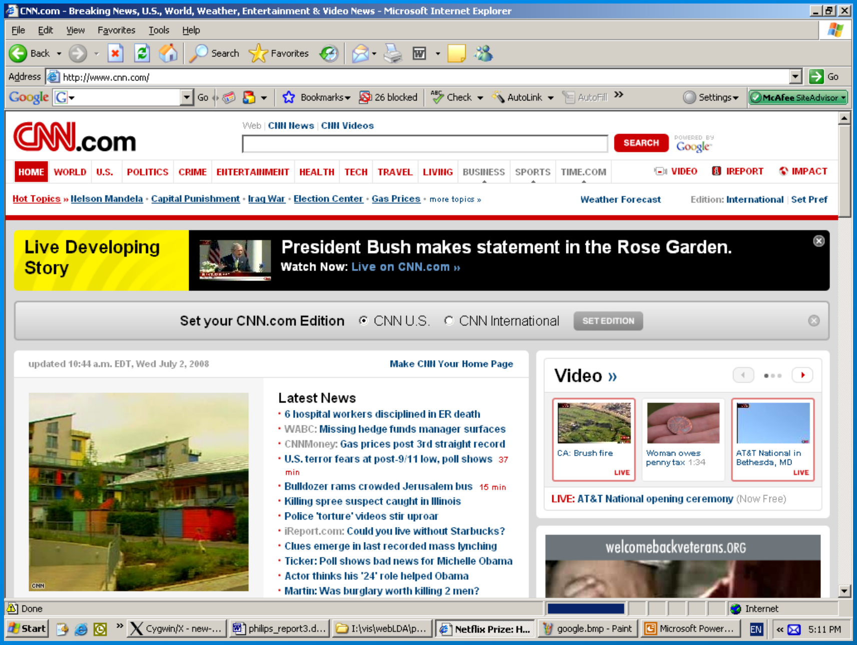
Latent Topic Models for Hypertext

Amit Gruber[1]
Michal Rosen-Zvi[2]
Yair Weiss[1]

[1]The Hebrew University of Jerusalem
[2]IBM Research Labs, Haifa

Introduction



CNN.com - Breaking News, U.S., World, Weather, Entertainment & Video News - Microsoft Internet Explorer

http://www.cnn.com/

CNN.com

Hot Topics » Nelson Mandela • Capital Punishment • Iraq War • Election Center • Gas Prices • more topics »

Weather Forecast

Edition: International | Set Pref

Live Developing Story

President Bush makes statement in the Rose Garden.

Watch Now: Live on CNN.com »

Set your CNN.com Edition ● CNN U.S. ○ CNN International

updated 10:44 a.m. EDT, Wed July 2, 2008

Make CNN Your Home Page

Latest News

- 6 hospital workers disciplined in ER death
- WABC: Missing hedge funds manager surfaces
- CNNMoney: Gas prices post 3rd straight record
- U.S. terror fears at post-9/11 low, poll shows 37 min
- Bulldozer rams crowded Jerusalem bus 15 min
- Killing spree suspect caught in Illinois
- Police 'torture' videos stir uproar
- iReport.com: Could you live without Starbucks?
- Clues emerge in last recorded mass lynching
- Ticker: Poll shows bad news for Michelle Obama
- Actor thinks his '24' role helped Obama
- Martin: Was burglary worth killing 2 men?

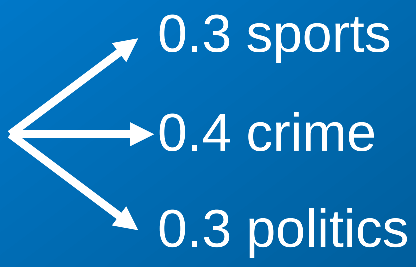
CA: Brush fire LIVE

Woman owes penny tax 1:34

AT&T National in Bethesda, MD LIVE

LIVE: AT&T National opening ceremony (Now Free)

welcomebackveterans.ORG

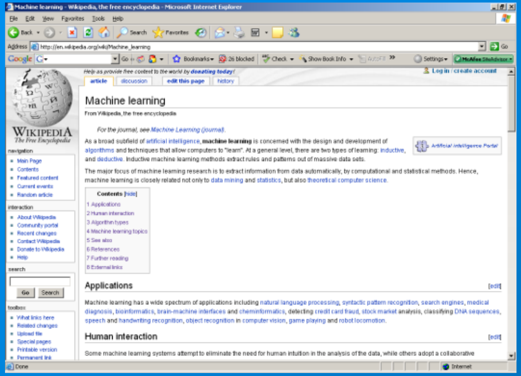

0.3 sports

0.4 crime

0.3 politics

- In this work we focus on hypertext documents, i.e. documents with links

Hypertext Documents


Machine learning - Wikipedia, the free encyclopedia - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://en.wikipedia.org/wiki/Machine_learning Go
Bookmarks 26 blocked Check Show Book Info AutoFill Settings McAfee SiteAdvisor
Help us provide free content to the world by donating today!
Log in / create account
article discussion edit this page history
Wikipedia The Free Encyclopedia
Search
Internet

Machine learning

From Wikipedia, the free encyclopedia

For the journal, see Machine Learning (journal).

As a broad subfield of artificial intelligence, **machine learning** is concerned with the design and development of algorithms and techniques that allow computers to "learn". At a general level, there are two types of learning: inductive, and deductive. Inductive machine learning methods extract rules and patterns out of massive data sets.

Artificial intelligence Portal

The major focus of machine learning research is to extract information from data automatically, by computational and statistical methods. Hence, machine learning is closely related not only to data mining and statistics, but also theoretical computer science.

Contents [hide]

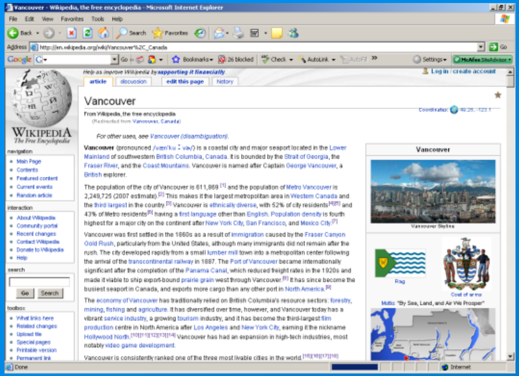
1 Applications
2 Human interaction
3 Algorithm types
4 Machine learning topics
5 See also
6 References
7 Further reading
8 External links

Applications [edit]

Machine learning has a wide spectrum of applications including natural language processing, syntactic pattern recognition, search engines, medical diagnosis, bioinformatics, brain-machine interfaces and cheminformatics, detecting credit card fraud, stock market analysis, classifying DNA sequences, speech and handwriting recognition, object recognition in computer vision, game playing and robot locomotion.

Human interaction [edit]

Some machine learning systems attempt to eliminate the need for human intuition in the analysis of the data, while others adopt a collaborative


Vancouver - Wikipedia, the free encyclopedia - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://en.wikipedia.org/wiki/Vancouver%2C_Canada Go
Bookmarks 26 blocked Check AutoLink AutoFill Settings McAfee SiteAdvisor
Help us improve Wikipedia by supporting it financially.
Log in / create account
article discussion edit this page history
Wikipedia The Free Encyclopedia
Search
Done
Internet

Vancouver

From Wikipedia, the free encyclopedia
(Redirected from Vancouver, Canada)

Coordinates: 49.25, -123.1

For other uses, see Vancouver (disambiguation).

Vancouver (pronounced /væn'ku ː vəɹ/) is a coastal city and major seaport located in the Lower Mainland of southwestern British Columbia, Canada. It is bounded by the Strait of Georgia, the Fraser River, and the Coast Mountains. Vancouver is named after Captain George Vancouver, a British explorer.

