Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Constructing public health evidence knowledge graph for decision-making support from COVID-19 literature of modelling study

Yunrong Yangb,a, Zhidong Caob,a,c,e, Pengfei Zhaoca,a,b,c, Dajun Daniel Zengb,a,c, Qingpeng Zhangd, Yin Lub,b,c

Keywords:
Evidence-based public health
Knowledge graph
COVID-19
Decision-making support

A R T I C L E   I N F O

The needs of mitigating COVID-19 epidemic prompt policymakers to make public health-related decision under the guidelines of science. Tremendous unstructured COVID-19 publications make it challenging for policymakers to obtain relevant evidence. Knowledge graphs (KGs) can formalize unstructured knowledge into structured form and have been used in supporting decision-making recently. Here, we introduce a novel framework that can extract the COVID-19 public health evidence knowledge graph (CPHE-KG) from papers relating to a modelling study. We screen out a corpus of 3096 COVID-19 modelling study papers by performing a literature assessment process. We define a novel annotation schema to construct the COVID-19 modelling study-related IE dataset (CPHIE). We also propose a novel multi-tasks document-level information extraction model SS-DYGIE++ based on the dataset. Leveraging the model on the new corpus, we construct CPHE-KG containing 60,967 entities and 51,140 relations. Finally, we seek to apply our KG to support evidence querying and evidence mapping visualization. Our SS-DYGIE++(SpanBERT) model has achieved a F1 score of 0.77 and 0.55 respectively in document-level entity recognition and coreference resolution tasks. It has also shown high performance in the relation identification task. With evidence querying, our KG can present the dynamic transmissions of COVID-19 pandemic in different countries and regions. The evidence mapping of our KG can show the impacts of variable non-pharmacological interventions to COVID-19 pandemic. Analysis demonstrates the quality of our KG and shows that it has the potential to support COVID-19 policy making in public health.

A B S T R A C T

1. Introduction

The emergence of the COVID-19 pandemic is the worst crisis at this century and is posing an unprecedented threat to global public health and the economy. Due to the lack of adequate vaccinations, implementing policy interventions is of significance with the guidance of scientific evidence to mitigate the pandemic [1]. Scientists in different disciplines have devoted themselves to investigating and understanding the disease, leading to the rapid increasing of COVID-19-related (COR) literature. It is in tremendous need to mine and integrate evidence contained in these papers to inform high-quality public health decision-making.

Knowledge graphs (KGs) are featured with representing structured knowledge and have recently been used for supporting decision-making, mostly for clinical decision-making support [2–4]. Constructing disease or health knowledge graph from texts has been an active field where previous studies rely highly on the annotations of domain experts. For example, Sheng et al. [5] proposes DEKGB, a framework to construct health knowledge graph with doctor-involved. Huang et al. [6] builds a disease-centric knowledge graph for depression from 10,190 trials, where psychiatric experts are employed to provide clinical use-cases. Finlayson et al. [7] extracts medical terms based on dictionary complied by clinical experts to build medicine graph from millions of clinical narratives. Moreover, some health graphs pay more attention to personal health rather than public health. Shen et al. [8] utilizes a clinical bayesian network to construct a medical ontology with over 17 aspects including patient’s condition, personal history, family medical history and etc. directly from over 10,000 unidentified patient records. Zhao et al. [9] constructs a medical knowledge graph manually from 992 electronic medical records (EMRs) and the graph is regarded as markov random field for inference. Rotmensch et al. [10] extracts medical con-
cepts from 273,174 deidentified patient records with three probabilistic models to automatically construct knowledge graphs.

To defence against the COVID-19 pandemic, researchers are working in developing COVID-19 knowledge graphs (KGs) to accelerate information collection and synthesis from COR scientific papers. Previous studies such as CKG [11] aims to present relationships between COR scientific articles and concepts. And Reese et al. creates KG-COVID-19 which covers mechanistic information associated with the novel coronavirus, such as drug-target interactions, biological processes and genes [12]. Other KGs explore relationships of COVID-19 pathophysiology [13] as well as application for drug repurposing [14, 15]. More similar work to ours is the KG released by Wang et al. which extracts fine-grained biomedical knowledge from scientific literature [16]. Hope et al. extracts sentence-level mechanism relations from COVID-19 papers with LM-driven model for knowledge base construction [17]. However, these KGs mostly focus on integrating the knowledge of coronavirus, but few present evidence such as transmission of epidemic, that can provide critical information for determining the type and intensity of disease interventions. Compared with the studies discussed above, our work aims to construct a KG that comprises evidence for public health decision-making. For COVID-19, it is unethical to conduct empirical trials. Hence, observational trials play an important role in providing evidence of epidemic trends. In this paper, we extract sufficiently plentiful evidence and their interrelationships from COR modelling study papers to construct the COVID-19 public health evidence knowledge graph (CPHE-KG) applied to decision-making support.

