School of
Information Systems

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Modeling Syntactic Structures of Topics with a Nested HMM-LDA

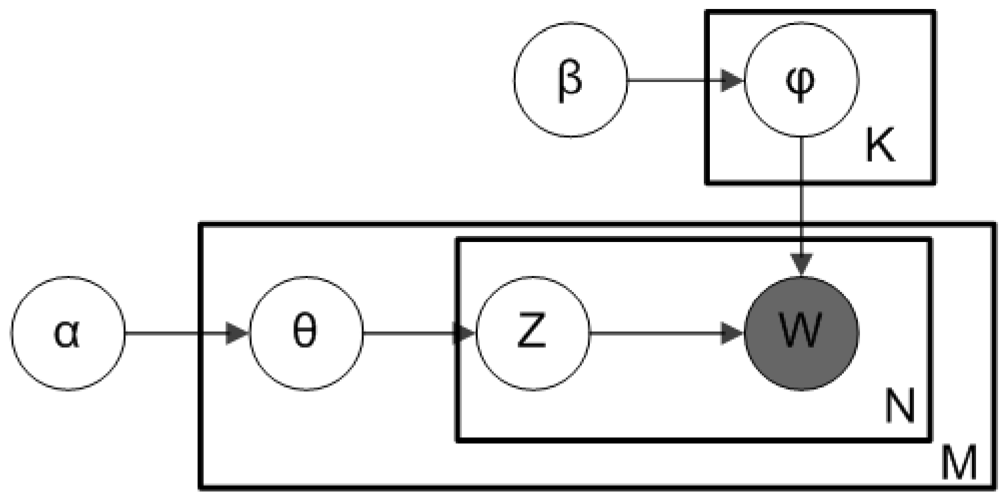
Jing Jiang

Singapore Management University

Topic Models

- A generative model for discovering hidden topics from documents
- Topics are represented as word distributions

This makes full synchrony of activated units the default condition in the model cortex, as in Brown s model [Brown and Cooke, 1996], so that the background activation is coherent, and can be read into high order cortical levels which synchronize with it.



How to Interpret Topics?

- List the top-*K* frequent words
  - Not easy to interpret
  - How are the top words related to each other?
- Our method: model the syntactic structures of topics using a combination of hidden Markov models (HMM) and topic models (LDA)
- A preliminary solution towards meaningful representations of topics

Related Work on Syntactic LDA

- Similar to / based on [Griffiths et al. 05]
  - More general, with multiple "semantic classes"
- [Boyd-Graber & Blei 09]
  - Combines parse trees with LDA
  - Expensive to obtain parse trees for large text collections
- [Gruber et al. 07]
  - Combines HMM with LDA
  - Does not model syntax

HMM to Model Syntax

- In natural language sentences, the syntactic class of a word occurrence (noun, verb, adjective, adverb, preposition, etc.) depends on its context
- Transitions between syntactic classes follow some structure
- HMMs can be used to model these transitions
  – HMM-based part-of-speech tagger

