Improving Temporal Knowledge Graph Reasoning with Hierarchical Semantic-Aware Contrastive Learning

Renning Pang[1], Yao Liu[1], Yanglei Gan[1], Tingting Dai[1], Yashen Wang[2], Xiaojun Shi[2], Tian Lan[1] ✉, and Qiao Liu[1]

[1] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China
{prn, yangleigan}@std.uestc.edu.cn,
{liuyao, lantian, qliu}@uestc.edu.cn,
ttdai_18@outlook.com
[2] Artificial Intelligence Institute of CETC, Beijing, China
{yswang1, XJSHI258}@outlook.com

Abstract. Temporal Knowledge Graph (TKG) reasoning seeks to infer the future evolution of incomplete facts from observed historical data. Although supervised contrastive learning has recently enhanced query representations for TKG reasoning, two critical challenges remain. First, current methods uniformly treat all negative samples, overlooking their semantic and temporal correlations. Second, these approaches do not fully exploit the hierarchical relationships between fine-grained events and higher-level event categories, thereby missing crucial event taxonomies. To address these limitations, we propose a Hierarchical Semantic-aware Contrastive Learning (HSCL) framework. Specifically, our Instance-level objective introduces a dynamic adaptive weighting mechanism that differentiates negative samples based on semantic similarity, while our Categ ory-level objective incorporates ontology-guided clustering to represent hierarchical event semantics. This dual-level design encourages cohesive embeddings within the same event category and clear separation across different categories. Extensive experiments on four real-world benchmarks demonstrate that HSCL consistently outperforms state-of-the-art baselines[1].

Keywords: Temporal knowledge graph · Contrastive learning · Graph Neural Network.

1 Introduction

Temporal Knowledge Graphs (TKGs) extend traditional Knowledge Graphs (KGs) by incorporating temporal information, associating event facts with timestamps or time intervals to indicate their validity periods. Unlike static KGs,

[1] The code is available at https://github.com/AONE-NLP/TKGR-HSCL

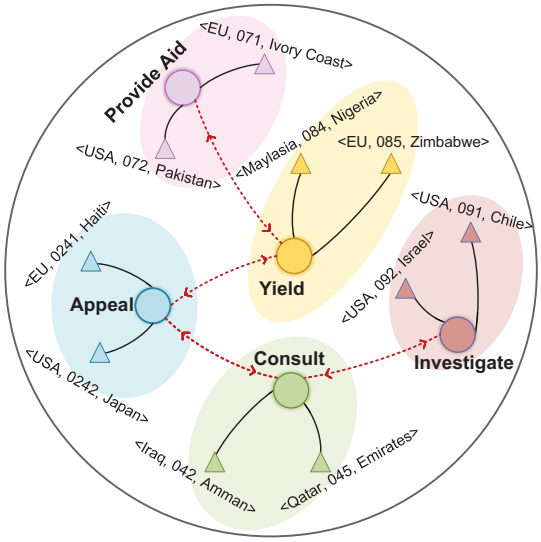


Fig. 1. An illustrative example highlighting the importance of capturing the hierarchical semantics related to the query. Event codes have been replaced with corresponding numbers from the codebook to emphasize the structured categorization. △ denotes fine-grained events, while ◯ denotes coarse-grained events.

which represent facts as triples (subject, relation, object), TKGs add a temporal dimension, forming quadruples $< s, r, o, t >$, where t specifies when the fact holds, such as $< USA, negotiate, Japan, 2019-6-29 >$.

Reasoning over TKGs involves inferring missing facts based on observed quadruples within a specified time range $[t_0,\ t_T]$. It is categorized into interpolation [2, 3, 11], which infers missing facts within the historical range ($t_0 \leq t \leq t_T$), and extrapolation [28, 32, 7], which predicts facts beyond the observed range ($t > t_T$). This study focuses on the extrapolation setting, emphasizing its forward-looking significance.

Accurately predicting future facts requires a thorough understanding of the developmental patterns embedded within historical facts. Existing research primarily focuses on aggregating structural information from adjacent entities and integrating temporal dynamics to construct historical representations of entities and relations [32, 37, 38]. These representations are then utilized within scoring functions, such as ConvTransE [52], to estimate the likelihood of future event facts. Building upon these foundational methods, recent advancements have introduced contrastive learning to enhance TKG reasoning by leveraging the discriminative capability of contrasting diverse types of information, such as local and global historical contexts [39], or historical and non-historical dependencies [36]. These techniques aim to refine feature representations by emphasizing the distinctions between positive and negative samples.

Despite the significant progress of contrastive learning in TKG reasoning, these methods exhibit inherent limitations. ***First, they uniformly treat all***

negative samples, failing to capture essential semantic and temporal correlations among queries [39, 36]. This uniform treatment overlooks critical distinctions necessary for optimal feature representation and contradicts the alignment principle of feature representations [42, 44, 43], which aims to position semantically similar samples closer together. As illustrated in Figure 1, events within the same coarse-grained category are treated uniformly as negative samples, ignoring their internal semantic and temporal relationships and potentially leading to suboptimal representations. ***Second, existing approaches fail to account for the hierarchical relations between fine-grained events and their broader categories***, as defined in the widely used CAMEO ontology, thereby preventing the model from capturing hierarchical semantic structures [48, 50]. Consider the event "*Investigate*" in Figure 1, which encompasses the fine-grained events "*091*" ("*Investigate crime, corruption*") and "*092*" ("*Investigate human rights abuses*"). While these fine-grained events exhibit distinct characteristics, they share the broader semantic context of "*Investigate*". We argue that incorporating these hierarchical semantics into contrastive learning objective would enable capture of latent semantic, resulting in more robust and generalized representations.

To bridge these gaps, we propose a Hierarchical Semantic-aware Contrastive Learning (HSCL) framework for TKG reasoning. Specifically, HSCL integrates short and long term sequence encoders with an entity-aware attention mechanism to capture query-specific temporal dynamics. Additionally, we design a hierarchical contrast module for short and long term queries. The instance-level contrastive learning refines representations by directly comparing individual event embeddings, distinguishing semantically similar and dissimilar events, and enabling the model to discern fine-grained differences between queries and their contexts; meanwhile, the category-level contrastive learning captures hierarchical relationships between detailed events and broader categories, encouraging events of the same type to cluster while keeping distinct types separated. This higher-level grouping enhances the generalization of the representations and supports the model in inferring broader patterns and relationships. The contribution of our work are three-fold:

- We propose two optimization objectives to maximize the mutual information of query-specific information in different granularities.
- We introduce a hierarchical contrastive learning framework in which an instance-level objective distinguishes subtle differences between events, while a category-level objective enforces hierarchical grouping into broader categories, enhancing generalization and robustness.
- Evaluations across four benchmark datasets highlight the effectiveness of the proposed method compared to the current state-of-the-art baselines in terms of reasoning accuracy and representation capability.

