Noether Embedding: Efficient Learning of Temporal Regularities

Chi Gao†, Zidong Zhou† and Luping Shi*

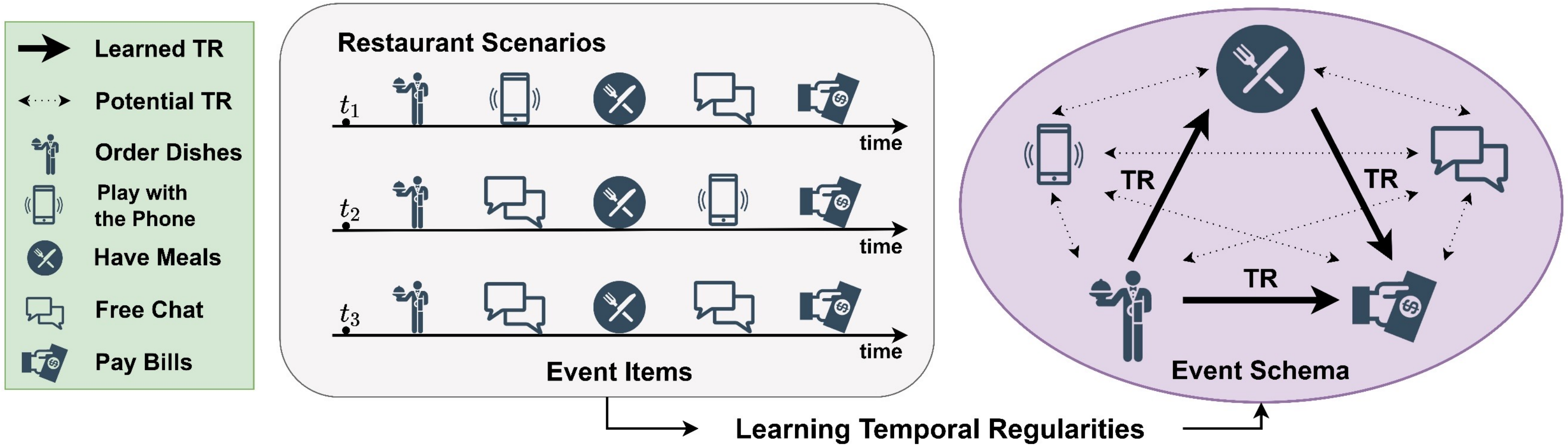
Center for Brain-Inspired Computing Research,
Department of Precision Instruments,
Tsinghua University.

Outline

- Background & Motivation
- Problem Formulation
- Inspiration from Noether's Theorem
- Noether Embedding
- NE's Effectiveness, Efficiency, and Potential Applications
- Comparison to Related Work

Background

Event schema is a **relational memory structure** in the brain that **supports generalization**.[1]



Temporal regularity (TR) is a **unit** structure of event schemas,[2]
which we formally define as **temporal associations that remain invariant to time shifts**.

$$(ev_b, t) \rightarrow (ev_h, t+\tau) \quad \forall t \in \mathbb{T}_a$$

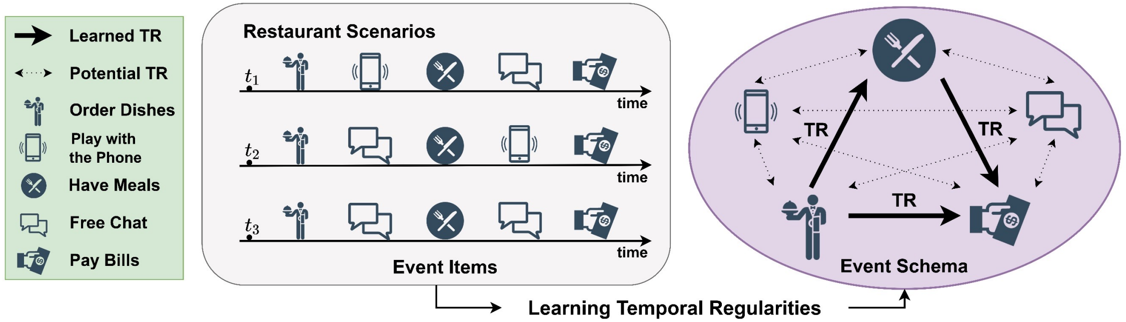
[1] Ghosh, V. E. and Gilboa, A. What is a memory schema? a historical perspective on current neuroscience literature. Neuropsychologia, 53:104–114, 2014.
[2] McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological review, 102(3):419, 1995.