The population of the city of Vancouver is 611,869 [1] and the population of Metro Vancouver is 2,249,725 (2007 estimate).[2] This makes it the largest metropolitan area in Western Canada and the third largest in the country.[3] Vancouver is ethnically diverse, with 52% of city residents[4][5] and 43% of Metro residents[6] having a first language other than English. Population density is fourth highest for a major city on the continent after New York City, San Francisco, and Mexico City.[7]

Vancouver was first settled in the 1860s as a result of immigration caused by the Fraser Canyon Gold Rush, particularly from the United States, although many immigrants did not remain after the rush. The city developed rapidly from a small lumber mill town into a metropolitan center following the arrival of the transcontinental railway in 1887. The Port of Vancouver became internationally significant after the completion of the Panama Canal, which reduced freight rates in the 1920s and made it viable to ship export-bound prairie grain west through Vancouver.[8] It has since become the busiest seaport in Canada, and exports more cargo than any other port in North America.[9]

The economy of Vancouver has traditionally relied on British Columbia's resource sectors: forestry, mining, fishing and agriculture. It has diversified over time, however, and Vancouver today has a vibrant service industry, a growing tourism industry, and it has become the third-largest film production centre in North America after Los Angeles and New York City, earning it the nickname Hollywood North.[10][11][12][13][14] Vancouver has had an expansion in high-tech industries, most notably video game development.

Vancouver is consistently ranked one of the three most livable cities in the world.[15][16][17][18]

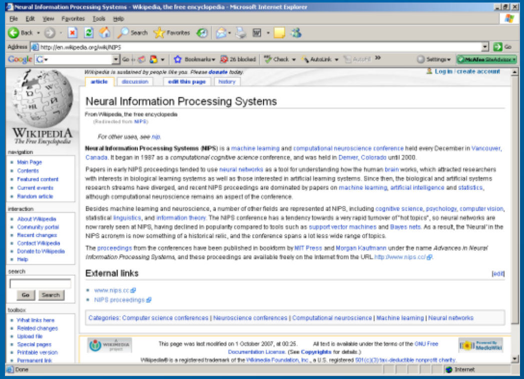
Vancouver

Vancouver Skyline

Flag

Coat of arms

Motto: "By Sea, Land, and Air We Prosper"


Neural Information Processing Systems - Wikipedia, the free encyclopedia - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://en.wikipedia.org/wiki/NIPS Go
Bookmarks 26 blocked Check AutoLink AutoFill Settings McAfee SiteAdvisor
Wikipedia is sustained by people like you. Please donate today.
Log in / create account
article discussion edit this page history
Wikipedia The Free Encyclopedia
Search
Done
Internet

Neural Information Processing Systems

From Wikipedia, the free encyclopedia
(Redirected from NIPS)

For other uses, see nip.

Neural Information Processing Systems (**NIPS**) is a machine learning and computational neuroscience conference held every December in Vancouver, Canada. It began in 1987 as a *computational cognitive science* conference, and was held in Denver, Colorado until 2000.

Papers in early NIPS proceedings tended to use neural networks as a tool for understanding how the human brain works, which attracted researchers with interests in biological learning systems as well as those interested in artificial learning systems. Since then, the biological and artificial systems research streams have diverged, and recent NIPS proceedings are dominated by papers on machine learning, artificial intelligence and statistics, although computational neuroscience remains an aspect of the conference.

Besides machine learning and neuroscience, a number of other fields are represented at NIPS, including cognitive science, psychology, computer vision, statistical linguistics, and information theory. The NIPS conference has a tendency towards a very rapid turnover of "hot topics", so neural networks are now rarely seen at NIPS, having declined in popularity compared to tools such as support vector machines and Bayes nets. As a result, the 'Neural' in the NIPS acronym is now something of a historical relic, and the conference spans a lot less wide range of topics.

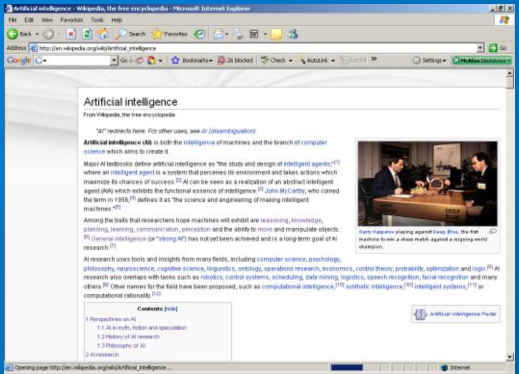
The proceedings from the conferences have been published in bookform by MIT Press and Morgan Kaufmann under the name *Advances in Neural Information Processing Systems*, and these proceedings are available freely on the Internet from the URL http://www.nips.cc/.

External links [edit]

- www.nips.cc
- NIPS proceedings

Categories: Computer science conferences | Neuroscience conferences | Computational neuroscience | Machine learning | Neural networks

This page was last modified on 1 October 2007, at 00:25. All text is available under the terms of the GNU Free Documentation License. (See Copyrights for details.)
Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a U.S. registered 501(c)(3) tax-deductible nonprofit charity.


Artificial intelligence - Wikipedia, the free encyclopedia - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://en.wikipedia.org/wiki/Artificial_intelligence Go
Bookmarks 26 blocked Check AutoLink AutoFill Settings McAfee SiteAdvisor
Opening page http://en.wikipedia.org/wiki/Artificial_intelligence...
Internet

Artificial intelligence

From Wikipedia, the free encyclopedia

"AI" redirects here. For other uses, see AI (disambiguation).

Artificial intelligence (AI) is both the intelligence of machines and the branch of computer science which aims to create it.

Major AI textbooks define artificial intelligence as "the study and design of intelligent agents,"[1] where an intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success.[2] AI can be seen as a realization of an abstract intelligent agent (AIA) which exhibits the functional essence of intelligence.[3] John McCarthy, who coined the term in 1956,[4] defines it as "the science and engineering of making intelligent machines."[5]

Among the traits that researchers hope machines will exhibit are reasoning, knowledge, planning, learning, communication, perception and the ability to move and manipulate objects.[6] General intelligence (or "strong AI") has not yet been achieved and is a long-term goal of AI research.[7]

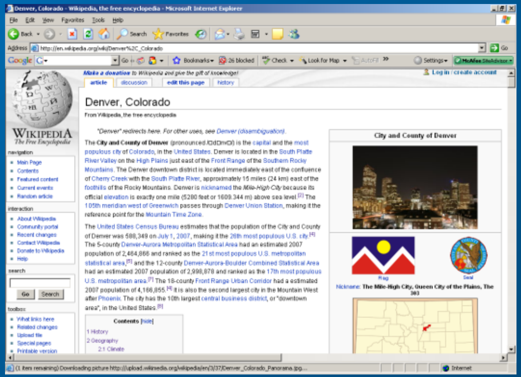
Garry Kasparov playing against Deep Blue, the first machine to win a chess match against a reigning world champion.

AI research uses tools and insights from many fields, including computer science, psychology, philosophy, neuroscience, cognitive science, linguistics, ontology, operations research, economics, control theory, probability, optimization and logic.[8] AI research also overlaps with tasks such as robotics, control systems, scheduling, data mining, logistics, speech recognition, facial recognition and many others.[9] Other names for the field have been proposed, such as computational intelligence,[10] synthetic intelligence,[10] intelligent systems,[11] or computational rationality.[12]

Contents [hide]

1 Perspectives on AI
1.1 AI in myth, fiction and speculation
1.2 History of AI research
1.3 Philosophy of AI
2 AI research

Artificial intelligence Portal


Denver, Colorado - Wikipedia, the free encyclopedia - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://en.wikipedia.org/wiki/Denver%2C_Colorado Go
Bookmarks 26 blocked Check Look for Map AutoFill Settings McAfee SiteAdvisor
Make a donation to Wikipedia and give the gift of knowledge!
Log in / create account
article discussion edit this page history
Wikipedia The Free Encyclopedia
Search
(1 items remaining) Downloading picture http://upload.wikimedia.org/wikipedia/en/3/37/Denver_Colorado_Panorama.jpg...
Internet

Denver, Colorado

From Wikipedia, the free encyclopedia

"Denver" redirects here. For other uses, see Denver (disambiguation).