2 Related Work

Temporal Knowledge Graph (TKG) reasoning seeks to predict facts at future timestamps by modeling observed historical facts. This task typically operates under two settings interpolation and extrapolation with our primary focus on the extrapolation scenario. Existing approaches mainly concentrate on either learning temporal patterns [55, 32] or capturing structural information [26, 34] to enrich entity and relation representations. RE-NET [55] models long-term historical interactions of target entities as sequences, integrating a Recurrent Neural Network (RNN) and a relational Graph Convolutional Network (GCN) to capture both temporal and structural dependencies. RE-GCN [32] similarly encodes temporal and relational dependencies among entities, leveraging graph convolutions to capture complex interactions. L^2TKG [40] exploits both intra-time relations (co-occurring entities at the same timestamp) and inter-time relations (entities appearing at different timestamps). DaeMon [57] adaptively retrieves temporal path information between query subjects and object candidates over time without relying on entity representations. TiPNN [56] constructs a unified history temporal graph to comprehensively represent past information, using query-aware temporal paths for historical path modeling. In addition, Tlogic [58] enhances interpretability through a symbolic framework built on temporal logical rules extracted via temporal random walks. Despite these achievements, existing methods often have shortcomings when modeling both long and short term temporal dependencies, as they frequently overlook explicit dependencies among distinct entities across different timestamps in long-term histories.

To address above limitation, TiRGN [35] proposes a global history encoder to identify entities associated with historical or unseen facts, complemented by a time-guided decoder. HGLS [38] directly connects the same entity at each timestamp to capture long and short term dependencies. CRAFT [41] leverages candidate-specific historical context via dual-attention-enhanced path encoding for long-term relevance, while using frequency-based contextual modeling for short-term repetitive patterns. Similarly, MGESL [60] incorporates a multi-granularity history module, capturing long-term coarse-grained multi-hop history alongside short-term fine-grained repetitive patterns. Building on these strategies, recent work integrates contrastive learning to merge historical and non-historical dependencies for enhanced temporal reasoning. CENET [36] applies historical contrastive learning in temporal knowledge graphs to improve entity prediction and event reasoning. LogCL [39] unifies local and global temporal embeddings through a contrastive loss function, ensuring robust historical context fusion. Dejavu [21] proposes a unified PLM-based framework that employs contrastive learning to model historical contextual information, effectively balancing temporal information and textual knowledge.

Despite these advances, existing contrastive learning methods often treat all negative samples uniformly, overlooking crucial semantic and temporal correlations that may arise across different queries. Moreover, they fail to leverage the hierarchical relationships between fine-grained events and their broader categories, thus limiting their ability to capture nuanced event structures.

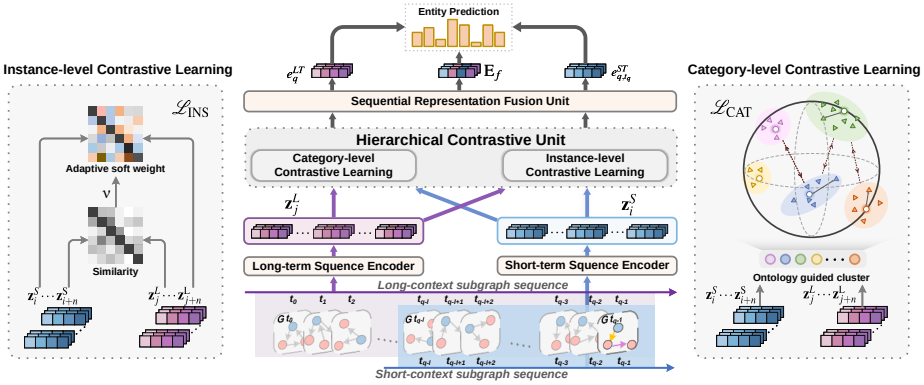


Fig. 2. Model overview of HSCL. Given a query $< s, r, ?, t{+}\Delta t >$, HSCL employs long and short term sequence encoders to capture query-specific temporal dynamics. The instance-level contrastive module refines representations by distinguishing subtle differences between individual events, while the category-level module organizes events into broader semantic groups. then the sequential representation fusion unit integrates the long and short term representations via a cross-attention to generate the final probability distribution over entities.

3 Methods

In this section, we detail our proposed framework, as shown in Figure 2 We begin by formulating the preliminaries of Temporal Knowledge Graph (TKG) reasoning. A TKG is formally represented as a sequence of snapshots $G = \{G_1, G_2, \ldots, G_{t-1}\}$, where each snapshot G_{t_i} encompasses all quadruples that occur at timestamp t_i. Each fact within a snapshot is denoted as a quadruple (e_s, r, e_o, t_i), indicating that the subject entity e_s interacts with the object entity e_o through relation r at time t_i. To ensure comprehensive representation [54], inverse relation quadruples (e_s, r^{-1}, e_o, t_i) are incorporated into G. The aim of TKG reasoning is to predict either the missing object $(e_s, r, ?, t)$ or subject $(?, r, e_o, t)$ at future timestamp by leveraging the historical snapshots $\{G_{t-l}, G_{t-l+1}, \ldots, G_{t-1}\}$ spanning the preceding l timestamps.

3.1 Sequential Representation Learning

To comprehensively capture the diverse temporal dynamics inherent in TKGs, we propose modeling both long-term and short-term representations of historical quadruple sequences. Long-term representations provide the essential context required to understand the evolution of entity relationships and interactions over extended periods. In contrast, short-term representations focus on recent interactions and transient relationships, which may indicate emerging trends or temporary states within the TKG.

Long-term Sequence Encoder. To model the long-term temporal dynamics, we adopt the R-GCN [9] as the encoder to capture historical subgraphs information. Given a query $(e_q, r_q, ?, t_q)$ and its corresponding temporal knowledge

graph $G^{t_q} = \{G_1, G_2, \ldots, G_{t_q-1}\}$, we first extract all one-hop historical facts associated with the query. Subsequently, for each of these target entities, we expand the sub-graph to include two-hop relational patterns to model historical long-term temporal information. The aggregated representation obtained from this expanded structure is then used for downstream processing.

To effectively encode the structural and semantic information embedded in this sub-graph, we update the entity embeddings at layer l as follows:

$$e_o^l = \sigma_1 \left(\frac{1}{c_o} \sum_{(e_s,r)\in\mathcal{E}_o} \mathrm{W}_1 \left({e_s}^{l-1} + r \right) \right) \tag{1}$$

Where $(e_s, r) \in \mathcal{E}_o$ represents the neighbors connected to the target node via edge c_o, and $\mathrm{W}_1 \in R^{d\times d}$ is the trainable parameter matrix at layer l layer for feature aggregation. σ_1 denotes the ReLU activation function, and e_o^l represents weighted summation of the neighboring entitys and relations by R-GCN at layer $l-1$. And for simplicity, the output of the R-GCN at the final layer L is represented as e^L.

