Deep Learning for Structuring Epidemiological Reports and Temporal Event Extraction: A TBNN Model Study

liang zhang, bing zhang, Ning Li, Yanru Chu, Xiaokang Jiao, Zengtao Jiao, yi chen

Submitted to: Journal of Medical Internet Research
on: March 06, 2024

Table of Contents

Deep Learning for Structuring Epidemiological Reports and Temporal Event Extraction: A TBNN Model Study

liang zhang[1]; bing zhang[2]; Ning Li[1]; Yanru Chu[3]; Xiaokang Jiao[4]; Zengtao Jiao[4]; yi chen[3]

[1]Division of Big Data Ningbo Center for Disease Control and Prevention Ningbo CN
[2]School of Public Health, Faculty of Medicine Ningbo University Ningbo CN
[3]Division of Acute Communicable Disease Control Ningbo Center for Disease Control and Prevention Ningbo CN
[4]Yidu Cloud (Beijing) Technology Co., Ltd. Beijing CN

Corresponding Author:
yi chen
Division of Acute Communicable Disease Control
Ningbo Center for Disease Control and Prevention
1166 Fan Jiangan Road
Haishu District
Ningbo
CN

Abstract

Background: The ongoing global health emergency has brought to the fore the need for quick and precise analysis of epidemiological investigations, as traditional methods have proven to be inadequate and limited in their ability to extract crucial information. To address this challenge, this study introduces an innovative deep learning approach: a Time-Based Neural Network (TBNN), which aims to efficiently extract and organize important data from epidemiological investigation reports.

Objective: To streamline and enhance the process of extracting essential information from epidemiological investigation reports, this study focuses on patient event data. The proposed TBNN model integrates temporal-aware deep learning algorithms to improve the efficiency of information processing. The anticipated outcome is a significant enhancement in the efficacy of epidemic response mechanisms.

Methods: The Time-Based Neural Network (TBNN) method employs a synergistic approach by integrating a pre-trained Bidirectional Encoder Representations from Transformers (BERT) language model with a Bi-directional Long Short-Term Memory (Bi-LSTM) network for information extraction. This technique involves the generation of time-enhanced textual representations, targeted extraction of information, and the assimilation of temporal dynamics, offering a holistic strategy for processing unstructured data in epidemiological investigation reports. To ensure robustness, the model training and validation are conducted using the ECR-COVID-19 dataset.

Results: Experimental results demonstrate that TBNN outperforms existing models in event extraction, specifically in terms of precision, recall, and F1 scores. The model's operational efficiency is highlighted by its capability to process reports efficiently on a commonly available 4-core Intel i5-class CPU, underscoring its practical feasibility for deployment in real-world epidemiological investigations.

Conclusions: The TBNN model offers a promising solution for refining data extraction and structuring within epidemiological investigation reports, providing a more efficient and accurate method for information extraction. The performance and efficiency of the model underscore its potential use in improving the overall capacity of epidemic responses during large-scale outbreaks.

(JMIR Preprints 06/03/2024:58135)
DOI: https://doi.org/10.2196/preprints.58135

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users. Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Deep Learning for Structuring Epidemiological Reports and Temporal Event Extraction: A TBNN Model Study

Liang Zhang[1*],BSc; Bing Zhang[4*],BA; Ning Li[1],MPH; Yanru Chu[2],PhD; Xiaokang Jiao[3],M.S; Zengtao Jiao[3],M.Sc; Yi Chen[2],MPH

[1] Division of Big Data, Ningbo Center for Disease Control and Prevention, Ningbo 315010, China

[2] Division of Acute Communicable Disease Control Ningbo Center for Disease Control and Prevention, Ningbo 315010, China

[3] Yidu Cloud (Beijing) Technology Co., Ltd.

[4] School of Public Health, Faculty of Medicine, Ningbo University

[*]These authors contributed equally

Corresponding Author:
Yi Chen, MPH
Division of Acute Communicable Disease Control
Ningbo Center for Disease Control and Prevention
1166 Fan Jiangan Road
Haishu District
Ningbo, 315010
China
Phone: 86 13967816045
Email: 30279068@qq.com

Abstract

Background: The ongoing global health emergency has brought to the fore the need for quick and precise analysis of epidemiological investigations, as traditional methods have proven to be inadequate and limited in their ability to extract crucial information. To address this challenge, this study introduces an innovative deep learning approach: a Time-Based Neural Network (TBNN), which aims to efficiently extract and organize important data from epidemiological investigation reports.

