707.009
Foundations of Knowledge Management
„Topic Modeling“

How can we uncover ~~semantic relationships~~ topics from natural language text?

Markus Strohmaier

Univ. Ass. / Assistant Professor
Knowledge Management Institute
Graz University of Technology, Austria

e-mail: markus.strohmaier@tugraz.at
web: http://www.kmi.tugraz.at/staff/markus

Acknowledgements: Course slides in part based on

…the following slide decks and papers:

- "Probabilistic Topic Models and Associative Memory"
  - Mark Steyvers, UC Irvine, Tom Griffiths, Brown University, Josh Tenenbaum,MIT
- "Topics in Semantic Representation"
  - Tom Griffiths, Brown University, Mark Steyvers, UC Irvine, Josh Tenenbaum,MIT
- Semantic Representations with Probabilistic Topic Models
  - Mark Steyvers, Joint work with: Tom Griffiths, UC Berkeley, Padhraic Smyth, UC Irvine
- "Modeling Documents"
  - Amruta Joshi, Department of Computer Science, Stanford University
- „Cognitive Modeling"
  - Lecture 14: Models of Semantic Processing, University of Edinburgh

Overview

Today‘s Agenda:

Topic Modeling

- Associative memory
- The topic model
- Applications to associative memory
- Applications in machine learning/text mining

Wissensorganisation – Zwei Herangehensweise

Formale vs. inhaltliche Struktur

Viele Informationen liegen in unstrukturierten Freitexten (ohne Struktur) vor. Aussagekräftig aber schlecht auswertbar

Zwei Herangehensweisen:

- Verwendung einer standardisierten Sprache a **priori** (stark formalisiert)
- Interpretation der heterogenen Sprache a **posteriori** (NLP, …)

Taxonomien, Ontologien, Semantische Netze

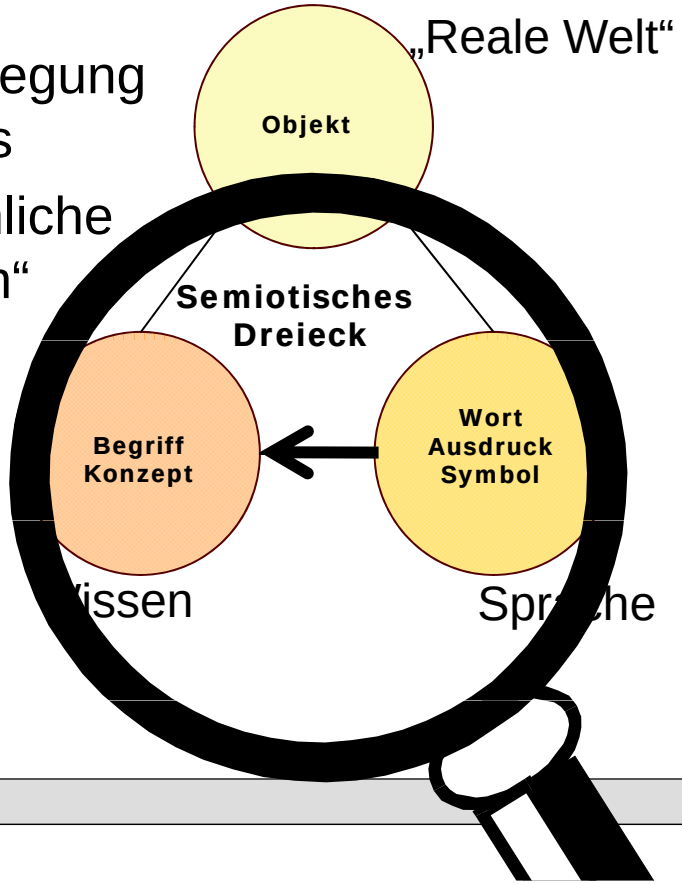
Schlüsselwort-extraktion, Folksonomies

Was sind Konzeptsysteme?

Konzeptsysteme sind Systeme von unterscheidbaren *Konzepten*, die mittels *Relationen* in Beziehung zueinander gesetzt werden und in einer natürlicheren *Sprache* formuliert werden können

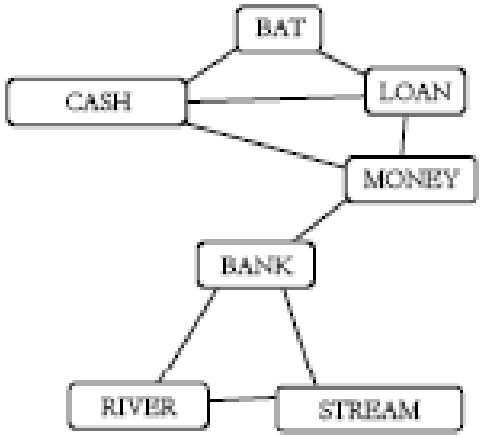
Zielsetzung: Entwicklung und Festlegung eines gemeinsamen Verständnisses

Repräsentationssysteme: menschliche Sprache, Logik, „Computersprachen“



Two approaches to semantic representation

Semantic networks

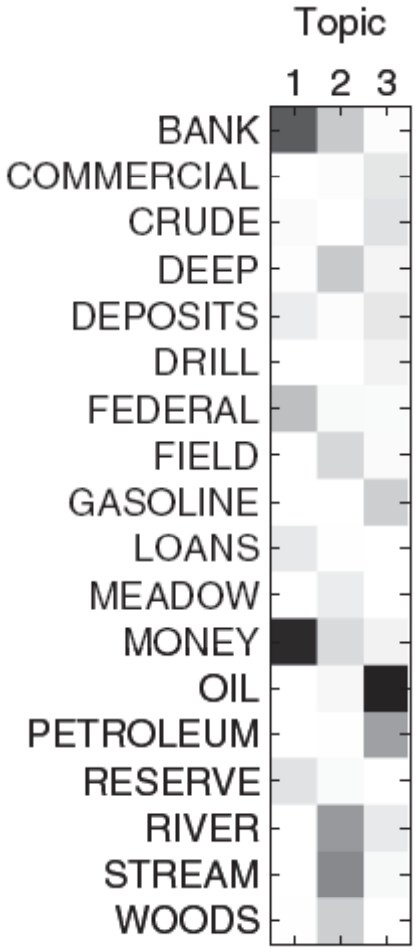


Semantic Spaces

How are these learned?

A third approach: Topic Modeling

(c)



MONEY	STREAM	OIL
BANK	RIVER	PETROLEUM
FEDERAL	BANK	GASOLINE
RESERVE	DEEP	CRUDE
LOANS	WOODS	COMMERCIAL
DEPOSITS	FIELD	DEPOSITS
COMMERCIAL	MONEY	RIVER
DEEP	MEADOW	DRILL
MEADOW	OIL	MONEY
OIL	FEDERAL	DEEP
RIVER	DEPOSITS	FIELD
CRUDE	LOANS	STREAM
DRILL	COMMERCIAL	RESERVE
FIELD	CRUDE	BANK
GASOLINE	DRILL	LOANS
PETROLEUM	GASOLINE	FEDERAL
STREAM	PETROLEUM	MEADOW
WOODS	RESERVE	WOODS

Overview

I Associative memory

II The topic model

III Applications to associative memory

IV Applications in machine learning/text mining

Example of associative memory: word association

CUE:

PLAY

RESPONSES:

FUN, BALL, GAME, WORK, GROUND, MATE, CHILD, ENJOY, WIN, ACTOR

Example of associative memory: free recall

STUDY THESE WORDS:

Bed, Rest, Awake, Tired, Dream, Wake, Snooze, Blanket, Doze, Slumber, Snore, Nap, Peace, Yawn, Drowsy

RECALL WORDS .....

FALSE RECALL: "Sleep" 61%

A theory for semantic association

Semantic association as probabilistic inference

Representation of semantic structure

Topic Models in Psychology

- Topic models address three computational problems for semantic memory system:

1) Gist extraction: what is this set of words about?

Infer g from **w**

2) Disambiguation: what is the sense of this word?

