Identifying civilians killed by police with distantly supervised entity-event extraction

Katherine A. Keith, Abram Handler, Michael Pinkham, Cara Magliozzi, Joshua McDuffie, and Brendan O’Connor

EMNLP 2017

College of Information and Computer Science
University of Massachusetts Amherst
Killings by police in the U.S.

- July 17, 2014: Eric Garner, New York, NY
- Aug 9, 2014: Michael Brown, Ferguson, MO
- July 5, 2016: Alton Sterling, Baton Rouge, LA
- July 6, 2016: Philando Castile, Falcon Heights, MN
Data needed for policy making
Data needed for policy making

• Fatality Statistics?
Data needed for policy making

- Fatality Statistics?
- Racial disparity/discrimination?
Data needed for policy making

- Fatality Statistics?
- Racial disparity / discrimination?
- Most effective police departments / policing methods?
Data needed for policy making

- Fatality Statistics?
- Racial disparity/discrimination?
- Most effective police departments/policing methods?

DATA!
Issues in government data

[Banks et al. 2015 (BJS/DOJ)]

Number of U.S. police killings 2003-2009, 2011
Issues in government data

[Banks et al. 2015 (BJS/DOJ)]
Issues in government data

Number of U.S. police killings 2003-2009, 2011

[Banks et al. 2015 (BJS/DOJ)]
Issues in government data

Number of U.S. police killings 2003-2009, 2011

[Banks et al. 2015 (BJS/DOJ)]
Issues in government data

Number of U.S. police killings 2003-2009, 2011

[Banks et al. 2015 (BJS/DOJ)]
Issues in government data

Number of U.S. police killings 2003-2009, 2011

[Estimated: 7,427]
[ARD: 3,620]
[SHR: 3,385]
[ARD or SHR: 5,324]

[Banks et al. 2015 (BJS/DOJ)]
Issues in government data

Number of U.S. police killings 2003-2009, 2011

[Banks et al. 2015 (BJS/DOJ)]
Alternative data: media reports

- Populate an entity-event database by manually reading news articles

- FatalEncounters.org, KilledByPolice.net, The Guardian, Washington Post...
Alternative data: media reports

• Populate an **entity-event database** by manually reading news articles

• FatalEncounters.org, KilledByPolice.net, The Guardian, Washington Post...
Alternative data: media reports

- Populate an entity-event database by manually reading news articles
- FatalEncounters.org, KilledByPolice.net, The Guardian, Washington Post...
- Fatal Encounters volunteers have read >2 million articles
Number of U.S. police killings 2003-2009, 2011

[Banks et al. 2015 (BJS/DOJ)]
Number of U.S. police killings 2003-2009, 2011

[Banks et al. 2015 (BJS/DOJ)]
Motivation:
Public data and government accountability
Overview

Motivation:
Public data and government accountability

Problems with existing approaches:
1. Manual updates are expensive
2. Continuous updates required
Overview

Motivation:
Public data and government accountability

Problems with existing approaches:
1. Manual updates are expensive
2. Continuous updates required

Goal:
Automatically update a police fatality database
Overview
Overview

sentence w/ entity
sentence w/ entity
sentence w/ entity
sentence w/ entity
sentence w/ entity
Overview

[Image of a newspaper]
Overview
Overview
Outline

1. Motivation and overview
2. Task and data
3. Model
4. Training
5. Evaluation
6. Conclusion
## Example Dataset

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>July 17, 2014</td>
<td>Eric Garner</td>
</tr>
<tr>
<td>Aug 9, 2014</td>
<td>Michael Brown</td>
</tr>
<tr>
<td>July 5, 2016</td>
<td>Alton Sterling</td>
</tr>
<tr>
<td>July 6, 2016</td>
<td>Philando Castile</td>
</tr>
</tbody>
</table>
Example Dataset

Corpus

- July 17, 2014: Eric Garner
- Aug 9, 2014: Michael Brown
- July 5, 2016: Alton Sterling
- July 6, 2016: Philando Castile

Database
Task: Database update

Corpus

Train time
(Distant supervision)