Objectives: To streamline and enhance the process of extracting essential information from epidemiological investigation reports, this study focuses on patient event data. The proposed TBNN model integrates temporal-aware deep learning algorithms to improve the efficiency of information processing. The anticipated outcome is a significant enhancement in the efficacy of epidemic response mechanisms.

Methods: The Time-Based Neural Network (TBNN) method employs a synergistic approach by integrating a pre-trained Bidirectional Encoder Representations from Transformers (BERT) language model with a Bi-directional Long Short-Term Memory (Bi-LSTM) network for information extraction. This technique involves the generation of time-enhanced textual representations, targeted extraction of information, and the assimilation of temporal dynamics, offering a holistic strategy for processing unstructured data in epidemiological investigation reports. To ensure robustness, the model training and validation are conducted using the ECR-COVID-19 dataset.

Results: Experimental results demonstrate that TBNN outperforms existing models in event extraction, specifically in terms of precision, recall, and F1 scores. The model's operational efficiency is highlighted by its capability to process reports efficiently on a commonly available 4-core Intel i5-class CPU, underscoring its practical feasibility for deployment in real-world

epidemiological investigations.
Conclusions: The TBNN model offers a promising solution for refining data extraction and structuring within epidemiological investigation reports, providing a more efficient and accurate method for information extraction. The performance and efficiency of the model underscore its potential use in improving the overall capacity of epidemic responses during large-scale outbreaks.

Keywords TBNN model; NER; information extraction; Bi-LSTM

Introduction

Since its emergence in late 2019, Corona Virus Disease 2019 (COVID-19) has exerted profound global impacts, underscoring the pivotal role of epidemiological investigations in tracing patient histories and managing the COVID-19 crisis. The swift transmissibility of the virus necessitates a large amount of manpower to analyze and systematize the data in epidemiological reports during an outbreak, which highlighted an urgent need for automation that promises accuracy and celerity. With the advancement of Natural Language Processing (NLP) technology, extracting vital information from unstructured texts has become more convenient. Typically, epidemiological reports contain basic patient information, social contact details, and recent events the patient has attended. Therefore, the structuring of epidemiological report texts can be categorized as a Named Entity Recognition, Relation Extraction, and Event Extraction problem within NLP. Given that the description of various events in these reports cannot be uniformly standardized, the effectiveness of patient event extraction has become a focal point in the structuring process of epidemiological reports.

Epidemiological investigation reports typically comprise textual records detailing fundamental information about individuals under investigation, which may include both confirmed infectious disease cases and those deemed at high risk. Such information encompasses the onset of illness, clinical visits, and the individual's social connections. These documents are vital for the Centers for Disease Control and Prevention (CDC) personnel to analyze the status of an outbreak and predict its development. However, these reports primarily utilize unstructured forms such as text and images for data recording and description, their interpretation is entirely dependent on the analyst's recall and experience. This dependence can result in inefficient data handling, gaps in data linkages, and the possibility of missing out on crucial insights. During widespread infectious disease outbreaks, CDC experts need to swiftly collect data and compile multiple epidemiological reports. Concurrently, they are responsible for synthesizing information from these reports to reveal patterns in patient contact networks and track disease transmission routes. Therefore, the application of Artificial Intelligence (AI) in interpreting and organizing these reports is paramount. AI facilitates the transformation of unstructured data into adaptable and analyzable formats, which is indispensable for constructing precise models that capture the temporal, personal, and spatial dynamics of an epidemic. Such models are crucial for gaining insights into the outbreak, enhancing data analysis, and improving the efficacy of emergency responses.

Named Entity Recognition (NER) is a classical task in information extraction, having evolved from the early Template Matching method to traditional machine learning such as Hidden Markov Model (HMM) and Conditional Random Field (CRF), and subsequently to the utilization of pre-trained models within deep learning frameworks. The pre-trained Bidirectional Encoder Representations from Transformer (BERT)[1] model, initially proposed by Google, is founded upon the Transformer[2] model, which boasts significantly accelerated computational speeds in comparison to Recurrent Neural Network (RNN). Owing to their superior performance, pre-trained models have become fundamental components in Natural Language Processing (NLP). Classical NER algorithms include FLAT[3], SoftLexicon[4], LEBERT[5], and RICON[6].