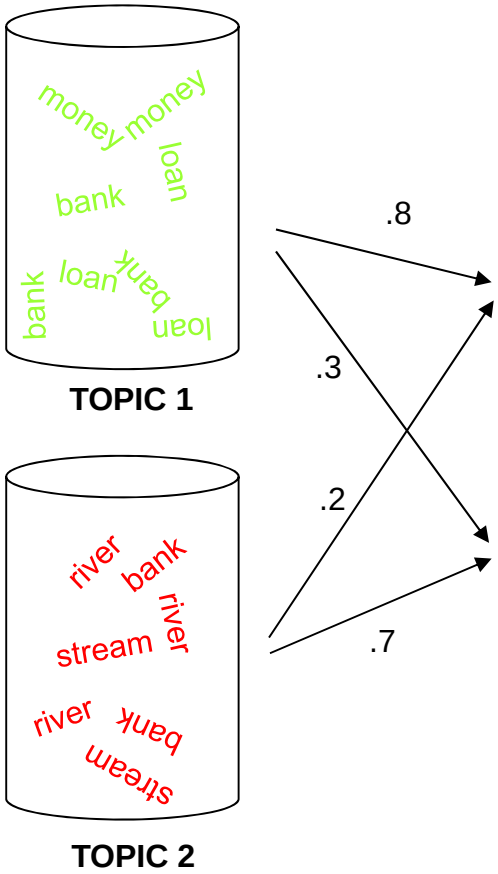
- E.g. "football field" vs. "magnetic field"

Infer z from **w**

3) Prediction: what fact, concept, or word is next?

Infer w_{n+1} from **w**

GENERATIVE PROCESS



DOCUMENT 1: money[1] bank[1] bank[1] loan[1] river[2] stream[2] bank[1] money[1] river[2] bank[1] money[1] bank[1] loan[1] money[1] stream[2] bank[1] money[1] bank[1] bank[1] loan[1] river[2] stream[2] bank[1] money[1] river[2] bank[1] money[1] bank[1] loan[1] bank[1] money[1] stream[2]

DOCUMENT 2: river[2] stream[2] bank[2] stream[2] bank[2] money[1] loan[1] river[2] stream[2] loan[1] bank[2] river[2] bank[2] bank[1] stream[2] river[2] loan[1] bank[2] stream[2] bank[2] money[1] loan[1] river[2] stream[2] bank[2] stream[2] bank[2] money[1] river[2] stream[2] loan[1] bank[2] river[2] bank[2] money[1] bank[1] stream[2] river[2] bank[2] stream[2] bank[2] money[1]

Mixture components

Mixture weights

- No notion of mutual exclusivity
- Capturing polysemy
- Bag of words

The probability of choosing a word:

$$P(w) = \sum_{z=1}^{T} P(w \mid z) P(z)$$

word probability in topic *j*

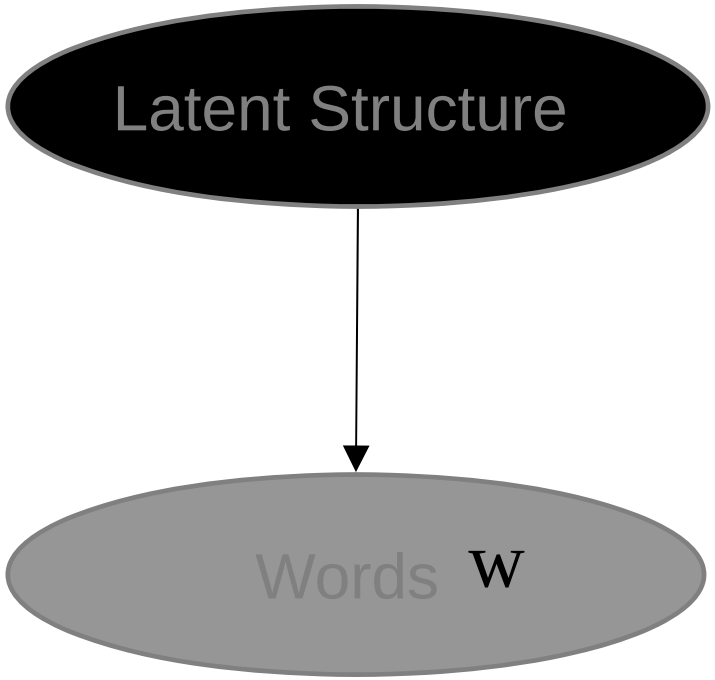
probability of topic *j* in document

T…Number of Topics

Bayes' rule:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Latent Semantic Structure



Distribution over words

$$P(\mathbf{w}) = \sum_{\ell} P(\mathbf{w}, \ell)$$

Inferring latent structure

$$P(\ell \mid \mathbf{w}) = \frac{P(\mathbf{w} \mid \ell) P(\ell)}{P(\mathbf{w})}$$

Prediction

$$P(w_{n+1} \mid \mathbf{w}) = \ldots$$

Overview

I Associative memory

II The topic model

III Applications to associative memory

IV Applications in machine learning/text mining

The Big Idea

Topic Model

- Model topics as distribution over words

Document Model

- Model documents as distribution over words

Document / Topic Model

- Probabilistic Model for both
- Model topics as distribution over words
- Model documents as distribution over topics

Topic Model

Unsupervised learning of topics ("gist") of documents:

- articles/chapters
- conversations
- emails
- .... any verbal context

Topics are useful latent structures to explain semantic association

Probabilistic Generative Model

Each topic is a probability distribution over words

From the TASA corpus, a collection of over 37,000 text passages from educational materials (e.g., language & arts, social studies, health, sciences) collected by Touchstone Applied Science Associates (see Landauer, Foltz, & Laham, 1998).

Topic 247

word	prob.
DRUGS	.069
DRUG	.060
MEDICINE	.027
EFFECTS	.026
BODY	.023
MEDICINES	.019
PAIN	.016
PERSON	.016
MARIJUANA	.014
LABEL	.012
ALCOHOL	.012
DANGEROUS	.011
ABUSE	.009
EFFECT	.009
KNOWN	.008
PILLS	.008

Topic 5

word	prob.
RED	.202
BLUE	.099
GREEN	.096
YELLOW	.073
WHITE	.048
COLOR	.048
BRIGHT	.030
COLORS	.029
ORANGE	.027
BROWN	.027
PINK	.017
LOOK	.017
BLACK	.016
PURPLE	.015
CROSS	.011
COLORED	.009

Topic 43

word	prob.
MIND	.081
THOUGHT	.066
REMEMBER	.064
MEMORY	.037
THINKING	.030
PROFESSOR	.028
FELT	.025
REMEMBERED	.022
THOUGHTS	.020
FORGOTTEN	.020
MOMENT	.020
THINK	.019
THING	.016
WONDER	.014
FORGET	.012
RECALL	.012

Topic 56

word	prob.
DOCTOR	.074
DR.	.063
PATIENT	.061
HOSPITAL	.049
CARE	.046
MEDICAL	.042
NURSE	.031
PATIENTS	.029
DOCTORS	.028
HEALTH	.025
MEDICINE	.017
NURSING	.017
DENTAL	.015
NURSES	.013
PHYSICIAN	.012
HOSPITALS	.011

Figure 1. An illustration of four (out of 300) topics extracted from the TASA corpus.

