A Consolidated Open Knowledge Representation for Multiple Texts

Rachel Wities, Vered Shwartz, Gabriel Stanovsky, Meni Adler, Ori Shapira, Shyam Upadhyay, Dan Roth, Eugenio Martinez Camara, Iryna Gurevych and Ido Dagan

Outline:

- Consolidated semantic representation for multiple texts
- Annotated dataset of news-related tweets
- Automatic baseline and results

Consolidated Representation

Single Sentence Semantic Representations

Semantic representations are focused on single sentences.

Single Sentence Semantic Representations

Semantic representations are focused on single sentences.

Example: Open IE pred-arg tuples:

3 people dead in shooting in Wisconsin.

1. *(**shooting in**, Wisconsin)*
2. *(three, **dead in**, shooting)*

Goal: Consolidated Representation

Applications often need to consolidate information from multiple texts:

Goal: Consolidated Representation

Applications often need to consolidate information from multiple texts:

3 people dead in shooting in Wisconsin.
Man kills three in Spa shooting.
Shooter was identified as Radcliffe Haughton, 45.

- Question answering
  - *How many people did Radcliffe Haughton shoot?*
- Abstractive summarization
  - *Radcliffe Haughton, 45, kills three in Spa shooting in Wisconsin.*

Goal: Consolidated Representation

Applications often need to consolidate information from multiple texts:

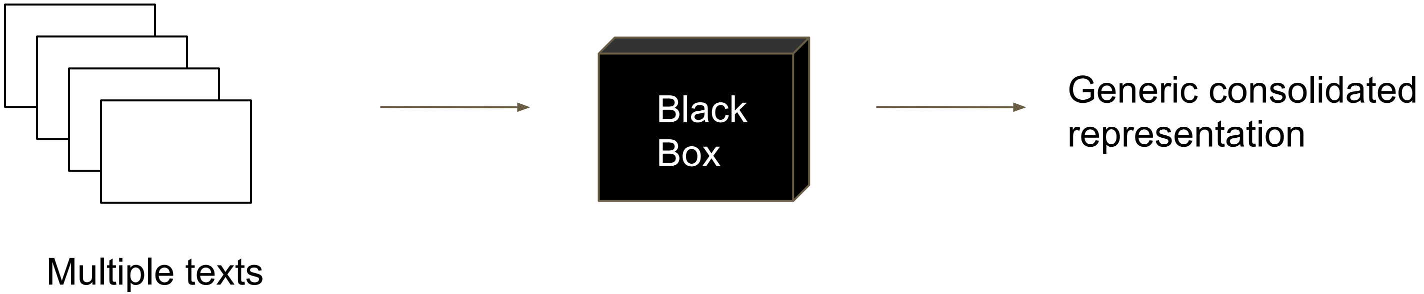
3 people dead in shooting in Wisconsin.
Man kills three in Spa shooting.
Shooter was identified as Radcliffe Haughton, 45.

- Question answering
  - *How many people did Radcliffe Haughton shoot?*
- Abstractive summarization
  - *Radcliffe Haughton, 45, kills three in Spa shooting in Wisconsin.*

Consolidation usually done at the application level, to a partial extent.

Our Proposal: Consolidated Propositions

- Generic semantic structures that represent multiple texts
- Can be used for various semantic applications
- “Out of the box” - another step in the semantic NLP pipeline



Our Solution

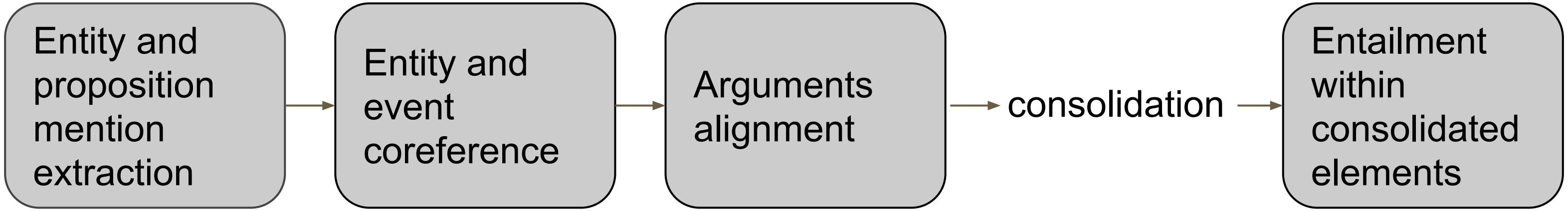
1. Predicate-argument structure for single sentences
    - Current scope: Open IE
2. Consolidating propositions based on coreference
3. Representing information overlap/containment via lexical entailments

Our Solution

1. Extract propositions for single sentences
    - Current scope: use Open IE proposition
2. Consolidating propositions based on coreference
3. Representing information overlap/containment via lexical entailments

⇒ Open Knowledge Representation structure (OKR)

OKR Pipeline



- Leverage **known** NLP tasks!

