Triple-view Event Hierarchy Model for Biomedical Event Representation

Jiayi Huang[1], **Lishuang Li**[1*], **Xueyang Qin**[1], **Yi Xiang**[1], **Jiaqi Li**[2], **Yubo Feng**[1]
[1] Dalian University of Technology, Dalian, China
[2] China Foreign Affairs University, Beijing, China
jyhuang1207@163.com, lils@dlut.edu.cn, qinxueyang@snnu.edu.cn
xiangyi_job@163.com, ljiaqi@mail.cfau.edu.cn, argamx@126.com

Abstract

Biomedical event representation can be applied to various language tasks. A biomedical event often involves multiple biomedical entities and trigger words, and the event structure is complex. However, existing research on event representation mainly focuses on the general domain. If models from the general domain are directly transferred to biomedical event representation, the results may not be satisfactory. We argue that biomedical events can be divided into three hierarchies, each containing unique feature information. Therefore, we propose the Triple-views Event Hierarchy Model (TEHM) to enhance the quality of biomedical event representation. TEHM extracts feature information from three different views and integrates them. Specifically, due to the complexity of biomedical events, We propose the Trigger-aware Aggregator module to handle complex units within biomedical events. Additionally, we annotate two similarity task datasets in the biomedical domain using annotation standards from the general domain. Extensive experiments demonstrate that TEHM achieves state-of-the-art performance on biomedical similarity tasks and biomedical event casual relation extraction.

1 Introduction

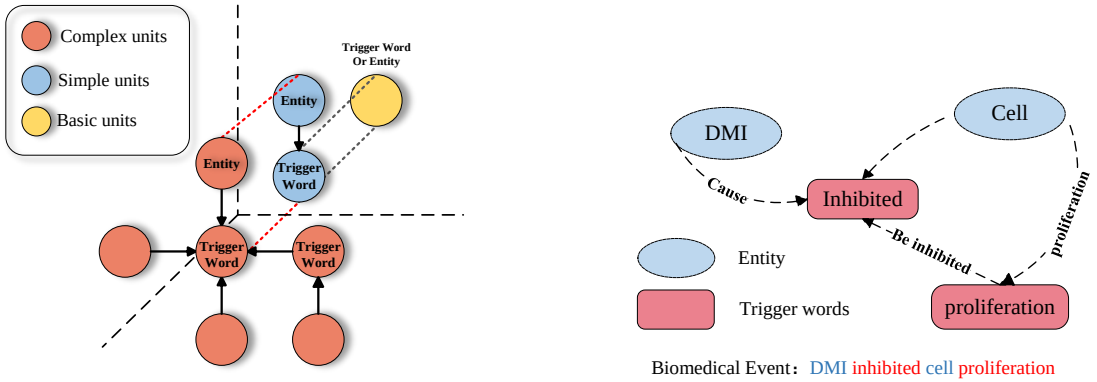
Biomedical events consist of trigger words, biomedical entities, and specific actions of entities (Frisoni et al., 2022). A complex biomedical event often contains multiple simple biomedical events, making the structure of event texts quite complex. Biomedical event representation refers to the process of converting biomedical event texts into machine-readable formats. Event representation is applied in tasks such as event prediction, event relation extraction, and event information extraction (Hwang et al., 2022), which can improve the performance of related tasks. Therefore, obtaining a representation of biomedical events is necessary.

Currently, in the general domain, the works on event representation focus on simple events, such as "He was frightened football" which typically follow a subject-verb-object structure. These works usually rely on pre-trained language models (PLMs) for text encoding (Gao et al., 2022). However, biomedical events involve multiple biomedical entities, and the event structure is not a simple subject-verb-object format. For example, in the biomedical event "cooperation among NF-kappa B-, AP-1- and NF-AT-binding sequences is required for induction of the GM-CSF gene through PKC- and Ca2+- signaling pathways downstream of T-cell activation" there are multiple verbs and the relationships between verbs are nested within each other (Weber et al., 2018). If we directly transfer models from the general domain for biomedical event representation, we cannot extract and integrate the structural features of biomedical events due to their complexity. Currently, existing works lack biomedical event representation models capable of handling complex structures.

We argue that although biomedical events are complex, they can be categorized into three hierarchies. As illustrated in Figure 1(a), we divide a complex biomedical event into complex units, simple units, and basic units. Hierarchically organizing biomedical events and deriving biomedical event representations from these three hierarchies is an effective approach. Therefore, based on this perspective, we

*Corresponding author: Lishuang Li(lils@dlut.edu.cn)

Proceedings of the 23rd China National Conference on Computational Linguistics, pages 1136-1147, Taiyuan, China, July 25 - 28, 2024.
Volume 1: Main Conference Papers



(a) Triple-views Event Hierarchy

(b) Complex units: Trigger-aware Graph

Figure 1: Figure 1(a) depicts the structure of a biomedical event. Trigger words and entities belong to the basic unit. A simple unit is an event that contains only one trigger word. A complex unit involves multiple trigger words. Figure 1(b) illustrates the event graph structure we constructed for processing complex units with graph neural networks, centered around trigger words.

propose the Triple-views Event Hierarchy Model (TEHM). The TEHM learns triple hierarchical features of biomedical events from three views. From the complex units view, we propose the Trigger-aware Aggregator module, which focuses on extracting features from complex units with the trigger word as the core. It is based on the architecture of graph neural networks and has been improved to better suit the characteristics of biomedical events for feature extraction. From the simple units view, we introduce the Context-aware Intensifier module, which integrates context information through prompt learning to explore implicit interaction features within simple events. From the basic units view, we present the Entity Relation Integrator module, which obtains feature encoding matrices for biomedical entities. This module constructs an entity relation network and employs the algorithm proposed in this paper BioER to train entity feature matrices.

