

NOETHER EMBEDDINGS: FAST TEMPORAL ASSOCIATION MINING

Anonymous authors

Paper under double-blind review

ABSTRACT

Simple and expressive in representing multi-relational events, temporal knowledge graphs (TKGs) have attracted increasing research interest. While temporal associations (TAs) reveal cause-effect relationships between event pairs across time, to the best of our knowledge, no previous work has paid attention to exploring such basic but meaningful regularities on TKGs. Despite the importance, temporal association mining (TAM) is prohibitively challenging due to its enormous time complexity. Inspired by Noether’s theorem in theoretical physics, we develop Noether Embeddings (NE), an embedding model of structured events that naturally encodes the time translation symmetry of TAs, and can efficiently decode conserved quantities to query TAs. Using NE, a three-stage TAM framework is developed, respectively of the encoding, decoding, and selecting stages. The major time complexity of TAM is therefore rearranged into the encoding stage, where the training time of NE can be ignored as long as it is shorter than the update period of the event stream. We successfully mined TAs both with semantic interpretability and statistical reliability. Experiments show that our method achieves an 11.0 times speedup over an optimized search algorithm for TAM on ICEWS14 dataset.

1 INTRODUCTION

Temporal knowledge graph (TKG), introducing the dimension of time into traditional knowledge graphs (KGs), is an emerging and promising field that has received much attention in recent years (García-Durán et al., 2018). Composed of (s, p, o, t) quadruples (Leblay & Chekol, 2018), where s, p, o, t indicates subject, predicate, object, and time, respectively, (Bob, visit, KFC, 2019-10-10), TKGs naturally contain scales of information over multi-relational events such as global political events (Trivedi et al., 2017), financial incidents (Yang et al., 2019), and user-item behaviors (Xiao et al., 2020), which are well represented in the form of KGs that incorporate abundant semantic information into sparse graph representations.

Efforts have been made to discover a variety of patterns over TKGs for different purposes. In particular, horn rules with temporal constraints and subgraphs connecting entities constitute the two main structures of the explored patterns for interpreting link forecasts (Liu et al., 2022), reasoning over events (Omran et al., 2019), etc. However, none of them found regularities invariant to time shifts. Patterns have been captured in embedding models constructed for TKGs in recent years (Cai et al., 2022). Specifically, inference patterns such as symmetry, inversion, and composition are implicitly represented in models of TeRo (Xu et al., 2020a), 3DRTE (Wang et al., 2020), TBox (Messner et al., 2022), etc. However, to the best of our knowledge, none of these embedding models can explicitly decode the patterns they encode.

Noticing the temporally associated nature of major events that repeatedly occur (Zhao, 2021), and aware that cause-effect pairs are common in social incidents (Radinsky et al., 2012; Lei et al., 2019), we propose a new problem of temporal association mining (TAM) on TKGs. For instance, State A may reduce diplomatic relations with State B in two days whenever State B declares war on State A. Information on entities, predicates, and relative time jointly determines the underlying causal relationship between the event pairs in the above example, implying the existence of massive similar but unknown temporal associations (TAs) in TKGs. TAM on TKGs can greatly expand human understanding of regularities in critical social events (Enthusiasts & Cartledge). In addition,

TAM also has great potential to aid downstream applications such as TKG completion (García-Durán et al., 2018) and event prediction (Ramakrishnan et al., 2014).

The fact that events in TKGs are dynamically-evolving (Jin et al., 2019; Trivedi et al., 2017) leads to a strong demand for efficient TAMs. Continuously generating events in TKGs, the social system behind is in general chaotic (Cherichi & Larodec, 2016), where event regularities emerge and break (Abul-Magd & Simbel, 2004). Therefore, TAs change occasionally. In addition, new types of events may emerge at any time, and the distribution of event occurrences is also highly heterogeneous (Mirtaheri et al., 2020). Such properties of TKGs jointly call for an efficient TAM towards its practical use in potential stream processing scenarios (Wrench et al., 2016). A prerequisite for TAM, is a more efficient implementation of temporal association querying (TAQ).

Despite the importance, performing efficient TAM on TKGs is prohibitively challenging. Primarily, the $O(N)$ time complexity and the large data volume of TKGs jointly result in computing difficulties. Taking GDELT as an example, a Google-made social sensor monitors the whole world and updates thousands of global events every 15 minutes (Leetaru & Schrodt, 2013). Larger TKGs are being constructed in our fast-growing big data era. It is far from satisfactory if we have to compute by a time complexity linear to the number of relevant events of a potential TA for efficient TAQ on such TKGs. Besides, the heterogeneous characteristics of semantic space and time in (s, p, o, t) quadruples demand a unified processing framework.

Inspired by Noether’s theorem, we develop a highly effective and computing-efficient mining framework based on a decodable embedding model to overcome these problems. Symmetry governs regularities in nature (Tanaka & Kunin, 2021), long before Noether proved the equivalence between symmetries and conservation laws in a physical system (Noether, 1918). By constructing Noether Embeddings (NE) whose decoding functions correspond to TAs, local symmetries are discovered by searching conserved quantities. Our embedding space allows the encoding process to convert the query time complexity into the storage space complexity of complex vectors. We thus reduce the time complexity in TAQ of $O(N)$ to a minor constant of $O(T_r D)$, where N refers to the number of relevant events of each queried TA, T_r refers to the queried range of relative time points and D refers to the dimension of vectors, and therefore significantly accelerate TAM.

Our main contributions are listed as follows:

1. (Methodological) (1) We develop an embedding model NE of structured events that naturally encodes the time translation symmetry of TAs, and can efficiently decode conserved quantities to query TAs. (2) We develop a TAM framework that supports efficient query of an approximate probability distribution of the relative time points within an arbitrary queried range of a potential TA. It rearranges the major time complexity of TAM into the encoding stage, where the training time of NE can be ignored as long as it is shorter than the update period of the event stream. (3) We develop a heuristic construction procedure, where axiomatized physical systems that satisfy Noether’s theorem can aid the discovery of regularities with general symmetries.
2. (Empirical) (1) We formally propose the problem of TAQ and TAM on TKGs and formulate related evaluation metrics. (2) Experiments reveal a significant speed advantage of our TAQ and TAM approach on both synthesized and real-world TKG data, compared to an optimized search algorithm. (3) We demonstrate the practical potential and advantage of our approach in stream processing scenarios. (4) We mine TAs both with semantic interpretability and statistical reliability.
3. (Theoretical) (1) We theoretically prove that our embedding model enforces convergence to a conserved quantity of decoding results implying time translation symmetries within associated event pairs. (2) We demonstrate that TKGs in our representation space are isomorphic to a physical system with action, which enables us to apply Noether’s theorem in theoretical physics.

2 PROBLEM FORMULATION

Here we formally present the problem of TAQ and TAM on TKGs after clarifying some key concepts and formulating relevant evaluation metrics.

2.1 CONCEPTS

Event Type For a TKG composed of (s, p, o, t) quadruples, we regard events with the same (s, p, o) triples as belonging to the same event type $ev : s, p, o$. For example, $(\text{Bob}, \text{visit}, \text{KFC}, 2019-10-10)$ and $(\text{Bob}, \text{visit}, \text{KFC}, 2020-11-12)$ both pertain to the event type of $(\text{Bob}, \text{visit}, \text{KFC})$.

TA TAs refer to associations invariant to time shifts. We define forward and reverse TAs as below:

$$(s, p_c, o, t) \rightarrow (s, p_e, o, t + \Delta) \quad \forall t \in \mathbb{T}_a \quad (1)$$

$$(s, p_c, o, t) \rightarrow (o, p_e, s, t + \Delta) \quad \forall t \in \mathbb{T}_a \quad (2)$$

In both formulas, $p_c, p_e, s, o, t, t + \Delta$ refer to the predicate of cause, the predicate of effect, subject, object, the time when the cause occurs, and the time when the effect occurs. \mathbb{T}_a refers to the complete collection of absolute time points and $\Delta = \Delta(\eta) = [\tau(1 - \eta), \tau(1 + \eta)]$ denotes any time interval whose width is proportional to its mean τ with ratio η .