Entity & Proposition Extraction

Entity and proposition mention extraction → Entity and event coreference → Arguments alignment → consolidation → Entailment within consolidated elements

- Extract entity and proposition mentions at single sentence level:

3 people dead in shooting in Wisconsin.

Man kills three in spa shooting .

Shooter was identified as Radcliffe Haughton, 45.

Entity mentions:

1. *3 people*
2. *Wisconsin*
3. *man*
4. *Three*
5. *…*

Proposition mentions:

1. *(3 people, **dead in**, shooting)*
2. *(**shooting in**, Wisconsin)*
3. *(Man, **kills**, three, shooting)*
4. *(spa, **shooting**)*
5. *…*

Entity Coreference

Entity and proposition mention extraction → **Entity and event coreference** → Arguments alignment → consolidation → Entailment within consolidated elements

- Create coreference chains of entity mentions

3 people dead in shooting in Wisconsin.

Man kills three in spa shooting .

Shooter was identified as Radcliffe Haughton, 45.

Entities:

E1: {3 people, three}

E2: {man, shooter, Radcliffe Haughton}

E3: ...

Event Coreference

Entity and proposition mention extraction → **Entity and event coreference** → Arguments alignment → consolidation → Entailment within consolidated elements

- Create coreference chains of entity mentions

3 people dead in shooting in Wisconsin.

Man kills three in spa shooting .

Shooter was identified as Radcliffe Haughton, 45.

*P1: {(3 people, **dead in**, shooting), (Man, **kills**, three, shooting)}*

*P2: {(**shooting in**, Wisconsin), (spa, **shooting**)}*

P3: …

Argument Alignment

Entity and proposition mention extraction → Entity and event coreference → **Arguments alignment** → consolidation → Entailment within consolidated elements

- Align arguments of corefering propositions based on semantic role:

a2 a3 a1 a2 a3

*P1: {(3 people, **dead in**, shooting), (Man, **kills**, three, shooting)}*

a1 a1

*P2: {(**shooting in**, Wisconsin), (spa, **shooting**)}*

Consolidation of propositions:

Entity and proposition mention extraction → Entity and event coreference → Arguments alignment → consolidation → Entailment within consolidated elements

*P1: {(3 people, **<u>dead in</u>**, shooting), (Man, **<u>kills</u>**, three, shooting)}*

{ [a2] dead in [a3],
[a1] kills [a2] in [a3] }

Consolidation of propositions:

Entity and proposition mention extraction → Entity and event coreference → Arguments alignment → consolidation → Entailment within consolidated elements

a2 a3 a1 a2 a3

*P1: {(3 people, **dead in**, shooting), (Man, **kills**, three, shooting)}*

{ [a2] dead in [a3],
[a1] kills [a2] in [a3] }

a1

a2

a3

Consolidation of propositions:

Entity and proposition mention extraction → Entity and event coreference → Arguments alignment → consolidation → Entailment within consolidated elements

