Modeling Historical Relevant and Local Frequency Context for Representation-Based Temporal Knowledge Graph Forecasting

Shengzhe Zhang[1], Wei Wei*[1], Rikui Huang[1], Wenfeng Xie[2], , Dangyang Chen[2]
[1]School of Computer Science & Technology, Huazhong University of Science and Technology
[2]Ping An Property & Casualty Insurance company of China
{zshengz, weiw, huangrk}@hust.edu.cn
julian_wind@163.com, chendangyang273@pingan.com.cn

Abstract

Temporal Knowledge Graph Forecasting (TKGF) aims to harness historical KG sequences to predict future facts. However, existing representation-based approaches indiscriminately incorporate all relevant historical facts as the temporal context of candidate entities for prediction, which may lead to serious information loss or homogeneous prediction, owing to neglecting candidate-specific temporal context. That is, each candidate entity is related to a unique temporal context, rather than an identical one. To address this issue, we propose a novel representation learning model for TKGF, namely **CRAFT**, which models the candidate-specific temporal context from two aspects: histori**C**al **R**elevant context and loc**A**l **F**requency contex**T**. Specifically, we design a Historical Relevant Context Encoder with the dual-attention mechanism to encode historical relevant paths of candidates and employ a Local Frequency Context Encoder to capture local frequency features in repetitive facts. Then, we apply a Context-Enhanced Decoder to facilitate more accurate predictions via the derived information. Experiments on six benchmark datasets validate the effectiveness of our model and further analysis proves that CRAFT can leverage temporal contextual information to achieve differential predictions.

1 Introduction

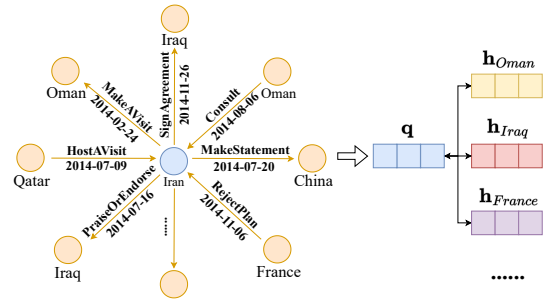
Temporal Knowledge Graphs (TKGs) (Boschee et al., 2015; Gottschalk and Demidova, 2018; Zhao, 2020) are constituted by a sequence of knowledge graphs partitioned according to timestamps and contain a vast amount of entity and relation information (Ding et al., 2024; Fan et al., 2024), garnering widespread attention in recent years. In these graphs, facts can be represented in the form of quadruples *(subject, relation, object, timestamp)*. However, real-world TKGs often contain substantial unexplored information. Consequently, Temporal Knowledge Graph Reasoning (TKGR) (Jiang et al., 2016; Leblay and Chekol, 2018; Jin et al., 2019; Li et al., 2021b; Huang et al., 2024a) was proposed, which employs link prediction to infer hidden facts within the graphs. Based on the temporal scope of available historical facts, TKGR is categorized into two settings: interpolation and extrapolation. The latter, also known as Temporal Knowledge Graph Forecasting (TKGF) (Li et al., 2021b, 2022a; Liu et al., 2022; Li et al., 2023; Huang et al., 2024b), aims to predict facts at future timestamps and plays a significant role in various intelligent applications, such as disaster relief (Signorini et al., 2011), financial analysis (Bollen et al., 2011), and even in foreseeing the future trends of artificial intelligence (Krenn et al., 2023).

TKGF is achieved by answering temporal queries about future timestamps. Aligning with human cognition, people need to rely on specific historical facts to answer temporal queries accurately, that is, for different responses, they depend on the particular temporal context. Therefore, it is crucial to model the candidate-specific temporal context for the TKGF task. Existing representation-based methods learn semantically rich embeddings and aggregate historical information into the single query representation to accomplish link prediction. However, they overlook the modeling of candidate-specific temporal context, which can easily result in information loss and homogeneous predictions.

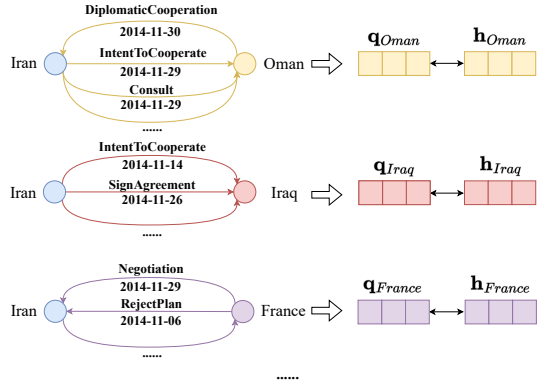
We take the temporal query *(Iran, Express intent to cooperate, ?, 2014-12-01)* for illustration. As depicted in Figure 1a, previous models excessively compress all historically related information into one query representation, inevitably leading to the loss of contextual information. Meanwhile, these approaches employ the same temporal context for predicting different candidate entities, making it difficult to achieve differential predictions, espe-

* Corresponding author.

Findings of the Association for Computational Linguistics: EMNLP 2024, pages 7675–7686



(a) Identical temporal context



(b) Candidate-specific temporal context

Figure 1: Different modeling paradigms of temporal context for query *(Iran, Express intent to cooperate, ?, 2014-12-01)*. $\mathbf{q}$ and $\mathbf{h}$ denote query representation and entity representation, respectively.

cially when the temporal semantics of candidate entities are similar. These factors collectively contribute to imprecise prediction results. We believe that leveraging candidate-specific temporal context rather than all historical facts will provide more convincing and differential reasoning evidence for different candidates, as illustrated in Figure 1b.

To this end, we propose a novel model, namely **CRAFT**, which models the candidate-specific temporal context from two aspects: histori**C**al **R**elevant context and loc**A**l **F**requency contex**T**. Specifically, for the **historical relevant context**, we first utilize a rule-based method to generate historical relevant paths that link the query with the candidates. Then, we design a Historical Relevant Context Encoder (HRCE) to encode the mined paths, thereby enhancing the contextual awareness capability of the model. We employ semantic alignment attention and time awareness attention to further enhance HRCE's ability to model paths in both semantic and temporal aspects. For the **local frequency context**, we employ a Local Frequency Context Encoder to encode local frequency features to capture contextual information in historical repetitive facts. These two modules jointly model the temporal context of queries. Ultimately, we apply a Context-Enhanced Decoder to facilitate more accurate predictions utilizing candidate-specific temporal contextual information.

