

Learning Script Knowledge with Web Experiments

Michaela Regneri, Alexander Koller, Manfred Pinkal

Hannah Seitz

Selected Topics in Semantics and Discourse

November 24, 2015

How do you order pizza?

- 1 Take menu of your favorite delivery service
- 2 Decide for a pizza
- 3 Call number on the menu
- 4 Tell which pizza you want
- 5 Wait some time
- 6 Get your pizza
- 7 Pay pizza
- 8 Eat pizza



How do you order pizza?

- 1 Take menu of your favorite delivery service
- 2 Decide for a pizza
- 3 Call number on the menu
- 4 Tell which pizza you want
- 5 Wait some time
- 6 Get your pizza
- 7 Pay pizza
- 8 Eat pizza

Script:

standardized sequence of events that describes some stereotypical human activity

Script Knowledge

- Implicit knowledge
- Shared between all members of a culture
- Benefit for text understanding systems
 - Manual creation not feasible
 - Must be created automatically

Focus of this paper:

- Data acquisition: Where to find training data?
- How to compute graphs that show:
 - Paraphrases: Which phrases describe the same event?
 - Temporal order: Which event happens before another one?

Terminology

- Script:
 - Stereotypical models of activities (actions 1-8 for ordering pizza)
- Scenario:
 - A class of activities (ordering pizza)
- Event:
 - Abstract part of a script (paying)
- Event Description:
 - Linguistic realization of an event (" pay the bill" , " pay for food")
- Event Sequence Description (ESD)
 - Ordered list of event descriptions (first slide)

Training Data

- Ask experts:
 - Not feasible on large scale
- Learn from corpora:
 - Most knowledge is implicit
 - Only special scenarios can be found in texts
- Ask non-experts:
 - Explicit data
 - Any domain, any granularity

Ask non-experts: Amazon Mechanical Turk

The screenshot shows the Amazon Mechanical Turk search results page. The browser address bar displays the URL: <https://www.mturk.com/mturk/searchbar?selectedSearchType=hitgroups&searchWords=&minReward=0.0&w=78y=>. The page header includes the Amazon Mechanical Turk logo, navigation tabs for 'Your Account', 'HITS', and 'Qualifications', and a notification for '655,264 HITS available now'. Below the header, there is a search filter section with a dropdown menu set to 'HITS', a search bar, and a filter for 'that pay at least \$ 0.0'. The main content area displays a list of HITs under the heading 'HITS that pay at least \$0.0'. The list shows 1-10 of 2890 results, sorted by 'HITS Available (most first)'. Each HIT entry includes a 'Transcribe data' or 'Classify Receipt' task, the requester's name (p9r or Jon Breig), the HIT expiration date, the time allotted, and the reward amount. A 'View a HIT in this group' link is provided for each entry.

amazonmechanicalturk Artificial Intelligence

Your Account HITS Qualifications 655,264 HITS available now Sign In

All HITS | HITS Available To You | HITS Assigned To You

Find HITS containing that pay at least \$ 0.0 for which you are qualified Require Master Qualification

HITS that pay at least \$0.0
1-10 of 2890 Results

Sort by: HITS Available (most first) Show all details | Hide all details 1 2 3 4 5 > Next >> Last

| | | | |
|------------------------------------------|--------------------------------------------------------------------------------------|----------------|------------------------------------------|
| Transcribe data Requester: p9r | HIT Expiration Date: Nov 19, 2015 (23 hours 55 minutes) Time Allotted: 45 minutes | Reward: \$0.03 | View a HIT in this group |
| Transcribe data Requester: p9r | HIT Expiration Date: Nov 19, 2015 (23 hours 55 minutes) Time Allotted: 45 minutes | Reward: \$0.04 | View a HIT in this group |
| Transcribe data Requester: p9r | HIT Expiration Date: Nov 19, 2015 (23 hours 55 minutes) Time Allotted: 45 minutes | Reward: \$0.02 | View a HIT in this group |
| Transcribe data Requester: p9r | HIT Expiration Date: Nov 19, 2015 (23 hours 58 minutes) Time Allotted: 45 minutes | Reward: \$0.05 | View a HIT in this group |
| Classify Receipt Requester: Jon Breig | HIT Expiration Date: Nov 25, 2015 (6 days 23 hours) Time Allotted: 20 minutes | Reward: \$0.02 | View a HIT in this group |
| Find the company address | | | View a HIT in this group |

