Learning Script
Knowledge from the Web

Natural Language Processing Seminar - Fall 2014
Lena Reed, PaUlo GomeS and Kendall Lewis
Introduction

- Script: “a standardized sequence of events that describes some stereotypical human activity such as going to a restaurant or visiting a doctor” (Barr and Feigenbaum, 1981)
- Event sequence descriptions (ESDs): temporal event structure that focuses on a single scenario.
- Goal is to create a *temporal script graph* by aligning and combining the ESDs.
Related Work

- Differences:
  - Acquires specific scripts from arbitrary domains
  - Control level of granularity
  - Their data will be more explicit which should be easier to learn from
  - Automatically learns different phrases that represent the same event
Scripts

- Scenarios - classes of human activities
- Events - must occur in a certain order
- Modeling and learning temporal order
  - directed graph $G_s = (E_s, T_s)$
  - $E_s$ - set of nodes representing events of scenario $s$
  - $T_s$ - set of edges $(e_i, e_k)$ indicating $e_i$ typically happens before $e_k$ in $s$
Data

- 22 scenarios
  - harder to describe scenarios
  - highly variable order
  - cultural differences

- Used MTurk to create ESDs
  - Bullet point style
  - 5 - 16 events per scenario
  - Can skip scenarios

- Collected 493 ESDs, discarded 15%
- Average 9 events per ESD, widely variable length
- 93% of individual events are unique
Algorithm takes sequences, a cost function for substitutions (alignments), and gap costs for insertions and deletions.

Multiple Sequence Alignment is a matrix
- If a row contains two non-gaps, those symbols are aligned
- Aligning a non-gap with a gap is an insertion or deletion
- Sum the alignment cost for any two symbols from $\Sigma$ that are aligned with each other and add the gap cost for each gap
Semantic similarity

- Intuitively, want the MSA to prefer the alignment of two phrases if they are semantically similar

- **Predicate** = first potential verb of the phrase
- **Subject** = the preceding noun (if any)
- **Objects** = all following potential nouns
- Calculate similarity as: 
  \[ sim = \alpha \cdot \text{pred} + \beta \cdot \text{subj} + \gamma \cdot \text{obj} \]

- If a constituent is not present in one of the phrases, set its weight to zero and redistribute it over the other weights
Semantic similarity

- Individual similarity values fixed depending on the WordNet relation between the most similar WordNet senses of the respective lemmas
- Optimized the similarity values and weights using a held-out development set of scenarios
  - Set $\alpha$ higher than other weights since the verb contributes most to the similarity
  - If verb contributes little to the meaning of the phrase, it’s assigned a lower $\alpha$
Building Graphs

- Construct an initial graph with one node for each row of the MSA
- Add an edge \((u,v)\) to the graph following some constraints
- Graph is automatically post-processed:
  - Prune spurious nodes that contain only one event description (i.e. was the only non-gap in a row in the MSA)
  - Merge nodes whose elements should have been aligned previously by were missed by the MSA
    - Must satisfy certain structural and semantic constraints in order to be merged
- Output of post-processing step is the *temporal script graph* (TSG)
Evaluation Methodology

- 10 scenarios: 5 crowd sourced and 5 from OMICS [Singh et al., 2002].
- Tasks
  - paraphrase: 30 system aligned and 30 random.
  - happens-before: 30 system sequential, 30 random and 60 in reverse.
- 5 turkers annotated, expert decides in 3:2 situations.
Baselines & Upper Bound

