Open-Domain Event Detection using Distant Supervision

Jun Araki and Teruko Mitamura
Language Technologies Institute, School of Computer Science, Carnegie Mellon University

Contributions

- We introduce **a new paradigm of open-domain event detection**. We target events in unrestricted domains with wider coverage than prior work.
- Our **distant supervision method** is able to generate high-quality training data. Despite no direct supervision, the distantly supervised model outperforms supervised models in both in-domain and out-domain settings.
- We **release the new corpus** of human-annotated events in 10 different domains such as geology and economics.

Introduction

- Events are a key component for natural language understanding.
- **Goal**: Detecting all kinds of events regardless of domains.
- **Motivation**: We need automatic event identification techniques with larger, wider, and more consistent coverage in order to advance natural language applications such as open-domain question answering (Sauri et al., 2005; Pradhan et al., 2007).

Research Problems

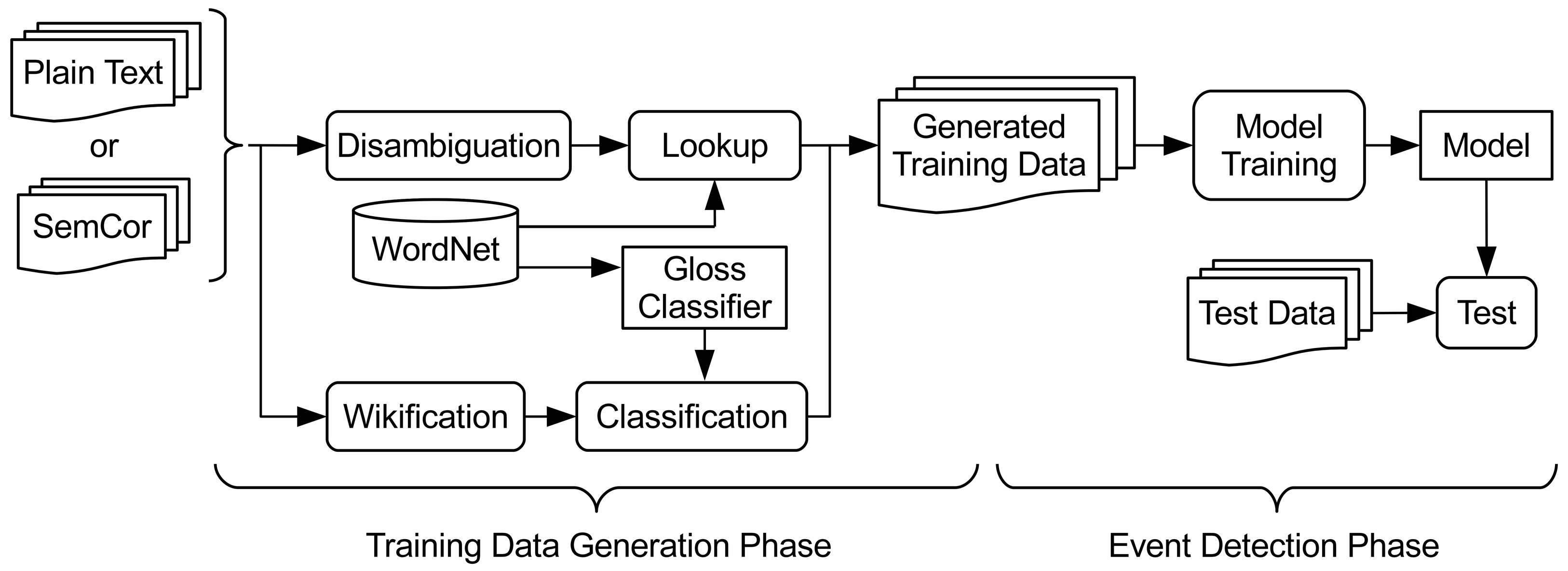
- Limited coverage of events
  - Most work focuses limited (closed-domain) event types, e.g., MUC, ACE, TAC KBP, GENIA, BioNLP, and ProcessBank.
  - Some work has conceptually different focuses, e.g., PropBank, NomBank, and FrameNet.
  - Other work focuses on limited syntactic types, e.g., OntoNotes, TimeML, ECB+, and RED.
- Lack of training data
  - Human annotation of events in the open domain is expensive.

