

T-FaNT2

Nonparametric Bayesian Approach for Distributional Hypothesis

2008/02/13

Ossei Sato

Joint work with

Minoru Yoshida

Hiroshi Nakagawa

University of Tokyo, Nakagawa lab

1

Table of contents

- **Background**
 - Distributional Hypothesis
- Nonparametric Bayesian Modeling
- Proposed Model
- Experimental results

2

Distributional Hypothesis

[Harris, 1954]

Words that occur in the same context
tend to have a similar meaning

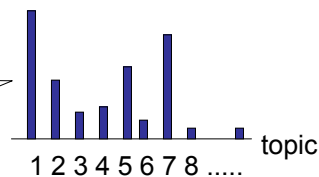
- Linguistics is the scientific study of language
- Linguistics is one of the scientific fields

context

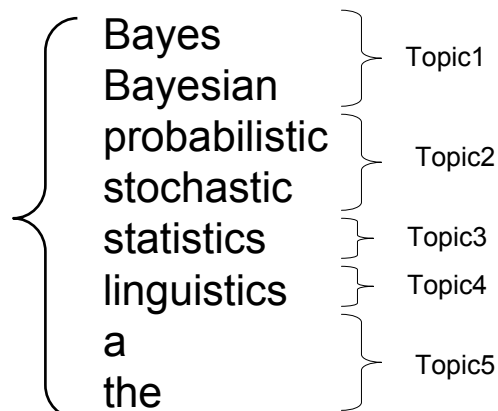
Why do these words have similar meanings?

3

Topic distribution of
context "is based on"

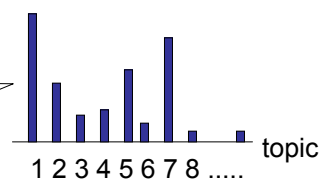


is based on



4

Topic distribution of context "is based on"



is based on

Bayes
Bayesian

Topic1

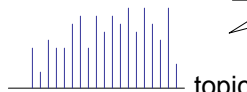
Topic2

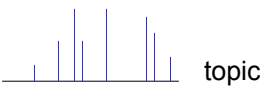
Extended interpretation

Distributional Hypothesis
contexts have their own topic distribution.
the words relating to the same topic have similar meanings

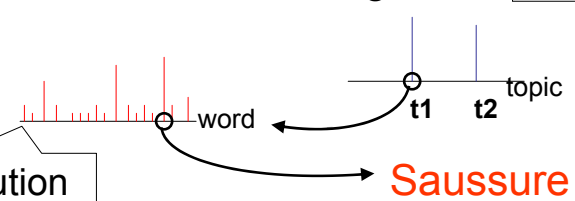
the

5

is  topic

Swiss linguist is  topic

A famous Geneva-born Swiss linguist is

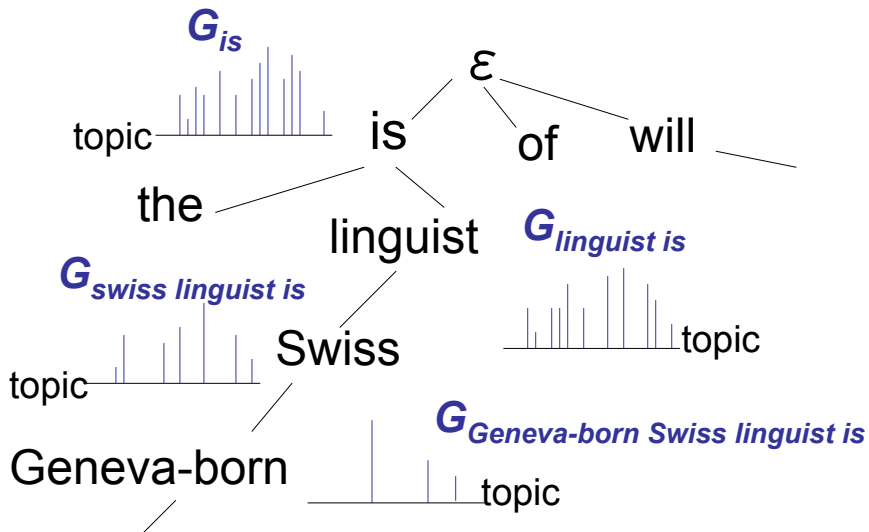
 word

word distribution related to topic:t1

Saussure

6

Context-based suffix tree



7

- How many topics are there?
- How to model the dependency of topic distributions.