For example, a TA could be: Whenever State A declares war against State B, State B will reduce diplomatic relations with State A in three to five days, where $\tau = 4, \eta = 0.25, \Delta = [3, 5]$.

Evaluation Metrics We consider only TAs whose Δ has the same η . For a TA ta , if a quadruple $q : (s, p, o, t)$ satisfies the body (indicating cause) of it, we denote it as $b(q; ta)$; if a quadruple satisfies the head (indicating effect) of it, we denote it as $h(q; ta)$. We define its support as:

$$sp(ta; \eta) = n(b(q; ta) \wedge h(q'; ta) \wedge (t' - t) \in \Delta(\eta)) \quad (3)$$

Support denotes the number of quadruple pairs respectively satisfying body and head of the TA.

We respectively define the standard confidence, head coverage, and general confidence of a TA as:

$$sc(ta; \eta) = \frac{sp(ta; \eta)}{n(b(ta))}, \quad hc(ta; \eta) = \frac{sp(ta; \eta)}{n(h(ta))}, \quad gc(ta; \eta) = \frac{2}{\frac{1}{sc(ta; \eta)} + \frac{1}{hc(ta; \eta)}} \quad (4)$$

where $n(b(ta))$ and $n(h(ta))$ respectively represent the number of events whose event type corresponds to the body and head of the TA.

Note that when calculating $sp(ta; \eta)$, we can only count one quadruple once and in one pair to avoid overcounting in the case of events occurring in consecutive periods.

TAs could then be evaluated not only by 0/1 judgement, but also continuously with gc from 0 to 1.

2.2 TASKS

Definition of TAQ We define TAQ as the problem of determining whether a TA holds given (s, o, p_c, p_e) and η , and giving the correct center point τ of $\Delta(\eta)$ so that the TA holds if possible. For a query (s, o, p_c, p_e) given the setting parameter η : if the query is not a TA then an answer of -1 is expected; else, a correct τ needs to be selected, that falls into the ground truth $\Delta(\eta)$. We calculate the total accuracy over all queries as an overall metric in the TAQ task.

Definition of TAM We define TAM as the problem of discovering TAs as fast and as many as possible given a lower bound gc_l of the gc and support sp_l of each discovered TA.

3 INSPIRATIONS FROM NOETHER'S THEOREM

In 1915, mathematician Emmy Noether proved one of the most fundamental theorems in theoretical physics: every differentiable symmetry of the action of a physical system with conservative forces has a corresponding conservation law. Meanwhile, any meaningful information in an axiomatized physical system can be deductively derived once we construct its action. Here we demonstrate the derivation process of Noether's theorem from the action.

Formal statement A physical system can be fully described by its action S .

$$S = \int_{t_1}^{t_2} L(\mathbf{q}(t), \dot{\mathbf{q}}(t), t) dt \quad (5)$$

In the above equation, \mathbf{q} and $\dot{\mathbf{q}}(t)$ represent generalized coordinates and velocities. L represents the Lagrangian, generally referring to the difference between kinetic and potential energy ($L = T - V$). The stationary points of action functional determine the evolution of the whole system. For a physical system with conservative forces, Euler-Lagrange Equation describes such evolution:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{q}}_i} = \frac{\partial L}{\partial \mathbf{q}_i}, i = 1, 2, \dots, n \quad (6)$$

Symmetry denotes that S remains the same under certain infinitesimal transformation with $\epsilon \rightarrow 0$:

$$t \rightarrow t' = t + \delta t = t + \epsilon X(\mathbf{q}, t), \quad \mathbf{q}(t) \rightarrow \mathbf{q}'(t) = \mathbf{q}(t) + \delta \mathbf{q}(t) = \mathbf{q}(t) + \epsilon \Psi(\mathbf{q}, t) \quad (7)$$

Here $\epsilon X(\mathbf{q}, t)$ and $\epsilon \Psi(\mathbf{q}, t)$ respectively denote a first-order approximation of δt and $\delta \mathbf{q}(t)$. By calculating $S[q'(t')] = S[q(t)]$ and Euler-Lagrange Equation, we derive the conserved quantity Λ :

$$\Lambda = \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \Psi(\mathbf{q}, t) + (L - \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \dot{\mathbf{q}}) X(\mathbf{q}, t) \quad (8)$$

Motivation If we regard the cause and effect event types over the relative time of a TA as a mapping across time: $(s, p_c, o) \rightarrow (s, p_e, o), t \rightarrow t + \tau: \forall t \in \mathbb{T}_a$, then TAs indicate invariance of such mappings under the transformation of time translation, which implies local symmetries.

We aim to construct an embedding model that naturally encodes the time translation symmetry of TAs, and a corresponding decoding function $g(\tau)$ that only requires traversing through a fixed range of relative time points $\tau \in \mathbb{T}_r$ to query each TA.

However, such construction is without a guide, and therefore hard. Here we demonstrate how this can be aided by repeating the derivation process of Noether’s theorem shown above.

Heuristic Construction Five stages are needed as below. Note that such a procedure is not limited to aid discovering TAs, but can be extended to aid in exploring regularities of general symmetries.

(1) Construction of action S . Denote $\mathbf{q}(t; s, p, o)$ as the event embeddings of a TKG quadruple. Noticing the symmetry conveyed by a TA with relative time τ , we can analogize the embeddings of a potential cause and effect event pair in a TA to a zero-mass two-particle system connected by a spring. It is a physical system with conservative forces that satisfies the constraints of applying Noether’s theorem. Therefore, the Lagrangian of them is $L = -\|\mathbf{q}_c(t) - \mathbf{q}_e(t + \tau)\|^2$, where the generalized coordinates $\mathbf{q}(t) = (\mathbf{q}_c(t), \mathbf{q}_e(t + \tau))$, and generalized velocities $\dot{\mathbf{q}}(t) = 0$. The action of event samples of a TA is therefore $S = \sum_{t \in \mathbb{T}_a} L$, where \mathbb{T}_a is the set of absolute time points.

(2) Applying Euler-Lagrange Equation. Since $\dot{\mathbf{q}}(t) = 0$, then $\frac{\partial L}{\partial \dot{\mathbf{q}}_i} = 0, i = 1, 2, \dots, n$. Therefore, we can derive $\frac{\partial L}{\partial \mathbf{q}_i} = \frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{q}}_i} = 0, i = 1, 2, \dots, n$.

(3) Applying transformation of symmetry. Since the action remains the same under the transformation of time translation, we can derive $\frac{\partial L}{\partial t} = 0$.

(4) Combining conclusions from (2) and (3). Considering $\frac{\partial L}{\partial \mathbf{q}_i} = 0, i = 1, 2, \dots, n, \frac{\partial L}{\partial t} = 0$, and that $L = -\|\mathbf{q}_c(t) - \mathbf{q}_e(t + \tau)\|^2$ is a quadratic function, we can derive that $\mathbf{q}_c(t) = \mathbf{q}_e(t + \tau)$, and that the conserved quantity is L , which is a constant of zero.

(5) Getting inspirations. First, we can map the event embeddings $\mathbf{q}(t; s, p, o)$ of positive and negative samples to two different constants. In this way, $\mathbf{q}_c(t) = \mathbf{q}_e(t + \tau)$ is satisfied because all sample pairs of the same relative time τ , be them both positive or both negative samples, are mapped to the same constant. $\mathbf{q}_c(t) \neq \mathbf{q}_e(t + (\kappa))$ is also satisfied because the positive and the negative sample in pairs of relative time $\kappa \neq \tau$ are mapped to two different constants. Second, we can set the decoding function as $g(\tau) = \|\mathbf{q}_c(t) - \mathbf{q}_e(t + \tau)\|$ to separate τ s within a queried range of relative time points, which maps τ s implying symmetries to a conserved quantity of $g(\tau) = 0$.

