Mining the Past with Dual Criteria: Integrating Three types of Historical Information for Context-aware Event Forecasting

Rong Ma[1,2,3], **Lei Wang**[1,2,3*], **Yating Yang**[1,2,3*], **Bo Ma**[1,2,3*], **Rui Dong**[1,2,3], **Fengyi Yang**[1,2,3], **Ahtamjan Ahmat**[1,2,3], **Kaiwen Lu**[1,2,3], **Xinyue Wang**[4]
[1]Xinjiang Technical Institute of Physics & Chemistry, Chinese Academy of Sciences, Urumqi, China [2]University of Chinese Academy of Sciences, Beijing, China
[3]Xinjiang Laboratory of Minority Speech and Language Information Processing, Urumqi, China
[4]Hohai University, Nanjing, China
{marong, wanglei, yangyt, mabo, dongrui, yangfy, ahtmj}@ms.xjb.ac.cn

Abstract

Event forecasting requires modeling historical event data to predict future events, and achieving accurate predictions depends on effectively capturing the relevant historical information that aids forecasting. Most existing methods focus on entities and structural dependencies to capture historical clues but often overlook implicitly relevant information. This limitation arises from overlooking event semantics and deeper factual associations that are not explicitly connected in the graph structure but are nonetheless critical for accurate forecasting. To address this, we propose a dual-criteria constraint strategy that leverages event semantics for relevance modeling and incorporates a self-supervised semantic filter based on factual event associations to capture implicitly relevant historical information. Building on this strategy, our method, termed ITHI (Integrating Three types of Historical Information), combines sequential event information, periodically repeated event information, and relevant historical information to achieve context-aware event forecasting. We evaluated the proposed ITHI method on three public benchmark datasets, achieving state-of-the-art performance and significantly outperforming existing approaches. Additionally, we validated its effectiveness on two structured temporal knowledge graph forecasting dataset [1].

1 Introduction

Event forecasting aims to predict future events based on observed historical data (Chang et al., 2024). This task is challenging and valuable, with significant research and practical application potential (Wang et al., 2025). It helps in comprehending the world's functioning and offers early alerts for significant occurrences like disasters (Li et al., 2024b; Shui et al., 2023) or regional conflicts (Li et al., 2024a).

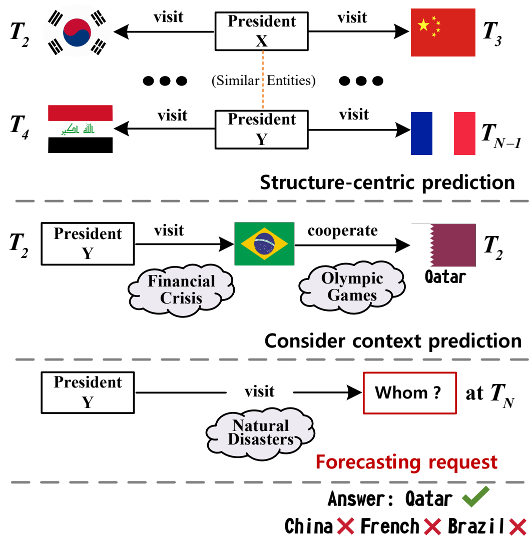

Figure 1: Comparison of structure-centric and consider context forecasting methods for a sample event forecasting request.

Existing event forecasting methods mainly include structured temporal knowledge graph reasoning and context-aware event forecasting. Structured methods leverage rule mining (Liu et al., 2022; Liao et al., 2024) or entity-based anchoring and model structural dependencies in temporal knowledge graphs to predict future events (Liao et al., 2024; Zhang et al., 2024; Chen et al., 2024). For instance, entities with similar behaviors are modeled to anticipate interactions (Chen et al., 2024; Mingcong et al., 2024). Context-aware methods address this limitation by incorporating event context into predictions. For example, Ma et al. (2023) improve prediction accuracy by fusing event context information through a graph disentanglement method. While these approaches perform well in an Explicit Prediction Scenario (i.e., when key events are discoverable via direct structural or semantic links), they falter in more complex cases. Structure-centric approaches usually consider only explicit dependencies and miss semantic differences, while

[*]Corresponding author.
[1]https://github.com/wooden070/ITHI

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 8148–8163
November 4-9, 2025 ©2025 Association for Computational Linguistics

context-aware approaches handle semantic features separately from structural links. This limitation becomes critical in an Implicit Prediction Scenario, where accurate forecasting is impossible by relying solely on structural connections or semantic filtering. As a result, current methods are limited in mining historical information that holds potential predictive value, failing to fully leverage the complete semantics of events to guide retrieval.

The challenge posed by the Implicit Prediction Scenario is clearly demonstrated in Figure 1. Consider the forecasting request: "where President Y will visit in the context of a natural disaster?" Structure-centered approaches rely on similarities in entity behavior, e.g., recommending France or Saudi Arabia by referring to previous visits by President Y and President X, leading to incorrect forecasting results. While context-aware approaches are capable of detecting semantic similarities under the contexts of "natural disasters" and "economic crises" and of identifying that President Y visited Brazil, their inability to incorporate structural dependency information prevents them from revealing the collaboration between Brazil and Qatar. Therefore, it is necessary to integrate event semantics with factual structural constraints to model historical information that is valuable for event forecasting. Addressing this integration to effectively identify such implicitly relevant historical events is the central focus of this paper.

To overcome these limitations, we propose a dual-criteria constraint strategy that leverages event semantics for relevance modeling and incorporates a self-supervised semantic filter based on factual event associations to capture implicitly relevant historical information. Building on this strategy, we introduce ITHI (Integrating Three types of Historical Information), which fuses sequential event information, periodically repeated event information, and relevant historical information for context-aware event forecasting. Using graph neural networks, we model different historical information to capture potential event correlations. We evaluated ITHI on three public benchmark datasets, achieving State-of-the-Art (SOTA) performance. Additionally, its effectiveness is verified in pure structured knowledge graph reasoning datasets (ICEWS14 and ICEWS18). Furthermore, ITHI can act as a plug-in to Large Language Models (LLMs), and their combination significantly enhances event forecasting performance.

Our main contributions are outlined below.

- We introduce a dual-criteria constrained strategy that leverages event semantics with factual event associations, utilizing an explicit self-supervised semantic filter to mine relevant historical information.
- We propose a new method, ITHI, which integrates sequential event information, periodically repeated event information, and relevant historical information to improve event forecasting accuracy.
- We validate ITHI on three public benchmarks, achieving new SOTA performance, and conduct extension experiments to further demonstrate its effectiveness.

2 Problem Formulation

Context-aware Event Forecasting: In this study, we define an event as a quintuple (s, r, o, t, c), where s and o denote the event subject and event object, respectively, both belonging to the entity set E ($s \in E$ and $o \in E$); r indicates the event relationship belonging to the relation set R ($r \in R$); t is the timestamp of the event occurrence, which denotes the temporal information of the event; and c denotes the context in which the current event is occurring, and c belongs to $C = \{c_1, c_2, ..., c_k\}$, and k is the total number of contexts. Under the same timestamp t, all event quintuples would form an event graph G_t, which is denoted as $G_t = \{(s_n, r_n, o_n, t_n, c_n)\}_{n=1}^{N}$, where c_n is the context of the n-th event. N is the total number of events occurring under the timestamp t. Given a query $Q = \{s, r, ?, t\}$ or $\{o, r^{-1}, ?, t\}$, and a context c corresponding to the query events, our goal is to predict the missing event object o in the Query, given multiple historical event graphs $G_{t'<t} = \{G_1, G_2, ..., G_{t'}\}$ prior to a certain timestamp t. Following the experimental setup and philosophy of Ma et al. (2023), we emphasize that specifying a categorization during inference context does not leak information about the predicted object. For example, given the context of the New Crown epidemic, the question "President Y will visit which country" will not be leaked to the model.

3 Methodology

This section provides a detailed explanation of the proposed ITHI event prediction method. The process is divided into three essential stages: historical information retrieval, node and relation representation, and prediction decoding. The ITHI framework addresses context-aware event prediction through multi-faceted historical evidence fusion. The overall model structure is shown in Figure 2.