After obtaining event representations, how to evaluate the quality of event representations is also a problem that needs to be addressed currently. Due to the significant differences in datasets between the general domain (which consists of simple events) and the biomedical domain, there is currently a notable lack of effective evaluation tasks to directly validate the quality of biomedical event representations. To address this issue, we adopt the same annotation standards (Weber et al., 2018; Kartsaklis and Sadrzadeh, 2014) as those used for the tasks in the general domain and manually annotate two validation tasks relevant to assessing the quality of biomedical event representations: Bio Hard Similarity Task and Bio Transtive Sentence Similarity Task. Additionally, we validate the effectiveness of biomedical event representation on standard biomedical event causal relation extraction datasets.

After extensive experiments, TEHM achieves improvements of 25.9% (accuracy) in Bio Hard Similarity and 0.045 (Spearman correlation coefficient) in Bio Transitive Sentence Similarity. Additionally, it improves 3.0% in the F1 score on Hahn Powell's Dataset for biomedical event causal relation extraction, reaching the current SOTA.

In summary, the contributions of this paper can be summarized as follows:

- To the best of our knowledge, we are the first to propose the task of biomedical event representation. We partition complex biomedical events into three hierarchies and propose the Triple-views Event Hierarchy Model (TEHM) for biomedical event representation. TEHM extracts features from three views, integrates them, and produces the final representation of biomedical events.

- Based on the characteristics of complex units in biomedical events, we propose a Trigger-aware Aggregator module with the trigger word as the core. This module is designed to explore the relational features within complex biomedical events.

- We construct a dataset for evaluating the quality of biomedical event representation. We will make it publicly available, laying the foundation for further research. Simultaneously, we achieve SOTA in similarity tasks and biomedical event relation causality extraction.

2 Triple-view Event Hierarchy Model(TEHM)

Figure 2 illustrates the framework of the TEHM model. The input to the TEHM model is the event text, which is simultaneously processed by the feature extraction modules of TEHM to obtain features from three different views of the event. These feature vectors are then concatenated to obtain the representation of the final biomedical event. TEHM employs a contrastive learning approach to training the model, which is a form of unsupervised learning.

2.1 Problem definition

Event Representation. Our goal is to enable the model to learn features of biomedical events, obtain embeddings for these events, and subsequently investigate the model's event representation capability. Specifically, given a similar event pairs $\left\{v_i^+, v_j^+\right\}$ and a dissimilar event pairs $\left\{v_i^-, v_j^-\right\}$, along with a similarity evaluation function S, where v refers to event representation. When evaluated using the function S, the great event representations should satisfy equation 1:

$$S\left\{v_i^+, v_j^+\right\} > S\left\{v_i^-, v_j^-\right\} \tag{1}$$

Biomedical Event Hierarchy. As shown in Figure 1(a), this paper defines three hierarchical units for biomedical events: basic units (which refer to **trigger words or entity elements**), simple units (which refer to **one simple event** consisting of a trigger word and entity elements), and complex units (which refer to a complex event consisting of **two or more simple events**)

2.2 Trigger-aware Aggregator (TAA)

For the complex units, we model it as a Trigger-aware graph (Figure 1 (b)), utilizing the structured information from the bioT2E dataset (Frisoni et al., 2022). We represent the Trigger-aware graph as: $\mathcal{G}_i = \{\nu_i, \varepsilon_i\}$, where ν_i, ε_i represents a set of nodes and edges for $\mathcal{G}_i$ respectively. A set of n distinct trigger words and entities is represented as $\nu_i = \{w_{i,1}, w_{i,2}, ..., w_{i,n-1}, w_{i,n}\}$. $\varepsilon_i = \{l_{i,1}, l_{i,2}, ..., l_{i,m-1}, l_{i,m}\}$ represents a set of m directed edges. Based on the structural information, we can directly establish edges between nodes in the Trigger-aware graph. The motivation behind our proposed Trigger-aware graph structure is as follows: for complex units, as the complexity of event texts increases and the number of entities and trigger words rises, directly modeling the event texts is not conducive to extracting implicit entity relationship features. We consider that different biomedical entities relate to each other due to trigger words. Therefore, using the trigger words as central hubs can uncover hidden relationships among different events that are not directly related.