Motivation

Humans

Data-efficient[1] TR formation

Time-efficient[2] TR retrieval



Existing Event Embeddings

Event completion[3]

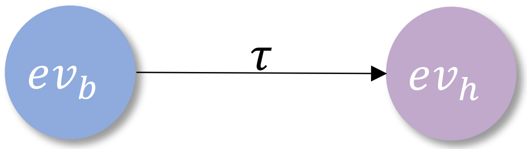
Event prediction[4]

[1] Pudhiyidath, A., Roome, H. E., Coughlin, C., Nguyen, K. V., and Preston, A. R. Developmental differences in temporal schema acquisition impact reasoning decisions. ognitive neuropsychology,37(1-2):25–45, 2020.
[2] Chaudhuri, R. and Fiete, I. Computational principles of memory. Nature neuroscience, 19(3): 394–403, 2016.
[3] Cai, B., Xiang, Y., Gao, L., Zhang, H., Li, Y., and Li, J. Temporal knowledge graph completion: A survey. arXiv preprint arXiv:2201.08236, 2022.
[4] Zhao, L. Event prediction in the big data era: A systematic survey. ACM Computing Surveys (CSUR), 54(5):1–37, 2021.

Problem Formulation

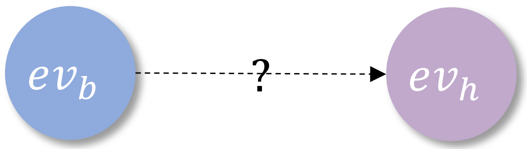
Aim: Learn TRs from event items $\{(ev, t)\}$

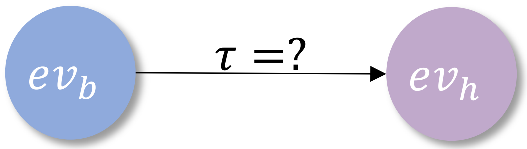
$$(ev_b, t) \rightarrow (ev_h, t + \tau) \quad \forall t \in \mathbb{T}_a$$



Tasks: TR Detection & TR Query

- **Detection**: Does (ev_h, ev_b) include a TR?
- **Query**: What is the τ for a TR (ev_h, ev_b)?





NEURAL INFORMATION PROCESSING SYSTEMS

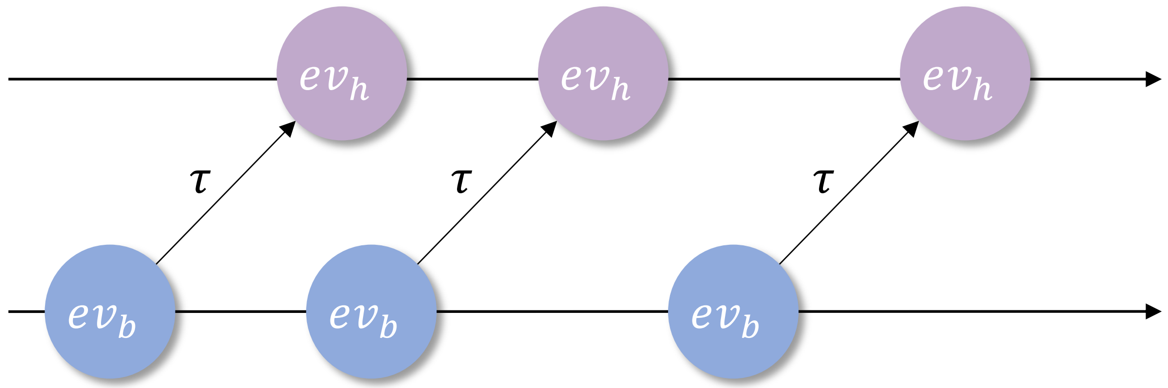
Inspiration from Noether’s Theorem

Noether’s Theorem

Every differentiable symmetry of the action of a physical system with conservative forces has a corresponding conservation law.
Specifically, time-translation symmetry corresponds to energy conservation.