To further exploit the long-term facts relevant to queries, we deploy an attention aggregation mechanism, as follows:

$$e_q^{LT} = \sigma_2(W_2 e^L + e') \cdot e^L \tag{2}$$

Where $W_2 \in R^{d\times d}$ is learnable weight matrix, while σ_2 denotes the RReLU activation function. e_q^{LT} denotes the final long-term representations with respect to the query.

Short-term Sequence Encoder. To capture the recent interactions and transient relationships with respect to the query. At timestamp t and in context c, the representation message $e_o^l \in R^d$ received by an object o at the l-th layer of graph propagation is defined as:

$$e_{o,t}^l = \sigma_2 \left(\frac{1}{c_o} \sum_{(s,r),\exists(s,r,o)\in\mathcal{E}_t} \mathrm{W}_3 \left(e_{s,t}^{l-1} + r \right) + \mathrm{W}_4 e_{o,t}^{l-1} \right) \tag{3}$$

where $\mathcal{E}_t$ denotes the set of edges for timestamp t; $e_{s,t}^{l-1}$ is the representation of a neighboring entity e_s in $(l-1)$-th layer at timestamp t. W_3,$\mathrm{W}_4 \in R^{d\times d}$ are learnable parameters for feature aggregation and self-loop connections. For simplicity, the output of the R-GCN at the final layer L is denoted as e_t^L.

To update entity representations over time, we employ a Gated Recurrent Unit (GRU) [4] to model short-term sequence dynamics:

$$\mathrm{E}_{t+1} = GRU(e_t, e_t^L) \tag{4}$$

Additionally, we apply a time gate unit over the relation embedding at $t-1$ to obtain the updated relation representation:

$$\mathrm{R}_t = \mathrm{U}_t \cdot \mathrm{R}_t^w + (1 - \mathrm{U}_t) \cdot \mathrm{R}_{t-1}; \tag{5}$$

$$\mathrm{U}_t = \sigma_3(\mathrm{W}_5\mathrm{R}_t^w + \mathrm{b}) \tag{6}$$

Where $\mathrm{U}_t \in R^{d\times d}$. σ_3 denotes sigmoid function, W_5 is learnable weight matrix. R_t^w denotes the relation representation at time t, calculated by aggregating the representations of entities interacting with the relation at time t:

$$\mathrm{R}_t^w = Meanpool(\mathrm{E}_{s,r}) + r, \forall e_s \in \mathcal{N}_r^t \tag{7}$$

where $\mathrm{E}_{s,r}$ is the embedding associated with r, $Meanpool(\cdot)$ aggregates the embeddings of entities in the set $\mathcal{N}_r^t$ that interact with relation r at time t.

Subsequently, we refine the entity embeddings using an attention mechanism to emphasize the importance of query-relevant KG snapshots. For a query at timestamp t_q, the query-specific features representation e_{t_q} is calculated as:

$$e_{t_q} = \mathrm{W}_6 \left[Meanpool(r_{t_q}) \parallel e_q\right] \tag{8}$$

where e_q denotes query entity, r_{t_q} denotes relations associated with e_{t_q} ,W_6 is learnable weight matrix, and $\parallel$ represents the concatenation operation.

For entity and relation within a time window $T_{\text{win}} = \{\tau \mid \tau \in [t_q - (p-1) \dots t_q]\}$, we apply an attention mechanism to weight the importance of query-relevant features:

$$e_{q,t_q}^{ST} = e_{t_q} + \sum_{\tau \in T_{\text{win}} \setminus \{t_q\}} \delta e_\tau \delta = \sigma_3 \left(\mathrm{W}_7 \left(e_\tau^L + e_{t_q}\right)\right), \tau \in T_{\text{win}} \setminus \{t_q\} \tag{9}$$

Where e_{q,t_q}^{ST} is the final representation of the short-term sequence, δ denotes attention scores for historical snapshots within T_{win}, then σ_3 is the Softmax function, and W_7 is learnable weight matrix. e_τ is representation of the entity associated with the query at historical timestamp τ.

3.2 Hierarchical Contrastive Learning

To improve the discriminative power and generalization of the embeddings, we propose a hierarchical contrastive learning unit comprising two complementary objectives: instance-level and category-level contrastive learning. These objectives address the challenges of capturing semantic correlations and hierarchical relationships, respectively.

Instance-level Contrastive Learning. To overcome the uniform treatment of negative samples, we incorporate a dynamic adaptive weighting mechanism into the instance-level contrastive learning objective. This mechanism differentiates negative samples based on semantic relations, thereby improving the model's ability to learn informative gradients. The short-term contextual embeddings e_i^S and the long-term contextual embeddings e_j^L are first projected into the same latent space using a shared projection model:

$$\mathbf{z}_i^{\mathrm{S}} = \mathrm{Norm}(\mathrm{Proj}(e_{q,t_q}^{ST})); \quad \mathbf{z}_j^{\mathrm{L}} = \mathrm{Norm}(\mathrm{Proj}(e_q^{LT})) \tag{10}$$

where Proj(·) is multi-layer perceptron (MLP), and Norm(·) denotes L_2 normalization. The adaptive soft weight $\mathcal{V}$ learned by 2-layer MLP accoding to the samples under the long and short term squences, formulated as:

$$\mathcal{V} = \sigma_3(W_8(\sigma_4(W_9 \cdot [\mathbf{z}_i^{\mathrm{S}} \parallel \mathbf{z}_j^{\mathrm{L}}] + b_1)) + b_2) \tag{11}$$

where W_8,W_9 are earnable parameter matrices, while b_1,b_2 are bias vectors, σ_4 is the Tanh activation function. The instance-level contrastive loss is then defined as:

$$\mathcal{L}_{\mathrm{INS}} = -\frac{1}{|B|} \sum_{i \in B} \log \frac{\exp\left(\mathbf{z}_i^{\mathrm{S}} \cdot \mathbf{z}_i^{\mathrm{L}} / \tau\right)}{\sum_{j \in B} \mathcal{V} \cdot \exp\left(\mathbf{z}_i^{\mathrm{S}} \cdot \mathbf{z}_j^{\mathrm{L}} / \tau\right)} \tag{12}$$

Where τ is a temperature parameter, B denotes batch size.