In general, our work makes the following contributions:

- We proposed a representation learning model for TKGF, which models historical relevant and local frequency context of candidates. To our best of knowledge, this is the first time to consider candidate-specific temporal context in-dpeth in representation-based approaches.
- We design a Historical Relevant Context Encoder and a Local Frequency Context Encoder to capture the contextual information in historical relevant paths and repetitive facts respectively. The encoded features are fed into a Context-Enhanced Decoder to achieve more precise predictions.
- Experiments on six widely-used TKG benchmarks validate the effectiveness of our proposed model. Further experimental analysis demonstrates that our model can efficiently capture the candidate-specific temporal context, thereby enabling differential predictions.

2 Related Work

In this section, we will introduce the related work of TKG Forecasting and how they utilize historical contextual information.

2.1 Representation-based TKG Forecasting

Representation-based approaches exploit historical KG sequences to learn semantically rich representations for TKG elements and integrate such representations with scoring functions to implement link prediction. RE-NET (Jin et al., 2019) and RE-GCN (Li et al., 2021b) employ RNNs and GCNs to capture temporal and structural dependencies in historical KG sequences. TANGO (Han et al., 2021) uses neural ordinary differential equations to model TKG in the continuous time domain. CEN (Li et al., 2022b) designs a length-aware convolutional neural network to handle the length diversity of evolutional patterns. TiRGN (Li et al., 2022a) simultaneously considers the sequential, repetitive, and cyclical modes to fully learn the characteristics of historical facts. HisMatch (Li et al., 2022c) regards query prediction as a matching task and utilizes query-related and candidate-related historical facts to enhance the representations. Nevertheless,

they incorporate excessive information while neglecting the contextual differentiation of specific candidates.

2.2 Path-based TKG Forecasting

Path-based methods generate connection paths that link the query and candidates to accomplish prediction. xERTE (Han et al., 2020) employs a subgraph sampling strategy to generate interpretable reasoning subgraphs. CluSTeR (Li et al., 2021a) and TITer (Sun et al., 2021) both use reinforcement learning to travel interpretable reasoning paths. However, they are all based on expansion methods, thus only considering the context of the query subject entity while neglecting the context of candidates. TECHS (Lin et al., 2023) incorporates structural dependencies, temporal dynamics, and hidden logical rules to learn differentiable logical rules for reasoning. TLogic (Liu et al., 2022) is a purely rule-based model that uses available rules of the query as context for probabilistic reasoning. Although these methods can generate connection paths to construct differential contextual representations for candidates, they all neglect non-connected information in the KG sequence, which may limit the model's reasoning performance (Li et al., 2023). In our work, we employ path-based methods to extract differential contextual information for query candidates and incorporate it into an enhanced representation-based model. This allows us to fully utilize specific contexts to achieve more accurate predictions.

3 Problem Formulation

Definition 1 (TKG & TKGF). A TKG can be modeled as a sequence of timestamped KGs, denoted as $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_t\}$. Each KG can be represented by $\mathcal{G}_t = (\mathcal{E}, \mathcal{R}, \mathcal{F}_t)$, where $\mathcal{E}$ denotes the set of entities, $\mathcal{R}$ represents the set of relations, and $\mathcal{F}_t$ signifies the collection of facts at timestamp t. A fact in $\mathcal{F}_t$ can be expressed as a quadruple $(s, r, o, t) \in \mathcal{F}_t$, and we follow (Li et al., 2021b) to generate a reverse quadruple (o, r^{-1}, s, t) for each quadruple. Given a query $(s, r, ?, t_q)$ at future timestamp t_q, TKGF aims to utilize historical KG sequences $\{\mathcal{G}_i\}_{i=1}^{t_q-1}$ to predict the missing entity.

Definition 2 (Query-specific temporal context). Given a temporal query $(s, r, ?, t_q)$, the relevant historical fact information that can serve as evidence for answering q constitutes the **Query-specific temporal context** of q.

Definition 3 (Candidate-specific temporal context). Given a temporal query $(s, r, ?, t_q)$ and a candidate o, the relevant historical fact information that can support o as the answer to q constitutes the **Candidate-specific temporal context** for o.

4 Methodology

CRAFT utilizes three principal components to model candidate-specific temporal context: Historical Relevant Context Encoder (Section 4.1), Local Frequency Context Encoder (Section 4.2), and Context-Enhanced Decoder (Section 4.3). The overview of CRAFT is illustrated in Figure 2.

4.1 Historical Relevant Context Encoder

We employ the Historical Relevant Context Encoder to capture contextual information in historical relevant paths of candidates. Specifically, we first generate historical relevant paths during the data preprocessing stage. Then, we derive the evolutional representations of TKG elements by capturing historical KGs' structural dependencies and sequential patterns. Subsequently, given a specific query, we utilize the evolutional representations to encode the relation-specific temporal rules obtained from preprocessing. Lastly, we combine the rule encodings of historical relevant paths with the dual-attention mechanism to generate aggregated path representations.

Data Preprocessing. We generate the historical relevant paths of candidates through temporal logical rule mining and application (Liu et al., 2022). The temporal logical rule is defined as follows:

$$(E_1, r_h, E_{l+1}, T_{l+1}) \leftarrow \wedge_{i=1}^{l}(E_i, r_i, E_{i+1}, T_i), \quad with \quad T_1 \leq T_2 \leq \ldots \leq T_l < T_{l+1} \quad (1)$$

where the left side of the arrow represents the rule head, while the right side denotes the rule body. E_i and T_i denote entity and timestamp variables respectively.

For a given query relation r, we first obtain the corresponding rule set TR through rule mining, which will be utilized in temporal rule encoding. Subsequently, for each temporal rule $tr \in TR$, we apply it by grounding the rule body to search for reachable object entities. After applying rules individually, we obtain the set of rule-reachable candidates O, where each candidate $o \in O$ corresponds to a set of applicable rules TR_o, and the groundings of applicable rules constitute historical

