Joint Event Trigger Identification and Event Coreference Resolution with Structured Perceptron

Jun Araki and Teruko Mitamura

Language Technologies Institute
School of Computer Science
Carnegie Mellon University

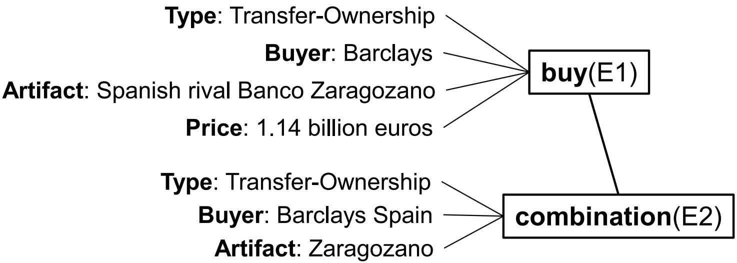
September 21, 2015

Carnegie Mellon University
Language Technologies Institute

Semantic and discourse aspects of events

- Events ⇒ who did what to whom where and when
- Event coreferences ⇒ discourse connections to form a coherent story

> British bank Barclays agreed to **buy**(E1) Spanish rival Banco Zaragozano for 1.14 billion euros. The **combination**(E2) of the banking operations of Barclays Spain and Zaragozano will bring together two complementary businesses.

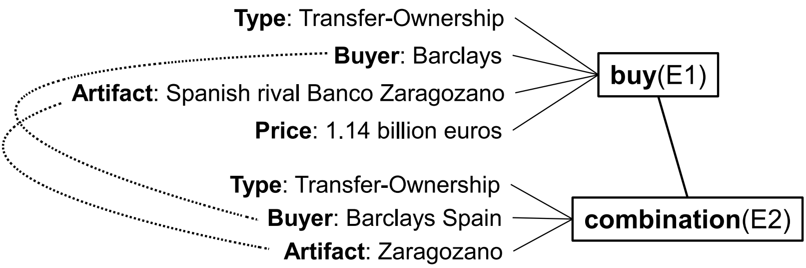


- Many NLP applications:
  - Question answering [Bikel+ 2008; Berant+ 2014]
  - Text summarization [Li+ 2006] etc.

Semantic and discourse aspects of events

- Events ⇒ who did what to whom where and when
- Event coreferences ⇒ discourse connections to form a coherent story

British bank Barclays agreed to **buy**(E1) Spanish rival Banco Zaragozano for 1.14 billion euros. The **combination**(E2) of the banking operations of Barclays Spain and Zaragozano will bring together two complementary businesses.

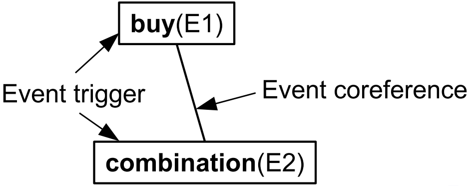


- Many NLP applications:
  - Question answering [Bikel+ 2008; Berant+ 2014]
  - Text summarization [Li+ 2006] etc.

Terminology

- We follow the definitions in the ProcessBank corpus [Berant+ 2014]

Term	Definition
Event	An abstract representation of a change of state, independent from particular texts
Event trigger	Main word(s) in text, typically a verb or a noun that most clearly expresses an event
Event arguments	Participants or attributes in text, typically nouns, that are involved in an event
Event mention	A clause in text that describes an event, and includes both a trigger and arguments
Event coreference	A linguistic phenomenon that two event mentions refer to the same event

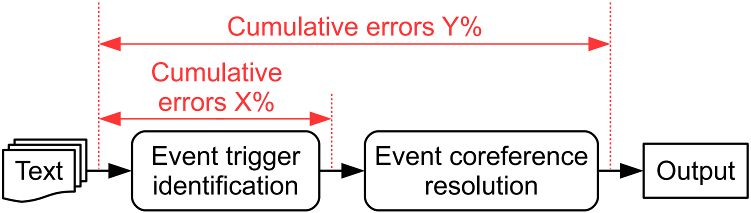


Research problem

- Event extraction and event coreference resolution have been addressed separately
- Some event triggers are relatively difficult to be identified

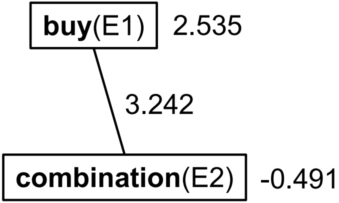
British bank Barclays agreed to **buy**(E1) Spanish rival Banco Zaragozano for 1.14 billion euros. The **combination**(E2) of the banking operations of Barclays Spain and Zaragozano will bring together two complementary businesses.