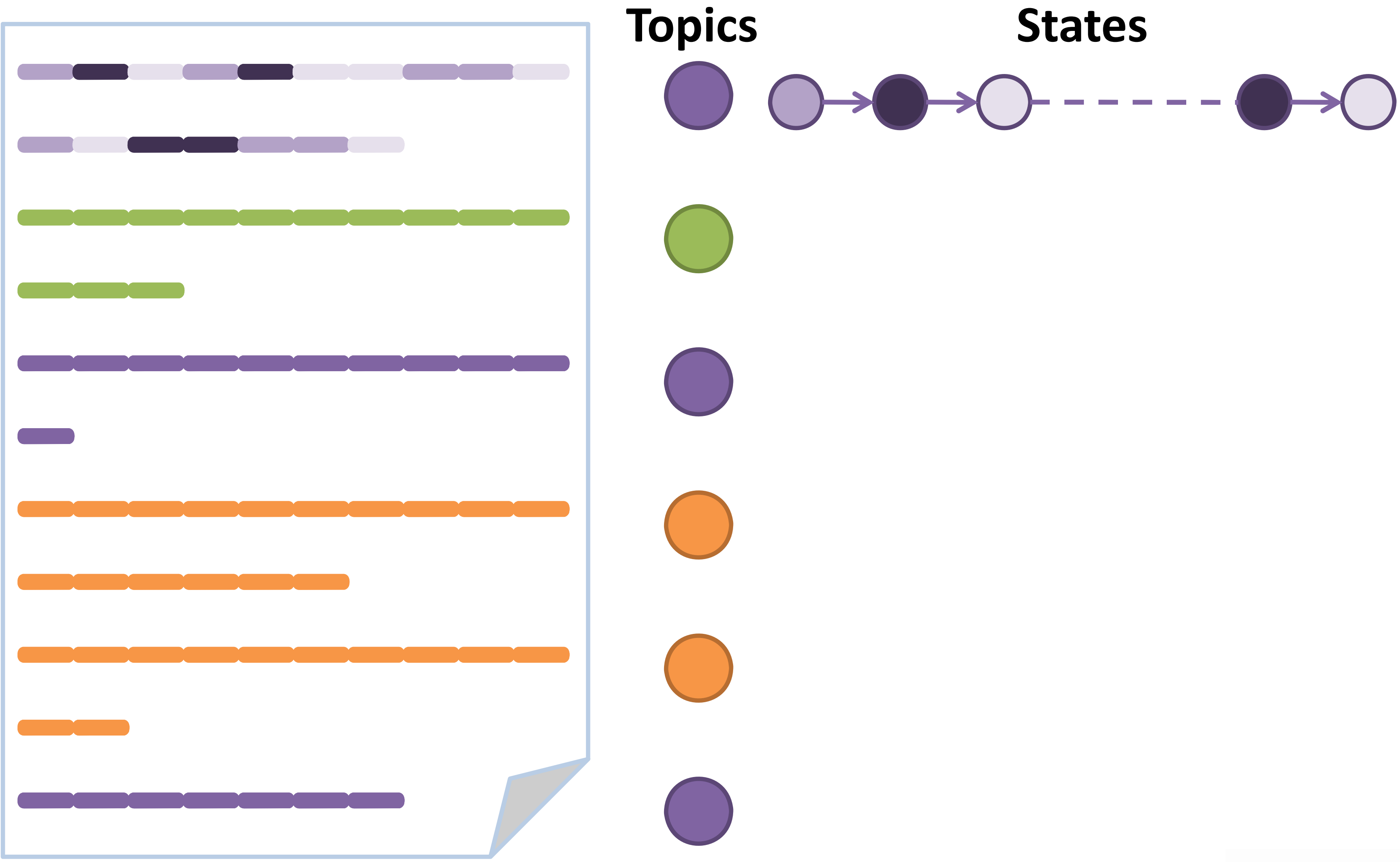
Overview of Our Model

- Assumptions
  - A topic is represented as an HMM
  - *C* ***Content*** states: convey semantic meanings of topics (likely to be nouns, verbs, adjectives, etc.)
  - *F* ***Functional*** states: serve linguistic functions (e.g. prepositions and articles)
    - Word distributions of these functional states are shared among topics
  - Each document has a mixture of topics
  - Each sentence is generated from a single topic