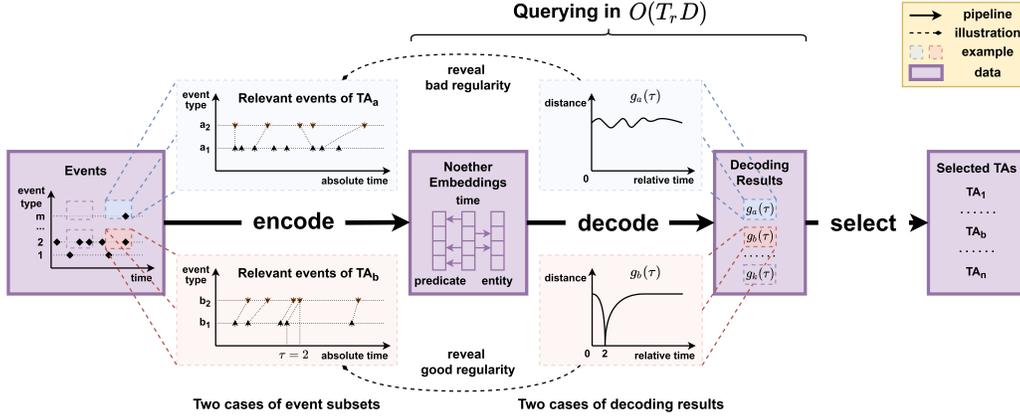


Figure 1: Illustration of our three-stage TAM framework. Solid lines and purple graphs in the middle jointly represent the data flow of the mining framework. In contrast, the two blue graphs above solid lines and the two red ones below demonstrate cases respectively of a TA with significant temporal associations and a TA with not.

4 NOETHER EMBEDDINGS

General Framework As illustrated in Figure 1, structured events are converted to NE of distributed complex vectors for storage at the encoding stage, either in batches or streams. At the decoding stage, TAQ is conducted by directly calculating the relevant vectors of each TA to derive the decoding results of each TA query. At the selecting stage, the extreme value of each decoding result is compared with a global threshold to select TAs with high approximate gcs .

4.1 DETAILED DESCRIPTIONS

Two technical concerns are central to constructing NE.

(1) The decoding function $g(\tau) = \|\mathbf{q}_c(t) - \mathbf{q}_e(t + \tau)\|$ needs to be irrelevant to t ($\forall t \in \mathbb{T}_a$). Only in this way can each TAQ be conducted in a time complexity of a minor constant. Only in this way can NE naturally embeds the time translation symmetry within potential TAs.

(2) The event embeddings $\mathbf{q}(t; s, p, o)$ of an event type is required to fit various functions of absolute time: with fixed s, p, o and changing t , embeddings of positive and negative samples are mapped to different constants, and an event type may occur at arbitrary and multiple time points.

Encoding We then construct the event embeddings of NE as below:

$$\mathbf{q}(t; ev) = \mathbf{u}(ev) \circ \mathbf{r}(t) \quad \text{where} \quad \mathbf{u}(ev : s, p, o) = \mathbf{s} \circ \mathbf{p} \circ \mathbf{o} \quad (9)$$

\circ denotes the Hadamard (or element-wise) product; $\mathbf{s}, \mathbf{p}, \mathbf{o}, \mathbf{r}(t)$ are complex vectors where $\mathbf{s}, \mathbf{p}, \mathbf{o}, \mathbf{r}(t) \in \mathbb{C}^d$, $\mathbf{r}(t) = e^{i\omega t}$, $\omega \in \mathbb{R}^d$, and d is the dimension of vectors; both subjects and objects share the same set of entity embeddings $\mathbf{e} \in \mathbb{C}^d$, while $\mathbf{s} = \mathbf{e}$ and $\mathbf{o} = \bar{\mathbf{e}}$. $\mathbf{u}(ev)$ is listed individually to show that we can extend NE to structured events of general forms. For example, $\mathbf{u}(ev : s, p) = \mathbf{s} \circ \mathbf{p}$ for (s, p, t) triples. The score and loss functions of NE are defined as follows:

$$f(t; ev, C) = \left| \sum_{i=1}^d \text{Re}(\mathbf{q}(t; ev)) - C \right| \quad (10)$$

$$L(\xi; C_p, C_n) = f(\xi; C_p)^2 + \frac{1}{N} \sum f(\xi'; C_n)^2 \quad (11)$$

ξ denotes a positive sample, while ξ' denotes its corresponding negative samples, whose number is N . For a positive sample (s, p, o, t) , its negative samples are the whole set of $\{(s, p, o, t' \neq t)\}$. We set $C = C_p$ and $C = C_n$ respectively for positive and negative samples, where $C_p \neq C_n$ are two different constants.

Decoding Since $\mathbf{r}(t + \tau) = \mathbf{r}(t) \circ \mathbf{r}(\tau)$, $\mathbf{r}(t) \circ \overline{\mathbf{r}(t)} = 1$, we derive the decoding function as:

$$g(\tau) = \|\mathbf{u}_c - \mathbf{u}_e \circ \mathbf{r}(\tau)\| \quad \text{where} \quad g(\tau) = \|(\mathbf{p}_c - \mathbf{p}_e \circ \mathbf{r}(\tau)) \circ \mathbf{s} \circ \mathbf{o}\| \quad \text{for TAs} \quad (12)$$

We can see from above that $g(\tau)$ can be extended to query general forms of one-to-one temporal associations other than the defined TAs, with the same trained embedding model of NE.

τ is traversed through set $\mathbb{T}_r : \{0, 1, 2, \dots, \tau_{max}\}$ to plot the original decoding results. The time complexity of each TAQ is then $O(T_r D)$ by calculating $g(\tau)$ for each TA query. We additionally calculate a converting function $p(\tau) = \frac{\exp[-g(\tau)]}{\sum_{\tau \in \mathbb{T}_r} \exp[-g(\tau)]}$ to form an approximate probability distribution of relative time points for actual decoding use and compute $\max_{\tau \in \mathbb{T}_r} p(\tau)$.

Selecting After obtaining $\max_{\tau \in \mathbb{T}_r} p(\tau)$, we compare it with a global threshold p_{th} to decide whether a (s, o, p_e, p_c) query corresponds to a TA. If $\max_{\tau \in \mathbb{T}_r} p(\tau) > p_{th}$, then a TA is justified as exists and τ is selected that maximizes $p(\tau)$. All potential TAs are queried one by one in this process, which can be regarded as an approximate gc based filtering, but without the need to search through the whole event set to find relevant event pairs and calculate the exact gc .

Data

$$Active : (s, p^{active}, o, t) \iff Passive : (o, p^{passive}, s, t) \quad (13)$$

Since a predicate can be illustrated in either active or passive voice as above, with doubled training quadruples, the decoding function can be calculated to discover both forward and reverse TAs.

4.2 WHY NE WORKS

Notations Within all event embeddings trained for a potential TA, suppose that m event pairs share the same relative time of τ , where m_1 pairs are both positive or negative samples. In each remaining $m_2 = m - m_1$ pair, one is a positive sample while the other is a negative one. Denote that $\mathbb{M} = \{1, 2, \dots, m\}$, $\mathbb{M}_1 = \{i_1, i_2, \dots, i_{m_1}\}$, $\mathbb{M}_2 = \{j_1, j_2, \dots, j_{m_2}\}$. There exists such least upper bound η that $f(t; ev) < \eta$ for the score function of all $2m$ samples. Denote that $\mathbf{a}_i = (\cos(\omega_1 t_i), \cos(\omega_2 t_i), \dots, \cos(\omega_d t_i), \sin(\omega_1 t_i), \sin(\omega_2 t_i), \dots, \sin(\omega_d t_i))^T$, $i \in \mathbb{M}$, where d is the dimension of NE vectors, t_i is time of the i th cause event in m sample pairs. Denote that in the decoding function, $\mathbf{u}_c - \mathbf{u}_e \circ \mathbf{r}(\tau) = \boldsymbol{\alpha} - i\boldsymbol{\beta}$ where $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are both real vectors, and $\mathbf{x} = (\alpha_1, \alpha_2, \dots, \alpha_d, \beta_1, \beta_2, \dots, \beta_d)^T$. Denote that $c_1 = \max_{i \in \mathbb{M}_1} |\cos(\mathbf{a}_i, \mathbf{x})|$, $c_2 = \min_{i \in \mathbb{M}_2} |\cos(\mathbf{a}_i, \mathbf{x})|$.

