# A Hierarchical Bayesian Model for Unsupervised Induction of Script Knowledge

Lea Frermann Ivan Titov Manfred Pinkal

April, 28th 2014

# Contents


1. Introduction
2. Technical Background
3. The Script Model
4. Evaluation & Results
5. Conclusion

# Scripts

*[...] A script is a predetermined, stereotyped sequence of actions that define a well-known situation.*[1]

[1] Schank and Abelson (1975)

# Scripts

*[...] A script is a predetermined, stereotyped sequence of actions that define a well-known situation.*[1]

Example situation: "Eating in a Restaurant"

| |
|---|
| Look at menu |
| Order your food |
| Wait for your food |
| Eat food |
| Pay |

Event Sequence Description (ESD): explicit instantiation of a script

[1]Schank and Abelson (1975)

# Scripts

*[...] A script is a predetermined, stereotyped sequence of actions that define a well-known situation.*[1]

Example situation: "Eating in a Restaurant"

| |
|---|
| Look at menu |
| Order your food |
| Wait for your food |
| Eat food |
| Pay |

a way of providing AI/NLP systems with world knowledge

- coherence estimation
- summarization

[1] Schank and Abelson (1975)

# Script Characteristics

| ESD 1 | ESD 2 |
| --- | --- |
| Look at menu | Check the menu |
| Order your food | Order the meal |
| Wait for your food | Wait for meal |
| | Talk to friends |
| Eat food | Have meal |
| Pay | Pay the bill |

**learn sequential ordering constraints**

# Script Characteristics

| ESD 1 | ESD 2 |
|---|---|
| Look at menu | Check the menu |
| Order your food | Order the meal |
| Wait for your food | Wait for meal |
| | Talk to friends |
| Eat food | Have meal |
| Pay | Pay the bill |

**learn sequential ordering constraints**

**learn event types (paraphrases)**

# Script Characteristics

| ESD 1 | ESD 2 |
|---|---|
| Look at menu | Check the menu |
| Order your food | Order the meal |
| Wait for your food | Wait for meal |
| | Talk to friends |
| Eat food | Have meal |
| Pay | Pay the bill |

**learn sequential ordering constraints**

**learn event types (paraphrases)**

**model participant types as latent variables**

# Event Sequence Descriptions (ESDs)

- from non-expert annotators (web experiments)
- noisy (grammar/spelling)
- variable number and detail of event descriptions
- few ESDs per scenario

"get menucard"
"search for items"
"order items"
"eat items"
"pay the bill"
"quit restaurant"

"enter the front door"
"let hostess seat you"
"tell waitress your drink order"
"tell waitress your food order"
"wait for food"
"eat and drink"
"get check from waitress"
"give waitress credit card"
"take charge slip from waitress"
"sign slip and add in a tip"
"leave slip on table"
"put up credit card"
"exit the restaurant"

# Unsupervised Learning of Scripts from Natural Text

## Chambers and Jurafsky (2008)

- infer script-like templates from news text ("Narrative Chains")
- learn event sets and ordering information in a 3-step process
  - identification of relevant events
  - temporal classification
  - clustering
- Chambers and Jurafsky (2009) learn events and participants jointly (but no ordering)

- script information often left implicit in natural text (world knowledge)

# Unsupervised Learning of Scripts from ESDs

### Regneri et al 2010

- collect sets of event sequence description for various scripts
- learn event types and orderings
  - align events across descriptions based on semantic similarity
  - compute graph representation using Multiple Sequence Alignment (MSA)
- Regneri et al. (2011) learn participant types based on those event graphs

- pipeline architecture
- MSA-based graphs cannot encode some script characteristics (e.g. event optionality)

# The Proposed Model

## (I) A Bayesian Script Model

- joint learning of event types and ordering constraints from ESDs
- generalized Mallows Model for modeling ordering constraints

**Bayesian Models of Ordering in NLP**

- document-level ordering constraints in structured text (Wikipedia Articles) (Chen et al., 2009)
- integrate a GMM into a standard topic model

# The Proposed Model

## (II) Informed Prior Knowledge from WordNet

- alleviates the problem of limited training data
- encode correlations between words based on WordNet similarity in the language model priors