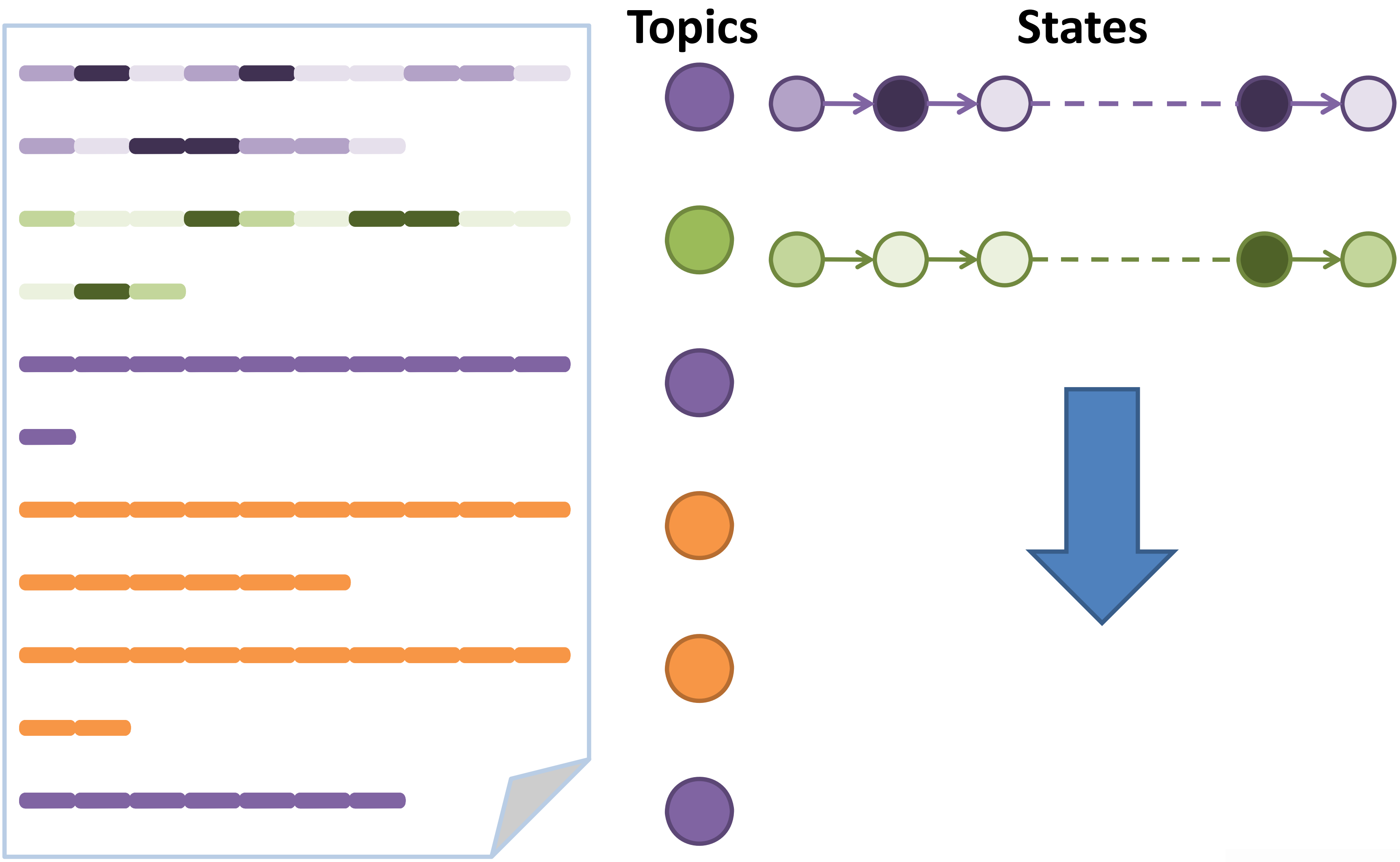
Overview of Our Model

Overview of Our Model

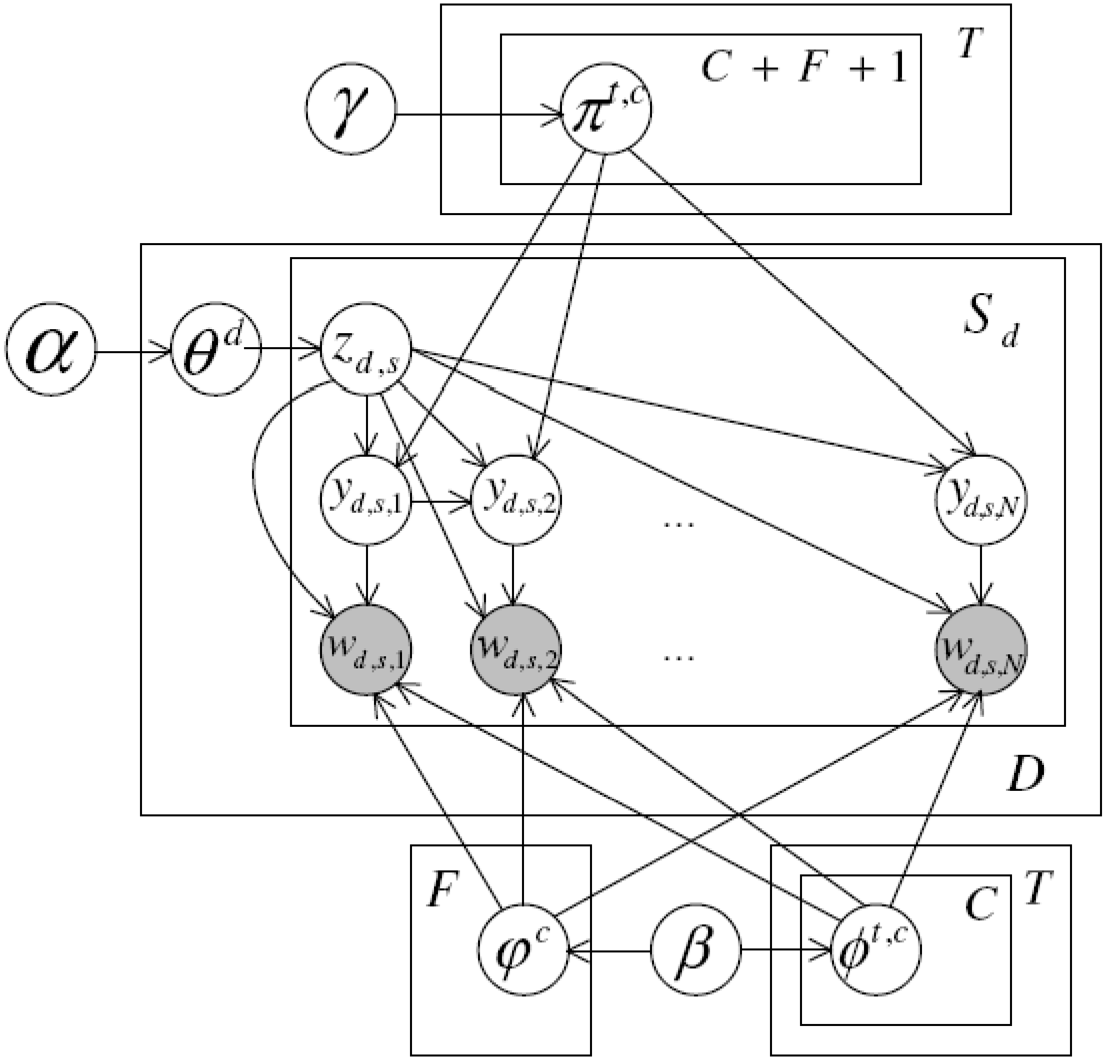
Overview of Our Model



Overview of Our Model



The n-HMM-LDA Model



School of
Information Systems

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Document Generation Process

1) For each global functional state $c = C+1, \ldots, C+F$, draw $\psi^c \sim \mathrm{Dir}(\beta)$
2) For each topic $t = 1, \ldots, T$
   a) For each state $c = 0, \ldots, G$, draw $\pi^{t,c} \sim \mathrm{Dir}(\gamma)$
   b) For each local content state $c = 1, \ldots, C$, draw $\phi^{t,c} \sim \mathrm{Dir}(\beta)$
3) For each document $d = 1, \ldots, D$
   a) Draw $\theta^d \sim \mathrm{Dir}(\alpha)$
   b) For each sentence $s = 1, \ldots, S_d$
      i) Draw $z_{d,s} \sim \mathrm{Multi}(\theta^d)$
      ii) For each word $n = 1, \ldots, N_{d,s}$
         A) Draw $y_{d,s,n} \sim \mathrm{Multi}(\pi^{z_{d,s}, y_{d,s,n-1}})$
         B) Draw $w_{d,s,n} \sim \mathrm{Multi}(\phi^{z_{d,s}, y_{d,s,n}})$ if $y_{d,s,n} \leq C$ or $w_{d,s,n} \sim \mathrm{Multi}(\psi^{y_{d,s,n}})$ if $y_{d,s,n} > C$

Sample topics and transition probabilities

Information Systems

SMU SINGAPORE MANAGEMENT UNIVERSITY

Document Generation Process

1) For each global functional state $c = C+1, \ldots, C+F$, draw $\psi^c \sim \mathrm{Dir}(\beta)$
2) For each topic $t = 1, \ldots, T$
   a) For each state $c = 0, \ldots, G$, draw $\pi^{t,c} \sim \mathrm{Dir}(\gamma)$