Test time

Gold Database = Fatal Encounters

Eric Garner
Michael Brown
Alton Sterling
Philando Castile

Corpus

Test time
Collecting data

- Keyword-querying web scraper running throughout 2016
- Preprocessing: text extraction, deduplication, spaCy NER+parsing, name cleanups
## Data

<table>
<thead>
<tr>
<th>Knowledge base</th>
<th>Historical</th>
<th>Test</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>News dataset</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
</table>
## Data

<table>
<thead>
<tr>
<th>Knowledge base</th>
<th>Historical</th>
<th>Test</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>News dataset</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>total docs.</td>
<td>793,010</td>
<td>317,345</td>
</tr>
</tbody>
</table>
Data

<table>
<thead>
<tr>
<th>Knowledge base</th>
<th>Historical</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>FE gold entities</td>
<td>17,219</td>
<td>452</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>News dataset</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>total docs.</td>
<td>793,010</td>
<td>317,345</td>
</tr>
<tr>
<td>total ments.</td>
<td>132,833</td>
<td>68,925</td>
</tr>
<tr>
<td>pos. ments.</td>
<td>11,274</td>
<td>6,132</td>
</tr>
<tr>
<td>total entities</td>
<td>49,203</td>
<td>24,550</td>
</tr>
<tr>
<td>pos. entities</td>
<td>916</td>
<td>258</td>
</tr>
</tbody>
</table>
## Data

<table>
<thead>
<tr>
<th>Knowledge base</th>
<th>Historical</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>FE gold entities</td>
<td>17,219</td>
<td>452</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>News dataset</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>total docs.</td>
<td>793,010</td>
<td>317,345</td>
</tr>
<tr>
<td>total ments.</td>
<td>132,833</td>
<td>68,925</td>
</tr>
<tr>
<td>pos. ments.</td>
<td>11,274</td>
<td>6,132</td>
</tr>
<tr>
<td>total entities</td>
<td>49,203</td>
<td>24,550</td>
</tr>
<tr>
<td>pos. entities</td>
<td>916</td>
<td>258</td>
</tr>
</tbody>
</table>

Data upper bound: 258 / 452 = 57% recall
Outline

1. Motivation and overview
2. Task and data
3. Model
4. Training
5. Evaluation
6. Conclusion
Test time

Historical data (Distant supervision)

Test time

Corpus
The Baton Rouge Police Department confirms that Alton Sterling, 37, died during a shooting at the Triple S Food Mart.

... the two officers involved in Tuesday's shooting of Alton Sterling ...

... Alton Sterling was a resident of Baton Rouge...
The Baton Rouge Police Department confirms that confirms **Alton Sterling**, 37, died during a shooting at the Triple S Food Mart.

... the two officers involved in Tuesday's shooting of **Alton Sterling** ...

... **Alton Sterling** was a resident of Baton Rouge...

(1) predict: describes police fatality?

0.4

0.8

0.01
The Baton Rouge Police Department confirms that confirms **Alton Sterling**, 37, died during a shooting at the Triple S Food Mart.

... the two officers involved in Tuesday’s shooting of **Alton Sterling** ...

... **Alton Sterling** was a resident of Baton Rouge...

(1) predict: describes police fatality? 0.4

(2) aggregate: add to database? 0.8

Corpus
Model

(1) Predict sentence-level **event** assertions

(2) Aggregate **entity**-level predictions
Model

(1) Predict sentence-level event assertions

(2) Aggregate entity-level predictions

\[ P(z_i = 1|x_i) = \sigma(\theta^T f(x_i)) \]
Model

(1) Predict sentence-level event assertions

(2) Aggregate entity-level predictions

\[ P(z_i = 1|x_i) = \sigma(\theta^T f(x_i)) \]

describes police killing event

sentence text

e.g. logistic regression, convolutional neural network
Model

(1) Predict sentence-level **event** assertions
(2) Aggregate **entity**-level predictions

1. Feature-engineered logistic regression
   - Syntactic dependency paths
   - N-grams
   - POS tags
Model

(1) Predict sentence-level **event** assertions

(2) Aggregate **entity**-level predictions

1. **Feature-engineered logistic regression**
   - Syntactic dependency paths
   - N-grams
   - POS tags

2. **Convolutional neural network**
   - [Kim 2014]
   - Used in other event detection work [e.g. Nguyen and Grishman 2015]
(1) Predict sentence-level event assertions

(2) Aggregate entity-level predictions

The Baton Rouge Police Department confirms that confirms Alton Sterling, 37, died during a shooting at the Triple S Food Mart

... the two officers involved in Tuesday 's shooting of Alton Sterling ...