Category-level Contrastive Learning. To capture hierarchical relationships and semantic context, we propose a Category-level contrastive learning task that models the relationships between fine-grained events and their broader categories. Specifically, We employ ontology-guided K-means clustering to group event based on their event types. The centroid C_c for each event type c is computed as:

$$\mathbf{C}_c = \frac{1}{|S_c|} \sum_{i \in S_c} \mathbf{z}_i, \tag{13}$$

where S_c denotes the set of events associated with type c, and $\mathbf{z}_i$ represents the embedding of event i. Each event embedding z_i treats its category centroid $\mathbf{C}_{c_i}$ as the positive sample, and the centroids of other categories $\mathbf{C}_{c(c \neq c_i)}$ as negative samples. The cosine similarity between event embedding and category centroid is computed as:

$$sim\left(\mathbf{z}_i, \mathbf{C}_c\right) = \frac{\mathbf{z}_i \cdot \mathbf{C}_c}{\|\mathbf{z}_i\| \, \|\mathbf{C}_c\|} \tag{14}$$

we dynamically select the hardest negative sample as the centroid with the highest similarity among negative categories. The category-level contrastive loss is then defined as:

$$L_{\mathrm{CAT}} = \frac{1}{N} \sum_{i-1}^{N} \max\left(0, \Delta_i\right) \Delta_i = \gamma - sim\left(\mathbf{z}_i, \mathbf{C}_{c_i}\right) + \max_{c \neq c_i} sim\left(\mathbf{z}_i, \mathbf{C}_c\right) \tag{15}$$

where γ is a hyper-parameter to control the hardness of probability assignment. This loss ensures that embeddings of events within the same type are closer together, while those of different categories are further apart, capturing hierarchical semantic structures.

3.3 Sequential Representation Fusion Unit

To capture latent signals from sequential representations, we introduce a weight-shared attention layer to optimally fuse diverse embeddings. For a clear explanation we denote the linearly-transformed long-term and short-term embeddings from e_q^{LT} and e_{q,t_q}^{ST} as E^L and E^S, respectively. We then evaluate the relevance

between each event embedding in one set and the other using a cross-attention mechanism:

$$\mathrm{E}_c^{\mathrm{SL}} = MHA(\underbrace{\mathrm{E}^{\mathrm{S}}}_{Q}, \underbrace{\mathrm{E}^{\mathrm{L}}}_{K}, \underbrace{\mathrm{E}^{\mathrm{L}}}_{V}); \mathrm{E}_{\mathrm{c}}^{\mathrm{LS}} = MHA(\underbrace{\mathrm{E}^{\mathrm{L}}}_{Q}, \underbrace{\mathrm{E}^{\mathrm{S}}}_{K}, \underbrace{\mathrm{E}^{\mathrm{S}}}_{V}) \tag{16}$$

where $\mathrm{MHA}(Q, K, V)$ denotes multi-head attention operator, Q, K and V are derived from the respective embedding sets.

These two features $\mathrm{E}_{\mathrm{c}}^{\mathrm{SL}}$ and $\mathrm{E}_{\mathrm{c}}^{\mathrm{LS}}$ are subsequently combined using a fusion layer to obtain the final representation:

$$\mathrm{E}_f = \mathrm{W}_9 \odot \mathrm{E}_{\mathrm{c}}^{\mathrm{LS}} + \mathrm{W}_{10} \odot \mathrm{E}_{\mathrm{c}}^{\mathrm{SL}} \tag{17}$$

W_9,W_{10} are learnable weight matrices, E_f deliver a holistic representation for reasoning.

3.4 Inference and Training

We utilize ConvTransE [10] to derive the representation of the query, we calculate the prediction scores for all candidate entities given the query at time t_q as follows:

$$\phi\left(e_q, r_q, e, q\right) = \sigma_3\left(\mathrm{E}_{t_q}^{e_q} ConvTransE\left(\hat{\mathrm{E}}_{t_q}^{e_q}, \mathrm{r}_{t_q}\right)\right) \tag{18}$$

$$\hat{\mathrm{E}}_{t_q}^{e_q} = \lambda \hat{e}_q^{LT} + (1-\lambda)\hat{e}_{q,t_q}^{ST} + \hat{\mathrm{E}}_f \tag{19}$$

where λ is a variable factor that is set at [0,1].

TKG reasoning task can be regarded as a multi-classification task. Following the past works [32], we employs Cross-Entropy as the loss function:

$$\mathcal{L}_{tkg} = \sum_{t=0}^{\mathcal{T}} \sum_{(c_s, r, c, t) \in \mathcal{F}_t} \sum_{e \in \mathcal{E}} y_t^e \log \phi\left(e_s, r, e, t\right) \tag{20}$$

where $\phi\left(e_s, r, e, t\right)$ is the entity prediction probabilistic scores $\mathbf{y}_t^e \in R^{|\mathcal{E}|}$ is the label $[0, 1]$. The final loss is then computed as:

$$\mathcal{L} = \mathcal{L}_{tkg} + \lambda_1 \mathcal{L}_{CAT} + \lambda_2 \mathcal{L}_{INS} \tag{21}$$

where λ_1 and λ_2 are the parameters that control the contrastive loss terms.

4 Experiments

4.1 Experimental Settings

Datasets We leverage four real-world event-based benchmark datasets to evaluate our proposed methods. The datasets include four from the Integrated Crisis Early Warning System (ICEWS) [3]: ICEWS14, ICEWS18, ICEWS05-15 and

Table 1. Dataset Statistics, with "Unseen Events" showing the proportion of test queries involving unseen events.

Dataset	ICEWS14	ICEWS18	ICEWS05-15	GDELT
Training Facts	74,845	373,018	368,868	1,734,399
Validation Facts	8,514	45,995	46,302	238,765
Test Facts	7,371	49,545	46,159	305,241
Entities	6,869	10,094	23,033	7,691
Relations	230	256	251	240
Time Interval	1 day	1 day	1 day	15 mins
Timestamps	365	365	4017	2975
Unseen Events	58.43%	55.69%	39.82%	43.72%

one from the Global Database of Events, Language, and Tone (GDELT) [27]. Following previous works [35, 37], all datasets are split into training, validation, and test sets in proportions of 80%, 10%, 10%. Detailed statistics for all datasets are presented in Table 1.

Evaluation Metrics We evaluate our approach on the task of entity prediction, where the objective is to predict the missing object entity given an entity-relation pair. The evaluation metrics include Mean Reciprocal Ranking (MRR), and Hits@1/3/10. Following standard practices in recent works [34, 38, 7], we report results under the time-aware filtered setting, which excludes quadruples occurring at the query time.

Baselines To provide a comprehensive comparison and intuitive proof, we compare it against with interpolation methods including TTransE [2], TA-DisMult [3] and DE-SimIE [11], as well as extrapolation methods RE-NET [28], CyGNet [29], TITer [31], RE-GCN [32], CEN [33], TiRGN [35], HisMatch [34], RETIA [37], CENET [36], HGLS [38], L^{2}TKG [40], CRAFT [41], THCN[6], LogCL [39]. We provide detailed model descriptions in Appendix A.

Implementation Details All experiments are conducted on NVIDIA A100 GPUs. We utilize Adam as the optimizer with a learning rate of 0.001. The embedding size d is set to 200. R-GCN layers on encoders is set to 2 and the dropout rate for each layer is set to 0.2. The optimal local historical KG snapshots sequences lengths of all datasets are set to 7. The number of heads of the multi-attention mechanism is set to 8, dropout is set to 0.3. The optimal temperature coefficient of all datasets are set to 0.05, 0.05, 0.03, 0.05.