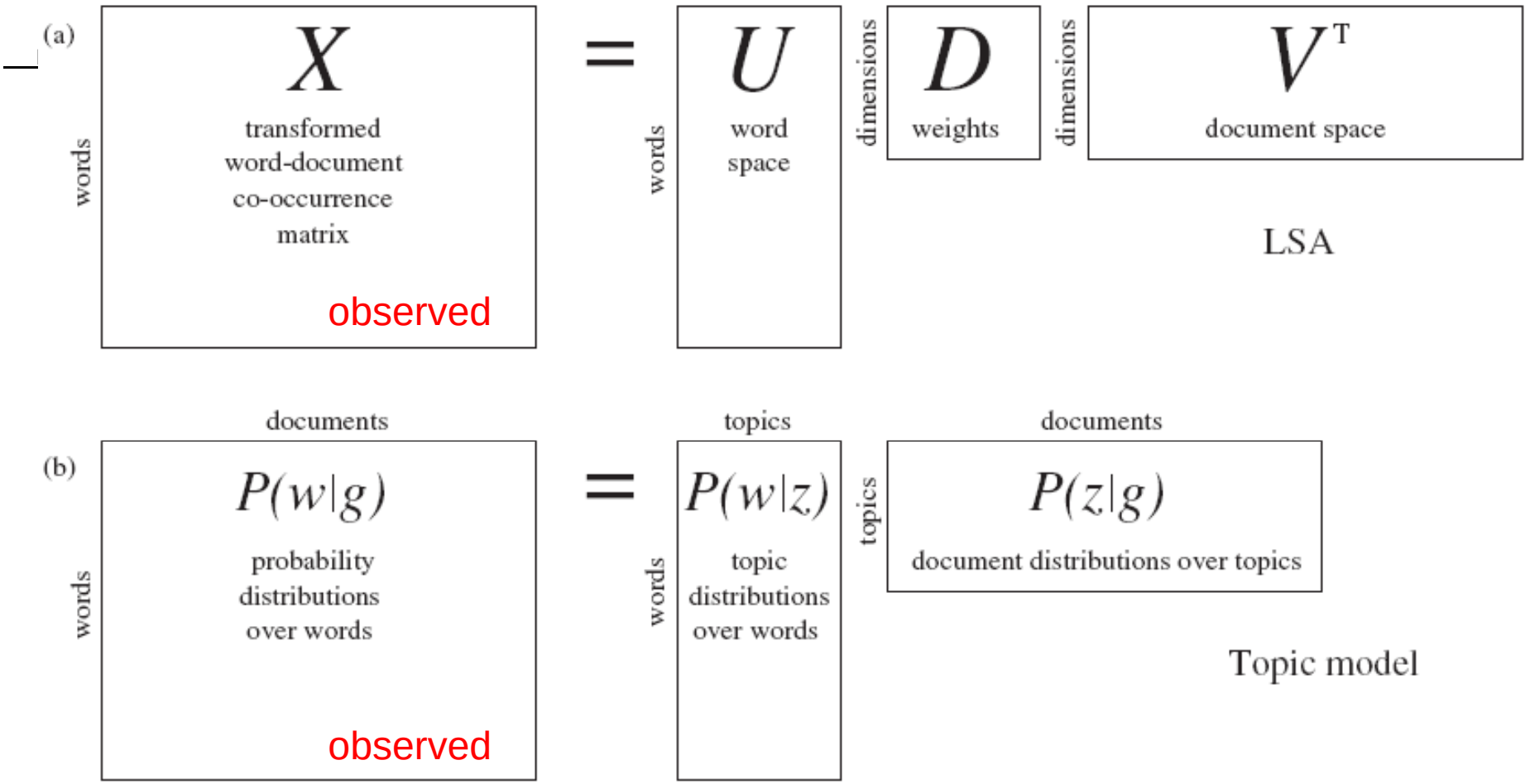


Figure 3. (a) Latent semantic analysis (LSA) performs dimensionality reduction using the singular value decomposition. The transformed word–document co-occurrence matrix, X, is factorized into three smaller matrices, U, D, and V. U provides an orthonormal basis for a spatial representation of words, D weights those dimensions, and V provides an orthonormal basis for a spatial representation of documents. (b) The topic model performs dimensionality reduction using statistical inference. The probability distribution over words for each document in the corpus conditioned on its gist, $P(w|g)$, is approximated by a weighted sum over a set of probabilistic topics, represented with probability distributions over words, $P(w|z)$, where the weights for each document are probability distributions over topics, $P(z|g)$, determined by the gist of the document, g.

Taken from „Topics in Semantic Representation“, Thomas L. Griffiths, Mark Steyvers,Joshua B. Tenenbaum

Inference – Constructing Topic Models

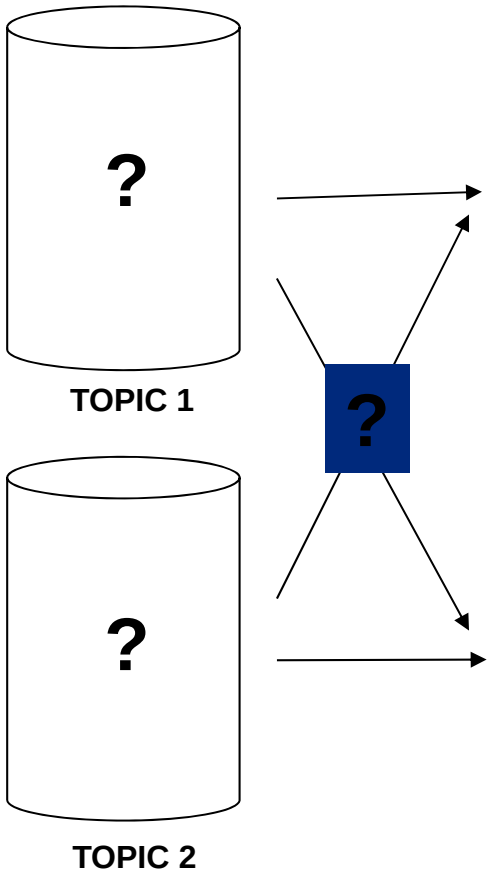
Expectation Maximization

- But poor results (local Maxima)

Gibbs Sampling

- Parameters: ϕ, θ
- Start with initial random assignment
- Update parameter using other parameters
- Converges after ‘n’ iterations
- Burn-in time

INVERTING THE GENERATIVE PROCESS

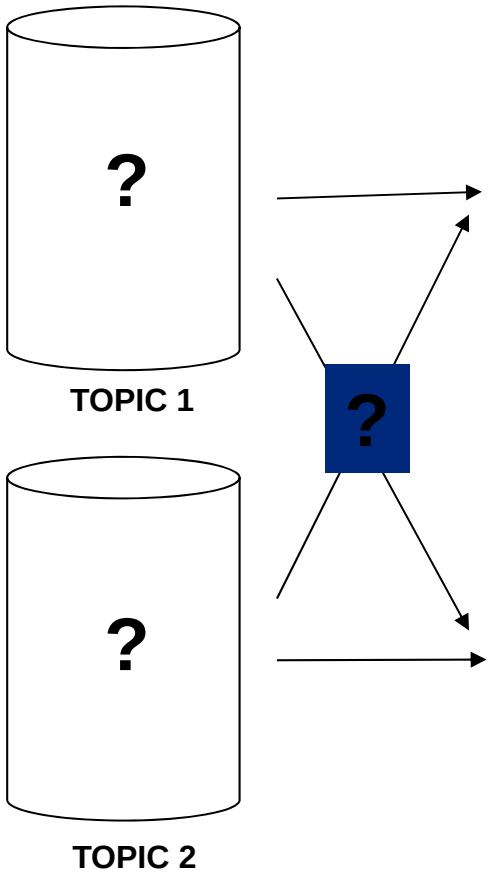


DOCUMENT 1: A Play is written to be performed on a stage before a live audience or before motion picture or television cameras (for later viewing by large audiences). A Play is written because playwrights have something ...

DOCUMENT 2: He was listening to music coming from a passing riverboat. The music had already captured his heart as well as his ear . It was jazz . Bix beiderbecke had already had music lessons . He wanted to play the cornet. And he wanted to play jazz .....

We estimate the assignments of topics to words

INVERTING THE GENERATIVE PROCESS



DOCUMENT 1: A Play[082] is written[082] to be performed[082] on a stage[082] before a live[093] audience[082] or before motion[270] picture[004] or television[004] cameras[004] (for later[054] viewing[004] by large[202] audiences[082]). A Play[082] is written[082] because playwrights[082] have something …

DOCUMENT 2: He was listening[077] to music[077] coming[009] from a passing[043] riverboat. The music[077] had already captured[006] his heart[157] as well as his ear[119]. It was jazz[077]. Bix beiderbecke had already had music[077] lessons[077]. He wanted[268] to play[077] the cornet. And he wanted[268] to play[077] jazz[077].....