Nonparametric Bayesian Modeling

8

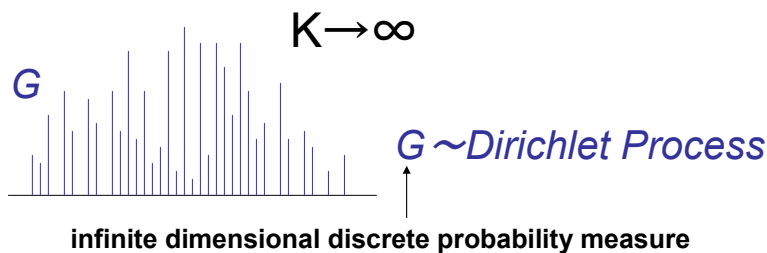
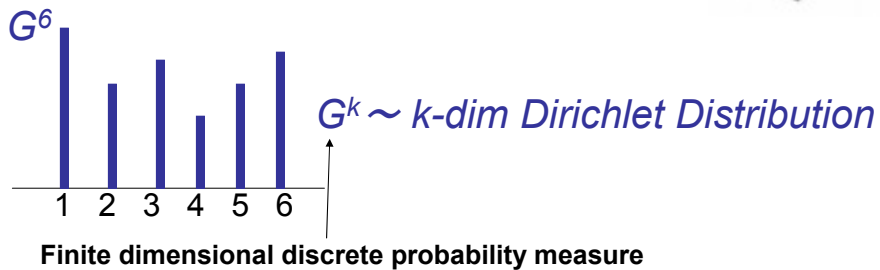
Table of contents

- Background
- **Nonparametric Bayesian Modeling**
 - Dirichlet Process
 - Pitman-Yor Process
 - Hierarchical Dirichlet Process
 - Hierarchical Pitman-Yor Process
- Proposed Model
- Experiment

9



Dice-K(=6) Distribution

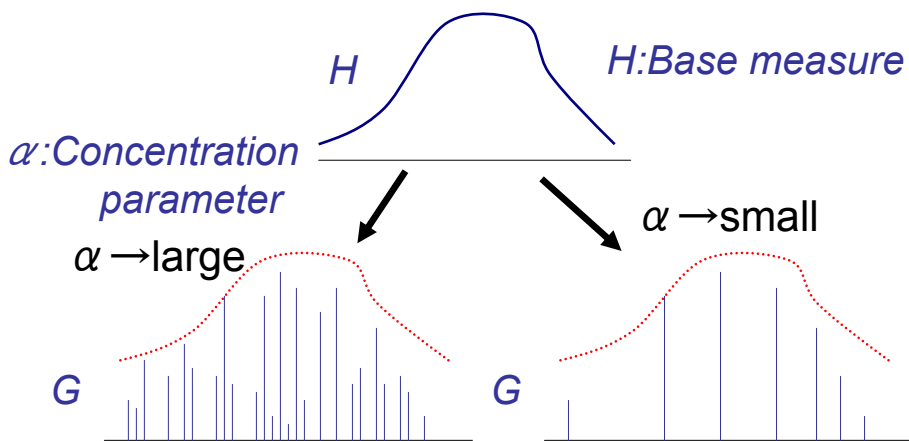


10

Dirichlet Process

[Ferguson, 1973]