4.2 Main Results

The experimental results for TKG reasoning are presented in Table 2, evaluated using time-aware filtered MRR and Hit@{1/3/10}. These results highlight the effectiveness of our proposed model, demonstrating robust performance and confirming the effectiveness of HSCL in addressing TKG reasoning tasks. Our model consistently outperforms all baseline methods, achieving state-of-the-art (SOTA) performance across the four TKG benchmark datasets. Notably, extrapolation models outperform interpolation models on all datasets. This is attributed to HSCL's ability to incorporate the temporal features of events, enabling it to predict future missing facts. When compared to models under the extrapolation

Table 2. The prediction performance of MRR and Hit@{1/3/10} are on all datasets with time-aware metrics. The best results are highlighted in **bold**, while the second-best results are underlined. All results are retrieved from [7, 39]. To highlight the effectiveness of our proposed HSCL method, we report its relative performance improvement in the final row of the table.

Models	ICEWS14				ICEWS18				ICEWS05-15				GDELT			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hits@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
TTransE	13.72	2.98	17.70	35.74	8.31	1.92	8.56	21.89	15.57	4.80	19.24	38.29	5.50	0.48	4.94	15.25
TA-DisMult	25.80	16.94	29.74	42.99	16.75	8.61	18.41	33.59	24.31	14.58	27.92	44.21	12.00	5.76	12.94	23.54
DE-SimIE	33.36	24.85	37.15	49.82	19.30	11.53	21.86	34.80	35.02	25.91	38.99	52.75	19.70	12.22	21.39	33.70
RE-NET	36.93	26.83	39.51	54.78	28.81	19.05	32.44	47.51	43.32	33.43	47.77	63.06	19.62	12.42	21.00	34.01
CyGNet	35.05	25.73	39.01	53.55	24.93	15.90	28.82	42.61	36.81	26.61	41.63	56.22	18.48	11.52	19.57	31.98
RE-GCN	40.39	30.66	44.96	59.21	30.58	21.01	34.34	48.75	48.03	37.33	53.85	68.27	19.64	12.42	20.99	34.81
CEN	42.20	32.08	46.07	61.31	31.50	21.70	35.44	50.59	46.84	36.38	52.45	67.40	20.39	12.69	21.77	37.67
TiRGN	44.04	33.48	48.95	63.84	33.19	23.91	37.90	54.22	50.71	41.62	56.10	70.71	21.67	13.53	23.37	37.60
HisMatch	46.42	35.91	51.63	66.84	33.99	23.91	37.90	53.94	52.85	42.01	59.05	73.28	22.01	14.45	23.80	36.61
CENET	39.02	29.62	43.23	57.49	27.85	18.15	31.63	46.93	47.13	37.25	47.13	67.61	20.23	12.69	21.70	34.92
L^2TKG	45.89	34.63	-	68.47	31.63	21.17	-	53.01	52.42	40.09	-	75.86	20.16	12.49	-	35.83
HGLS	47.00	35.06	-	70.41	29.32	19.21	-	49.83	46.21	35.32	-	67.12	19.04	11.79	-	33.23
RETIA	42.76	33.28	47.77	62.75	34.23	22.83	36.48	52.94	47.26	36.64	52.90	67.76	21.16	14.25	21.70	34.94
CRAFT	45.71	35.05	51.83	65.21	34.21	23.96	38.53	54.11	50.14	39.56	56.18	70.09	23.78	15.38	26.23	40.15
THCN	45.39	36.58	50.84	66.07	35.63	24.90	39.26	56.76	51.94	40.32	57.79	72.18	23.46	15.18	25.21	39.03
LogCL	48.87	37.76	54.71	70.26	35.67	24.53	40.32	57.74	57.04	46.07	63.72	77.87	23.75	14.64	25.60	42.33
HSCL	**50.03**	**39.24**	**55.65**	**71.29**	**36.52**	**25.37**	**41.61**	**58.22**	**58.98**	**48.33**	**65.93**	**79.62**	**25.08**	**15.51**	**27.31**	**44.40**
Improv.	2.37%	3.92%	1.72%	1.47%	2.46%	1.89%	3.20%	0.82%	3.40%	4.91%	3.47%	2.25%	5.47%	0.84%	4.12%	4.89%

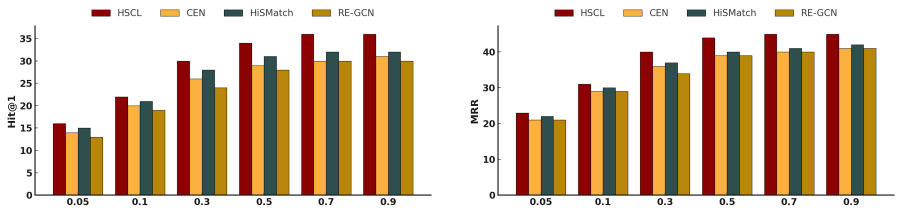
setting, our method consistently achieves superior results in both MRR and Hit scores, with improvements ranging from 0.82% to 5.47%. These results confirm that our hierarchical contrastive learning strategy effectively leverages both temporal and structural information, enabling the model to learn customized, query-specific representations that enhance overall reasoning accuracy. Furthermore, it is noteworthy that HSCL exhibits relatively better performance on ICEWS05-15 and GDELT compared to ICEWS14 and ICEWS18. This phenomenon may be due to more complex dynamic interactions among entities and relations in the ICEWS18 and GDELT datasets, underscoring HSCL's capability to model intricate temporal fact interactions.

Table 3. Ablation study of HSCL on ICEWS14 and ICEWS18.

Models	ICEWS14		ICEWS18	
	MRR	**Hit@1**	**MRR**	**Hit@1**
HSCL-w/o-CL	46.58	34.28	35.20	24.06
HSCL-w/o-L_{INS}	47.08	35.66	35.25	24.20
HSCL-w/o-L_{CAT}	47.42	35.96	35.38	24.36
HSCL-w/o-fusion	48.61	37.41	35.97	25.01
HSCL	**50.03**	**39.24**	**36.52**	**25.37**

4.3 Ablation Studies

To verify the effectiveness of each module in HSCL, ablation studies are carried out with ICEWS14 and ICEWS18, as shown in Table 3. Compared with HSCL-w/o-CL, which achieved an MRR of 46.58 and Hit@1 of 34.28 on ICEWS14,



(a) Hit@1 across training data scales. **(b)** MRR across training data scales.

Fig. 3. Performance of HSCL in terms of MRR and Hit@1 on ICEWS14 under different training data scale settings.

the full HSCL model improved these metrics to 50.03 and 36.52, respectively, demonstrating the positive impact of contrastive learning. Specifically, removing either the instance-level (HSCL-w/o-$\mathcal{L}_{\text{INS}}$) or category-level (HSCL-w/o-$\mathcal{L}_{\text{CAT}}$) loss led to decreases in performance of MRR dropped to 47.08 and 47.42 on ICEWS14, respectively, illustrating the complementary roles of both contrastive components, which certificates that incorporating hierarchical semantics into contrastive learning objective would enable capture of latent semantic, resulting in more robust and generalized representations. Furthermore, HSCL-w/o-fusion, which recorded 48.61 on MRR and 37.41 on Hit@1 on ICEWS14, was outperformed by the full model. The fusion unit, integrating long and short term sequence representations via cross-attention, contributed an absolute improvement of over 2.4% in MRR and nearly 3% in Hit@1 on ICEWS14. These improvements underscore how contrastive learning enhances feature differentiation and sample separation at dual semantic levels while the fusion unit facilitates positive knowledge interactions, ultimately leading to more robust reasoning.