We estimate the assignments of topics to words
Blue words represent stopwords/words not used

Statistical Inference

Fix number of topics T

We estimate the posterior over topic assignments

$$P(\mathbf{z} \mid \mathbf{w}) = \frac{P(\mathbf{w}, \mathbf{z})}{\sum_{\mathbf{z}} P(\mathbf{w}, \mathbf{z})}$$

Markov Chain Monte Carlo (MCMC) with Gibbs sampling

Topic Model Inference: Procedure

INPUT:

word-document counts

OUTPUT:

topic assignments to each word $\mathbf{z}$

likely words in each topic $P(w \mid z)$

likely topics for a document ("gist") $P(z \mid \mathbf{w})$

Example: topics from an educational corpus (TASA)

- 37K docs, 26K words
- 1700 topics, e.g.:

From the TASA corpus, a collection of over 37,000 text passages from educational materials (e.g., language & arts, social studies, health, sciences) collected by Touchstone Applied Science Associates (see Landauer, Foltz, & Laham, 1998).

PRINTING	PLAY	TEAM	JUDGE	HYPOTHESIS	STUDY
PAPER	PLAYS	GAME	TRIAL	EXPERIMENT	TEST
PRINT	STAGE	BASKETBALL	COURT	SCIENTIFIC	STUDYING
PRINTED	AUDIENCE	PLAYERS	CASE	OBSERVATIONS	HOMEWORK
TYPE	THEATER	PLAYER	JURY	SCIENTISTS	NEED
PROCESS	ACTORS	PLAY	ACCUSED	EXPERIMENTS	CLASS
INK	DRAMA	PLAYING	GUILTY	SCIENTIST	MATH
PRESS	SHAKESPEARE	SOCCER	DEFENDANT	EXPERIMENTAL	TRY
IMAGE	ACTOR	PLAYED	JUSTICE	TEST	TEACHER
PRINTER	THEATRE	BALL	EVIDENCE	METHOD	WRITE
PRINTS	PLAYWRIGHT	TEAMS	WITNESSES	HYPOTHESES	PLAN
PRINTERS	PERFORMANCE	BASKET	CRIME	TESTED	ARITHMETIC
COPY	DRAMATIC	FOOTBALL	LAWYER	EVIDENCE	ASSIGNMENT
COPIES	COSTUMES	SCORE	WITNESS	BASED	PLACE
FORM	COMEDY	COURT	ATTORNEY	OBSERVATION	STUDIED
OFFSET	TRAGEDY	GAMES	HEARING	SCIENCE	CAREFULLY
GRAPHIC	CHARACTERS	TRY	INNOCENT	FACTS	DECIDE
SURFACE	SCENES	COACH	DEFENSE	DATA	IMPORTANT
PRODUCED	OPERA	GYM	CHARGE	RESULTS	NOTEBOOK
CHARACTERS	PERFORMED	SHOT	CRIMINAL	EXPLANATION	REVIEW

Polysemy

PRINTING	**PLAY**	TEAM	JUDGE	HYPOTHESIS	STUDY
PAPER	PLAYS	GAME	TRIAL	EXPERIMENT	**TEST**
PRINT	STAGE	BASKETBALL	**COURT**	SCIENTIFIC	STUDYING
PRINTED	AUDIENCE	PLAYERS	CASE	OBSERVATIONS	HOMEWORK
TYPE	THEATER	PLAYER	JURY	SCIENTISTS	NEED
PROCESS	ACTORS	**PLAY**	ACCUSED	EXPERIMENTS	CLASS
INK	DRAMA	PLAYING	GUILTY	SCIENTIST	MATH
PRESS	SHAKESPEARE	SOCCER	DEFENDANT	EXPERIMENTAL	TRY
IMAGE	ACTOR	PLAYED	JUSTICE	**TEST**	TEACHER
PRINTER	THEATRE	BALL	**EVIDENCE**	METHOD	WRITE
PRINTS	PLAYWRIGHT	TEAMS	WITNESSES	HYPOTHESES	PLAN
PRINTERS	PERFORMANCE	BASKET	CRIME	TESTED	ARITHMETIC
COPY	DRAMATIC	FOOTBALL	LAWYER	**EVIDENCE**	ASSIGNMENT
COPIES	COSTUMES	SCORE	WITNESS	BASED	PLACE
FORM	COMEDY	**COURT**	ATTORNEY	OBSERVATION	STUDIED
OFFSET	TRAGEDY	GAMES	HEARING	SCIENCE	CAREFULLY
GRAPHIC	**CHARACTERS**	TRY	INNOCENT	FACTS	DECIDE
SURFACE	SCENES	COACH	DEFENSE	DATA	IMPORTANT
PRODUCED	OPERA	GYM	CHARGE	RESULTS	NOTEBOOK
CHARACTERS	PERFORMED	SHOT	CRIMINAL	EXPLANATION	REVIEW

Overview

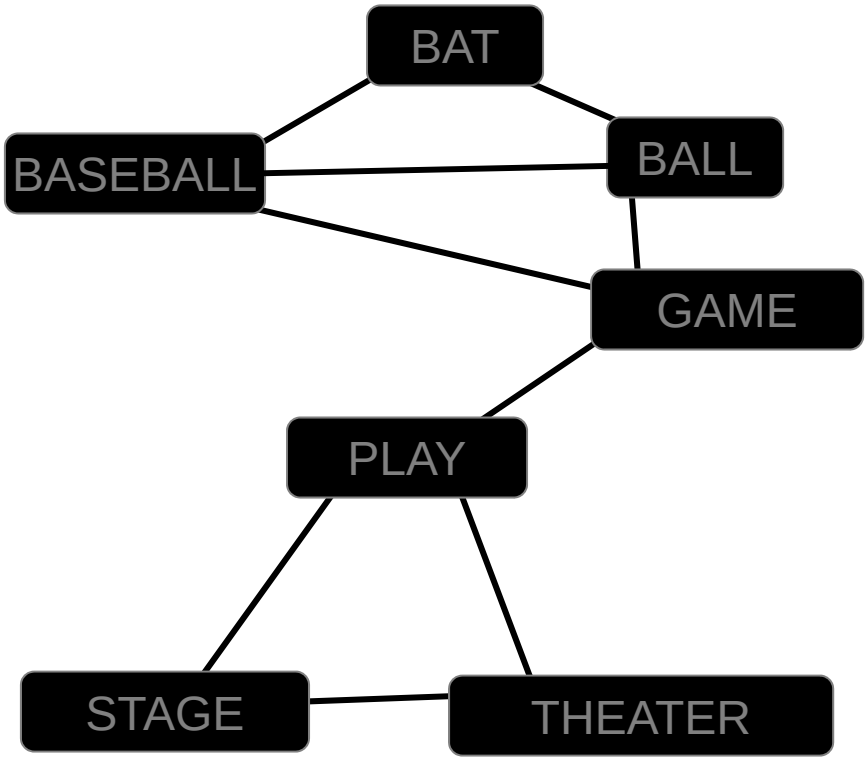
I Associative memory

II The topic model

III Applications to associative memory

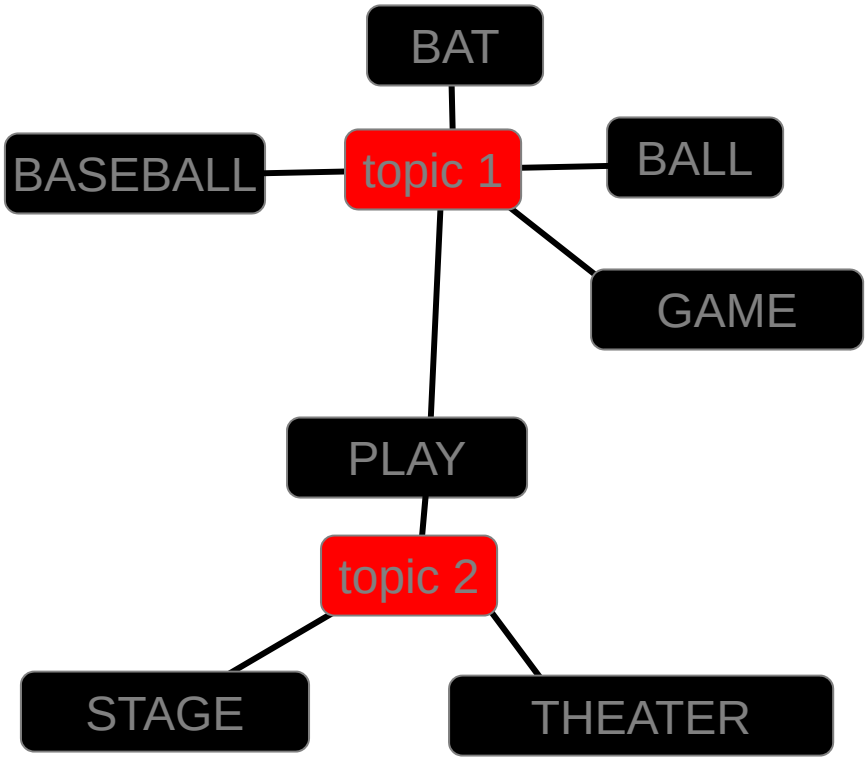
IV Applications in machine learning/text mining

Example associative structure



(Association norms by Doug Nelson et al. 1998)

Explaining structure with topics



Tasa corpus

Need a suitable corpus to model human associations

TASA

- an educational corpus of text
- 37K documents
- 26K words

Modeling Word Association

Word association modeled as prediction

Idea: The similarity between two words w_1 + w_2 can be measured by the extent that they share the same topics.