a2 a3 a1 a2 a3

*P1: {(3 people, **dead in**, shooting), (Man, **kills**, three, shooting)}*

E1: {3 people, three}

E2: {man, shooter, Radcliffe Haughton}

{ [a2] dead in [a3],
[a1] kills [a2] in [a3] }

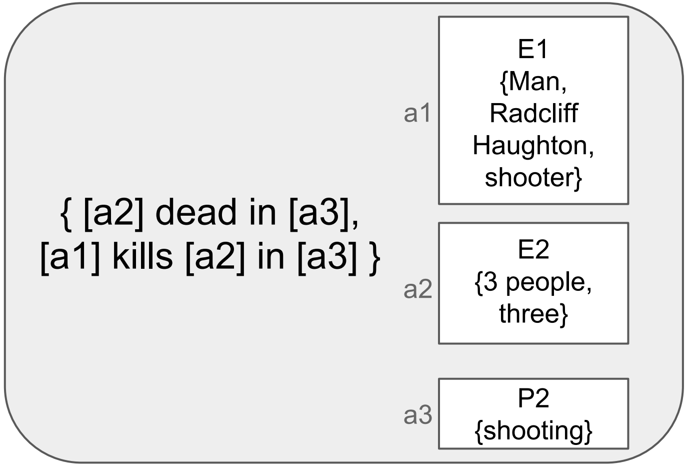
a1 E1 {Man, Radcliff Haughton, shooter}

a2 E2 {3 people, three}

a3 P2 {shooting}

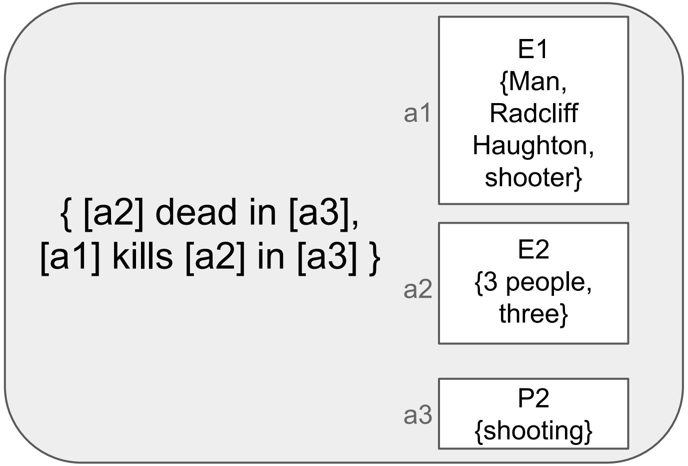
Consolidation Properties:

- All proposition information is concentrated in one structure
- No redundancy
- Tracking all original mentions
- Allow generation of new sentences
    - “Radcliff Haughton kills 3 people in shooting”



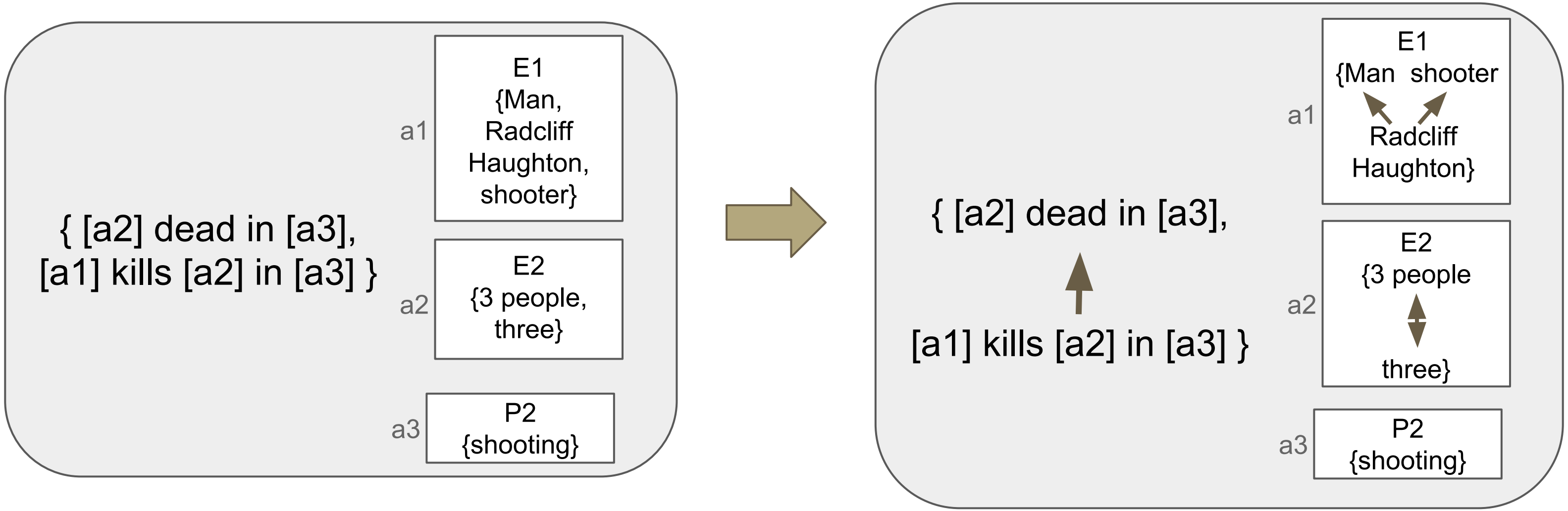
Still missing: modeling information overlap