Historical Information Retrieval: We argue that three crucial types of historical information are essential for predicting future events: 1) Sequential Event Information (SEI): Used to capture historical information that has a near-term correlation with the current forecast request, 2) Periodically Repeated Event Information (PREI): Users capture historical information with periodic and repetitive patterns, and 3) Relevant Historical Information (RHI): Relevant historical information associated with the current forecast request. Each type plays a unique role in capturing temporal, structural, and semantic dependencies between events.

For **SEI**, given a query on day t, the SEI retrieves the most recent n days of events to capture the continuous temporal progression. This information helps model the recent sequence of events and their direct impact on future predictions.

For **PREI**, given a query $Q = \{s, r, ?, t\}$, the process begins with retrieving historical event triples $\{s, r, o', t'\}$ from the temporal knowledge graph. To ensure temporal relevance, which are then ranked by their temporal proximity to the query timestamp t. The top-n most recent events are then selected based on this temporal distance metric. This temporal prioritization mechanism effectively captures recurring patterns and structural dependencies inherent in periodic events involving the same subject-relation pair (s, r).

For **RHI**, we propose a dual-criteria constraint architecture that first enforces semantic constraints through event information and employs similar semantic strategies to constrain historical data. Second, a self-supervised semantic filter for capturing relevant historical information is constructed by modeling semantic representation and fact structure patterns collaboratively.

First, existing event prediction methods often rely on structured relationships or explicit rules. We propose integrating a semantic embedding model with a historical event graph for cross-structural event retrieval. Using the semantic embedding model (GTE-base[2]) with the query $Q = \{s, r, ?, t, c\}$ is encoded to obtain the embedding representation $E_q \in \mathbb{R}^{3\times d}$. All historical events are similarly embedded, forming a resource library $E_{lib} \in \mathbb{R}^{3\times d}$. The top-$n$ most similar historical events E_{sim} are selected by calculating the Euclidean distance between E_q and E_{lib}.

Second, to establish reliable relevance signals under the premise of semantic similarity, we design a self-supervised scoring mechanism that prioritizes historical events exhibiting both semantic congruence (with the query and its contextual background) and structural proximity in the event graph. For each candidate event E_i retrieved through semantic similarity matching, we evaluate its association strength with the ground-truth answer in the graph $G_{current}$. Specifically: E_i receives a score of 3 if it directly connects to the ground-truth answer (0-hop), a score of 2 if it shares a direct neighbor (1-hop), and a score of 1 if linked via two intermediate nodes (2-hop). Events beyond 2-hop or without structural linkage are scored 0. This scheme quantifies the association strength between related events and the query event.

Third, we develop and train a specialized event-related information filter via self-supervised label construction, comprising three layers: input, hidden, and output.

$$\hat{E} = [E_i; E_q] \cdot W_1 + b_1, \hat{E} \in \mathbb{R}^d \tag{1}$$

$$\hat{y} = 3 \cdot \sigma((ReLU(\hat{E}) \cdot W_2 + b_2)), \hat{y} \in \mathbb{R}^1 \tag{2}$$

where $E_i \in E_{sim}$, $W_1 \in \mathbb{R}^{3d\times d}$, $W_2 \in \mathbb{R}^{d\times 1}$, and σ is a sigmoid function. The filter aims to minimize the error between predicted scores and self-supervised labels, using mean square error (MSE) as the loss function.

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{3}$$

where y_i is the labeling score obtained from self-supervised scoring mechanism. In order to avoid information leakage, we divide the new training set and test set from the original dataset training set to facilitate the learning of the filter.

Node and Relation Representation: After acquiring the three types of historical information, we perform representation learning on the nodes and edges of the corresponding graph to capture the structural features of event entities and their relationships. Specifically, we construct three subgraphs based on prediction requests: sequential history graph (G_s), history repetition subgraph (G_h),

[2]https://huggingface.co/thenlper/gte-base

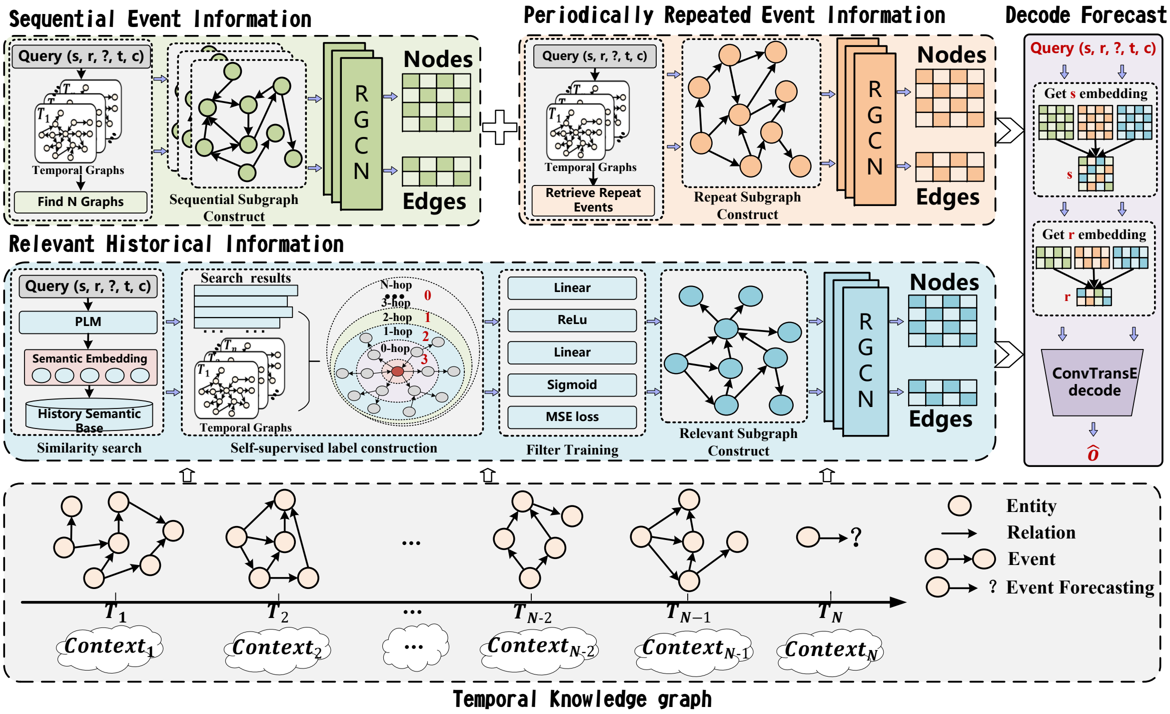

Figure 2: The overall structure of the proposed ITHI, which models sequential event information, periodically repeated event information, and relevant historical information for a query using independent GNNs.

and related history subgraph (G_r). The nodes and edges are randomly initialized and the representations of the nodes and edges in each subgraph are updated with a Relational Graph Convolutional Network (RGCN (Schlichtkrull et al., 2018)).

For SEI, denoted as G_s, nodes represent event entities and edges represent entity relations. For PREI and RHI, the historical repetitive subgraph G_h and relevant event subgraph G_r are constructed from the respective events obtained previously.

Initially, the node ($n_s, n_h, n_r \in \mathbb{R}^d$) and edge ($e_s, e_h, e_r \in \mathbb{R}^d$) features of the G_s, G_h and G_r respectively are initialized. In order to capture the relationships between nodes and their neighbors in the graph, we used three mutually independent RGCNs to update the node and edge representations for each of the three types of information.

$$h_{e_i}^{l+1} = \sigma\left(\sum_{r_j}\sum_{e_i,e_k}\frac{1}{|\mathcal{N}_i^j|}W_1^l h_{e_k}^l + W_2^l h_{e_i}^l\right) \quad (4)$$

where $h_{e_i}^l$ represents the embedding of node e_i at layer l; R and E denote the sets of relations and entities, respectively; $r_j \in R$ and $e_i, e_k \in E$; $\mathcal{N}_i^j$ represents the set of neighboring nodes connected to node e_i via relation r_j; $|\cdot|$ represents the number of nodes; W_1^l and W_2^l are the learnable parameters of RGCN in layer l; and σ is the activation function. The final node representation is generated through L-layer aggregation.

