Evaluating Contextual Dependency of Paraphrases using a Latent Variable Model

Kiyonori OHTAKE

ATR Spoken Language Communication Research Laboratories, Japan

October 14, 2005

Background

- Problem: *contextual dependency* of paraphrases
  - applying
  - collecting
  - etc.
- How to cope with contextual dependency?

Example

paraphrasing pair:
A: *I want to buy a pair of sandals.*
B: *I'm looking for sandals.*

Can we always paraphrase them?
from A to B: Yes
from B to A: No

Objective

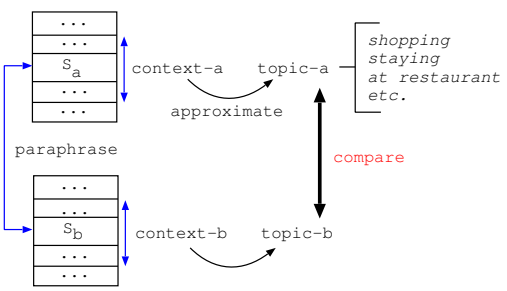
To construct an evaluation method for contextual dependency of paraphrases

If we can evaluate ...

- mis-paraphrasing
- mis-collecting paraphrases

will be avoided

Idea



Approach

Using latent variable text model

- modeling a text (unsupervised)
- each latent variable represents a *topic*

Context: sentence and surrounding sentences = window

Approximation of context: topic indicated by a latent variable

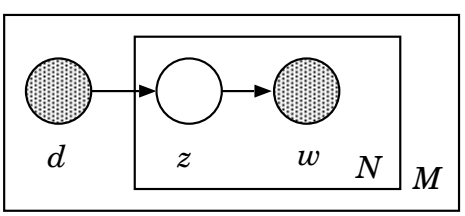
Latent variable text models

- PLSI $P(d, w)$
  Probabilistic replacement of LSI
- LDA $P(d|\alpha, \beta)$
  Bayesian replacement of pLSI

Latent (hidden) variable represents *topic*

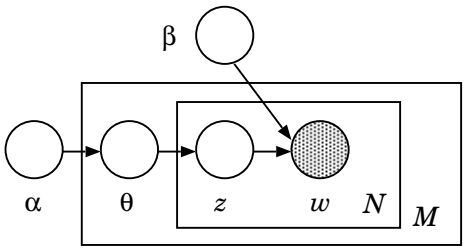
Note: # of latent variables (topics) is given

Latent variable model: pLSI



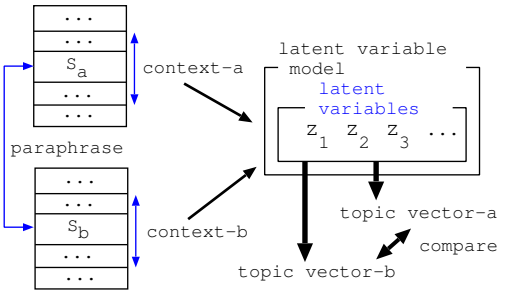
d: document, z: latent variable, w: word, N: vocaburary size, M: # of documents

Latent variable model: LDA



θ: Dirichlet random variable
α: parameter for θ, β: parameter for w

Evaluating method



Comparing topic vectors

- *Based on the largest element*
  Whether the largest element of topic vectors are the same
- *Cosine*
  Whether $cos\theta$ between topic vectors greater than threshold

Overview of Experiments

- *Comparing with labels by human*
  Matching the results of our method with the result based on the topic labels by human
- *Evaluation for paraphrasing*
  Evaluating our method based on collecting situation for paraphrases

Data set

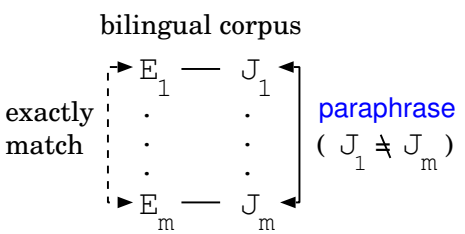
Bilingual corpus of travel conversation (162,000 sentence pairs)