- Pipeline models propagate errors $\Rightarrow$ normally $Y > X$



Joint model with event graph learning

- We formalize event trigger identification and event coreference resolution as a problem of document-level **joint structured learning**



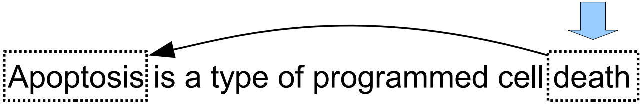
- x: input document
- y: event graph associated with x
  - Node $v \in V(y)$: event trigger
  - Edge $e \in E(y)$: event coreference link
- Node- and edge-factored scoring:

$$score(y) = \sum_{v \in V(y)} score(v) + \sum_{e \in E(y)} score(e)$$
$$= \sum_{v \in V(y)} \mathbf{w} \cdot \Phi(v) + \sum_{e \in E(y)} \mathbf{w} \cdot \Phi(e)$$

- Employ averaged perceptron [Collins 2002] for training
- Use 27 feature templates with a range of tools for feature extraction

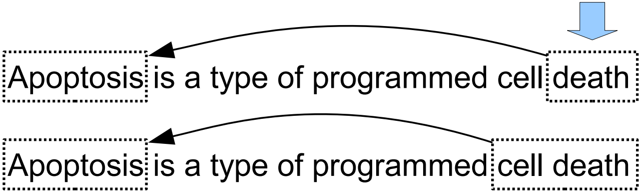
Our joint decoding

- Goal: output the best event graph $\hat{y}$ that maximizes $score(y)$
- Key idea: combine the following with multiple-beam search
  - **Segment-based decoding** [Zhang+ 2008a]
    - Uses previous beam states to form segments from previous positions
    - Computes the k-best partial structures (event subgraphs)
  - **Best-first clustering** [Ng+ 2002]
    - Selects the most likely antecedent for each trigger



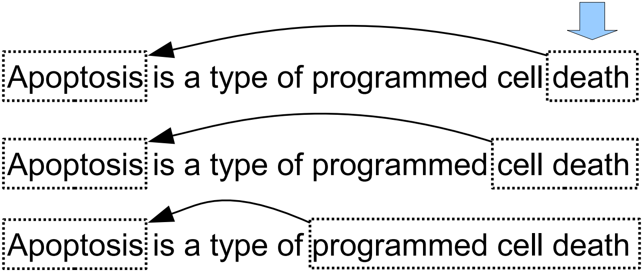
Our joint decoding

- Goal: output the best event graph $\hat{y}$ that maximizes $score(y)$
- Key idea: combine the following with multiple-beam search
  - **Segment-based decoding** [Zhang+ 2008a]
    - Uses previous beam states to form segments from previous positions
    - Computes the k-best partial structures (event subgraphs)
  - **Best-first clustering** [Ng+ 2002]
    - Selects the most likely antecedent for each trigger



Our joint decoding

- Goal: output the best event graph $\hat{y}$ that maximizes $score(y)$
- Key idea: combine the following with multiple-beam search
  - **Segment-based decoding** [Zhang+ 2008a]
    - Uses previous beam states to form segments from previous positions
    - Computes the k-best partial structures (event subgraphs)
  - **Best-first clustering** [Ng+ 2002]
    - Selects the most likely antecedent for each trigger



Other joint decoding which did not work well

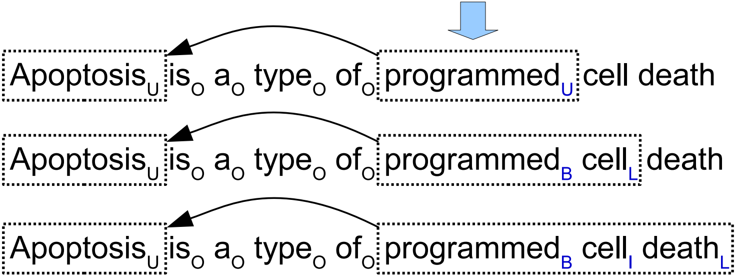
- Some initial tries (alternative approaches):
  - Token-level sequential labeling with `BILOU` scheme
    - Event coreference can be explored only from complete assignments
    - This makes token-level sequential labeling complicated

Apoptosis$_U$ is$_O$ a$_O$ type$_O$ of$_O$ programmed$_U$ cell death

  - Recall-oriented pre-filtering of event trigger candidates
    - Gained 97% recall $\Rightarrow$ 12,400 false positives
    - This makes it difficult to learn event triggers

