is implemented with two fully connected layers with a hidden size of 256 and ReLU activation. Fine-tuning used learning rate 6×10^{-2} and weight decay 1×10^{-3} with a batch size of 1,024. The from-scratch model used learning rate 2×10^{-3} and weight decay of 2×10^{-1} with a batch size of 1,024.

For country-level NPI prediction, we compared our proposed transfer-learning approach EpiTopics-stage 3 against two baseline models: (1) training from scratch; (2) model using randomly generated weights. For the random model, we randomly initialized the country-specific topic mixture η_s^t, topic embedding α and/or classifier parameters $\{\omega, \mathbf{b}\}$ and did not update their weights. To measure the effect of different levels of randomness, we experimented with three random baselines as follows:

- random country-topic mixture η, random topic embedding α, and random topic-NPI regression coefficients ω: setting η_s^t, α, and $\{\omega, \mathbf{b}\}$ to random numbers
- random η and ω: setting η_s^t and $\{\omega, \mathbf{b}\}$ to random numbers
- random ω: setting $\{\omega, \mathbf{b}\}$ to random numbers

Intuitively, the first random baseline preserves no information gained from the training of document-level NPI predictions, and thus represents a complete random model. The other two baselines preserve some level of information and thus may perform better than the first baseline.

Evaluation

Since NPI prediction is a multi-label classification problem, and that different NPIs can have different optimal thresholds, we used weighted AUPRC and macro AUPRC as evaluation metrics. We computed the AUPRC for each

NPI individually. We took either the unweighted average of them (macro AUPRC) or the weighted average of AUPRC by each classes' prevalence (weighted AUPRC).

Our contributions in contrast to the existing computational methods

Other researchers have applied supervised learning to online media[3] and Wikipedia articles[9] to identify COVID-19 NPIs. These approaches either entirely rely on the limited manually labeled data in a supervised learning framework or are difficult to interpret. In contrast, our approach has less demand for labeled data as it exploits large-scale unlabeled data via unsupervised and interpretable topic modeling combined with the transfer-learning strategy. Topic modeling has been used for surveillance of online media to detect epidemics[22–25] and to characterize news reports[26] and social media[27] in the context of COVID-19 or other outbreaks. Our work differs from these applications due to a focus on tracking NPIs as opposed to detecting epidemics. By adopting a dynamic topic model, EpiTopics goes beyond the document-level to support inference at the country level.

AUTHOR CONTRIBUTIONS

Y.L. and D.B. conceived the study. Y.L. and Z.W. developed the model with critical help from D.B. and G.P. I.C. collected and processed the data. Z.W. implemented the model and ran the experiments. Z.W., G.P., D.B., and Y.L. analyzed the results and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Winslow, C.-E. (1920). The untilled fields of public health. Science, 23–33.

2. Sanson-Fisher, R.W., Bonevski, B., Green, L.W., and D'Este, C. (2007). Limitations of the randomized controlled trial in evaluating population-based health interventions. Am. J. Prev. Med. *33*, 155–161.

3. Cheng, C., Barceló, J., Hartnett, A.S., Kubinec, R., and Messerschmidt, L. (2020). COVID-19 government response event dataset (coronanet v. 1.0). Nat. Hum. Behav. *4*, 756–768.

4. Brauner, J.M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenčiak, T., Stephenson, A.B., Leech, G., Altman, G., Mikulik, V., et al. (2021). Inferring the effectiveness of government interventions against COVID-19. Science *371*, eabd9338.

5. Blei, D.M., Ng, A.Y., and Jordan, M.I. (2003). Latent dirichlet allocation. J. Machine Learn. Res. *3*, 993–1022.

6. Li, Y., Nair, P., Wen, Z., Chafi, I., Okhmatovskaia, A., Powell, G., Shen, Y., and Buckeridge, D. (2020). Global surveillance of covid-19 by mining news media using a multi-source dynamic embedded topic model. In Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, pp. 1–14.

Patterns
Article

7. Dieng, A.B., Ruiz, F.J.R., and Blei, D.M. (2019). The dynamic embedded topic model. arXiv, 1907.05545.

8. Saito, T., and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10, e0118432. https://doi.org/10.1371/journal.pone.0118432.

9. Suryanarayanan, P., Tsou, C.-H., Poddar, A., Mahajan, D., Dandala, B., Madan, P., Agrawal, A., Wachira, C., Samuel, O.M., Bar-Shira, O., et al. (2021). AI-assisted tracking of worldwide non-pharmaceutical interventions for COVID-19. Sci. Data 8, 94.

10. Dion, M., AbdelMalik, P., and Mawudeku, A. (2015). Big data: big data and the global public health intelligence network (gphin). Can. Commun. Dis. Rep. 41, 209.

11. Brauner, J.M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenčiak, T., Stephenson, A.B., Leech, G., Altman, G., Mikulik, V., et al. (2021). Inferring the effectiveness of government interventions against COVID-19. Science (New York, NY) 371.

12. Clark, K., Luong, M.-T., Le, Q.V., and Manning, C.D. (2020). ELECTRA: Pre-training Text Encoders as Discriminators rather than Generators (ICLR).

13. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: pre-training of deep bidirectional transformers for language understanding. arXiv, 1810.04805.

14. Jain, S., and Wallace, B.C. (2019). Attention is not explanation. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, pp. 3543–3556.

15. Yasir Utomo, M.N., Adji, T.B., and Ardiyanto, I. (2018). Geolocation prediction in social media data using text analysis: a review. In 2018 International Conference on Information and Communications Technology (ICOIACT) (IEEE), pp. 84–89.

16. Dieng, A.B., Ruiz, F.J.R., and Blei, D.M. (2020). Topic modeling in embedding spaces. Trans. Assoc. Comput. Linguistics 8, 439–453.

17. Kingma, D.P., and Welling, M. (2014). Auto-encoding Variational Bayes. arXiv.

18. Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. Neural Comput. 9, 1735–1780.

19. Ranganath, R., Gerrish, S., and Blei, D. (2014). Black box variational inference. In Artificial intelligence and statistics (PMLR), pp. 814–822.

20. Hoffman, M.D., Blei, D.M., Wang, C., and Paisley, J. (2013). Stochastic variational inference. J. Machine Learn. Res. 14, 1303–1347.

21. D Kingma and J Ba. Adam: a method for stochastic optimization in: Proceedings of the 3rd International Conference for Learning Representations (ICLR'15). San Diego, 2015.

22. Walter, D., and Ophir, Y. (2019). News frame analysis: an inductive mixed-method computational approach. Commun. Methods Measures 13, 248–266.

23. Ophir, Y. (2018). Coverage of epidemics in American newspapers through the lens of the crisis and emergency risk communication framework. Health Secur. 16, 147–157.

24. Wang, Y., and Goutte, C. (2018). Real-time Change Point Detection Using On-Line Topic Models (COLING).

25. Ghosh, S., Chakraborty, P., Nsoesie, E.O., Cohn, E., Mekaru, S.R., Brownstein, J.S., and Ramakrishnan, N. (2017). Temporal topic modeling to assess associations between news trends and infectious disease outbreaks. Sci. Rep. 01, 1–12.

26. Poirier, W., Ouellet, C., Rancourt, M.-A., Béchard, J., and Dufresne, Y. (2020). (Un)covering the COVID-19 pandemic: framing analysis of the crisis in Canada. Can. J. Polit. Sci. Revue canadienne de Sci. politique 53, 365–371.

27. Han, X., Wang, J., Zhang, M., and Wang, X. (2020). Using social media to mine and analyze public opinion related to covid-19 in China. Int. J. Environ. Res. Public Health 17, 2788.