The **City and County of Denver** (pronounced /ˈdɛnvɚ/) is the capital and the most populous city of Colorado, in the United States. Denver is located in the South Platte River Valley on the High Plains just east of the Front Range of the Southern Rocky Mountains. The Denver downtown district is located immediately east of the confluence of Cherry Creek with the South Platte River, approximately 15 miles (24 km) east of the foothills of the Rocky Mountains. Denver is nicknamed the *Mile-High City* because its official elevation is exactly one mile (5280 feet or 1609.344 m) above sea level.[2] The 105th meridian west of Greenwich passes through Denver Union Station, making it the reference point for the Mountain Time Zone.

The United States Census Bureau estimates that the population of the City and County of Denver was 588,349 on July 1, 2007, making it the 26th most populous U.S. city.[4] The 5-county Denver-Aurora Metropolitan Statistical Area had an estimated 2007 population of 2,464,866 and ranked as the 21st most populous U.S. metropolitan statistical area,[5] and the 12-county Denver-Aurora-Boulder Combined Statistical Area had an estimated 2007 population of 2,998,878 and ranked as the 17th most populous U.S. metropolitan area.[7] The 18-county Front Range Urban Corridor had a estimated 2007 population of 4,166,855.[4] It is also the second largest city in the Mountain West after Phoenix. The city has the 10th largest central business district, or "downtown area", in the United States.[8]

Contents [hide]

1 History
2 Geography
2.1 Climate

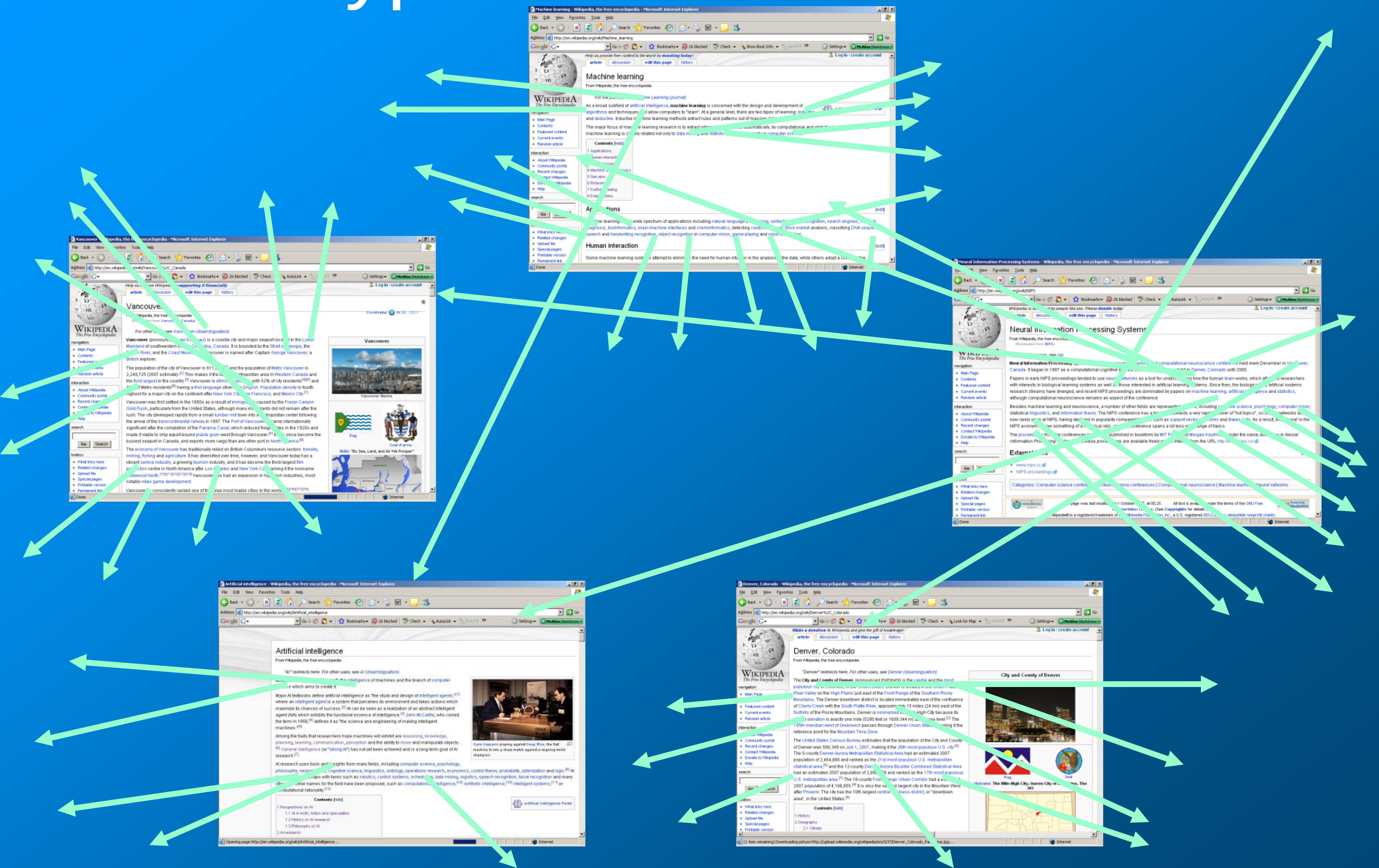
City and County of Denver

Flag

Seal

Nickname: The Mile-High City, Queen City of the Plains, The 303

Hypertext Documents



Machine learning - Wikipedia, the free encyclopedia - Microsoft Internet Explorer

http://en.wikipedia.org/wiki/Machine_learning

Machine learning

From Wikipedia, the free encyclopedia

As a broad subfield of artificial intelligence, **machine learning** is concerned with the design and development of algorithms and techniques that allow computers to "learn". At a general level, there are two types of learning: inductive and deductive. Inductive machine learning methods extract rules and patterns out of massive data sets.

The major focus of machine learning research is to extract information from data automatically, by computational and statistical methods. Hence, machine learning is closely related not only to data mining and statistics, but also theoretical computer science.

Contents [hide]
1 Applications
4 Machine learning topics
5 See also
6 References
7 Further reading
8 External links

Applications

Machine learning has a wide spectrum of applications including natural language processing, syntactic pattern recognition, search engines, medical diagnosis, bioinformatics, brain-machine interfaces and cheminformatics, detecting credit card fraud, stock market analysis, classifying DNA sequences, speech and handwriting recognition, object recognition in computer vision, game playing and robot locomotion.

Human interaction

Some machine learning systems attempt to eliminate the need for human intuition in the analysis of the data, while others adopt a collaborative

Vancouver - Wikipedia, the free encyclopedia - Microsoft Internet Explorer

Vancouver

From Wikipedia, the free encyclopedia

For other uses, see Vancouver (disambiguation).

Vancouver is a coastal city and major seaport located in the Lower Mainland of southwestern British Columbia, Canada. It is bounded by the Strait of Georgia, the Fraser River, and the Coast Mountains. Vancouver is named after Captain George Vancouver, a British explorer.