Alongside Named Entity Recognition (NER), relationship extraction also stands as a crucial component of information extraction. Typically, relationship extraction is performed together with named entities to form a joint extraction model. This model capitalizes on the interplay between entities and their interrelations to boost overall extraction capabilities. Notable classical joint extraction algorithms include CasRel[7], TPLinker[8], PRGC[9], PURE[10], and OneRel[11].

Event extraction poses a greater challenge when contrasted with NER and relation extraction. It involves a complex, four-step process: identifying event trigger words, determining event types, recognizing event elements, and determining the types of these elements. These stages proceed sequentially, yet their disconnection can lead to a cascade of errors due to the absence of integral linkages between them. Classical event extraction algorithms include dbRNN[12], JMEE[13], RCEE[14], and so on.

Methods

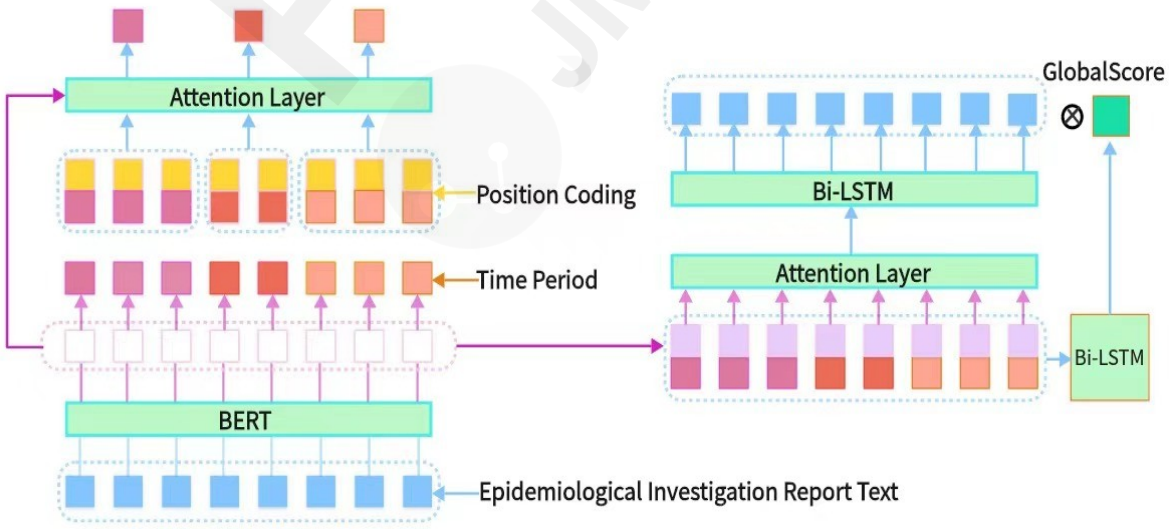
Overview of methods

The information extraction methods in the introduction have been applied across various standard datasets and are not tailored for the unique unstructured text found in epidemiological investigation reports. Consequently, this study proposes a novel information extraction approach, specifically designed to structure, and distill information from distinctive features of epidemiological investigation reports.

Epidemiological investigation reports typically begin with the patient's basic information and employ timestamps as keywords, delineating the patient's movements and close contacts over various time periods. The patient information is often the most standardized segment of text within these reports. It can be processed through simple regular expressions or may even be structured at the initial stage of data entry through forms. Considering the distinctive characteristics of the text in epidemiological reports, this paper focuses on addressing the challenge of event extraction within the structuring process. Drawing on the temporal attributes of events, we propose an extraction method centered on temporal information, named the Time-Based Neural Network (TBNN).