Theorem 1 Within the m_1 sample pairs, if $c_1 > 0$, then $g(\tau) < \frac{2\eta}{\sqrt{dc_1}}$.

Theorem 2 Within the m_2 sample pairs, if $c_2 > 0$ and $|C_p - C_n| > 2\eta$, then $g(\tau) > \frac{|C_p - C_n| - 2\eta}{\sqrt{dc_2}}$.

For quantities above, we can directly control η in the training process; C_p and C_n are artificially set constants. $c_2 > 0$ is strictly ensured and $c_1 > 0$ is ensured in most situations, which are shown in Appendix A

Implications Two conclusions are drawn from the theorems. (1) Convergence. We can see from theorem 1 that $g(\tau) \rightarrow 0$ as $\eta \rightarrow 0$. (2) Competition. Comparing these two theorems also tells us about the competing effect of well-trained sample pairs for the value of $g(\tau)$, generally affected by the ratio $\frac{m_1}{m_2}$.

Since $g(\tau) = \|\mathbf{u}_c - \mathbf{u}_e \circ \mathbf{r}(\tau)\| = \|\mathbf{u}_c \circ \mathbf{r}(t) - \mathbf{u}_e \circ \mathbf{r}(t + \tau)\| = \|q_c(t) - q_e(t + \tau)\|$ for any t , we can see that relevant sample pairs across time ('for any t ') are therefore automatically 'gathered' and 'averaged' when learning each event embeddings to form the same $g(\tau)$. This is due to the time translation symmetry introduced by $\mathbf{r}(t)$. These effects separate $g(\tau), \tau \in \mathbb{T}_r$ to form a cross-correlation-like function. We show the formal proof and further descriptions in Appendix A.

5 EXPERIMENT

We evaluate NE on two formulated tasks: TAQ and TAM. Baselines in TAQ are: Random Guess, an optimized search algorithm (Algorithm 1) and four classic embedding methods including DE-Simple (Goel et al., 2020), TeRo (Xu et al., 2020a), ATiSE (Xu et al., 2020b) and TNTComplex (Lacroix et al., 2020). Note that baseline embedding models are modified explicitly for experimental adaptations (details in Appendix B.3). Since baseline embedding models are found incapable in TAQ, we compare NE with only Search in TAM evaluation. Notably, the speedup ratio is calculated based on running time without regard to training. We demonstrate that the potential application of NE in practice use is TAQ and TAM in stream processing scenarios where the training time of NE can be ignored as long as it is shorter than the update period of the event stream. Detailed information on datasets, task settings and method implementation is presented in Appendix B.

5.1 RESULTS ON TAQ

Query accuracy Figure 2 shows the query accuracy of NE and other baselines on two real-world datasets added with synthesized samples for TA queries. Within all the TAs queried, 50% has the ground truth of 1 (the other 50% 0), where the η in $\Delta(\eta)$ is set as 0.1. NE achieves query accuracy of 0.812, 0.845, 0.837, 0.806, 0.786 and 0.788 with a different number of event pairs relevant to queried TAs on ICEWS18, and obtains the highest accuracy on ICEWS14, close to 1.0 by Search. On the contrary, other embedding models only achieve query accuracy of around 0.5, similar to Random Guess. These results imply that NE conducts accurate TAQs with high resolution of relative time, while traditional embedding models do not.

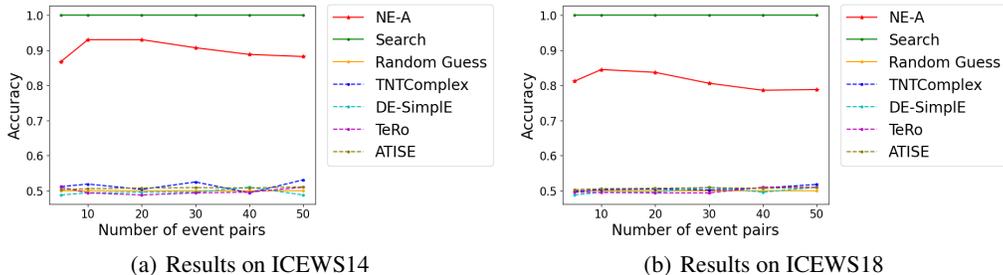


Figure 2: Comparisons of query accuracy over two datasets

Query speedup NE shows significant speed superiority over Search as presented in Table 1. NE (running on a CPU) is 14.1 times faster than Search in querying a synthesized TA of 50 event pairs on the mixed ICEWS18 dataset. We additionally conducted experiments because vectors compose our decoding functions, and vector computation is naturally implemented on a gpu. Since the running time of Search grows linearly with the scale of related event pair numbers, while that of NE remains a constant, our method holds potential for even larger datasets.

Table 1: NE’s query speedup to Search

Number of Event Pairs	Speedup Ratio on ICEWS14					Speedup Ratio on ICEWS18				
	5	10	20	30	50	5	10	20	30	50
NE (CPU) / Search (CPU)	3.47	4.78	6.37	8.69	13.3	5.99	7.71	9.17	11.2	14.1
NE (GPU) / Search (CPU)	182	237	301	445	590	277	327	423	523	678

Decoding results In Figure 3, we illustrate $g(\tau)$ (defined by Equation 12) and $p(\tau)$ (defined in Section 2.1) calculated by NE, from two cases indicating queries with high and low g_c respectively. The results indicate that NE can enforce strict convergence and also tell apart TAs with high and low g_c (defined by Equation 4) via comparing maximums of probability curves.

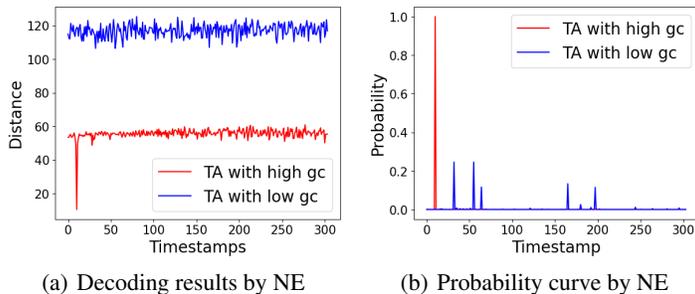


Figure 3: Decoding comparisons of TAs with high and low gc respectively

5.2 RESULTS ON TAM

Statistics Table 4 respectively presents the performance of NE for TAM on ICEWS14, ICEWS18, and GDELT datasets. It is shown that NE can mine an appreciable number of TAs with high confidence ($gc \geq 0.7$) and moderate reliability ($sp \geq 2$) on real-world data. Furthermore, NE achieves high speedup ratios of 11.0, 7.50, and 9.75 over the optimized search algorithm with the same number of TAs mined on ICEWS14, ICEWS18, and GDELT datasets. Note that we implement Search on a cpu; the encoding and decoding stage of NE is implemented on a gpu; the selecting stage is on a cpu, the same as Search. At the selecting stage of NE, we additionally add a time-costly filter of gc after the filter of p_{th} , because the convergence of NE is not guaranteed due to imbalanced reduction of training loss (averaged around $L = 40$) on large real-world datasets. This implies that the speedup has the potential to be further increased.

Table 2: NE’s speedup ratio over Search when reporting the same number of mined TAs.