... Alton Sterling was a resident of Baton Rouge...

\[ p(z \mid x) \]

0.4
0.8
0.01
aggregate
Model

(1) Predict sentence-level **event** assertions

(2) Aggregate **entity**-level predictions

---

The Baton Rouge Police Department confirms that confirms **Alton Sterling**, 37, died during a shooting at the Triple S Food Mart

... the two officers involved in Tuesday 's shooting of **Alton Sterling** ...

... **Alton Sterling** was a resident of Baton Rouge...

---

\[ p(z|x) \]

- max: 0.8
- mean: 0.403
- median: 0.4

aggregate
Model

(1) Predict sentence-level event assertions

(2) Aggregate entity-level predictions

The Baton Rouge Police Department confirms that confirms Alton Sterling, 37, died during a shooting at the Triple S Food Mart

... the two officers involved in Tuesday's shooting of Alton Sterling ...

... Alton Sterling was a resident of Baton Rouge...

\[ p(z \mid x) \]

<table>
<thead>
<tr>
<th>noisy-or</th>
<th>max</th>
<th>mean</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.881</td>
<td>0.8</td>
<td>0.403</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Noisy-Or

\[ \begin{align*}
&x \rightarrow z \\
&x \rightarrow z \\
&x \rightarrow z \\
&\text{OR} \rightarrow y
\end{align*} \]
Noisy-Or

\[ P(y_e = 1 | x_{M(e)}) = 1 - \prod_{i \in M(e)} (1 - P(z_i = 1 | x_i)) \]
Model

(1) Predict sentence-level event assertions

(2) Aggregate entity-level predictions

\[ p(z \mid x) \]

The Baton Rouge Police Department confirms that confirms Alton Sterling, 37, died during a shooting at the Triple S Food Mart

... the two officers involved in Tuesday's shooting of Alton Sterling ...

... Alton Sterling was a resident of Baton Rouge...

<table>
<thead>
<tr>
<th>noisy-or</th>
<th>max</th>
<th>mean</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.881</td>
<td>0.8</td>
<td>0.403</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Outline

1. Motivation and overview
2. Task and data
3. Model
4. Training
5. Evaluation
6. Conclusion
Michael Brown was killed by a white police officer in Ferguson, Mo.

Reporters interviewed the mother of Michael Brown last Sunday.

Katy Perry reacted on Twitter to the most recent police killing.
hand labeling is expensive  
\[\rightarrow\] distant supervision

**Corpus**

**Database**

*Michael Brown* was killed by a white police officer in Ferguson, Mo.

Reporters interviewed the mother of *Michael Brown* last Sunday.

*Katy Perry* reacted on Twitter to the most recent police killing.
Imputing training labels

1. “Hard” labeling

2. “Soft” labeling
Imputing training labels

1. “Hard” labeling
   Distant Supervision Assumption
   [Mintz et al., 2009]

2. “Soft” labeling
Michael Brown was killed by a white police officer in Ferguson, Mo.

Reporters interviewed the mother of Michael Brown last Sunday.

Katy Perry reacted on Twitter to the most recent police killing.
Michael Brown was killed by a white police officer in Ferguson, Mo.

Reporters interviewed the mother of Michael Brown last Sunday.

Katy Perry reacted on Twitter to the most recent police killing.
Michael Brown was killed by a white police officer in Ferguson, Mo.

Reporters interviewed the mother of Michael Brown last Sunday.

