

SPEECH: Structured Prediction with Energy-Based Event-Centric Hyperspheres

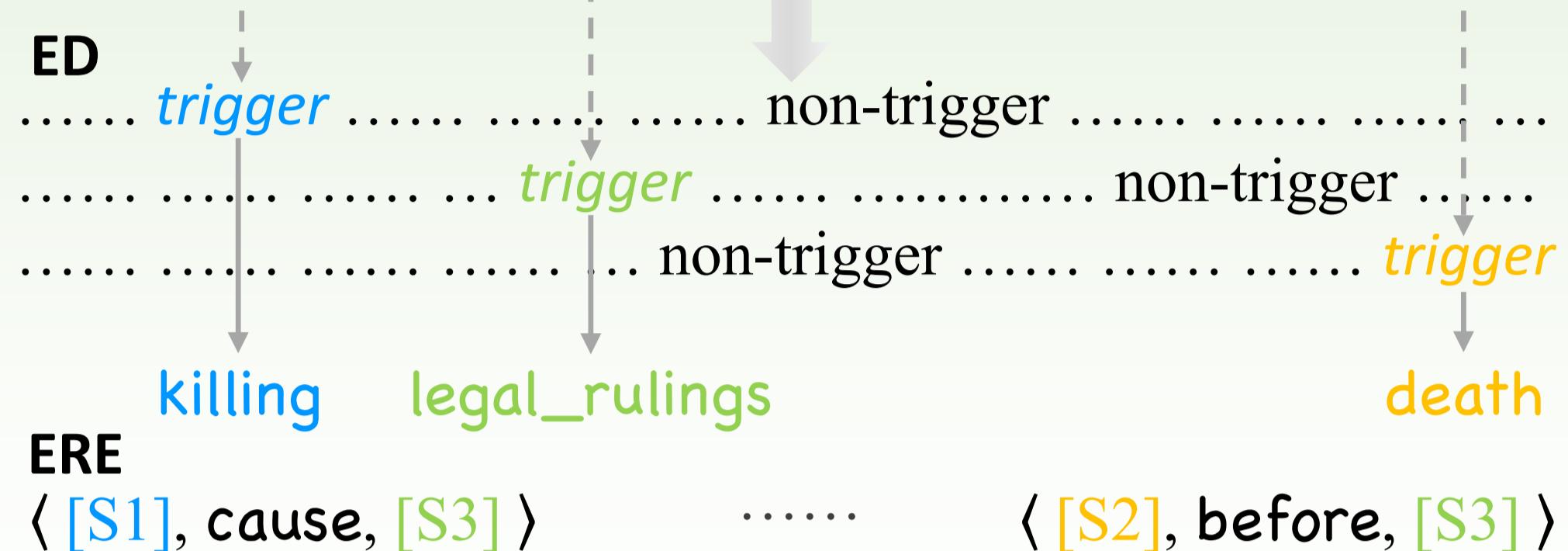
Shumin Deng¹, Shengyu Mao², Ningyu Zhang², Bryan Hooi¹

¹National University of Singapore, ²Zhejiang University



Introduction

..... [S1] Former NOPD police officer David Warren shot and killed Henry Glover. [S2] Five current and former officers of the NOPD were charged with Glover's death. [S3] David was convicted and sentenced to 25 years and 9 months in prison for shooting and killing Glover.



How to model manifold event structures?

How to efficiently represent these events?

Preliminary

Structured Prediction Energy Networks (SPENs)

- Input: $x \in \mathcal{X}$; Model: M_Φ ; Structured Outputs: $M_\Phi(x) \in \tilde{\mathcal{Y}}$
- score structured outputs with an **energy function** $E_\Theta : \mathcal{X} \times \tilde{\mathcal{Y}} \rightarrow \mathbb{R}$
- parameterized by Θ that iteratively optimize the energy between the input/output pair

Event-Centric Structured Prediction (ECSP) with Energy Networks

- Given a feature vector x belonging to one of T labels
- The model output: $M_\Phi(x) = \{0, 1\}^T \in \tilde{\mathcal{Y}}$
- Energy Function**