We initialize the representation for nodes using BERT. Given a node $\nu_{i,n}$ from the event graph $\mathcal{G}_i$, we employ the Dropout technique by feeding the same input $\nu_{i,j}$ into BERT twice, obtaining $\vec{h}_{i,j}$ and $\vec{h}_{i,j}^+$, where $\vec{h}_{i,j}, \vec{h}_{i,j}^+ \in R^F$, and F represents the feature dimension of the node input. It is important to note that different Dropout masks are applied for each of the two inputs:

$$\vec{h}_{i,j} = f_\theta\left(w_{i,j}, \phi_1\right), \vec{h}_{i,j}^+ = f_\theta\left(w_{i,j}, \phi_2\right), \tag{2}$$

where ϕ_1 and ϕ_2 represent two independent random masks, and f_θ represents the BERT encoding. We can obtain different vector representations for the same graph node by using different masks.

The subsequent steps involve facilitating information exchanges between nodes by utilizing the central node. For the Trigger-aware graph, we input a set of node features $h_i = \left\{\vec{h}_{i,1}, \vec{h}_{i,2}, ..., \vec{h}_{i,n}\right\}$, where n represents the number of nodes in the event graph. To enhance the importance of the central node, we incorporate the dimension information of the node itself in the aggregation layer of the graph. As illustrated

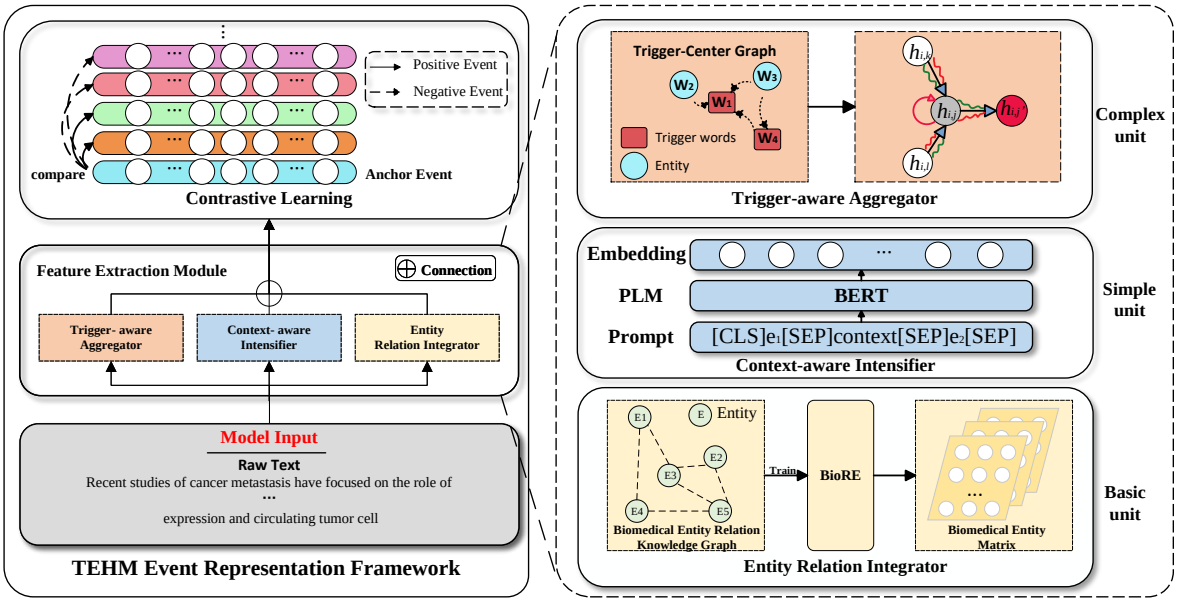


Figure 2: The overall framework of the TEHM model (left) and details the processing of biomedical events from triple views (right). After unsupervised learning, we obtain a well-trained TEHM model. The model's output serves as the representation of biomedical events.

in Figure 2, when $\vec{h}_{i,j}$ passes through the graph attention layer, it produces $h_i^{'} = \left\{\vec{h}_{i,1}^{'}, \vec{h}_{i,2}^{'}, ...\vec{h}_{i,n}^{'}\right\}$ where $\vec{h}_{i,j}^{'} \in R^{F'}$, and $F^{'}$ represents the feature dimension of the node output:

$$\vec{h}_{i,j}^{'} = sigmoid\left(\sum_{k \in N_i} \alpha_{ik} W_h \vec{h}_{i,k} + W_h \vec{h}_{i,j}\right), \tag{3}$$

where, N_{ij} represents the neighborhood formed by all nodes adjacent to $w_{i,j}$, and W_h is the feature matrix with dimensions $F \times F^{'}$. α_{ik} represents the attention coefficient for information exchange between node $h_{i,j}$ and node $h_{i,k}$, indicating the importance of node $h_{i,k}$ corresponding to node $h_{i,j}$. The calculation formula in this paper is as follows:

$$\alpha_{ik} = softmax\left(atten_{ik}\right) = \frac{exp\left(atten_{ik}\right)}{\sum_{j \in N_{ij}} exp\left(atten_{ik}\right)}, \tag{4}$$

where $atten_{ij}$ represents attention coefficient. Its calculation formula is shown as Equation 5:

$$atten_{ik} = \frac{exp\left\{\sigma\left(\vec{o}^T\left[W_a \vec{h}_{i,j} \oplus W_a \vec{h}_{i,k}\right]\right)\right\}}{\sum_{k \in N_i} exp\left\{\sigma\left(\vec{o}^T\left[W_a \vec{h}_{i,j} \oplus W_a \vec{h}_{i,k}\right]\right)\right\}}, \tag{5}$$

where $\oplus$ represents the concatenation of two matrices on both sides, σ represents LeakyRelu avtivate function, $\vec{o}$ is a weight vector with dimensions $\vec{o} \in R^{2F^{'}}$, W_a is the training weight matrix. The multiplication operation between them results in a scalar value. The calculation yields $atten_{ik}$. Substituting it into equations 4 and 3 obtains the enhanced node representation after TAA.