In general, KGs use a series of subject-predicate-object triples to represent different kinds of entities and relations in a domain. We generate the CPHE-KG based on structured evidence triples selected from COR papers of modelling study retrieved from CORD-19 dataset [18]. Fig. 1 illustrates the types of evidence and their relationships of a modelling study in the proposed KG. These evidence types include epidemic metrics (e.g., final infection numbers, or Rt), metric measures (e.g., greater, or 3,782,558), interventions (e.g., social distancing, or mandatory masking), modelling methods (e.g., SIR mathematical model), time nodes (e.g., the 92th day from the first infectious), regions (e.g., Bangladesh), etc. These relationships include within-sentence relations showing the context information and cross-sentence coreference.

Recently, information extraction (IE) has been used to extract mentions of evidence entities and relations from scientific articles [19–22]. In the biomedical and health domain, these extraction systems [23–26] mostly aim to extract evidence from papers of randomized controlled trials (RCTs) and the entity types are designed as populations, interventions/comparators, and outcomes, i.e., PICO frames. However, PICO frames are not adaptive for extracting evidence from modelling study papers whose structures are different from RCTs.

To resolve the aforementioned issues, we summarize a schema from COR evidence-based public health papers to annotate a dataset of evidence for COVID-19 modelling study papers to help knowledge discovery for the governance of epidemic. We propose a novel document-level information extraction model SS-DYGIE++ to extract evidence triples and coreference clusters. These triples and clusters form the foundation of KG of a single document. By utilizing knowledge fusion approaches, we merge the extracted KGs of documents into a comprehensive knowledge graph. The final CPHE-KG contains 60,967 entities and 51,140 relations. To further demonstrate the quality of the CPHE-KG, we provide a decision-making support service based on the CPHE-KG from two aspects: evidence querying and evidence mapping visualization. On the one hand, with evidence querying, the system returns the evidence subgraph relevant to query terms. On the other hand, evidence mapping visualization presents the evaluation results of the effectiveness of interventions in the COVID-19 pandemic.

The contributions of this paper are threefold.

1) We propose a novel annotation schema which includes critical policy-making-related evidence and relations in the abstract of COVID-19 modelling study literature. Moreover, to the best of our knowledge, our CPHE dataset guided with this schema is the first multi-tasks IE dataset on the basis of papers of modelling study in public health domain.

2) We introduce a novel document-level IE model SS-DYGIE++ to extract evidence from COVID-19 modelling study papers. The results demonstrate our model is feasible to achieve automatic KG construction. We release an COVID-19 public health evidence knowledge graph (CPHE-KG), which is an early exploration of KG applied to public health decision-making.

3) Our KG can provide two main functions: evidence querying and evidence mapping visualization, which has been proved to be effective in supporting decision-making relevant to COVID-19 public health.

The remainder of this paper is organized as follows. In Section 2, we describe the required pipelines for constructing a knowledge graph of evidence in COR papers of modelling study. In Section 3, we present the extracted results of evidence and the CPHE-KG. In Section 4, the case study to test the CPHE-KG is given. In Section 5, we conclude our framework and discuss future work. Finally, the annotation guideline of the IE dataset is proposed in Appendix.

2. Methodology

In this section, we present a systematic procedure to develop the CPHE-KG. As shown in Fig. 2, the procedure involves four main steps, including:

a) data preparation: CORD-19 dataset serves as the data sources of COVID-19 public health IE dataset and KG construction;

b) COVID-19 public health IE dataset construction: The coloured text spans represent entities; the solid arrows reflect the relations between the entities and the dotted arrows indicate co-reference relationships;
c) document-level knowledge extraction: The extracted knowledge and coreference clusters of a single abstract are first integrated into knowledge graph and we can get as many knowledge graphs as the abstracts;

d) knowledge fusion: After performing knowledge fusion to previously obtained knowledge graphs in step b), we acquire the final knowledge graph (CPHE-KG).

2.1. Data

In this paper, the public health KG is built based on the COR papers of modelling study in CORD-datasets. CORD-19 dataset is released by Kaggle, including more than 500,000 publications on COVID-19 and COR research from PubMed, WHO, bioRxiv and medRxiv. We take the query (COVID-19 OR coronavirus OR “sar cov”) AND (model OR forecast OR assess OR predict OR dynamic OR transmission) to retrieval papers of modelling study. As for COVID-19, empirical studies, such as RCTs are unethical and not feasible. Modelling study, especially the study of dynamic transmission models, can evaluate the effects that come from infectious disease interventions [27]. Therefore, we consider modelling study as the available evidence in scientific literature so far.