$$h_{e_i} = \sum_{l=0}^{L} h_{e_i}^{(l)} \quad (5)$$

Prediction Decode: The node and edge representations of the three types of historical information obtained are concatenated, resulting in nodes as $n_{all} = [n_s \oplus n_h \oplus n_r] \in \mathbb{R}^{3d}$, and edges as $e_{all} = [e_s \oplus e_h \oplus e_r] \in \mathbb{R}^{3d}$. Following existing work on the task of event prediction, we use a decoder based on ConvTransE. For a query $Q = (s, r, ?, t, c)$, a representation of the query is generated based on n_{all} and e_{all}, and then the candidate entities are scored by the inner product between the query and the candidate entities $\mathcal{E}$. The formalisation of this process is as follows:

$$\hat{p}(\mathcal{E}|s, r, c, G_{<t}) = \mathrm{f}(\hat{\mathcal{E}}_c \mathrm{CTE}(\hat{s}_c, \hat{r}_c)) \quad (6)$$

where CTE is the ConvTransE decoder; f$(\cdot)$ is the softmax function; $\hat{s}_c$ and $\hat{r}_c$ are the representations for s and r under the context c respectively.

$$\hat{o}_{(s,r,t,c)} = \arg\max_{\mathcal{E}} \hat{p}(\mathcal{E}|s, r, c, G_{<t}) \quad (7)$$

We optimise the whole framework by cross-entropy loss, defined as follows:

$$\mathcal{L} = \sum_{t=1}^{T} \sum_{c \in \mathcal{C}} \sum_{(s,r) \in G_c^t} o \log \hat{p}(\mathcal{E}|s, r, c, G_{<t}) \quad (8)$$

where T is the total number of training day steps; o is the query $Q(s, r, ?, c)$ of the missing target entity representation for one-hot.

4 Experiments

4.1 Datasets

We evaluate our method on three publicly available context-aware event forecasting datasets: EG, IR, and IS, covering Egypt, Iran, and Israel, respectively. These large-scale datasets, constructed by Ma et al. (2023) from February 2015 to March 2022, were filtered by geographic location and low-quality records, with political event-related news URLs extracted. Each dataset is split into training, validation, and test sets in an 8:1:1 ratio. In addition, we conduct extended experiments on two purely structured temporal knowledge graph forecasting datasets. We follow the experimental setup of Chen et al. (2024) to divide the training, validation, and test sets in the ratio of 8:1:1 on ICEWS14 and ICEWS18 datasets. Detailed statistics are shown in Table 1.

Datasets	EG	IR	IS	ICEWS14	ICEWS18
$\lvert\mathcal{V}\rvert$	2,594	2,988	3,456	7,128	23,033
$\lvert\mathcal{E}\rvert$	225	236	238	230	256
#urls	96,081	223,616	345,611	–	–
#days	2,584	2,584	2,584	365	365
#train	377,430	973,752	1,430,389	74,845	373,018
#valid	36,588	69,827	171,518	8,514	45,995
#test	28,644	76,239	156,695	7,371	49,545

Table 1: Detailed statistics of the datasets

4.2 Experimental Settings

We conduct experiments on three datasets (EG, IR, and IS) and evaluate performance using MRR and Hit@{1, 3, 10}, as in previous work (Li et al., 2021; Ma et al., 2023). The best-performing model on the validation set, based on MRR, is selected for testing. Following Ma et al. (2023), we remove frequent nodes from the graph during metric calculation to better assess model performance. To ensure fairness and maintain consistency with prior settings, we set d as 200, learning rate as $1e^{-3}$, weight decay as $1e^{-6}$, and employ cross-entropy loss. In our approach, we explore the number of RGCN propagation layers from $\{1, 2, 3\}$ and the training epoch is 20. For the range of days of sequential event information, we search from the range {1, 3, 7}. The Adam optimizer (Kingma and Ba, 2015) and Xavier initialization (Glorot and Bengio, 2010) are applied for all parameters. All experiments are carried out on eight 24GB NVIDIA GeForce GTX 4090. The LLMs involved in this paper are all Llama-3.1-8B-Instruct versions[3].

4.3 Baselines

To comprehensively evaluate the performance of our context-aware event forecasting method, we compare it with three baseline approaches representing distinct technical paradigms:

(1) Static Knowledge Graph Completion Methods treat event forecasting as a static link prediction task, ignoring temporal dynamics: RGCN (Schlichtkrull et al., 2018) and ConvTransE (Shang et al., 2019).

(2) Temporal Knowledge Graph Forecasting Methods. These approaches explicitly model temporal evolution patterns with structured historical retrieval, considering relevant information at the time of inference through path dependencies, rule constraints, and so on: RE-GCN (Li et al., 2021), HisMatch (Li et al., 2022), GenTKG (Liao et al., 2024), and LogCL (Chen et al., 2024).

(3) Context-Aware Event Forecasting Methods. These methods incorporate unstructured contextual signals for prediction: CMF_{ont} (Deng et al., 2021), CMF_{art} (Deng et al., 2021), SeCoGD (Ma et al., 2023), LLMs-ICL (Lee et al., 2023), and LLMs-COH (Xia et al., 2024). For the reproduction of the LLM-ICL and LLM-COH methods, we base on RE-GCN to feed the event context information to LLM for event prediction.

4.4 Performance Comparison

Replicating three RE-GCN-based models, we show that integrating In-Context Learning with Chain-of-History information improves performance, with additional gains from their combined application. However, despite these improvements, the gains from both methods still need enhancement. As shown in Table 2, our proposed ITHI method outperforms existing models across all datasets, achieving SOTA performance. For example, on the EG dataset, ITHI achieves an MRR improvement of 9.6% over the best existing method. On the IR dataset, ITHI outperforms the best existing

[3]https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct

Model	EG				IR				IS			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
RGCN	9.74	2.79	10.46	23.77	11.85	3.66	13.01	28.60	8.61	2.42	6.52	23.07
ConvTransE	12.05	3.77	13.05	29.21	14.05	4.62	15.29	34.12	10.79	2.87	9.94	29.30
RE-GCN*	11.99	3.97	12.90	28.65	14.41	4.94	14.77	34.21	10.66	2.60	9.85	29.27
HiSMatch	11.26	2.75	12.79	29.06	14.69	4.96	15.99	35.72	12.83	4.34	12.48	30.17
GenTKG*	–	1.96	5.56	15.02	–	0.46	10.14	20.12	–	1.35	5.16	10.01
LogCL*	<u>17.79</u>	<u>8.21</u>	<u>20.72</u>	<u>37.34</u>	17.13	4.96	18.79	<u>45.98</u>	<u>16.10</u>	5.45	15.57	<u>42.69</u>
CMF$_{ont}$	12.06	3.48	12.98	30.15	15.27	5.29	16.43	36.73	12.48	3.68	12.24	32.56
CMF$_{art}$	12.02	3.45	12.93	30.27	15.10	4.94	16.36	37.16	12.63	3.82	12.36	32.61
SeCoGD	14.64	5.93	16.05	32.36	<u>17.57</u>	<u>7.24</u>	<u>19.02</u>	39.75	15.52	<u>5.95</u>	<u>15.88</u>	36.93
RE-GCN+ICL*	12.15	4.03	13.20	28.92	13.73	3.76	15.38	34.45	11.09	2.95	10.51	29.61
RE-GCN+COH*	12.86	4.81	13.94	29.33	14.91	5.41	16.19	35.13	12.76	4.64	12.58	31.04
Ours	**27.39**	**19.03**	**30.00**	**43.80**	**31.65**	**23.21**	**33.27**	**49.35**	**30.05**	**22.84**	**30.13**	**46.10**

Table 2: Comparison of ITHI performance with three types of benchmark methods across three datasets, with results presented as percentages (%). * Indicates reproduced methods, with results presented as percentages (%).

method with an MRR improvement of 14.08%. On the IS dataset, ITHI attains an MRR improvement of 13.95% over the best existing method. Evaluations on three datasets using various metrics (MRR, Hit@1, Hit@3, and Hit@10) indicate that our proposed ITHI method significantly improves event prediction accuracy by effectively integrating SEI, PHEI, and RHI, ITHI significantly improves predictive performance over previous SOTA models.

Appendix Table 9 shows the results of the structural temporal knowledge graph forecasting task. Experimental findings indicate that leveraging event elements (e.g., subject and object) with a dual-criteria constraint strategy extracts relevant semantic clues even without event context, confirming that integrating event semantics enhances forecasting performance, especially in scenarios where structural connections alone are insufficient for accurate prediction.