Finally, we convert the node representations into integrated vectors $\vec{y}_i$ (likewise, using $\vec{h}_{i,j}^{+}$ as input into Equations (3 - 6) yields $\vec{y}_i^{+}$) the equation as follow:

$$\vec{y}_i = \frac{1}{n} \sum_{j \in h_i^{'}} \left[\vec{h}_{i,j}^{'}\right]^T, \qquad \vec{y}_i^{+} = \frac{1}{n} \sum_{j \in h_i^{'+}} \left[\vec{h}_{i,j}^{'+}\right]^T, \tag{6}$$

Proceedings of the 23rd China National Conference on Computational Linguistics, pages 1136-1147, Taiyuan, China, July 25 - 28, 2024.
Volume 1: Main Conference Papers

where, n denotes the nodes number of $\mathcal{G}_i$.

2.3 Context-aware Intensifier (CAI)

This part of the model is designed to handle simple units of biomedical events. We find that when people need to understand an event, they usually need to combine the context to grasp the meaning conveyed by the event. The context becomes even more important when we need to do reasoning. Therefore, injecting contextual information allows the model to uncover implicit event relationships and extract features of event relationships. This is also why TEHM can be applied to event relationship extraction. Additionally, annotation in biomedical event relationship data is expensive and challenging. However, our model is trained in unsupervised paradigms, addressing the annotation issue.

We place the contextual information between the head event e_1 and the tail event e_2. This effectively informs the model about two events along with the contextual information between them, enabling the model to explore their implicit relationships. We construct the Prompt as shown in the formula:

$$Prompt : [CLS]\, e_1\, [SEP]\, context\, [SEP]\, e_2\, [SEP]\,, \tag{7}$$

This Prompt helps the model learn event relationship information, enhances the features of the events themselves, and strengthens the capability of TEHM for event relationship extraction.

To obtain representation and its positive example, we also employ the Dropout technique to input the above template p_i into BERT twice:

$$\vec{z}_i = f_\theta\left(p_i, \phi_1\right), \vec{z}_i^{+} = f_\theta\left(p_i, \phi_2\right). \tag{8}$$

2.4 Entity Relation Integrator (ELI)

Algorithm 1 BioER

Input: Training set $S = \{h, r, t\}$, entities and relations sets. E and R, embeddings dim. k

initialize $r \leftarrow$ uniform$(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $r \in R$
  $r \leftarrow r/\|\mathrm{r}\|$ for each $r \in R$
  $e \leftarrow$ uniform $(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$
loop
  $e \leftarrow e/\|\mathrm{e}\|$ for each $e \in E$
  $S_{batch} \leftarrow$ sample(S,b)//sample a minibatch of size b
  $T_{batch} \leftarrow \emptyset$//initialize the set of pairs of triplets
  for $(h, r, t) \in S_{batch}$ **do**
    $(h^{'}, r^{'}, t^{'}) \leftarrow$sample$(S^{'}_{(h,r,t)})$ // sample a corrupted triplet
    $T_{batch} \leftarrow T_{batch} \bigcup \left\{((h, r, t), (h^{'}, r^{'}, t^{'}))\right\}$
  end for
  Updata embedding $\sum_{((h,r,t),(h^{'},r^{'},t^{'}))\in T_{batch}} -log\frac{d(h+r,t)}{d(h+r,t)+d(h^{'}+r^{'}-t^{'})}$
end loop

To extract the features of base units, we utilize entity relationship information provided by the Unified Medical Language System (UMLS) to establish a biomedical entity-relationship network. This paper posits that implicit entity relationship features can enhance event representation, thus facilitating the acquisition of high-quality event representations. As shown in Algorithm 1, to obtain knowledge embeddings for entity relations, we propose a new algorithm called BioER. Our training objective is to separate positive and negative examples as much as possible. The distance calculation formula is equation 9:

$$d(h + r, t) = \|h + r - t\|, \tag{9}$$

where the distance formula $d(h + r, t)$ can take either the L1 or L2 norm.

Proceedings of the 23rd China National Conference on Computational Linguistics, pages 1136-1147, Taiyuan, China, July 25 - 28, 2024.
Volume 1: Main Conference Papers

The core of the algorithm is to make the distance $d(h+r,t)$ of positive examples approach zero, while the distance $d(h^{'}+r^{'}-t^{'})$ of negative examples approach infinity. However, The traditional TransE's loss function is based on Hinge loss, which requires an additional margin distance parameter. This loss function cannot drive the negative value of $d(h^{'}+r^{'}-t^{'})$ to approach infinity, and its convergence heavily depends on the margin distance. Therefore, we improve the loss function here to address these issues.