Given that a single word is observed, what future other words might occur?

Under a single topic assumption:

$$P(w_{n+1} \mid \mathbf{w}) = \sum_z P(w_{n+1} \mid z) P(z \mid \mathbf{w})$$

Response

Cue

Probability that w2 is generated by z

Probability that w1 indicates topic z

What is the most likely topic z for w1, and what is the most likely word for this topic z. Multiply these two for all words w2.

Observed associates for the cue "play"

HUMANS

List generated by human subjects

Word	P(word)
FUN	.141
BALL	.134
GAME	.074
WORK	.067
GROUND	.060
MATE	.027
CHILD	.020
ENJOY	.020
WIN	.020
ACTOR	.013
FIGHT	.013
HORSE	.013
KID	.013
MUSIC	.013

Model predictions

List generated by topic models

HUMANS

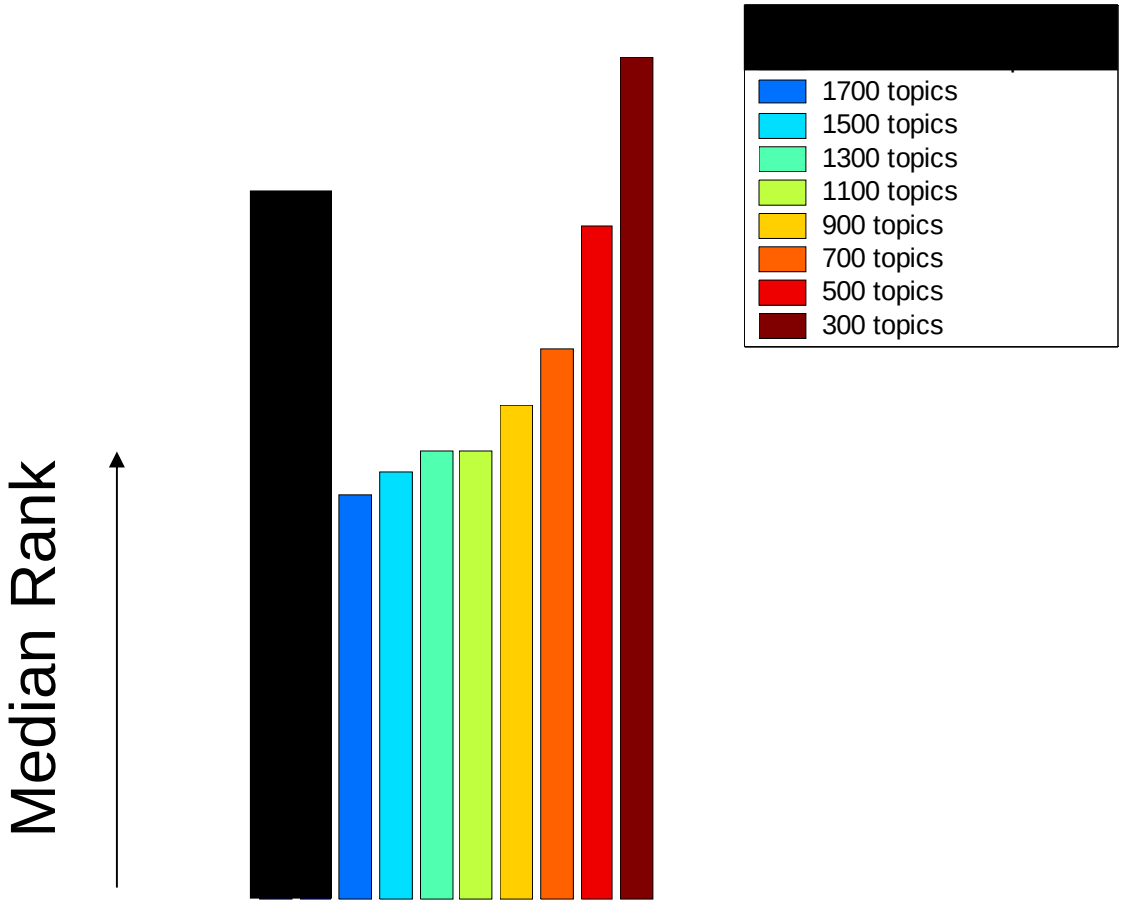
Word	P(word)
FUN	.141
BALL	.134
GAME	.074
WORK	.067
GROUND	.060
MATE	.027
CHILD	.020
ENJOY	.020
WIN	.020
ACTOR	.013
FIGHT	.013
HORSE	.013
KID	.013
MUSIC	.013

TOPICS (T=500)

Word	P(word)
BALL	.041
GAME	.039
CHILDREN	.019
ROLE	.014
GAMES	.014
MUSIC	.009
BASEBALL	.009
HIT	.008
FUN	.008
TEAM	.008
IMPORTANT	.006
BAT	.006
RUN	.006
STAGE	.005