Inspiration

TRs inherently possess
time-translation symmetries



Event embedding $\boldsymbol{q}(t; ev)$

Local energy $g\big(\boldsymbol{q}(t, ev_b), \boldsymbol{q}(t+\tau, ev_h)\big)$

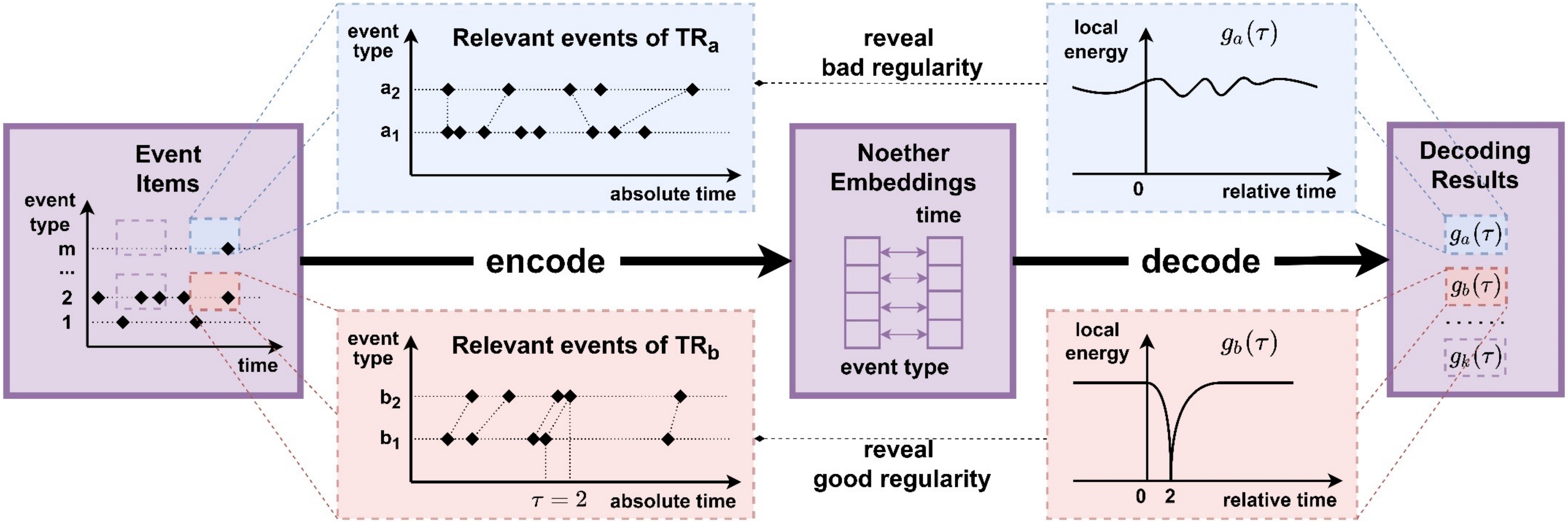
TR validity **Conserved**

Noether Embedding (NE)

Event embedding $\boldsymbol{q}(t;\ ev) = \boldsymbol{u}(ev) \circ \boldsymbol{r}(t)$

Score function $f(t;\ ev) = \sum_{i=1}^{d} \mathrm{Real}\big(\boldsymbol{q}(t;\ ev)\big)_i$