   b) For each local content state $c = 1, \ldots, C$, draw $\phi^{t,c} \sim \mathrm{Dir}(\beta)$
3) For each document $d = 1, \ldots, D$
   a) Draw $\theta^d \sim \mathrm{Dir}(\alpha)$
   b) For each sentence $s = 1, \ldots, S_d$
      i) Draw $z_{d,s} \sim \mathrm{Multi}(\theta^d)$
      ii) For each word $n = 1, \ldots, N_{d,s}$
         A) Draw $y_{d,s,n} \sim \mathrm{Multi}(\pi^{z_{d,s}, y_{d,s,n-1}})$
         B) Draw $w_{d,s,n} \sim \mathrm{Multi}(\phi^{z_{d,s}, y_{d,s,n}})$ if $y_{d,s,n} \leq C$ or $w_{d,s,n} \sim \mathrm{Multi}(\psi^{y_{d,s,n}})$ if $y_{d,s,n} > C$

Sample a topic distribution for the document (same as in LDA)

School of Information Systems

SMU Singapore Management University

Document Generation Process

1) For each global functional state $c = C+1, \ldots, C+F$, draw $\psi^c \sim \mathrm{Dir}(\beta)$
2) For each topic $t = 1, \ldots, T$
   a) For each state $c = 0, \ldots, G$, draw $\pi^{t,c} \sim \mathrm{Dir}(\gamma)$
   b) For each local content state $c = 1, \ldots, C$, draw $\phi^{t,c} \sim \mathrm{Dir}(\beta)$
3) For each document $d = 1, \ldots, D$
   a) Draw $\theta^d \sim \mathrm{Dir}(\alpha)$
   b) For each sentence $s = 1, \ldots, S_d$
      i) Draw $z_{d,s} \sim \mathrm{Multi}(\theta^d)$
      ii) For each word $n = 1, \ldots, N_{d,s}$
         A) Draw $y_{d,s,n} \sim \mathrm{Multi}(\pi^{z_{d,s}, y_{d,s,n-1}})$
         B) Draw $w_{d,s,n} \sim \mathrm{Multi}(\phi^{z_{d,s}, y_{d,s,n}})$ if $y_{d,s,n} \leq C$ or $w_{d,s,n} \sim \mathrm{Multi}(\psi^{y_{d,s,n}})$ if $y_{d,s,n} > C$