The population of the city of Vancouver is 611,869 and the population of Metro Vancouver is 2,249,725 (2007 estimate).[2] This makes it the largest metropolitan area in Western Canada and the third largest in the country.[3] Vancouver is ethnically diverse, with 52% of city residents[4][5] and 41% of Metro residents[6] having a first language other than English. Population density is fourth highest for a major city on the continent after New York City, San Francisco, and Mexico City.[7]

Vancouver was first settled in the 1860s as a result of immigration caused by the Fraser Canyon Gold Rush, particularly from the United States, although many immigrants did not remain after the rush. The city developed rapidly from a small lumber mill town into a metropolitan center following the arrival of the transcontinental railway in 1887. The Port of Vancouver became internationally significant after the completion of the Panama Canal, which reduced freight rates in the 1920s and made it viable to ship export-bound prairie grain west through Vancouver.[8] It has since become the busiest seaport in Canada, and exports more cargo than any other port in North America.[9]

The economy of Vancouver has traditionally relied on British Columbia's resource sectors: forestry, mining, fishing and agriculture. It has diversified over time, however, and Vancouver today has a vibrant service industry, a growing tourism industry, and it has become the third-largest film production centre in North America after Los Angeles and New York City, earning it the nickname Hollywood North.[10][11][12][13][14] Vancouver has had an expansion in high-tech industries, most notably video game development.

Vancouver is consistently ranked one of the three most livable cities in the world.

Vancouver

Vancouver Skyline

Flag

Coat of arms

Motto: "By Sea, Land, and Air We Prosper"

Neural Information Processing Systems - Wikipedia, the free encyclopedia - Microsoft Internet Explorer

http://en.wikipedia.org/wiki/NIPS

Neural Information Processing Systems

From Wikipedia, the free encyclopedia
(Redirected from NIPS)

Neural Information Processing Systems is a machine learning and computational neuroscience conference held every December in Vancouver, Canada. It began in 1987 as a computational cognitive science conference held in Denver, Colorado until 2000.

Papers in early NIPS proceedings tended to use neural networks as a tool for understanding how the human brain works, which attracted researchers with interests in biological learning systems as well as those interested in artificial learning systems. Since then, the biological and artificial systems research streams have diverged, and recent NIPS proceedings are dominated by papers on machine learning, artificial intelligence and statistics, although computational neuroscience remains an aspect of the conference.

Besides machine learning and neuroscience, a number of other fields are represented at NIPS, including cognitive science, psychology, computer vision, statistical linguistics, and information theory. The NIPS conference has a tendency towards a very rapid turnover of "hot topics", so neural networks are now rarely seen at NIPS, having declined in popularity compared to topics such as support vector machines and Bayesian nets. As a result, the "neural" in the NIPS acronym is now something of a historical relic, and the conference spans a lot less wide range of topics.

The proceedings from the conferences have been published in bookform by MIT Press and Morgan Kaufmann under the name *Advances in Neural Information Processing Systems*. These proceedings are available freely on the internet from the URL http://books.nips.cc/.

External links

- www.nips.cc
- NIPS proceedings

Categories: Computer science conferences | Neuroscience conferences | Computational neuroscience | Machine learning | Neural networks

Artificial intelligence - Wikipedia, the free encyclopedia - Microsoft Internet Explorer

http://en.wikipedia.org/wiki/Artificial_intelligence

Artificial intelligence

From Wikipedia, the free encyclopedia

"AI" redirects here. For other uses, see AI (disambiguation).

intelligence of machines and the branch of computer science which aims to create it.

Major AI textbooks define artificial intelligence as "the study and design of intelligent agents,"[1] where an intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success.[2] AI can be seen as a realization of an abstract intelligent agent (AIA) which exhibits the functional essence of intelligence.[3] John McCarthy, who coined the term in 1956,[4] defines it as "the science and engineering of making intelligent machines."[5]

Among the traits that researchers hope machines will exhibit are reasoning, knowledge, planning, learning, communication, perception and the ability to move and manipulate objects.[6] General intelligence (or "strong AI") has not yet been achieved and is a long-term goal of AI research.[7]

AI research uses tools and insights from many fields, including computer science, psychology, philosophy, neuroscience, cognitive science, linguistics, ontology, operations research, economics, control theory, probability, optimization and logic.[8] AI research also overlaps with tasks such as robotics, control systems, scheduling, data mining, logistics, speech recognition, facial recognition and many others. Other names for the field have been proposed, such as computational intelligence,[10] synthetic intelligence,[10] intelligent systems,[11] or computational rationality.[12]

Garry Kasparov playing against Deep Blue, the first machine to win a chess match against a reigning world champion.

Contents [hide]
1 Perspectives on AI
1.1 AI in myth, fiction and speculation
1.2 History of AI research
1.3 Philosophy of AI
2 AI research

Artificial intelligence Portal

Denver, Colorado - Wikipedia, the free encyclopedia - Microsoft Internet Explorer

http://en.wikipedia.org/wiki/Denver%2C_Colorado

Denver, Colorado

From Wikipedia, the free encyclopedia

"Denver" redirects here. For other uses, see Denver (disambiguation).

The **City and County of Denver** is the capital and the most populous city of Colorado, in the United States. Denver is located in the South Platte River Valley on the High Plains just east of the Front Range of the Southern Rocky Mountains. The Denver downtown district is located immediately east of the confluence of Cherry Creek with the South Platte River, approximately 15 miles (24 km) east of the foothills of the Rocky Mountains. Denver is nicknamed the *Mile-High City* because its official elevation is exactly one mile (5280 feet or 1609.344 m) above sea level.[2] The 105th meridian west of Greenwich passes through Denver Union Station, making it the reference point for the Mountain Time Zone.

The United States Census Bureau estimates that the population of the City and County of Denver was 588,349 on July 1, 2007, making it the 26th most populous U.S. city.[4] The 5-county Denver-Aurora Metropolitan Statistical Area had an estimated 2007 population of 2,464,866 and ranked as the 21st most populous U.S. metropolitan statistical area,[5] and the 12-county Denver-Aurora-Boulder Combined Statistical Area had an estimated 2007 population of 2,99 78 and ranked as the 17th most populous U.S. metropolitan area.[7] The 18-county Front Range Urban Corridor had a estimated 2007 population of 4,166,855.[4] It is also the second largest city in the Mountain West after Phoenix. The city has the 10th largest central business district, or "downtown area", in the United States.[8]

Contents [hide]
1 History
2 Geography
2.1 Climate

City and County of Denver

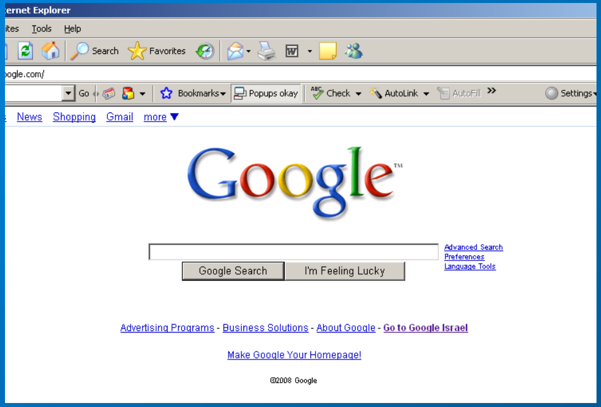
Flag

Seal

Nickname: The Mile-High City, Queen City of the Plains, The 303

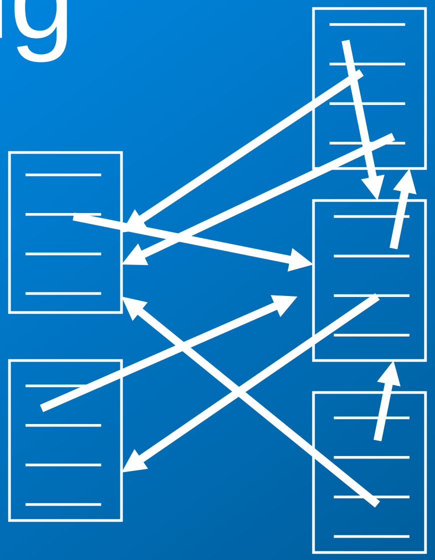
Introduction

- Hypertext is everywhere !
  - web pages, refs. in scientific publications
- Connectivity is important
  - PageRank