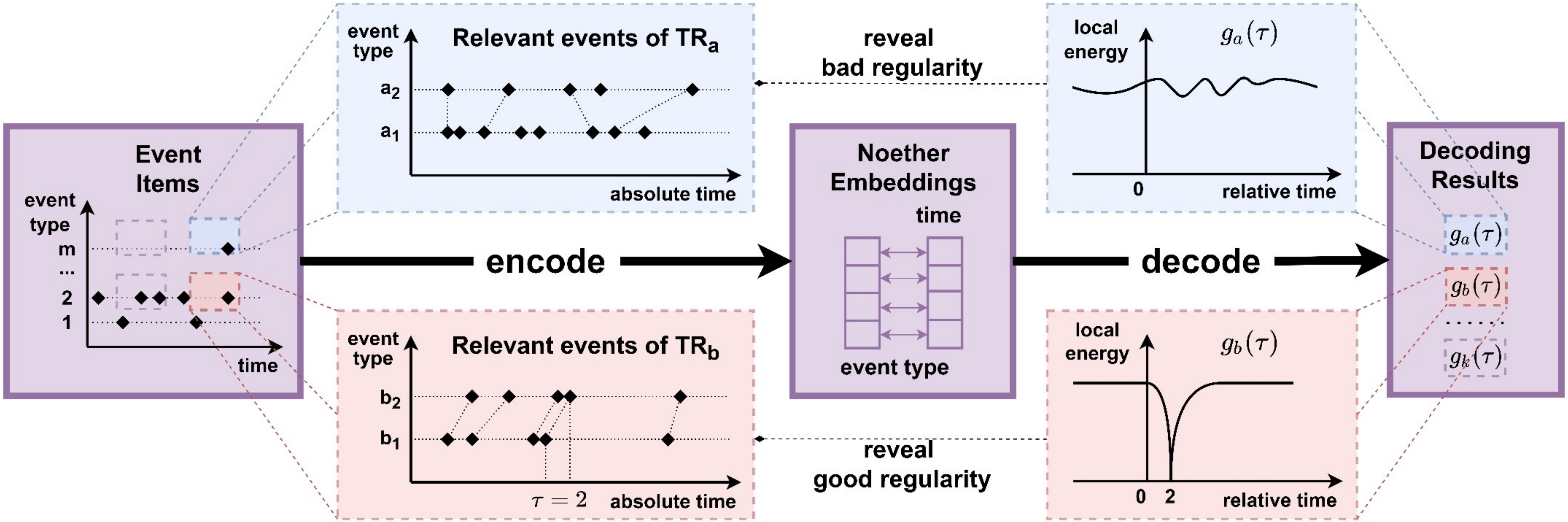
Loss function $L(\xi;\ C_p, C_n) = \left(\frac{1}{\sqrt{d}} f(\xi) - C_p\right)^2 + \frac{1}{N}\sum\left(\frac{1}{\sqrt{d}} f(\xi') - C_n\right)^2$



Noether Embedding (NE)

Decoding function

$$g(\tau) = |\boldsymbol{u}_b - \boldsymbol{u}_h \circ \boldsymbol{r}(\tau)|^2 \qquad \left(= |\boldsymbol{u}_b \circ \boldsymbol{r}(t) - \boldsymbol{u}_h \circ \boldsymbol{r}(t+\tau)|^2 = |\boldsymbol{q}_b(t) - \boldsymbol{q}_h(t+\tau)|^2 \right)$$