Sample a topic for a sentence

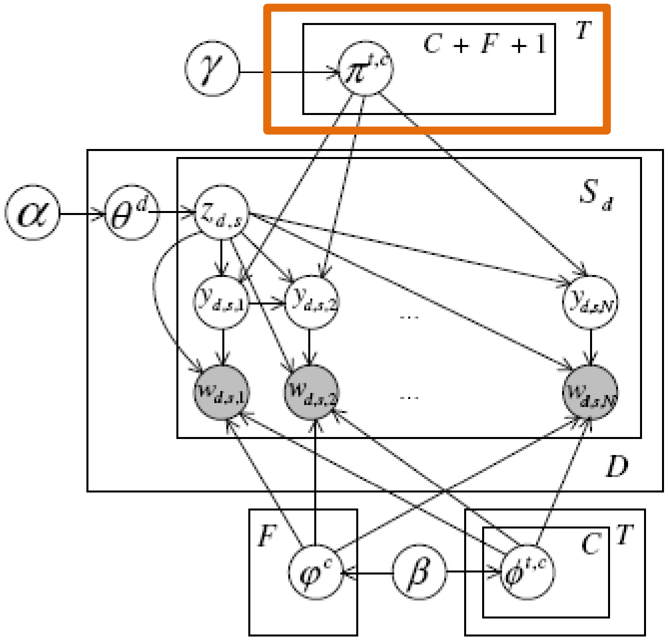
Document Generation Process

1) For each global functional state $c = C+1, \ldots, C+F$, draw $\psi^c \sim \mathrm{Dir}(\beta)$
2) For each topic $t = 1, \ldots, T$
   a) For each state $c = 0, \ldots, G$, draw $\pi^{t,c} \sim \mathrm{Dir}(\gamma)$
   b) For each local content state $c = 1, \ldots, C$, draw $\phi^{t,c} \sim \mathrm{Dir}(\beta)$
3) For each document $d = 1, \ldots, D$
   a) Draw $\theta^d \sim \mathrm{Dir}(\alpha)$
   b) For each sentence $s = 1, \ldots, S_d$
      i) Draw $z_{d,s} \sim \mathrm{Multi}(\theta^d)$
      ii) For each word $n = 1, \ldots, N_{d,s}$
         A) Draw $y_{d,s,n} \sim \mathrm{Multi}(\pi^{z_{d,s}, y_{d,s,n-1}})$
         B) Draw $w_{d,s,n} \sim \mathrm{Multi}(\phi^{z_{d,s}, y_{d,s,n}})$ if $y_{d,s,n} \leq C$ or $w_{d,s,n} \sim \mathrm{Multi}(\psi^{y_{d,s,n}})$ if $y_{d,s,n} > C$

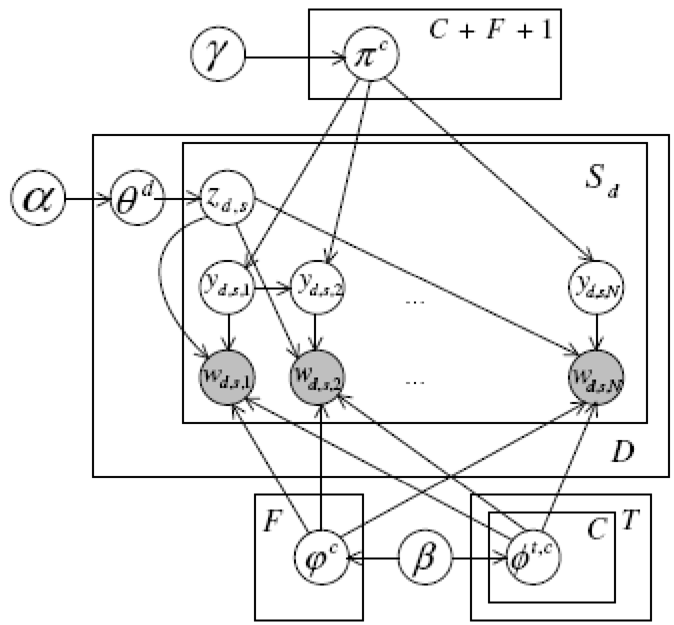
Generate the words in the sentence using the HMM corresponding to this topic

Variations

- Transition probabilities between states can be either topic-specific (left) or shared by all topics (right)



(a)



(b)