Definition of events

We use two notions from semantic and syntactic perspectives:

- Semantic perspective: **Eventualities** (Bach, 1986)
  - states: notions that are durative and changeless, e.g., want, own, love, resemble.
  - processes: notions that are durative and do not have any explicit goals, e.g., walking, sleeping, raining.
  - actions: notions that have explicit goals or are momentaneous happenings, e.g., build, walk to Pittsburgh, recognize, arrive, clap.
- Syntactic perspective: **Event nuggets** (Mitamura et al., 2015)
  - A semantically meaningful unit that expresses an event.
  - Can be either a single word (verb, noun, or adjective) or a phrase which is continuous or discontinuous.

Distantly Supervised Event Detection

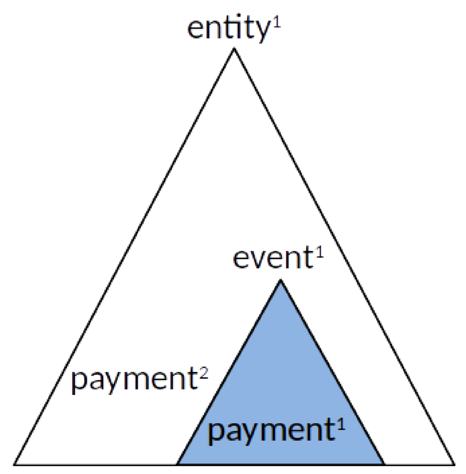


Assumption: Semantically adequate correspondence between components of eventualities and WordNet senses.

- state_n^2: the way something is with respect to its main attributes
- $\mathsf{process}_n^6$: a sustained phenomenon or one marked by gradual changes through a series of states
- event_n^1: something that happens at a given place and time

Examples (**events** vs. *non-events*):

(1) His **payment** was late.
(2) His *payment* was \$10.
(3) Snipers were **picking** them **off**.
(4) He **picked** an apple off the tree.



Enhancements with Wikipedia

- Problem: WordNet often lacks current terminology and proper nouns.

(5) Property damage by **Hurricane Katrina** around \$108 billion.

- Idea: Leverage the first sentence of a Wikipedia article.
  - "Hurricane Katrina was an extremely destructive and deadly tropical **cyclone** that is tied with Hurricane Harvey of 2017 as the costliest hurricane on record." (the underline portion: a *Wikipedia gloss*)
- Methods:

(A) Heuristics-based: HeadLookup
  - Get the syntactic head of a Wikipedia gloss.
  - Look up the head's sense in WordNet.

(B) Learning-based: Gloss classification
  - Collect gloss dataset $D = D_+ \cup D_-$ automatically from WordNet.
    - $D_+ = \{$gloss whose sense is under state_n^2, $\mathsf{process}_n^6$, or $\mathsf{event}_n^1\}$
    - $D_- = \{$all the other glosses of WordNet nouns$\}$
  - Train binary classifiers on D.

