CS Techniques for Social Media Analytics

2009.11.17.

Alice H Oh
alice.oh@cs.kaist.ac.kr

u+i lab
users and information

Today's Talk

- What is Social Media Analytics?
- CS in a Nutshell 1: Natural Language Processing
- CS in a Nutshell 2: Artificial Intelligence
- CS in a Nutshell 3: Applying NLP & AI to Social Media Analytics
- Examples

u+i lab
users and information

What is Social Media Analytics?

TISTORY twitter YouTube Broadcast Yourself me2DAY CYWORLD

delicious social bookmarking facebook flickr last.fm 아고라

Analyzing social media to find interesting phenomena caused by the interconnections among users and information

u+i lab
users and information

Internet as Read-Only

Internet as Read-Write

blog* 내가 보는 세상, 혹은 내가 보는 나

블로그 위치로그 미디어로그 태그 방명록 관리자

POST IT이야기

나도 할 수 있다! 구글웨이브(GoogleWave) – 1부 인터페이스편

1. 구글 웨이브란 무엇인가?

: 구글 웨이브는 한마디로 정의 내리기엔 그 가능성이 무한한 서비스입니다.
메신저, 메일, 카페, 블로그, SNS(소셜네트워크), 미디어매체, 웹하드, 개인적인 데이터백업소 등등
어떻게 활용할지 방법만 정하면 어떻게든 활용할 수 있습니다.
현재는 베타 프리뷰 서비스인고 너무나 부족한 모습을 보여주고 있지만 그 가능성에 큰 비중을 두고 기대를 가져봅니다.

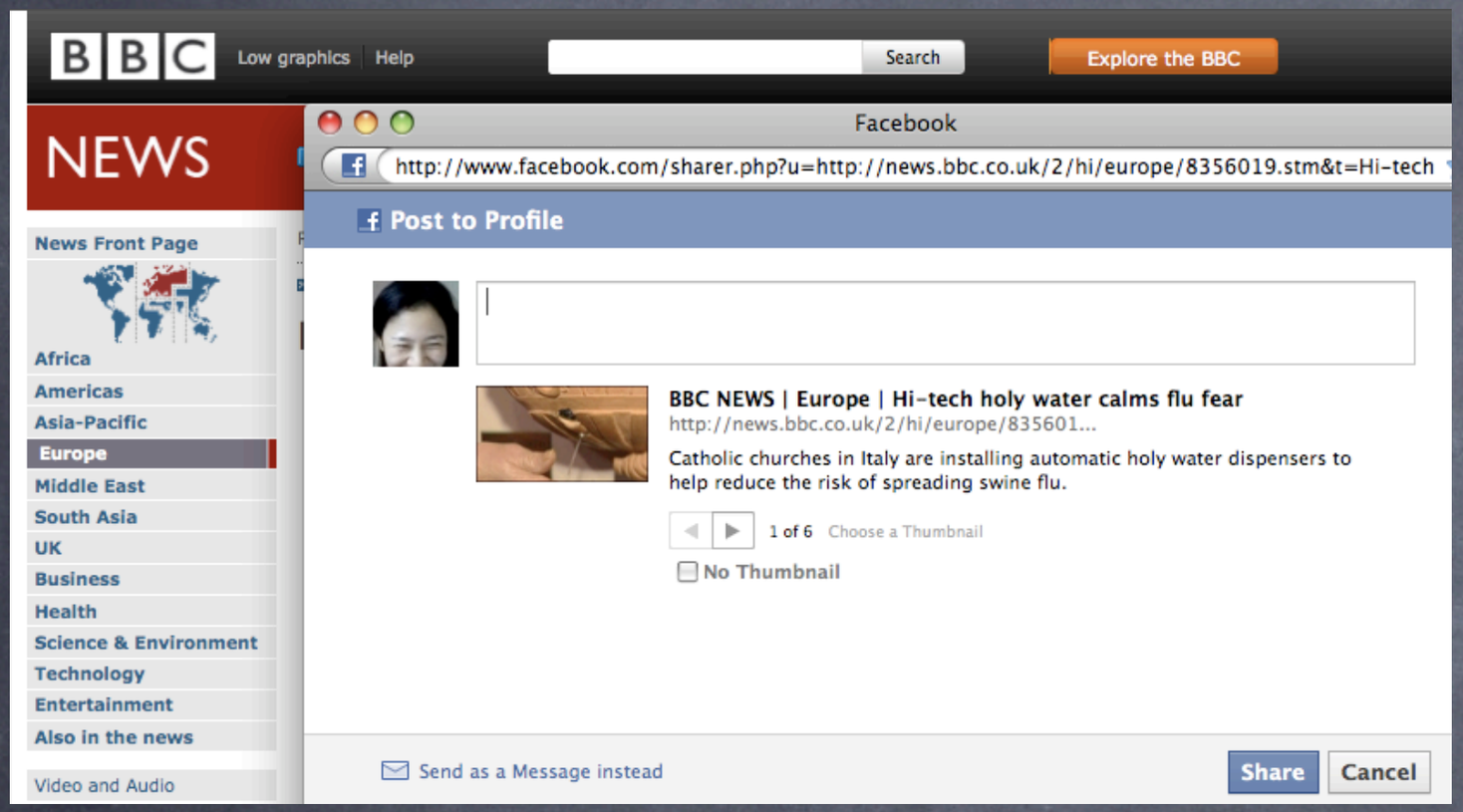
tags

googlewave, SNS, wave, 구글웨이브, 소셜네트워크

posted at

2009/11/13 15:25

Internet as Read-Write-Share



Sharing a BBC news article by posting to Facebook profile

Sharing a New York Times article by posting to Twitter

Internet as Read-Write-Share-Connect

u+i lab
users and information

facebook Home Profile Friends Inbox Alice Oh Settings Logout Search

News Feed
MIT
Carnegie Mellon
Status Updates
Photos
Links
More

Live Feed View News Feed

What's on your mind?

Alexander Koller Ok, I had my doubts, but this looks _very_ cool.
http://www.apple.com/trailers/sony_pictures/2012/extended_high.html

Apple - Trailers - 2012 - teaser - high
Source: www.apple.com
Never before has a date in history been so significant to so many cultures, so many religions, scientists, and governments. 2012 is an epic adventure about a global cataclysm that brings an end to the world and tells of the heroic struggle of the survivors.

35 minutes ago · Comment · Like · Share

Buddhika Kottahachchi was tagged in an album.