Model Inference: Gibbs Sampling

- Sample a topic for a sentence

$$
\begin{aligned}
& p(z_{d,s} = t | \mathbf{z}_{\neg\{d,s\}}, \mathbf{y}, \mathbf{w}) \\
\propto \; & \frac{M_{(t)}^{d} + \alpha}{M_{(\cdot)}^{d} + T\alpha} \cdot \prod_{c_1=0}^{G} \frac{\Gamma\left(M_{(\cdot)}^{t,c_1} + G\gamma\right)}{\Gamma\left(M_{(\cdot)}^{t,c_1} + L_{(\cdot)}^{c_1} + G\gamma\right)} \prod_{c_2=1}^{G} \frac{\Gamma\left(M_{(c_2)}^{t,c_1} + L_{(c_2)}^{c_1} + \gamma\right)}{\Gamma\left(M_{(c_2)}^{t,c_1} + \gamma\right)} \\
& \cdot \prod_{c=1}^{C} \frac{\Gamma\left(M_{(\cdot)}^{t,c} + V\beta\right)}{\Gamma\left(M_{(\cdot)}^{t,c} + L_{(\cdot)}^{c} + V\beta\right)} \prod_{v=1}^{V} \frac{\Gamma\left(M_{(v)}^{t,c} + L_{(v)}^{c} + \beta\right)}{\Gamma\left(M_{(v)}^{t,c} + \beta\right)}
\end{aligned}
$$

- Sample a state for a word

$$
\begin{aligned}
& P(y_{d,s,n} = c | \mathbf{z}, \mathbf{y}_{\neg\{d,s,n\}}, \mathbf{w}) \\
\propto \; & \left(M_{(c)}^{z_{d,s}, y_{d,s,n-1}} + \gamma\right) \cdot \frac{M(c) + \beta}{M_{(\cdot)}(c) + V\beta} \\
& \cdot \frac{M_{(y_{d,s,n+1})}^{z_{d,s},c} + \delta(y_{d,s,n-1}, c) \cdot \delta(c, y_{d,s,n+1}) + \gamma}{M_{(\cdot)}^{z_{d,s},c} + \delta(y_{d,s,n-1}, c) + C\gamma}
\end{aligned}
$$

School of
Information Systems

SMU SINGAPORE MANAGEMENT UNIVERSITY

Experiments – Data Sets

- NIPS publications (downloaded from http://nips.djvuzone.org/txt.html)
- Reuters-21578

	Date Sets	
	NIPS Publications*	Reuters-21578
Vocabulary	18,864	10,739
Words	5,305,230	1,460,666
documents for training	1314	8052
documents for testing	618	2665

Quantitative Evaluation

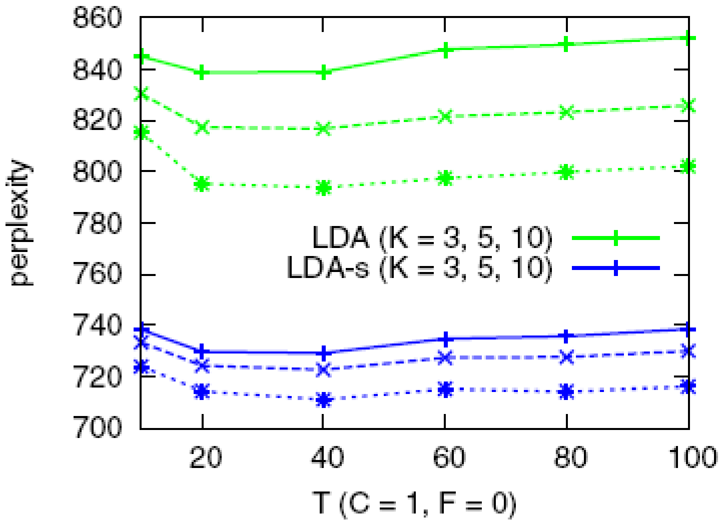
- Perplexity: a commonly used metric for the generalization power of language models

$$\text{perplexity}(\mathcal{D}_{\texttt{test}}) = \exp\left(\frac{-\log p(\mathcal{D}_{\texttt{test}})}{|\mathcal{D}_{\texttt{test}}|}\right).$$

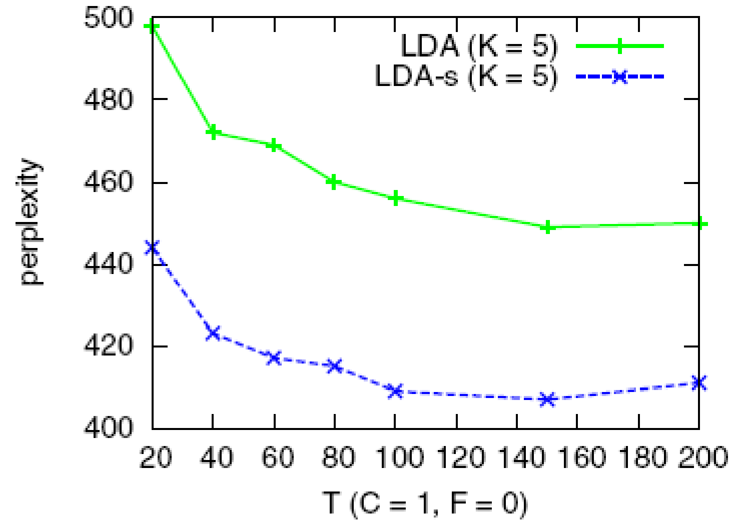
- For a test document, observe the first *K* sentences and predict the remaining sentences