Model	EG		IR		IS	
	MRR	Hit@1	MRR	Hit@1	MRR	Hit@1
ITHI	**27.39**	**19.03**	**31.65**	**23.21**	**30.05**	**22.84**
w/o SEI	26.82	18.61	28.68	20.57	28.05	20.95
w/o PREI	26.49	17.96	29.98	21.59	28.57	21.23
w/o RHI	11.34	2.70	12.88	4.20	11.86	3.68
w/o filter	24.34	15.87	27.89	18.90	26.82	18.00
w/o context	25.50	17.48	27.50	18.35	26.10	17.81

Table 3: Partial ablation experiment results across three datasets (%).

Model	EG			
	MRR	Hit@1	Hit@3	Hit@10
ITHI	27.39	**19.03**	30.00	43.80
ITHI+ICL	**27.43**	18.52	**30.67**	**44.84**

Table 4: Results of the Graph-LLMs Collaboration Experiment, reported in percentages (%).

4.5 Ablation Study

Table 3 presents the ablation experiments conducted on three datasets (complete results are given in Table 8), where different modules were removed to assess their impact on model performance. The results indicate that removing any single module leads to a significant drop in performance. Notably, removing RHI and SEI has the greatest impact, especially on the EG and IS datasets, where both MRR and Hit@1 show substantial declines, underscoring their critical roles in integrating structural dependencies and semantic relevance for accurate forecasting. In addition, removing the self-supervised filter (w/o filter) and the event context (w/o context) in the RHI module also lead to significant performance degradation. Experimental results show that the proposed dual-constraint strategy yields positive results. When the event context is removed, the performance degradation remains within acceptable limits, indicating that embedding only event subjects and predicates still provides effective semantic constraints.

Model	EG				IR				IS			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
SeCoGD*	14.63	5.79	15.97	32.64	17.29	6.66	19.07	40.07	15.42	5.86	15.89	36.84
SeCoGD-IPS	10.64	3.26	10.82	25.72	10.08	2.51	10.45	28.82	9.25	1.85	8.18	25.53
SeCoGD-EPS	27.00	13.67	31.97	54.11	30.58	15.17	36.78	63.18	27.28	13.56	30.71	58.58
ITHI	27.39	19.03	30.00	43.80	31.65	23.21	33.27	49.35	30.05	22.84	30.13	46.10
ITHI-IPS	18.66	10.97	20.39	33.68	19.67	11.78	20.29	36.00	18.31	11.84	17.48	32.32
ITHI-EPS	54.46	43.99	59.83	75.15	56.26	46.69	59.94	76.76	52.62	43.98	54.43	72.58

Table 5: Forecasting performance comparison between ITHI and SeCoGD models on EG, IR and IS test sets under Implicit Prediction Scenario (IPS) and Explicit Prediction Scenario (EPS).

4.6 Explore Graph-LLMs Collaboration

In this section, we explore the collaborative potential between ITHI and LLMs in event prediction tasks, and the experimental results are presented in Table 4. Experimental comparisons indicate that integrating the ITHI method with LLMs enhances event prediction performance. Detailed prompts are shown in Figure A.6 in the Appendix. Initially, the top 150 candidate answers from ITHI's predictions were selected, and prompts were constructed for each to aid LLMs in understanding and processing event prediction queries. Furthermore, considering event prediction involves both (s, r, ?) and (o, r', ?) queries, manual inversion of existing relationship descriptions was performed to leverage bidirectional relationships. For example, "A Express intent to cooperate B" can be reversed to read: "A Received an expression of intent to cooperate from B." Experimental results demonstrate that combining the ITHI method with LLMs improves performance metrics such as MRR, Hit@1, Hit@3, and Hit@10. Notably, Hit@10 increased from 43.80% to 44.84%, indicating that the proposed method can serve as a plugin for large models, further enhancing event prediction accuracy.

4.7 Analysis of Explicit and Implicit Event Forecasting Performance

We evaluate ITHI against the SeCoGD baseline on the EG, IR, and IS test sets under two distinct forecasting scenarios. In the Implicit Prediction Scenario (**IPS**), the correct answer does not appear among the historical events within the 50-day window preceding the prediction request; in the Explicit Prediction Scenario (**EPS**), it appears within that window. Experimental results are shown in Table 5.

Experimental results indicate that ITHI demonstrates superior forecasting performance in both explicit and implicit event forecasting scenarios by effectively modeling the corresponding event correlations. (1) In the IPS, the lack of explicit historical patterns poses a challenge for effective event forecasting. ITHI is significantly better at handling implicit event forecasting requests. While SeCoGD suffers a significant performance decline in the absence of explicit event repetitions (e.g., EG-MRR drops from 14.63% to 10.64%), ITHI maintains robust performance even when explicit patterns are unavailable (e.g., EG-MRR for the ITHI-IPS is 18.66% compared to 14.63% for SeCoGD*). This indicates that ITHI's modeling approach is more capable of revealing hidden relevant information between events. (2) In the EPS, both models benefit from the availability of explicit historical patterns. ITHI demonstrates a substantially stronger ability to leverage these cues. While SeCoGD performance improved over the original (e.g., EG-MRR from 14.63% to 27.00%), ITHI scored significantly higher on all metrics (e.g., EG-MRR from 27.39% to 54.46%). In summary, ITHI effectively models explicit and implicit historical event information,

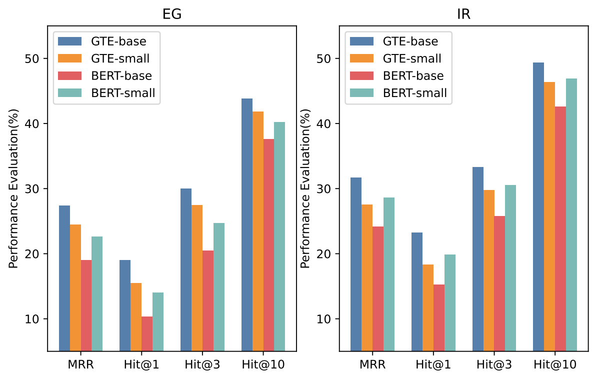

Figure 3: Performance of ITHI on EG and IR datasets using different pre-trained embedding models.

rendering it a more reliable and comprehensive event forecasting model than the SeCoGD.

4.8 Impact of Event Semantic Embedding Quality on Forecasting Performance

We evaluate the ITHI method on two datasets, EG and IR, using different pretrained semantic embedding models: GTE$_{base}$, GTE$_{small}$[4], BERT$_{base}$[5], and BERT$_{small}$[6]. All models train and evaluate under the same conditions to isolate the impact of embedding quality. The goal is to investigate the impact of event semantic representation quality on the predictive performance of ITHI. Experimental results are shown in Figure 3.

The experimental results reveal that the choice of semantic embedding model significantly impacts event forecasting performance. Within the ITHI model, GTE$_{base}$ exhibits superior performance to GTE$_{small}$ across all metrics. This is likely attributable to GTE$_{base}$'s higher embedding dimension and more robust semantic representation capabilities. BERT$_{small}$ outperforms BERT$_{base}$ despite its lower feature dimension. This arises because BERT$_{small}$ undergoes distillation during training, endowing it with semantic similarity encoding capabilities, whereas BERT$_{base}$ lacks specialized training for this task. These findings highlight the crucial role of high-quality semantic embeddings in effectively filtering relevant historical information, ultimately enhancing event forecasting accuracy.

4.9 Case Study

In order to validate the effectiveness of ITHI to retrieve relevant historical events, we analyze the PRHI and RHI retrieved by a specific event forecasting request. Case study is shown in Table 6. For the query "Egypt Make a visit Whom? at 2357th day" a series of repeat historical events is identified using the PRHI retrieval mechanism, such as "United Kingdom, Military Base, Kuwait." These events indicate multiple visits by Egypt around the queried time frame but do not provide clear clues to predict the specific visit target, resulting in ambiguous correlations. Applying the dual-constraint criteria filter to screen the RHI yields more relevant data. In Table 6, the event "Alexandria Make a visit Military Base at 2356th day" is highly relevant, clearly pointing to the potential visit target in the query. This demonstrates that the dual-constraint approach effectively enhances event forecasting by integrating implicit yet relevant historical information beyond simple event repetition.

