A Bayesian Model for Unsupervised Semantic Parsing

Ivan Titov
titov@mmci.uni-saarland.de

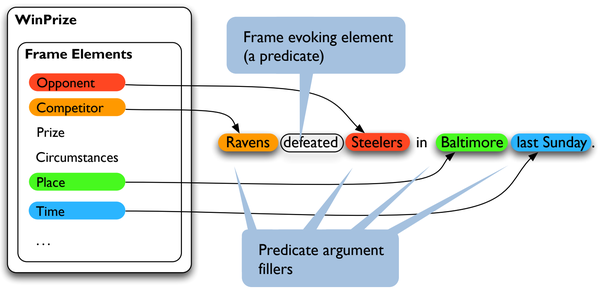
Alexandre Klementiev
aklement@jhu.edu

UNIVERSITÄT DES SAARLANDES

The Center For Language and Speech Processing
at the Johns Hopkins University

Task Definition

In frame semantics, a semantic frame is a conceptual structure describing a situation, object, or event along with associated properties and participants.



We treat semantic parsing as recursive prediction of predicate-argument structure and cluster labels for argument fillers.

- Decompose the sentence into lexical items (one or more words)
- Assign a cluster label to every lexical item
  - If predicate: assign a semantic frame
  - If argument: assign an argument filler cluster

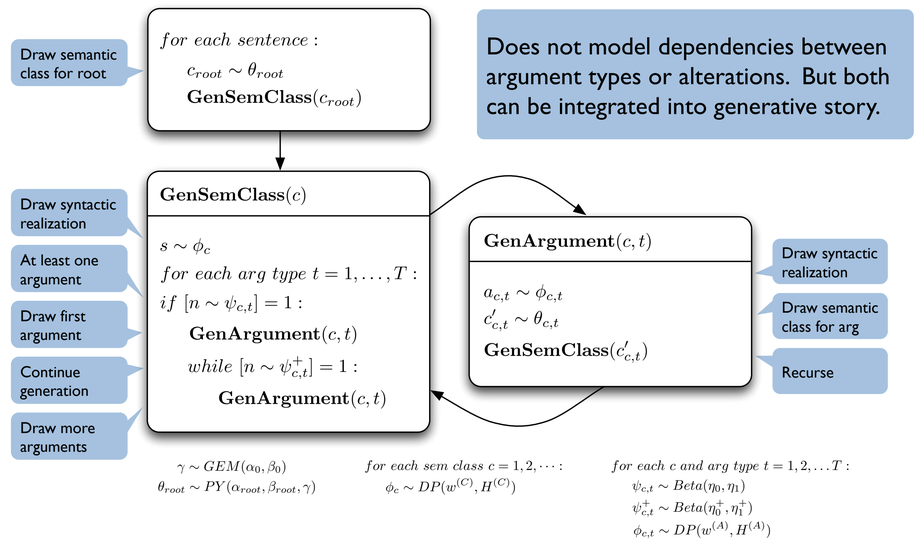
  We call them *semantic classes*
- Predict predicate-argument relations between lexical items

Our Assumptions

- Each lexical item corresponds to a subtree of the syntactic dependency graph of a sentence
- One lexical item can be an argument of another only if they are connected by an arc in the dependency tree
- We do not model polysemy: a syntactic fragment corresponds to a single semantic class

The Model

Our unsupervised model simultaneously induces frames and clusters of argument fillers by exploiting distributional context.



The model parameters control:

- The distribution of syntactic and lexical realizations (lexical items) for each semantic class (= frame evoking elements for each frame)
- The distribution over combinations of argument types (roles) for each semantic class
- Linking between semantic argument types and syntactic dependencies for each semantic class
- Selectional preferences for each argument type: a distribution over semantic classes of argument fillers

Inference

The model is estimated using sentences automatically annotated with syntactic structures and treating their semantic representation as latent

Metropolis-Hastings moves:

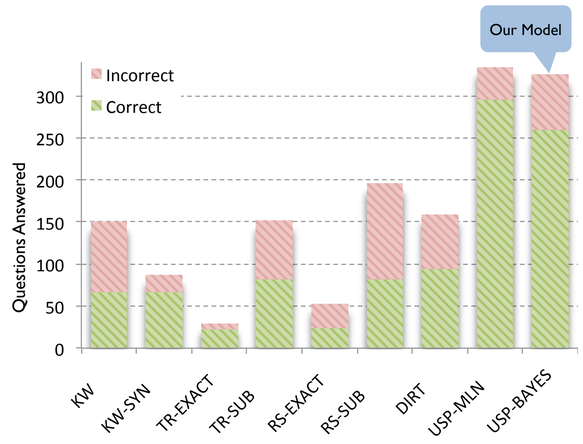
- Split-Merge: merges two semantics classes together or decomposes one class into two
- Compose-Decompose: composes two lexical items (dependency subtrees) or decomposes a lexical item
- Role-Syntax Alignment: generates a new linking between syntax and semantics for a frame

Proposal Distributions:

- For Split-Merge: based on cosine similarity of lexical and syntactic contexts
- For Compose-Decompose: based on how frequently the pair of fragments is connected

Evaluation

Question Answering about knowledge in a corpus of biomedical abstracts (GENIA)



More than 55% of mistakes are due to over coarse clustering in 3 semantic classes (antonymy / hyponymy)

Examples of induced semantic classes:

Roughly "cause change position on a scale" frame

Class	Variations
1	motif, sequence, regulatory element, response element, element, dna sequence
2	donor, individual, subject
3	important, essential, critical
4	dose, concentration
5	activation, transcriptional activation, transactivation
6	b cell, t lymphocyte, thymocyte, b lymphocyte, t cell, t-cell line, human lymphocyte, t-lymphocyte
7	indicate, reveal, document, suggest, demonstrate
8	augment, abolish, inhibit, convert, cause, abrogate, modulate, block, decrease, reduce, diminish, suppress, up-regulate, impair, reverse, enhance
9	confirm, assess, examine, study, evaluate, test, resolve, determine, investigate
10	nf-kappab, nf-kappa b, nfkappab, nf-kb

Example questions and answers:

Question	What does cyclosporin A suppress?
Answer	expression of EGR-2
Sentence	As with EGR-3 , expression of EGR-2 was blocked by cyclosporin A
Question	What inhibits tnf-alpha?
Answer	IL -10
Sentence	Our previous studies in human monocytes have demonstrated that interleukin (IL) -10 inhibits lipopolysaccharide (LPS) -stimulated production of inflammatory cytokines , IL-1 beta , IL-6 , IL-8 , and tumor necrosis factor (TNF) -alpha by blocking gene transcription .