**Encoding Correlations through a logistic normal distribution**

- "The correlated topic model" Blei and Lafferty (2006)

# Contents


1. Introduction
2. Technical Background
3. The Script Model
4. Evaluation & Results
5. Conclusion

# The Mallows Model (Mallows, 1957)

**A probability distribution over permutations of items**

- distance measure between two permutations $\pi_1$ and $\pi_2$: $d(\pi_1, \pi_2)$
- parameters:
  $\sigma$, the canonical ordering (identity ordering [1,2,3,...,n])
  $\rho > 0$, a dispersion parameter ($\approx$ distance penalty)

**The probability of an observed permutation $\pi$**

$$P(\pi; \rho, \sigma) = \frac{e^{-\rho * d(\pi, \sigma)}}{\psi(\rho)}$$

# The Generalized Mallows Model (Fligner and Verducci, 1986)

**Generalization to item-specific dispersion parameters**

$\boldsymbol{\rho} = [\rho_1, \rho_2, ...]$ for items in $\boldsymbol{\pi} = [\pi_1, \pi_2, ...]$

$$\begin{aligned} GMM(\pi; \boldsymbol{\rho}, \sigma) &\propto e^{-\sum_i -\rho_i d(\pi_i, \sigma_i)} \\ &\propto \prod_i e^{-\rho_i d(\pi_i, \sigma_i)} \end{aligned}$$

**Relation to our model**

- items $(\pi_i) \hat{=}$ event types
- model event type-specific temporal flexibility

# Contents


1. Introduction
2. Technical Background
3. The Script Model
4. Evaluation & Results
5. Conclusion

# Generative Story I: Ordering Generation

**for** esd $d$ **do**

COOKING PASTA

# Generative Story I: Ordering Generation

draw an event type permutation $\boldsymbol{\pi}$

**for** esd $d$ **do**
$\boldsymbol{\pi} \sim GMM(\boldsymbol{\rho}, \boldsymbol{\nu})$

COOKING PASTA

| $\boldsymbol{\pi}$ | |
|---|---|
| 1 | get |
| 3 | boil |
| 4 | put |
| 7 | wait |
| 2 | grate |
| 5 | add |
| 6 | drain |

Conjugate prior ($GMM_0$)

# Generative Story II: Event Type Generation

realize event type $e$ with success probability $\theta^e$

**for** esd $d$ **do**

$\pi \sim GMM(\rho, \nu)$

$\mathbf{t} : t_e \sim Binomial(\theta^e)$

COOKING PASTA

| $\boldsymbol{\pi}$ | |
|---|---|
| 1 | get |
| 3 | boil |
| 4 | put |
| $\cancel{7}$ | ~~wait~~ |
| 2 | grate |
| $\cancel{5}$ | ~~add~~ |
| 6 | drain |

Conjugate prior (*Beta*)

# Generative Story II: Event Type Generation

realize event type $e$ with success probability $\theta^e$

**for** esd $d$ **do**

$\pi \sim GMM(\rho, \nu)$

$\mathbf{t} : t_e \sim Binomial(\theta^e)$

COOKING PASTA

| t | |
|---|---|
| 1 | get |
| 3 | boil |
| 4 | put |
| 2 | grate |
| 6 | drain |

Conjugate prior ($Beta$)

# Generative Story III: Participant Type Generation

for each event $e \in \mathbf{t}$, realize participant type $p$ with success probability $\varphi_p^e$

**for** esd $d$ **do**
  $\pi \sim GMM(\rho, \nu)$
  $\mathbf{t} : t_e \sim Binomial(\theta^e)$
  **for** event $e \in$ **t do**
    $\mathbf{u_e} : u_e^p \sim Binomial(\varphi_p^e)$

COOKING PASTA

| **t** | |
|---|---|
| 1 | get |
| 3 | boil |
| 4 | put |
| 2 | grate |
| 6 | drain |

| | |
|---|---|
| $\mathbf{u_{get}}$ | ~~pasta~~ ~~water~~ ~~cheese~~ **pot** ~~salt~~ ~~stove~~ ~~strainer~~ |
| $\mathbf{u_{put}}$ | **pasta** ~~water~~ ~~cheese~~ ~~pot~~ ~~salt~~ ~~stove~~ ~~strainer~~ |
| ... | |
| $\mathbf{u_{drain}}$ | ~~pasta~~ **water** ~~cheese~~ **pot** ~~salt~~ ~~stove~~ ~~strainer~~ |