- **Clustering** (*cl*):
  - paraphrase if in the same cluster.
  - happens-before *e*-f if some event from *e*’s cluster precedes some event from *f*’s cluster.
- **Levenshtein** (*lev*): string distance for similarity and node merging.
- **Upper Bound** (*upper*): random human annotation for each pair.
| SCENARIO                        | | SCENARIO                        | |
|--------------------------------|--------------------------------|
| **MC-TEK**                     | | **OMIN**                        | |
| **sys**                        | **base_cdt**                   | **base_clev**                   | **sys**                        | **base_cdt**                   | **base_clev**                   |
| pay with credit card           | 0.52 0.43 0.50                 | 0.84 0.89 0.11                 | 0.64 0.58 0.17                 | 0.60                           |                                  |
| eat in restaurant              | 0.70 0.42 0.75                 | 0.88 1.00 0.25                 | 0.78 0.59 0.38                 | 0.79                           |                                  |
| iron clothes I                 | 0.52 0.32 1.00                 | 0.94 1.00 0.12                 | 0.67 0.48 0.21                 | 0.82                           |                                  |
| cook scrambled eggs            | 0.58 0.34 0.50                 | 0.86 0.95 0.10                 | 0.69 0.50 0.16                 | 0.91                           |                                  |
| take a bus                     | 0.65 0.42 0.40                 | 0.87 1.00 0.09                 | 0.74 0.59 0.14                 | 0.88                           |                                  |
| **OMIN**                       | | **OMIN**                        | |
| answer the phone               | 0.93 0.45 0.70                 | 0.85 1.00 0.21                 | 0.89 0.71 0.33                 | 0.79                           |                                  |
| buy from vending machine       | 0.59 0.43 0.59                 | 0.83 1.00 0.54                 | 0.69 0.60 0.57                 | 0.80                           |                                  |
| iron clothes II                | 0.57 0.30 0.33                 | 0.94 1.00 0.22                 | 0.71 0.46 0.27                 | 0.77                           |                                  |
| make coffee                    | 0.50 0.27 0.56                 | 0.94 1.00 0.31                 | 0.65 0.42 0.40                 | 0.82                           |                                  |
| make omelette                  | 0.75 0.54 0.67                 | 0.92 0.96 0.23                 | 0.83 0.69 0.34                 | 0.85                           |                                  |
| **AVERAGE**                    | **0.63 0.40 0.60**             | **0.89 0.98 0.22**             | **0.73 0.56 0.30**             | **0.82**                       |                                  |

Figure 4: Results for paraphrasing task; significance of difference to sys: ⋄ : p ≤ 0.01, ● : p ≤ 0.1

| SCENARIO                        | | SCENARIO                        | |
|--------------------------------|--------------------------------|
| **MC-TEK**                     | | **OMIN**                        | |
| **sys**                        | **base_cdt**                   | **base_clev**                   | **sys**                        | **base_cdt**                   | **base_clev**                   |
| pay with credit card           | 0.86 0.49 0.65                 | 0.84 0.74 0.45                 | 0.85 0.59 0.53                 | 0.92                           |                                  |
| eat in restaurant              | 0.78 0.48 0.68                 | 0.84 0.98 0.75                 | 0.81 0.64 0.71                 | 0.95                           |                                  |
| iron clothes I                 | 0.78 0.54 0.75                 | 0.72 0.95 0.53                 | 0.75 0.69 0.62                 | 0.92                           |                                  |
| cook scrambled eggs            | 0.67 0.54 0.55                 | 0.64 0.98 0.69                 | 0.66 0.70 0.61                 | 0.88                           |                                  |
| take a bus                     | 0.80 0.49 0.68                 | 0.80 1.00 0.37                 | 0.80 0.66 0.48                 | 0.96                           |                                  |
| **OMIN**                       | | **OMIN**                        | |
| answer the phone               | 0.83 0.48 0.79                 | 0.86 1.00 0.96                 | 0.84 0.64 0.87                 | 0.90                           |                                  |
| buy from vending machine       | 0.84 0.51 0.69                 | 0.85 0.90 0.75                 | 0.84 0.66 0.71                 | 0.83                           |                                  |
| iron clothes II                | 0.78 0.48 0.75                 | 0.80 0.96 0.66                 | 0.79 0.64 0.70                 | 0.84                           |                                  |
| make coffee                    | 0.70 0.55 0.50                 | 0.78 1.00 0.55                 | 0.74 0.71 0.53                 | 0.83                           |                                  |
| make omelette                  | 0.70 0.55 0.79                 | 0.83 0.93 0.82                 | 0.76 0.69 0.81                 | 0.92                           |                                  |
| **AVERAGE**                    | **0.77 0.51 0.68**             | **0.80 0.95 0.65**             | **0.78 0.66 0.66**             | **0.90**                       |                                  |

Figure 5: Results for happens-before task; significance of difference to sys: ⋄ : p ≤ 0.01, ● : p ≤ 0.1

Results
Summary

- Learn generalized scripts of everyday life scenarios.
- Approach
  - Crowdsource event sequence descriptions.
  - Minimize cost of aligning them with MSA.
  - Construct graph representing the temporal dependencies using semantic and structural constraints.
- Better F1 scores on average than chosen baselines.
Discussion