4.4 Performance under Low-resource Setting

In this section, we analyze the impact of different scales of training data setting. We randomly generate different proportions $r\%$ $(5\%, 10\%, 20\%, 30\%, 50\%, 90\%)$ of training data to study the influence of the size of the training set and use the original vaild and test set for evaluation. As shown in Figure 3, we compare the performance of HisMatch, CEN and RE-GCN in various size of training set. Generally, predictions under low-resource conditions yield poorer results due to the greater difficulty in optimizing representations from limited data. However, HSCL consistently outperforms all baselines across various $r\%$ settings in terms of MRR and Hit@1. This performance can be attributed to the hierarchical contrastive learning module, which is particularly effective in low-resource scenarios. The hierarchical contrastive learning module operates by leveraging both instance-level and category-level distinctions to refine representations. Under data scarcity, the instance-level contrastive objective encourages the model to discern subtle differences between facts using dynamic adaptive weighting, which magnifies important relational and temporal cues even when direct exam-

Table 4. Performance on predicting unseen events in terms of MRR and Hit@1 on both ICEWS14 and ICEWS18 datasets.

Models	ICEWS14		ICEWS18	
	MRR	Hit@1	MRR	Hit@1
RE-GCN	23.26	13.91	15.08	7.09
CEN	22.06	13.28	15.41	8.20
RETIA	24.17	14.67	16.62	9.08
HisMatch	27.49	19.04	17.51	11.13
LogCL	29.19	18.72	18.40	11.74
HSCL	**30.73**	**20.00**	**19.43**	**12.17**
Improv.	**5.28%**	**5.04%**	**5.60%**	**3.66%**

ples are few. Simultaneously, the category-level contrastive component groups events into broader semantic categories, allowing the model to draw on shared structural and temporal patterns across similar events.

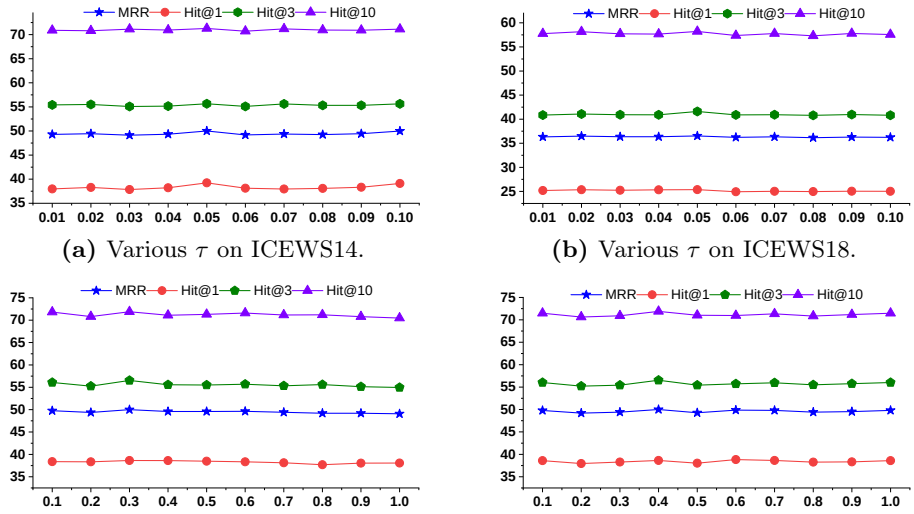
4.5 Performance under Inductive Setting

To further validate HSCL's capacity for learning discriminative and generalized query-specific representations, we evaluate its performance on ICEWS14 and ICEWS18 datasets featuring unseen events absent from historical TKGs. Chosen for their higher rate of unseen events, as shown in Table 1, these datasets pose a rigorous test for generalization. Experimental results in Table 4 demonstrate that HSCL outperforms all SOTA baselines, with relative improvements of 5.28% and 5.60% on MRR and 5.04% and 3.66% on Hit@1 for ICEWS14 and ICEWS18, respectively. This underscores HSCL's ability to integrate both structural dependencies and temporal dynamics of entities and relations. Its superior performance on unseen events suggests that HSCL's hierarchical contrastive learning effectively captures underlying patterns, enabling robust reasoning even when encountering unfamiliar events. By synthesizing structural information with evolving temporal patterns, HSCL constructs nuanced, query-specific representations that differentiate subtle semantic variations and relational changes critical for reasoning in unseen events.

4.6 Sensitivity Analysis

We run our model with two sets of different important hyper-parameters to explore weight impacts. The temperature coefficient τ is crucial in contrastive learning, affecting how aggressively the model differentiates between positive and negative pairs. A smaller τ encourages more assertive separation of negative pairs, which can accelerate learning but may also lead to instability or overfitting. To assess the training stability of HSCL, we conducted a sensitivity analysis of τ across a range from 0.01 and 0.09 on both ICEWS14 and ICEWS18 datasets. As shown in Figure 4a and Figure 4b, our findings reveal that the model's effectiveness remains consistent across various settings, without any significant decline in performance metrics in terms of MRR and Hit@$\{1/3/10\}$. Furthermore, we analyze the impact of the loss balancing terms λ_1 and λ_2 within our multi-task

learning paradigm. To systematically investigate these terms, we conducted experiments varying λ_1 and λ_2 from 0.1 to 0.9 in increments of 0.1 as depicted in Figure 4c and Figure 4d. Our findings confirm that appropriate balancing of task contributions is crucial for maximizing model effectiveness, with optimal performance observed when λ_1 and λ_2 are set to 0.3 and 0.4, respectively.



(a) Various τ on ICEWS14. **(b)** Various τ on ICEWS18.

(c) Various λ_1 on ICEWS14. **(d)** Various λ_2 on ICEWS14.

Fig. 4. Performance of HSCL under different hyper-parameters

5 Conclusion

In this study, we propose a novel Hierarchical Semantic-aware Contrastive Learning (HSCL) framework for TKG reasoning. HSCL leverages two complementary contrastive objectives that operate at different semantic levels to effectively address the challenges of fine-grained differentiation and broader hierarchical structuring. Specifically, the instance-level contrastive objective refines representations by distinguishing subtle differences between individual events, ensuring that semantically similar events remain adequately differentiated. Meanwhile, the category-level contrastive objective clusters events into broader semantic groups, capturing hierarchical relationships among events and improving generalization. Our empirical evaluations across four benchmarks confirm HSCL's superior representation capability and inference accuracy under various experimental setups.