Katy Perry reacted on Twitter to the most recent police killing.
Imputing training labels

1. “Hard” labeling

Distant Supervision Assumption
[Mintz et al., 2009]

2. “Soft” labeling

“At least one” assumption
[Bunescu and Mooney 2007]
Michael Brown was killed by a white police officer in Ferguson, Mo.

Reporters interviewed the mother of Michael Brown last Sunday.

Katy Perry reacted on Twitter to the most recent police killing.

(2) "Soft" labeling

Corpus

Database

Eric Garner

Michael Brown
(2) “Soft” labeling

EM Training [Dempster et al. 1977]
(2) “Soft” labeling

EM Training [Dempster et al. 1977]

Initialize with hard distant labels
EM Training [Dempster et al. 1977]

Initialize with hard distant labels

E-Step:
Marginal posterior probability for each $z_i$

\[
q(z_i = 1) = \frac{P(z_i = 1, y_{e_i} = 1 | \mathcal{X}(e_i))}{P(y_{e_i} = 1 | \mathcal{X}(e_i))}
\]

- $y_{e_i} = 1$ identifies sentence $i$ as a police fatality event
- $z_i = 1$ is the entity label
- $\mathcal{X}(e_i)$ is the set of all sentences for the given entity
(2) “Soft” labeling

EM Training \[\text{[Dempster et al. 1977]}\]

Initialize with hard distant labels

E-Step:
Marginal posterior probability for each \(z_i\)

\[
q(z_i = 1) = \frac{P(z_i = 1, y_{e_i} = 1 | x_{M(e_i)})}{P(y_{e_i} = 1 | x_{M(e_i)})}
\]

M-Step:

\[
\max_\theta \sum_i \sum_{z \in \{0,1\}} q(z_i = z) \log P_\theta(z_i = z | x_i).
\]
Outline

1. Motivation and overview
2. Task and data
3. Model
4. Training
5. Evaluation
6. Conclusion
Model results

![Model results graph](image)

**Table 6: Precision, recall, and F1 scores for test data using event extractors SEMAFOR and RPI-JIE under the full R3 system.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Rule Prec.</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEMAFOR</td>
<td>0.011</td>
<td>0.436</td>
<td>0.022</td>
</tr>
<tr>
<td>RPI-JIE</td>
<td>0.016</td>
<td>0.447</td>
<td>0.030</td>
</tr>
<tr>
<td>RPI-JIE (R1)</td>
<td>0.016</td>
<td>0.447</td>
<td>0.030</td>
</tr>
<tr>
<td>RPI-JIE (R2)</td>
<td>0.031</td>
<td>0.162</td>
<td>0.051</td>
</tr>
<tr>
<td>RPI-JIE (R3)</td>
<td>0.172</td>
<td>0.168</td>
<td>0.170</td>
</tr>
</tbody>
</table>

**Figure 4: Precision-recall curves for the given models.**


---

As in Open et al. (2014), we aggregate mention-level predictions to obtain entity-level predictions with a deterministic OR of (R1), (R2), and (R3).

In experiments, we use the ACE 'life/die' event type/subtype with roles 'victim' and 'agent'. For RPI-JIE, we use the ACE 'life/die' event type/subtype with frame elements 'Victim' and 'Killer'. For RPI-JIE under the full R3 system, we use the full event arguments.

**Data upper bound:**

- SEMAFOR R1: 0.011, 0.436, 0.022
- SEMAFOR R2: 0.031, 0.162, 0.051
- SEMAFOR R3: 0.172, 0.168, 0.170

**Parsons et al. (1990):**

- Rule Prec. R1: 0.57
- Recall R1: 0.73
- F1 R1: 0.57

---

We want the system to detect (1) killing events, (2) victims of a killing event, and 'agent'. SEMAFOR defines a token span for every argument in the sentence. RPI-JIE predicts spans as event arguments, while RPI-JIE also predicts spans as event arguments.

Our results indicate that hard-LR, n-gram feats. 0.134 0.257, hard-LR, dep. feats. 0.117 0.229, soft-LR (EM) 0.193 0.316, and soft-CNN (EM) 0.164 0.267 perform better than hard-LR, all feats. 0.142 0.266.

