GENEVA: Benchmarking Generalizability for Event Argument Extraction with Hundreds of Event Types and Argument Roles

Tanmay Parekh[1], I-Hung Hsu[2], Kuan-Hao Huang[1], Kai-Wei Chang[1], Nanyun Peng[1]

[1]University of California, Los Angeles [2]University of Southern California

Introduction

Introducing a diverse dataset GENEVA for Event Argument Extraction (EAE) to evaluate generalizability of EAE models

Event Type: Destroying

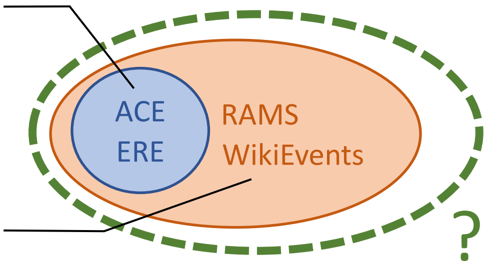
The subsequent explosions obliterated the museum

Cause | Patient

EAE = Extracting event-specific arguments and roles from text

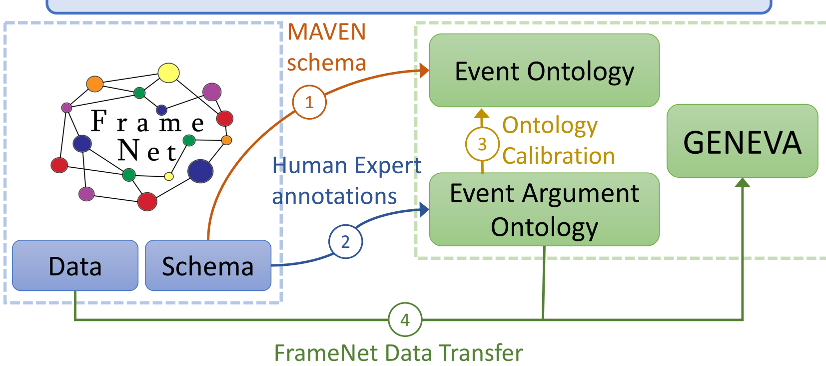
1. Limited number of event types and argument roles
2. Entity-only argument roles
3. Less diversity in events

1. Vast number of events
2. Moderate number of argument roles
3. Entity-only argument roles
4. Less diversity in events



GENEVA Creation

Transferring from existing SRL dataset – FrameNet for EAE



Quality Checks: Quality Assessment | Comprehensiveness Assessment

Qualitative Examples

Sentence	Event + Trigger	Arguments
With rail service in place and forty blocks of private property , it was ready to become a real town .	Event: Becoming Trigger: become	Entity: it Final_category: a real town
Canadian companies sent $28.5 billion in goods to the United States in February , up 1.6% from January revised level , while they imported $20.9 billion worth , up 2.4% .	Event: Sending Trigger: sent	Sender: Canadian companies Theme: $28.5 billion in goods Recipient: United States

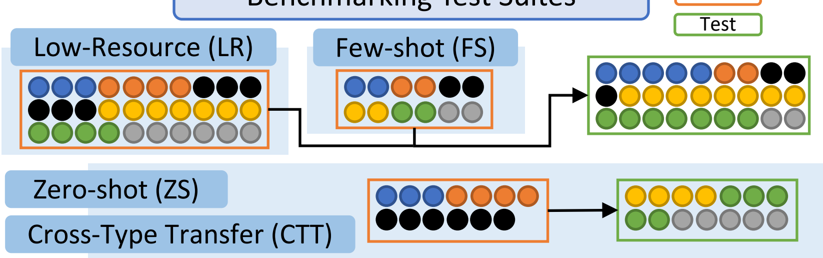
Experimental Setup

Classification-based: DyGIE++ [1], OneIE [2], Query-Extract [3]

Question-Answering: BERT-QA [4], TE [5]

