Nathanael Chambers and Dan Jurafsky (2008): Unsupervised Learning of Narrative Event Chains

Feraena Bibyna

January 6, 2014

Outline

Narrative Event Chains

- Partially ordered set of events centered around a common **protagonist**
- Related to structured sequences of participants and events (**scripts**)
- Can be filled in and instantiated to draw inferences

Narrative Chains Induction

_ accused X	W joined _
X claimed _	W served _
X argued	W oversaw _
_ dismissed X	W resigned

...X denied _? _ replaces W?

This Paper

- Narrative chains offer the structure and power to directly infer new subevents by providing background knowledge
- This paper addresses the three task of chain induction:
  - narrative event induction
  - temporal ordering of events
  - structured selection

Steps to narrative induction

1. use entity based model for learning narrative relation
2. order events in the same narrative chains
3. cluster and prune the space to create discrete sets of narrative chains

Narrative Chain Model

- Entity is in focus through a sequence of sentences
  - Centering (Grosz et al., 1995)
  - Barzilay and Lapata, 2005
- A narrative chain...
  - has a central actor (protagonist)
  - is structured by the protagonist's grammatical role
  - consists of narrative events that are ordered by some theory of time
- **Task**: learn event that constitute narrative chains

Formally...

Narrative event ($e_1 : (event, dependency)$)

Tuple of an event (verb) and its participants, represented as typed dependencies

Narrative Chains

Partially ordered set of narrative events $e_1, e_2, ..., e_n$ where n is the size of the chain, and a relation $B(e_i, e_j)$ that is true if narrative event e_1 occurs strictly before e_j in time

Narrative Coherence

Assumption

"Verbs sharing coreferring arguments are semantically connected by virtue of narrative discourse structure"

Captures grammatical constraints, e.g. the object of *push* is the subject of *fall*

Learning Narrative Relations

Learn basic information about narrative chains (protagonist and subevents, but not their ordering)

1. Extract pairwise relation between events unsupervised
2. Build global narrative score

Pairwise Relation between Events

A distributional score based on how often two events share grammatical arguments (using PMI) is used.

$$pmi(e(w,d), e(v,g)) = \log \frac{P(e(w,d), e(v,g))}{P(e(w,d))P(e(v,g))}$$

...where the numerator is defined by:

$$P(e(w,d), e(v,g)) = \frac{C(e(w,d), e(v,g))}{\sum_{x,y}\sum_{d,f} C(e(x,d), e(y,f))}$$

...where $C(e(x,d), e(y,f))$ is the number of times the two events has a coreferring entity filling values of dependencies d and f

Pairwise Relation between Events (Cont'd)

- Adopt 'discount score' to penalize low occurring words (Pantel and Ravichandran, 2004)
- Experimented with t-test.
  - PMI outperforms t-test by itself and when interpolated together using various mixture weights.

Global Narrative Score

- A global narrative score can be built such that all events provide feedback on the event in question
- Given all narrative events in a document, we can find the next most likely event to occur by maximizing:

$$\max_{j:0<j<m} \sum_{i=0}^{n} pmi(e_i, f_j)$$

- A ranked list of guesses can be built from this summation
- The more events in the chain, the more informed the ranked output

Global Narrative Score (Cont'd)

Known events:

(pleaded subj), (admits subj), (convicted obj)

Likely Events:

sentenced obj	0.89	indicted obj	0.74
paroled obj	0.76	fined obj	0.73
fired obj	0.75	denied subj	0.73

Evaluation Metric: Narrative Cloze

- The **cloze task** (Taylor, 1953)

  I forgot to ___ the waitress for the good service

- **Narrative cloze** requires narrative knowledge to solve
  1. McCann threw two interceptions early
     - **(threw subject)**
  2. Toledo pulled McCann aside and told him he'd start
     - **(pulled object)**
     - **(told object)**
     - **(start subject)**
  3. McCann quickly completed his first two passes
     - **(completed object)**

Narrative Cloze Experiment

- **Training data**: years 1994-2004 Gigaword Corpus (Graff, 2002) (1,007,227 documents)
  - parsed into dependency graphs with Stanford Parser (de Marneffe et al., 2006)
  - resolved entity mentions with OpenNLP
- 740 cloze tests (69 narratives with 740 events)
- Position of the correct event is averaged over all 740 tests for the final score
- Unseen event is penalized by setting their ranked position to the length of guess list

Narrative Cloze Experiment (Cont'd)

Known events:

(pleaded subj), (admits subj), (convicted obj)

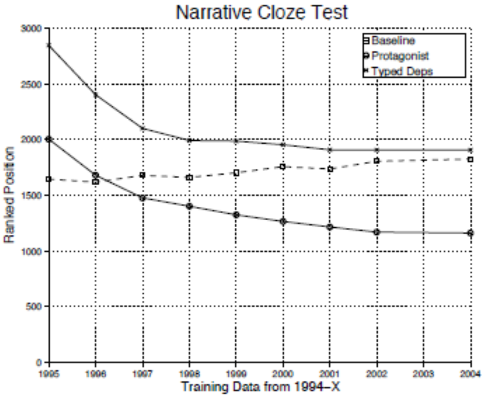
Likely Events:

sentenced obj	0.89	indicted obj	0.74
paroled obj	0.76	fined obj	0.73
fired obj	0.75	denied subj	0.73

Baseline

- Learn relatedness strictly based upon verb
- Performing very poorly due to the huge amount of data involved in counting all possible verb pairs
  - all pairs occurring less than 10 times in training data is removed
- Can't directly compare to narrative approach due to the lack of typed dependency
  - Modified the narrative model to ignore typed dependencies, but still count events with shared arguments
  - Calculate PMI across verbs that shared arguments