Other joint decoding which did not work well

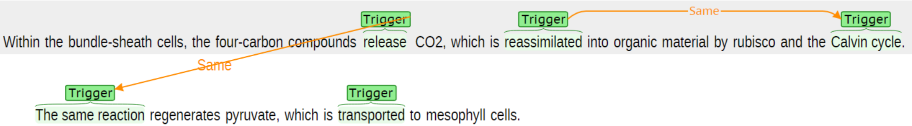
- Some initial tries (alternative approaches):
  - Token-level sequential labeling with `BILOU` scheme
    - Event coreference can be explored only from complete assignments
    - This makes token-level sequential labeling complicated



  - Recall-oriented pre-filtering of event trigger candidates
    - Gained 97% recall ⇒ 12,400 false positives
    - This makes it difficult to learn event triggers

Experimental settings (1/2): ProcessBank corpus

- 200 paragraphs from a textbook in biology



- Event coreference is annotated as a link
- 13.4% of event triggers comprise multiple tokens
- Corpus statistics:

	Train	Dev	Test	Total
# of paragraphs	120	30	50	200
# of event triggers	823	224	356	1403
# of event coreferences	73	28	30	131

Experimental settings (2/2)

- Our baseline
  - Two-stage pipelined model using averaged perceptron
    - 1st stage: event trigger identification
    - 2nd stage: event coreference resolution
    - Same parameters and feature templates as the joint model
- Parameters
  - Number of iterations $T = 20$
    - 20-iteration training almost reached convergence
  - Maximum length of an event trigger $l_{max} = 6$ tokens
    - Specifies how far one can go back in the joint decoding
    - The longest event trigger has 6 tokens in the corpus
  - Beam size $k = 1$
    - A larger beam size did not improve the performance
    - This seems to be due to the small size of dev data

Experimental results

- Evaluation using a reference scorer [Pradhan+ 2014; Luo+ 2014]
- Results of event trigger identification

System	Recall	Precision	F1
Baseline (1st stage)	57.02	64.85	60.68
Joint	55.89	65.24	60.21

- Results of event coreference resolution

	MUC			B^3			$CEAF_m$		
System	R	P	F1	R	P	F1	R	P	F1
Baseline (2nd stage)	26.66	19.51	22.53	55.47	58.64	57.01	53.08	60.38	56.50
Joint	20.00	37.50	**26.08**	53.37	63.36	**57.93**	53.93	62.95	**58.09**

	$CEAF_e$			BLANC			CoNLL
System	R	P	F1	R	P	F1	F1
Baseline (2nd stage)	52.68	63.14	57.44	30.13	25.10	25.05	45.66
Joint	55.06	62.11	**58.38**	27.51	38.43	**31.91**	**47.45**

Observations

- Event coreference resolution
  - The joint model outperforms the baseline
    - Precision $\nearrow \Leftarrow$ false positives $\searrow$
    - Explores a larger number of false positives in its search process
    - Learns to penalize false positives more adequately
- Event trigger identification
  - The joint model does not outperform the baseline
    - This seems to be due to the small size of the corpus
- Some error cases
  - Difficult in the both tasks

When the cell is stimulated, gated channels open that facilitate Na+ **diffusion**(E5). Sodium ions then **"fall"**(E6) down their electrochemical gradient, . . .

The next seven steps **decompose**(E7) the citrate back to oxaloacetate. It is this **regeneration**(E8) of oxaloacetate that makes this process a cycle.

Related work

- **Event extraction**
  - Pipelined approaches for event triggers and arguments [Ji+ 2008; Liao+ 2010; Hong+ 2011]
  - Approaches to joint dependencies [Poon+ 2010; McClosky+ 2011; Riedel+ 2011; Li+ 2013; Venugopal+ 2014]
- **Event coreference resolution**
  - As a starting point, most work uses event triggers from:
    - Human annotation in a corpus [Bejan+ 2014; Liu+ 2014]
    - Output of an event extraction system [Lee+ 2012]
  - Joint learning for event arguments and coreferences [Berant+ 2014]
- **Joint structured learning** in NLP
  - Idea: capturing interactions between two relevant tasks via structure
    - Word segmentation and POS tagging [Zhang+ 2008b]
    - POS tagging and dependency parsing [Bohnet+ 2012]
    - Dependency parsing and semantic role labeling [Johansson+ 2008]
    - Extraction of event triggers and arguments [Li+ 2013]
    - Extraction of entity mentions and relations [Li+ 2014]

Conclusion and future work

- **Conclusion**
  - The first work that solves event trigger identification and event coreference resolution simultaneously
    - Combines the segment-based decoding and best-first clustering
  - The proposed model outperformed a pipelined model in event coreference resolution
- **Future work**
  - Use larger corpora while reducing training time
  - Incorporate other components of events
    - Event types, event arguments, and other relations
  - Neural network based approaches to the joint dependencies