Modelling Methodology and Framework

Figure 1. The basic structure of the TBNN model



Obtain Text Input Vector

Figure 1 illustrates the initial step: inputting the original epidemiological investigation report text into the pre-trained BERT language model. After extensive pre-training on a large-scale corpus, BERT yields a precise vectorized representation of the text without the need for annotation.

$$T = BERT(Text) \quad (1)$$

Obtain Time-enhanced and Location-encoded Text Representations

In the second step, the correlation between temporal information and events is taken into account. The model introduces positional encoding P to the text of various temporal passages T_{ti}, and then conducts a pooling operation for feature dimensionality reduction, resulting in a query vector q.

$$T = [T_{t1}, T_{t2}, \cdots, T_{tn}] \quad (2)$$

$$q_{ti} = pooling(T_{ti} + P), i = 1,2,\cdots,n \quad (3)$$

$$Q = [q_{t1}, q_{t2}, \cdots, q_{tn}] \quad (4)$$

Extract Focused Text Information

In the third step, an Attention Mechanism is employed to extract information that is of greater significance across various time periods.

$$K = W_k T \quad (5)$$

$$V = W_v T \quad (6)$$

$$a_{ti} = \frac{q_{ti} K^T}{\sqrt{d}} V, i = 1,2,\cdots,n \quad (7)$$

Integrate Information

In the fourth step, the features extracted by the Attention Mechanism are combined with the corresponding encoded text, which serves as input for subsequent information extraction.

$$A = [a_{t1}, a_{t2}, \cdots, a_{tn}] \quad (8)$$

$$input = concat(T, A) \quad (9)$$

Refine Focused Information After Splicing

In the fifth step, the processed input information is subjected to the self-attention operation so that the feature vector can focus more on the text features that are more relevant to the current location features.

$$Q = W_q input \quad (10)$$

$$K = W_k input \quad (11)$$

$$V = W_v input \quad (12)$$

$$hidden_{attenion} = \frac{QK^T}{\sqrt{d}} V \quad (13)$$

Event Reorganization

The sixth step identifies features through a Bi-directional Long Short-Term Memory (Bi-LSTM) model, which filters events in the text involving the patient.

$$pos = BiLSTM(hidden_{attention}) \quad (14)$$

In addition, to make the model global, a global score GlobalScore is derived for the input features to fine-tune the output probability distribution of the model and improve the final recognition performance.

$$globalscore = BiLSTM(input) \quad (15)$$

$$pos = globalscore \times pos \quad (16)$$

Summary of the Model

The model proposed in this study comprises two main components. The first part is dedicated to the integration of word vectors, which are encoded by a pre-trained model and refined to take into account the unique structure and textual nuances of epidemiological reports, of which the temporal aspect is a key dimension. This integration is enhanced with positional encoding, enriching each word vector with relevant context, thereby paving the way for more effective information extraction. The second part, the information extraction model, employs the fused feature vectors and utilizes the Bi-directional Long Short-Term Memory (Bi-LSTM) model as its construction method. The Bi-LSTM, a classic model frequently utilized in information extraction, is capable of leveraging feature information for the efficacious extraction of information.

Training the Model

In the model training process, the integration of the pre-trained language model BERT provides a robust initialization. Consequently, the model does not employ cross-validation for the determination of hyper-parameters. Rather, it depends on a pre-determined fixed validation set to affirm the hyper-parameters, and then the model undergoes adequate training with a setting of 6 epochs.

Data Set

To evaluate the performance of the proposed model, this study utilized the ECR-COVID-19 dataset[15]. This dataset comprises a total of 2264 original Chinese epidemiological investigation reports of confirmed cases from the National CDC and other mainstream media websites. The dataset encompasses various entities used for statistical and analytical purposes in actual epidemiological tasks and includes annotations for corresponding entity locations. Regarding the proposed event extraction component in this study, the dataset includes eight categories of events and seven types of event tuples. For model training and validation, according to segment of the ECR-COVID-19 dataset in the original corresponding researches, the training set contains 1811 reports, the test set contains 227 reports, and the validation set contains 226 reports.

More specifically, the event information within this dataset can be summarized as follows: hospital visits, confirmations of diagnosis, observations in quarantine, deaths, modes of travel, and other detailed information. These data aid CDC professionals in determining the recent trajectory of the patient and the timeline of their diagnosis. Subsequently, NER-related techniques are applied for the identification of various event entities, which are then used in subsequent epidemiological investigation analysis and decision-making.

The specific data structure is shown in Table 1.

