Event Argument Extraction for TAC KBP Events 2014

Results on ACE2005 and Lessons Learned from the Task
Yangqiu Song et al.
UIUC
Outline

• Event Annotation and Task

• Event Extraction Pipeline

• Inference of World Knowledge
Examples of Event (1)

They also deployed along the border with Israel.

**Event Type:** Movement.Transport

**Argument Role:**
- **Person**
- **Trigger:** Transport
- **Destination**

**Argument Role:** Event Type: Movement.Transport
After Sept. 11, 2001, Indonesia was quick to sign onto U.S. President George W. Bush's global war on terror.

Event Type: Conflict.Attack
In Baghdad, a cameraman died when an American tank fired on the Palestine Hotel.

Event Type: Life.Die
Event Type: Conflict.Attack
## Event Definition

- **8 Types, 33 sub-types**

<table>
<thead>
<tr>
<th>Type</th>
<th>Subtype</th>
<th>Role</th>
<th>Valid Entity Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conflict</td>
<td>Attack</td>
<td>Agent[1]</td>
<td>PER, ORG, GPE</td>
</tr>
<tr>
<td></td>
<td>Demonstrate</td>
<td>Artifact[1]</td>
<td>PER, WEA, VEH</td>
</tr>
<tr>
<td></td>
<td>Meet</td>
<td>Attacker</td>
<td>PER, ORG, GPE</td>
</tr>
<tr>
<td></td>
<td>Phone-Write</td>
<td>Destination</td>
<td>GPE, LOC, FAC</td>
</tr>
<tr>
<td>Life</td>
<td>Marry</td>
<td>Instrument</td>
<td>WEA, VEH</td>
</tr>
<tr>
<td></td>
<td>Divorce</td>
<td>Price[2]</td>
<td>NUM</td>
</tr>
<tr>
<td></td>
<td>Injure</td>
<td>Target</td>
<td>PER, ORG, VEH, FAC, WEA</td>
</tr>
<tr>
<td></td>
<td>Die</td>
<td>Vehicle</td>
<td>VEH</td>
</tr>
<tr>
<td></td>
<td>Transport</td>
<td>Victim</td>
<td>PER</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event Subtype</th>
<th>ARG1</th>
<th>ARG2</th>
<th>ARG3</th>
<th>ARG4</th>
<th>ARG5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conflict.Attack</td>
<td>Attacker</td>
<td>Target</td>
<td>Instrument</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The world's most powerful fine art auction houses, Sotheby's and Christie's, have agreed to pay 40 million dollars to settle an international price-fixing scam, Sotheby's said. The payment, if approved by the courts, would settle a slew of suits by clients over auctions held between 1993 and 2000....

In December 2001, a Manhattan jury convicted Sotheby's main shareholder and former chairman Alfred Taubman of entering into a price-fixing agreement with Christie's. Four months later, the billionaire was sentenced to one year in prison and a 7.5-million-dollar fine over the price-fixing, which prosecutors said had cost clients of both firms 400 million dollars.
### A Comparison of the Events and Relations Across ACE, ERE, TAC-KBP, and FrameNet Annotation Standards

J. Aguilar et al. 2014. ACL Workshop.
Structure of ACE Annotation

• **Entity:**
  – Named entity: GPE; FACILITY; WEA; VEH; LOCATION; PERSON; ORGANIZATION
    • Entity mention list
  – Time
  – Value: Crime; Contact-Info; Numeric; Sentence; Job-Title

• **Relation:**

• **Event:**
  – Event mention list
    • Anchor/trigger
    • Argument list/role list
Outline

• Event Annotation and Task

• Event Extraction Pipeline

• Inference of World Knowledge
A Two-step Approach

• Trigger detection
  – Trigger identification
  – Trigger classification

• Argument detection
  – Mention detection
  – Argument identification
  – Argument classification

But after spending 40 days in prison, Jordan's King Abdullah II pardoned the former legislator known for her harsh criticism of the state.
Feature Engineering