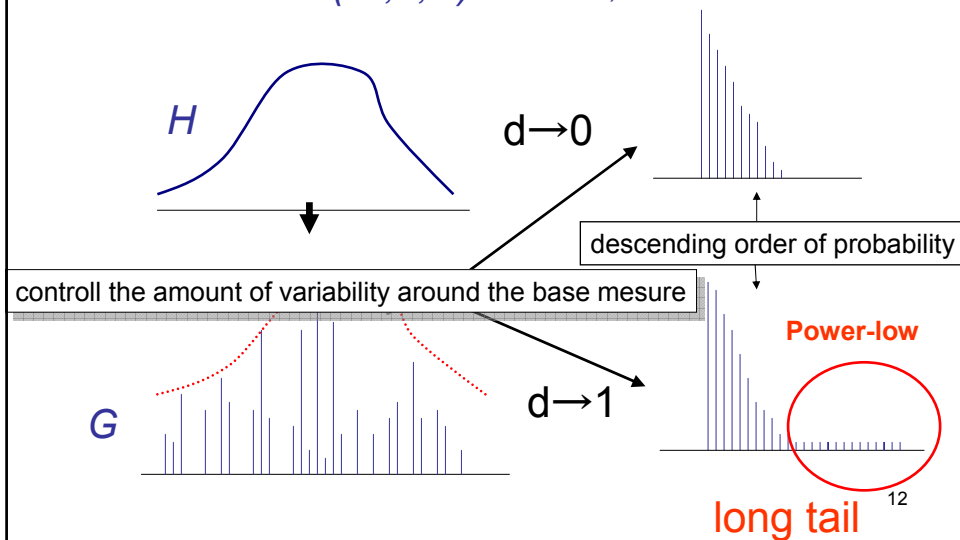
$$G \sim DP(\alpha, H) \quad 0 < \alpha$$



11

Pitman-Yor Process

$$G \sim PY(\alpha, d, H) \quad -d < \alpha, \quad 0 \leq d < 1$$



12

- Hierarchical Dirichlet Process [Teh,2003]

$$G_0 \sim DP(\alpha_0, H)$$

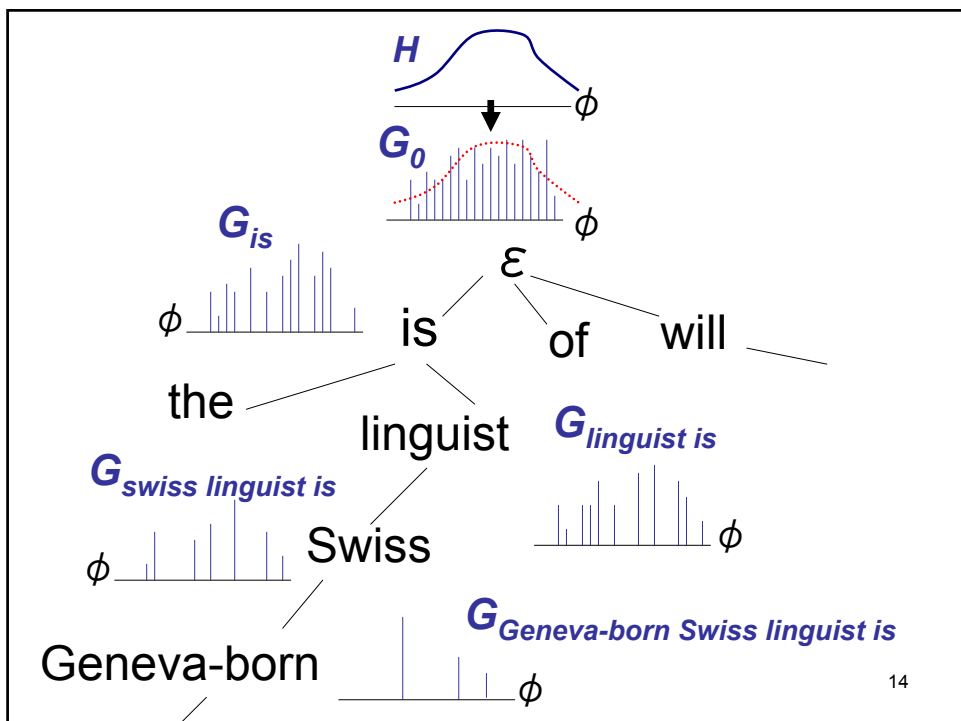
$$G_1 \sim DP(\alpha_1, G_0)$$

- Hierarchical Pitman-Yor Process [Teh,2006]