- Topology of the WWW is complicated

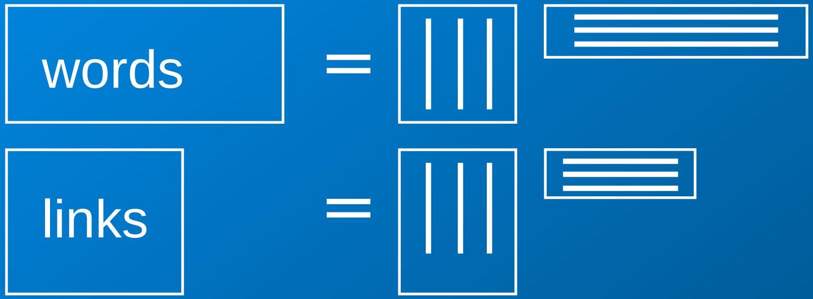
Problem Setting



- Input: documents and links
- Estimate:
  - – Document topic mixture
  - – Pr(word | topic)
  - – Document importance
- Unsupervised

Previous Work:
Topic Models for Hypertext

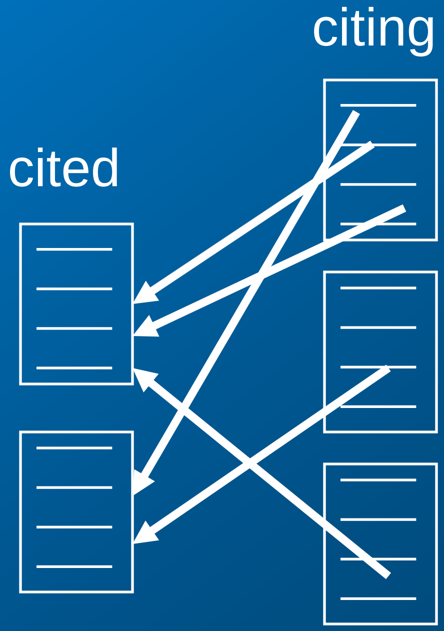
- Cohn and Hofmann, ’01.
- Erosheva et al. ’04.
  - Links are modeled similar to words
  - Links are not associated with words



words =

links =

- Dietz et al. ’07.
- Nallapati and Cohen, ’08.
  - Distinguish between citing and cited docs



citing

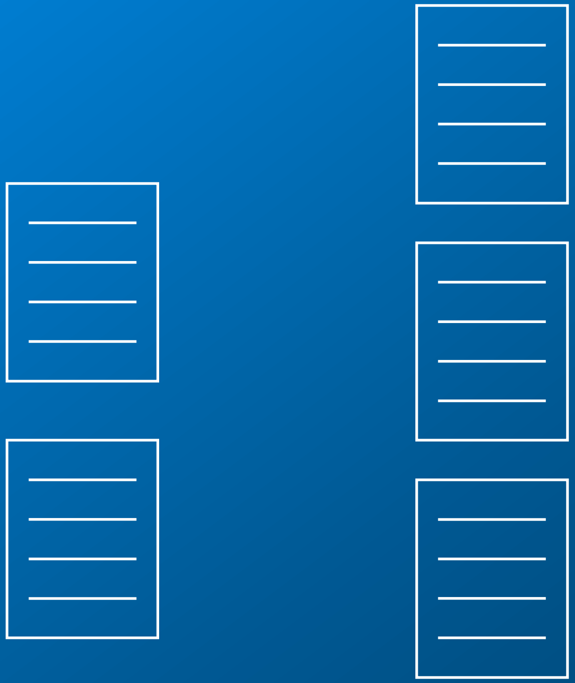
cited

The Latent Topic Hypertext Model

(LTHM)

LTHM: Generative Model

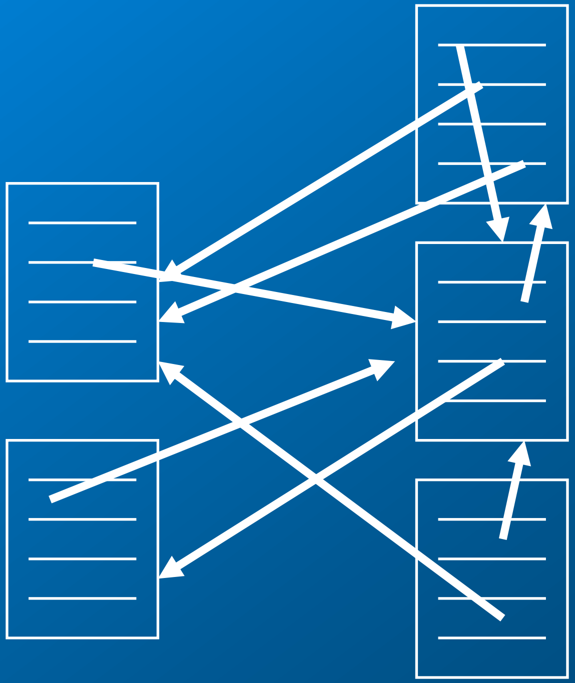
1. Words are created (by LDA)



LTHM: Generative Model

1. Words are created (by LDA)

2. Links are created (our contribution)



LTHM: Modeling Links

- Allows for arbitrary topology of the citation graph (including self links)
- A link points from a word to a document