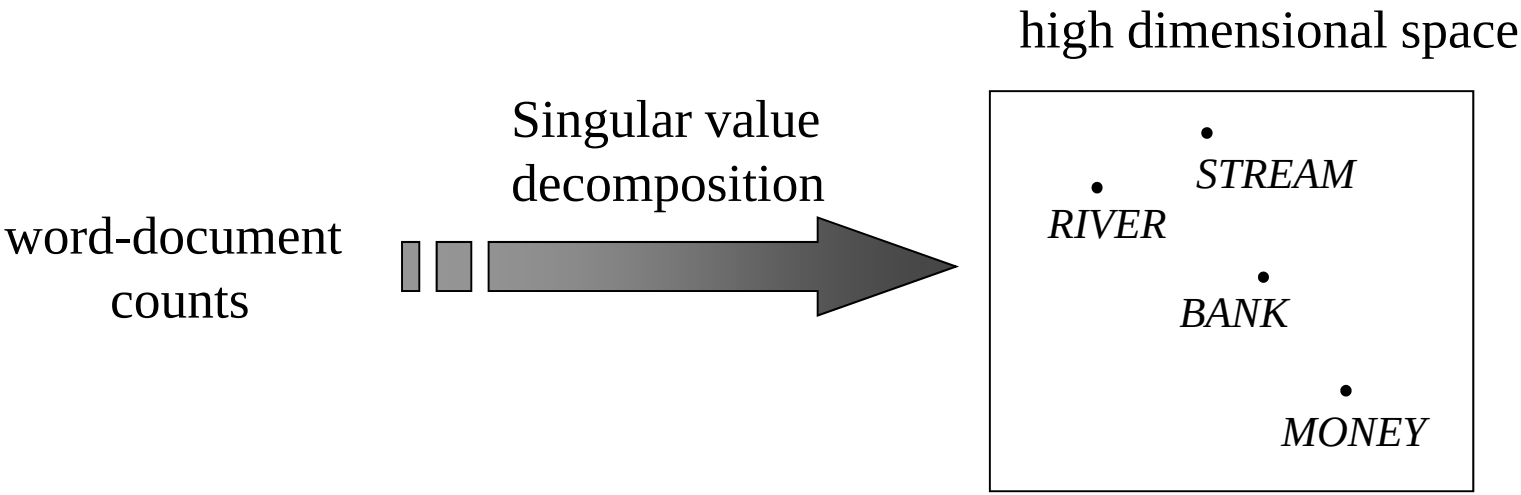
RANK 9

Median rank of first associate



Latent Semantic Analysis

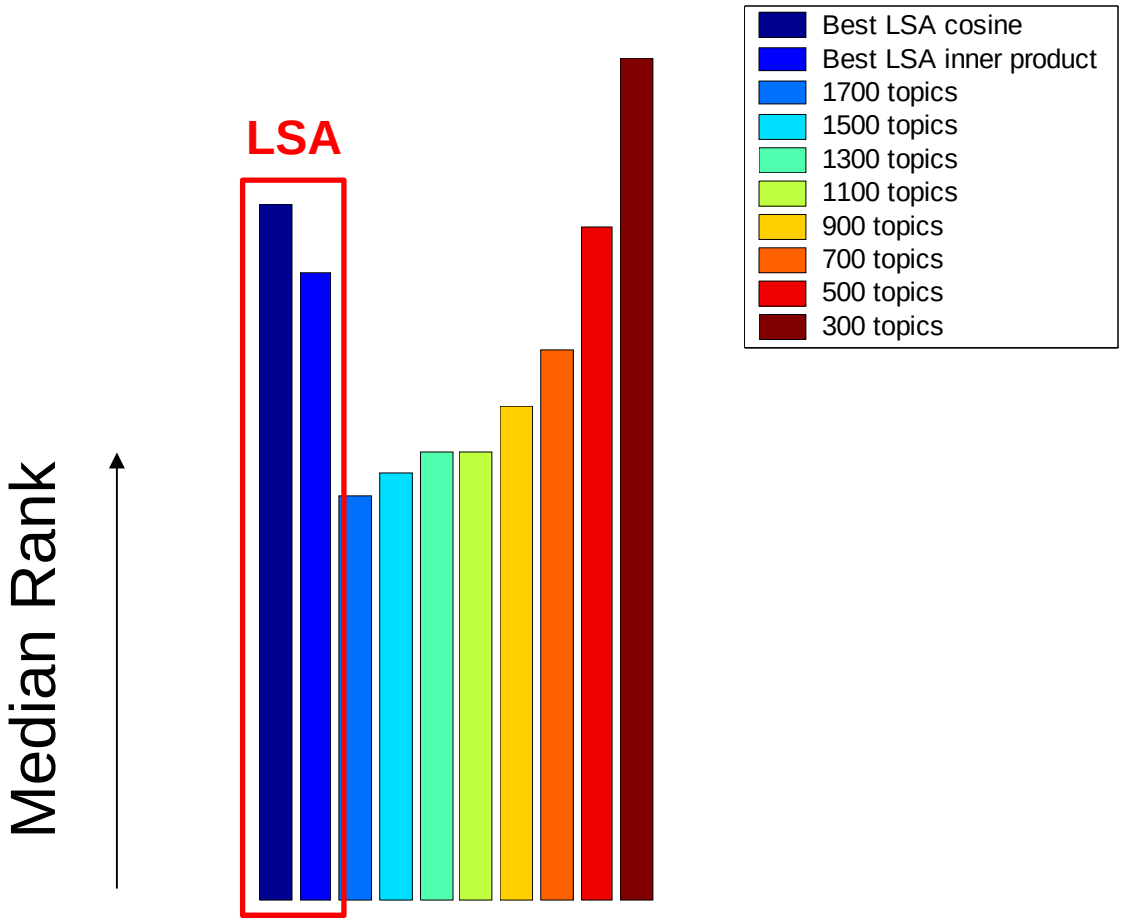
(Landauer & Dumais, 1997)



Each word is a single point in semantic space

Similarity measured by cosine of angle between word vectors

Median rank of first associate



Recall: example study List

STUDY:

Bed, Rest, Awake, Tired, Dream, Wake, Snooze, Blanket, Doze, Slumber, Snore, Nap, Peace, Yawn, Drowsy

FALSE RECALL: “Sleep” 61%

Recall as a reconstructive process

Reconstruct study list based on the stored “gist”

The gist can be represented by a distribution over topics

Under a single topic assumption:

$$P(w_{n+1} \mid \mathbf{w}) = \sum_z P(w_{n+1} \mid z) P(z \mid \mathbf{w})$$

Retrieved word

Study list

$$P(w_{\text{recall}} | \mathbf{w}_{\text{study}}) = \sum_z P(w_{\text{recall}} | z) P(z | \mathbf{w}_{\text{study}}).$$

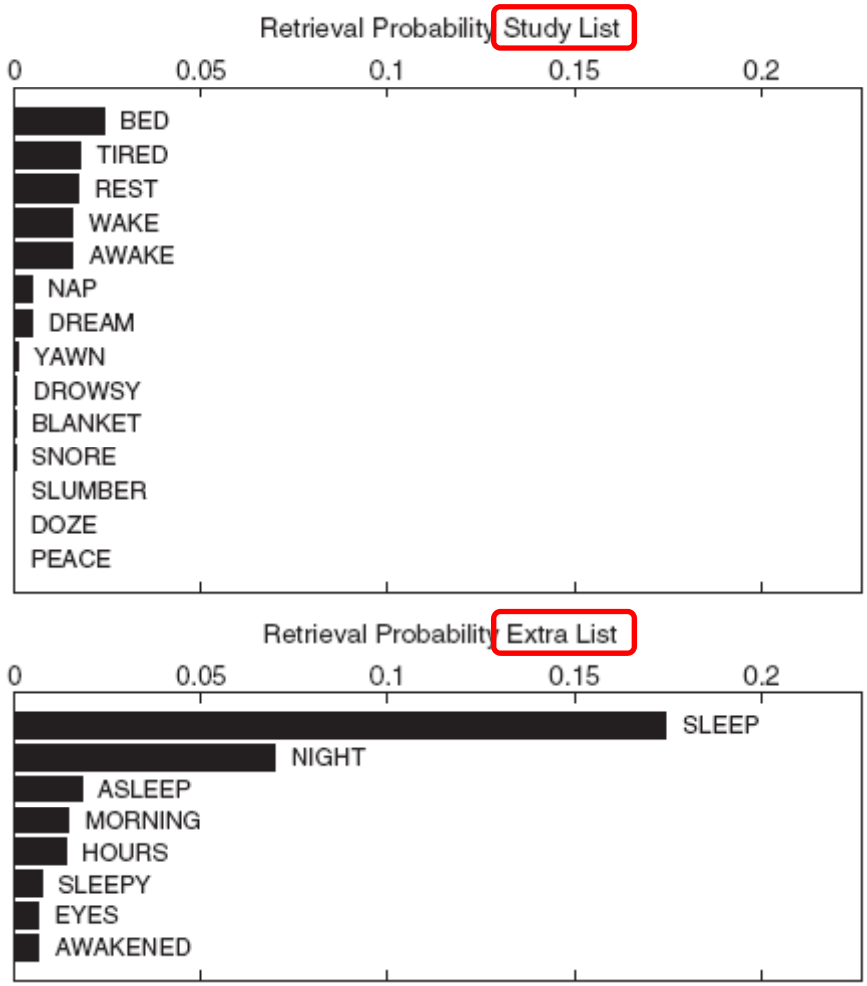


Figure 12. Retrieval probabilities, $P(w_{recall}|w_{study})$, for a study list containing words semantically associated with *sleep*. The upper panel shows the probabilities of each of the words on the study list. The lower panel shows the probabilities of the most likely extra-list words. *sleep* has a high retrieval probability and would thus be likely to be falsely recalled.