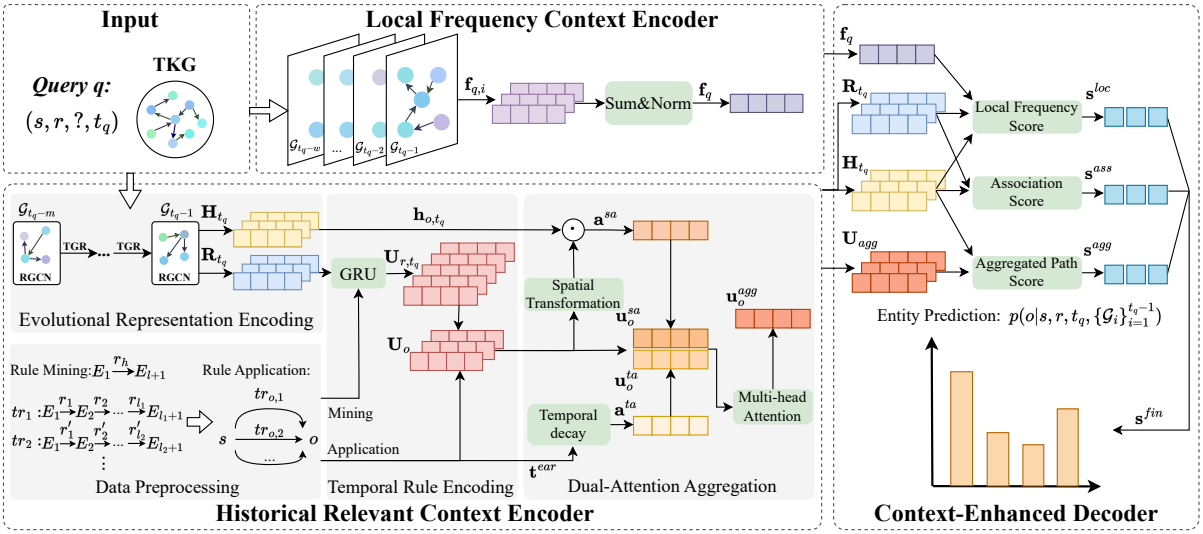


Figure 2: Overview of CRAFT. The Historical Relevant Context Encoder utilizes evolutional representations and historical relevant paths, combined with the dual-attention mechanism to generate aggregated path representations. The Local Frequency Context Encoder encodes the local frequency features. The outputs of both encoders are fed into the Context-Enhanced Decoder to calculate the decoding scores and accomplish entity prediction.

relevant paths of entities. More details about path generation can be found in the Appendix A. Moreover, we calculate a score for each relevant path and candidate pair to facilitate path aggregation.

Evolutional Representation Encoding. For queries at timestamp t_q, We encode the historical KG sequence $\{\mathcal{G}_i\}_{i=t_q-m}^{t_q-1}$ of length m to obtain the evolutional representations. The initial representations for entities and relations are designated as $\mathbf{H} \in \mathbb{R}^{|\mathcal{E}|\times d}$ and $\mathbf{R} \in \mathbb{R}^{|\mathcal{R}|\times d}$, respectively, where d denotes the dimension of the embedding. Following (Li et al., 2022a), we evolve the KG in chronological order, utilizing RGCNs to capture intra-graph structural dependencies, and employing time gate recurrent (TGR) modules to capture inter-graph sequential patterns:

$$\mathbf{R}_t^{GCN} = RGCN_{rel}\left(\mathbf{R}_t, \mathbf{H}_t, G_t\right) \quad (2)$$

$$\mathbf{R}_{t+1} = TGR_{rel}\left(\mathbf{R}_t, \mathbf{R}_t^{GCN}\right) \quad (3)$$

$$\mathbf{H}_t^{GCN} = RGCN_{ent}\left(\mathbf{H}_t, \mathbf{R}_{t+1}, G_t\right) \quad (4)$$

$$\mathbf{H}_{t+1} = TGR_{ent}\left(\mathbf{H}_t, \mathbf{H}_t^{GCN}\right) \quad (5)$$

where timestamp $t \in [t_q - m, t_q - 1]$. $\mathbf{R}_{t_q-m}$ and $\mathbf{H}_{t_q-m}$ are set to $\mathbf{R}$ and $\mathbf{H}$, respectively.

Temporal Rule Encoding. We utilize the derived evolutional representations to encode temporal rules. Given the query $(s, r, ?, t_q)$, we obtain the set of rules TR corresponding to the relation r. For each rule $tr \in TR$, we extract the relation chain $\{r_i\}_{i=1}^{l}$ in the rule body, where l denotes the length of the relation chain. For all temporal rules, we take the evolutional representations of the relations in the embedding matrix $\mathbf{R}_{t_q}$ to obtain the relation chain tensor $\mathbf{C}_{r,t_q} \in \mathbb{R}^{|TR|\times l_{max}\times d}$, where l_{max} represents the maximum length of the relation chains. Chains with insufficient length are padded to facilitate batch processing. Subsequently, we utilize GRU to encode temporal rules and obtain the rule representation matrix $\mathbf{U}_{r,t_q}$:

$$\mathbf{U}_{r,t_q} = GRU\left(\mathbf{C}_{r,t_q}, \mathbf{Z}\right) \quad (6)$$

where $\mathbf{U}_{r,t_q} \in \mathbb{R}^{|TR|\times d}$ and $\mathbf{Z}$ denotes the zero matrix. Notably, we adopt the output of the recurrent unit corresponding to the rule's length as the representation of the rule.

Dual-Attention Aggregation. We employ the dual-attention mechanism to integrate historical relevant paths and generate aggregated path representations. For each rule-reachable candidate $o \in O$ and its applicable rule set TR_o, we rank the corresponding historical relevant paths in descending order by their scores and select the top n for path aggregation. We obtain the path matrix $\mathbf{U}_o \in \mathbb{R}^{n\times d}$ for candidate o from the encoded temporal rules $\mathbf{U}_{r,t_q}$. For each relevant path, we utilize semantic alignment attention and time awareness attention to capture important temporal contextual information.

Semantic alignment attention aims to reinforce contextual features relevant to query candidate entities, and we achieve it by capturing the association between paths and candidate entities. Due to the disparity between path representation and entity representation, inspired by (Lin et al., 2015), we employ a linear layer to transform the path rep-

resentations into a shared space. The semantic alignment attention is computed as follows:

$$\mathbf{a}^{sa} = \text{softmax}((\mathbf{U}_o\mathbf{W_1} + \mathbf{b_1})\,\mathbf{h}_{o,t_q}) \quad (7)$$

where $\mathbf{W_1} \in \mathbb{R}^{d\times d}$ and $\mathbf{b_1} \in \mathbb{R}^d$ are learnable parameters, and $\mathbf{h}_{o,t_q}$ denotes the representation of candidate o in the timestamp t_q. Subsequently, we weighted the path representations to obtain semantic alignment aggregated representation:

$$\mathbf{u}_o^{sa} = \sum_{i=1}^{n} \alpha_i^{sa}\mathbf{u}_i \quad (8)$$

where $\mathbf{u}_i$ denotes the ith path representation.