Acknowledgments. We would like to thank the reviewers for their constructive comments. This project is supported by the National Natural Science Foundation of China (U2336204, U22B2061), Key Development Projects of the Sichuan Provincial Science and Technology Plan (2024YFG0005), Natural Science Foundation of Sichuan, China (2024NSFSC0496).

References

1. Chen, T., Kornblith, S., Norouzi, M., Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PmLR.
2. Leblay, J., Chekol, M. W. (2018, April). Deriving validity time in knowledge graph. In Companion proceedings of the the web conference 2018 (pp. 1771-1776).
3. García-Durán, A., Dumančić, S., Niepert, M. (2018). Learning sequence encoders for temporal knowledge graph completion. arXiv preprint arXiv:1809.03202.
4. Chung, J., Gulcehre, C., Cho, K., Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.
5. Van der Maaten, L., Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11), 2579-2605.
6. Chen, T., Long, J., Wang, Z., Luo, S., Huang, J., Yang, L. (2024). THCN: A Hawkes Process Based Temporal Causal Convolutional Network for Extrapolation Reasoning in Temporal Knowledge Graphs. IEEE Transactions on Knowledge and Data Engineering.
7. Cai, Y., Liu, Q., Gan, Y., Li, C., Liu, X., Lin, R., Luo, D., Yang, J. (2024). Predicting the Unpredictable: Uncertainty-Aware Reasoning over Temporal Knowledge Graphs via Diffusion Process. Findings of the Association for Computational Linguistics ACL 2024, 5766-5778.
8. Vashishth, S., Sanyal, S., Nitin, V., Talukdar, P. (2019). Composition-based multi-relational graph convolutional networks. arXiv preprint arXiv:1911.03082.
9. Schlichtkrull, M., Kipf, T. N., Bloem, P., Van Den Berg, R., Titov, I., Welling, M. (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 15 (pp. 593-607). Springer.
10. Shang, C., Tang, Y., Huang, J., Bi, J., He, X., Zhou, B. (2019). End-to-end structure-aware convolutional networks for knowledge base completion. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, pp. 3060-3067).
11. Goel, R., Kazemi, S. M., Brubaker, M., Poupart, P. (2020). Diachronic embedding for temporal knowledge graph completion. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 04, pp. 3988-3995).
12. Ren, X., Wu, Z., He, W., Qu, M., Voss, C. R., Ji, H., Abdelzaher, T. F., Han, J. (2017). Cotype: Joint extraction of typed entities and relations with knowledge bases. In Proceedings of the 26th international conference on world wide web (pp. 1015-1024).
13. Ma, Y., Ye, C., Wu, Z., Wang, X., Cao, Y., Chua, T. S. (2023). Context-aware event forecasting via graph disentanglement. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 1643-1652).
14. Zheng, S., Chen, W., Zhao, P., Liu, A., Fang, J., Zhao, L. (2021). When hardness makes a difference: Multi-hop knowledge graph reasoning over few-shot relations. In Proceedings of the 30th ACM International Conference on Information Knowledge Management (pp. 2688-2697).
15. Wang, R., Zhang, Y., Li, J., Liu, S., Sun, D., Wang, T., Wang, T., Chen, Y., Kara, D., Abdelzaher, T. (2024). Metahkg: Meta hyperbolic learning for few-shot temporal reasoning. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 59-69).
16. Sun, H., Geng, S., Zhong, J., Hu, H., He, K. (2022). Graph hawkes transformer for extrapolated reasoning on temporal knowledge graphs. In Proceedings of the

2022 Conference on Empirical Methods in Natural Language Processing (pp. 7481-7493).
17. Ding, Z., Wu, J., He, B., Ma, Y., Han, Z., Tresp, V. (2022). Few-shot inductive learning on temporal knowledge graphs using concept-aware information. arXiv preprint arXiv:2211.08169. year=2022
18. Liu, K., Zhao, F., Chen, H., Li, Y., Xu, G., Jin, H. (2022). Da-net: Distributed attention network for temporal knowledge graph reasoning. In Proceedings of the 31st ACM International Conference on Information Knowledge Management (pp. 1289-1298).
19. Gao, Y., Feng, L., Kan, Z., Han, Y., Qiao, L., Li, D. (2022). Modeling precursors for temporal knowledge graph reasoning via auto-encoder structure. In Proceedings of the International Joint Conference on Artificial Intelligence (pp. 2044-2051).
20. Wang, R., Li, Z., Sun, D., Liu, S., Yin, B., Abdelzaher, T. (2022). Learning to sample and aggregate: Few-shot reasoning over temporal knowledge graphs. In Advances in Neural Information Processing Systems (Vol. 35, pp. 16863-16876).
21. Peng, M., Liu, B., Xu, W., Jiang, Z., Zhu, J., Peng, M. (2024). Deja vu: Contrastive Historical Modeling with Prefix-tuning for Temporal Knowledge Graph Reasoning. In Findings of the Association for Computational Linguistics: NAACL 2024 (pp. 1178-1191).
22. Feng, X., Liu, X., Yang, Y., Wang, W., Wang, J. (2024). Learning Rules in Knowledge Graphs via Contrastive Learning. In International Conference on Database Systems for Advanced Applications (pp. 408-424). Springer.
23. Li, J., Selvaraju, R. R., Gotmare, A. S., Joty, S., Xiong, C., Hoi, S. C. H. (2021). Align before fuse: Vision and language representation learning with momentum distillation. Advances in Neural Information Processing Systems, 34, 9694-9705.
24. Gunel, B., Du, J., Conneau, A., Stoyanov, V. (2020). Supervised contrastive learning for pre-trained language model fine-tuning. arXiv preprint arXiv:2011.01403.
25. Wu, J., Wang, X., Feng, F., He, X., Chen, L., Lian, J., Xie, X. (2021). Self-supervised graph learning for recommendation. In Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval (pp. 726-735).
26. Han, Z., Ding, Z., Ma, Y., Gu, Y., Tresp, V. (2021). Learning neural ordinary equations for forecasting future links on temporal knowledge graphs. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (pp. 8352-8364).
27. Trivedi, R., Dai, H., Wang, Y., Song, L. (2017). Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In International Conference on Machine Learning (pp. 3462-3471). PMLR.
28. Jin, W., Qu, M., Jin, X., Ren, X. (2019). Recurrent event network: Autoregressive structure inference over temporal knowledge graphs. arXiv preprint arXiv:1904.05530.
29. Zhu, C., Chen, M., Fan, C., Cheng, G., Zhang, Y. (2020). Learning from history: Modeling temporal knowledge graphs with sequential copy-generation networks. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 04, pp. 3988-3995).
30. Han, Z., Chen, P., Ma, Y., Tresp, V. (2021). Explainable subgraph reasoning for forecasting on temporal knowledge graphs. In International Conference on Learning Representations.
31. Sun, H., Zhong, J., Ma, Y., Han, Z., He, K. (2021). Timetraveler: Reinforcement learning for temporal knowledge graph forecasting. arXiv preprint arXiv:2109.04101.