	ICEWS14		ICEWS18		GDELT	
	Number	Speedup	Number	Speedup	Number	Speedup
NE / Search	479	11.0	1203	7.50	24316	9.75

Results in online settings In Table 3, we show the training time of NE and the number of mined TAs. The update period is 24h which is essentially larger than the training time of 2.2, 2.9, 5.4, and 6.8 seconds over one batch. Moreover, cases of mined TAs with $\tau > w_{batch}$ are shown in Appendix B.2.3. Therefore, we demonstrate that NE has significant potential for speed and the temporal scope of results in online TAM.

Table 3: TAM results on ICEWS14 under online settings

Temporal Width of One Batch w_{batch}	10	20	50	100
Number of TAs mined	153	171	205	278
Training Time of One Batch(s)	2.2	2.9	5.4	6.8

Case study The examples of mined TAs in ICEWS18 are demonstrated. (1) Whenever *Donald Trump* discusses by telephone with *China* (day 16 and 86 in 2018), then in around 24 days, he will consider policy options to *China* (day 43 and 109 in 2018). (2) *Syria* will provide military protection or peacekeeping to *Turkey* (day 26 and 33 in 2018) back in around two days whenever *Syria* fights with artillery and tanks against *Turkey* (day 23, 30 and 31 in 2018). It is shown that the mined TAs accord with commonsense and is reliable statistically. More examples (including generalized TAs mined by NE) are listed in Appendix C.

6 RELATED WORK

Embedding models Complex vectors have shown great potential in embedding models. RotatE (Sun et al., 2019) represents each entity and relationship as a complex vector, and defines the score function as $f(s, p, o) = \|s \circ p - o\|$. It is designed to model and infer various relation patterns, including symmetry, antisymmetry, inversion, composition on static knowledge graphs. TeRo (Xu et al., 2020a) represents time as rotation of entities in the complex vector space: $s_t = s \circ e^{i\omega t}$, $o_t = o \circ e^{i\omega t}$, and defines the score function as $f(s, p, o, t) = \|s_t + p - \bar{o}_t\|$. It can learn and infer various time relationship patterns on temporal knowledge graphs. However, both methods do not encode the time translation symmetry of TAs, and can not query TAs without traversing the absolute time set. In addition to directly for link prediction tasks, embedding models are also used to decode rules. IterE (Zhang et al., 2019), for instance, was proposed based on the observation that embeddings learned with linear map assumption can fully support rule learning in static knowledge graphs. It represents each entity as a real vector, each relationship as a real matrix, and the score function as $f(s, p, o) = \|sP - o\|$. It can learn rules like $p_1(s, o) \rightarrow p_2(o, s), \forall s, o$ in static knowledge graphs. However, this model can only decode commonsense rules in static knowledge graphs. In contrast, ours can decode unknown TAs, and can be extended for decoding generalized temporal associations of various structured events.

Temporal association mining Various methods in different fields mine temporal associations. Namaki et al. developed an effective search algorithm to discover graph temporal association rules in real-world event networks (Namaki et al., 2017). Chen et al. developed a hierarchical temporal association mining approach for video event detection in video databases (Chen et al., 2007). Yoo and Shekhar developed a similarity-profiled algorithm to mine temporal associations in a time-stamped transaction database (Yoo & Shekhar, 2008). West and Lee described and tested a model for temporal association learning over spam blacklist data (West & Lee, 2011). Guillame-Bert and Crowley presented a novel algorithm for extracting temporal interval tree association rules from large datasets expressed as symbolic time sequences (Guillame-Bert & Crowley, 2012). Li et al. developed an approach to discover calendar-based temporal association rules in transactional data (Li et al., 2003). However, none of these works supports an efficient query of an approximate probability distribution of the relative time points within the arbitrary queried range of a potential TA.

Noether’s theorem empowered AI Noether’s theorem has inspired numerous works in artificial intelligence. Tanaka and Kunin established a theoretical foundation to discover geometric design principles for the learning dynamics of neural networks by generalizing Noether’s theorem to Noether’s Learning Dynamics (Tanaka & Kunin, 2021). Alet et al. provided a general framework for discovering inductive biases in sequential problems by Noether Networks that reduces finding inductive biases to meta-learning useful conserved quantities (Alet et al., 2021). Mototake proposed a Noether’s theorem-inspired framework to achieve interpretable conservation law estimation by extracting the symmetries of dynamics from trained DNNs (Mototake, 2021). Santini and Sanz presented a theory of image matching based on elastic deformation, which allows a unified treatment of invariance using Noether’s theorem (Santini & Sanz). Despite these efforts, none of these works utilizes the axiomatized property of a physical system to empower AI.

7 CONCLUSIONS AND FUTURE WORK

We develop Noether Embeddings, an embedding model of structured events that naturally encodes the time translation symmetry of TAs, and can efficiently decode conserved quantities to query TAs. A three-stage TAM framework is developed with NE, which rearranges the major time complexity of TAM into the encoding stage, and has strong potential for online TAM. We also develop a heuristic embedding construction framework, which aids AI by Noether’s theorem by utilizing a physical system’s axiomatized property.

Despite the effectiveness of NE, it only supports decoding a 1-1 TA. Since multiple reasons may cause an event, embedding models are to be constructed that allow decoding of n-1 TAs for better exploration of event regularities. In addition, since our mining framework has great potential in stream processing scenarios, further improvements should be made to relieve NE’s training convergence and forgetting problem.

REFERENCES

- AY Abul-Magd and MH Simbel. Phenomenological model for symmetry breaking in a chaotic system. *Physical Review E*, 70(4):046218, 2004.
- Ferran Alet, Dylan Doblar, Allan Zhou, Josh Tenenbaum, Kenji Kawaguchi, and Chelsea Finn. Noether networks: meta-learning useful conserved quantities. *Advances in Neural Information Processing Systems*, 34:16384–16397, 2021.
- Elizabeth Boschee, Jennifer Lautenschlager, Sean O’Brien, Steve Shellman, James Starz, and Michael Ward. Icews coded event data. *Harvard Dataverse*, 12, 2015.
- Borui Cai, Yong Xiang, Longxiang Gao, He Zhang, Yunfeng Li, and Jianxin Li. Temporal knowledge graph completion: A survey. *arXiv preprint arXiv:2201.08236*, 2022.
- Min Chen, Shu-Ching Chen, and Mei-Ling Shyu. Hierarchical temporal association mining for video event detection in video databases. In *2007 IEEE 23rd International Conference on Data Engineering Workshop*, pp. 137–145. IEEE, 2007.
- Soumaya Cherichi and Rim Faiz Larodec. Using big data values to enhance social event detection pattern. In *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, pp. 1–8. IEEE, 2016.
- Tidewater Big Data Enthusiasts and Chuck Cartledge. Big data potential of the global database of events, language, and tone (gdelt).
- Alberto García-Durán, Sebastijan Dumančić, and Mathias Niepert. Learning sequence encoders for temporal knowledge graph completion. *arXiv preprint arXiv:1809.03202*, 2018.
- Rishab Goel, Seyed Mehran Kazemi, Marcus Brubaker, and Pascal Poupard. Diachronic embedding for temporal knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 3988–3995, 2020.
- Mathieu Guillaume-Bert and James L Crowley. Learning temporal association rules on symbolic time sequences. In *Asian conference on machine learning*, pp. 159–174. PMLR, 2012.
- Zhen Han, Peng Chen, Yunpu Ma, and Volker Tresp. xerte: Explainable reasoning on temporal knowledge graphs for forecasting future links. *arXiv preprint arXiv:2012.15537*, 2020.
- Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. Recurrent event network: Autoregressive structure inference over temporal knowledge graphs. *arXiv preprint arXiv:1904.05530*, 2019.
- Timothée Lacroix, Guillaume Obozinski, and Nicolas Usunier. Tensor decompositions for temporal knowledge base completion. *arXiv preprint arXiv:2004.04926*, 2020.
- Julien Leblay and Melisachew Wudage Chekol. Deriving validity time in knowledge graph. In *Companion Proceedings of the The Web Conference 2018*, pp. 1771–1776, 2018.
- Kalev Leetaru and Philip A Schrodt. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, volume 2, pp. 1–49. Citeseer, 2013.
- Lei Lei, Xuguang Ren, Nigel Franciscus, Junhu Wang, and Bela Stantic. Event prediction based on causality reasoning. In *Asian Conference on Intelligent Information and Database Systems*, pp. 165–176. Springer, 2019.
- Yingjiu Li, Peng Ning, X Sean Wang, and Sushil Jajodia. Discovering calendar-based temporal association rules. *Data & Knowledge Engineering*, 44(2):193–218, 2003.
- Yushan Liu, Yunpu Ma, Marcel Hildebrandt, Mitchell Joblin, and Volker Tresp. Tlogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 4120–4127, 2022.
- Johannes Messner, Ralph Abboud, and Ismail Ilkan Ceylan. Temporal knowledge graph completion using box embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 7779–7787, 2022.