We train a well-performing BioER model Using the aforementioned algorithm. Subsequently, we utilize BioER to encode the biomedical entities within the events, obtaining a biomedical entity matrix, as illustrated in Figure 2. When inputting corresponding biomedical entities, we search for the corresponding vectors in the biomedical entity matrix. Then, we perform average pooling on the entity vectors to obtain the feature vector $\vec{r}_i$ for the biomedical event i.

2.5 Model Training Objection

Finally, we obtain the final representation $\vec{v}_i$ of the biomedical event along with its positive sample $\vec{v}_i^{+}$ by performing the following operations on $\vec{z}_i$, $\vec{y}_i$, $\vec{r}_i$, and $\vec{z}_i^{+}$, $\vec{y}_i^{+}$, $\vec{r}_i^{+}$. As shown in Equation 10:

$$\vec{v}_i = \vec{y}_i \oplus \vec{z}_i \oplus \vec{r}_i, \qquad \vec{v}_i^{+} = \vec{y}_i^{+} \oplus \vec{z}_i^{+} \oplus \vec{r}_i^{+}. \tag{10}$$

Contrastive learning allows the model to better capture subtle features of positive and negative examples. Therefore, similar to (Gao et al., 2022), we adopt the multi-positive-sample InfoNCE(Gao et al., 2022) loss function for training the model. When presenting a set containing events x_i, the loss function is defined as follows:

$$\mathcal{L} = \sum_{a \in A_i} -\log \frac{g\left(\vec{v}_i, \vec{v}_{ia}^{+}\right)}{g\left(\vec{v}_i, \vec{v}_{ia}^{+}\right) + \sum_{k \in M_i} g\left(\vec{v}_i, \vec{v}_k\right)}, \tag{11}$$

where A(i) and M(i) respectively denote the sets of positive and negative examples for the biomedical event x_i. We use $A_i = \left\{\vec{v}_i^{+1}, \vec{v}_i^{+2}\right\}$. Here, k represents the index of negatives within a sample batch, and g is defined as: $g(\vec{v}_i, \vec{v}_k) = exp(\vec{v}_i^{T}\vec{v}_k/\tau)$, where τ is a hyperparameter known as temperature.

3 Biomedical Event Representation Evaluation Dataset

BioT2E (Frisoni et al., 2022): This dataset provides structured biomedical event text information, including event triggers, biomedical entities, and entity-specific attributes. After filtering out texts without events and single-trigger events, we obtain a total of 36,662 biomedical events.

- **Bio Hard Similarity**: Similar to the hard similarity task(Weber et al., 2018), we have one annotator create similar/dissimilar pairs from BioT2E, while three different annotators give the similarity/dissimilarity rankings. We keep pairs where the annotators agree completely. Finally, we select 240 event pairs from BioT2E. It includes pairs of biomedical events that are semantically similar but have minimal lexical overlaps, as well as pairs that are semantically different but have significant lexical overlaps. For example, "accelerated resistance TGF-beta cells | speedy resistibility TGF-beta cells | accelerated resistance TGF-beta cells | accelerated activation TGF-beta cells". "|" indicates the separator. We use accuracy as the metric to evaluate this task, measuring the accuracy of the model in correctly identifying similar event pairs with higher cosine similarity representations compared to dissimilar event pairs.

- **Bio Transtive Sentence Similarity**: We select 120 event pairs from BioT2E and use standard annotations from the general domain (Kartsaklis and Sadrzadeh, 2014) for this task. Each pair of data is manually annotated with a similarity score ranging from 1 to 7. For example, "leading secretion IL-8 induce transcription IL-8 | leading secretion IL-8 cause tumor apoptosis | 4.1". The final score, 4.1 in this case, represents the similarity score. This dataset comprises a total of 120 similar data instances. A higher score indicates greater similarity between the events. We compute the Spearman correlation coefficient ($\rho \in [-1, 1]$) between the similarity scores predicted by the model and the similarity scores manually annotated.

Hahn-Powell's(Hahn-Powell et al., 2016): The dataset is about extracting causal relations in biomedical events, and it's currently one of the most commonly used datasets for biomedical event causal relation extraction. Given two events in the text and determining whether there is a causal relationship between the two events, we compute the precision (P), recall (R), and F1-score (F) for this task.

4 Experiments

Implementation Details. During the training process, the batch size is set to 32, and our model uses the Adam optimizer. The learning rate for the Adam optimizer of the Context-aware Intensifier is set to 3e-5, while the Adam optimizer used for the Trigger-aware Aggregator has a learning rate of 3e-7. The temperature parameter τ is set to 0.3.

4.1 Main results

We compare the TEHM model with several PLMs in the field of event representation: BERT (Kenton and Toutanova, 2019), Roberta (Liu et al., 2019), BioBERT (Lee et al., 2020), and T5 (Raffel et al., 2020). We also compare two models that use contrastive learning for event representation: Simcse (Chen and He, 2021), SWCC (Gao et al., 2022) and PromptCL (Feng et al., 2023). Performance of various models in Bio Hard Similarity and Bio Transitive Sentence Similarity is reported in Table 1.