- **Trigger Features**
  - Context/lemma/postag (left, right, center) unigram, bigram
  - WordNet extension (hypernym, hyponym, lemma, similar)
  - Is in FrameNet, VerbNet, PropBank
  - VerbNet Class
  - Concept from ESA
  - Word2vec

- **Mention Features**
  - Context/lemma/postag (left, right, center) unigram, bigram
  - WordNet extension (hypernym, hyponym, lemma, similar)
  - Concept from ESA
  - Word2vec
  - Head word
  - Coreference mention
  - Named entity types
  - Entity type/subtypes from ACE

- **Pair Features**
  - Path (through relation/node) from dependency parser
  - Path (through word/postag) from parse tree
  - Common ancestor (left/right path)
  - Position of argument (right, left, overlapped)
## Feature Evaluation

<table>
<thead>
<tr>
<th>View</th>
<th>Feature Number</th>
<th>F1 (Each View)</th>
<th>F1 (Remove Each)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Context</td>
<td>13535</td>
<td>0.6206</td>
<td>0.7277</td>
</tr>
<tr>
<td>Stanford Lemma</td>
<td>11020</td>
<td>0.6475</td>
<td>0.7279</td>
</tr>
<tr>
<td>Concept (ESA 500)</td>
<td>172047</td>
<td>0.6129</td>
<td>0.7208</td>
</tr>
<tr>
<td>Word2vector (500)</td>
<td>500</td>
<td>0.5704</td>
<td>0.7243</td>
</tr>
<tr>
<td>WordNet Synset Words</td>
<td>11992</td>
<td>0.6329</td>
<td>0.7301</td>
</tr>
<tr>
<td>WordNet Hyponym</td>
<td>18241</td>
<td>0.5669</td>
<td>0.7285</td>
</tr>
<tr>
<td>WordNet Hypernym</td>
<td>3856</td>
<td>0.5607</td>
<td>0.7301</td>
</tr>
<tr>
<td>WordNet Synonym</td>
<td>20378</td>
<td>0.5398</td>
<td>0.7213</td>
</tr>
<tr>
<td>WordNet Lemma List</td>
<td>5215</td>
<td>0.6527</td>
<td>0.7255</td>
</tr>
<tr>
<td>WordNet Stem List</td>
<td>5066</td>
<td>0.6564</td>
<td>0.7240</td>
</tr>
<tr>
<td>VerbNet Class</td>
<td>262</td>
<td>0.3962</td>
<td>0.7287</td>
</tr>
<tr>
<td>Check Verb List (FrameNet/VerbNet/PropBank)</td>
<td>6</td>
<td>0.0000</td>
<td>0.7098</td>
</tr>
</tbody>
</table>

All: 0.7315
Colum-wise Normalization

- For each column (each feature)
  - $f \leftarrow 2 \ast (f - \text{min}(f)) / (\text{min}(f) - \text{max}(f)) - 1 : f \text{ in } [-1, +1]$
  - $f \leftarrow (f - \text{mean}(f)) / \text{std}(f) : f \sim \text{N}(0, 1)$
Row-wise Normalization

- For each row
- \( x \leftarrow x / \|x\| : \|x\| = 1 \)
Row-wise Normalization for Each View

- For each view
- $x \leftarrow x / \|x\| : \|x\| = 1$
Feature Normalization/Selection

• No normalization
  – 0.6165 (F1)

• Column-wise normalization
  – 0.6442

• Row-wise normalization
  – 0.7257

• View based row-wise normalization
  – TF: 0.7315
  – Multiplication with IDF(Wiki): **0.7408**

• Feature selection with information gain
  – 0.7336
They also deployed along the border with Israel.