# Generative Story III: Participant Type Generation

for each event $e \in \mathbf{t}$, realize participant type $p$ with success probability $\varphi_p^e$

**for** esd $d$ **do**
  $\pi \sim GMM(\rho, \nu)$
  $\mathbf{t} : t_e \sim Binomial(\theta^e)$
  **for** event $e \in$ **t do**
    $\mathbf{u_e} : u_e^p \sim Binomial(\varphi_p^e)$

COOKING PASTA

| **t** | | **u** |
|---|---|---|
| 1 | get | pot |
| 3 | boil | water |
| 4 | put | pasta |
| 2 | grate | cheese |
| 6 | drain | water, pot |

| | |
|---|---|
| $\mathbf{u_{get}}$ | ~~pasta~~ ~~water~~ ~~cheese~~ **pot** ~~salt~~ ~~stove~~ ~~strainer~~ |
| $\mathbf{u_{put}}$ | **pasta** ~~water~~ ~~cheese~~ ~~pot~~ ~~salt~~ ~~stove~~ ~~strainer~~ |
| ... | |
| $\mathbf{u_{drain}}$ | ~~pasta~~ **water** ~~cheese~~ **pot** ~~salt~~ ~~stove~~ ~~strainer~~ |

# Generative Story IV: Lexical Realization

draw a lexical realizations for each realized event and participant

**for** esd $d$ **do**
  $\pi \sim GMM(\rho, \nu)$
  $\mathbf{t} : t_e \sim Binomial(\theta^e)$
  **for** event $e \in \mathbf{t}$
    $\mathbf{u_e} : u_e^p \sim Binomial(\varphi_p^e)$
  **for** $e \in \mathbf{t}$ **do**
    $w_e \sim Mult(\vartheta_e)$
    **for** $p \in \mathbf{u_e}$ **do**
      $w_p \sim Mult(\varpi_p)$

COOKING PASTA

| t | | u |
|---|---|---|
| 1 | get | pot |
| 3 | boil | water |
| 4 | put | pasta |
| 2 | grate | cheese |
| 6 | drain | water, pot |

# Generative Story IV: Lexical Realization

draw a lexical realizations for each realized event and participant

**for** esd $d$ **do**
  $\pi \sim GMM(\rho, \nu)$
  $\mathbf{t} : t_e \sim Binomial(\theta^e)$
  **for** event $e \in \mathbf{t}$
    $\mathbf{u_e} : u_e^p \sim Binomial(\varphi_p^e)$
  **for** $e \in \mathbf{t}$ **do**
    $w_e \sim Mult(\vartheta_e)$
    **for** $p \in \mathbf{u_e}$ **do**
      $w_p \sim Mult(\varpi_p)$

COOKING PASTA

| | |
|---|---|
| “fetch | saucepan” |
| “boil | water” |
| “add | noodles” |
| “grate | cheese“ |
| ”drain | water from pot “ |

# Generative Story IV: Lexical Realization

draw a lexical realizations for each realized event and participant

**for** esd $d$ **do**
  $\pi \sim GMM(\rho, \nu)$
  $\mathbf{t} : t_e \sim Binomial(\theta^e)$
  **for** event $e \in \mathbf{t}$
    $\mathbf{u_e} : u_e^p \sim Binomial(\varphi_p^e)$
  **for** $e \in \mathbf{t}$ **do**
    $w_e \sim Mult(\vartheta_e)$
    **for** $p \in \mathbf{u_e}$ **do**
      $w_p \sim Mult(\varpi_p)$

COOKING PASTA

| | |
|---|---|
| “fetch | saucepan” |
| “boil | water” |
| “add | noodles” |
| “grate | cheese“ |
| ”drain | water from pot “ |