We remove papers irrelevant with the transmission of COVID-19 and keep papers with enough information we need. As there is no available ‘Risk of bias’ checklist for modelling study, we follow assessment guidelines from the International Society for Pharmacoeconomics and Outcomes (ISPOR) and the Society for Medical Decision making (SMDM) to grade the transmission models. Our grade criteria are similar to the criteria in the COVID-19 rapid systematic review [28], which are presented as follows:

- Is the model a dynamic transmission model?
- Does the paper conduct uncertainty analyses?
- Does the paper provide measure values of epidemic metrics related to the interventions?

The final corpus contains 3096 papers, which is the foundation of KG. Inclusion procedure is shown in Fig. 3. Firstly, 5960 papers are retrieved from CORD-19 datasets with the retrieval mentioned above. We then exclude papers that are public health irrelevant, repetitive, and too short or too long by hand search. Further, we assess the quality of these papers with the grade criteria. Finally, the remaining corpus are formed with 877 papers from PMC, 1576 papers from WHO, 570 papers from MedRxiv, and 73 papers from BioRxiv.

2.2. Dataset construction

We construct a dataset (called CPHIE: COVID-19 Public Health Information Extraction) of 597 richly annotated abstracts of COVID-19 modelling study papers taken from the CORD-19 dataset. Our dataset annotates public health-related entities, relations and coreference clusters. The statistics of the CPHIE dataset are shown in Table 1.

Referring to the systematic reviews and meta-analysis related to evidence-based public health of COVID-19, we summarize the most concerned entity and relation types as the foundation of our dataset. We define 10 entity types (Task, Region, Intervention, Method, Indicator, Measure, Time, Variable, Generic, OtherTerm) and 10 relation types (USED-FOR, EQUAL-TO, ASSOCIATE-WITH, HAPPEN-IN, LOCATED-IN, CONJUNCTION, PART-OF, COMPARED-TO, FEATURE-OF, EVALUATE-FOR) that can embody characteristics of transmissions of infectious diseases and are useful in public health decision-making. Typically, we consider entity type of Time, which is significant to analyse the temporal condition of COVID-19 epidemic.

To annotate a high-quality dataset efficiently, we take a semi-automatic labelling scheme. We first utilize the natural language process tool spaCy (https://spacy.io/) library to automatically label the general entity types: Region entities and Time entity. We also use it to tag the Arabic numbers. The noisy tagging results are checked manually subsequently. Domain-related entities and relations are primarily labelled referring to rules and templates then are manually annotated using the BRAT visual text annotation platform [29] according to the Annotation Guidelines proposed in Appendix A. Since there are few nested entities in the abstracts, nested entities are not annotated in our dataset. A public health domain annotator annotated all abstracts in the dataset, and another public health domain annotator checked and revised the annotations. Fig. 4 shows an annotation example of CPHIE dataset.

Table 1

| Total | per Abstract | per Sentence |
|-------|--------------|--------------|
| # of sentences | 5507 | 9.22 | - |
| # of words | 139,213 | 233.19 | 25.27 |
| # of entities | 11,616 | 19.46 | 2.11 |
| # of relations | 8400 | 14.07 | 1.53 |
| # of coreference clusters | 1771 | 2.97 | - |
2.3. Document-level information extraction

For COVID-19 public health domain document-level information extraction task, assume there are two corpuses: a) a small richly labelled seed document set \( D_r = \{ D_1, D_2, \ldots, D_n \} \), which is CPHIE dataset. b) a small unlabelled raw document set \( D_u \) used for semi-supervised learning. Given a document \( D_i \), which is first represented as a sequence of tokens \( D_i = \{ t_1, t_2, \ldots, t_m \} \) for subsequent processing. Then we construct the set of spans \( S = \{ s_1, s_2, \ldots, s_N \} \) which are the ordered sequence of within-sentence tokens. Our task includes three subtasks, entity recognition, relation extraction, and coreference resolution. Entity recognition implicates predicting the best entity type label for each span. Relation extraction implicates predicting the best relation type label between a pair of spans. And coreference resolution implicates clustering spans that refer to the same entity.

We propose a novel multi-tasks IE model SS-DYGIE++, which is improved on the foundation of DYGIE++ model [21]. DYGIE++ model is the state-of-art multi-task model in many document-level information extraction datasets. Since the scale of the annotated corpus is relatively small, in this study, we take full advantage of the unlabelled raw in-domain corpus with the semi-supervised learning method. Our SS-DYGIE++ model include two improvements, language model (LM) post-training and semi-supervised learning. The architecture of our approach is depicted in Fig. 5.

**Language Model (LM) Post-training.** To adapt pretrained language model to COVID-19 public health domain, we do language model post-training on the abstracts of large-scale public health papers of COVID-19. The pre-trained language models adapted in this paper is SpanBERT [30]. SpanBERT is pretrained with span boundary objective (SBO) – which tries to predict the entire masked span using only the representations of the tokens at the span’s boundary. It outperforms BERT [31] in coreference resolution task. We choose it as a basic language model, considering the importance of coreference entities for knowledge graph construction. We follow the schema of pretraining SpanBERT to post-train it on 140K+ raw COVID-19 public health papers and then finetune it on document-level information extraction task.