- Mehrnoosh Mirtaheri, Mohammad Rostami, Xiang Ren, Fred Morstatter, and Aram Galstyan. One-shot learning for temporal knowledge graphs. *arXiv preprint arXiv:2010.12144*, 2020.
- Yoh-ichi Mototake. Interpretable conservation law estimation by deriving the symmetries of dynamics from trained deep neural networks. *Physical Review E*, 103(3):033303, 2021.
- Mohammad Hossein Namaki, Yinghui Wu, Qi Song, Peng Lin, and Tingjian Ge. Discovering graph temporal association rules. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1697–1706, 2017.
- E. Noether. Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918. URL <http://eudml.org/doc/59024>.
- Pouya Ghiasnezhad Omran, Kewen Wang, and Zhe Wang. Learning temporal rules from knowledge graph streams. In *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering*, 2019.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
- Kira Radinsky, Sagie Davidovich, and Shaul Markovitch. Learning causality for news events prediction. In *Proceedings of the 21st international conference on World Wide Web*, pp. 909–918, 2012.
- Naren Ramakrishnan, Patrick Butler, Sathappan Muthiah, Nathan Self, Rupinder Khandpur, Parang Saraf, Wei Wang, Jose Cadena, Anil Vullikanti, Gizem Korkmaz, et al. ‘beating the news’ with embers: forecasting civil unrest using open source indicators. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1799–1808, 2014.
- Simone Santini and Jorge Sanz. Multidimensional elastic pattern matching.
- Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*, 2019.
- Hidenori Tanaka and Daniel Kunin. Noether’s learning dynamics: Role of symmetry breaking in neural networks. *Advances in Neural Information Processing Systems*, 34:25646–25660, 2021.
- Rakshit Trivedi, Hanjun Dai, Yichen Wang, and Le Song. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *international conference on machine learning*, pp. 3462–3471. PMLR, 2017.
- Jingbin Wang, Wang Zhang, Xinyuan Chen, Jing Lei, and Xiaolian Lai. 3drte: 3d rotation embedding in temporal knowledge graph. *IEEE Access*, 8:207515–207523, 2020.
- Andrew G West and Insup Lee. Towards the effective temporal association mining of spam blacklists. In *Proceedings of the 8th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference*, pp. 73–82, 2011.
- Chris Wrench, Frederic Stahl, Giuseppe Di Fatta, Vidhyalakshmi Karthikeyan, and Detlef D Nauck. Data stream mining of event and complex event streams: A survey of existing and future technologies and applications in big data. In *Enterprise Big Data Engineering, Analytics, and Management*, pp. 24–47. IGI Global, 2016.
- Chunjing Xiao, Leilei Sun, and Wanlin Ji. Temporal knowledge graph incremental construction model for recommendation. In *Asia-pacific web (apweb) and web-age information management (waim) joint international conference on web and big data*, pp. 352–359. Springer, 2020.
- Chengjin Xu, Mojtaba Nayyeri, Fouad Alkhoury, Hamed Shariat Yazdi, and Jens Lehmann. Tero: A time-aware knowledge graph embedding via temporal rotation. *arXiv preprint arXiv:2010.01029*, 2020a.

Chenjin Xu, Mojtaba Nayyeri, Fouad Alkhoury, Hamed Yazdi, and Jens Lehmann. Temporal knowledge graph completion based on time series gaussian embedding. In *International Semantic Web Conference*, pp. 654–671. Springer, 2020b.

Yiyang Yang, Zhongyu Wei, Qin Chen, and Libo Wu. Using external knowledge for financial event prediction based on graph neural networks. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2161–2164, 2019.

Jin Soung Yoo and Shashi Shekhar. Similarity-profiled temporal association mining. *IEEE Transactions on Knowledge and Data Engineering*, 21(8):1147–1161, 2008.

Wen Zhang, Bibek Paudel, Liang Wang, Jiaoyan Chen, Hai Zhu, Wei Zhang, Abraham Bernstein, and Huajun Chen. Iteratively learning embeddings and rules for knowledge graph reasoning. In *The World Wide Web Conference*, pp. 2366–2377, 2019.

Liang Zhao. Event prediction in the big data era: A systematic survey. *ACM Computing Surveys (CSUR)*, 54(5):1–37, 2021.

A THEORETICAL ANALYSIS

A.1 PROOF

Theorem 1 Denote that $|\cos(\mathbf{a}_k, \mathbf{x})| = c_1 = \max_{i \in \mathbb{M}_1} |\cos(\mathbf{a}_i, \mathbf{x})|$, where $k \in \mathbb{M}_1$. Suppose that the k th pair are both positive samples, we have

$$\begin{aligned} f(t_k) &= |\sum_{i=1}^d \text{Re}(\mathbf{u}_c \circ \mathbf{r}(t_k))_i - C_p| < \eta \\ f(t_k + \tau) &= |\sum_{i=1}^d \text{Re}(\mathbf{u}_e \circ \mathbf{r}(t_k + \tau))_i - C_p| < \eta \end{aligned}$$

By the trigonometric inequality, we can derive

$$|\sum_{i=1}^d \text{Re}((\mathbf{u}_c - \mathbf{u}_e \circ \mathbf{r}(\tau)) \circ \mathbf{r}(t_k))_i| < 2\eta$$

Therefore, we get the result below, the same when the sample pair are both negative samples.

$$|\mathbf{a}_k^T \mathbf{x}| < 2\eta$$

Noticing that $|\mathbf{a}_k^T \mathbf{x}| = \|\mathbf{a}_k\| \|\mathbf{x}\| |\cos(\mathbf{a}_k, \mathbf{x})| = \|\mathbf{a}_k\| \|\mathbf{x}\| c_1$, and $\|\mathbf{a}_i\| = \sqrt{d}$, $c_1 > 0$ we can derive

$$g(\tau) = \|\mathbf{x}\| < \frac{2\eta}{\sqrt{d}c_1}$$

Theorem 2 Denote that $|\cos(\mathbf{a}_l, \mathbf{x})| = c_2 = \min_{i \in \mathbb{M}_2} |\cos(\mathbf{a}_i, \mathbf{x})|$, where $l \in \mathbb{M}_2$. Suppose that the cause event of the l th pair is a positive sample, and the effect event is a negative sample, we have

$$\begin{aligned} f(t_l) &= |\sum_{i=1}^d \text{Re}(\mathbf{u}_c \circ \mathbf{r}(t_l))_i - C_p| < \eta \\ f(t_l + \tau) &= |\sum_{i=1}^d \text{Re}(\mathbf{u}_e \circ \mathbf{r}(t_l + \tau))_i - C_n| < \eta \end{aligned}$$

By the trigonometric inequality, and noticing that $|C_p - C_n| > 2\eta$, we can derive

$$|\sum_{i=1}^d \text{Re}((\mathbf{u}_c - \mathbf{u}_e \circ \mathbf{r}(\tau)) \circ \mathbf{r}(t_l))_i| > |C_p - C_n| - 2\eta$$

Therefore, we get the result below, the same when the cause event is a negative sample, and the effect event is a positive sample.

$$|\mathbf{a}_l^T \mathbf{x}| > |C_p - C_n| - 2\eta$$

Noticing that $|\mathbf{a}_l^T \mathbf{x}| = \|\mathbf{a}_l\| \|\mathbf{x}\| |\cos(\mathbf{a}_l, \mathbf{x})| = \|\mathbf{a}_l\| \|\mathbf{x}\| c_2$, and $\|\mathbf{a}_l\| = \sqrt{d}$, $c_2 > 0$, we can derive

$$g(\tau) = \|\mathbf{x}\| > \frac{|C_p - C_n| - 2\eta}{\sqrt{d}c_2}$$

A.2 CONDITIONS

Denote $P(\mathbf{v})$ as the probability density function of the m values $|\cos(\mathbf{a}_i, \mathbf{v})|, i \in \mathbb{M}$ for an arbitrary vector \mathbf{v} of dimension $2d, d = 400$ as set for all experiments, $m \in M = 0, 1, \dots, 364$ and ω is also set as a fixed distribution used in all experiments.