$$E_\Theta(x, y) = E_\Theta^{\text{local}}(x, y) + E_\Theta^{\text{label}}(y)$$

$$= \sum_{i=1}^T y_i V_i^\top f(x) + w^\top g(Wy)$$

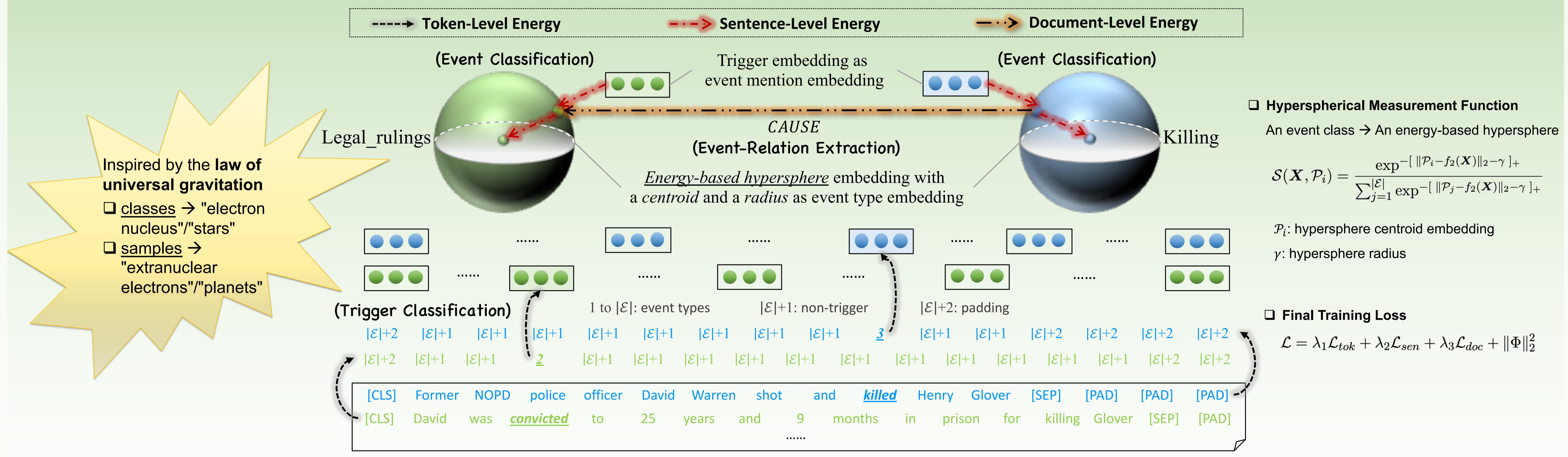
After learning the energy function, prediction minimizes energy $\tilde{y} = \arg \min_{y \in \tilde{\mathcal{Y}}} E_\Theta(x, y)$

final theoretical optimum for SPENs:

$$\min_{\Theta} \max_{\Phi} \sum [\Delta(M_\Phi(x_i), y_i) - E_\Theta(x_i, M_\Phi(x_i)) + E_\Theta(x_i, y_i)]_+$$

$$[a]_+ = \max(0, a) \quad \Delta(\tilde{y}, y): \text{"margin-rescaled" structured hinge loss}$$

Model



Trigger Classification

Token-level Energy Function

$$E_\Theta(x, y) = - \left(\sum_{n=1}^L \sum_{i=1}^{|\mathcal{E}|+2} y_n^i \left(V_{1,i}^\top f_1(x_n) \right) + \sum_{n=1}^L y_{n-1}^\top W_1 y_n \right)$$

Token-level Loss Function

$$\mathcal{L}_{\text{tok}} = \sum_{i=1}^L [\Delta(\tilde{y}_i, y_i) - E_\Theta(x_i, \tilde{y}_i) + E_\Theta(x_i, y_i)]_+ + \mu_1 \mathcal{L}_{\text{CE}}(\tilde{y}_i, y_i)$$

Sentence-level Energy Function

$$E_\Theta(\mathbf{X}, \mathbf{Y}) = - \left(\sum_{i=1}^{|\mathcal{E}|} \mathbf{Y}_i \left(V_{2,i}^\top f_2(\mathbf{X}) \right) + w_2^\top g(W_2 \mathbf{Y}) \right)$$

Sentence-level Loss Function

$$\mathcal{L}_{\text{sen}} = \sum_{i=1}^K [\Delta(\tilde{Y}_i, Y_i) - E_\Theta(\mathbf{X}_i, \tilde{Y}_i) + E_\Theta(\mathbf{X}_i, Y_i)]_+ + \mu_2 \mathcal{L}_{\text{CE}}(\tilde{Y}_i, Y_i)$$

Event Classification

Document-level Energy Function

$$E_\Theta(\tilde{\mathbf{X}}, z) = - \left(\sum_{i=1}^{|\mathcal{R}|} z_i \left(V_{3,i}^\top f_3(\tilde{\mathbf{X}}) \right) + w_3^\top g(W_3 z) \right)$$

Document-level Loss Function

$$\mathcal{L}_{\text{doc}} = \sum_{k=1}^N [\Delta(\tilde{z}_k, z_k) - E_\Theta(\tilde{\mathbf{X}}_k, \tilde{z}_k) + E_\Theta(\tilde{\mathbf{X}}_k, z_k)]_+ + \mu_3 \mathcal{L}_{\text{CE}}(\tilde{z}_k, z_k)$$

Experiments

Datasets

	MAVEN-ERE	ONTOEVENT-DOC
# Document	4,480	4,115
# Mention	112,276	60,546
# Temporal	1,216,217	5,914
# Causal	57,992	14,155
# Subevent	15,841	/
(Use Valid Set)		
(Use Test Set)		