Contextual information is time-sensitive. Hence, we employ time awareness attention to model the temporal validity of context. For each candidate o, we take the timestamp of the first edge of top n historical relevant paths to obtain the timestamp vector $\mathbf{t}^{ear} \in \mathbb{R}^n$. Intuitively, the time validity of facts will decay over time, and we model this decay using the exponential distribution. The time awareness attention is calculated as follows:

$$\mathbf{a}^{ta} = \text{softmax}(-e^{\lambda(t_q-\mathbf{t}^{ear})}) \quad (9)$$

where $\lambda > 0$ represents the time decay coefficient. Similarly, we weighted the path representations to obtain time awareness aggregated representation:

$$\mathbf{u}_o^{ta} = \sum_{i=1}^{n} \alpha_i^{ta}\mathbf{u}_i \quad (10)$$

After obtaining the two attention representations, we employ multi-head attention (Vaswani et al., 2017) to enable the model to focus on salient contextual features, and generate the final candidate-specific aggregated path representation:

$$\mathbf{u}_o^{agg} = \mathbf{W_2}[\mathbf{u}_o^{sa}; \mathbf{u}_o^{ta}] + \mathbf{b}_2 \quad (11)$$

where $\mathbf{W_2} \in \mathbb{R}^{d\times 2d}$ and $\mathbf{b_2} \in \mathbb{R}^d$ are learnable parameters. [;] denotes the vector concatenation operation.

4.2 Local Frequency Context Encoder

For specific query candidates, their occurrences over history constitute a significant part of their contexts. Therefore, we designed a Local Frequency Context Encoder to capture these features. For the query $(s, r, ?, t_q)$, due to the time validity of facts, we count the frequency of query candidates appearing on the local rather than global historical KG sequence $\{\mathcal{G}_i\}_{i=t_q-w}^{t_q-1}$ with the window size of w, and limit the subject entities and relations of facts to be the same as the query. We obtain the query-specific historical frequency representation $\mathbf{f}_q$ by cumulatively aggregating over timestamps:

$$\mathbf{f}_q = \text{Norm}(\sum_{i=t_q-w}^{t_q-1} \mathbf{f}_{q,i}) \quad (12)$$

where $\mathbf{f}_{q,i} \in \mathbb{R}^{|\mathcal{E}|}$ denotes the historical frequency of query q at timestamp i, and Norm represents the normalization operation. During the decoding phase, we leverage the frequency representation to constrain the candidate entity space, thereby enhancing the contextual features of the query.

4.3 Context-Enhanced Decoder

We employ a Context-Enhanced Decoder to facilitate more accurate predictions via the derived encoded temporal contextual information. Specifically, we utilize the obtained representations to calculate association, local frequency, and aggregated path decoding scores for each candidate. These scores are then aggregated to yield the final scores.

We employ ConvTransE (Shang et al., 2019) to obtain the decoding scores. The association score captures solely the relevance between the query conditions and the candidate. Based on this, the local frequency score takes into account the local historical frequency information and masks candidates that have not appeared in the local history. The two scores are calculated as follows:

$$\mathbf{s}^{ass} = \mathbf{H}_{t_q}\text{ConvTransE}(\mathbf{h}_{s,t_q}, \mathbf{r}_{t_q}) \quad (13)$$

$$\mathbf{s}^{loc} = \mathbf{H}_{t_q}\text{ConvTransE}(\mathbf{h}_{s,t_q}, \mathbf{r}_{t_q}) \odot \mathbf{f}_q \quad (14)$$

where $\odot$ denotes element-wise multiplication. Additionally, we calculate the aggregated path score by leveraging the aggregated path representation of the specific candidate:

$$s_o^{agg} = \mathbf{h}_{o,t_q}\text{ConvTransE}(\mathbf{h}_{s,t_q}, \mathbf{u}_o^{agg}) \quad (15)$$

For the candidates unreachable by any rule, we set the corresponding $\mathbf{u}_o^{agg}$ to a zero vector and ultimately yield the aggregated path score vector $\mathbf{s}^{agg}$. Subsequently, we aggregate the scores to compute the final scores for the candidates:

$$\mathbf{s}^{fin} = \alpha^{ass} \times f_a(\mathbf{s}^{ass}) + \alpha^{loc} \times f_a(\mathbf{s}^{loc}) + \alpha^{agg} \times f_a(\mathbf{s}^{agg}) \quad (16)$$

where $\mathbf{s}^{fin} \in \mathbb{R}^{|\mathcal{E}|}$, $\alpha^{ass}; \alpha^{loc}; \alpha^{agg} \in [0, 1]$ and $\alpha^{ass} + \alpha^{loc} + \alpha^{agg} = 1$. f_a denotes the softmax activation function.

Model	ICEWS14				ICEWS18				ICEWS05-15			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
RE-NET	39.86	30.11	44.02	58.21	29.78	19.73	32.55	48.46	43.67	33.55	48.83	62.72
RE-GCN	42.00	31.63	47.20	61.65	32.62	22.39	36.79	52.68	48.03	37.33	53.90	68.51
TANGO	36.48	26.90	41.03	54.82	28.97	19.51	32.61	47.51	42.86	32.72	48.14	62.34
CEN	41.90	31.74	46.97	61.55	31.84	21.95	35.89	51.26	46.96	36.50	52.56	67.13
TiRGN	44.04	33.83	48.95	63.84	33.66	23.19	37.99	**54.22**	50.04	39.25	56.13	70.71
HisMatch	45.15	**35.12**	49.88	64.51	33.26	23.17	37.05	52.69	**52.79**	**42.15**	**58.93**	**73.22**
xERTE	40.79	32.70	45.67	57.30	29.31	21.03	33.51	46.48	46.62	37.84	52.31	63.92
TITer	41.73	32.74	46.46	58.44	29.98	22.05	33.46	44.83	47.60	38.29	52.74	64.86
TLogic	42.53	33.20	47.61	60.29	29.59	20.42	33.60	48.05	46.94	36.16	53.24	67.21
TECHS	43.88	34.59	49.36	61.95	30.85	21.81	35.39	49.82	48.38	38.34	54.69	68.92
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

Table 1: Performance for entity prediction task on ICEWS14, ICEWS18 and ICEWS05-15. The best results are bolded, and the second-best results are underlined.

Model	YAGO				WIKI				GDELT			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
RE-NET	66.93	58.59	71.48	86.84	58.32	50.01	61.23	73.57	19.55	12.38	20.80	34.00
RE-GCN	82.30	78.83	84.27	88.58	78.53	74.50	81.59	84.70	19.69	12.46	20.93	33.81
TANGO	63.34	60.04	65.19	68.79	53.04	51.52	53.84	55.46	19.66	12.50	20.93	33.55
CEN	83.51	79.86	85.88	89.80	78.88	74.98	81.91	84.98	20.41	12.94	21.79	35.08
TiRGN	87.95	84.34	91.37	92.92	**81.65**	**77.77**	85.12	**87.08**	21.67	13.63	23.27	37.60
HisMatch	76.55	71.96	79.28	85.39	71.94	67.12	74.86	80.49	22.23	14.56	24.12	36.93
xERTE	84.19	80.09	88.02	89.78	73.60	69.05	78.03	79.73	19.45	11.92	20.84	34.18
TITer	87.47	80.09	89.96	90.27	73.91	71.70	75.41	76.96	18.19	11.52	19.20	31.00
TLogic	78.76	74.31	83.38	83.72	78.93	73.05	84.97	86.91	19.83	12.27	21.74	35.72
TECHS	89.24	-	-	92.39	75.98	-	-	82.39	-	-	-	-
CRAFT	**90.23**	**87.56**	**92.81**	**93.19**	81.32	77.21	**85.36**	86.86	**23.78**	**15.38**	**26.23**	**40.15**

Table 2: Performance for entity prediction task on YAGO, WIKI and GDELT. The best results are bolded, and the second-best results are underlined.

