Semantic Frame Induction

Eli Drumm, Bill Noble, Jacob Verdegaal

University of Amsterdam

12/16/14

Overview

Semantic Frames

- Frame semantics – Charles J. Fillmore (1976).
- A semantic frame is "a description of a type of event, relation, or entity and the participants in it." (Framenet)
- There is no mapping between frames and verbs.

An example:

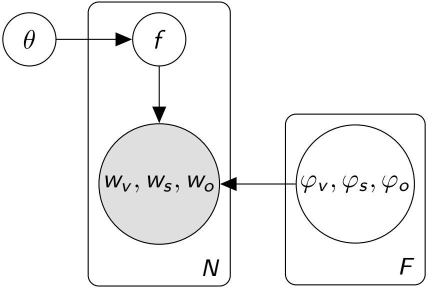
predicate:	Transaction
role 1:	baker
role 2:	Alice
role 3:	bread

- Alice buys a bread from the baker
- The baker sells Alice a bread

Google Syntactic-Ngrams

- A syntactic-ngram is a *k*-word rooted subtree for some sentence.
- Google ngrams come from a corpus of 3.5 million English books.
- We trimmed the "verb args" dataset to consider only subject-verb-object triples (VSO's).
- The dataset contains 1,629,120 unique VSO's with a total of 96,245,401 by count.
- In our final results we may only use the most common %20 of these...

Model 0 - EM



N	number of data points
F	number of frames
θ	distribution over frames
φ_a	distributions (for each frame) over vocabulary of argument a
w_a	datapoint (verb, subject, object)

E-Step:

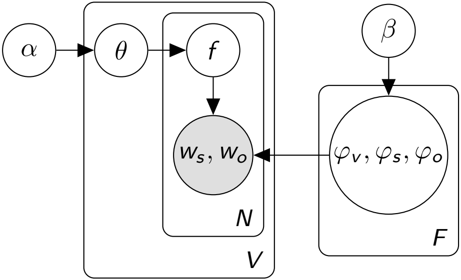
$$P(f|w_i) = \frac{\prod_a^{\{v,s,o\}} P(w_i^a|\varphi_f^a)P(f|\theta)}{\sum_{f'}^F \prod_a^{\{v,s,o\}} P(w_i^a|\varphi_{f'}^a)}$$

M-Step:

$$P(f|\theta) = \frac{\sum_{i=1}^N \mu_i(f)}{\sum_{f'=1}^F \sum_{i=1}^N \mu_i(f')}$$

$$P(w|\varphi_f^a) = \frac{\sum_{i=1}^N \mu_i(f)C(w,a)}{\sum_{w'=1}^{V^a} \sum_{i=1}^N \mu_i(f)C(w',a)}$$

Model 1 - Gibbs Sampling



N	number of data points
F	number of frames
V	number of documents (= verb vocabulary)
θ	distribution over frames
φ_a	distributions (for each frame) over vocabulary of argument a
w_a	datapoint (verb, subject, object)
β	hyperprior for φ (for each argument)
α	hyperprior for θ

Gibbs:

$$P(z_{ij} = f | \mathbf{f}_{-ij}, \alpha, \beta, w) \propto \prod_{a}^{\{v,s,o\}} \frac{\beta^a + \tilde{C}(f, w_{ij})}{V^a + \tilde{C}(f)} \cdot \frac{\alpha + \tilde{C}(i, f)}{F\alpha + \tilde{C}(i)}$$

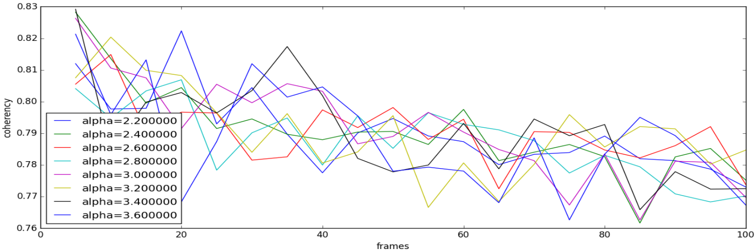
$\tilde{C}(f, w_{ij})$	count times w_{ij} is assigned to frame f
$\tilde{C}(f)$	count total data points assigend to f
$\tilde{C}(i, f)$	count sentences in document i assigned to f
$\tilde{C}(i)$	count sentences in document i

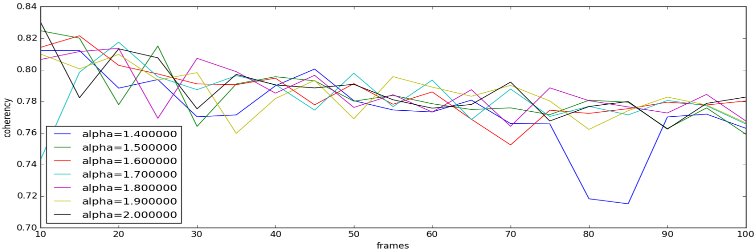
Evaluation metrics

- Frame coherency: for a datapoint (v, s, o) and a tuple (v^r, s, o) where v^r is a random choses verb: $P(v \mid s, o) \geq P(v^r \mid s, o)$
- Frame correctness: for the top 25 most probable verbs per frame TV and framenet classes of verbs FN:

$$\frac{2|TV \cap FN|}{|TV| + |FN|}$$

Results EM - model 0





Results Gibbs sampling - model 1

[1] Yoav Goldberg and Jon Orwant. A dataset of syntactic-ngrams over time from a very large corpus of english books. In *Second Joint Conference on Lexical and Computational Semantics (* SEM)*, volume 1, pages 241–247, 2013.

[2] Brendan O'Connor. Learning frames from text with an unsupervised latent variable model. 2013.

[3] Mats Rooth, Stefan Riezler, Detlef Prescher, Glenn Carroll, and Franz Beil. Inducing a semantically annotated lexicon via em-based clustering. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 104–111. Association for Computational Linguistics, 1999.