Query	**Egypt Make a visit Whom? at 2357th day.**
HREI	Egypt Make a visit United Kingdom at 2356th day. Egypt Make a visit Military Base at 2356th day. Egypt Make a visit Kuwait at 2355th day. Egypt Make a visit Criminal at 2355th day. ...
RHI	Egypt Host a visit Russia at 2331th day. Egypt Make a visit **Military Base** at 2356th day. **Alexandria** Make a visit **Military Base** at 2356th day. Egypt Host a visit Foreign Minist at 2341th day. ...
Answer	**Alexandria**

Table 6: Case study to validate the effectiveness of ITHI for retrieving historical events.

5 Related Work

Natural language processing applications are attracting increasing attention (Ahmat et al., 2025), including event forecasting tasks. A dominant approach utilizes Temporal Knowledge Graphs ($e.g.$ ICEWS George et al. (2019) and GDELT Leetaru and Schrodt (2013)) , to model events with quadruples (entity, relation, entity, timestamp) (Cai et al., 2024). The primary objective of TKG forecasting is to predict future events, which is often formulated as completing a missing entity or relation in a future quadruple (e.g.,(s,r,?,t) where t is a future timestamp) (Liang et al., 2024).

Early researchers extended static graph methods such as DistMult (Yang et al., 2015), ConvE (Dettmers et al., 2018), and ConvTransE (Shang et al., 2019) to temporal settings, typically employing interpolation or extrapolation to predict future event relations. Some scholars, embracing the notion of historical repetition, modeled the sequentiality, repetitiveness, and periodicity of event occurrences to study temporal knowledge graph reasoning tasks (Xu et al., 2023; Gastinger et al., 2024; Lv et al., 2024; Mirtaheri et al., 2023). Recently, additional researchers have explored how to recognize relevant historical information that contributes to the task of temporal knowledge graph forecasting (Liao et al., 2024; Chen et al., 2024). Specifically, Mingcong et al. (2024) use the entity as an anchor site to predict future events by modeling the interactions between entities with similar behaviors. Liao et al. (2024) retrieve relevant information based on temporal logic rules (Liu et al., 2022), in generative forecasting by ef-

[4]https://huggingface.co/thenlper/gte-small
[5]https://huggingface.co/google-bert/bert-base-uncased
[6]https://huggingface.co/prajjwal1/bert-small

ficiently fine-tuning a LLM through parameters. Leveraging temporal dependencies of entity association paths, Chen et al. (2024) utilized contrastive learning on entity-centric multi-hop historical subgraphs to mine relevant historical information for forecasting. However, these methods fail to fully capture the semantic and contextual information required for complex events.

Context-aware event forecasting integrates event context, setting it apart from methods that rely solely on structural information; for instance, Ma et al. (2023) improved forecasting accuracy by fusing event context through a graph disentanglement method. Xia et al. (2024) used a question-answer approach with LLMs to filter relevant historical information for event forecasting.

6 Conclusion

In this paper, we propose ITHI, a novel method for context-aware event forecasting that systematically integrates three types of historical information: sequential event information, periodically repeated event information, and relevant historical information. Our dual-criteria constrained architecture synergizes semantic constraints with factual structural patterns through a self-supervised semantic filter, which enables the evidence-based retrieval of historically relevant events. Experimental results across three public benchmark datasets demonstrate the superiority of ITHI, achieving a new SOTA performance, and furthermore, extension experiments provide further evidence that ITHI is also effective in purely structured temporal knowledge graph reasoning task. Moreover, ITHI can function as a plug-in for LLMs, and the integration improves the performance of event forecasting.

Limitations

Although the ITHI method significantly improves Context-aware Event Forecasting, it has several limitations. First, ITHI relies on existing information in historical knowledge graphs, leading to insufficient generalization when handling new or unseen events. This limits the model's ability to adapt to new scenarios. Second, due to its dependence on semantic and temporal propagation, the self-supervised semantic filter might miss implicit connections in complex event chains, especially with indirectly related events. Additionally, the current design of ITHI is focused on event prediction tasks, not fully addressing the diversity of different types of events or tasks, which limits its application in other fields or more complex scenarios. Future research could deeply integrate graph methods with LLMs to enhance the model's generalization ability and expand its application to complex prediction tasks across various domains.

Ethics Statement

This work fully complies with the ACL Ethics Policy. We declare that there are no ethical issues in this paper, to the best of our knowledge.

Acknowledgement

This work is mainly supported by the Tianshan Talent Training Program (Grant No. 2023TSYCCX0041), in part by the Tianshan Elite Science and Technology Innovation Leading Talents Program (Grant No. 2022TSYCLJ0046), in part by the Tianshan Elite Science and Technology Topnotch Youth Talents Program (Grant No. 2022TSYCCX0059 and 2023TSYCCX0044), in part by the Natural Science Foundation of Xinjiang Uyghur Autonomous Region (Grant No. 2022D01D81, 2023D01D17 and 2022D01D04), in part by the Outstanding Member Program of the Youth Innovation Promotion Association of Chinese Academy of Sciences (Grant No. Y2023118), and the Key Research and Development Program of Xinjiang Uyghur Autonomous Region (Grant No. 2023B03024 and 2024B03026).

References

Ahtamjan Ahmat, Lei Wang, Yating Yang, Bo Ma, Rui Dong, Kaiwen Lu, Rong Ma, and Xinyue Wang. 2025. M^2-vlp: Enhancing multilingual vision-language pre-training via multi-grained alignment. In *Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025-2 May 2025*, pages 3438–3450. ACM.

Li Cai, Xin Mao, Yuhao Zhou, Zhaoguang Long, Changxu Wu, and Man Lan. 2024. A survey on temporal knowledge graph: Representation learning and applications. *CoRR*, abs/2403.04782.

He Chang, Chenchen Ye, Zhulin Tao, Jie Wu, Zhengmao Yang, Yunshan Ma, Xianglin Huang, and Tat-Seng Chua. 2024. A comprehensive evaluation of large language models on temporal event forecasting. *CoRR*, abs/2407.11638.

Wei Chen, Huaiyu Wan, Yuting Wu, Shuyuan Zhao, Jiayaqi Cheng, Yuxin Li, and Youfang Lin. 2024. Local-global history-aware contrastive learning for temporal knowledge graph reasoning. In *40th IEEE International Conference on Data Engineering, ICDE 2024, Utrecht, The Netherlands, May 13-16, 2024*, pages 733–746. IEEE.

Songgaojun Deng, Huzefa Rangwala, and Yue Ning. 2021. Understanding event predictions via contextualized multilevel feature learning. In *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021*, pages 342–351. ACM.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *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 1811–1818. AAAI Press.

Julia Gastinger, Christian Meilicke, Federico Errica, Timo Sztyler, Anett Schuelke, and Heiner Stuckenschmidt. 2024. History repeats itself: A baseline for temporal knowledge graph forecasting. *CoRR*, abs/2404.16726.

Olawunmi George, Rizwana Rizia, Md Fitrat Hossain, Nadiyah Johnson, Carla Echeveste, Jose Lizarraga Mazaba, Katinka Hooyer, Zeno Franco, Mark Flower, Praveen Madiraju, and Lisa Rein. 2019. Visualizing early warning signs of behavioral crisis in military veterans: Empowering peer decision support. In *Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management, València, Spain, May 19-22, 2019*. ISCRAM Association.

Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2010, Chia Laguna Resort, Sardinia, Italy, May 13-15, 2010*, volume 9 of *JMLR Proceedings*, pages 249–256. JMLR.org.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Dong-Ho Lee, Kian Ahrabian, Woojeong Jin, Fred Morstatter, and Jay Pujara. 2023. Temporal knowledge graph forecasting without knowledge using in-context learning. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, pages 544–557. Association for Computational Linguistics.

Kalev Leetaru and Philip A Schrodt. 2013. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, volume 2, pages 1–49. Citeseer.

Guopeng Li, Zirui Li, Victor L. Knoop, and J. W. C. van Lint. 2024a. How far ahead should autonomous vehicles start resolving predicted conflicts? exploring uncertainty-based safety-efficiency trade-off. *IEEE Trans. Intell. Transp. Syst.*, 25(10):14183–14195.

Jing Li, Feng Chu, Ada Che, and Yunqiang Yin. 2024b. A three-stage relief network design approach for predictable disasters considering time-dependent uncertainty. *IEEE Trans. Intell. Transp. Syst.*, 25(6):5418–5434.