**Informed asymmetric Dirichlet priors**

# Informed Asymmetric Dirichlet Priors

Tie together the prior values ("pseudo counts") of semantically related words

Multivariate Normal
$\mathcal{N}(0, \Sigma)$

# Informed Asymmetric Dirichlet Priors

Tie together the prior values (”pseudo counts“) of semantically related words

Multivariate Normal
$\mathcal{N}(0, \Sigma)$

| | saucepan | pot | cheese | noodles | pasta |
|---|---|---|---|---|---|
| saucepan | 7 | 6 | 1 | 0 | 1 |
| pot | 6 | 12 | 0 | 0 | 1 |
| cheese | 1 | 0 | 13 | 3 | 3 |
| noodles | 0 | 0 | 3 | 7 | 5 |
| pasta | 1 | 1 | 3 | 5 | 6 |
| ... | | | | | |

# Informed Asymmetric Dirichlet Priors

Tie together the prior values ("pseudo counts") of semantically related words

Multivariate Normal
$\mathcal{N}(0, \Sigma)$

semantic similarity:
♯ shared WordNet synsets

| | saucepan | pot | cheese | noodles | pasta |
|---|---|---|---|---|---|
| saucepan | 7 | 6 | 1 | 0 | 1 |
| pot | 6 | 12 | 0 | 0 | 1 |
| cheese | 1 | 0 | 13 | 3 | 3 |
| noodles | 0 | 0 | 3 | 7 | 5 |
| pasta | 1 | 1 | 3 | 5 | 6 |
| ... | | | | | |

$\delta \sim \mathcal{N}(0, \Sigma)$ $\phi \sim Dirichlet(\delta)$ $w \sim Multinomial(\phi)$

# Inference

**Collapsed Gibbs Sampling for approximate inference Slice**

**Sampling for continuous distributions** $GMM$ **and** $\mathcal{N}(0, \Sigma)$

**Parameters to be estimated (after collapsing)**

- latent ESD labels $\mathbf{z} = \{\mathbf{t}, \mathbf{u}, \boldsymbol{\pi}\}$
- GMM dispersion parameter $\boldsymbol{\rho}$
- language model parameters $\boldsymbol{\delta}_{(partic)}, \boldsymbol{\gamma}_{(event)}$

# Contents

1 Introduction

2 Technical Background

3 The Script Model

4 Evaluation & Results

5 Conclusion

# Data I

## Collection of ESDs (Regneri et al., 2010)

- sets of explicit descriptions of event sequences
- created from non-experts via web experiments

| Scenario Name | ♯ESDs | Avg ♯events |
|---|---|---|
| Answer the telephone | 55 | 4.47 |
| Buy from vending machine | 32 | 4.53 |
| Make scrambled eggs | 20 | 10.3 |
| Eat in fast food restaurant | 15 | 8.93 |
| Take a shower | 21 | 11.29 |
| ... | ... | ... |

- test set (10 scenarios)
- separate development set (5 scenarios)

# Evaluation Setup (Regneri et al., 2010)

**Binary event paraphrase classification**

- [`get pot` , `fetch saucepan`] $\Rightarrow$ true
- [`add pasta` , `add water`] $\Rightarrow$ false

**Binary follow-up classification**

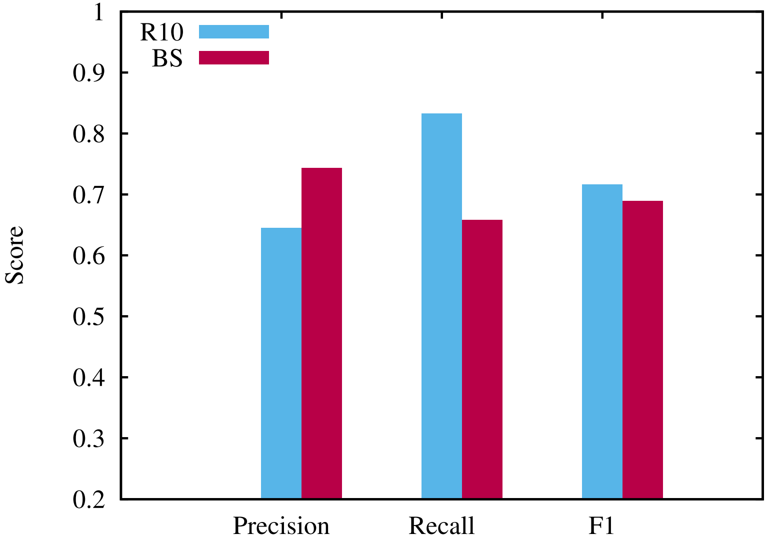
- [`fetch pot` , `put pasta into pot`] $\Rightarrow$ true
- [`put pasta into pot` , `fetch pot`] $\Rightarrow$ false