$$G_0 \sim PY(\alpha_0, d_0, H)$$

$$G_1 \sim PY(\alpha_1, d_1, G_0)$$

13

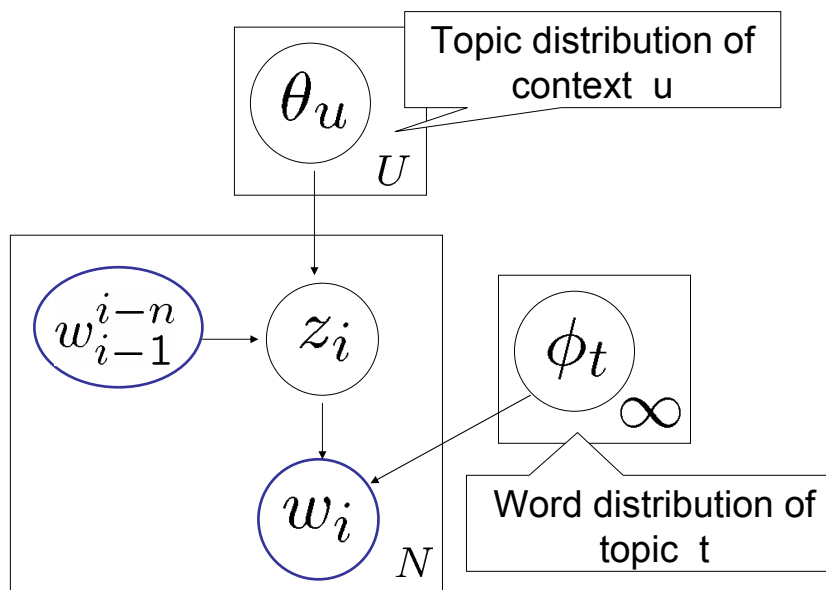


14

Table of contents

- Background
- Nonparametric Bayesian Modeling
- Proposed Model
 - Model 1
 - Model 2
- Experiment