4.4 Parameter Learning

We regard entity prediction as a multi-label learning task, and we employ the cross-entropy function to compute the loss:

$$L = -\sum_{t=1}^{T} \sum_{(s,r,o,t)\in\mathcal{G}_t} \log p(o|s,r,t,\{\mathcal{G}_i\}_{i=1}^{t-1}) \tag{17}$$

where T represents the maximum timestamp in the training set, and $p(o|s,r,t,\{\mathcal{G}_i\}_{i=1}^{t-1}) = s_o^{fin}$ denotes the prediction probability of query answer entity o.

5 Experiments

5.1 Experimental Setup

In this section, we introduce the basic experimental setup for our model, and more details are presented in Appendix B.

Datasets & Evaluation Metrics. We conducted experiments on six TKG benchmarks, including ICEWS14, ICEWS18, ICEWS05-15 (García-Durán et al., 2018), YAGO (Mahdisoltani et al., 2013), WIKI (Leblay and Chekol, 2018) and GDELT (Jin et al., 2019). Moreover, we adopted two widely used metrics to evaluate the performance of the models on TKGF: Mean Reciprocal Rank (MRR) and Hits@k (H@k) in percentage (%). Meanwhile, we perform experiments under time-aware filtering settings (Han et al., 2020) to filter out other correct entities.

Baselines. We compare CRAFT with previous well-performing TKGF models, which can be divided into representation-based and path-based. The former includes RE-NET (Jin et al., 2019), RE-GCN (Li et al., 2021b), TANGO (Han et al., 2021), CEN (Li et al., 2022b) , TiRGN (Li et al., 2022a) and HisMatch (Li et al., 2022c). The latter includes xERTE (Han et al., 2020), TITer (Sun et al., 2021), TLogic (Liu et al., 2022) and TECHS (Lin et al.,

2023).

5.2 Results

The results are shown in Table 1 and Table 2. We can observe that CRAFT outperforms the representation-based models on most metrics, indicating that modeling candidate-specific temporal context can effectively enhance the accuracy of predictions. For path-based approaches, CRAFT performs better on all metrics, as it fully exploits the associated information within the KG sequences. Notably, although TECHS also attempts to combine rule learning and entity representations, it decodes based on the simple aggregated representations of entity and time information, which limits the expressiveness of entities. In contrast, CRAFT utilizes semantically rich evolutional representations, achieving superior performance.

Model	ICEWS14		ICEWS18	
	MRR	H@3	MRR	H@3
base	42.84	47.70	32.58	36.63
+HRCE	43.96	49.70	33.18	37.63
+LFCE	44.54	50.14	33.66	38.10
+LFCE+saa	45.38	50.79	33.91	38.27
+LFCE+taa	45.29	50.80	33.96	38.44
CRAFT	**45.71**	**51.83**	**34.21**	**38.53**

Table 3: Ablation studies of CRAFT on ICEWS14, ICEWS18. **base** means only utilize evolutional representation encoding. **HRCE** is Historical Relevant Context Encoder, **LFCE** is Local Frequency Context Encoder, **saa** is semantic alignment attention, **taa** is time awareness attention.

5.3 Ablation Study

To verify the effectiveness of each component of CRAFT, we conducted ablation experiments, and the results are shown in Table 3. Augmenting the base model with either the Historical Relevant Context Encoder (+HRCE) or the Local Frequency Context Encoder (+LFCE) led to notable improvements in model performance, indicating that capturing temporal contextual features from historical relevant paths and contextual information from repetitive facts facilitates more accurate predictions. Furthermore, the addition of semantic alignment attention (+saa) and time awareness attention (+taa) to the encoded path representations resulted in better model performance, suggesting the necessity of capturing more crucial contextual information, and utilizing the dual-attention mechanism (CRAFT) can further elevate the model capability.

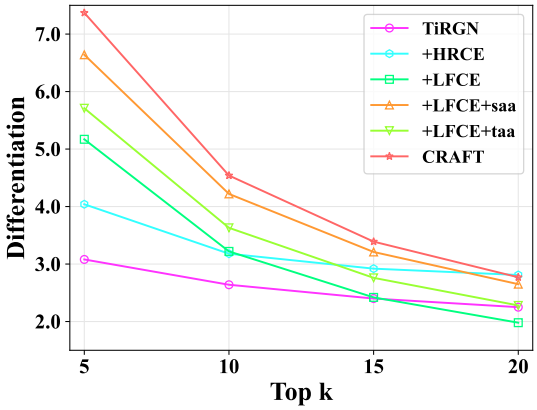


Figure 3: Differentiation analysis of diverse models. We quantify the differentiation by calculating the variance of the scores of the top k candidates for all queries in the test set.

5.4 Why Can Contextual Information Enhance Representation Learning?

Previous representation-based methods use the same temporal context for predictions. We believe that by modeling the candidate-specific temporal context, we can assign higher scores to candidates with effective contextual information, thus enabling differential prediction and improving the model's performance. To validate this assumption, we conduct a differentiation analysis, which is illustrated in Figure 3. It can be observed that the differentiation of the SOTA model TiRGN is minimal. By incrementally incorporating components onto the base model, the differentiation of the prediction scores progressively increases, reaching its apex when all components are engaged, which is consistent with our analysis. Specifically, when k is small, only a few top-ranked candidates have sufficient temporal contextual support for prediction, leading to a pronounced distinctiveness in scores. As k increases, the temporal contextual information available to lower-ranked candidates becomes relatively sparse, and thus, the score variance decreases accordingly. For the +LFCE variant, given that most candidates exhibit similar or identical local historical frequencies, the differentiation diminishes as k increases.