- "killed" is more specific than "dead"
- "man" is more general than "Radcliff Haughton"
- Need to model level of specificity of mentions
- Our proposal: entailment graphs within structure components



Entailment between Elements

Entity and proposition mention extraction → Entity and event coreference → Arguments alignment → consolidation → Entailment within consolidated elements



Dataset and Baselines

News-Related Tweets Dataset

- OKR Annotation of 1257 news-related tweets from 27 event clusters
  - Collected from the Twitter Event Detection Dataset (McMinn et al., 2013)

- Annotated Dataset characteristics:
  - High proportion of nominal predicates - 39%
    - Example: *accident, demonstration*
  - High entailment connectivity within coreference chains
    - 96% of our entailment graphs (entity and proposition) form a connected component

Inter-Annotator Agreement

	Entity Extraction *(avg. accuracy)*	Entity Coref. *(CoNNL F1)*	Proposition extraction *(avg. accuracy)*				Predicate coreference *(CoNNL F1)*	Entailment *(F1)*	
			Predicates			Arguments		Entities	Predicates
agreement	**.85**	**.90**	**.74**	**Verbal .93**	**Non verbal .72**	**.85**	**.83**	**.70**	**.82**

Inter-annotator agreement

	Entity Extraction *(avg. accuracy)*	Entity Coref. *(CoNNL F1)*	Proposition extraction *(avg. accuracy)* Predicates			Proposition extraction Arguments	Predicate coreference *(CoNNL F1)*	Entailment *(F1)* Entities	Entailment *(F1)* Predicates
agreement	**.85**	**.90**	**.74**	**Verb.: .93**	**Non verb. .72**	**.85**	**.83**	**.70**	**.82**

- Entity or Predicate?
    - Examples: *terror, hurricane*

Baselines

- Perform pipeline tasks independently
- A simple baseline for each task:
  - **Entity extraction** – spaCy NER model and all nouns.
  - **Proposition extraction** - Open IE propositions extracted from PropS (Stanovsky et al., 2016).
  - **Proposition and Entity coreference -** clustering based on simple lexical similarity metrics
    - lemma matching, Levenshtein distance, Wordnet synset.
  - **Argument alignment –** align all mentions of the same entity
  - **Entity Entailment -** knowledge resources (Shwartz et al., 2015) and a pre-trained model for HypeNET (Shwartz et al., 2016)
  - **Predicate Entailment -** rules extracted by Berant et al. (2012)

Baselines - results

	Entity Extraction (avg. accuracy)	Entity Coref. (CoNNL F1)	Proposition extraction (avg. accuracy)				Predicate coreference (CoNNL F1)	Entailment (F1)	
			Predicates			Arguments		Entities	Predicates
agreement	.85	.90	.74	Verb. .93	Non verb. .72	.85	.83	.70	.82
predicted	**.58**	**.85**	**.41**	**Verb. .73**	**Non verb. .25**	**.37**	**.56**	**.44**	**.56**

Baselines - results

	Entity Extraction *(avg. accuracy)*	Entity Coref. *(CoNNL F1)*	Proposition extraction *(avg. accuracy)* Predicates			Arguments	Predicate coreference *(CoNNL F1)*	Entailment *(F1)* Entities	Predicates
agreement	.85	.90	.74	Verb. .93	Non verb. .72	.85	.83	.70	.82
predicted	**.58**	**.85**	**.41**	**Verb. .73**	**Non verb. .25**	**.37**	**.56**	**.44**	**.56**

- Main challenges:
  - Recognize arguments for nominal predicates - current systems are verb-centric (well known)

Baselines - results

	Entity Extraction *(avg. accuracy)*	Entity Coref. *(CoNNL F1)*	Proposition extraction *(avg. accuracy)*				Predicate coreference *(CoNNL F1)*	Entailment *(F1)*	
			Predicates			Arguments		Entities	Predicates
agreement	.85	.90	.74	Verb. .93	Non verb. .72	.85	.83	.70	.82
predicted	**.58**	**.85**	**.41**	**Verb. .73**	**Non verb. .25**	**.37**	**.56**	**.44**	**.56**

