Statistical Scripts

Statistical Scripts are models of co-occurring events which allow us to infer additional events from a document.

**Multi-Argument Events**

How to represent events?

*Bob called Alice but she ignored him.*

Previous work uses a protagonist model with verb-dependency pairs [1; 2]:

- Bob (call, subj) (ignore, obj)
- Alice (call, obj) (ignore, subj)

Instead, we use multi-argument events:

- call(B, A)
- ignore(A, B)

These events capture entity interactions that we couldn’t before.

Learning

- Run parser and coreference engine on unlabeled corpus.
- Extract one event sequence per document.
- Abstract entity mentions into variables, with one variable per coreference class.
- Count co-occurrences between events $a$ and $b$ to estimate $P(a,b)$.

Inference

Following [2], infer event $a$ at position $p$ by maximizing probability of $a$ following earlier events and preceding later events:

$$S(a) = \sum_{i=1}^{p-1} \log P(a|a_i) + \sum_{i=p+1}^{A} \log P(a_i|a)$$

Evaluation

**Narrative Cloze** [1; 2]: evaluate a system by inferring held-out events from documents.

**Results**

Systems compared:

- **Random**: guess events at random.
- **Unigram**: guess events according only to frequency
- **Multi-protagonist**: combine inferred (verb, dependency) pairs into multi-argument events.
- **Joint**: directly model multi-argument events.

**Inferring Held-out Multi-Argument Events**:

Directly modeling entity interactions provides better prediction of held-out events, in both multi-argument and verb-dependency-pair inference.

Previous work uses a protagonist model with verb-dependency pairs [1; 2]:


This research was supported in part by the DARPA DEFT program under AFRL grant F29601-13-2-0026.