KL Divergence Randomly (by angle) sample $h=1000$ \mathbf{v} s from the $2d$ space, we calculated an average KL divergence

$$D_{sample} = \frac{1}{h^2} \sum D_{KL}(P(\mathbf{v}_i)||P(\mathbf{v}_j)) = 0.088$$

We can see from this result that $P(\mathbf{v})$ is sparse towards a single distribution for all \mathbf{v} s in space.

Max, Mean, Min, Var(standard variance) Randomly (by angle) sample $h=1000000$ \mathbf{v} s from the $2d$ space, we plot the distribution of $c_{min} = \min_{i \in \mathbb{M}} |\cos(\mathbf{a}_i, \mathbf{v})|, c_{mean} = \text{mean}_{i \in \mathbb{M}} |\cos(\mathbf{a}_i, \mathbf{v})|, c_{max} = \max_{i \in \mathbb{M}} |\cos(\mathbf{a}_i, \mathbf{v})|, c_{var} = \text{var}_{i \in \mathbb{M}} |\cos(\mathbf{a}_i, \mathbf{v})|$ as below.

We can see that $c_{min}, c_{mean}, c_{var}$ are respectively sharp towards around 0, 0.03, 0, while c_{max} is distributed in $[0.05, 0.2]$. These results also imply that $P(\mathbf{v})$ is sparse towards a single distribution for all \mathbf{v} s in space.

Reasons for $c_1 > 0, c_2 > 0$ Here we demonstrate a randomly selected sample of the cumulative distribution function $F(\mathbf{v})$ for the m values $|\cos(\mathbf{a}_i, \mathbf{v})|, i \in \mathbb{M}$, which is similar to $F(\mathbf{x})$ for reasons above.

Since $\frac{m_1}{m} > 0.1$ in most situations, we can see that $c_1 = \max_{i \in \mathbb{M}_1} |\cos(\mathbf{a}_i, \mathbf{x})| > 0$ is therefore satisfied.

Meanwhile, since $\|\mathbf{a}_l\| \|\mathbf{x}\| c_2 > |C_p - C_n| - 2\eta > 0$ is satisfied during the training process (l as denoted in the proof of theorem 2), and $\|\mathbf{a}_l\| \|\mathbf{x}\| < +\infty, c_2 > 0$ is therefore strictly ensured.

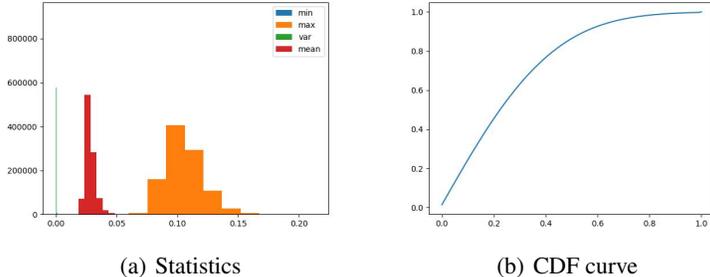


Figure 4: Statistics and CDF via Random sampling

B EXPERIMENTAL DETAILS

B.1 DATASETS

ICEWS (Boschae et al., 2015) and GDELT (Leetaru & Schrodt, 2013) are two popular data sources in the TKG research community, which provide datasets containing political events with timestamps. Our experiments use ICEWS from the following two subsets: ICEWS14 and ICEWS18, which contain events in 2014 and 2018, respectively, the same as in (Han et al., 2020). We use the version of GDELT released by Jin et al. (2019), which contains events from 2018/1/1 to 2018/1/31.

B.2 TASK SETTINGS

B.2.1 TAQ

Evaluation Following the formulation stated in Section 2, we set $\eta = 0.1$ when performing evaluations. For convenience, we force the lengths of $\Delta(\eta)$ to be integers by setting its endpoints as $\tau - \lceil \tau \times \eta \rceil$ and $\tau + \lceil \tau \times \eta \rceil$ (τ is the center point of $\Delta(\eta)$).

We calculate the overall accuracy of a dataset by averaging the accuracy of all queries. We also record the query time to examine the efficiency of methods. For Search, we measure the whole running time since it requires no training procedure. For Noether Embeddings, we record the time of processing all queries regardless of the training time beforehand. For a fair comparison, we report the speedup ratio of NE running on a GPU and a CPU to Search running on the same CPU, respectively.

Data Generation We generate training data and queries from real-world datasets mixed with synthesized ground truth data. For a selected dataset, all data from its training set is included in our training set. Then we generate 1000 queries by randomly choosing (s, o, p_c, p_e) for each query. To avoid interference between queries, we ensure that each query points to event types that have never occurred in the real-world dataset. Queries do not overlap on event types to avoid false generations of relevant events of TAs queried. We set the proportion of positive queries as $r = 0.5$, the same as that of the random guess baseline. For queries with positive results, we randomly choose a t out of all time intervals as the golden answer for the query. Afterwards, we produce n event pairs of n samples as type of (s, p_c, o) and another n samples of (s, p_e, o) for each queries (during evaluation $n = 5, 10, 20, 30, 50$). For queries (s, o, p_c, p_e) with positive results, we guarantee that for each sample of (s, p_c, o, t) , there exists a sample $(s, p_e, o, t + \tau)$. For queries with negative results, we implement samples with no temporal associations through random sampling.

We generate datasets based on ICEWS14, ICEWS18, and GDELTA, respectively.

B.2.2 TAM

Evaluation We use the formulation the same as in Section 2.2, in which $\eta = 0.1$. We conduct experiments over ICEWS14, ICEWS18, and GDELTA, respectively. We set $gc_l = 0.7$, $sp_l = 2$ here. Only TAs with $gc \geq gc_l$ and $sp \geq sp_l$ are considered reliable here. We compare the running time of NE and the search algorithm when they mine the same number of TAs ($gc \geq gc_l$, $sp \geq sp_l$).

B.2.3 ONLINE SETTINGS

The difference between online TAM and normal TAM is that the training data is provided in streaming form in online settings. Models only access a batch of data whose timestamps belong to $[t, t + w_{batch}]$, and t increases by $l_{interval}$ with every update. It is assumed that the interval time between two adjacent updates equals $l_{interval}$ in real-world applications.

Evaluation Methods like NE that require training are trained on data whose timestamps belong to $[t, t + w_{batch}]$, with t increasing by $l_{interval}$ with every update. Mining is performed after the training process of NE, while directly for Search. Algorithms can only access the final batch of data when conducting online TAM in realistic situations. Since TAM with NE still requires an additional filter by gc for TAM on real-world datasets, we allow the filter by gc to access data batches beforehand for NE. Note that this relaxation does not change the conclusions demonstrated in the main text, since the filter of gc can be removed if the imbalanced fitting problem is solved in the future. Other settings are the same as in normal TAM.

B.3 METHOD IMPLEMENTATION

Search We construct an optimized search algorithm (Algorithm 1) as a baseline. To handle a query, the algorithm first establishes the dictionary that enables fast indexing event occurrences from event type (s, p, o) . Then our algorithm answers TA queries via traversing all possible center points of $\Delta(\eta)$ and finding the best gc ever calculated. We calculate gc as fastest as $O(N)$ (N is the

number of all events related to the query). Moreover, we optimize our algorithm to return results early if a golden center point is found with $gc = 1$.