Old Photos

46 minutes ago

Francisco Rojas My lab advisor's birthday is celebrated, followed

Requests See All
- 1 event invitation
- 4 friend suggestions
- 4 group invitations
- 35 other requests

Suggestions See All

Steph Kim
Add as Friend

Hyemin Chung
Help make Facebook better for her.
Write on Her Wall

Sponsored

Connect With More Friends
Share the Facebook experience with more of your friends. Use our simple invite tools to start connecting.

Enter a friend's email address...

Chat (9)

u+i lab
users and information

The Changing Landscape of the Internet

User	Information	User Needs	Relevance
Consumer	Web Pages	Information	Topic-Based
+	+	+	+
Creator	Reviews, Weblogs	Opinions	Polarity
+	+	+	+
Member of Communities	Network of Users & Information	Ideas, Issues, Viewpoints, People	Non-Topical Dimensions

u+i lab
users and information

Interesting Problems

TISTORY twitter YouTube Broadcast Yourself me2DAY cyworld

delicious social bookmarking facebook flickr last.fm 다음 아고라

- What are people talking about in blogs? Online communities? Microblogs?
- What are the "thoughts" and "opinions" expressed in those media?
- Why are some blogs and communities more popular than others?
- Who are the "friends" in our social web?

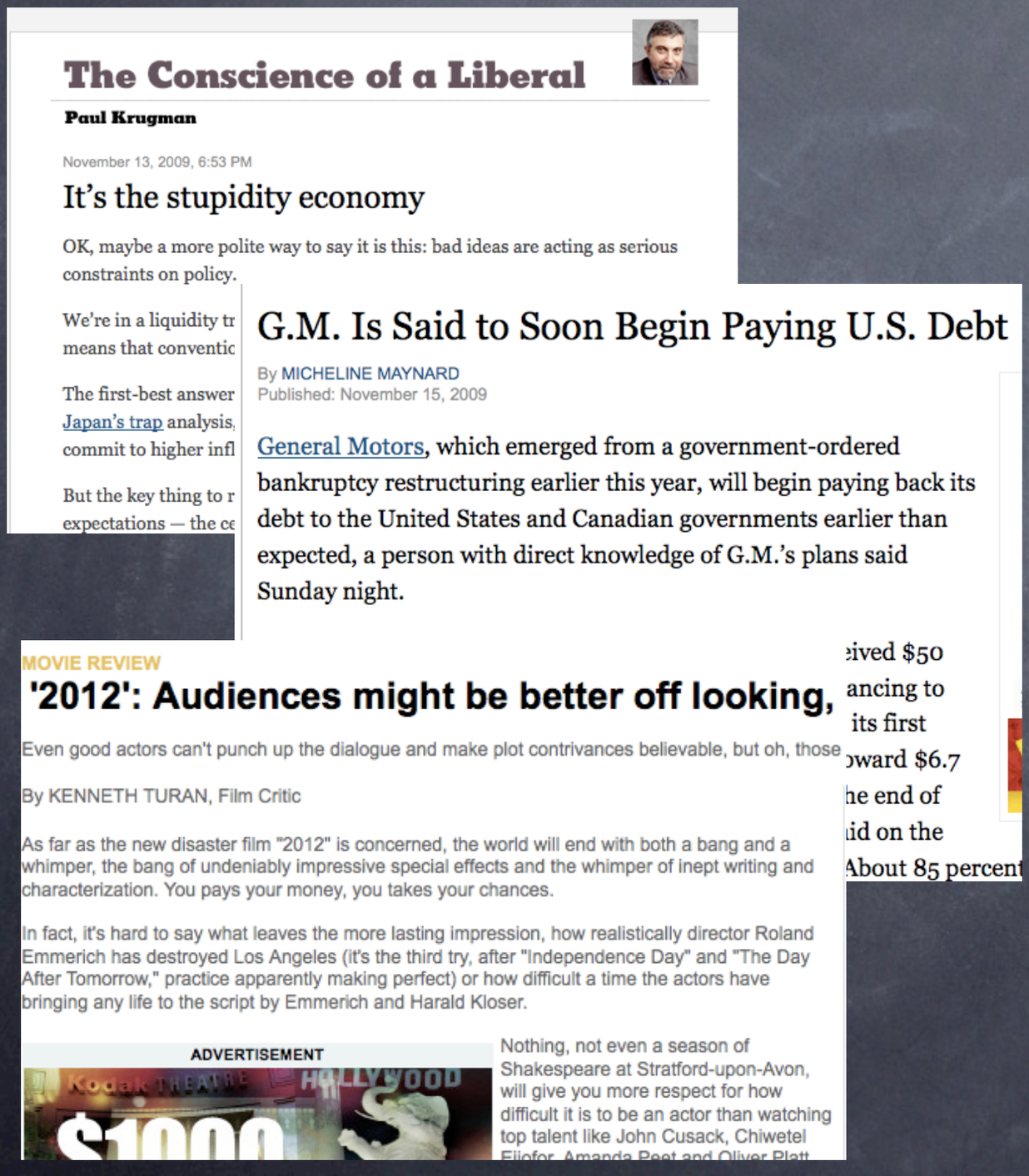
u+i lab
users and information

Interesting Problems

u+i lab
users and information

CS in a Nutshell
Part 1:
Natural Language Processing (NLP)

NLP Problems



The Conscience of a Liberal

Paul Krugman

November 13, 2009, 6:53 PM

It's the stupidity economy

OK, maybe a more polite way to say it is this: bad ideas are acting as serious constraints on policy.

We're in a liquidity tr
means that conventic

The first-best answer
Japan's trap analysis,
commit to higher infl

But the key thing to r
expectations — the ce

G.M. Is Said to Soon Begin Paying U.S. Debt

By MICHELINE MAYNARD
Published: November 15, 2009

General Motors, which emerged from a government-ordered bankruptcy restructuring earlier this year, will begin paying back its debt to the United States and Canadian governments earlier than expected, a person with direct knowledge of G.M.'s plans said Sunday night.

eived $50
ancing to
its first
ward $6.7
he end of
id on the
About 85 percent

MOVIE REVIEW

'2012': Audiences might be better off looking,

Even good actors can't punch up the dialogue and make plot contrivances believable, but oh, those

By KENNETH TURAN, Film Critic

As far as the new disaster film "2012" is concerned, the world will end with both a bang and a whimper, the bang of undeniably impressive special effects and the whimper of inept writing and characterization. You pays your money, you takes your chances.

In fact, it's hard to say what leaves the more lasting impression, how realistically director Roland Emmerich has destroyed Los Angeles (it's the third try, after "Independence Day" and "The Day After Tomorrow," practice apparently making perfect) or how difficult a time the actors have bringing any life to the script by Emmerich and Harald Kloser.