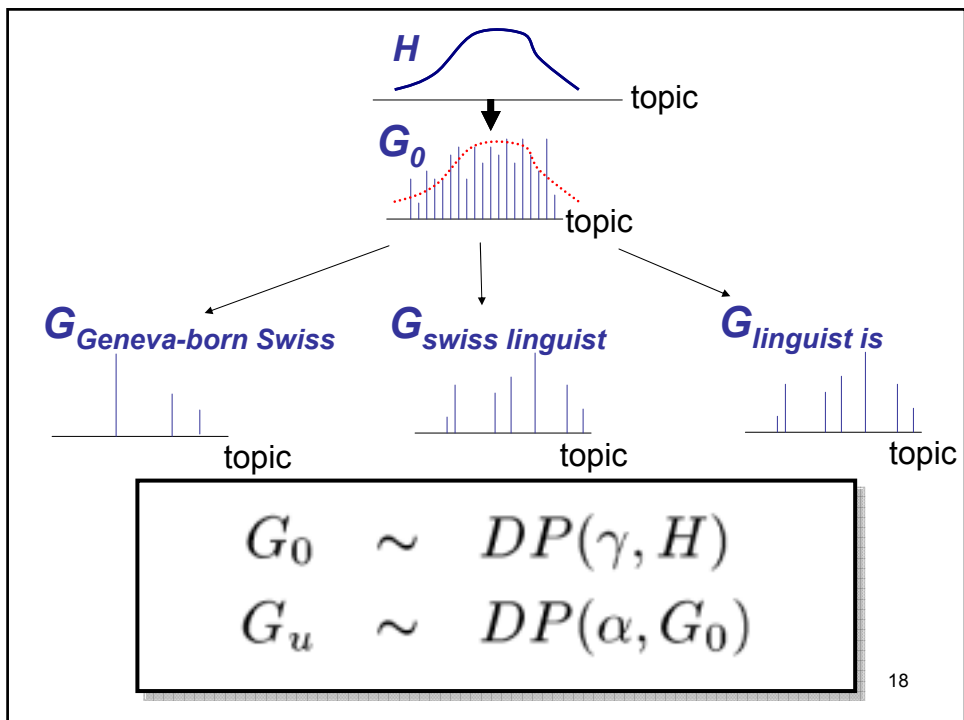
15



16

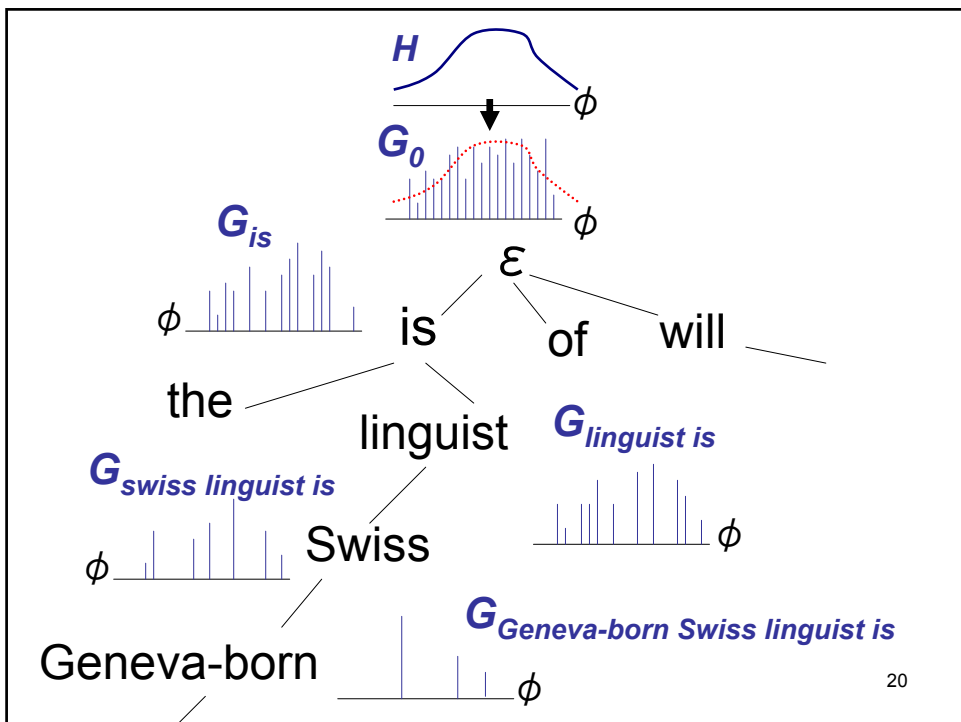
Model 1

17

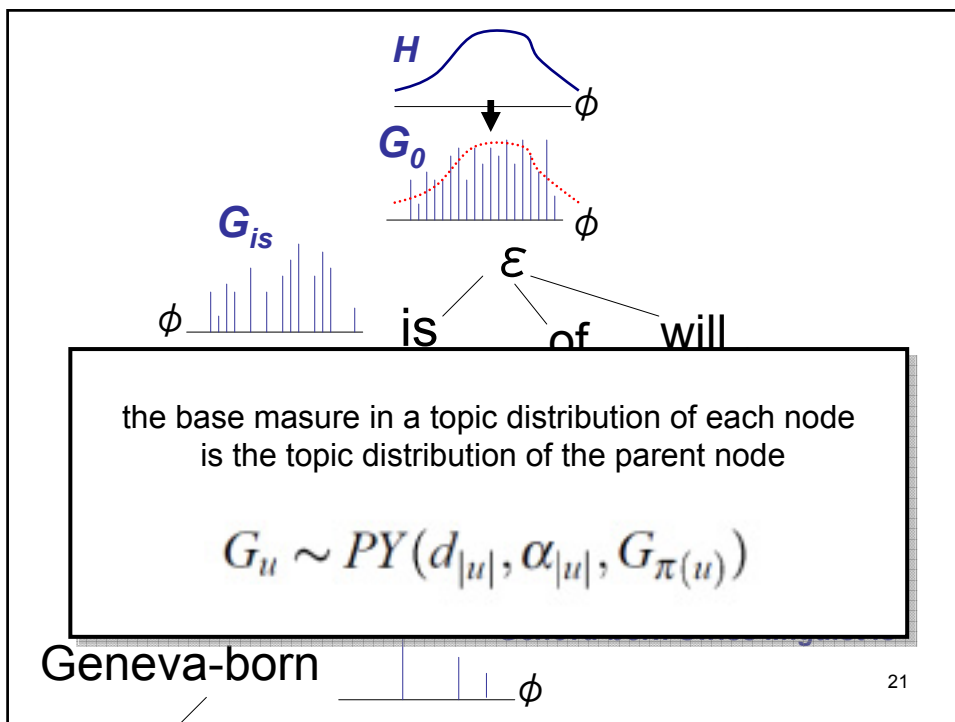


Model 2

19



20



Experiment

Learning algorithm of our models

- (Collapsed) Gibbs sampling

Dataset

-Reuters (39,149 sentences, 41,366 words)

Evaluation - word clustering

baseline

- Sentence co-occurrence(HDP-LDA)
Document co-occurrence(HDP-LDA)
- Proximity(Multinomial Mixture model)

23

Baseline 1:HDP-LDA [Teh,2003]

- Adopte HDP-LDA in a sentence-level and in a document-level
- Cluster words by estimating topics of words in HDP-LDA
- Identify topic as follows:

$$z_{ji} = \operatorname{argmax}(P(z_{ji}|w_{ji}))$$