**Metric**

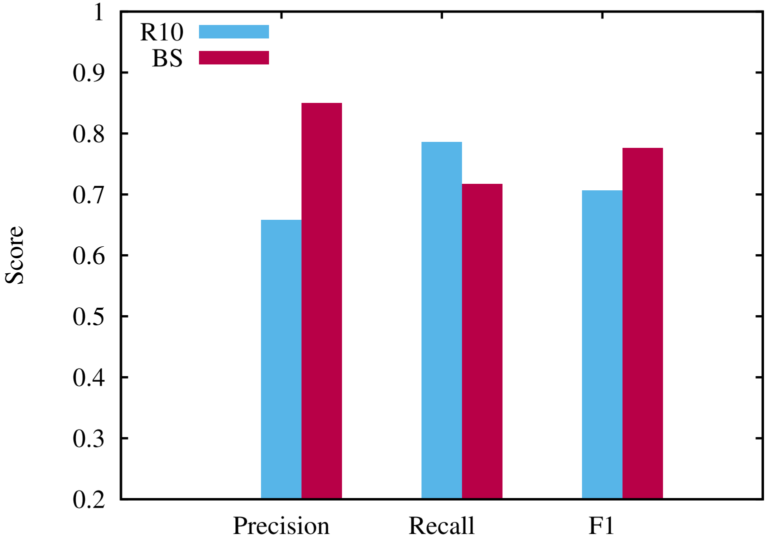
$$precision = \frac{true_{system} \cap true_{gold}}{true_{system}}$$

$$recall = \frac{true_{system} \cap true_{gold}}{true_{gold}} \qquad F = \frac{2 * precision * recall}{precision + recall}$$