Generation-based: TANL [7], DEGREE [8], GPT-3.5 turbo

Benchmarking Test Suites



Code + Data | **Preprint**

SCAN ME | SCAN ME

Correspondence

Tanmay Parekh
tparekh@cs.ucla.edu
I-Hung Hsu
ihunghsu@isi.edu

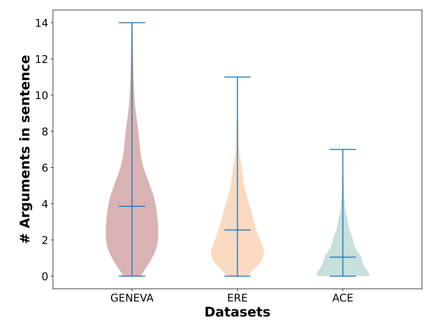
Data Statistics

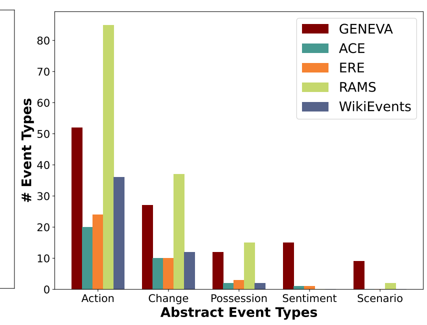
	ACE	RAMS	Full	GENEVA
# Event Types	33	139	179	115
# Abstract Event Types	2	3	5	5
# Argument Roles	22	65	362	220
Avg. Roles per Event	4.75	3.76	4.82	3.97
% Entity Argument Roles	100%	100%	65%	63%
% Non-Entity Argument Roles	0%	0%	35%	37%

- Vast number of events and argument roles
- Introduction of non-entity argument roles
- More diverse events

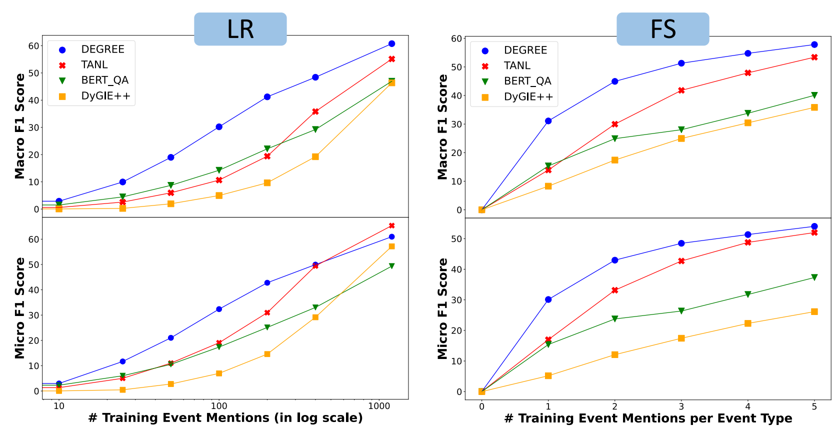
Dataset	# Event Types	# Arg Types	Avg. Event Mentions	Avg. Arg Mentions
ACE	33	22	153.18	274.55
ERE	38	21	191.76	499
GENEVA	115	220	65.26	55.77

Diverse | Dense | High Coverage | Difficult to Learn





Results



Model	ZS-1	ZS-5	ZS-10	CTT
TE	7.54	7.54	7.54	6.39
BERT-QA	5.05	21.53	24.24	11.17
DEGREE	**24.06**	**34.68**	**39.43**	**27.9**

DEGREE and generative-models establish superior generalizability

GENEVA is Challenging!

	LR-400		ZS-10	
	GENEVA	ACE	GENEVA	ACE
BERT-QA	33	-	24.2	46.7
DEGREE	49.9	57.3	39.4	53.3

Model Performance on ACE > GENEVA

Model	Entity	Non-Entity	Diff
DEGREE	54.46	39.89	14.57
TANL	52.54	42.4	10.14
BERT-QA	36.71	24.86	11.85

Non-entity argument roles are tough!

GPT – 22.73 (GPT 3.5-turbo)

DEGREE – 39.43 (ZS-10)

GPT Prompt

Passage: Assistance in the establishment of a factory to assemble the DPRK Scud variant missiles .
Event: creating. Trigger: The event trigger word is establishment
Query: The created entity is some created entity. The creator is some creator. The cause is some cause.
Output: The created entity is of a factory. The creator is some creator. The cause is some cause.

In-context Examples ... Test Example

Passage: In the case of North Korea , determining the status of its nuclear weapons program is especially difficult .
Event: confronting problem. Trigger: The event trigger word is difficult
Query: The activity is some activity. The experiencer is some experiencer.

[1] Wadden, David, et al. "Entity, relation, and event extraction with contextualized span representations." (2019).
[2] Lin, Ying, et al. "A joint neural model for information extraction with global features." (2020).
[3] Wang, Sijia, et al. "Query and extract: Refining event extraction as type-oriented binary decoding." (2021).
[4] Du, Xinya, and Claire Cardie. "Event extraction by answering (almost) natural questions." (2020).
[5] Lyu, Qing, et al. "Zero-shot event extraction via transfer learning: Challenges and insights." (2021).
[6] Paolini, Giovanni, et al. "Structured prediction as translation between augmented natural languages." (2021).
[7] Hsu, I., et al. "DEGREE: A data-efficient generative event extraction model." (2021).