**Table 5: Area under precision-recall curve (AUPRC) and F1 (its maximum value from the PR curve) for entity prediction on the test set.**

Source: Banarescu et al. (2013), filled curves in green (R3); Kingsbury and Palmer (2014), filled curves in blue (R3); Daan et al. (2014), filled curves in red (R3); and Fillmore et al. (2004), filled curves in black (R3).

---

Part of this is due to inherent difficulty of the task, and 'agent'. SEMAFOR only determines R2 and R3, whereas we allow a match on any of an entity's extents. For RPI-JIE, we use the ACE 'life/die' event type/subtype with roles 'victim' and 'agent'. For RPI-JIE under the full R3 system, we use the full event arguments.

---

For each mention, we need an extraction system to handle different trigger constructions like "killed" versus "shot dead." We implement a small progression of these rules to classify the event type as 'kill.'
Model results

<table>
<thead>
<tr>
<th>Model</th>
<th>AUPRC</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data upper bound</td>
<td>0.57</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Table 5: Area under precision-recall curve (AUPRC) and F1 (its maximum value from the PR curve).

Table 6: Precision, recall, and F1 scores for test data using event extractors SEMAFOR and RPI-JIE.

Model R1 0.016 0.447 0.030
Model R2 0.044 0.327 0.078
Model R3 0.098 0.009 0.016

For each mention, we evaluate two freely available, off-the-shelf event extractors SEMAFOR and RPI-JIE to extract event tuples of the form (event type, agent, patient) from the sentence. For SEMAFOR, we use the FrameNet ‘Killing’ frame with frame elements ‘Victim’ and ‘Killer’. For RPI-JIE, we use the ACE ‘life/die’ event type/subtype with roles ‘victim’ and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR tends to be small, but it does sometimes resolve pronouns. For RPI-JIE, we predict spans as event arguments, while RPI-JIE also predicts spans as event arguments. RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. SEMAFOR defines a token span for every argument; RPI-JIE/ACE defines two spans, both a head word and ‘agent’. RPI-JIE refers to the event type as ‘kill.’
Model results

![Model results graph]

<table>
<thead>
<tr>
<th>Model</th>
<th>AUPRC</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>hard-LR, dep. feats.</td>
<td>0.117</td>
<td>0.229</td>
</tr>
<tr>
<td>hard-LR, n-gram feats.</td>
<td>0.134</td>
<td>0.257</td>
</tr>
<tr>
<td>hard-LR, all feats.</td>
<td>0.142</td>
<td>0.266</td>
</tr>
<tr>
<td>Data upper bound</td>
<td>0.57</td>
<td>0.73</td>
</tr>
</tbody>
</table>
We need an extraction system to handle different trigger constructions like "killed" versus "shot dead." We assume such systems are too narrow for our purposes, since event structures in text, but with lighter ontologies where event classes directly correspond with lexical items—including their associated frames—pay off.

We use RPI-JIE to identify instances of gun violence. For each mention of a killed person that contains a police keyword, we want the system to detect (1) killing events, (2) the killed person is the target mention, and (3) the person who killed them is a police officer. We implement a small progression of these rules to classify entities as event arguments and gives each a within-text role.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUPRC</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>hard-LR, dep. feats.</td>
<td>0.117</td>
<td>0.229</td>
</tr>
<tr>
<td>hard-LR, n-gram feats.</td>
<td>0.134</td>
<td>0.257</td>
</tr>
<tr>
<td>hard-LR, all feats.</td>
<td>0.142</td>
<td>0.266</td>
</tr>
<tr>
<td>hard-CNN</td>
<td>0.130</td>
<td>0.252</td>
</tr>
<tr>
<td>Data upper bound</td>
<td>0.57</td>
<td>0.73</td>
</tr>
</tbody>
</table>

As in (Parsons 2016), we aggregate mention-level predictions with a deontic operator (R1) if: (R1) the event type is 'kill.' Part of this is due to inherent difficulty of the task, though all results are relatively poor (Table 5). We suspect a major issue is that these systems heavily rely on their annotated training sets and may have significant performance loss on news, suggesting domain transfer for future work.