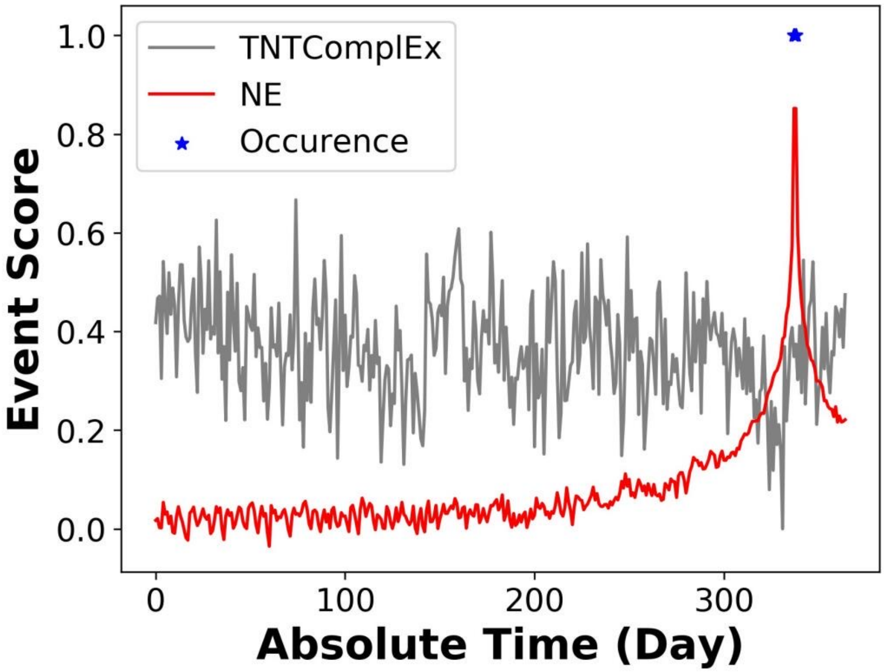
Effectiveness of NE

Table 1: Statistical results on ICEWS14, ICEWS18, and GDELT

Embedding	TR Detection (F1)			TR Query (r)		
	D14	D18	GDELT	D14	D18	GDELT
TNTComplEx	0.26	0.18	0.08	0.08	0.08	0.01
DE-SimplE	0.22	0.20	-	0.09	0.09	-
TASTER	0.18	0.15	0.08	0.09	0.09	0.00
TeRo	0.43	0.64	0.16	0.08	0.08	0.01
BoxTE	0.40	0.40	0.18	0.08	0.08	0.01
ATISE	0.40	0.44	0.18	0.08	0.08	0.01
NE with $g'(\tau)$	0.78	0.79	0.48	0.85	0.83	0.83
NE with $g(\tau)$	**0.82**	**0.83**	**0.51**	**0.87**	**0.86**	**0.85**

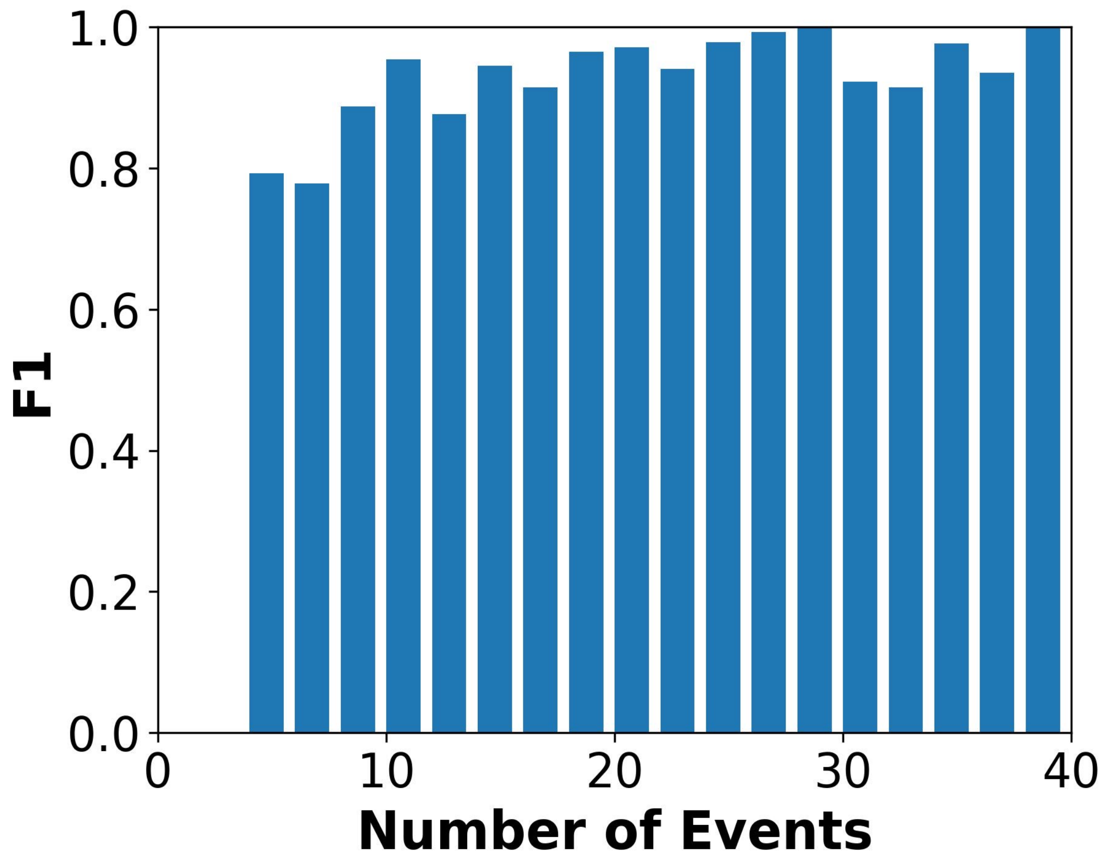
$$g'(\tau) = \frac{\sum_{t,t+\tau\in\mathbb{T}_a} f'_b(t) f'_h(t+\tau)}{\sum_{t\in\mathbb{T}_a} f'_b(t)\cdot\sum_{t\in\mathbb{T}_a} f'_h(t)}$$

Baselines **over-apply** the generalization capabilities of distributed representations, which **hinders the fit of event occurrences**.



