

A huge volume of textual data is produced every day by traditional and social media. It is not possible to keep track of the many events that are mentioned, nor to recognize known patterns of events. This project uses and develops natural language processing (machine learning) techniques to automatically extract events from text. In future work we hope to relate these events to known patterns called scripts.

For this work, we pursued 3 strategies for extracting events from textual data:

1. Event extraction from sentences
2. Macro-event extraction from documents
3. Minimally supervised extraction of events from Twitter

Trigger Identification:

Bi-directional LSTM + CRF

- Concatenate character embeddings with pre-trained word embeddings to get vectors representing each word
- Run a bi-LSTM over the sequence of word vectors to obtain the two hidden states
- Use a CRF to find the sequence with the highest probability
- Comparison of state-of-the-art JointEventEntity strategy with our Bi-directional LSTM + CRF strategy

Results	P	R	F1
State-of-the-art	61.4	71.0	65.9
Our strategy	66.7	66.0	66.4

Trigger Replicated:

- Train SVMs to label triggers as belonging to one of 33 subtypes (e.g., attack, sentence, convict, etc.)
- However, with perfect trigger identification, F1 for trigger classification is 80.0.

Results	P	R	F1
JointEventEntity (replicated)	61.5	71.2	66.0
SVM	81.6	54.5	65.3

Bottom line: Neither technique should be trusted to perform accurate trigger identification

Macro-Event Extraction

Document-level analysis – extract only the most important events in a text

ATTACK Macro-Event

Perpetrator	Michael Dunn
Victim - Dead	Jordan Davis
Victim - Injured	None
Time	Nov. 23, 2012
Location	Jacksonville

ARREST Macro-Event

Arrestee	Michael Dunn
Time	(Unknown)
Location	(Unknown)

TRIAL Macro-Event

Defendant	Michael Dunn
Crime	Murder
Verdict	Guilty
Sentence	Prison
Time	Friday
Location	Florida

We are developing two novel ML-based algorithms for solving this problem

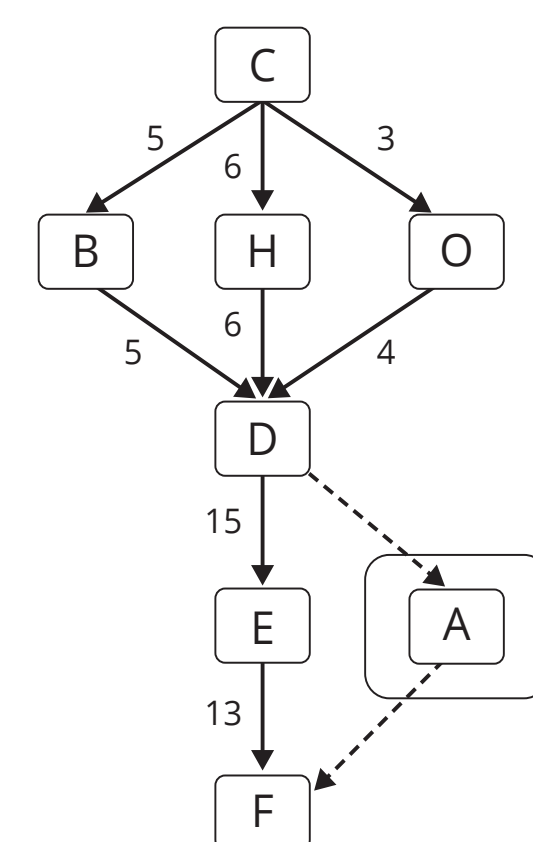
- A structured prediction model based on Learning to Search
- A deep neural network based on machine comprehension with no reliance on target domain training data

Preliminary results on the attack and elections domains show significantly improved performance against baseline methods