Predictions for the “Sleep” list

Word Sense Disambiguation

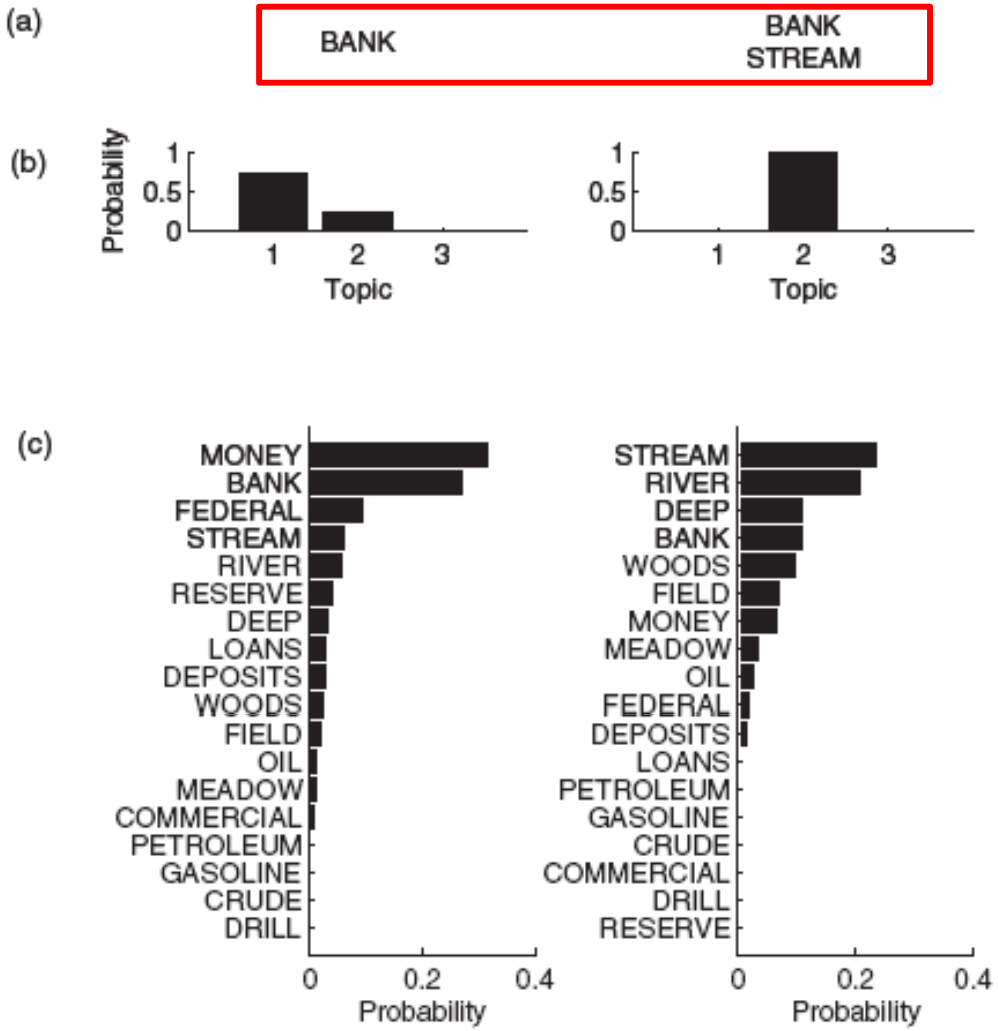


Figure 5. Prediction and disambiguation. (a) Words observed in a sentence, **w**. (b) The distribution over topics conditioned on those words, $P(z|\mathbf{w})$. (c) The predicted distribution over words resulting from summing over this distribution over topics, $P(w_{n+1}|\mathbf{w}) = \sum_z P(w_{n+1}|z)P(z|\mathbf{w})$. On seeing *bank*, the model is unsure whether the sentence concerns finance or the countryside. Subsequently seeing *stream* results in a strong conviction that *bank* does not refer to a financial institution.

Latent Semantic Analysis vs. Topics

The topic model and LSA use the same input—a word–document co-occurrence matrix—but they differ in how this input is analyzed and in the way that they represent the gist of documents and the meaning of words.

Quantitative differences

Qualitative differences

- probabilistic generative models can work with more structured representations
- Extensions of topic models:
  - hierarchies
  - syntax-semantics

Overview

I Associative memory

II The topic model

III Applications to associative memory

IV Applications in machine learning/text mining

Applications in Machine Learning

Automatically learn topics from large text collections

- NSF/NIH grant proposals
- 18th century newspapers
- Enron email

Topics provide quick overview of content

Enron email data

From: PGE News
To: ALL PGE EMPLOYEES
Date: 8/14/01 2:54PM
Subject: Jeff Skilling resigns as CEO of Enron

PGE News ..... August 14, 2001

Jeff Skilling resigns as CEO of Enron

Enron today announced that President and CEO Jeff Skilling has resigned, effective immediately, and that the Enron Board of Directors has asked Ken Lay to resume his role as Chairman and CEO.

"Stan Horton called this afternoon to inform me of Jeff's decision to step down for personal reasons," says PGE CEO and President Peggy Fowler. Horton, CEO of Enron Transportation, is Fowler's executive connection to the Enron team. "He wanted to let me know that Mr. Skilling's departure will not in any way impact Enron's ongoing strategy for success and we should expect no near-term dramatic organizational changes."

"Clearly, Enron will continue to focus on increasing the company's stock value," Fowler added. "PGE can help in this effort by remaining committed to our Scorecard goals and operational excellence."

Below is the letter Ken Lay is sending to Enron employees this afternoon announcing the decision:

To: Enron Employees Worldwide
From: Ken Lay

It is with regret that I have to announce that Jeff Skilling is leaving Enron. Today, the Board of Directors accepted his resignation as President and CEO of Enron. Jeff is resigning for personal reasons and his decision is voluntary. I regret his decision, but I accept and understand it. I have worked closely with Jeff for more than 15 years, including 11 here at Enron, and have had few, if any, professional relationships that I value more. I am pleased to say that he has agreed to enter into a consulting arrangement with the company to advise me and the Board of Directors.

Now it's time to look forward.

With Jeff leaving, the Board has asked me to resume the responsibilities of President and CEO in addition to my role as Chairman of the Board. I have agreed. I want to assure you that I have never felt better about the prospects for the company. All of you know that our stock price has suffered substantially over the last few months. One of my top priorities will be to restore a significant amount of the stock value we have lost as soon as possible. Our performance has never been stronger; our business model has never been more robust; our growth has never been more certain; and most importantly, we have never had a better nor deeper pool of talent throughout the company. We have the finest organization in American business today. Together, we will make Enron the world's leading company.

CC: Kathy & George Wyatt; Kathy Wyatt

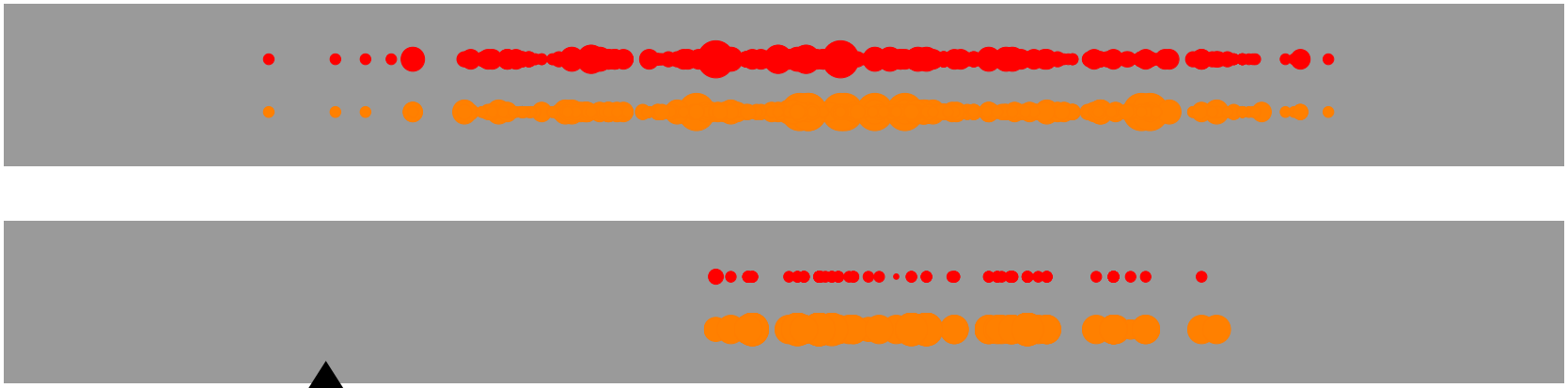
500,000 emails

5000 authors

1999-2002

Enron topics

TEXANS	GOD	ENVIRONMENTAL	FERC	POWER	STATE
WIN	LIFE	AIR	MARKET	CALIFORNIA	PLAN
FOOTBALL	MAN	MTBE	ISO	ELECTRICITY	CALIFORNIA
FANTASY	PEOPLE	EMISSIONS	COMMISSION	UTILITIES	DAVIS
SPORTSLINE	CHRIST	CLEAN	ORDER	PRICES	RATE
PLAY	FAITH	EPA	FILING	MARKET	BANKRUPTCY
TEAM	LORD	PENDING	COMMENTS	PRICE	SOCAL
GAME	JESUS	SAFETY	PRICE	UTILITY	POWER
SPORTS	SPIRITUAL	WATER	CALIFORNIA	CUSTOMERS	BONDS
GAMES	VISIT	GASOLINE	FILED	ELECTRIC	MOU



May 22, 2000
Start of California energy crisis

TIMELINE

NSF & NIH grant abstracts

Analyze 22,000+ active grants during 2002

- NIH – NIMH, NCI
- NSF – BIO, SBE

What topics are funded?