Table 1. ECR-COVID-19 Dataset Summary

Type of information	Number of entities	Entity	
Event	8types	Event	Verb entities corresponding to the various types of activities produced by the patient
		Onset	Verb entities corresponding to disease onset events
		Hospital Visit	Verb entity corresponding to the event of the patient's visit to the doctor

Extraction	7tuples	Diagnosis Confirmed	Word entities corresponding to patient diagnostic events
		Inpatient	Verb entities corresponding to patient admission events
		Discharge	Verb entities corresponding to patient discharge events
		Death	Verb entities corresponding to patient death events
		Observed	The patient was observed as a suspected case for the verb entity corresponding to the event
		Date	Time or date entity
		Symptom	Describe the entity of the symptom
		Lab Test	Laboratory testing of relevant entities
		Imaging Examination	Image examination of relevant entities
		Location	Provincial-level location entities
		Spot	Location entities such as guesthouses, hotels, homes, etc.
		Vehicle	Vehicle entities such as trains, cars, etc.

Results

Performance on Test Set

The TBNN model proposed in this paper is validated using the ECR-COVID-19 dataset described above, and the final experimental results are shown in Table 2 in comparison with the results of other methods.

Table 2. Comparison of information extraction results

Model method	Precision rate	Recall rate	F1
TMT-NN[15]	77.6	80.0	78.8
Bi-LSTM+CRF	75.1	84.2	79.6
TBNN	84.1	83.0	83.6

The experiments demonstrate that the model shows better performance on the epidemiological investigation report dataset, and can accurately identify most of the key information.

Compared to the TMT-NN model and the classic Bi-LSTM+CRF model, the current model has achieved optimal results in both precision and F1 scores. The Bi-LSTM+CRF model has a slightly better recall rate, but its precision is significantly lower than that of the TBNN method presented in this paper. Precision refers to the proportion of events correctly identified by the model out of all the events it recognized, indicating the model's accuracy in identifying outcomes. Recall rate indicates the proportion of events the model identified out of all actual events, representing the model's capability to comprehensively screen for factual events. Precision and recall are mutually restrictive and generally inversely related. Hence, the introduction of the F1 score as a composite metric serves to balance the impacts of precision and recall, providing a more comprehensive evaluation of a model's extraction abilities.

Ablation Study

To validate the efficacy of the model structure proposed in this study, it is necessary to carry out ablation experiments to ascertain whether there is a significant difference when features are extracted in segments based on temporal information and concatenated with the original textual features, thereby verifying the rationality of the model structure. The results show that using temporal features to process the output of BERT enhances the extraction of textual information from the epidemiological investigation report, yielding a marked improvement in precision, recall, and F1 score, as shown in Table 3.

Table 3. Overall statistics of ablation study results

Model method	Precision rate	Recall rate	F1
TBNN	84.1	83.0	83.6
Ablation Study- No temporal feature processing	82.9	69.6	75.7

For entities directly associated with temporal elements—specifically symptoms, locations, and transportation entities—this paper independently assessed the impact of temporal feature processing on the extraction of these three categories of entities. The findings indicate that incorporating temporal features can significantly enhance the recall rate of certain entities, thereby noticeably improving the overall F1 score. See Table 4 for details.

Table 4. Ablation study for selected entities

Model method	Type of entity	Precision rate	Recall rate	F1
TBNN	location	83.2	79.0	81.0
	Symptom	91.1	82.7	86.7
	Vehicle	81.7	79.2	80.3
Ablation Study- No temporal feature processing	location	81.1	71.2	75.5
	Symptom	90.7	74.3	81.7
	Vehicle	82.5	70.7	76.2

Discussion

Extracting Key Information from Epidemiological Investigation Reports

Epidemiological investigations Report structuring can provide important value for obtaining key clues about the epidemic. Considering that in the actual epidemiological investigation and judgment process, the basic patient information has a standardized writing format or has been formed into structured data by filling out forms, while the various types of patient trip events are difficult to

extract directly, this model concentrates on structuring the extraction of patient event information and verifies the model's effectiveness by using a publicly available Chinese dataset of epidemiological investigation reports.