Zixuan Li, Zhongni Hou, Saiping Guan, Xiaolong Jin, Weihua Peng, Long Bai, Yajuan Lyu, Wei Li, Jiafeng Guo, and Xueqi Cheng. 2022. Hismatch: Historical structure matching based temporal knowledge graph reasoning. In *Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 7328–7338. Association for Computational Linguistics.

Zixuan Li, Xiaolong Jin, Wei Li, Saiping Guan, Jiafeng Guo, Huawei Shen, Yuanzhuo Wang, and Xueqi Cheng. 2021. Temporal knowledge graph reasoning based on evolutional representation learning. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 408–417. ACM.

Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Sihang Zhou, Xinwang Liu, Fuchun Sun, and Kunlun He. 2024. A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(12):9456–9478.

Ruotong Liao, Xu Jia, Yangzhe Li, Yunpu Ma, and Volker Tresp. 2024. Gentkg: Generative forecasting on temporal knowledge graph with large language models. In *Findings of the Association for Computational Linguistics: NAACL 2024, Mexico City, Mexico, June 16-21, 2024*, pages 4303–4317. Association for Computational Linguistics.

Yushan Liu, Yunpu Ma, Marcel Hildebrandt, Mitchell Joblin, and Volker Tresp. 2022. Tlogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 4120–4127. AAAI Press.

Ao Lv, Yongzhong Huang, Guige Ouyang, Yue Chen, and Haoran Xie. 2024. Rlgnet: Repeating-local-global history network for temporal knowledge graph reasoning. *CoRR*, abs/2404.00586.

Yunshan Ma, Chenchen Ye, Zijian Wu, Xiang Wang, Yixin Cao, and Tat-Seng Chua. 2023. Context-aware event forecasting via graph disentanglement. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, Long Beach, CA, USA, August 6-10, 2023*, pages 1643–1652. ACM.

Shi Mingcong, Chunjiang Zhu, Detian Zhang, Shiting Wen, and Qing Li. 2024. Multi-granularity history and entity similarity learning for temporal knowledge graph reasoning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024*, pages 5232–5243. Association for Computational Linguistics.

Mehrnoosh Mirtaheri, Mohammad Rostami, and Aram Galstyan. 2023. History repeats: Overcoming catastrophic forgetting for event-centric temporal knowledge graph completion. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 7740–7755. Association for Computational Linguistics.

Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, volume 10843 of *Lecture Notes in Computer Science*, pages 593–607. Springer.

Chao Shang, Yun Tang, Jing Huang, Jinbo Bi, Xiaodong He, and Bowen Zhou. 2019. End-to-end structure-aware convolutional networks for knowledge base completion. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 3060–3067. AAAI Press.

Ruihao Shui, Yixin Cao, Xiang Wang, and Tat-Seng Chua. 2023. A comprehensive evaluation of large language models on legal judgment prediction. In *Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6-10, 2023*, pages 7337–7348. Association for Computational Linguistics.

Zhen Wang, Xi Zhou, Yating Yang, Bo Ma, Lei Wang, Rui Dong, and Azmat Anwar. 2025. Openforecast: A large-scale open-ended event forecasting dataset. In *Proceedings of the 31st International Conference on Computational Linguistics, COLING 2025, Abu Dhabi, UAE, January 19-24, 2025*, pages 5273–5294. Association for Computational Linguistics.

Yuwei Xia, Ding Wang, Qiang Liu, Liang Wang, Shu Wu, and Xiao-Yu Zhang. 2024. Chain-of-history reasoning for temporal knowledge graph forecasting. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024*, pages 16144–16159. Association for Computational Linguistics.

Yi Xu, Junjie Ou, Hui Xu, and Luoyi Fu. 2023. Temporal knowledge graph reasoning with historical contrastive learning. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 4765–4773. AAAI Press.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Shengzhe Zhang, Wei Wei, Rikui Huang, Wenfeng Xie, and Dangyang Chen. 2024. Modeling historical relevant and local frequency context for representation-based temporal knowledge graph forecasting. In *Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA, November 12-16, 2024*, pages 7675–7686. Association for Computational Linguistics.

A Appendix

The appendix offers additional materials that further substantiate our approach. Section A.1 discusses reverse prediction scenarios in event forecasting via manual inversion of relationship descriptions. Section A.2 presents a case study of our self-supervised relevant information filter, demonstrating its effectiveness in selecting pertinent historical events. Section A.3 visualizes the impact of key hyperparameters—specifically GNN depth and history window length—thereby elucidating essential design principles. Section A.4 reports comprehensive ablation experiments that confirm the importance of the SEI, PREI, and RHI modules, as well as the utility of the self-supervised filter and event context. Section A.5 extends our evaluation to temporal knowledge graph forecasting on the ICEWS14 and ICEWS18 benchmarks, highlighting the robustness of our dual-criteria constraint strategy. Finally, Section A.6 analyzes reasoning time consumption.

A.1 Handling of Relations in Reverse Prediction Scenarios

In event forecasting, handling reverse prediction scenarios is crucial due to the bidirectional nature

Query	Cairo Engage in diplomatic cooperation Whom? at 2356th day.
History Repeat Information	Cairo Engage in diplomatic cooperation South Sudan at 2351th day. Cairo Engage in diplomatic cooperation Hamas at 2349th day. Cairo Engage in diplomatic cooperation Israel at 2349th day. Cairo Engage in diplomatic cooperation Gaza at 2334th day. ...
Filtered Relevant Information	Egypt Engage in diplomatic cooperation Libya at 2354th day. Israel Cooperated with diplomatically by Cairo at 2349th day. Egypt Engage in diplomatic cooperation Qatar at 2353th day. Gaza Cooperated with diplomatically by Cairo at 2334th day. ...
Answer	**Qatar**

Table 7: Case study to validate the effectiveness of ITHI for retrieving historical events.

of relationships between entities. Since relationships are not always symmetrical, predicting inverse relations can be challenging.

To address this, we process both (s, r, ?) and (o, r', ?) queries by manually inverting existing relationship descriptions. For example, "A expresses intent to cooperate with B" can be reversed to "A received an expression of intent to cooperate from B." This inversion allows the model to account for both directions of interaction, improving its ability to predict events in reverse scenarios.

A.2 Relevant Information Filter Case Study

Table 7 presents another example comparing "History Repeat Information" and "Filtered Relevant Information" results. The comparison demonstrates that the self-supervised semantic filter more effectively identifies historical information pertinent to queries. Specifically, the former includes broad information on Cairo's diplomatic cooperation with various countries or organizations, making it difficult to answer the query "Cairo Engage in diplomatic cooperation with Whom? on the 2356th day". The latter focuses on specific diplomatic interactions between Cairo and entities like Egypt, Libya, and Israel, closely aligning with the query and providing more targeted support for predictions, thereby significantly enhancing event prediction accuracy and relevance.

A.3 Study of the visualization of hyperparameter selection in ITHI

Our ablation study systematically evaluates the impact of two architectural parameters: 1) The depth of graph neural networks ($n = \{1, 2, 3\}$); 2) The temporal span of historical graphs ($l = \{1, 3, 7\}$). The results as shown in Figure 4. For the EG and IR datasets, the best performance is achieved when using a shorter sequential history window ($l = 1$) and three GNN layers ($n = 3$) to effectively capture event correlation. For the IR dataset, the best performance was achieved when using a shorter sequential history window ($l = 1$) and two GNN layers ($n = 2$). Further, when the sequential history window is 1, extending the number of layers of the GNN further ($n : 2 \rightarrow 3$) reduces the performance. In general, a shorter continuous history window effectively filters out noise, enabling the model to focus on relevant information. In addition, the results of all trials indicate that the single GNN layer performs worst, likely due to its inability to capture deep structural dependencies. This analysis highlights key design principles for this task: while the optimal GNN depth is dataset-specific, single-layer GNNs are generally insufficient for modeling structural relationships, and shorter history windows are preferred over longer ones.

Additionally, we investigate how the number of relevant historical events, selected by the semantic event filter, impacts event prediction performance on the EG dataset (Figure 5). We conduct experiments with Top-N values of 60, 80, 90, and 100. The results reveal that forecasting performance declines as the number of semantic filtering samples diminishes, indicating that event semantic constraints alone are insufficient and further confirming the synergistic effectiveness of dual constraints.