Manually and roughly labeled with topics (hierarchical; level-1: **19**, level-2: 218)

Textual cohesion
→ Used fixed window to clip a context

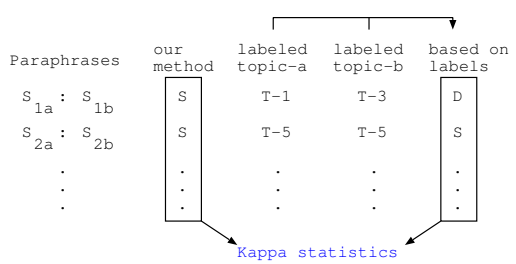
Data format: bag-of-words

Extracting paraphrases



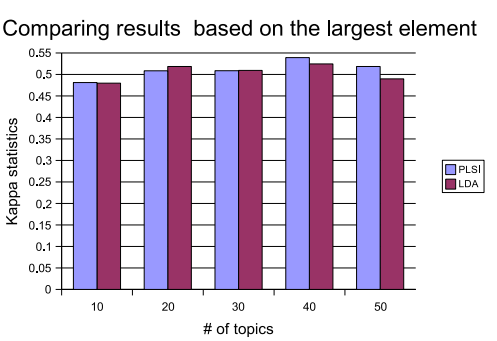
Obtained 944,547 Japanese paraphrasing pairs

Comparing with labels by human

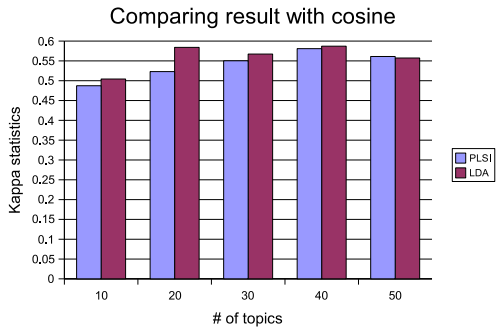


Measurement: Kappa statistics
Two comparing method: largest, cosine

Result 1/2



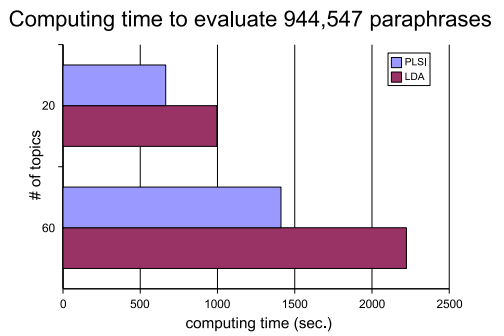
Result 2/2



Discussion

- Kappa statistics: 0.59 (highest)
  moderate: 0.4-0.6
  substantial: 0.6-0.8
  perfect: over 0.8
- No major performance difference
- LDA was good at with cosine
- PLSI sometimes outperformed LDA

Computing time

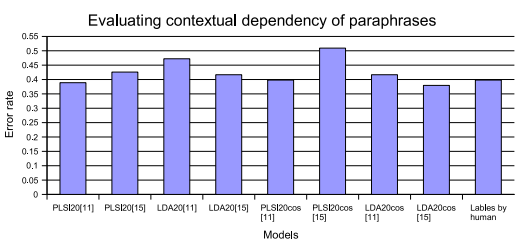


Evaluation for paraphrasing

1. Fed randomly sampled 108 paraphrasing pairs into our method
2. Manually evaluated whether they were contextually independent

Measurement: error rate

Result



Discussion

- Almost the same as the result based on labels by human
- *25 unavoidable errors*
  Potential upper bound based on topic information: 77% (0.23 error rate)

Conclusion

- Proposed evaluation method for contextual dependency of paraphrases using pLSI and LDA
- No major performance difference between pLSI and LDA
- Potential upper bound using *only* topic: 77%
  Achieved: 62%

Future works

- Introducing a topic-boundary detection technique
- Employing more complicated data (e.g., dependency structure) not b.o.w.
- Investigating difference of paraphrase:
  - What makes contextual dependency?
  - What contexts are possible?

Thank you very much.