A Bayesian Approach to Unsupervised Semantic Role Induction

Ivan Titov and Alexandre Klementiev

Frame Semantics

- Frame semantics and the nature of language (Fillmore, 1977)
- The Berkeley FrameNet Project (Baker et al., 1998)
- Automatic Labeling of Semantic Roles (Gildea and Jurafsky, 2002)
- Frequent task at CoNLL, SensEval/SemEval

What is a frame?

- Structures that describe particular situations, objects, or events

- Each frame contains a set of roles for participating elements

Frame: Ingestion

Michael eats a sandwhich

A sandwich is eaten by Michael

Frame: Ingestion

Michael eats a sandwhich

$[\text{Michael}]_{Ingestor}$ **eats** $[\text{a sandwich}]_{Ingestible}$

A sandwich is eaten by Michael

$[\text{A sandwich}]_{Ingestible}$ **is eaten** $[\text{by Michael}]_{Ingestor}$

A point of confusion!
(And there are many!)

- FrameNet vs. PropBank

A point of confusion!
(And there are many!)

- FrameNet vs. PropBank

FrameNet

<u>Frame</u>: Giving
<u>Frame Elements(Roles)</u>:
Donor
Recipient
Theme
Circumstances
Depictive
Manner
...

PropBank

<u>Frame</u>: Give
<u>Frame Elements(Roles)</u>:
Arg0 (Typically Agent)
Arg1 (Typically Patient)
Arg2
...

Semantic Role Labeling
The Task

2 Stages (Frame Induction)

1. Identification of arguments (Syntax based heuristics)

2. Role labeling (What this paper is all about!)

Semantic Role Labeling

Inputs or how is this unsupervised?

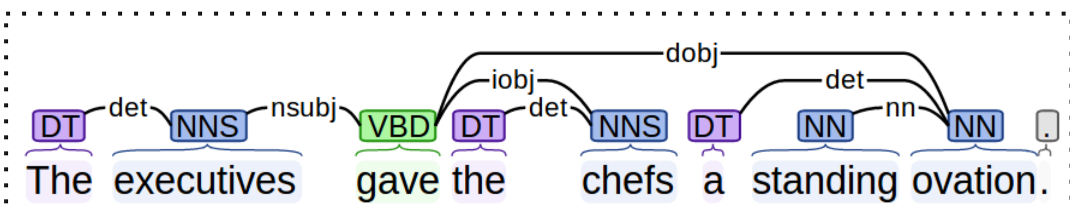
1. A sentence

2. An automatically generated syntactic dependency graph of that sentence

Semantic Role Labeling

An Example

The executives gave the chefs a standing ovation.



Frame: Give

Roles:

- Arg0: giver
- Arg1: thing given
- Arg2: entity given to

Semantic Role Labeling

An Example

The executives **gave** the chefs a standing ovation.

Rel: gave

Frame: Give

Roles:

- Arg0: giver
- Arg1: thing given
- Arg2: entity given to

Semantic Role Labeling

An Example

{The executives} **gave** {the chefs} {a standing ovation}

Rel: gave

Frame: Give

Roles:

- Arg0: giver
- Arg1: thing given
- Arg2: entity given to

Semantic Role Labeling

An Example

{The executives} **gave** {the chefs} {a standing ovation}.

Rel: gave

- Arg0: The executives
- Arg1: a standing ovation
- Arg2: the chefs

Frame: Give

Roles:

- Arg0: giver
- Arg1: thing given
- Arg2: entity given to

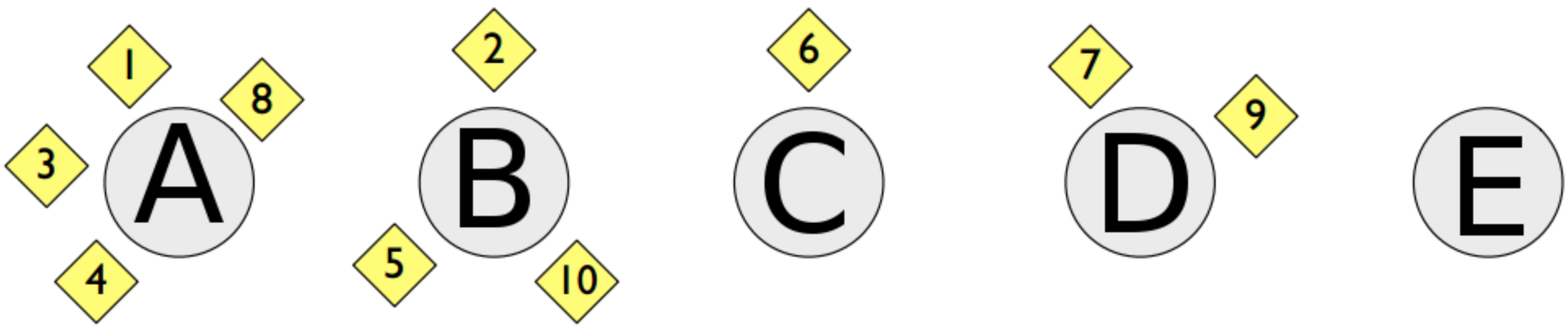
Argument Representation

- Argument Keys
  - Verb Voice (ACT/PASS)
  - Argument position relative to verb (LEFT/RIGHT)
  - Syntactic relation to governor
  - Preposition used for argument (if any)

- Argument Keys for 'Michael' in the sentences:
  a. Michael ate a sandwich.
     - ACT:LEFT:SBJ
  b. The sandwich was eaten by Michael.
     - PASS:RIGHT:LGS->BY

CRP and DD-CRP

- Chinese Restaurant Process



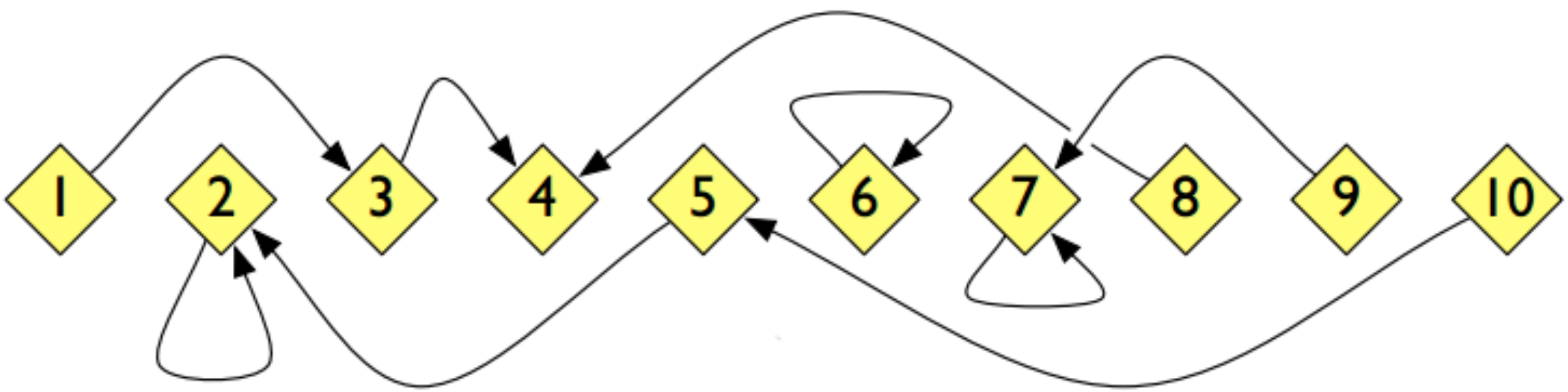
Given $i-1$ customers seated at K tables,

$$\mathbf{Pr}[c_i = k] = \frac{N_k}{i-1+\alpha} \qquad \mathbf{Pr}[c_i = K+1] = \frac{\alpha}{i-1+\alpha}$$

where c_i is table assingment of customer i and N_k is the number of customers already seated at table k.

CRP and DD-CRP

- Distance Dependant Chinese Restaurant Process



Given $i-1$ customers already seated, customer i chooses a partner with probability

$$\mathbf{Pr}[c_i = j | D, \alpha] = \frac{d_{i,j}}{\sum_{j'=1}^{i} d_{i,j'}}$$

where D is the entire similarity graph, $d_{i,j}$ is the similarity between customers i and j, and $d_{i,i} = \alpha$.

Models

Clustering of argument keys:

Factored model:

for each predicate $p = 1, 2, \ldots :$

$B_p \sim CRP(\alpha)$ [partition of arg keys]

Coupled model:

$D \sim NonInform$ [similarity graph]

for each predicate $p = 1, 2, \ldots :$

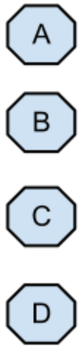
$B_p \sim dd\text{-}CRP(\alpha, D)$ [partition of arg keys]

MAP Search

- A simple greedy procedure
  - Each argument key is assigned to its own cluster
  - On first iteration, choose argument keys by their frequency in the corpus

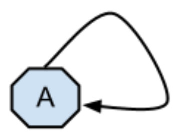
1. Choose an argument key at random
2. (Factored) Assign it to the most probable cluster, including a new cluster
   (Coupled) Assign it to the most probable partner, including itself

MAP Search



Order by frequency and assign seats.