ADVERTISEMENT

Nothing, not even a season of Shakespeare at Stratford-upon-Avon, will give you more respect for how difficult it is to be an actor than watching top talent like John Cusack, Chiwetel

What is this article talking about?

Summarize the article in a few sentences.

Are these two articles talking about the same thing?

Is this saying that I should buy product A?

Is this saying that I should see movie B?

What does the author think about issue C?

Is this article for an expert or an everyday reader?

NLP Problems

G.M. Is Said to Soon Begin Paying U.S. Debt

By MICHELINE MAYNARD
Published: November 15, 2009

General Motors, which emerged from a government-ordered bankruptcy restructuring earlier this year, will begin paying back its debt to the United States and Canadian governments earlier than expected, a person with direct knowledge of G.M.'s plans said Sunday night.

Enlarge This Image

The company, which received $50 billion in government financing to avoid collapse, will make its first payment of $1.2 billion toward $6.7 billion of senior debt at the end of December, this person said on the condition of anonymity. About 85 percent

Source: nytimes.com

What is this article talking about?

Summarize the article in a few sentences.

Are these two articles talking about the same thing?

Is this saying that I should buy product A?

Is this saying that I should see movie B?

What does the author think about issue C?

Is this article for an expert or an everyday reader?

Automakers

GM Plans to Pay Back $6.7B Loan

By Joseph Woelfel 11/16/09 - 01:09 AM EST

Add Comment

DETROIT (TheStreet) -- **General Motors** plans to begin paying back a $6.7 billion loan it owes the U.S. government starting late this year and could repay the entire loan by the middle of 2011, according to published reports.

The government debt represents about 13% of the $52 billion that U.S. taxpayers have invested in General Motors, the majority of which was exchanged for a 61% ownership stake in the company, the *Associated Press* reports.

The U.S. automaker is expected to announce the repayment play Monday when it releases its preliminary third-quarter earnings, *AP* and the *Wall Street Journal* report.

Under the plan to pay back the $6.7 billion, GM will make quarterly payments of $1 billion to the U.S. government and $200 million to the Canadian

More on Automakers

TRW Automotive Gets Moody's Debt Lift

Source: thestreet.com

NLP Problems

Kindle Review After 1 Year of Use–The Good and The Bad…01.26.09

This is an excerpt from this past weekend's post A year with an Amazon Kindle (and new Kindle Cases) by Scott Hanselman on his SCOTT HANSELMAN'S COMPUTERZEN.COM blog:

26 01 2009

"…**The Good**:

- **Coverage.** Anyway I've gone in the states, I've had good coverage and no trouble getting new books. There isn't complete coverage, but if you make sure to download whatever books you want for your trip before you head into the boonies, you'll cool. I loaded up before a trip to South Africa, turned off the radio, and used it happily disconnected for weeks.
- **Battery Life.** The battery really lasts for thousands of page-turns. Remember that it doesn't use really any power at all if the pages aren't turning. It'll stay on standby (with the radio switch off) for days and days.
- **Flexibility.** I read lots of books, some purchased from Amazon (the rule of thumb is that they are 25%-50% less because there's no molecules) and some free books formatted for the Kindle. It also like that your kindle gets an email address so you can email <yourusername>@kindle.com and whatever you send it (PDF, DOC, etc) will just show up on your Kindle.

Source: lonewolflibrarian.wordpress.com

MOVIE REVIEW

'2012': Audiences might be better off looking,

Even good actors can't punch up the dialogue and make plot contrivances believable, but oh, those

By KENNETH TURAN, Film Critic

As far as the new disaster film "2012" is concerned, the world will end with both a bang and a whimper, the bang of undeniably impressive special effects and the whimper of inept writing and characterization. You pays your money, you takes your chances.

In fact, it's hard to say what leaves the more lasting impression, how realistically director Roland Emmerich has destroyed Los Angeles (it's the third try, after "Independence Day" and "The Day After Tomorrow," practice apparently making perfect) or how difficult a time the actors have bringing any life to the script by Emmerich and Harald Kloser.

ADVERTISEMENT

Nothing, not even a season of Shakespeare at Stratford-upon-Avon, will give you more respect for how difficult it is to be an actor than watching top talent like John Cusack, Chiwetel

Source: nytimes.com

What is this article talking about?

Summarize the article in a few sentences.

Are these two articles talking about the same thing?

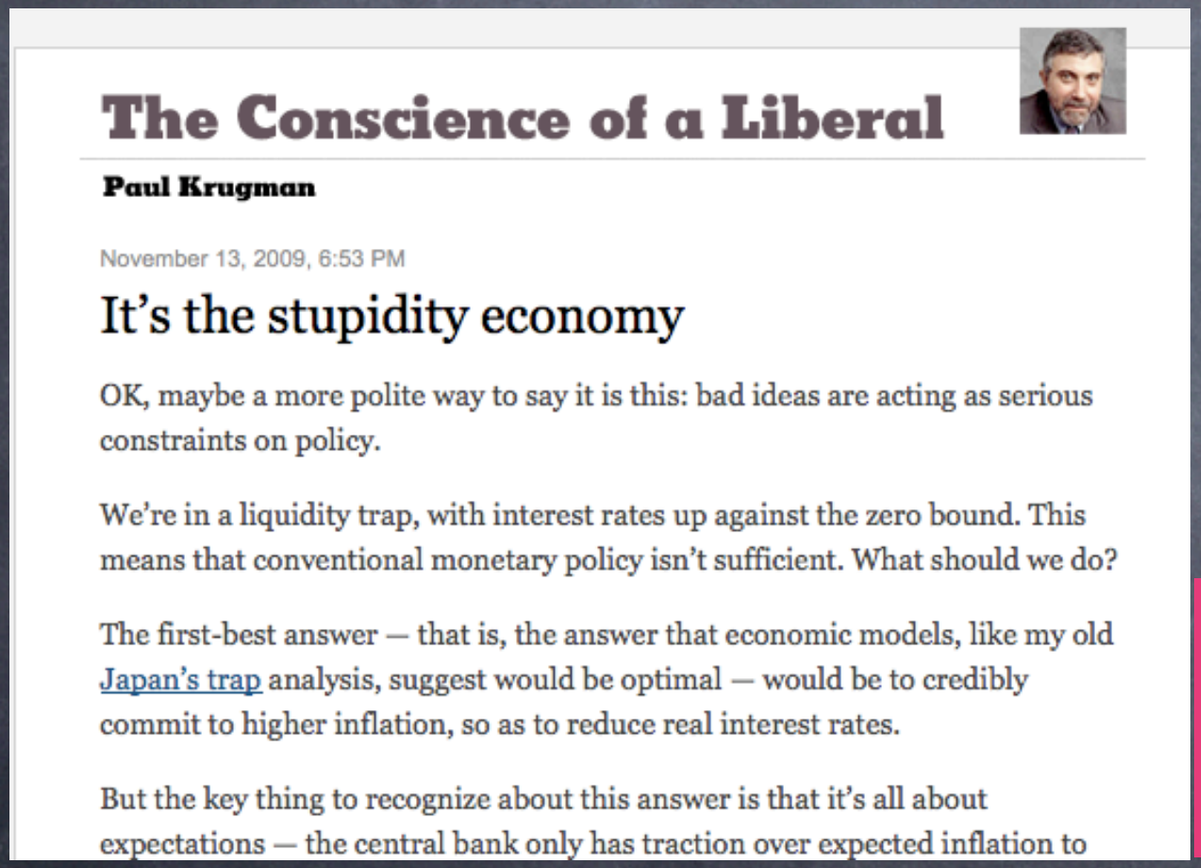
Is this saying that I should buy product A?