Topic map of funding programs

Example topics

BRAIN IMAGING	
brain	.101
fmri	.054
imaging	.054
functional	.046
mri	.033
subjects	.033
magnetic	.031
resonance	.029
neuroimaging	.028
structural	.018

CHILD PARENT INTERACTION	
children	.153
child	.089
parent	.038
parents	.032
family	.032
families	.022
early	.020
problems	.019
mothers	.017
risk	.017

HIV INTERVENTION	
hiv	.121
intervention	.064
risk	.050
sexual	.043
prevention	.037
aids	.024
interventions	.018
reduction	.015
behavior	.015
men	.013

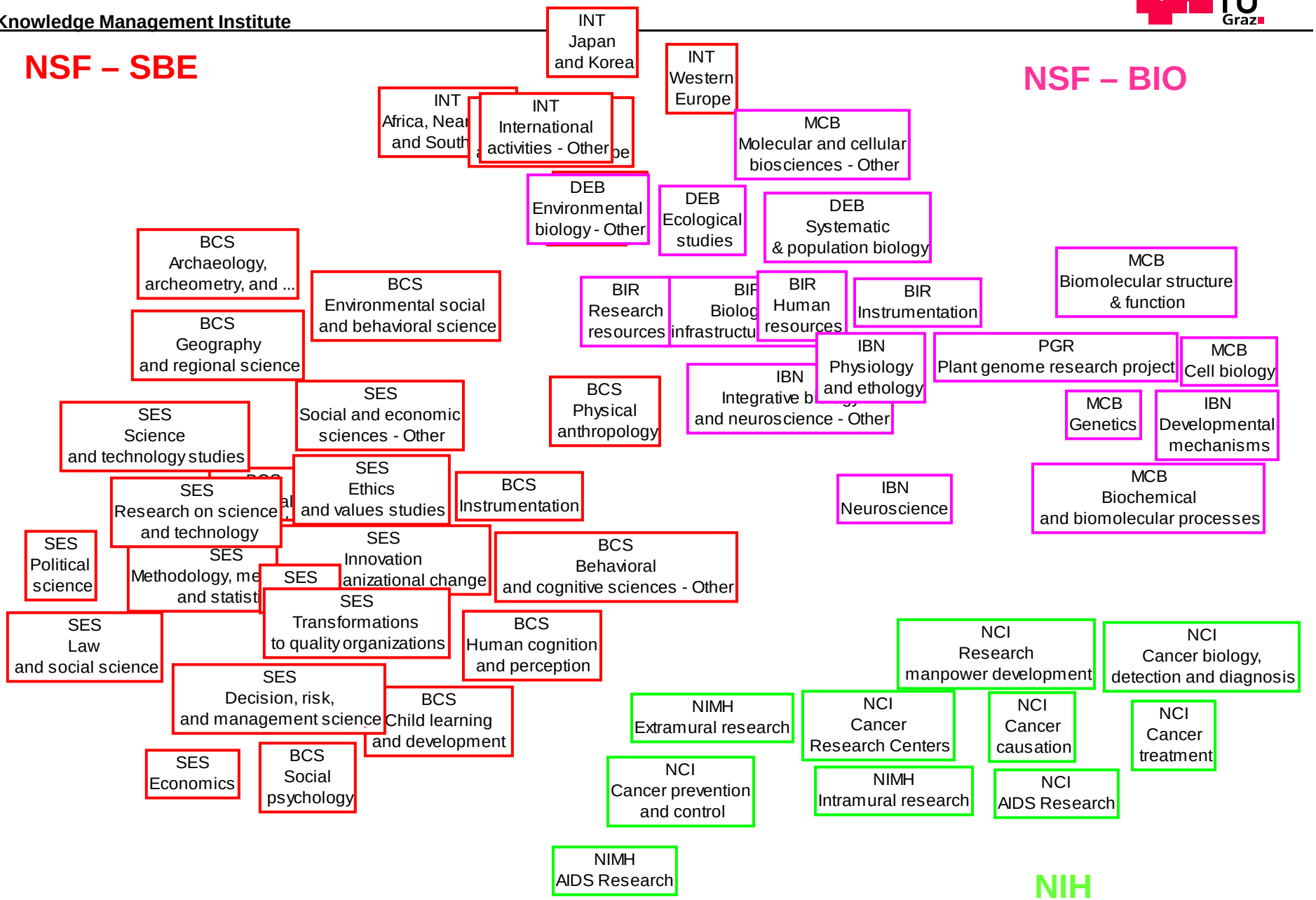
SCHIZOPHRENIA	
schizophrenia	.226
patients	.067
deficits	.054
schizophrenic	.027
psychosis	.024
subjects	.023
psychotic	.022
dysfunction	.019
abnormalities	.017
clinical	.015

VISUAL PROCESSING	
visual	.075
processing	.048
sensory	.035
spatial	.034
information	.022
eye	.020
stimuli	.020
object	.019
objects	.019
perception	.018

MEMORY	
memory	.237
working	.049
memories	.022
tasks	.022
retrieval	.021
encoding	.020
cognitive	.019
processing	.019
recognition	.018
performance	.016

AGING	
older	.083
adults	.071
age	.066
elderly	.041
geriatric	.041
life	.039
aging	.033
late	.032
cognitive	.028
aged	.022

ALZHEIMER DISEASE	
disease	.102
ad	.074
alzheimer	.043
diabetes	.025
cardiovascular	.016
insulin	.015
vascular	.015
blood	.013
clinical	.012
individuals	.012


NSF – SBE
NSF – BIO
INT Japan and Korea
INT Western Europe
INT Africa, Near and South
INT International activities - Other
MCB Molecular and cellular biosciences - Other
DEB Environmental biology - Other
DEB Ecological studies
DEB Systematic & population biology
BCS Archaeology, archeometry, and ...
BCS Environmental social and behavioral science
BCS Geography and regional science
BIR Research resources
BIR Biolog infrastructu
BIR Human resources
BIR Instrumentation
MCB Biomolecular structure & function
IBN Physiology and ethology
PGR Plant genome research project
MCB Cell biology
IBN Integrative b and neuroscience - Other
BCS Physical anthropology
MCB Genetics
IBN Developmental mechanisms
SES Science and technology studies
SES Social and economic sciences - Other
SES Ethics and values studies
SES Research on science and technology
BCS Instrumentation
IBN Neuroscience
MCB Biochemical and biomolecular processes
SES Political science
SES Methodology, me and statist
SES
SES Innovation anizational change
BCS Behavioral and cognitive sciences - Other
SES Transformations to quality organizations
SES Law and social science
BCS Human cognition and perception
NCI Research manpower development
NCI Cancer biology, detection and diagnosis
SES Decision, risk, and management science
BCS Child learning and development
NIMH Extramural research
NCI Cancer Research Centers
NCI Cancer causation
NCI Cancer treatment
SES Economics
BCS Social psychology
NCI Cancer prevention and control
NIMH Intramural research
NCI AIDS Research
NIMH AIDS Research
NIH

Conclusion

Semantic association as probabilistic inference

Generative models are useful

- makes modeling assumptions explicit
- flexible

Cognitive Science ←→ Machine Learning

Questions?

See you!