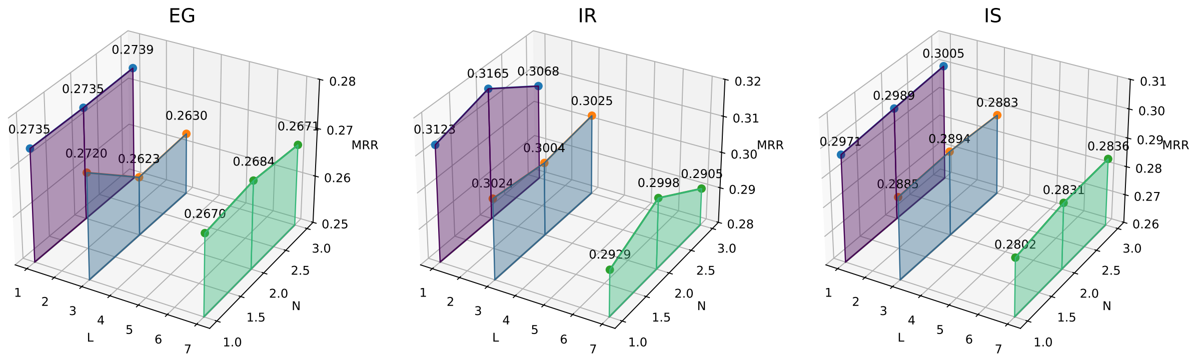

Figure 4: Visualization results for two critical modules explored on three datasets, EG, IR, and IS, where L denotes the number of search days for selecting successive events, and N represents the number of layers in the GNN.

Model	EG				IR				IS			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
ITHI	**27.39**	**19.03**	**30.00**	**43.80**	**31.65**	**23.21**	**33.27**	**49.35**	**30.05**	**22.84**	**30.13**	**46.10**
w/o SEI	26.82	18.61	29.42	43.06	28.68	20.57	29.94	46.39	28.05	20.95	28.60	44.91
w/o PREI	26.49	17.96	29.43	43.00	29.98	21.59	31.61	47.50	28.57	21.23	29.03	45.16
w/o RHI	11.34	2.70	12.13	29.75	12.88	4.20	13.64	31.41	11.86	3.68	12.56	29.40
w/o filter	24.34	15.87	26.55	41.48	27.89	18.90	29.81	47.03	26.82	18.00	28.74	43.11
w/o context	25.50	17.48	27.65	41.63	27.50	18.35	29.47	46.33	26.10	17.81	27.94	42.84

Table 8: Results of complete ablation experiments, reported in percentages. "w/o SEI" indicates the removal of the sequential event information module; "w/o PREI" denotes the omission of the periodically repeated event information module; "w/o RHI" represents the exclusion of the relevant historical information module; "w/o filter" signifies the removal of the self-supervised filter within the RHI module; and "w/o context" means that event contextual information is not used during RHI selection.

A.4 Complete ablation experiments

The ablation experiments (Table 8) reveal that eliminating the SEI, PREI, or RHI modules induces a significant performance drop, underscoring the crucial role of sequential, repetitive, and relevant historical event information for accurate event prediction. Notably, removing relevant historical information (w/o RHI) caused significant declines in Hit@3 and Hit@10 across all datasets, with Hit@10 on the EG dataset dropping from 43.80% to 29.75%, highlighting the importance of relevant historical information. Moreover, removing either the self-supervised filter or event contextual information within the RHI module leads to additional performance degradation. This indicates that both the self-supervised semantic filtering mechanism and the integration of rich event semantic embeddings facilitate the effective selection of pertinent historical events. Overall, these findings validate the effectiveness of the dual-criteria constraint strategy employed in ITHI.

A.5 Extension experiments on the task of temporal knowledge graph forecasting

To validate our approach on purely structured knowledge raph forecasting, we evaluated LogCL(Chen et al., 2024) combined with our self-supervised RHI-mining strategy on the standard ICEWS14 and ICEWS18 benchmarks (Detailed statistics analysis of the Datasets is shown in Table 3). Due to the absence of explicit event contexts in these datasets, we randomly initialize each event's context and incorporate semantic information from its subject and object. Dual-criteria constraints capture latent inter-event correlations for temporal knowledge graph forecasting.

Specifically, we utilize the GTE-base model for semantic embedding and set the initial Top-n to 100. Experimental hyperparameters are kept consistent with LogCL: a feature dimension of 200; a learning rate of 0.001; 50 decoder kernels across all datasets with a kernel size of 2×3; and a dropout rate of 0.2. Experimental results in Table 9 are

Model	ICEWS14				ICEWS18			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
GenTKG (Liao et al., 2024)	–	36.85	47.95	53.50	–	24.25	37.25	42.10
TiRGN + COH (Xia et al., 2024)	43.94	33.07	49.64	64.90	32.98	21.83	37.79	54.92
MGESK (Mingcong et al., 2024)	45.88	35.43	51.54	65.70	34.18	23.66	38.64	54.89
LogCL (Chen et al., 2024)	48.87	37.76	54.71	70.26	35.67	24.53	40.32	57.74
+ Ours	**49.65**	**38.45**	**55.68**	**71.12**	**36.30**	**25.11**	**40.97**	**58.54**

Table 9: Extension of experimental results in temporal knowledge graph forecasting task (%).

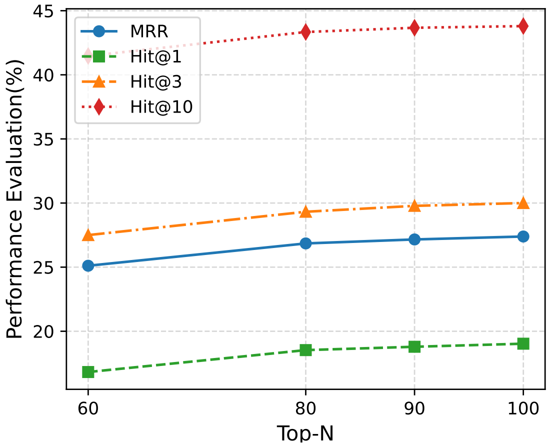

Figure 5: Experiment on the effect of initial number of events on forecasting performance for semantic filtering of events in EG dataset

model	RE-GCN	SeCoGD	ITHI
reasoning time spent	17.48	156.4	162.22

Table 10: Reasoning time spent experiments in seconds.

averaged over three trials. Results indicate that coupling the LogCL method with our dual-criteria constraint strategy yields consistent performance gains. On the ICEWS14 dataset, the MRR increases from 48.87% to 49.65% and Hit@1 from 37.76% to 38.45%, with corresponding improvements in Hit@3 and Hit@10. On the ICEWS18 dataset, MRR rises from 35.67% to 36.30% and Hit@10 from 57.74% to 58.54%. These findings validate that integrating event semantics for mining latent inter-event information reliably enhances performance in both event forecasting and temporal knowledge graph forecasting tasks.

A.6 Time consumption analysis experiment

We compared the time consumption of RE-GCN, SeCoGD, and ITHI on EG dataset using 10000 samples (Table 10). There are two observations to be made: (1) RE-GCN's time efficiency far surpasses other methods. This stems from its ability to acquire representations of all same-timestamp queries in bulk during learning, unlike query-by-query methods. (2) Only slightly less time-consuming than ITHI, SeCoGD categorizes topics into multiple types (5 types in SeCoGD), using separate RE-GCN for each category and hypergraphs for cross-category interactions. Like ITHI, it uses multiple RE-GCN for graph modeling. However, ITHI performs significantly better on multiple datasets and metrics. We think the time consumption brought about is acceptable.