Ask non-experts: Amazon Mechanical Turk

Transcribe data

Requester: p9r

Qualifications Required: HIT approval rate (%) is greater than 90

Reward: \$0.06 per HIT

Duration: 45 minutes

ANSWERS

Day[09] Month[09] Year[2015]

Enter value:

Blank

Not clear / not allowed

q c

ANSWERS

Day[09] Month[09] Year[2015]

Enter value:

Blank

Not clear / not allowed

Ask non-experts

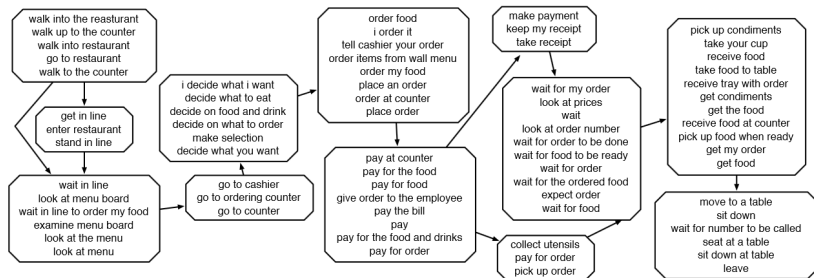
- 25 people
- 22 scenarios
- Task:
 - Describe scenario
 - 5-16 bulletin points
 - Temporal ordering
- Results:
 - Event Sequence Descriptions
- Manual correction

1. look at menu
2. decide what you want
3. order at counter
4. pay at counter
5. receive food at counter
6. take food to table
7. eat food

1. walk to the counter
2. place an order
3. pay the bill
4. wait for the ordered food
5. get the food
6. move to a table
7. eat food
8. exit the place

1. walk into restaurant
2. find the end of the line
3. stand in line
4. look at menu board
5. decide on food and drink
6. tell cashier your order
7. listen to cashier repeat order
8. listen for total price
9. swipe credit card in scanner
10. put up credit card
11. take receipt
12. look at order number
13. take your cup
14. stand off to the side
15. wait for number to be called
16. get your drink

Temporal Script Graphs



Computing Temporal Script Graphs

- Positive:
 - Each ESD is already ordered
- But:
 - Different ESDs use different words
 - Not every ESD contains every step

Task:

- Identify phrases from different ESDs that describe the same event
- Multiple Sequence Alignment (MSA)

Multiple Sequence Alignment (MSA)

| row | S ₁ | S ₂ | S ₃ | S ₄ |
|-----|-------------------------|--------------------------------|---------------------------|--------------------|
| 1 | ∅ | walk into restaurant | ∅ | enter restaurant |
| 2 | ∅ | ∅ | walk to the counter | go to counter |
| 3 | ∅ | find the end of the line | ∅ | ∅ |
| 4 | ∅ | stand in line | ∅ | ∅ |
| 5 | look at menu | look at menu board | ∅ | ∅ |
| 6 | decide what you want | decide on food and drink | ∅ | make selection |
| 7 | order at counter | tell cashier your order | place an order | place order |
| 8 | ∅ | listen to cashier repeat order | ∅ | ∅ |
| 9 | pay at counter | ∅ | pay the bill | pay for food |
| 10 | ∅ | listen for total price | ∅ | ∅ |
| 11 | ∅ | swipe credit card in scanner | ∅ | ∅ |
| 12 | ∅ | put up credit card | ∅ | ∅ |
| 13 | ∅ | take receipt | ∅ | ∅ |
| 14 | ∅ | look at order number | ∅ | ∅ |
| 15 | ∅ | take your cup | ∅ | ∅ |
| 16 | ∅ | stand off to the side | ∅ | ∅ |
| 17 | ∅ | wait for number to be called | wait for the ordered food | ∅ |
| 18 | receive food at counter | get your drink | get the food | pick up order |
| 19 | ∅ | ∅ | ∅ | pick up condiments |
| 20 | take food to table | ∅ | move to a table | go to table |
| 21 | eat food | ∅ | eat food | consume food |
| 22 | ∅ | ∅ | ∅ | clear tray |
| 22 | ∅ | ∅ | exit the place | ∅ |

Multiple Sequence Alignment (MSA)

- 1 Calculate pairwise similarity
- 2 For 2 sequences
- 3 Find cheapest alignment:

minimize:

$$c(A) = c_{gap} * |gap| + \sum_{i=1}^n \sum_{j=1, a_{ij} \neq gap}^m \sum_{k=j+1, a_{ki} \neq gap}^m cost_m(a_{ij}, a_{ki})$$