**Semi-supervised Learning.** We try to further improve the performance of DYGIE++ with bootstrapping, which is a classic semi-
supervised learning algorithm [32] and is proven to be effective in information extraction tasks [33, 34]. After training DYGIE++ on CPHIE dataset $D_s$, we save the trained model and apply it to the unlabelled corpus $D_u$. Next $D_u$ with noisy predicted labels is combined with $D_s$ to update DYGIE++ model iteratively. The overall process is shown in Table 2.

First, we train an initial DYGIE++ model $M_0$ based on the training set of CPHIE dataset. Fig. 5 depicts the DYGIE++ model architecture, which has a token encoding layer, a span enumeration layer, a span graph propagation layer, and a multi-task classification layer where the feedforward neural network (FFNN) concatenated with a SoftMax function is applied for predicting the result. We use the post-trained SpanBERT for token representation. Each sentence is imported to SpanBERT together with a neighbourhood of surrounding sentences. We enumerate and construct spans $S = \{s_1, s_2, ..., s_N\}$ of document by concatenating the tokens representing their left and right endpoints, together with a learned span width embedding.

As for different subtasks: relation extraction and coreference resolution, we adopt different variants of graph propagation to update each span representation by integrating neighbour span representations. And the final updated span representations are input to a two-layer feedforward neural network (FFNN) with SoftMax function which outputs predictions for each of the end tasks and output probabilities are represented as $p(y)$, where $y$ is the predicted result.

Further, we use the initial DYGIE++ model $M_0$ to assign labels on $D_u$, and get a weakly-annotated dataset $D_u^\ast$. Moreover, we combine $D_s$ with $D_u^\ast$ and get an augmented dataset. Different from training, we set thresholds $\theta$ to filter tags with low probabilities output by the classification layer. We random divide $D_s$ into $K$ subsets and iteratively train the DYGIE++ model with $D_u^\ast$, until the model has achieved an acceptable level of accuracy, or until the maximum number of iterations.

2.4. Knowledge fusion

In this paper, knowledge fusion includes two subtasks, entity normalization and entity alignment. The expression for the same entity in different papers could be various, hence entity normalization is needed to map the original entity into a standard one. And entity alignment is required to merge entities that have the same meaning.

2.4.1. Entity normalization

With respect to generic entities and entities relevant in terms of public health field, we construct a dictionary for them according to Medical Subject Headings (MeSH) terms and match the training set with this dictionary. However, as for Time entity, it is difficult to solve with a dictionary. Since learning-based methods usually need more training instances as a result, therefore, we opt for a rule-based method the same as Ning et al. [35] that is simple and efficient enough for time normalization.

2.4.2. Entity alignment

To merge the set of small size knowledge graphs, we perform entity alignment by utilizing the cluster algorithm. This subtask is performed because there exit some entities that are not in the dictionary. We perform TF-IDF to transform entities into vector representation and further measure the cosine similarity between the entity pairs. An entity pair is considered to refer to the same entity when the similarity is higher than 0.5.
Table 3

| Model             | Entity Identification | Relation Identification | Coreference Resolution |
|-------------------|-----------------------|-------------------------|------------------------|
|                   | metric                | P | F1 | R    | P | F1 | R    | P | F1 | R    |
| SS-DYGIE++        | strict                | 0.66 | 0.60 | 0.56 | 0.52 | 0.34 | 0.25 | 0.47 | 0.26 | 0.17 |
| (SciBERT)         | strict                | 0.78 | 0.65 | 0.64 | 0.63 | 0.43 | 0.36 | 0.51 | 0.41 | 0.28 |
| SS-DYGIE++        | rough                 | 0.67 | 0.67 | 0.67 | 0.46 | 0.40 | 0.35 | 0.38 | 0.31 | 0.26 |
| (BioMed-RoBERTa)  | rough                 | 0.72 | 0.70 | 0.71 | 0.72 | 0.53 | 0.51 | 0.44 | 0.37 | 0.31 |
| SS-DYGIE++        | strict                | 0.67 | 0.67 | 0.67 | 0.48 | 0.41 | 0.35 | 0.43 | 0.31 | 0.25 |
| (SpanBERT)        | rough                 | 0.88 | 0.77 | 0.77 | 0.78 | 0.46 | 0.46 | 0.61 | 0.55 | 0.58 |

3. Results

3.1. Results of document-level information extraction

We compare our SS-DYGIE++(SpanBERT) model with the model variations below. These variations have the same architecture with SS-DYGIE++(SpanBERT) model, and the difference lines in the pretrained language model.

