Investigating the Challenges of **Temporal Relation Extraction from Clinical Text** Diana Galvan¹, Naoaki Okazaki², Koji Matsuda¹, Kentaro Inui^{1,3} ¹Tohoku University ²Tokyo Institute of Technology ³RIKEN Center for Advanced Intelligence Project

Why?



тоноки UNIVERSITY

Overview

- Temporal reasoning remains as an unsolved NLP task, particularly in the clinical domain.
- State-of-the-art systems are good at EVENT and TIME EXPRESSION identification but perform poorly in Temporal Relation Extraction.

Temporal relations (TLINK)



Adapting Miwa and Bansal (2016) tree-based LSTM model we obtained a 2 points improvement over the state-of-the-art.

Related Work

Best 2016 Clinical TempEval System Method: HMM and SVM's

	Precision	Recall	F1
(Lee et al., 2016)	0.588	0.559	0.573
(Lin et al., 2016)	0.669	0.534	0.594
(Leeuwenberg and Moens, 2017)	_	_	0.608
This work	0.983	0.462	0.629
Human performance	_	-	0.817

2 points gap below human performance

5 points improvement using neural networks

000	rurre	h d	lur i	ing	EVE chem	NT Othe	eran				
·	20110			G., I		EVENT		<u>y</u> .			Patient's timeline
	Ex	pe	rir	nei	nts						······································
	Bin	ary c	lassi	ficatio	n						· · · · · · · · · · · · · · · · · · ·
	TLINI	K	Р	R	F1		Tra	inad a		bcot of	iournal abstracts
	CONTAIN	IS (0.983	0.462	0.629		i	n Onc	ology	and Ga	stroenterology
		I	I								
	Multi	-clas	s clas	sifica	tion						
		Wiki	pedia wo	ord emb	PubN	Med word	emb	PubMed	word em	b + FNE	
1	FLINK	Р	R	F1	Р	R	F1	Р	R	F1	Filtered Negative
BEFC	ORE	0.698	0.185	0.292	0.708	0.198	0.310	0.683	0.202	0.312	removed pairs that
BEGI	NS-ON	0.585	0.062	0.112	0.615	0.103	0.177	0.608	0.116	0.195	according to the
CON	TAINS	0.905	0.472	0.621	0.908	0.471	0.620	0.889	0.479	0.623	guidelines, should
				+						<u> </u>	never be reinikeu.

> Findings

of instances TLINK Accuracy Overlap(e_1, e_2) 1291 0.14 BeginsOn(e_2 , e_1) 0.14 176 Same accuracy, 7x instances

Dev set: Misclassified Event-Event pairs

ENDS-ON

OVERLAP

TLINK	V-V	V-NV	*NV-V	NV-NV
CONTAINS	6	47	24	103
OVERLAP	6	55	27	193
Total	12	102	51	296

– Verb NV – Non-Verb

Annotated events are mostly **nouns**

0.126 0.213 0.760 0.126

0.504 0.134 0.211 0.497

0.216

0.218

0.140

Domain knowledge: Itching is an irritating and uncontrollable sensation.

Time: Naturally associated to **verbs**

TLINK	Accuracy	# of instances
Overlap(e ₂ , e ₁)	0.024	> 290
Before(e ₂ , e ₁)	0.34	353

In <u>timex1:today</u>'s <u>e1:visit</u> the patient said he had some <u>e2:itching</u>, mostly timex2:this morning. timex1: today

0.704

Temporality of nominal events

0.148

0.211

0.520 0.086

0.504 0.134

Contains (timex1, e1)

el: visit V **Definite time interval:** There is a natural endpoint. Complete containment is clear.



Conclusion and Future work

- Event Event relations are the more frequent and most difficult pairings to handle.
- High use of domain knowledge to understand the temporal properties of an event.
- We plan to analyze further Event Event relations differentiating events as verbal and non-verbal events.

Contact / Acknowledgement

- Contact author: Diana Galvan, dianags@ecei.tohoku.ac.jp
- This work was supported by JST CREST Grant Number 1513