Comparison and Interpretation of Model Results

The experimental results reveal that the TBNN model surpasses the TMT-NN model with respect to the F1 score for event extraction, achieving a score above 0.8. Event extraction is typically the most critical aspect in the structuring of actual epidemiological reports. By extracting events and the contacts of patients, it is possible to infer the patients' movement trajectories and close-contact situations, which provide an essential foundation for subsequent analyses of risk areas, spatiotemporal patterns, and transmission chains. The demonstrated absolute and relative effectiveness of the TBNN method in the test dataset confirms its viability for structuring real-world epidemiological reports.

Rationalization of the Model Structure

To validate the impact of each component of the structure on the final information extraction results, this study conducted an ablation study on the temporal enhancement module. The study confirmed that enhancing textual features by time periods contributes positively to the overall model effectiveness, particularly in terms of the model's recall rate for extracting each entity.

Incorporation of Temporal Features

Events targeted for extraction have a temporal dimension, indicating a clear correlation between events and time. For example, a patient may visit a particular location repeatedly during similar time frames, or there may be a causal relationship between the timing of a patient's visits to various locations and the temporal evolution of their symptoms. The integration of temporal features provides practical interpretability. Thus, this model incorporates temporal features to augment information, diverging from conventional models that directly encode text. Empirical validation during ablation experiments has also showcased the practical effectiveness of introducing temporal features.

Entity Identification with GlobalScore Mechanism

The main structure of the model is based on the self-attention mechanism of the Bi-LSTM network, which is a commonly used approach in entity recognition tasks. The Bi-LSTM takes the coded representation of the text as input and outputs the event category corresponding to each character, achieving event annotation ability as verified in major datasets. To enhance the traditional event annotation model's consideration of global event information, this model introduces the concept of GlobalScore. This concept refines the boundary between different event entities, extending beyond the currently recognized event entity itself. It provides the model with a global concept in recognition.

Model Efficiency and Practical Applicability

Beyond the model's recognition capabilities, the efficiency of the structured output is a critical aspect of the practical epidemiological investigation process. It is widely acknowledged that the performance of deep learning experiments is directly linked to Central Processing Unit (CPU) capabilities. However, establishing a CPU environment and accessing CPU hardware can be challenging, given the varied circumstances of epidemiological fieldwork. To enhance the practicality of this model in epidemiological investigations, this study evaluates its efficiency based on the processing power of a commercially available 4-core Intel i5 CPU. The experiments demonstrate that the proposed TBNN model operates on a 4-core Intel i5 CPU with a batch size of

16, processing approximately 38 epidemiological investigation reports per minute. In contrast, manually structuring patient information and event details from an investigation report usually takes around 29 minutes. Additionally, secondary manual verification frequently reveals a significant number of incorrect annotations, especially when structuring information about individual trajectories and timelines. In comparison, the model clearly provides significantly better information extraction efficiency than human capabilities while also ensuring high accuracy. During times of epidemic crises, this can help relevant personnel extract patient movement trajectories and related event information more efficiently, conserving human resources.

Conclusions

This paper introduces a new method for extracting key information from epidemiological reports, the TBNN model, to optimize the use of limited resources and save manpower in the analysis of epidemiological reports. The experiments conducted indicate that this method has commendable information extraction capabilities, providing robust and effective support for the analysis of epidemiological reports. By streamlining the processing of information in these reports, the saved manpower can be redirected to more critical aspects of epidemic prevention. This could potentially enhance the capacity to respond to health crises.

During testing, the model proposed in this study was able to process approximately 38 epidemiological investigation reports per minute on a 4-core Intel i5-class CPU. This outcome indicates that the model outperforms humans in the volume of reports it can process per minute, even without reliance on costly high-performance computers. This reaffirms the model's ability to fully leverage the advantages of information extraction algorithms.

Moreover, the model presented in this paper requires the initial extraction of time-related entities. The accuracy of time entity recognition will impact the results of subsequent information extraction. Therefore, there is scope for optimizing and enhancing the model's performance. Future work could address the issue of error propagation in this pipeline approach and propose a more suitable model for structuring epidemiological reports based on their distinct characteristics.

Acknowledgments

This study was supported by the Zhejiang Province Basic Public Welfare Research Program Project (LGF20H260007) and Ningbo Top Medical and Health Research Program(No.2023020713）

Authors' Contributions

LZ and BZ were responsible for research design and paper writing and contributed equally to this work as co-first authors.XJ and ZJ were responsible for Modeling.NL and YanC were responsible for Modeling Tests and Applications. YiC was responsible for research design and dissertation quality control.