The experimental results indicate that our TEHM demonstrates superior performance in similarity tasks for biomedical event representation. It not only outperforms current large-scale language models but also surpasses the SOTA model PromptCL (which be used in the general domain) by 26.1% in the Bio Hard Similarity task.

Model	**Bio Hard Similarity** Acc.%	**Bio Transitive Sentence Similarity**(ρ)
BERT(Kenton and Toutanova, 2019)	42.5	-0.011
BioBERT(Lee et al., 2020)	41.7	0.148
RoBERTa(Liu et al., 2019)	38.3	0.525
T5(Raffel et al., 2020)	50.8	0.284
SimCSE(Chen and He, 2021)	41.7	0.643
SWCC(Gao et al., 2022)	49.2	0.412
PromptCL(Feng et al., 2023)	50.6	0.410
TEHM(Our)	**76.7**	**0.688**

Table 1: Performance in Bio Hard Similarity task and Bio Transitive Sentence Similarity task

These results adequately prove the effectiveness of the TEHM model designed for the structure of biomedical events in improving the quality of biomedical event representation. We argue that there are three reasons why TEHM achieves state-of-the-art performance in biomedical event representation:

- **From a holistic perspective:** We partition biomedical events into three hierarchies and enable TEHM to capture features from three views, then merge them. This helps TEHM to comprehend events at different hierarchies, thereby enhancing the quality of event representation.
- **From a local perspective:** Due to the complexity of biomedical events, we propose the Trigger-aware Aggregator module. It focuses on the trigger word to extract the relationships between various entities and the trigger word within complex events. This architecture enables the model to emphasize the importance of trigger words when learning event representations.
- **From an external knowledge perspective:** We integrate context to capture implicit relationship features within events. Here's an intuitive explanation: rich contextual information aids our understanding of event text when we read it, enabling us to analyze the inherent relationships among multiple simple events within complex events. Additionally, the relationships between biomedical

entities that we integrate are specific to biomedical entities, helping the model understand events from the most basic hierarchies.

4.2 Biomedical Event Casual Relation Extraction

To further validate the representation capability of TEHM in biomedical events, we apply TEHM to the SOTA biomedical causal relation extraction model, MKFN. Table 2 demonstrates the superior performance of TEHM in biomedical event casual relation extraction on Hahn Powell's dataset. This is because the given data text typically contains multiple events. Handling such complex event texts is precisely the problem that the TEHM model aims to address. TEHM can perceive nearby entities and other trigger words centered around the trigger word while incorporating contextual information to explore implicit causal relationships within the text. This enhancement results in a 3.0% performance improvement when integrating event representations into the MKFN on Hahn Powell's dataset.

Model	P(%)	R(%)	F1(%)
LSTM (Akkasi and Moens, 2021)	40.0	25.0	31.0
SVM+L1 (Akkasi and Moens, 2021)	54.0	35.0	43.0
Graph state LSTM (Akkasi and Moens, 2021)	57.0	45.0	50.0
Multiview-CNN (Akkasi and Moens, 2021)	60.0	44.0	52.0
BiLSTM+ATTENTION (Akkasi and Moens, 2021)	54.0	49.0	52.0
BERT (Akkasi and Moens, 2021)	53.7	51.2	52.0
BioBERT+CNN (Liang et al., 2022)	63.0	53.0	58.0
MKFN (Hao et al., 2023)	69.7	58.9	63.9
MKFN+TEHM(Our)	**72.4**	**62.2**	**66.9**

Table 2: Biomedical Event Casual Relation Extraction on Hahn Powell's Dataset

4.3 Ablation Experiments

To study the influence of different modules in TEHM on its overall performance, we conduct a series of ablation experiments on Bio Hard similarity and Bio Transtive Sentence similarity. We can draw the following conclusions from Table 3:

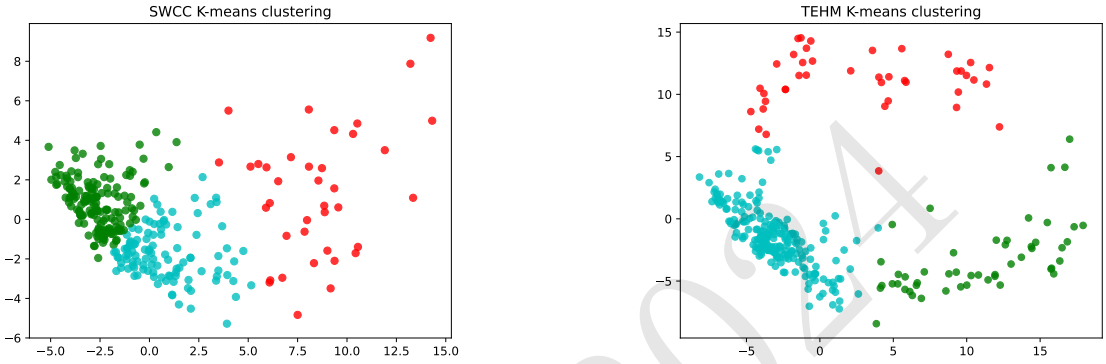
Model	Bio Hard similarity Acc.%	Bio Transitive sentence similarity(ρ)
TEHM	76.7	0.688
–w/o Trigger-aware Aggregator	54.2	0.664
–w/o Event-Context Intensifier	65.8	0.648
–w/o Entity Relation Integrator	75.8	0.688
–w/o Prompt	55.0	0.661