Is this saying that I should see movie B?

What does the author think about issue C?

Is this article for an expert or an everyday reader?

NLP Problems



The Conscience of a Liberal

Paul Krugman

November 13, 2009, 6:53 PM

It's the stupidity economy

OK, maybe a more polite way to say it is this: bad ideas are acting as serious constraints on policy.

We're in a liquidity trap, with interest rates up against the zero bound. This means that conventional monetary policy isn't sufficient. What should we do?

The first-best answer — that is, the answer that economic models, like my old Japan's trap analysis, suggest would be optimal — would be to credibly commit to higher inflation, so as to reduce real interest rates.

But the key thing to recognize about this answer is that it's all about expectations — the central bank only has traction over expected inflation to

Source: nytimes.com

What is this article talking about?

Summarize the article in a few sentences.

Are these two articles talking about the same thing?

Is this saying that I should buy product A?

Is this saying that I should see movie B?

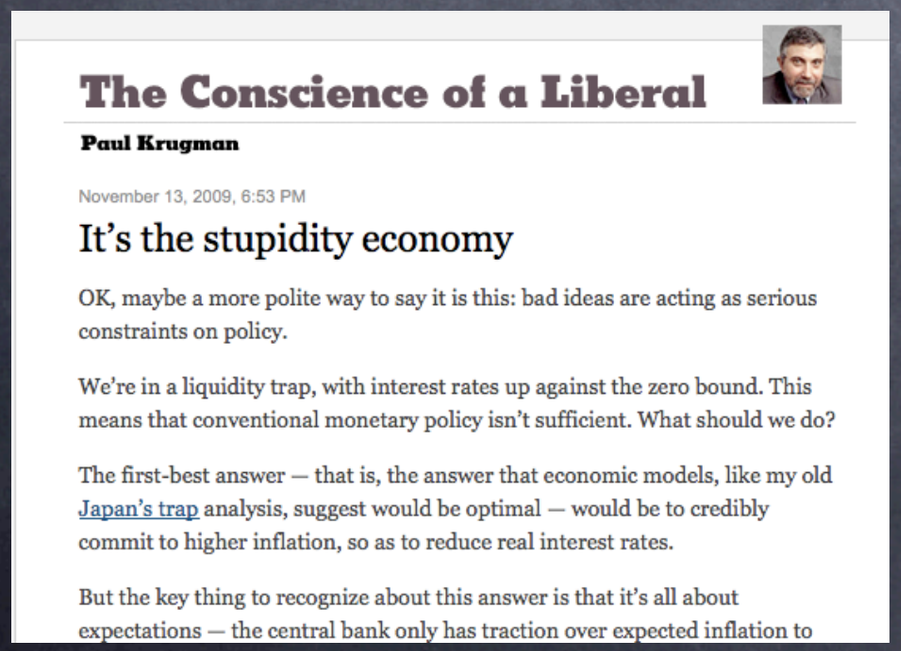
What does the author think about issue C?

Is this article for an expert or an everyday reader?

Document Representation

- We need to turn a document into an instance of computational representation so that we can compute with it

- Simplest Representation: Document as a set of words (with frequencies and probabilities)



The Conscience of a Liberal

Paul Krugman

November 13, 2009, 6:53 PM

It's the stupidity economy

OK, maybe a more polite way to say it is this: bad ideas are acting as serious constraints on policy.

We're in a liquidity trap, with interest rates up against the zero bound. This means that conventional monetary policy isn't sufficient. What should we do?

The first-best answer — that is, the answer that economic models, like my old Japan's trap analysis, suggest would be optimal — would be to credibly commit to higher inflation, so as to reduce real interest rates.

But the key thing to recognize about this answer is that it's all about expectations — the central bank only has traction over expected inflation to

word	freq	prob
ok	1	0.002
maybe	3	0.006
economic	5	0.010
inflation	4	0.008

Pre-Processing Steps

- Stop-Word Removal -- Words that occur frequently in any document is not helpful (e.g., "the", "this", "you", etc)
- Stemming -- Unifying all tenses and morphological variations of a word (e.g., "going, goes, went" --> "go")
- Named Entity Recognition -- Identifying names of people, organizations, places, etc.

u+i lab
users and information

Bag-Of-Words

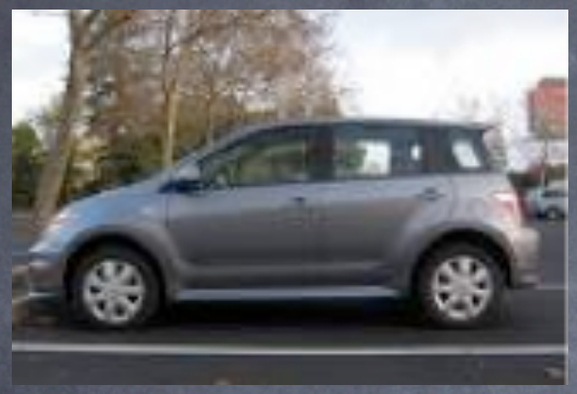
word	freq	prob
ok	1	0.002
maybe	3	0.006
economic	5	0.010
inflation	4	0.008

u+i lab
users and information

CS in a Nutshell
Part 2:
Artificial Intelligence

u+i lab
users and information

A Problem in AI



Machine Learning

car

car

car

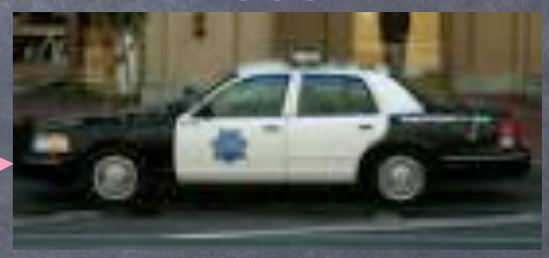
car

training

testing

???