- Main challenges:
    - Recognize arguments for nominal predicates - current systems are verb-centric (well known)
    - Distinguish entity nouns from predicate nouns (*organization* vs. *elections*)

Baselines - results

	Entity Extraction *(avg. accuracy)*	Entity Coref. *(CoNNL F1)*	Proposition extraction *(avg. accuracy)* Predicates			Proposition extraction *(avg. accuracy)* Arguments	Predicate coreference *(CoNNL F1)*	Entailment *(F1)* Entities	Entailment *(F1)* Predicates
agreement	.85	.90	.74	Verb. .93	Non verb. .72	.85	.83	.70	.82
predicted	**.58**	**.85**	**.41**	**Verb. .73**	**Non verb. .25**	**.37**	**.56**	**.44**	**.56**

- Main challenges:
    - Recognize arguments for nominal predicates - current systems are verb-centric (well known)
    - Distinguish entity nouns from predicate nouns (*organization* vs. *elections*)
    - Entity entailment is hard for multi-word expressions

Baselines - results

	Entity Extraction *(avg. accuracy)*	Entity Coref. *(CoNNL F1)*	Proposition extraction *(avg. accuracy)*				Predicate coreference *(CoNNL F1)*	Entailment *(F1)*	
			Predicates			Arguments		Entities	Predicates
agreement	.85	.90	.74	Verb. .93	Non verb. .72	.85	.83	.70	.82
predicted	**.58**	**.85**	**.41**	**Verb. .73**	**Non verb. .25**	**.37**	**.56**	**.44**	**.56**

- Main challenges:
  - Recognize arguments for nominal predicates - current systems are verb-centric (well known)
  - Distinguish entity nouns from predicate nouns (*organization* vs. *elections*)
  - Entity entailment is hard for multi-word expressions
  - Predicate coreference is harder

Future work:

- Using OKR for summarization and for for interactive text exploration
- OKR Version 2
  - Avoid distinguishing entities from predicates
  - Knowledge-graph perspective
- Consolidation of other types of predicate-argument structures:
  - SRL
  - AMR

Summary

- We present a generic semantic representation for multiple texts
- Consolidating propositions using coreference and entailment
- 1257 annotated tweets
- Our dataset is available at:
  http://u.cs.biu.ac.il/~nlp/resources/downloads/twitter-events/


ThaNK
YOU

Outline:

- Intro: motivation & positioning
- Our solution:
  - Focus in this work: Open -IE predicate-argument structure for single sentences
  - Consolidation of propositions using coreference
  - Representing information overlap/containment via lexical entailments
- Pipeline:
  - OIE extraction (show for a sentence, with same visual output - for single extractions)
  - Entity and event coref (same visual)
  - Consolidation - final visual (as in intro teaser)
- Notes bullet slides - phenomena addressed - see paper: (2-3 points)
  - Nested propositions, implicit predicates, predicate representation as templates
- Dataset and baseline slides - like in Saarland presentation
- Conclusions
  - ?yes KG perspective
  - We focused on creating multi-text representations from OIE single sentence; future work may explore analogous representations based on other single sentence representations (e.g. AMR)

Other phenomena addressed (see paper for more details)

- Implicit and relation predicates
  - Examples: *Radcliffe Haughton, 45 ⇒ IMPLICIT (Radcliffe Haughton;45)*
- Support
  - Number of mentions of each proposition is indicative to factuality and salience.
- Predicate representation as templates
  - DIRT-like propositions

Proposition Consolidation

Proposition and entity mention extraction → Entity and event coreference → Proposition and entity consolidation → Entailment between elements

Propositions mentions:

1. *Dead in (At least 2; shooting)*
2. *Shooting in (Wisconsin)*
3. *Kills in (Man; three; shooting)*
4. *Shooting (Spa)*
5. *...*

Propositions:

P1

{ shooting in [a1],
[a1] shooting] }

a1

E1
{Wisconsin}
E3
{spa}

OKR pipeline

Proposition and entity mention extraction → Entity and event coreference → **Entity and proposition consolidation** → Entailment between elements

Entities mentions:

1. *At least 2*
2. *Wisconsin*
3. *Man*
4. *Three*
5. *Spa*
6. *Radcliffe Haughton*
7. ...

Entities:

E1: {Man, Radcliff Haughton)

E2: {At least 2}

...