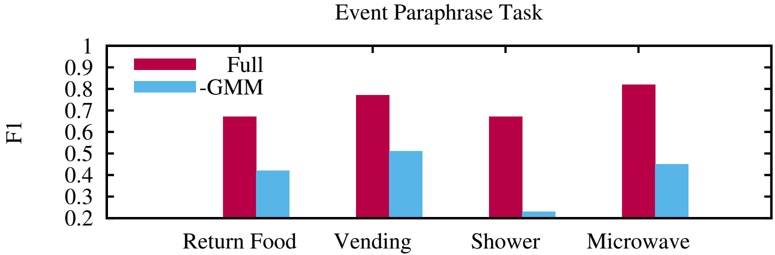
# The Event Paraphrase Task

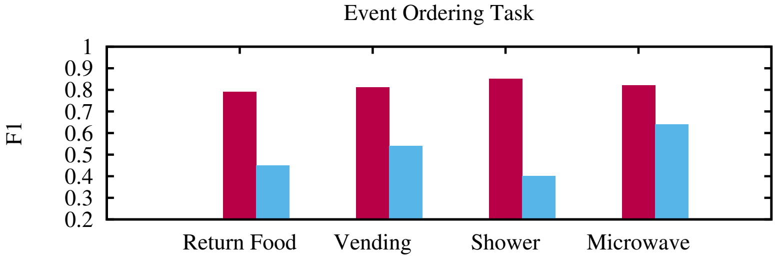


# The Event Ordering Task

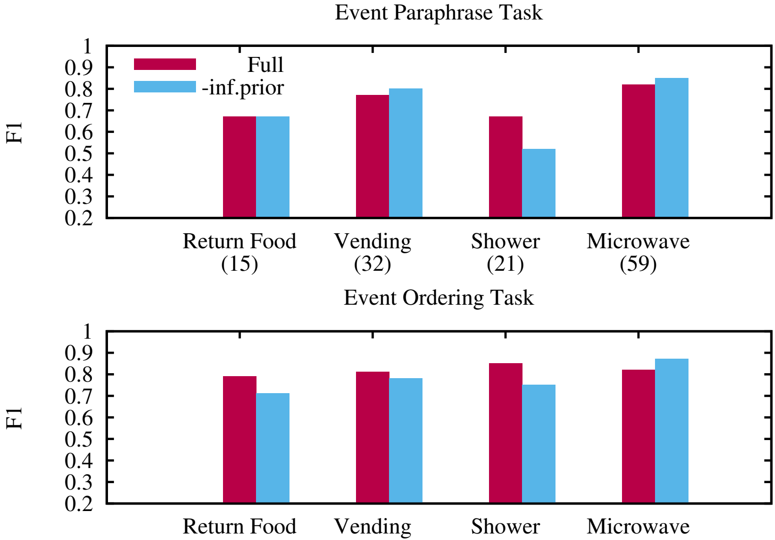


# Influence of Model Components



# Influence of Model Components



# Induced Clustering

**Cooking food in the microwave**

$\{$get$\}$ $\rightarrow$ $\{$open,take$\}$ $\rightarrow$ $\{$put,place$\}$ $\rightarrow$ $\{$close$\}$

$\rightarrow$ $\{$set,select,enter,turn$\}$ $\rightarrow$ $\{$start$\}$ $\rightarrow$ $\{$wait$\}$

$\rightarrow$ $\{$remove,take,open$\}$ $\rightarrow$ $\{$push,press,turn$\}$

# Contents

1 Introduction

2 Technical Background

3 The Script Model

4 Evaluation & Results

5 Conclusion

# Conclusion

**A hierarchical Bayesian Script model**

- joint model of event types and ordering constraints
- competitive performance with a recent pipeline-based model
- inclusion of word similarity knowledge as correlations in the language model priors
- explicitly target apparent script characteristics (event optionality, event type-specific temporal flexibility)
- GMM as an effective model for event ordering constraints

# Conclusion

**A hierarchical Bayesian Script model**

- joint model of event types and ordering constraints
- competitive performance with a recent pipeline-based model
- inclusion of word similarity knowledge as correlations in the language model priors
- explicitly target apparent script characteristics (event optionality, event type-specific temporal flexibility)
- GMM as an effective model for event ordering constraints

**Thank you!**

# References I

A. Barr and E.A. Feigenbaum. 1986. *The handbook of artificial intelligence. 1 (1981)*. The Handbook of Artificial Intelligence. Addison-Wesley.

D. Blei and J. Lafferty. 2006. Correlated topic models. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 147–154. MIT Press, Cambridge, MA.

N. Chambers and D. Jurafsky. 2008. Unsupervised learning of narrative event chains. In *Proceedings of ACL-08: HLT*, pages 789–797. Association for Computational Linguistics, Columbus, Ohio.

# References II

N. Chambers and D. Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proceedings of ACL-09 and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 602–610. Association for Computational Linguistics, Stroudsburg, PA, USA.

H. Chen, S. R. K. Branavan, R. Barzilay, and D. R. Karger. 2009. Content modeling using latent permutations. *J. Artif. Int. Res.*, 36(1):129–163.

M. Fligner and J. Verducci. 1986. Distance based ranking models. *Journal of the Royal Statistical Society, Series B*, 48:359–369.

C. L. Mallows. 1957. Non-null ranking models. *Biometrika*, 44:114–130.

# References III

M. Regneri, A. Koller, and M. Pinkal. 2010. Learning script knowledge with web experiments. In *Proceedings of ACL 2010*. Association for Computational Linguistics, Uppsala, Sweden.

M. Regneri, A. Koller, J. Ruppenhofer, and M. Pinkal. 2011. Learning script participants from unlabeled data. In *Proceedings of RANLP 2011*. Hissar, Bulgaria.

R. C. Schank and R. P. Abelson. 1975. Scripts, plans and knowledge. In PN Johnson-Laird and PC Wason, editors, *Thinking: Readings in Cognitive Science, Proceedings of the Fourth International Joint Conference on Artificial Intelligence*, pages 151–157.