Conflicts of Interest

None declared.

References

1. Devlin J, Chang M-W, Lee K,et al.BERT: pretraining of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North. Association for Computational Linguistics; 2019.[doi: 10.18653/v1/n19-1423]

2. Vaswani A , Shazeer N , Parmar N ,et al.Attention Is All You Need. arXiv, 2017. [doi:10.48550/arXiv.1706.03762]

3. Li X, Yan H, Qiu X, Huang X. FLAT: Chinese NER Using Flat-Lattice Transformer. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics; 2020.[doi: 10.18653/v1/2020.acl-main.611]
4. Ma R, Peng M, Zhang Q, Wei Z, Huang X. Simplify the Usage of Lexicon in Chinese NER. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics; 2020.[doi: 10.18653/v1/2020.acl-main.528]
5. Liu W, Fu X, Zhang Y, Xiao W. Lexicon Enhanced Chinese Sequence Labeling Using BERT Adapter. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics; 2021.
[doi: 10.18653/v1/2021.acl-long.454]
6. Gu Y, Qu X, Wang Z, Zheng Y, Huai B, Yuan NJ. Delving Deep into Regularity: A Simple but Effective Method for Chinese Named Entity Recognition. Findings of the Association for Computational Linguistics: NAACL 2022. Association for Computational Linguistics; 2022.[doi: 10.18653/v1/2022.findings-naacl.143]
7. Wei Z, Su J, Wang Y, Tian Y, Chang Y. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics; 2020.[doi: 10.18653/v1/2020.acl-main.136]
8. Wang Y, Yu B, Zhang Y, Liu T, Zhu H, Sun L. TPLinker: Single-stage Joint Extraction of Entities and Relations Through Token Pair Linking. Proceedings of the 28th International Conference on Computational Linguistics. International Committee on Computational Linguistics; 2020.[doi: 10.18653/v1/2020.coling-main.138]
9. Zheng H, Wen R, Chen X, Yang Y, Zhang Y, Zhang Z, et al. PRGC: Potential Relation and Global Correspondence Based Joint Relational Triple Extraction. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics; 2021.
[doi: 10.18653/v1/2021.acl-long.486]
10. Zhong Z, Chen D. A Frustratingly Easy Approach for Entity and Relation Extraction. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics; 2021.[doi: 10.18653/v1/2021.naacl-main.5]
11. Shang Y M , Huang H , Mao X L . OneRel:Joint Entity and Relation Extraction with One Module in One Step. 2022.[doi:10.48550/arXiv.2203.05412]
12. Sha L, Qian F, Chang B, Sui Z. Jointly Extracting Event Triggers and Arguments by Dependency-Bridge RNN and Tensor-Based Argument Interaction. AAAI Association for the Advancement of Artificial Intelligence (AAAI); 2018 Apr 26;32(1).[doi: 10.1609/aaai.v32i1.12034]
13. Liu X, Luo Z, Huang H. Jointly Multiple Events Extraction via Attention-based Graph Information Aggregation. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics; 2018.[doi: 10.18653/v1/d18-1156]
14. Liu J, Chen Y, Liu K, Bi W, Liu X. Event Extraction as Machine Reading Comprehension. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics; 2020.[doi: 10.18653/v1/2020.emnlp-main.128]
15. Wang J, Wang K, Li J, Jiang J, Wang Y, Mei J, et al. Accelerating Epidemiological Investigation Analysis by Using NLP and Knowledge Reasoning: A Case Study on COVID-19. AMIA Annu Symp Proc. 2020;2020:1258-67.[Medline: 33936502]

Abbreviations

NER：Named Entity Recognition
Bi-LSTM：Bi-directional Long Short-Term Memory
COVID-19：Corona Virus Disease 2019
NLP: Natural Language Processing
CDC：Centers for Disease Control and Prevention
HMM: Hidden Markov Model
CRF: Conditional Random Field
RNN: Recurrent Neural Network
TBNN: Time-Based Neural Network
CPU: Central Processing Unit
BERT: Bidirectional Encoder Representations from Transformer