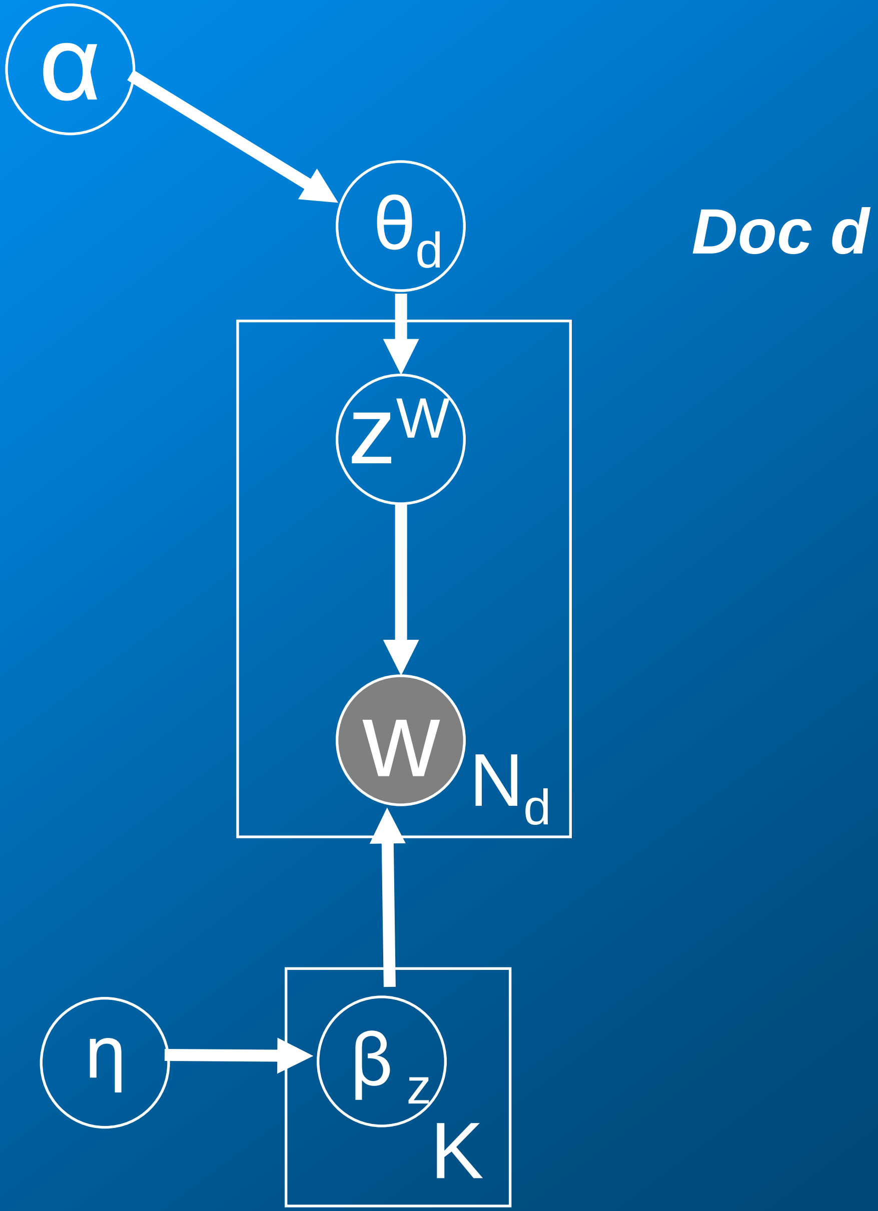
LTHM: Link Generation

- Depends on:
  - The topic of the anchor word
  - The topic mixture of the target document
  - The importance of the target document

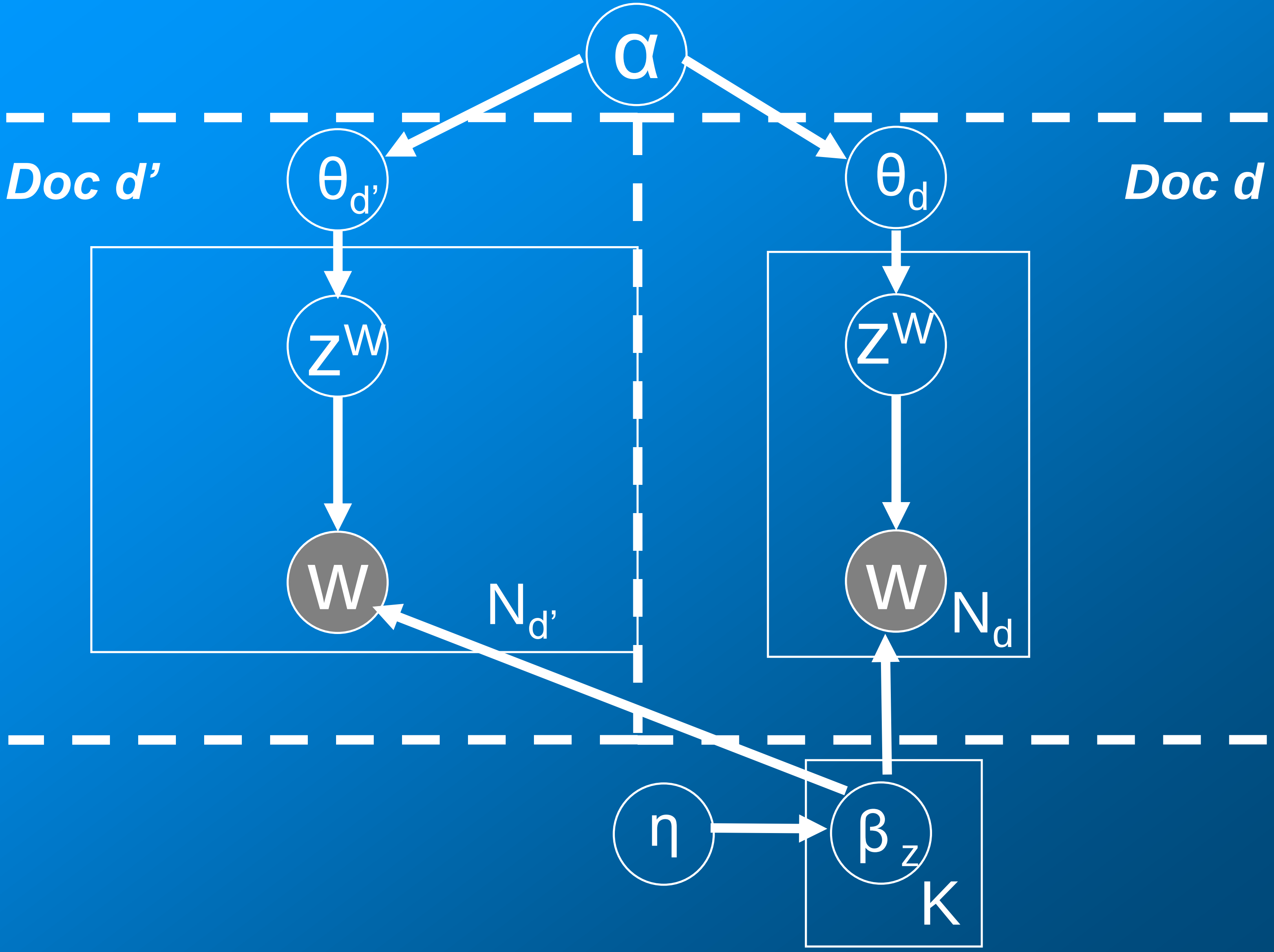
$$\Pr(\text{link} = d \mid \text{topic} = z) = \underbrace{\lambda_d}_{\text{importance of d}} \underbrace{\theta_d}_{\text{prevalence of z in d}}(z)$$

Generating a single document d

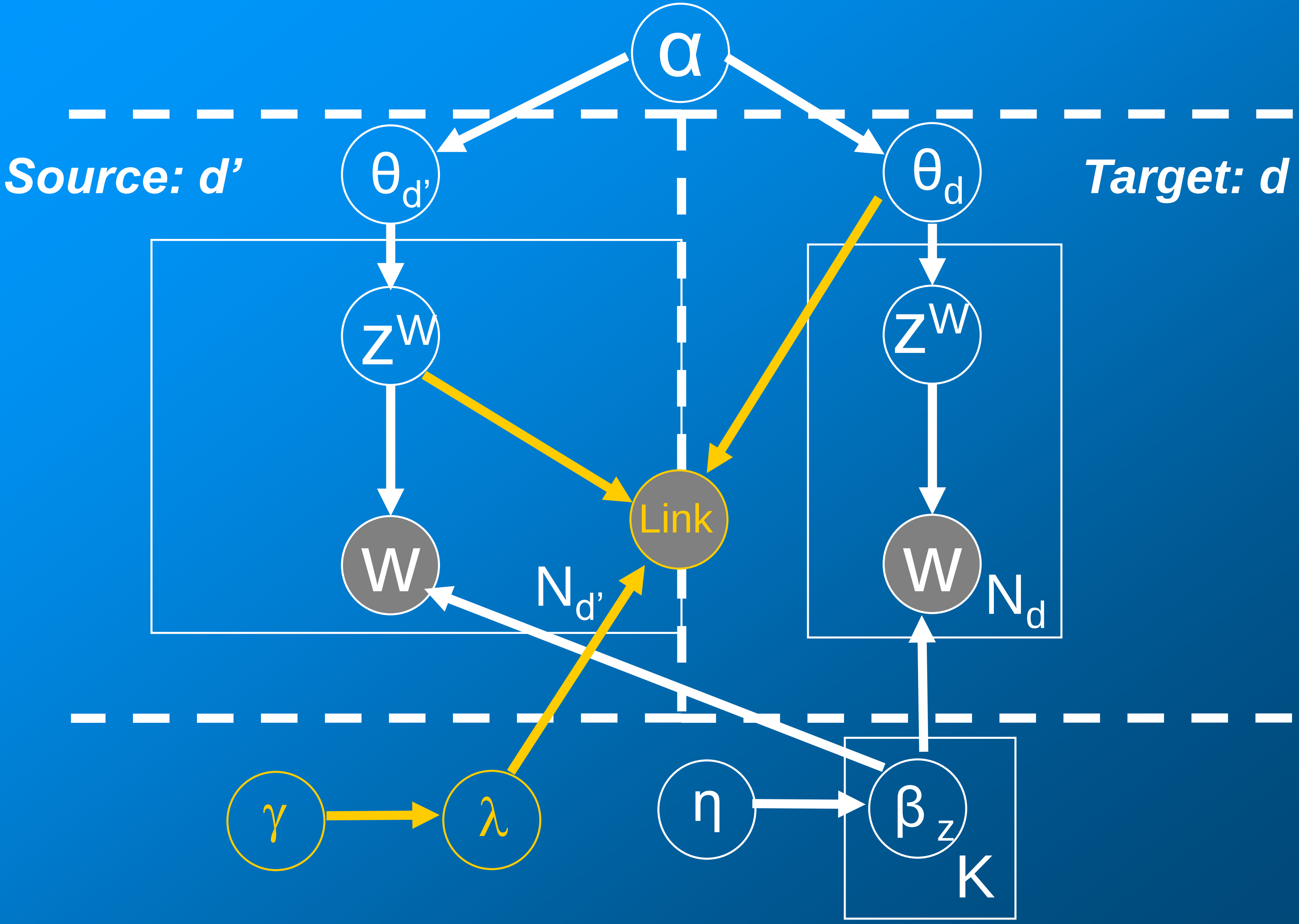
d is generated by Latent Dirichlet Allocation