**SS-DYGIE++(SciBERT):** SciBERT [36] is pretrained on a corpus of which 82% are biomedical literature and 18% are computer science literature.

**SS-DYGIE++(BioMed-RoBERTa):** BioMed-RoBERTa [37] is pretrained with the Masked language modelling (MLM) objective and was adapted to 2.68 million scientific papers from the Semantic Scholar corpus via continued pretraining.

For entity recognition, we experiment two evaluation metrics to adapt to our dataset. To begin, we evaluate entity recognition with a rigid evaluation metric, based on which the prediction is considered correct only when both the span and span’s label match with the gold entity. Empirically, analysing the entity prediction results on testing set, we observe most of the long span predictions that did not match with the gold entities accurately, could however represent entity meaning correctly. For instance, the **Indicator** entity “basic reproductive number (RO)” could be recognized as “basic reproductive number” or “basic reproductive number (RO).” This is unsurprising given the complexity of the spans presented in our dataset. In this scenario, the rigid evaluation metric is excessively stringent. Motivated by this observation, we further consider relaxing the equivalence criteria. In the context of true predicted entity label, if gold entity span contains predicted span or vice versa, the entity prediction is recognized as correct. The rough metric has the significant advantage of explicitly illustrating the improvement of our IE approach.

For relation extraction, a relation is correct if both the span pairs and relation labels match with a gold relation triple. For coreference resolution, coreference cluster is correct if the predicted cluster match with the gold coreference cluster.

We utilize raw texts of COVID-19 public health paper to post-train SpanBERT, SciBERT and Biomed-RoBERTa for 10 epochs respectively under the Transformers framework of HuggingFace. The maximum sequence length is set to 128 tokens, which is consistent with most sentence length of raw texts. The batch size is set to 16 when post-training on a single TITAN GPU. For document-level IE, we use DYGIE++ library and search the optimal hyper-parameters with Optuna [38].

Table 3 reports the performance of the DYGIE++ model finetuned with SpanBERT and of the different model variants. Our model outperforms baselines in entity identification and coreference resolution when evaluated with rough metric.

3.2. Results of constructed public health KG

We apply our model to 3096 abstracts of COVID-19 public health papers to construct a KG of 60,967 entities and 51,140 relations. The statistics of entities and relations could be found in Table 4.

Table 4

| Entity types       | Entity numbers | Relations | Relation numbers |
|--------------------|----------------|-----------|------------------|
| Task               | 5673           | USED-FOR  | 10,913           |
| Region             | 5802           |EQUAL-TO  | 10,326           |
| Intervention       | 5279           | ASSOCIATE-WITH | 7640 |
| Method             | 10,961         | HAPPEN-IN | 6214            |
| Indicator          | 10,664         | LOCATED-IN | 12,311         |
| Measure            | 10,931         | CONJUNCTION | 3206            |
| Time               | 7771           | PART-OF  | 531              |
| Variable           | 2035           | COMPARED-TO | 332              |
| Generic            | 1012           | FEATURE-OF | 233             |
| OtherTerm          | 839            | EVALUATE-FOR | 434            |

4. Case study

To illustrate the anticipated decision-making support of our COVID-19 public health knowledge graph, we focus on the main functions of CPHE-KG: evidence querying and evidence mapping visualization.

4.1. Evidence querying

With respect to evidence querying, CPHE-KG allows users to query conditions of interest and filter relevant components of knowledge graph. Once a set of search terms is assigned, relevant entities and relations are retrieved from the CPHE-KG. The interface of Neo4j then displays the recovered graph by the filter. Each node and link in the graphs can be clicked to explicitly exhibit the complete content.

For example, the user specifies a Region entity of “Italy” in Fig. 6, the system returns a graph about the COVID-19 epidemic development in Italy. Furthermore, users can browse the COVID-19 epidemic in temporal dimension as illustrated in Fig. 7. Additionally, users can provide interventions of interest to further narrow the search. The association of “lockdown” and “quarantine” with COVID-19 epidemic is shown in Fig. 8 and Fig. 9 respectively.

4.2. Evidence mapping

Featured with straightforward, evidence mapping [39] is a significant evidence synthesis method for evidence-based public health, which is usually employed to evaluate and integrate the evidence related to intervention measures, etc., namely the evidence mapping of interventions.

In this section, we return to the question in Section 1: which is the most effective non-pharmacological intervention against COVID-19 epidemic? We follow a process in line with international practices [39] to construct evidence mapping. To begin, we select and count (Intervention, ASSOCIATED-WITH, Indicator) triples. We further sort the triples by Intervention and filter the top eight non-pharmacological interventions.