32. Li, Z., Jin, X., Li, W., Guan, S., Guo, J., Shen, H., Wang, Y., Cheng, X. (2021). Temporal knowledge graph reasoning based on evolutional representation learning. In Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval (pp. 408-417).
33. Li, Z., Guan, S., Jin, X., Peng, W., Lyu, Y., Zhu, Y., Bai, L., Li, W., Guo, J., Cheng, X. (2022). Complex evolutional pattern learning for temporal knowledge graph reasoning. arXiv preprint arXiv:2203.07782.
34. Li, Z., Hou, Z., Guan, S., Jin, X., Peng, W., Bai, L., Lyu, Y., Li, W., Guo, J., Cheng, X. (2022). Hismatch: Historical structure matching based temporal knowledge graph reasoning. arXiv preprint arXiv:2210.09708.
35. Li, Y., Sun, S., Zhao, J. (2022). TiRGN: Time-guided recurrent graph network with local-global historical patterns for temporal knowledge graph reasoning. In Proceedings of the International Joint Conference on Artificial Intelligence (pp. 2152-2158).
36. Xu, Y., Ou, J., Xu, H., Fu, L. (2023). Temporal knowledge graph reasoning with historical contrastive learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 4, pp. 4765-4773).
37. Liu, K., Zhao, F., Xu, G., Wang, X., Jin, H. (2023). RETIA: Relation-Entity Twin-Interact Aggregation for Temporal Knowledge Graph Extrapolation. In 2023 IEEE 39th International Conference on Data Engineering (ICDE) (pp. 1761-1774). IEEE.
38. Zhang, M., Xia, Y., Liu, Q., Wu, S., Wang, L. (2023). Learning long-and short-term representations for temporal knowledge graph reasoning. In Proceedings of the ACM Web Conference 2023 (pp. 2412-2422).
39. Chen, W., Wan, H., Wu, Y., Zhao, S., Cheng, J., Li, Y., Lin, Y. (2023). Local-global history-aware contrastive learning for temporal knowledge graph reasoning. In 2024 IEEE 40th International Conference on Data Engineering (ICDE) (pp. 733-746). IEEE.
40. Zhang, M., Xia, Y., Liu, Q., Wu, S., Wang, L. (2024). Learning latent relations for temporal knowledge graph reasoning. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 12617-12631).
41. Zhang, S., Wei, W., Huang, R., Xie, W., Chen, D. (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 (pp. 7675-7686).
42. Wang, T., Isola, P. (2020). Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In International conference on machine learning (pp. 9929-9939). PMLR.
43. Luo, D., Gan, Y., Hou, R., Lin, R., Liu, Q., Cai, Y., Gao, W. (2024). Synergistic Anchored Contrastive Pre-training for Few-Shot Relation Extraction. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 17, pp. 18742-18750).
44. Xu, H., Zhang, X., Li, H., Xie, L., Dai, W., Xiong, H., Tian, Q. (2022). Seed the views: Hierarchical semantic alignment for contrastive representation learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(3), 3753-3767.
45. Yang, C., Wu, Q., Jin, J., Gao, X., Pan, J., Chen, G. (2022). Trading hard negatives and true negatives: A debiased contrastive collaborative filtering approach. arXiv preprint arXiv:2204.11752.

46. Kalantidis, Y., Sariyildiz, M. B., Pion, N., Weinzaepfel, P., Larlus, D. (2020). Hard negative mixing for contrastive learning. Advances in Neural Information Processing Systems, 33, 21798-21809.
47. Zhang, H., Zhang, J., Molybog, I. (2024). HaSa: Hardness and Structure-Aware Contrastive Knowledge Graph Embedding. In Proceedings of the ACM Web Conference 2024 (pp. 2116-2127).
48. Liu, Y., Shu, L., Chen, C., Zheng, Z. (2024). Fine-Grained Semantics Enhanced Contrastive Learning for Graphs. IEEE Transactions on Knowledge and Data Engineering.
49. Guo, Q., Guo, Y., Zhao, J. (2024). HRCL: Hierarchical Relation Contrastive Learning for Low-Resource Relation Extraction. IEEE Transactions on Neural Networks and Learning Systems, 35(1), 1-12. doi:10.1109/TNNLS.2024.3281234.
50. Wang, J., Li, W., Hou, C., Tang, X., Qiao, Y., Fang, R., Li, P., Gao, P., Xie, G. (2022). HCL: Improving graph representation with hierarchical contrastive learning. In International Semantic Web Conference (pp. 108-124). Springer.
51. Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., Starz, J., Ward, M. (2015). CAMEO.CDB.09b5.pdf. In ICEWS Coded Event Data (Version V37). Harvard Dataverse. doi:10.7910/DVN/28075/SCJPXX. Retrieved from https://doi.org/10.7910/DVN/28075/SCJPXX.
52. Dettmers, T., Minervini, P., Stenetorp, P., Riedel, S. (2018). Convolutional 2D knowledge graph embeddings. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 32, No. 1).
53. Kalantidis, Y., Sariyildiz, M. B., Pion, N., Weinzaepfel, P., Larlus, D. (2020). Hard negative mixing for contrastive learning. Advances in Neural Information Processing Systems, 33, 21798-21809.
54. Kazemi, S. M., Poole, D. (2018). Simple embedding for link prediction in knowledge graphs. In Advances in neural information processing systems (pp. 4284-4295).
55. Jin, W., Qu, M., Jin, X., Ren, X. (2019). Recurrent event network: Autoregressive structure inference over temporal knowledge graphs. arXiv preprint arXiv:1904.05530.
56. Dong, H., Wang, P., Xiao, M., Ning, Z., Wang, P., Zhou, Y. (2024). Temporal inductive path neural network for temporal knowledge graph reasoning. Artificial Intelligence, 329, 104085. doi:10.1016/j.artint.2024.104085.
57. Dong, H., Ning, Z., Wang, P., Qiao, Z., Wang, P., Zhou, Y., Fu, Y. (2023). Adaptive path-memory network for temporal knowledge graph reasoning. In Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (pp. 2086-2094).
58. Liu, Y., Ma, Y., Hildebrandt, M., Joblin, M., Tresp, V. (2022). Tlogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 4, pp. 4120-4127).
59. He, Y., Zhang, P., Liu, L., Liang, Q., Zhang, W., Zhang, C. (2024). Hip network: Historical information passing network for extrapolation reasoning on temporal knowledge graph. arXiv preprint arXiv:2402.12074.
60. Mingcong, S., Zhu, C., Zhang, D., Wen, S., Qing, L. (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 (pp. 5232-5243).
61. Chen, B., Xiao, C., Zhou, F. (2024). Natural Evolution-based Dual-Level Aggregation for Temporal Knowledge Graph Reasoning. In Findings of the Association for Computational Linguistics: EMNLP 2024 (pp. 9274-9284).