???



???

Classification

- Many machine learning problems are classification problems
  - Does this belong to the "car" class?
- Each object we wish to recognize forms a class
  - car class, people class, tree class
  - spam mail class, non-spam mail class

u+i lab
users and information

Features

- We use "features" to simplify the classification problem
- Features turn objects (images, documents, etc) into a set of numbers so that we can do computation over them
- Examples: color histogram, bag-of-words, number of wheels, face/no face

u+i lab
users and information

Naive Bayes Classifier

- A probabilistic model
  1. compute the probability of an object belonging to a class, given the features
  2. find the class with the highest probability

$$y = \text{argmax}_y \; P(y|f_1 \dots f_n)$$

Naive Bayes Classifier

- Naive Bayes Classifier assumes that all features are independent given the class

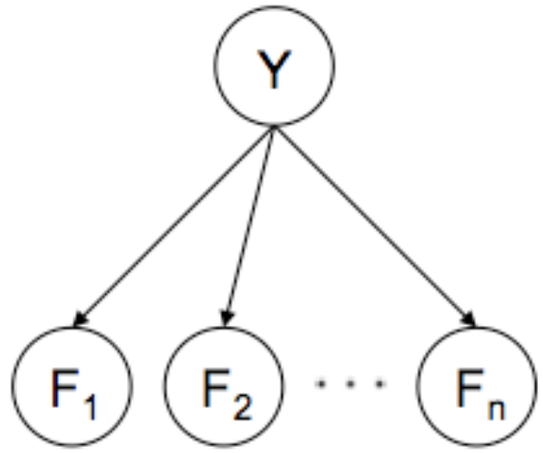
$|Y| \times |F|^n$ parameters

$$P(Y, F_1 \dots F_n) =$$

$$P(Y) \prod_i P(F_i|Y)$$

$|Y|$ parameters

$n \times |F| \times |Y|$ parameters



- Compute the probabilities for all classes, and choose the class with the highest probability

$$y = \text{argmax}_y \; P(y|f_1 \dots f_n)$$

u+i lab
users and information

Linear Classifier

negative instances






positive instances







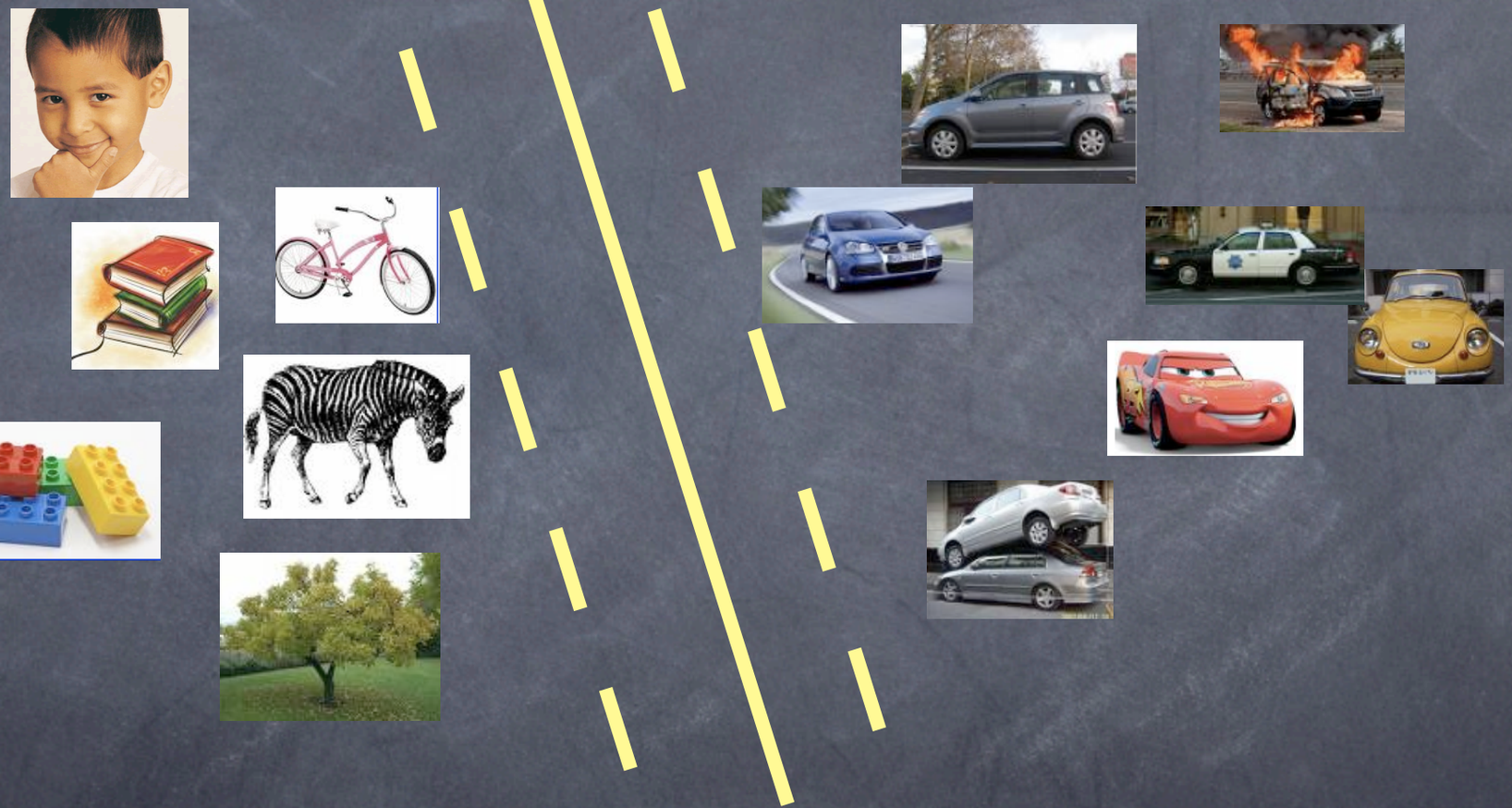
A linear classifier is a hyperplane that separates the positive and negative instances