Event-Relation Extraction

ERE Task	RoBERTa	SPEECH
Temporal	MAVEN-ERE +joint ONTOEVENT-DOC	49.21 ± 0.33 49.91 ± 0.58 37.68 ± 0.47 52.36 ± 0.71 35.63 ± 0.70 65.69 ± 0.39
	MAVEN-ERE +joint ONTOEVENT-DOC	29.91 ± 0.34 29.03 ± 0.91 35.48 ± 1.77 79.29 ± 2.15 44.99 ± 0.29 67.76 ± 1.28
Causal	MAVEN-ERE +joint ONTOEVENT-DOC	19.80 ± 0.44 19.14 ± 2.81 19.91 ± 0.52 21.96 ± 1.24
	All Joint	MAVEN-ERE ONTOEVENT-DOC

Table 4: F1 (%) performance of ERE on MAVEN-ERE *valid set* and ONTOEVENT-DOC *test set*. “+joint” in the 2nd column denotes jointly training on all ERE tasks and evaluating on the specific one, with the same setting as Wang et al. (2022). “All Joint” in the last two rows denotes treating all ERE tasks as one task.

Energy-based modeling
enables capture long-range dependency of tokens
places no limits on the size of event structures

SPEECH excels in modeling classes with fewer samples
SPEECH is advantageous in the scenario with more classes

Energy functions enable the model to directly capture complicated dependency between event mentions and event types, instead of implicitly inferring from data

Trigger Classification

Model	MAVEN-ERE			ONTOEVENT-DOC		
	P	R	F1	P	R	F1
DMCNN [†]	60.09 ± 0.36	60.34 ± 0.45	60.21 ± 0.21	50.42 ± 0.99	52.24 ± 0.46	51.31 ± 0.39
BiLSTM-CRF [†]	61.30 ± 1.07	64.95 ± 1.03	63.06 ± 0.23	48.86 ± 0.81	55.91 ± 0.56	52.10 ± 0.43
DMBERT [†]	56.79 ± 0.54	76.24 ± 0.26	65.09 ± 0.32	53.82 ± 1.01	66.12 ± 1.02	59.32 ± 0.24
BERT-CRF [†]	62.79 ± 0.34	70.51 ± 0.94	65.73 ± 0.57	52.18 ± 0.81	62.31 ± 0.45	56.80 ± 0.53
MLBInet [†]	63.50 ± 0.57	63.80 ± 0.47	63.60 ± 0.52	56.09 ± 0.93	57.67 ± 0.81	56.87 ± 0.87
TANL [‡]	68.66 ± 0.18	63.79 ± 0.19	66.13 ± 0.15	57.73 ± 0.65	59.93 ± 0.31	59.13 ± 0.52
TEXT2EVENT [‡]	59.91 ± 0.83	64.62 ± 0.65	62.16 ± 0.25	52.93 ± 0.94	62.27 ± 0.49	57.22 ± 0.75
CorED-BERT [‡]	67.62 ± 1.03	69.49 ± 0.63	68.49 ± 0.42	60.27 ± 0.55	62.25 ± 0.66	61.25 ± 0.19
SPEECH w/o energy	78.82 ± 0.82	79.37 ± 0.75	79.09 ± 0.82	74.67 ± 0.58	74.73 ± 0.62	74.70 ± 0.58
	76.12 ± 0.32	76.66 ± 0.25	76.38 ± 0.28	71.76 ± 0.38	72.17 ± 0.39	71.96 ± 0.38

Table 2: Performance (%) of trigger classification on MAVEN-ERE *valid set* and ONTOEVENT-DOC *test set*. [†]: results are produced with codes referred to Wang et al. (2020b); [‡]: results are produced with official implementation. Best results are marked in bold, and the second best results are underlined.

Event Classification

Model	MAVEN-ERE			ONTOEVENT-DOC		
	P	R	F1	P	R	F1
DMCNN	61.74 ± 0.32	63.11 ± 0.34	62.42 ± 0.15	51.52 ± 0.87	52.84 ± 0.61	52.02 ± 0.36
DMBERT	59.45 ± 0.48	77.77 ± 0.21	67.39 ± 0.25	57.06 ± 1.04	72.97 ± 1.11	65.03 ± 0.45
HPN	62.80 ± 0.72	62.62 ± 0.99	62.71 ± 0.85	61.18 ± 0.81	60.88 ± 0.79	61.03 ± 0.81
OntoED	67.82 ± 1.70	67.72 ± 1.52	67.77 ± 1.61	64.32 ± 1.15	64.16 ± 1.31	64.25 ± 1.22
TANL	68.73 ± 0.16	65.65 ± 0.63	67.15 ± 0.29	60.34 ± 0.71	62.52 ± 0.43	61.42 ± 0.51
TEXT2EVENT	61.14 ± 0.80	65.93 ± 0.69	63.44 ± 0.19	56.76 ± 0.97	66.78 ± 0.48	61.36 ± 0.77
SPEECH w/o energy	72.91 ± 0.76	72.81 ± 0.76	72.86 ± 0.77	58.92 ± 0.96	58.45 ± 1.08	58.69 ± 1.40
	71.22 ± 0.58	71.07 ± 0.45	71.12 ± 0.45	56.12 ± 1.87	55.69 ± 1.66	55.91 ± 1.76

Table 3: Performance (%) of event classification on MAVEN-ERE *valid set* and ONTOEVENT-DOC *test set*.