<table>
<thead>
<tr>
<th>Event Type</th>
<th>ARG1</th>
<th>ARG2</th>
<th>ARG3</th>
<th>ARG4</th>
<th>ARG5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conflict.Attack</td>
<td>Attacker</td>
<td>Target</td>
<td>Instrument</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Event Detection Pipeline

### Event Detection Pipeline

- **Trigger Identification (1 Binary)**
- **Trigger Classification (1 33-class)**
- **Each Subtype: Argument Identification (33 Binary)**
- **Each Subtype: Argument Classification (33 Multi-class)**
- **Mention Type Checking**

---

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Step (1 34-class)</td>
<td>0.7869</td>
<td>0.5478</td>
<td>0.6460</td>
</tr>
<tr>
<td>Identification (0.3)</td>
<td>0.6704</td>
<td>0.6520</td>
<td>0.6611</td>
</tr>
<tr>
<td>Classification</td>
<td>0.9242</td>
<td>0.9242</td>
<td>0.9242</td>
</tr>
<tr>
<td>Identification + Classification</td>
<td>0.6853</td>
<td>0.6148</td>
<td>0.6482</td>
</tr>
</tbody>
</table>
Event Detection Pipeline

- Trigger Identification (1 Binary)
- Trigger Classification (1 33-class)
- Each Subtype: Argument Identification (33 Binary)
- Each Subtype: Argument Classification (33 Multi-class)
- Mention Type Checking

<table>
<thead>
<tr>
<th>Configuration</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Step</td>
<td>0.6237</td>
<td>0.3465</td>
<td>0.4455</td>
</tr>
<tr>
<td>Identification (Best on validation set)</td>
<td>0.5473</td>
<td>0.4835</td>
<td>0.5134</td>
</tr>
<tr>
<td>Identification (Best on validation set) + Classification</td>
<td><strong>0.4993</strong></td>
<td><strong>0.4265</strong></td>
<td><strong>0.4601</strong></td>
</tr>
<tr>
<td>Identification (0.1)</td>
<td>0.4976</td>
<td>0.5354</td>
<td>0.5159</td>
</tr>
<tr>
<td>Identification (0.1) + Classification</td>
<td><strong>0.4531</strong></td>
<td><strong>0.4731</strong></td>
<td><strong>0.4629</strong></td>
</tr>
<tr>
<td>Identification (0.2)</td>
<td>0.5453</td>
<td>0.4949</td>
<td>0.5189</td>
</tr>
<tr>
<td>Identification (0.2) + Classification</td>
<td><strong>0.4986</strong></td>
<td><strong>0.4373</strong></td>
<td><strong>0.4659</strong></td>
</tr>
</tbody>
</table>
Event Detection Pipeline

<table>
<thead>
<tr>
<th>Stage</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Step</td>
<td>0.6237</td>
<td>0.3465</td>
<td>0.4455</td>
</tr>
<tr>
<td>One Step (Check types)</td>
<td>0.6332</td>
<td>0.3465</td>
<td>0.4479</td>
</tr>
<tr>
<td>Identification (0.2)</td>
<td>0.5453</td>
<td>0.4949</td>
<td>0.5189</td>
</tr>
<tr>
<td>Identification (0.2) + Classification</td>
<td>0.4986</td>
<td>0.4373</td>
<td>0.4659</td>
</tr>
<tr>
<td>Identification (0.2) (Check types)</td>
<td>0.5626</td>
<td>0.4949</td>
<td>0.5266</td>
</tr>
<tr>
<td>Identification (0.2) + Classification (Check types)</td>
<td>0.5140</td>
<td>0.4373</td>
<td>0.4726</td>
</tr>
</tbody>
</table>
## Classifier Voting

### Motivation

All the event types share "place" and "time"

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Agents</th>
<th>Roles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life.Marry</td>
<td>Person</td>
<td>Person</td>
</tr>
<tr>
<td>Life.Divorce</td>
<td>Person</td>
<td>Person</td>
</tr>
<tr>
<td>Personnel.Start-Position</td>
<td>Person</td>
<td>Entity[1]</td>
</tr>
<tr>
<td>Personnel.End-Position</td>
<td>Person</td>
<td>Entity[1]</td>
</tr>
<tr>
<td>Justice.Arrest-Jail</td>
<td>Agent[1]</td>
<td>Person</td>
</tr>
<tr>
<td>Justice.Trial-Hearing</td>
<td>Prosecutor</td>
<td>Adjudicator</td>
</tr>
<tr>
<td>Justice.Sentence</td>
<td>Adjudicator</td>
<td>Defendant</td>
</tr>
<tr>
<td>Justice.Charge-Indict</td>
<td>Prosecutor</td>
<td>Adjudicator</td>
</tr>
<tr>
<td>Justice.Sue</td>
<td>Plaintiff</td>
<td>Adjudicator</td>
</tr>
</tbody>
</table>
Event Detection Pipeline (Voting)