Usefulness for practitioners: For practitioners, our model is better than existing methods to extract event arguments. Comparing SEMAFOR R1 and RPI-JIE under the full R3 system performs best, though our task-specific model still outperforms SEMAFOR R1.

Our results indicate that even model AUPRC and F1 are small, but it does sometimes resolve pronouns. For SEMAFOR, we use the FrameNet 'Killing' frame and (3) the person who killed them is a police officer. We implement a small progression of these rules to obtain entity-level predictions with a deontic operator (R1) if: (R1) the event type is 'kill.' Part of this is due to inherent difficulty of the task, though all results are relatively poor (Table 5). We suspect a major issue is that these systems heavily rely on their annotated training sets and may have significant performance loss on news, suggesting domain transfer for future work.
Model results

Table 5: Area under precision-recall curve

<table>
<thead>
<tr>
<th>Model</th>
<th>AUPRC</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>hard-LR, dep. feats.</td>
<td>0.117</td>
<td>0.229</td>
</tr>
<tr>
<td>hard-LR, n-gram feats.</td>
<td>0.134</td>
<td>0.257</td>
</tr>
<tr>
<td>hard-LR, all feats.</td>
<td>0.142</td>
<td>0.266</td>
</tr>
<tr>
<td>hard-CNN</td>
<td>0.130</td>
<td>0.252</td>
</tr>
<tr>
<td>soft-CNN (EM)</td>
<td>0.164</td>
<td>0.267</td>
</tr>
<tr>
<td><strong>soft-LR (EM)</strong></td>
<td><strong>0.193</strong></td>
<td><strong>0.316</strong></td>
</tr>
<tr>
<td>Data upper bound</td>
<td>0.57</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Figure 4: Precision-recall curves for the given models.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUPRC</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>soft-LR (EM)</td>
<td>0.193</td>
<td>0.316</td>
</tr>
<tr>
<td>Data upper bound</td>
<td>0.57</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Usefulness for practitioners:

- The system needs to detect killing events, which are often difficult to distinguish.
- The system predicts event arguments but doesn't handle pronouns.
- The system doesn't explicitly handle pronouns.
- The system faces challenges with messy text and new domains.

Different trigger conditions like "killed" versus "shot dead."
Off-the-shelf event extractors
Off-the-shelf event extractors

SEMAFOR
(trained for FrameNet)
[Das et al. 2014]
Off-the-shelf event extractors

SEMAFOR
(trained for FrameNet)
[Das et al. 2014]

RPI-JIE
(trained for ACE)
[Li and Ji 2014]
Off-the-shelf event extractors

SEMAFOR
(trained for FrameNet)
[Das et al. 2014]

RPI-JIE
(trained for ACE)
[Li and Ji 2014]

Used in gun violence database pipeline
[Pavlick and Callison-Burch 2016]
## Off-the-shelf event extractors

<table>
<thead>
<tr>
<th>Rule</th>
<th>Prec.</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEMAFOR (trained for FrameNet) [Das et al. 2014]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPI-JIE (trained for ACE) [Li and Ji 2014]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Off-the-shelf event extractors

<table>
<thead>
<tr>
<th>Model</th>
<th>Rule</th>
<th>Prec.</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEMAFOR (trained for FrameNet) [Das et al. 2014]</td>
<td>R1</td>
<td>0.011</td>
<td>0.436</td>
<td>0.022</td>
</tr>
<tr>
<td>RPI-JIE (trained for ACE) [Li and Ji 2014]</td>
<td>R1</td>
<td>0.016</td>
<td>0.447</td>
<td>0.030</td>
</tr>
</tbody>
</table>

R1: killing event
## Off-the-shelf event extractors

<table>
<thead>
<tr>
<th></th>
<th>Rule</th>
<th>Prec.</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEMAFOR</td>
<td>R1</td>
<td>0.011</td>
<td>0.436</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>R2</td>
<td>0.031</td>
<td>0.162</td>
<td>0.051</td>
</tr>
<tr>
<td>RPI-JIE</td>
<td>R1</td>
<td>0.016</td>
<td>0.447</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>R2</td>
<td>0.044</td>
<td>0.327</td>
<td>0.078</td>
</tr>
</tbody>
</table>