LDA vs. LDA-s

- LDA-s: n-HMM-LDA with a single state for each HMM.
  - Same as standard LDA with each sentence having



NIPS

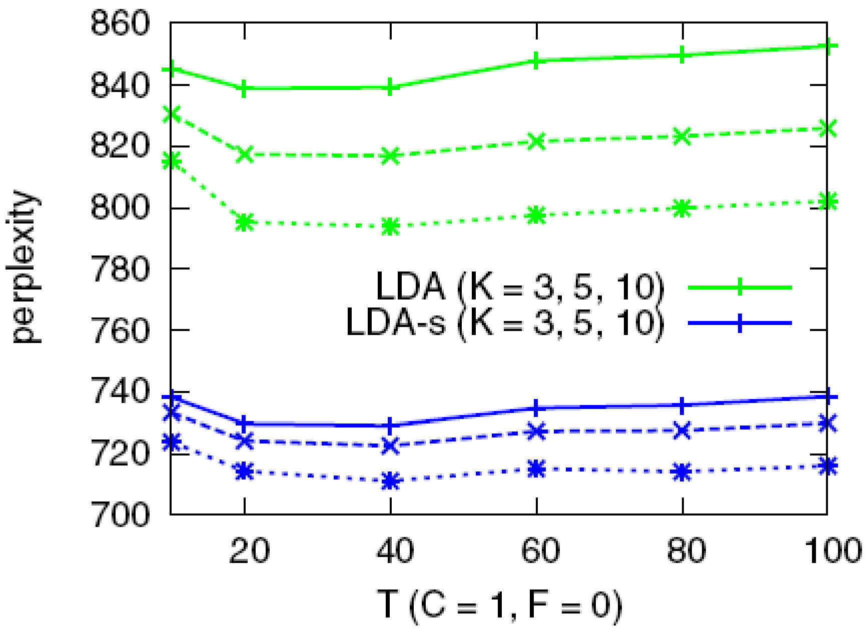


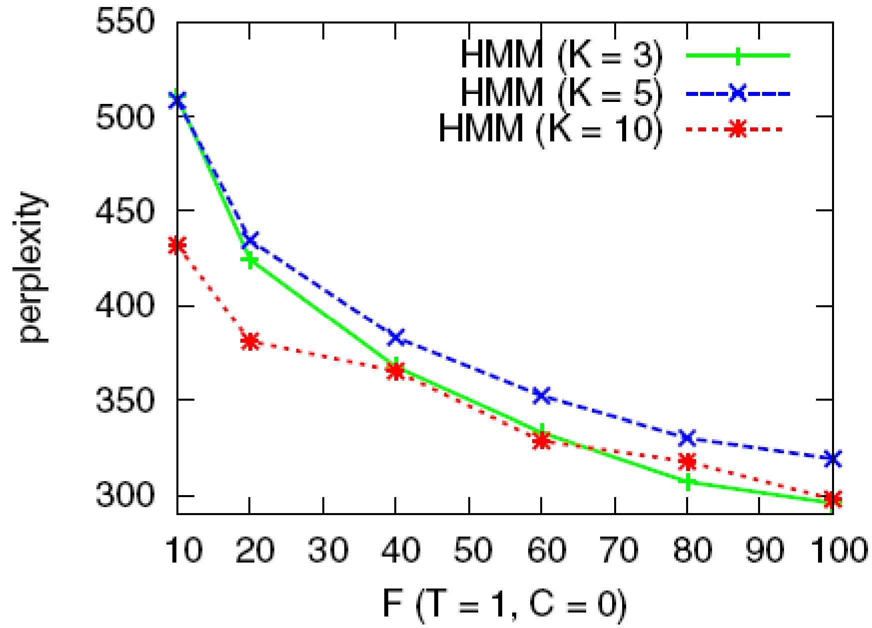
Reuters

One-topic-per-sentence assumption helps.