- Trigger Identification (1 Binary)
- Trigger Classification (1 8-class, 1 33-class)
- Each type: Argument Identification (8 Binary)
- Each Subtype: Argument Identification (33 Binary)
- All: Argument Identification (1 Binary)
- All: Argument Classification (1 Multi-class)
- Each type: Argument Classification (8 Multi-class)
- Each Subtype: Argument Classification (33 Multi-class)
- Mention Type Checking

Two of the three agree
### Event Detection Pipeline (Voting)

<table>
<thead>
<tr>
<th>Without voting</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Step (Check types)</td>
<td>0.6332</td>
<td>0.3465</td>
<td>0.4479</td>
</tr>
<tr>
<td>With voting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Step (Check types)</td>
<td>0.6629</td>
<td>0.3501</td>
<td>0.4582</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Without voting</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification (0.2) (Check types)</td>
<td>0.5626</td>
<td>0.4949</td>
<td>0.5266</td>
</tr>
<tr>
<td>Identification (0.2) + Classification (Check types)</td>
<td>0.5140</td>
<td>0.4373</td>
<td>0.4726</td>
</tr>
<tr>
<td>With voting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification (0.2) (Check types)</td>
<td>0.5914</td>
<td>0.4544</td>
<td>0.5140</td>
</tr>
<tr>
<td>Identification (0.2) + Classification (Check types)</td>
<td>0.5456</td>
<td>0.4074</td>
<td>0.4665</td>
</tr>
</tbody>
</table>
Baselines: we used a different test set

<table>
<thead>
<tr>
<th>Methods</th>
<th>Trigger Identification (%)</th>
<th>Trigger Identification + classification (%)</th>
<th>Argument Identification (%)</th>
<th>Argument Role (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F1</td>
<td>P</td>
</tr>
<tr>
<td>Sentence-level in Hong et al. (2011)</td>
<td>N/A</td>
<td></td>
<td></td>
<td>67.6</td>
</tr>
<tr>
<td>Staged MaxEnt classifiers</td>
<td>76.2</td>
<td>60.5</td>
<td>67.4</td>
<td>74.5</td>
</tr>
<tr>
<td>Joint w/ local features</td>
<td>77.4</td>
<td>62.3</td>
<td>69.0</td>
<td>73.7</td>
</tr>
<tr>
<td>Joint w/ local + global features</td>
<td>76.9</td>
<td>65.0</td>
<td>70.4</td>
<td>73.7</td>
</tr>
<tr>
<td>Cross-entity in Hong et al. (2011)†</td>
<td>N/A</td>
<td></td>
<td></td>
<td>72.9</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Trigger</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Step</td>
<td>0.7869</td>
<td>0.5478</td>
<td>0.6460</td>
</tr>
<tr>
<td>Identification (0.3)</td>
<td>0.6704</td>
<td>0.6520</td>
<td>0.6611</td>
</tr>
<tr>
<td>Classification</td>
<td>0.9242</td>
<td>0.9242</td>
<td>0.9242</td>
</tr>
<tr>
<td>Identification + Classification</td>
<td>0.6853</td>
<td>0.6148</td>
<td>0.6482</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Argument</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without Voting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Step (Check types)</td>
<td>0.6332</td>
<td>0.3465</td>
<td>0.4479</td>
</tr>
<tr>
<td>Identification (0.2) (Check types)</td>
<td>0.5626</td>
<td>0.4949</td>
<td>0.5266</td>
</tr>
<tr>
<td>Identification (0.2) + Classification (Check types)</td>
<td>0.5140</td>
<td>0.4373</td>
<td>0.4726</td>
</tr>
<tr>
<td>With Voting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Step (Check types)</td>
<td>0.6629</td>
<td>0.3501</td>
<td>0.4582</td>
</tr>
<tr>
<td>Identification (0.2) (Check types)</td>
<td>0.5914</td>
<td>0.4544</td>
<td>0.5140</td>
</tr>
<tr>
<td>Identification (0.2) + Classification (Check types)</td>
<td>0.5456</td>
<td>0.4074</td>
<td>0.4665</td>
</tr>
</tbody>
</table>
Outline