- **R1**: killing event
- **R2**: R1 and patient = entity
## Off-the-shelf event extractors

<table>
<thead>
<tr>
<th>Rule</th>
<th>Prec.</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.011</td>
<td>0.436</td>
<td>0.022</td>
</tr>
<tr>
<td>R2</td>
<td>0.031</td>
<td>0.162</td>
<td>0.051</td>
</tr>
<tr>
<td>R3</td>
<td>0.098</td>
<td>0.009</td>
<td>0.016</td>
</tr>
<tr>
<td>R1</td>
<td>0.016</td>
<td>0.447</td>
<td>0.030</td>
</tr>
<tr>
<td>R2</td>
<td>0.044</td>
<td>0.327</td>
<td>0.078</td>
</tr>
<tr>
<td>R3</td>
<td>0.172</td>
<td>0.168</td>
<td>0.170</td>
</tr>
</tbody>
</table>

R1: killing event
R2: R1 and patient = entity
R3: R2 and agent = police
Off-the-shelf event extractors

<table>
<thead>
<tr>
<th>Extractor</th>
<th>Rule</th>
<th>Prec.</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEMAFOR (trained for FrameNet)</td>
<td>R1</td>
<td>0.011</td>
<td>0.436</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>R2</td>
<td>0.031</td>
<td>0.162</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>R3</td>
<td>0.098</td>
<td>0.009</td>
<td>0.016</td>
</tr>
<tr>
<td>RPI-JIE (trained for ACE)</td>
<td>R1</td>
<td>0.016</td>
<td>0.447</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>R2</td>
<td>0.044</td>
<td>0.327</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>R3</td>
<td>0.172</td>
<td>0.168</td>
<td>0.170</td>
</tr>
<tr>
<td>soft-LR (EM)</td>
<td></td>
<td></td>
<td></td>
<td>0.316</td>
</tr>
</tbody>
</table>

- **R1**: killing event
- **R2**: R1 and patient = entity
- **R3**: R2 and agent = police

- Data upper bound ($\mu$): 1.0 0.57 0.73
- Soft-LR (EM): 0.172 0.168
- Soft-CNN (EM): 0.164 0.267
- Hard-CNN: 0.130 0.252
- Hard-LR, all feats.: 0.142 0.266
- Hard-LR, n-gram feats.: 0.134 0.257
- Hard-LR, dep. feats.: 0.117 0.229

Table 5: Area under precision-recall curve (AUPRC) and F1 (its maximum value from the PR curve) for entity prediction on the test set.
## Top entities at test time

<table>
<thead>
<tr>
<th>rank</th>
<th>name</th>
<th>positive</th>
<th>analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Keith Scott</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Terence Crutcher</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Alfred Olango</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Deborah Danner</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Carnell Snell</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Kajuan Raye</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Terrence Sterling</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Francisco Serna</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Sam DuBose</td>
<td>false</td>
<td>name mismatch</td>
</tr>
<tr>
<td>10</td>
<td>Michael Vance</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Tyre King</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Joshua Beal</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Trayvon Martin</td>
<td>false</td>
<td>killed, not by police</td>
</tr>
<tr>
<td>14</td>
<td>Mark Duggan</td>
<td>false</td>
<td>non-US</td>
</tr>
<tr>
<td>15</td>
<td>Kirk Figueroa</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Anis Amri</td>
<td>false</td>
<td>non-US</td>
</tr>
<tr>
<td>17</td>
<td>Logan Clarke</td>
<td>false</td>
<td>shot not killed</td>
</tr>
<tr>
<td>18</td>
<td>Craig McDougall</td>
<td>false</td>
<td>non-US</td>
</tr>
<tr>
<td>19</td>
<td>Frank Clark</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Benjamin Marconi</td>
<td>false</td>
<td>name of officer</td>
</tr>
</tbody>
</table>
## Top entities at test time