topic of i-th word in sentence or document j

24

Baseline2: Proximity [Hagiwara+,2006]

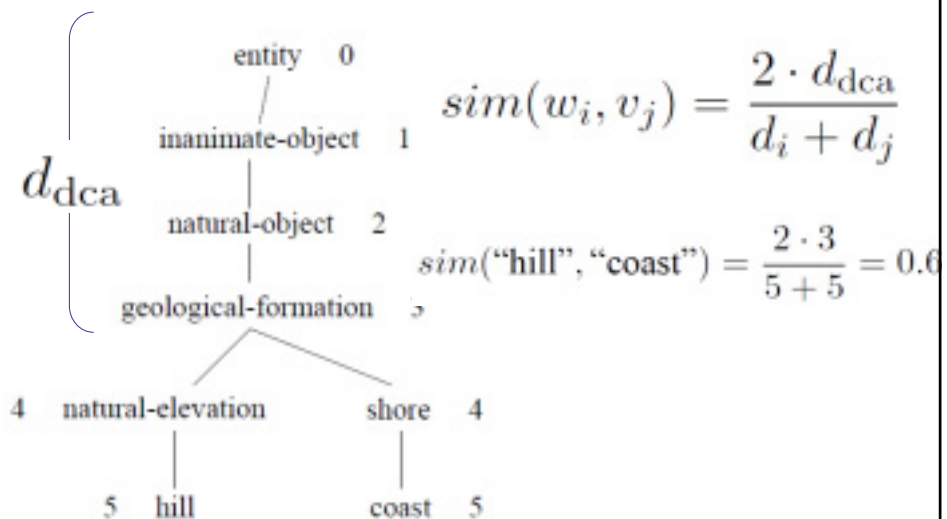
words that appeared in the neighborhood
of the target word in a sentence

Linguistics is the scientific study of language
L:3 L:2 L:1 study R:1 R:2

study → L:3:is 1
 L:2:the 1
 L:1:scientific 1 → Vector
 R:1:of 1 Space
 R:2:language 1 Model

25

WordNet Similarity



26

Evaluation scheme

Target word ← Top 1000 (noun) words
in descending order of frequency

Positive example

- Word1 +
 - Word2 +
 - Word3 -
- ↑ high
similarity

Negative example

- Word1 +
 - Word2 -
 - Word3 -
- ↑ low
similarity
- } Top N

If the target word and word i is in the same cluster, set label “+”.