Graph + LLM Prompt

You need to predict the missing event object {Whom} in the query based on the given event context, query information, and the corresponding list of candidate answers. The query follows the format "{subject} {relation} {whom}? at {Time}." The 150 candidate answers you received corresponding to the current query have been sorted from highest to lowest probability. You need to prioritize the most likely answers at the top and keep the rest in the same order as much as possible, considering the query and event context.
Output example:
1: xxx\n2: xxx\n3: xxx\n...\n150: xxx\n
You will now receive the background context of an event, the query, and the corresponding 150 results. The candidate answers are sorted by probability (high to low).
Event background: egyptian minister accuses ethiopia of intransigence over gerd\nQuery: Egypt Express intent to cooperate Whom? at 2326th day.
Candidate Whom (probability sorted from high to low):
1: Ethiopia\n2: Sudan\n3: Saudi Arabia\n4: Addis Ababa\n5: Iraq\n6: Bahrain\n7: Egypt\n8: Greece\n9: China\n10: Khartoum\n11: Chinese\n12: Business\n13: Africa\n14: Israel\n15: United Kingdom\n16: Kuwait\n17: Jordan\n18: United States\n19: Ethiopian\n20: Djibouti\n21: Egyptian\n22: Sudanese\n23: Saudi\n24: Libya\n25: Russia\n26: Qatar\n27: South Sudan\n28: President\n29: Cyprus\n30: Lebanon\n31: Unsc\n32: United Arab Emirates\n33: Kenya\n34: Minist\n35: Washington\n36: Yemen\n37: Iran\n38: Company\n39: Oman\n40: African Union\n41: Algeria\n42: Geneva\n43: Britain\n44: The Eu\n45: Somalia\n46: Tunisia\n47: Companies\n48: Government\n49: Military\n50: France\n51: Cairo\n52: Italy\n53: Ministry\n54: United Nations\n55: French\n56: Pakistan\n57: The Un\n58: Syria\n59: Greek\n60: Turkey\n61: Israeli\n62: Arab League\n63: British\n64: Germany\n65: World Bank\n66: Japan\n67: Amman\n68: The Us\n69: Libyan\n70: Bank\n71: Russian\n72: Foreign Minist\n73: Syrian\n74: Italian\n75: Yemeni\n76: Uganda\n77: Hamas\n78: Moscow\n79: Sweden\n80: Prison\n81: Gaza\n82: Lebanese\n83: Qatari\n84: Palestinian\n85: International Monetary Fund\n86: Azerbaijan\n87: Muslim\n88: Switzerland\n89: Deputy\n90: Paris\n91: Tanzania\n92: Iraqi\n93: Journalist\n94: Morocco\n95: Kingdom\n96: Australia\n97: Administration\n98: Islamic\n99: Citizen\n100: Cabinet\n101: Eritrea\n102: Baghdad\n103: The European Union\n104: Serbia\n105: Canada\n106: Industry\n107: Belarus\n108: University\n109: Indonesia\n110: Ghana\n111: Ankara\n112: Prime Minister\n113: Common Market For Eastern And Southern Africa\n114: Sinai\n115: Nigeria\n116: Kenyan\n117: Romania\n118: Turkish\n119: Palestine\n120: Religion\n121: Protester\n122: Dubai\n123: Media\n124: Europe\n125: Irrigation Minist\n126: Jordanian\n127: Unesco\n128: Student\n129: South Sudanese\n130: Field Marshal\n131: Jewish\n132: Omar Al Bashir\n133: Professor\n134: Armenia\n135: Hungary\n136: Terrorist Group\n137: Commander\n138: Obama\n139: Istanbul\n140: Iranian\n141: Somali\n142: Population\n143: Vietnam\n144: The Au\n145: Minist Of Electricity\n146: Singapore\n147: Sunni\n148: Representatives\n149: Amr Moussa\n150: Reuters
Please prioritize the most likely results while preserving the original order as much as possible. Ensure that all thirty candidates are presented in the specified format, avoiding omissions or repetitions. Output only the adjusted results without any additional content.\n

Figure 6: Example of Graph collaborating with LLM for a prompt.

Graph + LLM + COH Prompt

You need to predict the missing event object {Whom} in the query based on the given event context, query information, historic events, and the corresponding list of candidate answers. The query follows the format "{subject} {relation} {whom}? at {Time}." The 100 candidate answers you received corresponding to the current query have been sorted from highest to lowest probability. You need to prioritize the most likely answers at the top and keep the rest in the same order as much as possible, considering the query and event context.
Output example:\n1: xxx\n2: xxx\n3: xxx\n...\n100: xxx
You will now receive the background context of an event, the query, and the corresponding 100 results. The candidate answers are sorted by probability (high to low).
Query: Egypt Express intent to cooperate Whom? at 2326th day.
30 candidate events:\n1: Egypt Express intent to cooperate Iraq at 2301th day.\n\n2: Egypt Express intent to cooperate Sudan at 2308th day.\n\n3: Egypt Express intent to cooperate Libya at 2316th day.\n\n4: Egypt Express intent to cooperate Israeli at 2310th day.\n\n5: Egypt Express intent to cooperate Jordan at 2323th day.\n\n6: Egypt Express intent to cooperate The International Community at 2314th day.\n\n7: Egypt Express intent to cooperate Bank at 2303th day.\n\n8: Egypt Express intent to cooperate Sudan at 2298th day.\n\n9: Egypt Express intent to cooperate Qatar at 2308th day.\n\n10: Egypt Express intent to cooperate Company at 2317th day.\n\n11: Egypt Express intent to cooperate Sudan at 2306th day.\n\n12: Egypt Express intent to cooperate Prosecutor at 2324th day.\n\n13: Egypt Express intent to cooperate Ethiopia at 2308th day.\n\n14: Egypt Express intent to cooperate Hamas at 2300th day.\n\n15: Egypt Express intent to cooperate Saudi Arabia at 2305th day.\n\n16: Egypt Express intent to cooperate Commander at 2316th day.\n\n17: Egypt Express intent to cooperate Ethiopia at 2321th day.\n\n18: Egypt Express intent to cooperate Palestine at 2317th day.\n\n19: Egypt Express intent to cooperate Mayor at 2315th day.\n\n20: Egypt Express intent to cooperate Sudan at 2313th day.\n\n21: Egypt Express intent to cooperate Sudan at 2306th day.\n\n22: Egypt Express intent to cooperate Greece at 2314th day.\n\n23: Egypt Express intent to cooperate Bahrain at 2308th day.\n\n24: Egypt Express intent to cooperate Zawahiri at 2313th day.\n\n25: Egypt Express intent to cooperate Ethiopian at 2324th day.\n\n26: Egypt Express intent to cooperate Sudan at 2302th day.\n\n27: Egypt Express intent to cooperate Saudi Arabia at 2305th day.\n\n28: Egypt Express intent to cooperate Iraq at 2301th day.\n\n29: Egypt Express intent to cooperate Prime Minister at 2324th day.\n\n30: Egypt Express intent to cooperate Qatari at 2306th day.
100 candidate events:\n1: Sudan\n2: Ethiopia\n3: Israel\n4: Russia\n5: Greece\n6: Saudi Arabia\n7: Cyprus\n8: Jordan\n9: Egypt\n10: Hamas\n11: Turkey\n12: France\n13: United States\n14: Palestinian\n15: Egyptian\n16: United Kingdom\n17: Iran\n18: Syria\n19: China\n20: Khartoum\n21: President\n22: Gaza\n23: Libya\n24: Russian\n25: Bahrain\n26: Qatar\n27: United Arab Emirates\n28: Iraq\n29: Washington\n30: Tunisia\n31: Japan\n32: Germany\n33: Italy\n34: Saudi\n35: Lebanon\n36: Yemen\n37: Kuwait\n38: Israeli\n39: The Us\n40: French\n41: World Bank\n42: Libyan\n43: Business\n44: Djibouti\n45: Government\n46: Cairo\n47: Islamic\n48: International Monetary Fund\n49: Military\n50: Syrian\n51: Company\n52: Uganda\n53: Africa\n54: Ethiopian\n55: Chinese\n56: Oman\n57: Britain\n58: Algeria\n59: West Bank\n60: Addis Ababa\n61: Morocco\n62: Arab League\n63: Bank\n64: Minist\n65: Parliament\n66: Greek\n67: United Nations\n68: Kingdom\n69: Moscow\n70: Palestine\n71: Azerbaijan\n72: Paris\n73: Italian\n74: Tel Aviv\n75: Muslim\n76: The White House\n77: Brazil\n78: Prime Minister\n79: Representatives\n80: Sudanese\n81: American\n82: Riyadh\n83: Switzerland\n84: Investor\n85: Army\n86: Geneva\n87: New York\n88: German\n89: South Korea\n90: South Sudan\n91: The European Union\n92: Somalia\n93: Congress\n94: Media\n95: The Eu\n96: Pakistan\n97: British\n98: Tourist\n99: Jerusalem\n100: Kenya
Please prioritise the most likely candidates based on the above historical events and queries, whilst retaining the original order as far as possible. Ensure that all 100 candidates are presented in the specified format to avoid omissions or duplicates. Output only the adjusted results without any additional content.\n

Figure 7: Example of a Graph with LLMs collaborating via Chain-Of-History for a prompt.