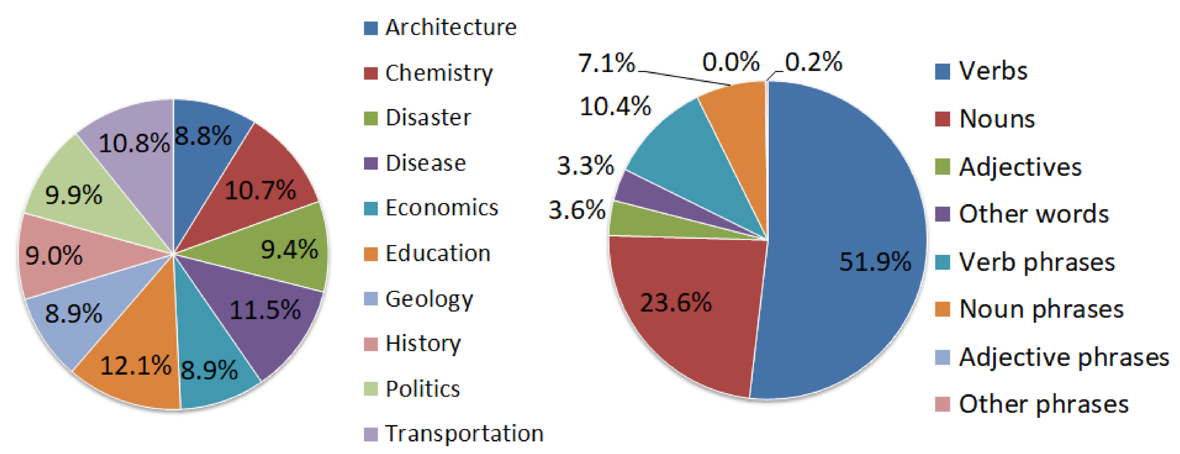
Learning for event detection

- Generate training data from SemCor (Miller et al., 1993).
- Formalize event detection as a sequence labeling problem.
  - Train a BLSTM on the generated training data by minimizing cross-entropy loss with a tagging scheme {B, I, DB, DI, O}.

Results and Future Work

Dataset: Open-domain event corpus (SW100)

- Manually annotated 100 articles in Simple English Wikipedia.
- 5397 event nuggets in 10 different domains.
- Inter-annotator agreement (average of pairwise F1 scores): 80.7% (strict match) and 90.3% (partial match).



Results

- Gloss classifiers: Bag-of-words model with logistic regression, deep average network (Iyyer et al., 2015), BLSTM with (or without) self-attention (Lin et al., 2017)
- Rule-based event detectors: All single-word main verbs (VERB), all predicates extracted by PathLSTM (Roth and Lapata, 2016), our WordNet-based algorithm (RULE)

Model	WordNet	Wikipedia
HeadLookup	77.80	73.50
BoW-LR	79.50	73.00
DAN	83.15	64.00
GC-BLSTM	90.10	80.00
GC-BLSTM-Attn	**91.65****	**85.00***

Table: Accuracy of gloss classification (*: $p < 0.05$; **: $p < 0.005$).

Model	Strict			Partial		
	P	R	F1	P	R	F1
VERB (Baseline)	79.5	51.7	62.7	95.4	62.0	75.2
PRED (Baseline)	55.1	62.4	58.5	67.6	76.6	71.8
RULE	80.1	77.0	78.5	89.0	85.5	87.2
RULE-WP-HL	80.5	77.5	79.0	88.6	85.3	86.9
RULE-WP-GC	80.8	77.7	**79.2**	89.1	85.7	**87.3**

Table: Performance of rule-based event detection on SW100.

- Performance of our distantly supervised model
  - The model performs better with larger synthesized training data.
  - The model outperforms supervised models.

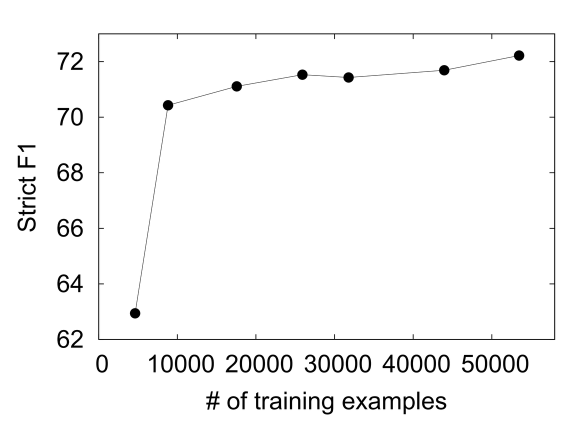


Figure: Performance of distantly supervised event detection on SW100.

Setting	Model	S-F1	P-F1
In-domain	BLSTM	73.8	85.9
	DS-BLSTM	**76.1**	**88.0**
Out-domain	BLSTM	67.9	82.8
	DS-BLSTM	**71.3**	**86.6**

Table: Results of event detection (S-F1: Strict F1, P-F1: Partial F1).

Future Work

- Conduct experiments on normal English text, e.g., newspaper articles.
- Event coreference resolution to detect eventive pronouns and demonstrative determiners.

Acknowledgements

This publication was partly made possible by grant NPRP-08-1337-1-243 from the Qatar National Research Fund (a member of the Qatar Foundation).