5.5 Case Study

We demonstrate how CRAFT models temporal context through a case study, which is presented in Table 4. For the answer entity *Oman*, CRAFT leverages the dual-attention mechanism to assign higher weights to important historical relevant

Candidate	Contextual Information	saa(%)	taa(%)	s_C	s_T
Oman	*Iran* $\xrightarrow[\text{2014-11-29}]{\text{Express intent to cooperate}}$ *Oman*	33.51	11.52	**13.49** (0.00%)	13.26 (↓0.45%)
	Iran $\xrightarrow[\text{2014-11-29}]{\text{Consult}}$ *Oman*	32.70	11.52		
	Iran $\xrightarrow[\text{2014-11-30}]{\text{Engage in diplomatic cooperation}^{-1}}$ *Oman*	15.04	13.37		
Iraq	*Iran* $\xrightarrow[\text{2014-11-14}]{\text{Express intent to cooperate}}$ *Iraq*	32.71	7.35	10.07 (↓25.35%)	**13.32** (0.00%)
	Iran $\xrightarrow[\text{2014-11-12}]{\text{Make Statement}^{-1}}$ *Iraq*	19.39	7.36		
	Iran $\xrightarrow[\text{2014-11-26}]{\text{Sign formal agreement}}$ *Iraq*	11.56	10.42		
France	*Iran* $\xrightarrow[\text{2014-11-29}]{\text{Engage in negotiation}^{-1}}$ *France*	35.76	52.97	8.78 (↓34.91%)	12.77 (↓4.13%)
	Iran $\xrightarrow[\text{2014-11-06}]{\text{Reject plan, agreement to settle dispute}^{-1}}$ *France*	21.23	23.52		

Table 4: Case study for query *(Iran, Express intent to cooperate, Oman, 2014-12-01)*. We report part of contextual information, attention scores, and prediction scores for the top 3 candidates. The contextual information is sorted in descending order according to the average of the two attentions. s_C denotes the score given by the CRAFT model, and s_T represents the score given by the TiRGN model.

paths, thereby fully utilizing related contextual information to achieve correct prediction. For the candidates *Iraq* and *France*, CRAFT also captures semantically and temporally critical contextual features. However, the temporal validity of fact *(Iran, Express intent to cooperate, Iraq, 2014-11-14)* is relatively low, and facts *(Iran, Sign formal agreement, Iraq, 2014-11-26)* and *(Iran, Reject plan, France, 2014-11-06)* exhibit weak semantic correlation with the query. These factors contribute to the temporal contexts of candidates *Iraq* and *France* not sufficiently supporting them as the correct answer. Furthermore, the scoring of candidates by different models aligns with our analysis in Section 5.4. Due to the utilization of identical temporal context, the SOTA model TiRGN struggles to achieve distinctive predictions. In contrast, within the framework of the CRAFT, candidates with effective temporal contexts receive higher scores, while those without receive lower scores. Additional case studies are presented in Appendix C.

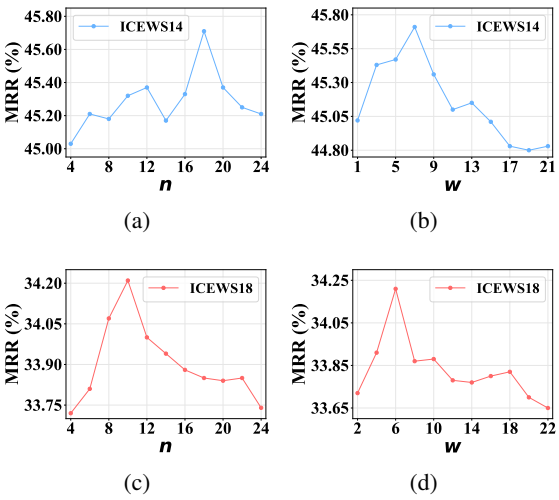


Figure 4: Hyperparameter analysis of the number of aggregated paths n and the window size of Local Frequency Context Encoder w on datasets ICEWS14 and ICEWS18.

5.6 The Impact of Contextual Information Quantity

To investigate the impact of the amount of contextual information used on the prediction performance of the CRAFT model, we conduct further analysis of two key hyperparameters: the number of aggregated paths n and the window size of Local Frequency Context Encoder w. The results are shown in Figure 4. We observe that the model's performance deteriorates when n is either too small or too large. We believe this is because when the number of aggregated paths is inadequate, the model fails to fully leverage the contextual information pertinent to the candidates; conversely, an excessive amount of paths inevitably introduces noise, thereby affecting the precision of predictions. Additionally, w exhibits a parallel trend, indicating that a window that is too narrow prevents the exploitation of effective historical frequency information, while events that occurred far in the past struggle to support the recurrence of future events. Therefore, it is necessary to use an appropriate amount of temporal contextual information to sufficiently support the prediction results while simultaneously mitigating the introduction of noise, thereby facilitating more accurate predictions.

6 Conclusion

In this paper, we propose a novel representation-based model for TKGF, namely CRAFT, to model the candidate-specific temporal context from two aspects: historical relevant context and local frequency context. We first employ a Historical Relevant Context Encoder to extract contextual features in historical relevant paths of candidates and utilize a Local Frequency Context Encoder to capture contextual information in repetitive facts. Then, we apply a Context-Enhanced Decoder to achieve more accurate predictions via the obtained representation. The experiments on six widely-used TKG datasets demonstrate the effectiveness of CRAFT and can achieve differential predictions.

7 Limitations

Our proposed model has the following limitations: First, we only utilized the relation chains and timestamps in the historical relevant paths of candidates while neglecting the entity chain information within, which may also be beneficial for temporal context modeling. Second, we designed two attention functions that effectively capture important contextual information in relevant paths; however, more complex and expressive attention functions may be available. Last but not least, due to the introduction of historical relevant paths, both the spatial and temporal costs of model training have increased.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant No. 62276110, No. 62172039 and in part by the fund of Joint Laboratory of HUST and Pingan Property & Casualty Research (HPL). The authors would also like to thank the anonymous reviewers for their comments on improving the quality of this paper.

References

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8.

Elizabeth Boschee, Jennifer Lautenschlager, Sean O'Brien, Steve Shellman, James Starz, and Michael Ward. 2015. ICEWS Coded Event Data.

Zhuojun Ding, Wei Wei, Xiaoye Qu, and Dangyang Chen. 2024. Improving pseudo labels with global-local denoising framework for cross-lingual named entity recognition. *arXiv preprint arXiv:2406.01213*.

Chenghao Fan, Wei Wei, Xiaoye Qu, Zhenyi Lu, Wenfeng Xie, Yu Cheng, and Dangyang Chen. 2024. Enhancing low-resource relation representations through multi-view decoupling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17968–17976.

Alberto García-Durán, Sebastijan Dumančić, and Mathias Niepert. 2018. Learning sequence encoders for temporal knowledge graph completion. *arXiv preprint arXiv:1809.03202*.