u+i lab
users and information

Support Vector Machine

negative instances

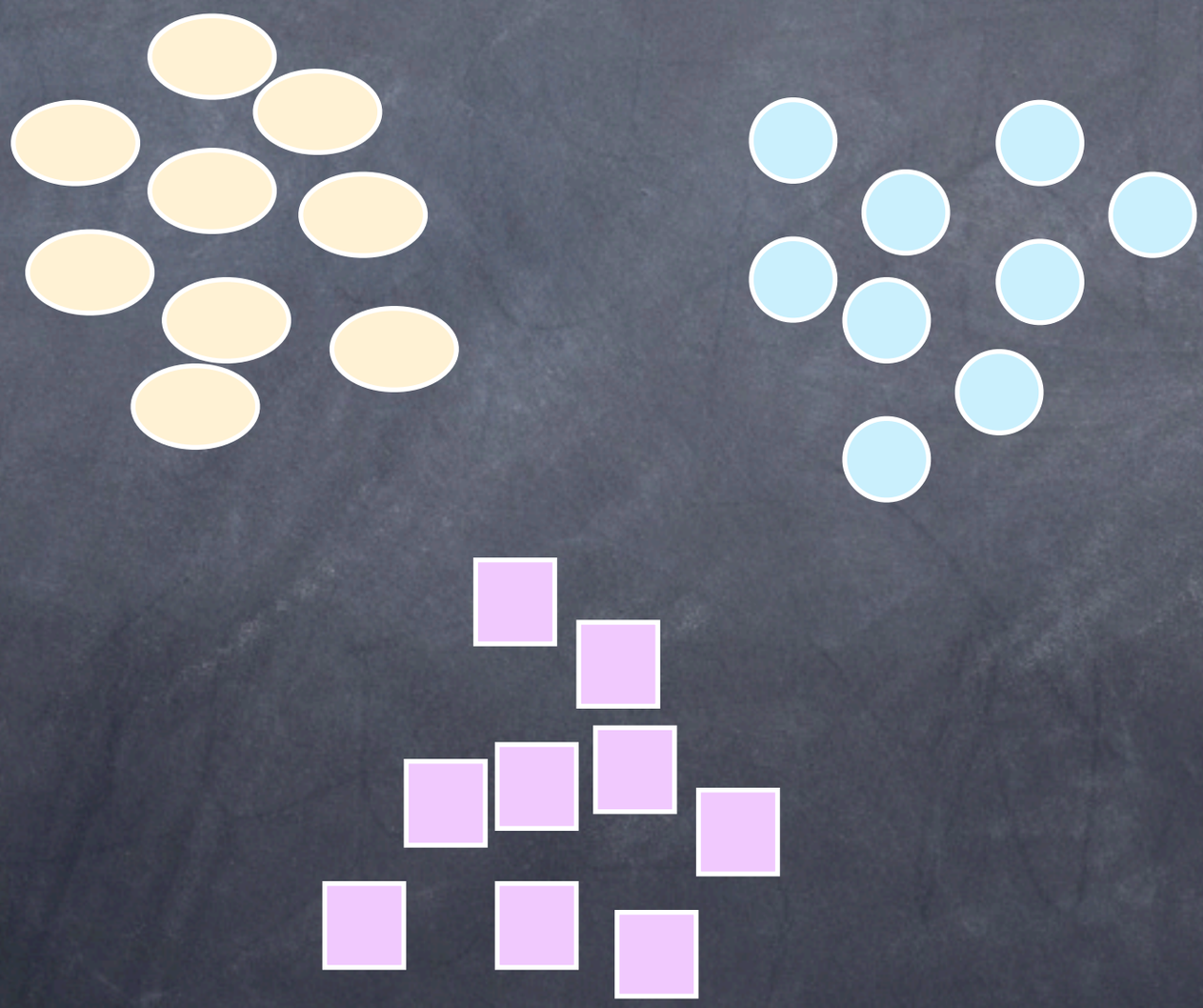
positive instances



Support Vectors are training instances that lie on the hyperplanes that maximize the margins between the classes

u+i lab
users and information

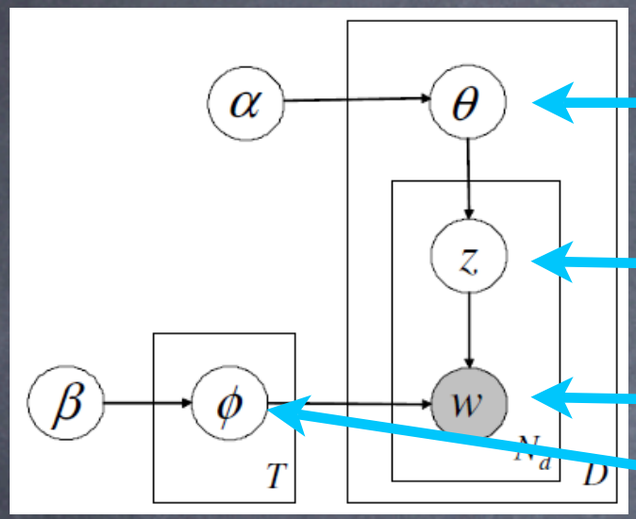
Clustering



Supervised vs Unsupervised

- Supervised Learning
  - Labelled data for known classes
    - Example: object recognition
- Unsupervised Learning
  - No labelled data
  - Unknown classes
  - Example: topic discovery

Latent Dirichlet Allocation (LDA)



topic distribution

topic

word

word distribution

The William Randolph Hearst Foundation **will give** $1.25 million **to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our** board **felt that we had a real opportunity to make a mark on the future of the performing arts with these** grants **an act every bit as important as our traditional areas of** support **in health, medical** research, **education and the social** services," **Hearst** Foundation President **Randolph A. Hearst said Monday in** announcing **the** grants. **Lincoln Center's share will be** $200,000 **for its new** building, **which will** house **young artists and** provide **new** public facilities. **The Metropolitan Opera Co. and New York Philharmonic will** receive $400,000 **each. The Juilliard School, where music and the performing arts are taught, will get** $250,000. **The Hearst** Foundation, **a leading supporter of the Lincoln Center Consolidated Corporate** Fund, **will make its usual** annual $100,000 **donation, too.**