If the target word and word i is **not** in the same cluster, set label “+”.

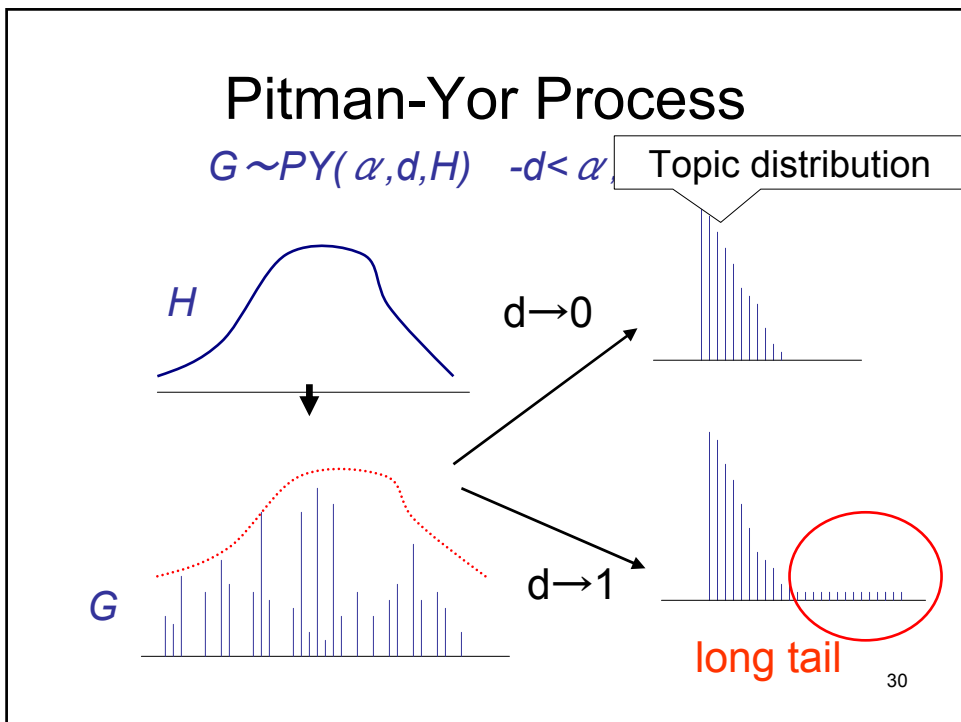
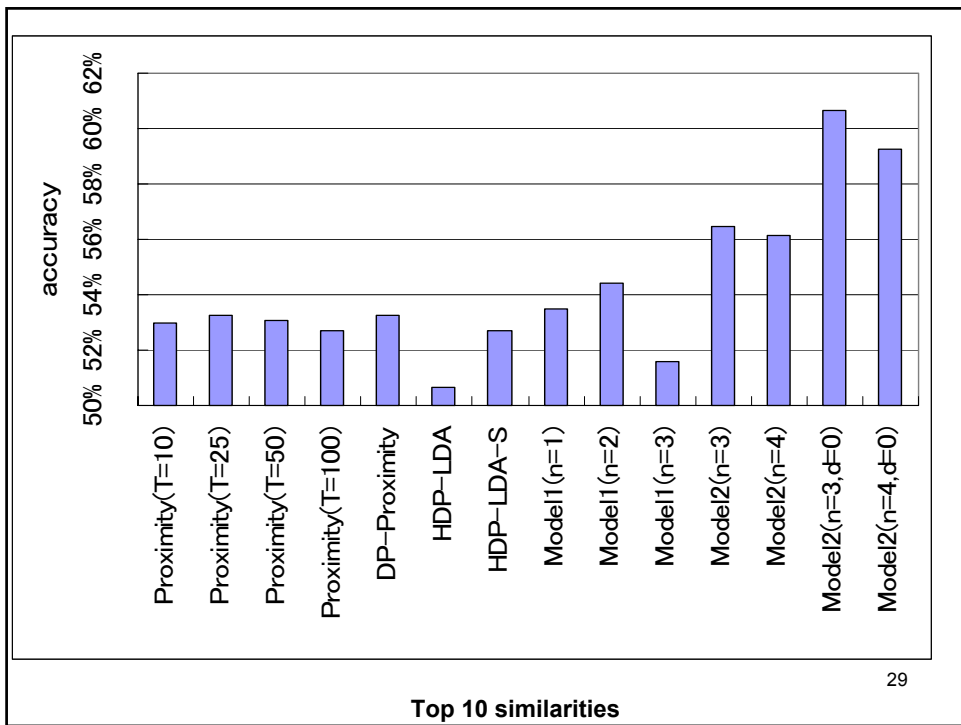
accuracy = the proportion of “+”