Algorithm 1: The Optimized Search Algorithm

Input: $queries = [q^{(1)}, q^{(2)}, \dots]$ as all queries, $q^{(i)} = (s^{(i)}, o^{(i)}, p_c^{(i)}, p_e^{(i)}, type^{(i)})$

Output: $[res^{(1)}, res^{(2)}, \dots]$ as results for all queries sequentially.

$res^{(i)} \in \{-1\} \cup \{\text{all possible time intervals}\}$. $res^{(i)} = -1$ means no TA is found, else means the center of found $\Delta(\eta)$

Data: A directory $triple_occur$ that maps every triple of (s, p, o) to its event occurrence $[(s, p, o, t_1), (s, p, o, t_2), \dots]$

Build $triple_occur$ while traversing through all events.

for $q^{(i)}$ **in** $queries$ **do**

$gc_{max}, res = 0, -1$

for res' **in** $\{\text{all possible time intervals}\}$ **do**

$occurrence_c = triple_occur[(s^{(i)}, p_c^{(i)}, o^{(i)})]$

if $type = \text{"forward"}$ **then**

$occurrence_e = triple_occur[(s^{(i)}, p_e^{(i)}, o^{(i)})]$

else

$occurrence_e = triple_occur[(o^{(i)}, p_e^{(i)}, s^{(i)})]$

 Sort $occurrence_c$ and $occurrence_e$ by event time

$support, head, tail = 0, len(occurrence_c), len(occurrence_e)$

while $occurrence_c$ is not empty and $occurrence_e$ is not empty **do**

$(s_c, p_c, o_c, t_c) = occurrence_c[0]$

$(s_e, p_e, o_e, t_e) = occurrence_e[0]$

if $t_e - t_c \in [res' - \eta \times res', res' + \eta \times res']$ **then**

 Pop (s_c, p_c, o_c, t_c) from $occurrence_c$

 Pop (s_e, p_e, o_e, t_e) from $occurrence_e$

$support += 1$

else if $t_e - t_c < res' - \eta \times res'$ **then**

 Pop (s_e, p_e, o_e, t_e) from $occurrence_e$

else if $t_e - t_c > res' + \eta \times res'$ **then**

 Pop (s_c, p_c, o_c, t_c) from $occurrence_c$

 Calculate gc using $support, head$ and $tail$.

If $gc > gc_{max}$ $gc = gc_{max}$

$res = res'$

if $gc = 1$ **then**

 Break;

 // Find best answer, break directly.

 Store res as the i -th query's answer

Return $[res^{(1)}, res^{(2)}, \dots]$

Embedding Methods In contrast to NE that encodes every timestamp with one complex vector as a basis for complex Fourier expansions and gives $g(\tau)$ with all $\tau \in \mathbb{T}_r$ for each query (s, o, p_c, p_e) , baseline embedding models (TNTComplex, TeRo, DE-Simple, and ATiSE) embed timestamps with independent parameters and provide $score(s, p, o; t)$ (stands for the possibility that an event occurs at time t) for each event type of (s, p, o) . In order to answer the query (s, o, p_c, p_e) , we thus develop a decoding function $g'(\tau)$ and converting function $p'(\tau)$ in functionality similar to $g(\tau)$ and $p(\tau)$ of NE, for these embedding models to calculate cross-correlation coefficients of $score(s, p_c, o; t)$ and $score(s, p_e, o; t)$ with $\tau \in \mathbb{T}_r$. Note that since $g'(\tau)$ here directly represents correlations, unlike in NE where $g(\tau)$ represents distances, we do not take the opposite number of $g'(\tau)$ in the equation of $p'(\tau)$.

$$g'(\tau) = \sum_{t \in \mathbb{T}_a} score(s, p_c, o; t) \times score(s, p_e, o; t + \tau) \quad (14)$$

$$p'(s, o, p_c, p_e; \tau) = \frac{\exp g'(s, o, p_c, p_e; \tau)}{\sum_{\tau \in \mathbb{T}_r} \exp g'(s, o, p_c, p_e; \tau)} \quad (15)$$

Extra processing on TAM When performing mining on TAM, we first filter all candidate TAs (s, o, p_c, p_e) . Then we consider the justification of each candidate TA as a query. Thus, the following processing is the same as in TAQ. Note that the same process is implemented for both Search and NE for a fair comparison.

B.4 HYPERPARAMETER SELECTION

All the embedding methods are trained for 100 epochs on TAQ and 200 epochs on TAM with the learning rate of 0.1, optimized with Adagrad (Paszke et al., 2019). We set the embedding dimension of all embedding methods as $d = 400$. For the baseline methods, we adopt the settings of other hyperparameters reported in the corresponding papers. The p_{th} on TAQ for DE-Simple, TeRo, ATiSE, and TNTComplex are 0.004, 0.0035, 0.004, and 0.3 tuned to separate the maximum and minimum point, which turns out that $p'(\tau)$ produced by these models can hardly distinguish the golden τ via comparing. For NE, we tune hyperparameters over grid search for $\omega_{max} \in \{200, 300, 400, 500, 600\}$, $\alpha_p \in \{0.5, 1, 1.5\}$ and $p_{th} \in \{0.5, 0.7, 0.9\}$ on TAQ and $p_{th} \in \{0.05, 0.1, 0.15\}$ on TAM. ω_{max} is a hyperparameter used to control the range of time embeddings where $\omega_i = (2\pi \times \omega_{max})^{\frac{i}{d}} - 1$, $0 \leq i < d$. The performance hardly changes if we fix ω_i from the beginning, so it is not trained during training. α_p is related to C_p : $C_p = \alpha_p \times \sqrt{d}$ because C_p is of the same magnitude as the length of a d -dimensional vector. C_n is set as 0 to distinguish negative samples from positive ones.

B.5 EXPERIMENTAL ENVIRONMENTS

Experiments are conducted on a single GPU (GeForce RTX 3090) and a single CPU (Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz).

C GENERALIZED TAs

Table 4: Cases of generalized TAs mined by NE

Cause Event			Effect Event			Relative Time
s_c	p_c	o_c	s_e	p_e	o_e	τ
Barack Obama	Engage in diplomatic cooperation	Iran	Iran	Make optimistic comment	Iraq	4
United Arab Emirates	Host a visit	John Kerry	John Kerry	Make pessimistic comment	Military (Russia)	38
Pokot	Use conventional military force	Citizen (Kenya)	Citizen (Kenya)	Criticize or denounce	Hassan Joho	45
Combatant (Mali)	Use unconventional violence	Military (Mali)	Military (Mali)	Use conventional military force	Armed Rebel (Mali)	1

D ABLATION STUDY

Here we present the analysis of hyperparameters in TAQ: ω_{max} , α_p and p_{th} in Figure 5. The ablation study is conducted on ICEWS14 with each query relevant to 50 event pairs.

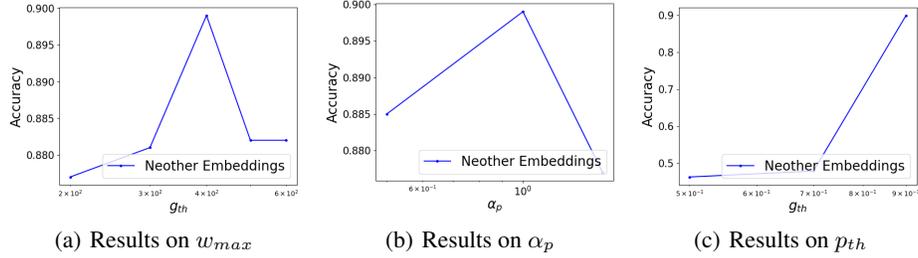


Figure 5: The accuracy with ω_{max} , α_p and p_{th} variant while other hyperparameters fixed.