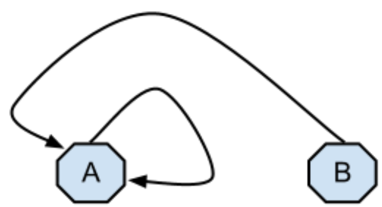
MAP Search

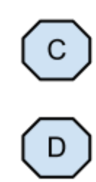


Order by frequency and assign seats.

MAP Search

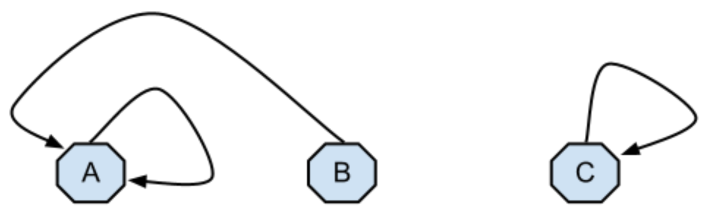
$$\mathbf{Pr}(c_B = A) > \mathbf{Pr}(c_B = B)$$

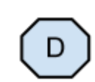




Order by frequency and assign seats.

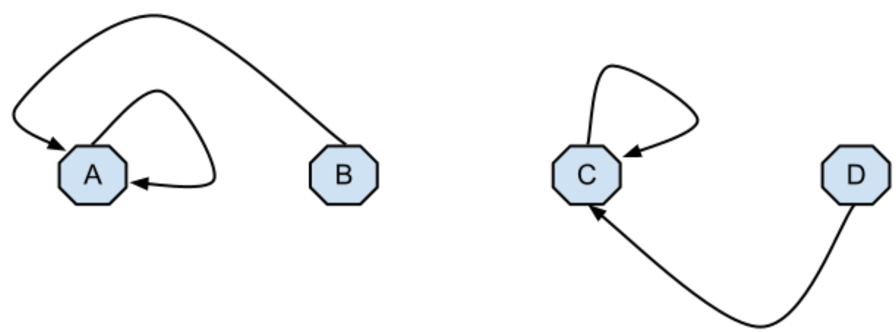
MAP Search





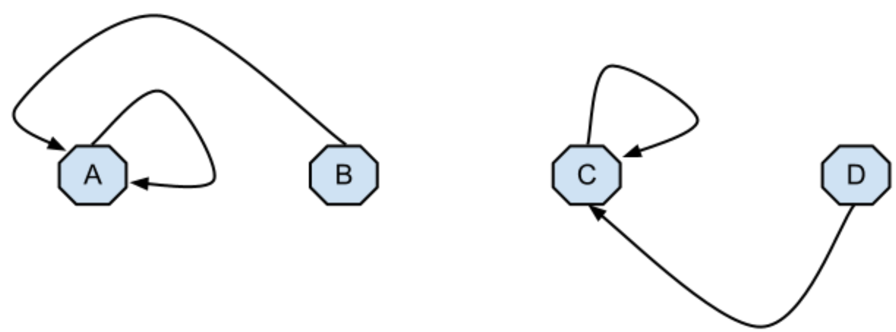
Order by frequency and assign seats.

MAP Search



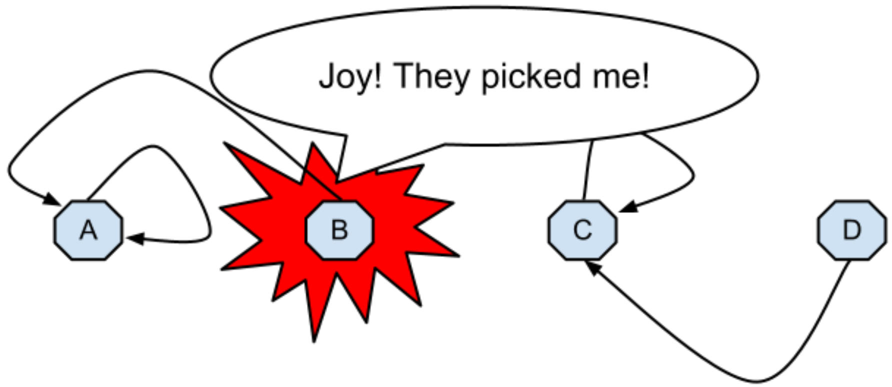
Order by frequency and assign seats.