The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

u+i lab
users and information

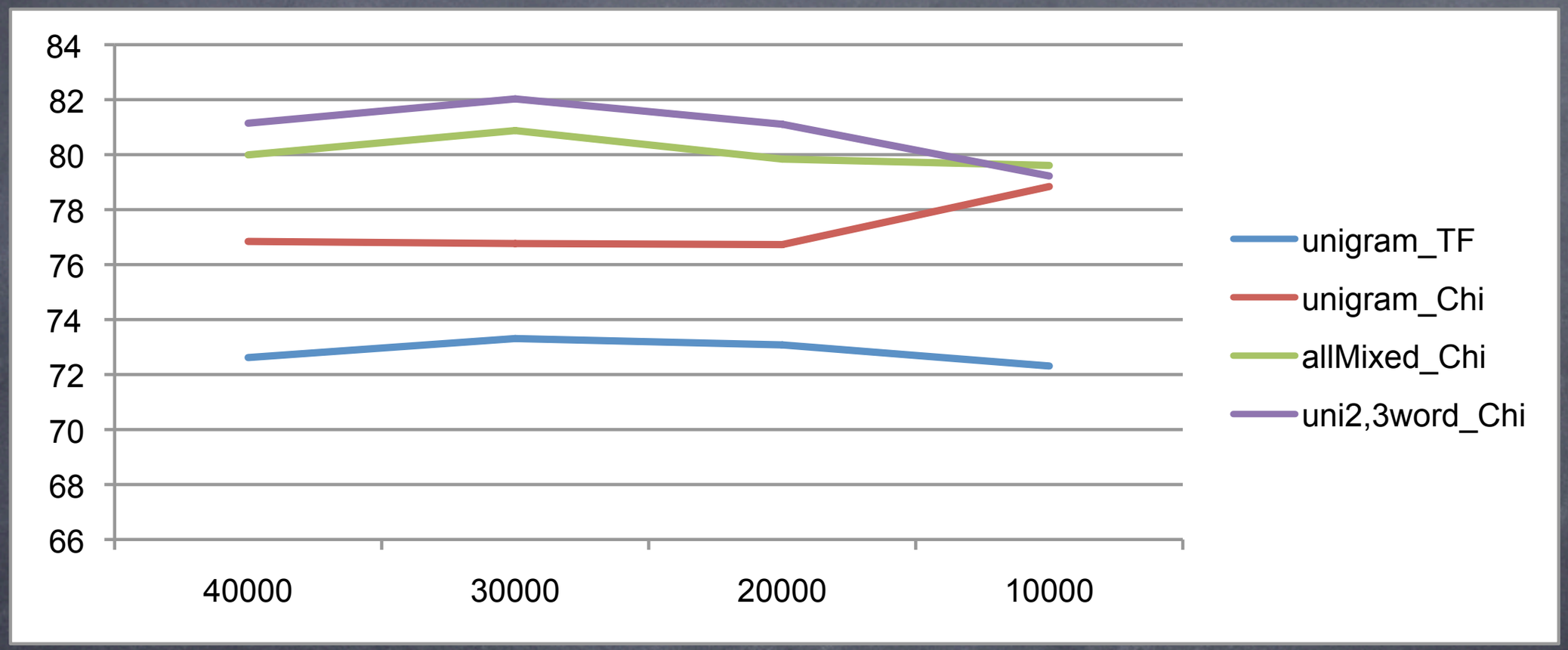
CS in a Nutshell
Part 3:
Applying NLP & AI
Techniques to Social Media Analytics

u+i lab
users and information

Ex 1: Political Blogs

- Search political blogs with keywords such as "Obama", "bailout", "Iraq"
- Classify the posts as Liberal or Conservative
- Present the results in 3 different views for user study

Classification



F-scores for different feature sets used in SVM for Political Blog Classification

User Study

View 1: Normal View (no tag)

View 2: Tagged View

Palestine: ?How many deaths in Gaza is enough?? **LIBERAL**
As the Israeli attacks on Gaza continue, in this roundup of Gaza's blogs we hear about food shortages, the frustration of being stuck at home, the humour of medical workers - and a question from a young boy: ?Mama ? why don't the ...
http://globalvoicesonline.org/2009/01/10/palestine-how-many-deaths-in-gaza-is-enough/

Amira Haas in Gaza **CONSERVATIVE**
The Israeli bombardment of Gaza is in its third day, and now the Islamic University in Gaza ? which The New York Times calls a ?Hamas stronghold? and Israel says helps to manufacture some of the Qassam rockets that hit Israeli villages ...
http://washingtonindependent.com/23081/amira-haas-in-gaza

View 3: Two-Column View

Political Blog Search [economy] Search

Used keywords : obama, **economy**, gaza
Unused keywords : president, senate, congress, election, tax, bailout, iraq, afghanistan, bush, hamas, israel, middle east

Ukraine: Politics Versus the Economy **LIBERAL**
Throughout the past week, Edward Hugh of Ukraine Economy Watch has been reporting on the repeated downgrades of Ukraine's rankings by Moody's (Oct. 20), Fitch (Oct. 21), and Standard & Poor's (Oct. 24). But on Tuesday, Oct. ...
http://globalvoicesonline.org/2008/10/25/ukraine-politics-versus-the-economy/

Can the US economy afford a Keynesian stimulus? **LIBERAL**
Economic policy is based on a collection of half-truths. The nature of these half-truths changes occasionally. Economics as a scholarly discipline consists in the periodic rediscovery and refinement of old half-truths. ...
http://blogs.ft.com/maverecon/2009/01/can-the-us-economy-afford-a-keynesian-stimulus/

Biden on economy: 'We're at war' **CONSERVATIVE**
Biden likened the country?s economic crisis to the attacks of 9/11in a privat Capitol Hill.
http://www.politico.com/news/stories/0109/17097.html

'Jolting' the Economy **CONSERVATIVE**
Barack Obama says that we have to "jolt" the economy. That certainly makes take the media's account of the economy seriously-- but should the media be Amid all the political and media hysteria, ...
http://www.realclearpolitics.com/articles/2008/11/jolting_the_economy.html

Results

	plain	tagged	2-col
ease-of-use	**2.73**	2.27	**1.93**
satisfied-results	**2.66**	**2.06**	2.2
satisfied-posts	2.33	2.26	2.2

Users' Answers to 3 Questions about the 3 views

View	#	Reasons
plain	5	I don't like the tags. They're too polarizing It provided a less biased option I like to decide for myself what is lib/cons
tagged	1	liked the tags
two-col	9	Easier to sort out ones you want to read Nice to see comparison I like to read a couple from each side I like to know something about the author

Users' Preferences Among 3 Views and Reasons for the Preference

Ex 2: Topic Detection on a Very Large Dataset

- Discovering newly emerging topics
- Previous Work
  - Small (3,000~50,000 posts) dataset
  - Topics given
- Our approach: Use a real-world scale dataset (millions of posts), with no given topics (pure discovery)