Generating two documents d’ and d



Generating Links from d’ to d



Properties of the Model

- D additional parameters (λ) for links vs. DxK parameters in previous models
- The existence (or non-existence) of a link is an observation
- A link shares the same topic with the word
- Link affects topic estimation in both the source and target documents

Approximate Learning

- Exact inference is intractable in hierarchical models such as LDA
- Approximate inference in LTHM is even more challenging as non-links are also observations
- Using symmetries, we derived an O(K*corpus size) EM algorithm

Experiments

- WebKB dataset

  8282 documents

  12911 links

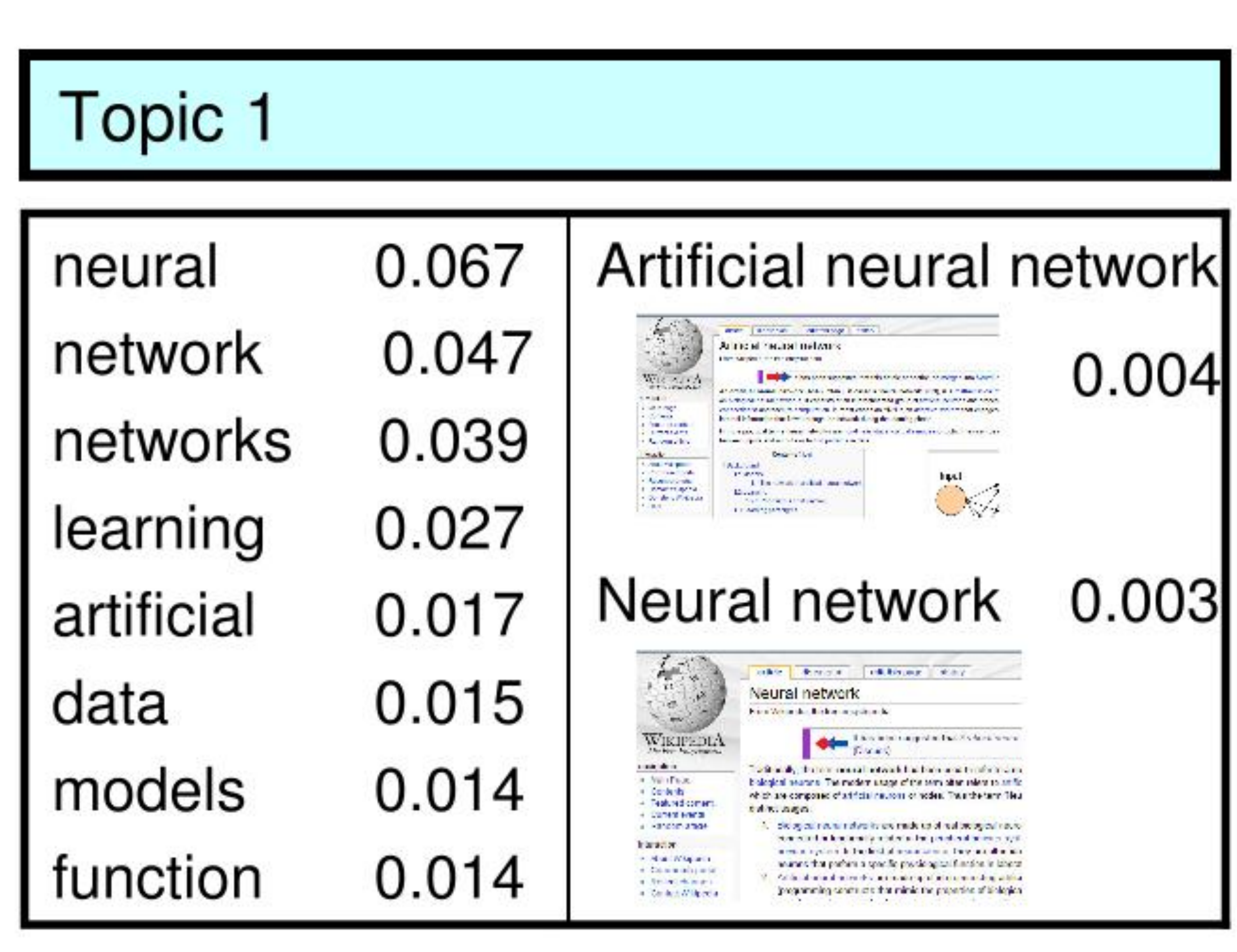
- Wikipedia

  A new data set, collected by crawling from the NIPS Wikipedia page

  105 documents

  790 links

Experiments: Wikipedia



Topic 1

neural	0.067	Artificial neural network	0.004
network	0.047		
networks	0.039		
learning	0.027		
artificial	0.017	Neural network	0.003
data	0.015		
models	0.014		
function	0.014		

Experiments: Wikipedia

Topic 2

recognition	0.058
speech	0.033
language	0.015
pattern	0.012
handwriting	0.011
evaluation	0.010
robots	0.010
systems	0.009

Speech recognition

0.004

Pattern recognition

0.004

Experiments: Wikipedia

Topic 3

vancouver	0.051
denver	0.043
city	0.041
retrieved	0.024
colorado	0.011
area	0.009
population	0.009
canada	0.008

Denver, Colorado

Denver, Colorado

0.0008

Vancouver 0.0002

Vancouver

Experiments: Wikipedia

Topic 4

brain	0.047
cognitive	0.026
science	0.016
press	0.011
neurons	0.010
mind	0.010
systems	0.010
human	0.010

Cognitive science 0.003

Cognitive science

Neuroscience 0.002

navigation

- Main Page
- Contents
- Featured content
- Current events
- Random article

Neuroscience

From Wikipedia, the free encyclopedia

Neuroscience is a field that is devoted to the sci
evolutionary history, development, genetics, bioch
pathology of the nervous system. Traditionally it i
the convergence of interest from many allied disci
physics, and medicine. The scope of neuroscienc
investigation of the central and peripheral nervous
been enormously expanded, from biochemical an
to imaging representations of perceptual and moto