HMM

- Achieves much lower perplexity, but cannot be used to discover topics

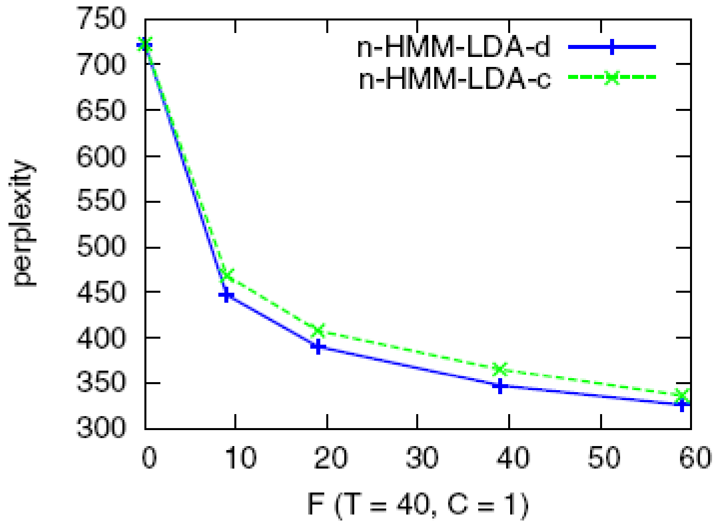




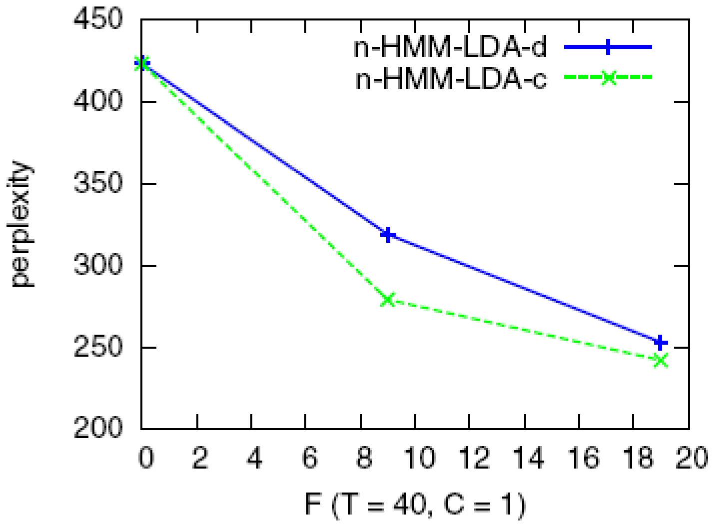
NIPS

Increase Number of Functional States

- Fixing the number of content states to 1 and the number of topics to 40



NIPS



Reuters

More functional states decreases perplexity.

Qualitative Evaluation

- Use the top frequent words to represent a topic/state

School of
Information Systems

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Sample Topics/States from LDA/HMM

LDA	the of a figure i local two in	the algorithm of and to for gradient in	the a of and to * for on	the of a in and time signal frequency
HMM	used shown trained given * found obtained defined	of between and in for over take where	the a this an each these our its	function units set learning space layer functions recognition

NIPS

School of
Information Systems

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Sample States from n-HMM-LDA-d

n-HMM-LDA-d (content states)	voltage circuit * current output input v gate	word * character characters recognition words training segmentation	in neural morgan systems processing information kaufmann san	* chip and analog digital bit vlsi hardware
n-HMM-LDA-d (functional states)	we it which that this there they and	in for by with on as from at	network model algorithm system results problem approach method	can will may have would must should could

NIPS

School of Information Systems

Different Content States

THE TOP-10 WORDS FROM THE TWO CONTENT STATES OF TWO TOPICS.

A topic from NIPS		A topic from Reuters	
cortex	*	and	united
*	visual	gas	states
cells	orientation	oil	union
neurons	cortical	*	soviet
connections	ocular	natural	kong
activity	receptive	exploration	hong
organization	eye	development	*
maps	cell	ltd	south
patterns	lateral	corp	party
fields	dominance	production	san

School of
Information Systems

Case Study (LDA)

This makes full synchrony of activated units the default condition in the model cortex, as in Brown s model [Brown and Cooke, 1996], so that the background activation is coherent, and can be read into high order cortical levels which synchronize with it.

is	the	of
the	of	the
that	a	in
be	signal	and
are	and	cells
can	to	to
one	in	cell
it	frequency	model
to	is	a
for	*	response

School of
Information Systems

Case Study (n-HMM-LDA)

This makes full synchrony of **activated** **units** the default condition in the **model** **cortex**, as in Brown s **model** [Brown and Cooke, 1996], so that the **background** **activation** is coherent, and can be read into high order **cortical** **levels** which synchronize with it.

*
receptive
synaptic
inhibitory
head
excitatory
direction
cell
visual
pyramidal

cells
cell
*
neurons
field
input
response
model
activity
synapses

Case Study (Comparison)

This makes full synchrony of activated units the default condition in the model cortex, as in Brown s model [Brown and Cooke, 1996], so that the background activation is coherent, and can be read into high order cortical levels which synchronize with it.

LDA

This makes full synchrony of **activated units** the default condition in the **model cortex**, as in Brown s **model** [Brown and Cooke, 1996], so that the **background activation** is coherent, and can be read into high order **cortical levels** which synchronize with it.

n-HMM-LDA

School of
Information Systems

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Conclusion

- We proposed a nested-HMM-LDA to model the syntactic structures of topics
  - Extension of [Griffiths et al. 05]
- Experiments on two data sets show that
  - The model achieves perplexity between LDA and HMM
  - The model can provide more insights into the structures of topics than standard LDA

Thank You!

- Questions?

School of
Information Systems

SMU
SINGAPORE MANAGEMENT UNIVERSITY