Simon Gottschalk and Elena Demidova. 2018. Eventkg: A multilingual event-centric temporal knowledge graph. In *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15*, pages 272–287. Springer.

Zhen Han, Peng Chen, Yunpu Ma, and Volker Tresp. 2020. Explainable subgraph reasoning for forecasting on temporal knowledge graphs. In *International Conference on Learning Representations*.

Zhen Han, Zifeng Ding, Yunpu Ma, Yujia Gu, and Volker Tresp. 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*, pages 8352–8364.

Rikui Huang, Wei Wei, Xiaoye Qu, Wenfeng Xie, Xianling Mao, and Dangyang Chen. 2024a. Joint multi-facts reasoning network for complex temporal question answering over knowledge graph. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 10331–10335. IEEE.

Rikui Huang, Wei Wei, Xiaoye Qu, Shengzhe Zhang, Dangyang Chen, and Yu Cheng. 2024b. Confidence is not timeless: Modeling temporal validity for rule-based temporal knowledge graph forecasting. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10783–10794.

Tingsong Jiang, Tianyu Liu, Tao Ge, Lei Sha, Sujian Li, Baobao Chang, and Zhifang Sui. 2016. Encoding temporal information for time-aware link prediction. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 2350–2354.

Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. 2019. Recurrent event network: Autoregressive structure inference over temporal knowledge graphs. *arXiv preprint arXiv:1904.05530*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Mario Krenn, Lorenzo Buffoni, Bruno Coutinho, Sagi Eppel, Jacob Gates Foster, Andrew Gritsevskiy, Harlin Lee, Yichao Lu, João P Moutinho, Nima Sanjabi, et al. 2023. Forecasting the future of artificial intelligence with machine learning-based link prediction in an exponentially growing knowledge network. *Nature Machine Intelligence*, 5(11):1326–1335.

Julien Leblay and Melisachew Wudage Chekol. 2018. Deriving validity time in knowledge graph. In *Companion proceedings of the the web conference 2018*, pages 1771–1776.

Ningyuan Li, E Haihong, Shi Li, Mingzhi Sun, Tianyu Yao, Meina Song, Yong Wang, and Haoran Luo. 2023. Tr-rules: Rule-based model for link forecasting on temporal knowledge graph considering temporal redundancy. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 7885–7894.

Yujia Li, Shiliang Sun, and Jing Zhao. 2022a. Tirgn: Time-guided recurrent graph network with local-global historical patterns for temporal knowledge graph reasoning. In *IJCAI*, pages 2152–2158.

Zixuan Li, Saiping Guan, Xiaolong Jin, Weihua Peng, Yajuan Lyu, Yong Zhu, Long Bai, Wei Li, Jiafeng Guo, and Xueqi Cheng. 2022b. Complex evolutional pattern learning for temporal knowledge graph reasoning. *arXiv preprint arXiv:2203.07782*.

Zixuan Li, Zhongni Hou, Saiping Guan, Xiaolong Jin, Weihua Peng, Long Bai, Yajuan Lyu, Wei Li, Jiafeng Guo, and Xueqi Cheng. 2022c. Hismatch: Historical structure matching based temporal knowledge graph reasoning. *arXiv preprint arXiv:2210.09708*.

Zixuan Li, Xiaolong Jin, Saiping Guan, Wei Li, Jiafeng Guo, Yuanzhuo Wang, and Xueqi Cheng. 2021a. Search from history and reason for future: Two-stage reasoning on temporal knowledge graphs. *arXiv preprint arXiv:2106.00327*.

Zixuan Li, Xiaolong Jin, Wei Li, Saiping Guan, Jiafeng Guo, Huawei Shen, Yuanzhuo Wang, and Xueqi Cheng. 2021b. 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*, pages 408–417.

Qika Lin, Jun Liu, Rui Mao, Fangzhi Xu, and Erik Cambria. 2023. Techs: Temporal logical graph networks for explainable extrapolation reasoning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1281–1293.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29.

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 *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 4120–4127.

Farzaneh Mahdisoltani, Joanna Biega, and Fabian M Suchanek. 2013. Yago3: A knowledge base from multilingual wikipedias. In *CIDR*.

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 *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3060–3067.

Alessio Signorini, Alberto Maria Segre, and Philip M Polgreen. 2011. The use of twitter to track levels of disease activity and public concern in the us during the influenza a h1n1 pandemic. *PloS one*, 6(5):e19467.

Haohai Sun, Jialun Zhong, Yunpu Ma, Zhen Han, and Kun He. 2021. Timetraveler: Reinforcement learning for temporal knowledge graph forecasting. *arXiv preprint arXiv:2109.04101*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

LIANG Zhao. 2020. Event prediction in big data era: A systematic survey. arxiv preprint. *ArXivorg*.

A Generation of Historical Relevant Paths

Taking query $(s, r, ?, t_q)$ as an example. For each temporal rule $tr \in TR$, we ground the rule body by instantiating the variables with historical facts in a KG sequence and ensuring $E_1 = s$. The object entity of the last edge in body grounding, which is in correspondence with the object entity in the rule head, is the rule-reachable candidate entity. Notably, a single rule can reach multiple candidates, concurrently, a candidate can be accessible via multiple rules, and the body groundings of reachable rules constitute the historical relevant paths of candidates. For example, for the query *(Iran, Express intent to cooperate, ?, 2014-12-01)* and its corresponding temporal rule: $(E_1, Express\ intent\ to\ cooperate, E_2, T_2) \leftarrow (E_1, Consult, E_2, T_1)$, we can obtain the historical relevant path *(Iran, consult, Oman, 2014-11-29)* and the rule-reachable candidate *Oman* by rule application.

B Details of Experimental Setup

Datasets & Evaluation Metrics. Following (Jin et al., 2019), we split each dataset chronologically into training, validation, and test sets with a proportion of 80%, 10%, and 10%. The statistics of the datasets are presented in Table 5. For each query, the model calculates a rank for the correct entity. MRR denotes the average of the reciprocal rank of all query answers, and Hits@k is the proportion of query answers that rank in the top k. Specifically, we report Hits@{1, 3, 10}.