Journal of Machine Learning Research

LDA

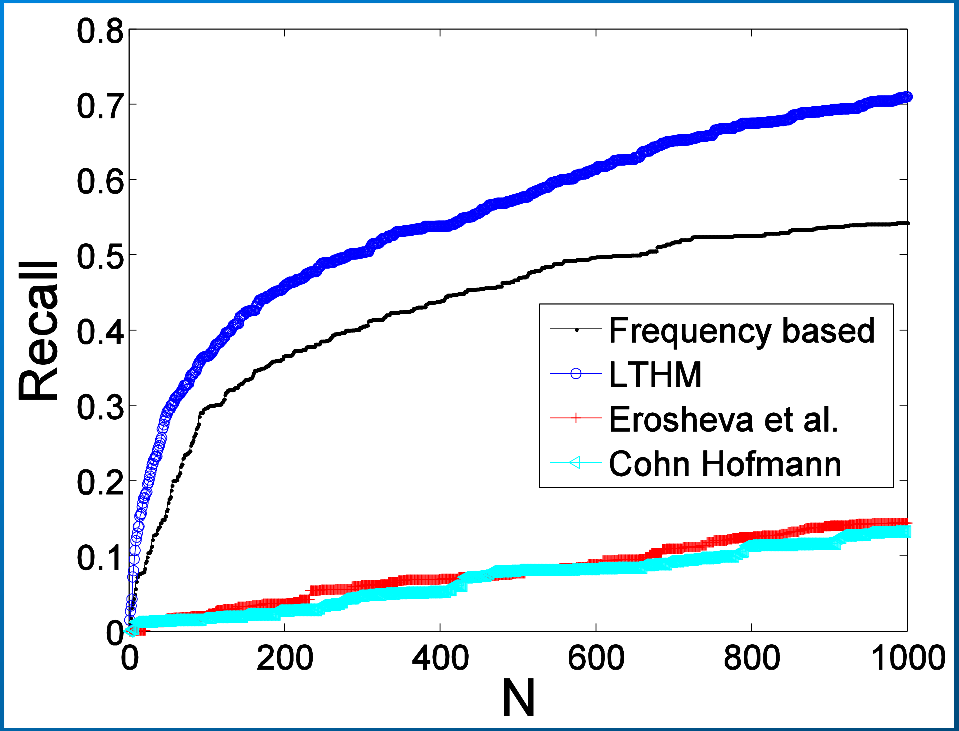
Topic prob	Top words
0.1504	search, article, navigation
0.0798	press, university, new
0.0652	learning, machine, algorithms
0.0594	fixes, skins, import
0.0533	model, regression, reasoning

LTHM

Topic prob	Top words
0.4136	learning, machine, engineering
0.0943	card, conference, credit

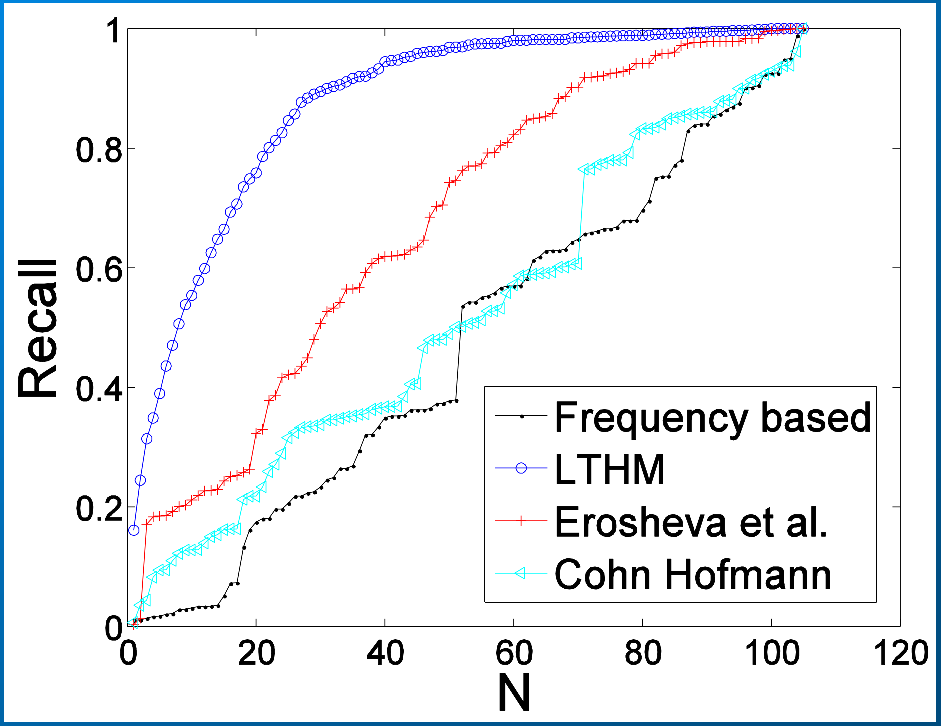
Experiments – link prediction on test set

- Wikipedia corpus: 105 documents with 790 links
- 20 hidden aspects
- Test set: 11 documents, outgoing links are invisible



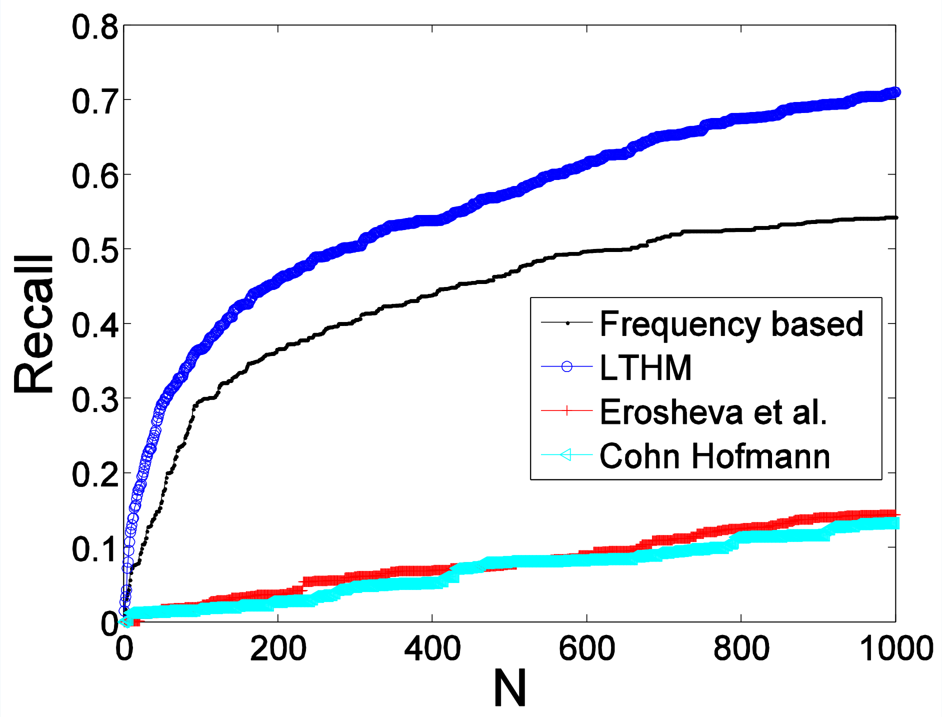
Experiments – link prediction on train set

- Wikipedia corpus: 105 documents with 790 links
- 20 hidden aspects



Experiments – link prediction

- Webkb corpus: 8282 documents with 12911 links
- 20 hidden aspects
- Test set: 10%



Summary

- Explicit modeling of link generation in an LDA like model
- Efficient approximate inference algorithm
- Performs better than previous topic models in link recommendation

- Code and data available online at http://www.cs.huji.ac.il/~amitg/lthm.html