MAP Search



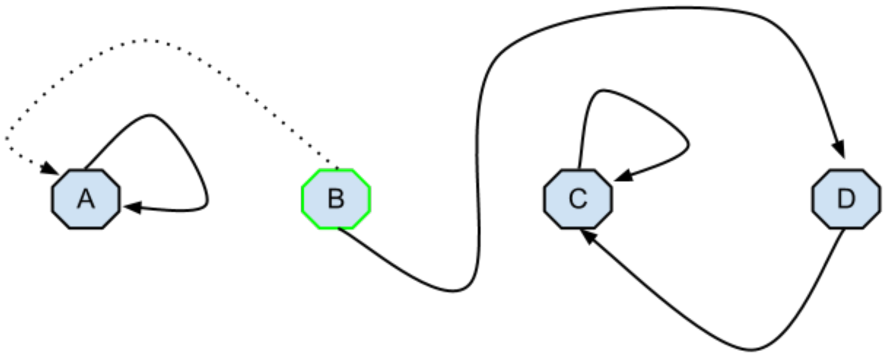
Now randomly select an argument key.

MAP Search



Now randomly select an argument key.

MAP Search



Reassign the argument key to the most probable argument key.

Models

Parameters:

for each predicate $p = 1, 2, \ldots$:
  for each role $r \in B_p$:
    $\theta_{p,r} \sim DP(\beta, H^{(A)})$ [distrib of arg fillers]
    $\psi_{p,r} \sim Beta(\eta_0, \eta_1)$ [geom distr for dup roles]

Data Generation:

for each predicate $p = 1, 2, \ldots$:
  for each occurrence l of p:
    for every role $r \in B_p$:
      if $[n \sim Unif(0,1)] = 1$: [role appears at least once]
        GenArgument(p, r) [draw one arg]
        while $[n \sim \psi_{p,r}] = 1$: [continue generation]
          GenArgument(p, r) [draw more args]

GenArgument(p, r):
  $k_{p,r} \sim Unif(1, \ldots, |r|)$ [draw arg key]
  $x_{p,r} \sim \theta_{p,r}$ [draw arg filler]

Gibbs Sampling for DD-CRP (Blei and Frazier, 2011)

Let model hyperparameters $\eta = \{D, \alpha, f, G_0\}$. $z(\mathbf{c})$ is the partition resultant from seating assignments $\mathbf{c}$.

$$\mathbf{Pr}(c_i^{(\text{new})}|\mathbf{c}_{-i}, \mathbf{x}, \eta) \propto \mathbf{Pr}(c_i^{(\text{new})}|D, \alpha)\mathbf{Pr}(\mathbf{x}|z(\mathbf{c}_{-i} \cup c_i^{(\text{new})}), G_0)$$

$$\begin{array}{rl} \mathbf{Pr}(c_i^{(\text{new})}|D, \alpha) & \text{prior} \\ \mathbf{Pr}(\mathbf{x}|z(\mathbf{c}_{-i} \cup c_i^{(\text{new})}), G_0) & \text{likelihood} \end{array}$$

Gibbs Sampling for DD-CRP (Blei and Frazier, 2011)

Let model hyperparameters $\eta = \{D, \alpha, f, G_0\}$. $z(\mathbf{c})$ is the partition resultant from seating assignments $\mathbf{c}$.

$$\mathbf{Pr}(c_i^{(\text{new})} | \mathbf{c}_{-i}, \mathbf{x}, \eta) \propto \mathbf{Pr}(c_i^{(\text{new})} | D, \alpha) \mathbf{Pr}(\mathbf{x} | z(\mathbf{c}_{-i} \cup c_i^{(\text{new})}), G_0)$$

$$\mathbf{Pr}(c_i^{(\text{new})} | \mathbf{c}_{-i}, \mathbf{x}, \eta) \propto \begin{cases} \alpha & \text{if } c_i^{(\text{new})} \text{ is equal to } i \\ f(d_{ij}) & \text{if } c_i^{(\text{new})} = j \text{ does not join two tables.} \\ f(d_{ij}) \dfrac{\mathbf{Pr}(\mathbf{x}_{z^k(\mathbf{c}_{-i}) \cup z^\ell(\mathbf{c}_{-i})} | G_0)}{\mathbf{Pr}(\mathbf{x}_{z^k(\mathbf{c}_{-i})} | G_0) \mathbf{Pr}(\mathbf{x}_{z^\ell(\mathbf{c}_{-i})} | G_0)} & \text{if } c_i^{(\text{new})} = j \text{ joins tables } k \text{ and } \ell. \end{cases}$$