<table>
<thead>
<tr>
<th>rank</th>
<th>name</th>
<th>positive</th>
<th>analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Sam DuBose</td>
<td>false</td>
<td>name mismatch</td>
</tr>
<tr>
<td>13</td>
<td>Trayvon Martin</td>
<td>false</td>
<td>killed, not by police</td>
</tr>
<tr>
<td>14</td>
<td>Mark Duggan</td>
<td>false</td>
<td>non-US</td>
</tr>
<tr>
<td>16</td>
<td>Anis Amri</td>
<td>false</td>
<td>non-US</td>
</tr>
<tr>
<td>17</td>
<td>Logan Clarke</td>
<td>false</td>
<td>shot not killed</td>
</tr>
<tr>
<td>18</td>
<td>Craig McDougall</td>
<td>false</td>
<td>non-US</td>
</tr>
<tr>
<td>20</td>
<td>Benjamin Marconi</td>
<td>false</td>
<td>name of officer</td>
</tr>
</tbody>
</table>
Outline

1. Motivation and overview
2. Task and data
3. Model
4. Training
5. Evaluation
6. Conclusion
Goal: database update
(1) Walter Scott

- A group prayer is held on April 12, 2015 at the site where Walter Scott was killed by a North Charleston police officer in North Charleston, South Carolina. View photos. A group prayer is held on April 12, 2015 at the site where Walter Scott was killed by a North Charleston police officer in North Charleston, South Carolina. (AFP Photo/JOE RAEDLE) (BUTTON)
dl date 2016-12-06 Doc 2173194_36_4 pred=0.998
- The shooting happened just months after Walter Scott, an unarmed black man, was killed by white police officer Michael Slager when he fled a traffic stop in North Charleston.
dl date 2016-12-16 Doc 2203135_323_0 pred=0.991
- A man walks past the lot where Walter Scott was killed by a North Charleston police officer Saturday after a traffic stop in North Charleston, S.C., Thursday, April 9, 2015.
dl date 2016-12-06 Doc 2172211_194_0 pred=0.99

(2) Keith Scott

- News of the jury’s failure to reach a verdict came just a few days after a prosecutor in Charlotte, N.C., announced no charges would be filed against a police officer in the September shooting of Keith Scott, an African American man whose death inspired violent protests in North Carolina.
dl date 2016-12-02 Doc 2163436_27_0 pred=0.97
- Nation/World Keith Lamont Scott, pictured at right in a photo released by his family, was fatally shot by police in Charlotte, North Carolina on Sept. 20, 2016.
dl date 2016-12-02 Doc 2163074_100_0 pred=0.951
- People march in Charlotte, N.C., on Sept. 23 to protest the fatal police shooting of Keith Lamont Scott.
dl date 2016-12-20 Doc 2213883_298_0 pred=0.947

(3) Alton Sterling

- Hundreds of miles away, protesters marched outside a convenience store in Baton Rouge, Louisiana, where Alton Sterling was fatally shot Tuesday while police tackled him in a parking lot.
dl date 2016-12-29 Doc 2241447_83_0 pred=0.995
- [rtsh3xr.jpg?quality=80&strip=all&w=50] Ieshia L. Evans, a demonstrator protesting the shooting death of Alton Sterling is detained by law enforcement near the headquarters of the Baton Rouge Police Department in Baton Rouge, Louisiana, on July 9.
dl date 2016-12-27 Doc 2234040_59_0 pred=0.995
- old Alton Sterling, a black man killed by white Baton Rouge officers after a confrontation at a convenience store.
dl date 2016-12-27 Doc 2235302_71_0 pred=0.995
Future Work

- Other model architectures (e.g. LSTMs)
- Other domains for database update problem
- Extract additional event information
- Build interactive interface for practitioners
Contributions

• Distant supervision approach much cheaper
• Public data for the social good
• New NLP task, released data publicly
• Progress towards fully-automatic system
Thanks!

Code and data:
http://slanglab.cs.umass.edu/PoliceKillingsExtraction/

Acknowledgements:
• Amazon Web Services (AWS) Cloud Credits for Research program.
• D. Brian Burghart for advice on police fatalities tracking