The evidence mapping shown in Fig. 10 is displayed, providing a comprehensive presentation of the evidence available for the effectiveness of the top eight non-pharmacological interventions with respect to their outcomes for ten indicators measuring COVID-19 epidemic. We observe that, the impact of non-pharmacological interventions is largely
reflected in two indicators “transmission rate” and “confirmed cases”. We also observe that “social distancing”, “lockdown” and “quarantine” are the three most effective non-pharmacological interventions according to evidence mapping. The conclusion of our evidence mapping is similar to the results in systematic review [1], which has demonstrated the feasibility of our knowledge graph.

5 Conclusions and discussions

In this paper, we develop a procedure that can transform massive modelling study COR papers into an organized, structured, and actionable knowledge graph (CPHE-KG). In particular, we proposed a document-level information extraction model, SS-DYGIE++ to identify entities, relations and coreference clusters related to COVID-19 public health evidence. This module is of importance for knowledge graph construction. To enable the multi-task information extraction, we also created a high quality semi-automatically annotated dataset (CPHIE) and provided a series of results on the SS-DYGIE++ model and its variants.

In our SS-DYGIE++ model, we introduced a semi-supervised learning approach, bootstrapping, that dispenses with the need to have a large scale of labelled texts for information extraction by utilizing the unlabelled domain texts. We also introduced the language model post-training for token representation, which allows language models adaptive to COR public health domains and negates the need to pretrain a language model in public health domain from the scratch. However, our approach has a limitation on the length of processed documents as our language model is based on BERT, which specifies the inputs are no longer than 512 tokens. We also release two applications based on our CPHE-KG. The applications of evidence querying and evidence mapping demonstrated how CPHE-KG contribute to COR public health decision-making.

In the future work, we hope to extract evidence from long documents, such as the full text of scientific literature, rather than focusing on
Fig. 8. Intervention “lockdown” associated with COVID-19 pandemic.

Fig. 9. Intervention “quarantine” associated with COVID-19 pandemic.
abstracts. One possible solution is replacing our BERT-based language model with a language model that can process documents without length limitation. We also aim to expand the proposed KG with more evidence from modelling study papers of various epidemic diseases, such as severe acute respiratory syndrome (SARS), middle east respiratory syndrome (MERS), etc. In addition, we will target at expanding the proposed KG with multimedia evidence, e.g., figures and tables in papers to improve cross-media evidence grounding and inference.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grants No. 72025404 and No. 71621002), Beijing Natural Science Foundation (L192012) and Beijing Nova Program (Z201100006820085).

Appendix A

The COVID-19 Public Health Information Extraction dataset (CPHIE) annotation guidelines are projected primarily to provide a simple description of the various entities and their relations in modelling study papers relating to dynamic transmission of COVID-19. Referring to these guidelines, it is easy and efficaciously used by anyone who may want to use this dataset for various natural language processing tasks such as entity recognition or relation extraction. In the following subsections we summarize the guidelines that were used in annotating the 597 abstracts of COVID-19 public health papers as we explore the entities and relations that were chosen to be labelled in this dataset.

A.1. Entities

Table A.1 shows examples of entities annotated in CPHIE dataset.

| Region          | China, Italy, globe, temperate region, etc. |
|-----------------|--------------------------------------------|
| Indicator       | R0, CFR, infected cases, fatality cases, etc. |
| Method          | SEIR, machine learning, etc. |
| Interventions   | lockdown, isolation, quarantine, etc. |
| Measure         | 0.98, 2.34, greater, etc. |
| Time            | 4 April 2020, 12 days, etc. |
| Variable        | temperature, climate factors, mobility, etc. |
| Generic         | it, that, this, model, etc. |
| Other Term      | polynomial growth, pre-pandemic levels, etc. |

Table A.1

| Entity Types   | Examples                                                                 | Frequency |
|----------------|--------------------------------------------------------------------------|-----------|
| Region         | China, Italy, globe, temperate region, etc.                              | 2783      |
| Indicator      | R0, CFR, infected cases, fatality cases, etc.                            | 2313      |
| Method         | SEIR, machine learning, etc.                                            | 1430      |
| Interventions  | lockdown, isolation, quarantine, etc.                                    | 1351      |
| Measure        | 0.98, 2.34, greater, etc.                                                | 1191      |
| Time           | 4 April 2020, 12 days, etc.                                              | 901       |
| Variable       | temperature, climate factors, mobility, etc.                            | 376       |
| Generic        | it, that, this, model, etc.                                              | 256       |
| Other Term     | polynomial growth, pre-pandemic levels, etc.                             | 94        |

Method: Methods, models, systems to use, or tools, components of a system, frameworks.

Intervention: Public health control policies to mitigate COVID-19 epidemic, especially, non-pharmaceutical interventions.

Measure: Numbers or measure words that can quantify indicators or interventions etc.

Task: Goals, problems, or projects to solve, mostly are phrases.