Efficiency of NE

Data-efficient TR formation



Time-efficient TR retrieval

$$g(\tau) = \left\| \boldsymbol{u}_b - \boldsymbol{u}_h \circ \boldsymbol{r}(\tau) \right\|^2$$

$$g'(\tau) = \frac{\sum_{t,t+\tau \in \mathbb{T}_a} f'_b(t) f'_h(t+\tau)}{\sum_{t \in \mathbb{T}_a} f'_b(t) \cdot \sum_{t \in \mathbb{T}_a} f'_h(t)}$$

Storage-efficient TR memory

$$\boldsymbol{S}(\boldsymbol{NE}) = S(ev - vector) + S(time - vector)$$

$$= (2Nd + 2d) \times 64bit$$

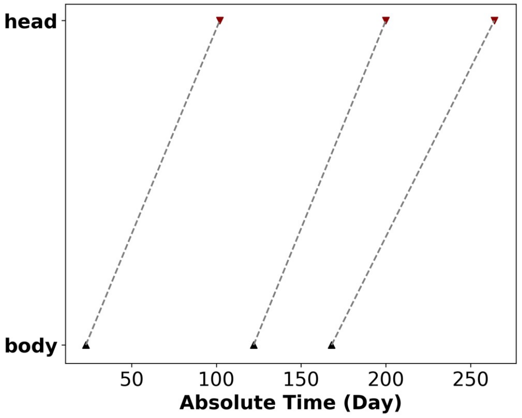
$$\boldsymbol{S}(\boldsymbol{CT}) = S(TR) + S(event)$$

$$= N^2 \times T_a \times \log_2 \frac{n}{N} bit + N \times \frac{n}{N} \times \log_2 T_a bit$$

NEURAL INFORMATION PROCESSING SYSTEMS

NE's Potential Applications

Social event prediction

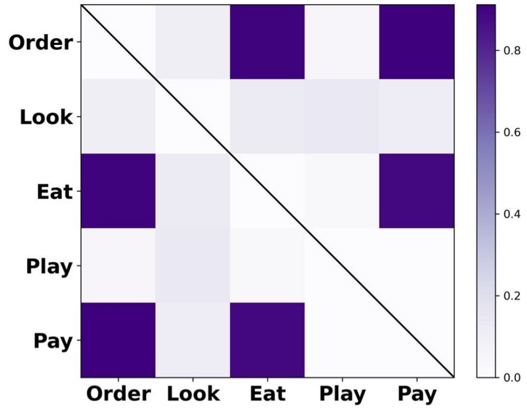


Narendra Modi
Appeal for diplomatic cooperation to
Citizen (India)

~ 87 days →

Citizen (India)
Reject to
Narendra Modi

Personal decision making



Order dishes

~ 10 mins →

have meals

Memory-constrained scenarios, such as the edge

NEURAL INFORMATION PROCESSING SYSTEMS

Comparison to Related Work

- **Event schema induction (in natural language processing)**:

  organize known event regularities **already given as priors** for the extracting algorithm and focus on the schemas for use

  → Our tasks are designed to learn event regularities **directly from experience without supervision.**

- **Temporal rule mining (in data mining)**

  the mined rules and source events are generally stored as symbolic representations in list form

  → Using event embeddings, NE strikes a balance between **storage efficiency and storage accuracy.**

- **Embedding models of structured data (in representation learning)**

  embed (temporal) knowledge graphs, generally supporting completion, inference, and prediction tasks

  → NE for the first time enables event embeddings to achieve
  data-efficient TR formation, storage-efficient TR memory, and time-efficient TR retrieval.

Thank you