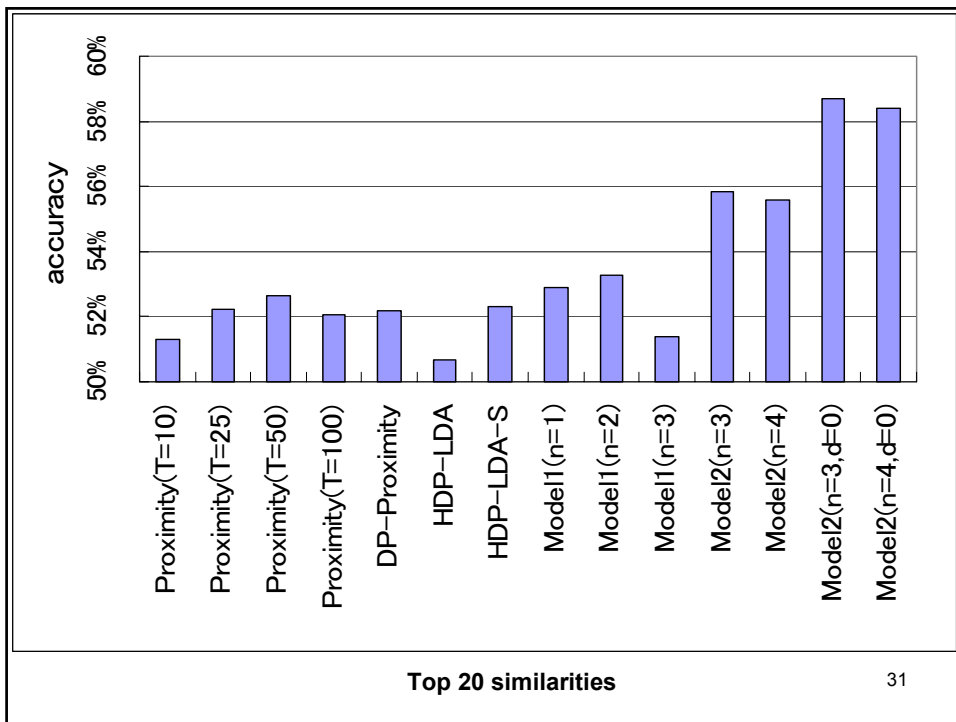
27

Models

- **Proximity**: T is the number of clusters
- **DP-Proximity** : estimate T using DP
- **HDP-LDA**: document-level
- **HDP-LDA-S**: sentence-level
- **Model_{1,2}** : n is the length of context

28





抽出例 1/2

[growth in the]	
topic	word
10	pc, gross
80	economy, market, u.s., company first, second, third
85	march, january
100	student
485	latter

抽出例 2/2

"language"	
topic	Context
61	[swahili and Somali] [(the inca] [local nahuatl indian] [most widely speak] [not understand the] [which provide foreign] [, whose body] [, the international] [china 's national]
90	[statement say that] ['s shift in] [in munich adopt] [in education and]

33

Future work

How to estimate the length of the context

identifying the semantic meaning of a word?



Infinite Morkov Model[Mochihasi+,2007]

34