Time: Date, durations, or time-related words

Variable: Factors that associated with COVID-19 epidemic

Generic: General words that usually used to connect coreference entities.

Other Term: Public health related terms that do not belong to entity types listed ahead.

A.2. Relations

Table A.2 shows examples of relations annotated in CPHIE dataset.

USED-FOR: Associates any type entities to a Method entity that indicates solutions.

EQUAL-TO: This relation marks the Measure entity that the Indicator entity equals.

ASSOCIATE-WITH: Links an Intervention or a Variable entity to a Task or a Method entity, marks the existing relationship of the entities.

HAPPEN-IN: Associates entities to a Time entity that represents the entity takes place.

LOCATED-IN: A link that associates a Region entity to an entity. It indicates that the Region entity is the site where the entity is performed.
Table A.2
Relations annotated in CPHIE dataset.

| Relations         | Labels                                         | Examples |
|-------------------|-----------------------------------------------|----------|
| USED-FOR          | Method→Indicator|Task|Intervention       |                      |
| EQUAL-TO          | Indicator|Variable→Measure                          |                      |
| ASSOCIATE-WITH    | Intervention→Task|Indicator|Variable          |                      |
| HAPPEN-IN         | Indicator|Measure|Task|Variable→Time     |                      |
| LOCATED-IN        | Indicator|Variable|Intervention|Task|Measure→Region   |                      |
| CONJUNCTION       | Entity→Entity                                   |                      |
| PART-OF           | Method|Intervention|Variable→Method|Intervention|Variable |                      |
| COMPARED-TO       | Intervention|Method→Intervention|Method            |                      |
| FEATURE-OF        | OtherTerm→Task|Method|Intervention|Variable           |                      |
| EVALUATE-FOR      | Indicator→Task|Intervention|Method           |                      |

CONJUNCTION: Links two coordinate entities that usually are linked by words and, or belong to the same entity type.

PART-OF: This relation marks an entity includes another entity.

COMPARED-TO: Compares two entities that usually belong to the same entity type.

FEATURE-OF: Characteristics associated with a Method, Task, Variable, or Indicator entity.

EVALUATE-FOR: A link that associated an Indicator entity to a Task entity or an Entity. The former indicates that the Indicator entity is associated with COVID-19 epidemic, the latter indicates that the Indicator entity is the evaluation metric of the Method entity.

A.3. Coreference

Coreference: A link that associates two entities refer to the same entity.

A.4. Annotation examples

Fig. A.1 shows an example of annotations in CPHIE dataset.

Fig. A.1. Example sentences from CPHIE dataset as shown in the BRAT annotation interface.

References

[1] J.M. Brauner, et al., Inferring the effectiveness of government interventions against COVID-19, Science 371 (6531) (2021).
[2] L. Li, et al., Real-world data medical knowledge graph: construction and applications, Artif. Intell. Med. 103 (2020) 101817.
[3] L. Chen, M. Agrawal, S. Horng, D. Sontag, Robustly extracting medical knowledge from EHRs: a case study of learning a health knowledge graph, Pac. Symp. Biocomput. 25 (2020) 19–30.
[4] H. Sun, et al., Medical knowledge graph to enhance fraud, waste, and abuse detection on claim data: model development and performance evaluation, JMIR Med. Inform. 8 (2020).
[5] M. Sheng, et al., DEKGB: an extensible framework for health knowledge graph, ICHS (2019).
[6] Z. Huang, J. Yang, F.V. Harmelen, Q. Hu, Constructing disease-centric knowledge graphs: a case study for depression (short version), AIME (2017).
[7] S. Finlayson, P. LePendu, N. Shah, Building the graph of medicine from millions of clinical narratives, Sci. Data 1 (2014).
[8] Y. Shen, et al., CBN: constructing a clinical Bayesian network based on data from the electronic medical record, J. Biomed. Inform. 88 (2018) 1–10.
[9] C. Zhao, J. Jiang, Y. Guan, X. Guo, B. He, EMR-based medical knowledge representation and inference via Markov random fields and distributed representation learning, Artif. Intell. Med. 87 (2018) 49–59.
[10] M. Rotmensz, Y. Halpern, A. Tilman, S. Horng, D. Sontag, Learning a health knowledge graph from electronic medical records, Sci. Rep. 7 (2017).
Y. Yang, Z. Cao, P. Zhao et al.

Journal of Safety Science and Resilience 2 (2021) 146–156

[11] C. Wise, et al., COVID-19 knowledge graph: accelerating information retrieval and discovery for scientific literature, in: Proceedings of Knowledgeable NLP: the First Workshop on Integrating Structured Knowledge and Neural Networks for NLP, 2020, pp. 1–10.

[12] J. Reese, et al., KG-COVID-19: a framework to produce customized knowledge graphs for COVID-19 response, Patterns 2 (2020).