Implementation. For the computation of all decoding scores, we employ the time vectors as introduced in (Li et al., 2022a). Following (Li et al., 2021b), we add static graph constraints to all ICEWS datasets. The embedding dimension d is set to 200 for all datasets. We obtain the optimal values of some hyperparameters by grid search on the validation sets and grid search range is as follows: the number of paths to aggregate $n \in \{5, 6, ..., 20\}$, the time decay coefficient $\lambda \in \{0.1, 0.2, ..., 0.9\}$, the window size of Local Frequency Context Encoder $w \in \{1, 2, ..., 15\}$, score coefficients $\alpha^{ass}; \alpha^{loc}; \alpha^{agg} \in \{0.1, 0.2, ..., 0.9\}$. Notably, for GDELT, the search range of w is set to $\{20, 30, ..., 100\}$. The specific values of hyperparameters set on various datasets are presented in Table 6. The maximum training epoch is set to 40 for all datasets and the early stopping strategy is utilized to mitigate the risk of overfitting. We utilize Adam (Kingma and Ba, 2014) for parameter learning and the learning rate is set to 0.001. Regarding the remaining hyperparameters, we adhered to the settings prescribed by (Li et al., 2022a) and (Liu et al., 2022).

C Case Study

The results of the case study are presented in Tables 7 and 8. For query *(Japan, Consult, China, 2014-12-08)*, the temporal context of the candidates *China* and *South Korea* both effectively support them as prediction results, hence CRAFT calculates similar scores for them. Taking into account both the semantic relevance and temporal validity of the historical relevant paths, CRAFT ultimately predicts the correct answer as *China*. For the candidate *North Korea*, due to insufficient contextual information, CRAFT calculates a lower score.

For query *(Cambodia, Engage in diplomatic cooperation, Vietnam, 2014-12-25)*, CRAFT similarly captures important contextual features for each candidate. The candidates *Thailand* and *China* not only exhibit relevant paths semantically correlated with the query, such as *(Thailand, Cooperate economically, Cambodia, 2014-12-15)* and *(Cambodia, Engage in diplomatic cooperation, Thailand, 2014-12-02)*, but also demonstrate closer temporal interactions with the subject entity of query *Cambodia*, including *(Cambodia, Host a visit, Thailand, 2014-12-24)* and *(China, Sign formal agreement, Cambodia, 2014-12-24)*. However, CRAFT concurrently considers the significance of both semantic and temporal aspects rather than evaluating them separately. Therefore, by employing the dual-attention aggregation and context-enhanced decoding, CRAFT predicts the candidate *Vietnam* as the correct answer.

Datasets	#Entities	#Relations	#Train	#Valid	#Test	#Timestamps	#Time interval
ICEWS14	6,869	230	74,845	8,514	7,371	365	24 hours
ICEWS18	23,033	256	373,018	45,995	49,995	304	24 hours
ICEWS05-15	10,094	251	368,868	46,302	46,159	4,017	24 hours
YAGO	10,623	10	161,540	19,523	20,026	189	1 year
WIKI	12,554	24	539,286	67,538	63,110	232	1 year
GDELT	7,691	240	1,734,399	238,765	305,241	2,976	15 minutes

Table 5: The Statistics of datasets. #Train, #Valid and #Test are the numbers of quadruples in the training, validation and test sets.

Datasets	n	λ	w	α^{ass}	α^{loc}	α^{agg}
ICEWS14	18	0.2	7	0.4	0.5	0.1
ICEWS18	10	0.7	6	0.8	0.1	0.1
ICEWS05-15	16	0.9	10	0.7	0.2	0.1
YAGO	15	0.3	1	0.2	0.7	0.1
WIKI	8	0.7	3	0.6	0.3	0.1
GDELT	10	0.2	100	0.5	0.4	0.1

Table 6: The hyperparameter settings for the dataset.

Candidate	Contextual Information	saa(%)	taa(%)	s_C	s_T
China	*Japan* $\xrightarrow[2014\text{-}12\text{-}01]{\text{Express intent to meet or negotiate}}$ *China* $\xrightarrow[2014\text{-}12\text{-}04]{\text{Make a visit}^{-1}}$ *Japan* $\xrightarrow[2014\text{-}12\text{-}05]{\text{Consult}}$ *China*	34.47	5.15	**14.03** (0.00%)	13.91 (↓2.66%)
	Japan $\xrightarrow[2014\text{-}12\text{-}01]{\text{Engage in diplomatic cooperation}}$ *China*	27.04	5.15		
	Japan $\xrightarrow[2014\text{-}11\text{-}29]{\text{Express intent to cooperate}}$ *Iran*	18.81	4.75		
South Korea	*Japan* $\xrightarrow[2014\text{-}12\text{-}05]{\text{Consult}}$ *South Korea*	59.44	7.79	13.98 (↓0.36%)	**14.29** (0.00%)
	Japan $\xrightarrow[2014\text{-}11\text{-}27]{\text{Engage in negotiation}}$ *South Korea*	14.26	5.03		
	Japan $\xrightarrow[2014\text{-}11\text{-}22]{\text{Express intent to cooperate}}$ *South Korea*	8.80	4.69		
North Korea	*Japan* $\xrightarrow[2014\text{-}11\text{-}20]{\text{Express intent to meet or negotiate}^{-1}}$ *North Korea*	99.89	20.18	6.84 (↓51.25%)	13.69 (↓4.20%)

Table 7: Case study for query *(Japan, Consult, China, 2014-12-08)*.

Candidate	Contextual Information	saa(%)	taa(%)	s_C	s_T
Vietnam	*Cambodia* $\xrightarrow[2014\text{-}12\text{-}15]{\text{Cooperate economically}^{-1}}$ *Vietnam*	69.74	4.48	**17.36** (0.00%)	16.41 (↓7.34%)
	Cambodia $\xrightarrow[2014\text{-}12\text{-}18]{\text{Engage in diplomatic cooperation}}$ *Vietnam*	28.23	5.00		
Thailand	*Cambodia* $\xrightarrow[2014\text{-}12\text{-}15]{\text{Cooperate economically}^{-1}}$ *Thailand*	68.34	4.57	8.44 (↓51.38%)	**17.71** (0.00%)
	Cambodia $\xrightarrow[2014\text{-}12\text{-}24]{\text{Host a visit}}$ *Thailand*	10.31	9.06		
	Cambodia $\xrightarrow[2014\text{-}12\text{-}02]{\text{Engage in diplomatic cooperation}}$ *Thailand*	13.99	4.69		
China	*Cambodia* $\xrightarrow[2014\text{-}12\text{-}24]{\text{Sign formal agreement}^{-1}}$ *China*	74.39	17.26	8.04 (↓53.69%)	15.92 (↓10.11%)
	Cambodia $\xrightarrow[2014\text{-}12\text{-}18]{\text{Engage in diplomatic cooperation}^{-1}}$ *Vietnam* $\xrightarrow[2014\text{-}12\text{-}22]{\text{Host a visit}}$ *China*	14.37	9.74		

Table 8: Case study for query *(Cambodia, Engage in diplomatic cooperation, Vietnam, 2014-12-25)*.