Result


Narrative Cloze Test
Baseline
Protagonist
Typed Deps
Ranked Position
3000
2500
2000
1500
1000
500
0
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
Training Data from 1994–X

Temporal Relationship Between Two Events

- Many works on temporal relationship between two items are trained on Timebank Corpus (Pustejovsky et al., 2003)
- Currently highest performing: Chambers et al., 2007
  - 59.4% accuracy
  - Relationships: *before, immediately-before, included-by, simultaneous, begins, ends*
  - For this purpose: merge *immediately-before* with *before*, merge the rest as *other*
    - 72.1% accuracy on Timebank

Temporal Relationship Between Two Events (Cont'd)

Two stage machine learning architecture

1. Uses supervised machine learning to label temporal attribute of events
   - relies on features such as neighboring POS tag, neighboring auxiliaries and modals, WordNet synsets
   - uses SVM
2. Uses classification from (1) combined with other linguistic features to classify temporal relationship between two events

Training a Temporal Classifier

- Many potential event pairs in Timebank are unlabeled: *none* relation, *overlap* relation, or not tagged at all
  - All untagged relation is considered as *other*, and experimented including none, half, and all of them in training
- **Transitivity**: **if** run BEFORE fall **and** fall BEFORE injured **then** run BEFORE injured
- But... the expanded version performed worse than the original!
  - Inconsistent labeling
  - Results are from training *without* transitivity

Temporal Classfier in Narrative Chains

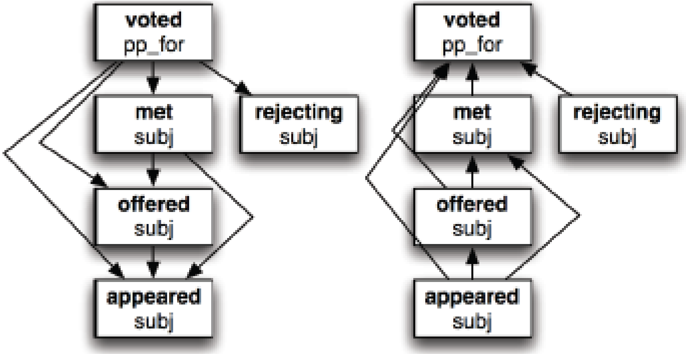
- Gigaword corpus is classified in two stages
  1. for temporal features on each events → trained on Timebank
  2. between all pairs of events that share arguments
- Count the resulting number of labeled *before* relation between each verb/dependency pair
- Confidence of two generic events A and B can be measured by comparing how many *before* labels have been seen versus their inverted order B and A

Temporal Evaluation

Evaluate temporal decision within a coherence framework

- test set uses actual narrative chains from documents, hand labeled for partial ordering
- evaluate coherence of true chains against random ordering
- Evaluation data: 69 documents used in test set for learning narrative relations

A Narrative Chain and Its Reverse Order



Coherence Score

Sum of all relations that the classified corpus agree with, weighted by how certain we are

$$\sum_{E:x,y} \begin{cases} \log(D(x,y)), & \text{if } x\beta y \text{ and } B(x,y) > B(y,x) \\ -\log(D(x,y)), & \text{if } x\beta y \text{ and } B(y,x) > B(x,y) \\ -\log(D(x,y)), & \text{if } !x\beta y \& !y\beta x \\ \& D(x,y) > 0 & \\ 0, & \text{otherwise} \end{cases}$$

Result

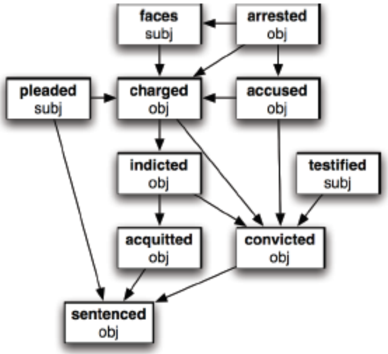
	All	**$\geq$ 6**	**$\geq$ 10**
correct	8086 75%	7603 78%	6307 89%
incorrect	1738	1493	619
tie	931	627	160

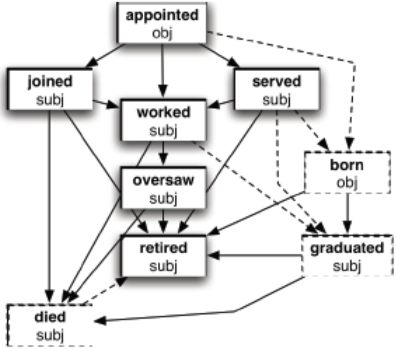
- Accuracy improves the larger the ordered narrative chains
- Highest accuracy on documents whose chains most progressed through time
- Training without *none* relations resulted in high recall

Discrete Narrative Event Chains

- Discrete sets have the drawback of shutting out unseen and unlikely events from consideration
- Easily achievable by using PMI scores in an agglomerative clustering algorithm, then apply ordering relations

Discrete Narrative Events Chains (Cont'd)





Conclusions

- Narrative Event Induction
  - Narrative relations show improvement offer baseline
  - Potential implementation in many area of NLP
  - New measure of similarity, using the protagonist to extract list of related events
  - May not need presorted topic of documents to learn inferences
- Temporal ordering of Events
  - Applied state of the art temporal classification to show that sets of events can be partially ordered
- Structured Selection
  - Event space of narrative relations can be clustered to create discrete sets
  - Offer insight into quality of narratives
- Important area not discussed: use for semantic role learning