[13] D. Domingo-Fernández, et al., COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology, Bioinformatics (2020).

[14] R. Zhang, D. Hristovski, D. Schutte, A. Kastrin, M. Fisman, H. Kilicoglu, Drug repurposing for COVID-19 via knowledge graph completion, J. Biomed. Inform. (2021) 103696.

[15] Y. Zhou, Y. Hou, J. Shen, Y. Huang, W.R. Martin, F. Cheng, Network-based drug repurposing for novel coronavirus 2019-nCov/SARS-CoV-2, Cell. Discov. 6 (2020).

[16] Q. Wang, et al., COVID-19 literature knowledge graph construction and drug repurposing report generation, NAACL (2021).

[17] T. Hope, et al., Extracting a knowledge base of mechanisms from COVID-19 papers, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 4488–4503.

[18] L.L. Wang, et al., CORD-19: the COVID-19 open research dataset, in: Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020, 2020.

[19] Y. Luan, L. He, M. Ostendorf, H. Hajishirzi, Multi-task identification of entities, relations, and conferrure for scientific knowledge graph construction, EMNLP, 2018.

[20] Y. Luan, D. Wadden, L. He, A. Shah, M. Ostendorf, H. Hajishirzi, A general framework for information extraction using dynamic span graphs, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, 2019, pp. 3036–3046, (Long and Short Papers).

[21] D. Wadden, U. Wernberg, Y. Luan, H. Hajishirzi, Entity, relation, and event extraction with contextualized span representations, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 5788–5793.

[22] T. Jiang, T. Zhao, B. Qin, T. Liu, N. Chawla, M. Jiang, The role of “condition”: a novel scientific knowledge graph representation and construction model, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019.

[23] N. Stylianou, G. Ratz, D.G. Goulis, I. Vlahavas, EBDM+: advancing evidence-based medicine via two level automatic identification of populations, outcomes in medical literature, Artif. Intell. Med. 108 (2020) 101949.

[24] T. Kang, S. Zou, C. Weng, Pretraining to recognize PIDO elements from randomized controlled trial literature, Stud. Health Technol. Inform. 264 (2019) 188.

[25] B. Nye, et al., A corpus with multi-level annotations of patients, interventions and outcomes to support language processing for medical literature, in: Proceedings of the Conference. Association for Computational Linguistics. Meeting, 2018, 2019, p. 197. NIH Public Access.

[26] M. Zabinger, M. Sabou, S. Hofstätter, A. Hanbury, Effective crowd-annotation of participants, interventions, and outcomes in the text of clinical trial reports, EMNLP, 2020.

[27] R. Pitman, et al., Dynamic transmission modeling: a report of the ISPOR-SMDM modeling good research practices task force-5, Value Health 15 (2012) 828–834.

[28] B. Nussbaum-Streit, et al., Quarantine alone or in combination with other public health measures to control COVID-19: a rapid review, Cochrane. Database. Syst. Rev. 4 (2020) CD013574.

[29] P. Stenetorp, S. Prysalo, G. Topić, T. Ohta, S. Ananiadou, J.I. Taijii, BRAT: a web-based tool for NLP-assisted text annotation, in: Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, 2012, pp. 102–107.

[30] M. Joshi, D. Chen, Y. Liu, D.S. Weld, L. Zettlemoyer, O. Levy, Spanbert: improving pre-training by representing and predicting spans, Trans. Amoc. Comput. Linguist. 8 (2020) 64–77.

[31] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, 2019, pp. 4171–4186, Volume(Long and Short Papers).

[32] S. Abney, Bootstrapping, in: Proceedings of the 40th annual meeting of the Association for Computational Linguistics, 2002, pp. 360–367.

[33] E. Agichtein, L. Gravano, Snowball: extracting relations from large plain-text collections, in: Proceedings of the fifth ACM conference on Digital libraries, 2000, pp. 85–94.

[34] S. Brin, Extracting patterns and relations from the world wide web, in: International Workshop on the World Wide Web and Databases, Springer, 1998, pp. 172–183.

[35] Q. Ding, B. Zhou, Z. Feng, H. Peng, D. Roth, CogCompTime: a tool for understanding time in natural language, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2018, pp. 72–77.

[36] I. Beltagy, K. Lo, A. Cohan, SciBERT: a pretrained language model for scientific text, EMNLP/IJCNLP, 2019.

[37] S. Gururangan, et al., Don’t stop pretraining: adapt language models to domains and tasks, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 8342–8360.

[38] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: a next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019.

[39] P. Bragge, O. Clavisi, T. Turner, E. Tavender, A. Collie, R.L. Gruen, The global evidence mapping initiative: scoping research in broad topic areas, BMC Med. Res. Method. 11 (1) (2011) 1–12.