Table 3: Ablation Study For TEHM On Similarity Tasks

- We remove the Trigger-aware Aggregator (TAA) module, resulting in a performance decrease of 22.5% in Bio Hard similarity and 0.024 in Bio Transtive Sentence similarity. Our findings suggest that the TAA plays a significant role in the Bio Hard similarity task, indicating its effectiveness in capturing subtle semantic features. This capability allows the model to distinguish events with high word overlaps but different semantics, mapping them to different vector spaces.
- We remove the Context-aware Intensifier (CAI) module, resulting in a significant decrease in performance by 0.04 for Bio Transitive Sentence similarity, which is the largest decrease among all ablation experiments. We argue that continuous event text contents are beneficial for enhancing performance in this task. This is because the task requires the model to understand the overall semantic meaning of events and calculate the Spearman correlation coefficient.

- We remove the Entity Relation Integrator (ELI) module, resulting in a 0.9% decrease in performance for Bio Hard similarity and no decline in performance for Bio Transtive Sentence similarity. This is consistent with our aforementioned analysis, indicating that Bio Transtive Sentence similarity task requires the model to comprehend events from the whole event. The entity-relationship matrix only provides entity relationship information, aiding in distinguishing subtle semantic features of events.
- To investigate the impact of the Prompt we proposed, we conduct experiments without connecting context information. The results indicate a decrease in performance for Bio Hard similarity and Bio Transtive Sentence similarity tasks by 21.7% and 0.027, respectively. This suggests that injecting context information helps the model better understand the semantic information of biomedical events both at a detailed and holistic level.

4.4 Visualization of Learned Representations



(a) SWCC (b) TEHM

Figure 3: K-means clustering results of biomedical events

We randomly select 300 biomedical events and encode them using both the SWCC and TEHM. Subsequently, we apply Principal Component Analysis (PCA) to map the obtained vectors to a two-dimensional space. Using an unsupervised approach, we perform cluster analysis on the biomedical events with a specified cluster count of 3. The results in Figure 3 show that the vectors encoded by the SWCC lack effective clustering. We also report the clustering performance metric for both the TEHM and the SWCC model (Table 4).

Model	Sil	CH	DBI
SWCC	0.424	341.33	0.811
TEHM	0.623	408.4	0.610

Table 4: Clustering performance metrix

The evaluation involves computing three key metrics: Silhouette Score (Sil), Calinski-Harabasz Score (CH), and Davies Bouldin Score (DB). Biomedical event representations encoded by the TEHM exhibit superior performance in clustering experiments. The cluster analysis demonstrates that the biomedical event representations by the TEHM effectively bring semantically similar events closer in vector space and further separate semantically distinct biomedical events.

4.5 Case Study

To further analyze the performance of TEHM in learning biomedical event representations, we conduct a case study. Table 5 provides the cosine similarity of the vectors of biomedical event pairs encoded by BERT, SWCC, and TEHM.

To illustrate, we take the third and fourth as examples. In the pair of biomedical events, “transformation mammary epithelial cells” and “conversion breast epithelial cells”, which are semantically similar but have fewer overlapping words, TEHM encodes them with a cosine similarity higher than BERT by 0.0739. This case study highlights TEHM’s ability to capture more accurate and contextually meaningful

event representations. In another pair "induced production TNF-alpha" and "induced production cancer", although these events have high word overlap, they are fundamentally different because TNF-alpha is a biologically active cytokine, while "cancer" refers to a disease. However, while BERT encodes these events with a cosine similarity score as high as 0.9490, TEHM yields a score of only 0.2106. Clearly, 0.2106 is more semantically meaningful.

Event Pair	Ground Truth	BERT	SWCC	TEHM
induces B box localization leads B box positioning	Similar	0.6905	0.6562	0.8245
induces B box localization induces B box expression	Dissimilar	0.9570	0.6217	0.1915
transformation of mammary epithelial cells conversion breast epithelial cells	Similar	0.9048	0.8582	0.9833
induced production of TNF-alpha induced production cancer	Dissimilar	0.9490	0.4780	0.2106

Table 5: Cosine Similarity Scores Case Study: We analyze **cosine similarity scores** for biomedical events encoded by different models.

5 Related Work

Due to the significant role of event representations in downstream applications, various methods for event representation have been introduced in current research (Zheng et al., 2024). Long Short-Term Memory (LSTM) (Huang et al., 2015) effectively integrates the sequential information of events into a unified model. Additionally, the fusion of external knowledge in jointly trained models provides a promising direction for capturing potential relationships among events (Ding et al., 2019). It's worth noting that the application of contrastive learning techniques in the field of computer vision, such as Deep InfoMax (Hjelm et al., 2019), MoCo (Long et al., 2024), and SimCSE (Chen and He, 2021), has significantly improved the performance of unsupervised image classification, even surpassing supervised methods. SWCC(Gao et al., 2022) and PromptCL (Feng et al., 2023) both use contrastive learning architecture. PromptCL has achieved SOTA performance in event representation. Recent studies (Gao et al., 2022; Wu et al., 2021) have introduced dropout noise as a data augmentation technique, highlighting its effectiveness in NLP tasks compared to many traditional augmentation methods.