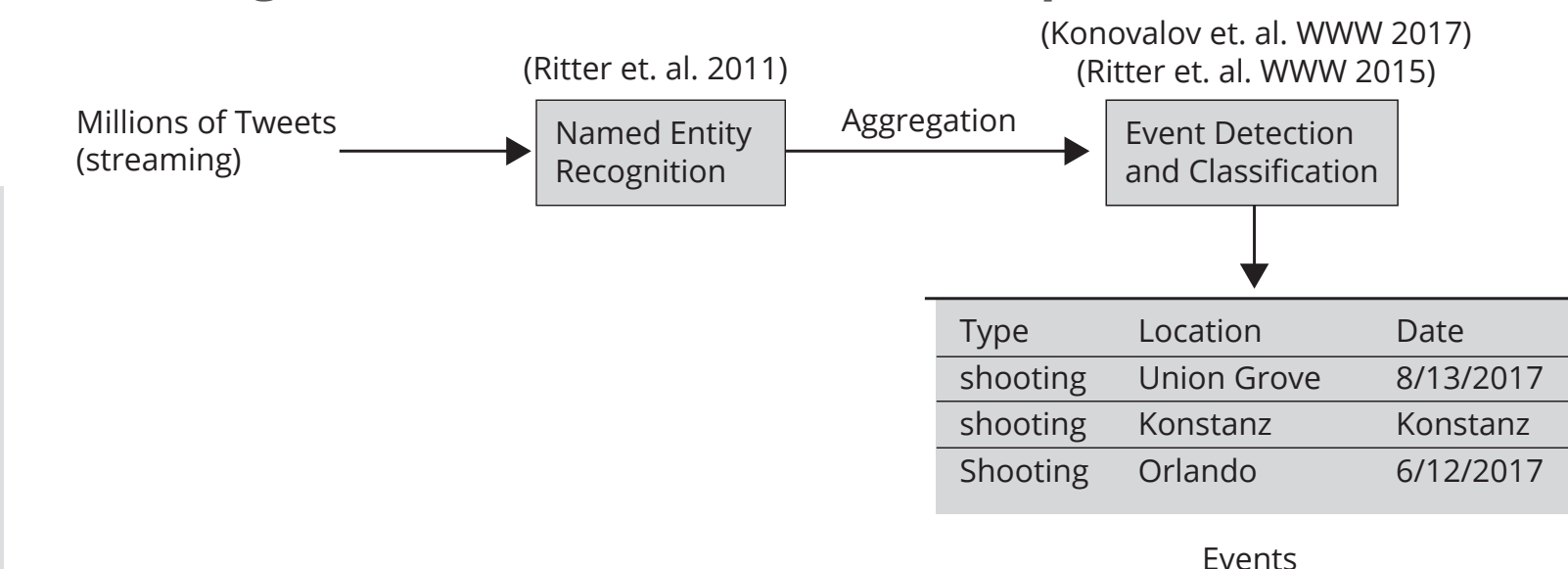
Currently gathering annotated data via Mechanical Turk

Next Step: Recognize and create scripts

- Better representation of the world
 - Finer-grain event representation (actor, location, etc.)
 - Probability distributions over the possible arguments
 - A refined similarity metric that can reflect the new representation
 - More robust script manipulation
 - Better script addition (avoiding adding rare instances, etc.)
 - Splitting / pruning existing scripts
- Use of macro-event knowledge for inferring constraints



Learning to Extract Events Minimal Supervision



Gun Violence Archive

Type: Shooting
Location: Orlando
Date: 6/12/2017

Time

My prayers sent out to all the victims in the Orlando shooting and a special prayer to my friend Mario

Orlando Shooting: 50 Killed, Shooter Pledged ISIS Allegiance

victims in the X shooting

X Shooting: 50 Killed

Germany Konstanz nightclub shooting: 2 killed, 4 seriously injured. Shoo...
youtu.be/47EDd2U6s6A via @YouTube

Type: Shooting
Location: Union Grove, WI
Date: 8/13/2017

Type: Shooting
Victim: Konstanz
Date: 7/30/2017

Twitter Shooting Events

Date: Aug-17-2017

Entity	Representative Tweet	Count
sherwaygardens	RT @680NEWS: MORE: Police are searching for two suspects after a shooting outside a coffee shop at Sherway Gardens in Etobicoke...	20
denver	1 hurt in shooting at Denver apartment complex #DIA #VCinvesting #COS #jobs https://t.co/sm230QmBvb	31
toronto	RT @AIBoeNEWS: #BREAKING: UPDATE: One victim in the Toronto shooting was rushed to a nearby trauma center for his injuries https://t.co/JJ0...	35
newark	From https://t.co/2NaY2S3QM Shooting - Newark, NJ: U/D PD looking for 1 tall b/m wearing dreads, white shirt and blue jeans... #BreakingNews	9
sheboygan	Sheboygan police investigate shooting near 19th and Broadway https://t.co/bq0SbGg32Q	7
nj	From https://t.co/2NaY2S3QM Shooting - Newark, NJ: U/D PD looking for 1 tall b/m wearing dreads, white shirt and blue jeans... #BreakingNews	12
democrat	RT @gerfingerpoken: The Russians Hacked My Homework: Democrat's Lame Excuse - American Thinker https://t.co/YxPAXuEu72 #PUNET 999 https://t.co/...	12
phoenix	RT @GMAZMargaret: #Breaking We're learning about an officer-involved shooting in Phoenix. @GibbyParra on scene.	12
charlotte	Keith Lamont Scott's Family Releases Video in Fatal Shooting by Charlotte Police KTLA https://t.co/kl8OR9587E	7
aurora	RT @HeartedSayings: I'm saddened to learn about the mass shooting in Aurora, CO. My prayers go out to all the broken hearts. Life is more p...	8

Copyright 2017 Carnegie Mellon University. All Rights Reserved.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

Internal use:* Permission to reproduce this material and to prepare derivative works from this material for internal use is granted, provided the copyright and "No Warranty" statements are included with all reproductions and derivative works.

External use:* This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other external and/or commercial use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

* These restrictions do not apply to U.S. government entities.

Carnegie Mellon® is registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM17-0731
Events, Relationships, and Script Learning