- 4 Consider result as one sequence
- 5 Back to 2

Semantic Similarity ($cost_m$)

$$sim = \alpha * pred + \beta * subj + \gamma * obj$$

- Identify subject, object(s) and predicate
- Similarity scores based on Wordnet
- Optimized weights, $\alpha > \beta$, $\alpha > \gamma$

Multiple Sequence Alignment (MSA)

| row | S ₁ | S ₂ | S ₃ | S ₄ |
|-----|-------------------------|--------------------------------|---------------------------|--------------------|
| 1 | ∅ | walk into restaurant | ∅ | enter restaurant |
| 2 | ∅ | ∅ | walk to the counter | go to counter |
| 3 | ∅ | find the end of the line | ∅ | ∅ |
| 4 | ∅ | stand in line | ∅ | ∅ |
| 5 | look at menu | look at menu board | ∅ | ∅ |
| 6 | decide what you want | decide on food and drink | ∅ | make selection |
| 7 | order at counter | tell cashier your order | place an order | place order |
| 8 | ∅ | listen to cashier repeat order | ∅ | ∅ |
| 9 | pay at counter | ∅ | pay the bill | pay for food |
| 10 | ∅ | listen for total price | ∅ | ∅ |
| 11 | ∅ | swipe credit card in scanner | ∅ | ∅ |
| 12 | ∅ | put up credit card | ∅ | ∅ |
| 13 | ∅ | take receipt | ∅ | ∅ |
| 14 | ∅ | look at order number | ∅ | ∅ |
| 15 | ∅ | take your cup | ∅ | ∅ |
| 16 | ∅ | stand off to the side | ∅ | ∅ |
| 17 | ∅ | wait for number to be called | wait for the ordered food | ∅ |
| 18 | receive food at counter | get your drink | get the food | pick up order |
| 19 | ∅ | ∅ | ∅ | pick up condiments |
| 20 | take food to table | ∅ | move to a table | go to table |
| 21 | eat food | ∅ | eat food | consume food |
| 22 | ∅ | ∅ | ∅ | clear tray |
| 22 | ∅ | ∅ | exit the place | ∅ |

Evaluation

- 10 scenarios
- Paraphrase set
 - 30 pairs classified as paraphrases
 - 30 pairs random selected
- Happens-before-set
 - 30 pairs classified as happens-before
 - 30 pairs random selected
 - All 60 pairs in reverse order
- Gold standard: 5 non-experts

Baselines

- Upper bound
 - Random selected human annotation
- Clustering
 - One node per event description
 - Semantic similarity: as before
- Levenshtein
 - Semantic similarity: Levenshtein Distance
 - Character-wise distance / length

Paraphrase task

| SCENARIO | PRECISION | | | RECALL | | | F-SCORE | | | | |
|----------|--------------------------|--------------------|---------------------|--------|--------------------|---------------------|-------------|--------------------|---------------------|--------|--------|
| | sys | base _{cl} | base _{lev} | sys | base _{cl} | base _{lev} | sys | base _{cl} | base _{lev} | upper | |
| MTURK | 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 | ● 0.92 |
| | 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 |
| OMICS | 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: ● : $p \leq 0.01$, ○ : $p \leq 0.1$

Happens-before-task

| SCENARIO | PRECISION | | | RECALL | | | F-SCORE | | | upper | |
|----------|--------------------------|--------------------|---------------------|--------|--------------------|---------------------|-------------|--------------------|---------------------|-------------|--------|
| | sys | base _{cl} | base _{lev} | sys | base _{cl} | base _{lev} | sys | base _{cl} | base _{lev} | | |
| MTURK | 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 |
| OMICS | 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: ● : $p \leq 0.01$, ○ : $p \leq 0.1$


Summary


- Better than baselines
 - Levenshtein: too restrictive: low recall, high precision
 - Clustering: can't use sequential information: high recall, low precision
 - Close to upper bound in most cases
- Results depend on scenarios
 - Complexity
 - Variable orderings

Improvements

- Scale to any number of scenarios
- Less supervision:
 - Restrict user input
 - Heuristics for filtering data
- More elaborate data structures
 - Optional events
 - Alternative events
 - Events with arbitrary order

References

 M. Regneri, A. Koller, M. Pinkal (2009)
Learning Script Knowledge with Web Experiments
Proceedings of ACL 2010

 Desmond G. Higgins and Paul M. Sharp (1988)
CLUSTAL: a package for performing multiple sequence alignment on a microcomputer
Gene. Volume 73, 1988, Pages 237-244

Thank you for your attention!