6 Conclusion

We introduce the task of biomedical event representation for the first time and establish the evaluation tasks for biomedical event representation. We will make the relevant datasets publicly available for researchers to conduct further studies. Due to the complexity of biomedical events, we propose a hierarchical structure for biomedical events and design the TEHM model for biomedical event representation. TEHM is capable of obtaining features of biomedical events from three perspectives and capturing subtle semantic features through contrastive learning. Particularly, based on the characteristics of biomedical events, we propose a Trigger-aware Aggregator to extract relational features within complex biomedical events. Additionally, we inject context information into the model using Prompt learning. Experimental results demonstrate that we achieve state-of-the-art performance in biomedical similarity tasks. We also validate the effectiveness of event representation in improving task performance in biomedical causal relation extraction, which also achieves state-of-the-art performance.

Acknowledgements

This work is supported by grant from the National Natural Science Foundation of China (No. 62076048).

References

Abbas Akkasi and Mari-Francine Moens. 2021. Causal relationship extraction from biomedical text using deep neural models: A comprehensive survey. *Journal of biomedical informatics*, 119:103820.

Xinlei Chen and Kaiming He. 2021. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15750–15758.

Xiao Ding, Kuo Liao, Ting Liu, Zhongyang Li, and Junwen Duan. 2019. Event representation learning enhanced with external commonsense knowledge. pages 4893–4902.

Yubo Feng, Lishuang Li, Yi Xiang, and Xueyang Qin. 2023. PromptCL: Improving event representation via prompt template and contrastive learning. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 261–272. Springer.

Giacomo Frisoni, Gianluca Moro, and Lorenzo Balzani. 2022. Text-to-text extraction and verbalization of biomedical event graphs. In Nicoletta Calzolari, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na, editors, *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022*, pages 2692–2710. International Committee on Computational Linguistics.

Jun Gao, Wei Wang, Changlong Yu, Huan Zhao, Wilfred Ng, and Ruifeng Xu. 2022. Improving event representation via simultaneous weakly supervised contrastive learning and clustering. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 3036–3049. Association for Computational Linguistics.

Gus Hahn-Powell, Dane Bell, Marco Antonio Valenzuela-Escárcega, and Mihai Surdeanu. 2016. This before that: Causal precedence in the biomedical domain. In Kevin Bretonnel Cohen, Dina Demner-Fushman, Sophia Ananiadou, and Jun'ichi Tsujii, editors, *Proceedings of the 15th Workshop on Biomedical Natural Language Processing, BioNLP@ACL 2016, Berlin, Germany, August 12, 2016*, pages 146–155. Association for Computational Linguistics.

Jing Hao, Lishuang Li, and Xueyang Qin. 2023. Biomedical causal relation extraction via data augmentation and multi-source knowledge fusion. In Xingpeng Jiang, Haiying Wang, Reda Alhajj, Xiaohua Hu, Felix Engel, Mufti Mahmud, Nadia Pisanti, Xuefeng Cui, and Hong Song, editors, *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2023, Istanbul, Turkiye, December 5-8, 2023*, pages 1112–1117. IEEE.

R. Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Philip Bachman, Adam Trischler, and Yoshua Bengio. 2019. Learning deep representations by mutual information estimation and maximization.

Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional LSTM-CRF models for sequence tagging. *CoRR*, abs/1508.01991.

EunJeong Hwang, Jay-Yoon Lee, Tianyi Yang, Dhruvesh Patel, Dongxu Zhang, and Andrew McCallum. 2022. Event-event relation extraction using probabilistic box embedding. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 235–244. Association for Computational Linguistics.

Dimitri Kartsaklis and Mehrnoosh Sadrzadeh. 2014. A study of entanglement in a categorical framework of natural language. 172:249–261.

Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.

Zhengzhong Liang, Enrique Noriega-Atala, Clayton Morrison, and Mihai Surdeanu. 2022. Low resource causal event detection from biomedical literature. In *Proceedings of the 21st Workshop on Biomedical Language Processing*, pages 252–263.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Xianzhong Long, Han Du, and Yun Li. 2024. Two momentum contrast in triplet for unsupervised visual representation learning. volume 83, pages 10467–10480.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.

Noah Weber, Niranjan Balasubramanian, and Nathanael Chambers. 2018. Event representations with tensor-based compositions. In Sheila A. McIlraith and Kilian Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 4946–4953. AAAI Press.

Lijun Wu, Juntao Li, Yue Wang, Qi Meng, Tao Qin, Wei Chen, Min Zhang, Tie-Yan Liu, et al. 2021. R-drop: Regularized dropout for neural networks. *Advances in Neural Information Processing Systems*, 34:10890–10905.

Jianming Zheng, Fei Cai, Jun Liu, Yanxiang Ling, and Honghui Chen. 2024. Multistructure contrastive learning for pretraining event representation. *IEEE Trans. Neural Networks Learn. Syst.*, 35(1):842–854.