• Event Annotation and Task

• Event Extraction Pipeline

• Inference of World Knowledge
World Knowledge

• Inferences of arguments may include inferring *casuality/part-of relations* between the verbal-events in a passage, inferring locations through part-of relations.
  
  – Agent argument of Life.Injure from the Attacker argument of Conflict.Attack.

• The assessor is instructed to judge “*Does this document support the claim* of the (EventType, Role, NormalizedArgumentString, Realis) tuple?”
Motivation

In Baghdad, a cameraman died when an American tank fired on the Palestine Hotel.

Event Type: Life.Die
Event Type: Conflict.Attack
New Arguments with World Knowledge

Source Event
Gadhafi's visit to Italy continued that process of emergence from international isolation.

Target Event
The Libyans say the retaliatory attacks killed 41 people, including Gadhafi's adopted daughter, and injured 226 others.
Word Knowledge Inference

• Step 1: Co-occurrence of subtypes in a document: \( f_1 \)
  – **Attack**:11821, **Die**:2808, Transport:2783, Transfer-Ownership:428, Injure:830, Arrest-Jail:280

• Step 2: Number of identical arguments (also coreference) of two subtypes: \( f_2 \)

• Step 3: Rank(Attack, Die)
  
  S1: \( \frac{f_2(\text{Die}, \text{Attack})}{f_1(\text{Die}, \text{Attack})} = \frac{505}{2808} \)
  
  S2: \( \frac{f_1(\text{Die}, \text{Attack})}{f_1(\text{Die}, \text{Die})} = \frac{2808}{11821} \)
  
  Rank = \( 2 \times S1 \times S2 / (S1 + S2) \)

  – For Attack: We select Die:0.2047, Transport:0.1759, Injure:0.1015

• Step 4: Check the argument co-reference frequency
  – **Attack@Attacker**: Die@Agent:122, Transfer-Ownership@Buyer:117, Transport@Person:95, Injure@Agent:42, Charge-Indict@Defendant:25, Sentence@Defendant:24, End-Position@Person:22, Arrest-Jail@Person:20, Convict@Defendant:18, Phone-Write@Entity:15
  – **Attack@Target**: Die@Victim:143, Transport@Person:53, Injure@Victim:35
  – **Attack@Place**: Transport@Destination:108, Die@Place:70, Injure@Place:31
  – **Attack@Instrument**: Die@Instrument:29, Injure@Instrument:13, Transport@Artifact:3
New Arguments with World Knowledge

Source Event
Gadhafi's visit to Italy continued that process of emergence from international isolation.

Target Event
The Libyans say the retaliatory attacks killed 41 people, including Gadhafi's adopted daughter, and injured 226 others.
Outline

• Event Annotation and Task

• Event Extraction Pipeline

• Inference of World Knowledge
Other Issues

• Mention Detection
  – Haoruo’s system: with named entities; without time and value
  – IllinoisNER/StanfordNER/OpenNLP
  – SRL arguments (Thanks to Christos and Mark)
  – Kai-Wei’s Co-reference System

• Mention Classification
  – 13 Types: Acc: 0.8045
  – 53 Subtypes: Acc: 0.7400

• Event Classification:
  – Generic/Actual/Other (Ray)

• Co-reference of entities
  – Kai-Wei’s system

• Time expression normalization
  – Illinois Time from Stephen

Thank you!