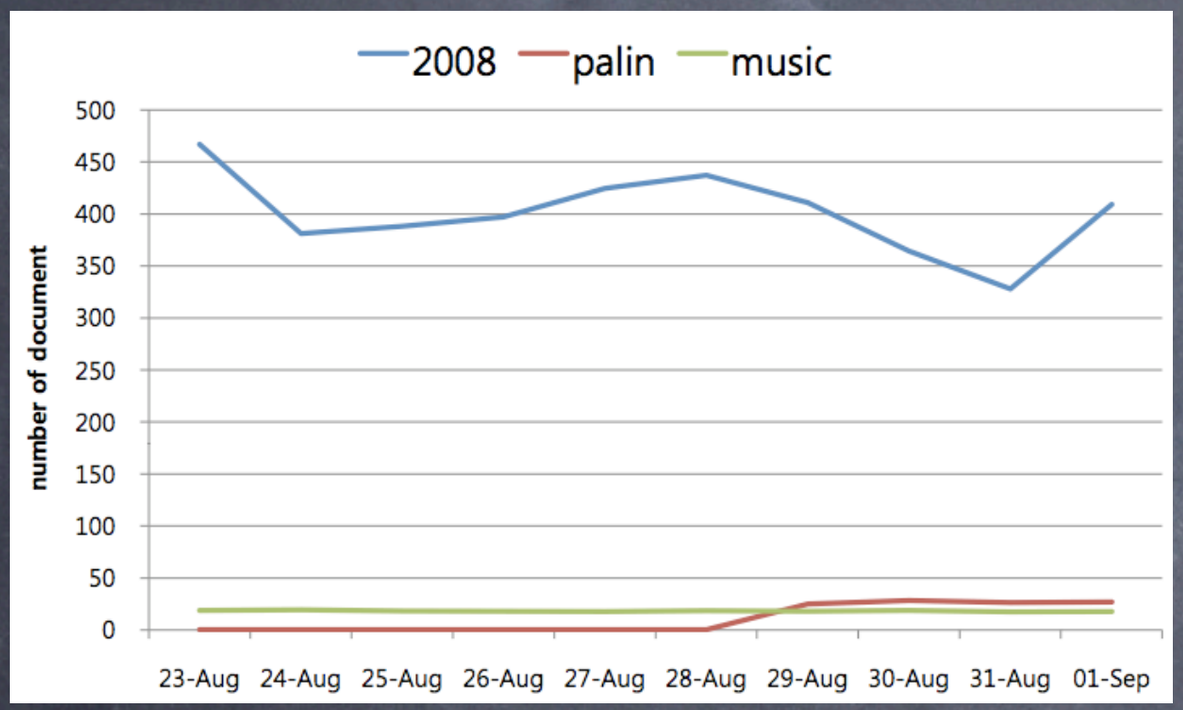
u+i lab
users and information

Large Dataset

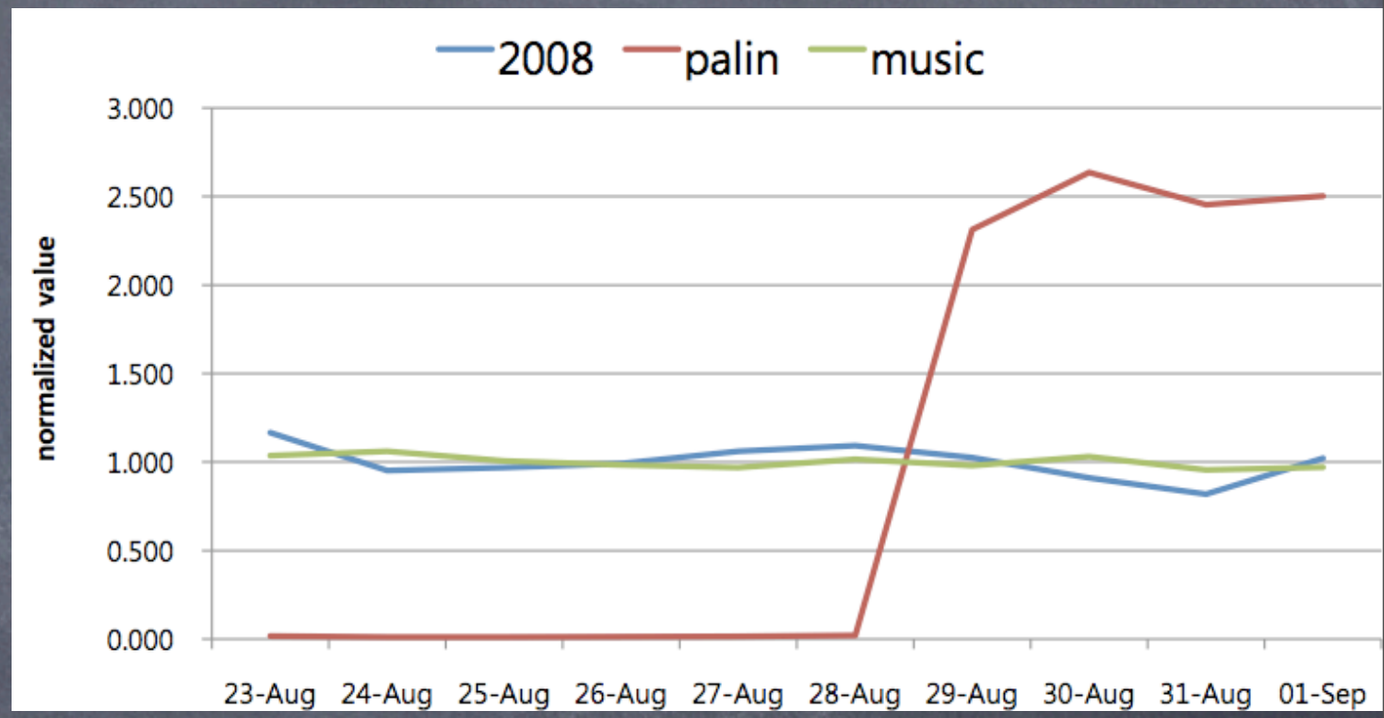
Time span	Aug. 1. 2008 – Sep. 31. 2008 (62 days)
Number of sites	1,071,156
Number of posts	14,970,428
Average Number of words/post	69.2
Number of unique words/post	51.2
Average indegree	175.62
Size	60 GB

Simple Metric for Detecting New Topics

u+i lab
users and information



word frequencies



normalized word frequencies

Finding New Topics

	2008	palin	music
Average Frequency	390	16	18
STDEV	0.156154	1.088005	0.057355
IIM Ranking	2	1	3
		Topic word	
Freq/Fluctuation	High frequency Low fluctuation	High fluctuation Low-medium frequency	Low-medium frequency Low fluctuation

u+i lab
users and information

Results

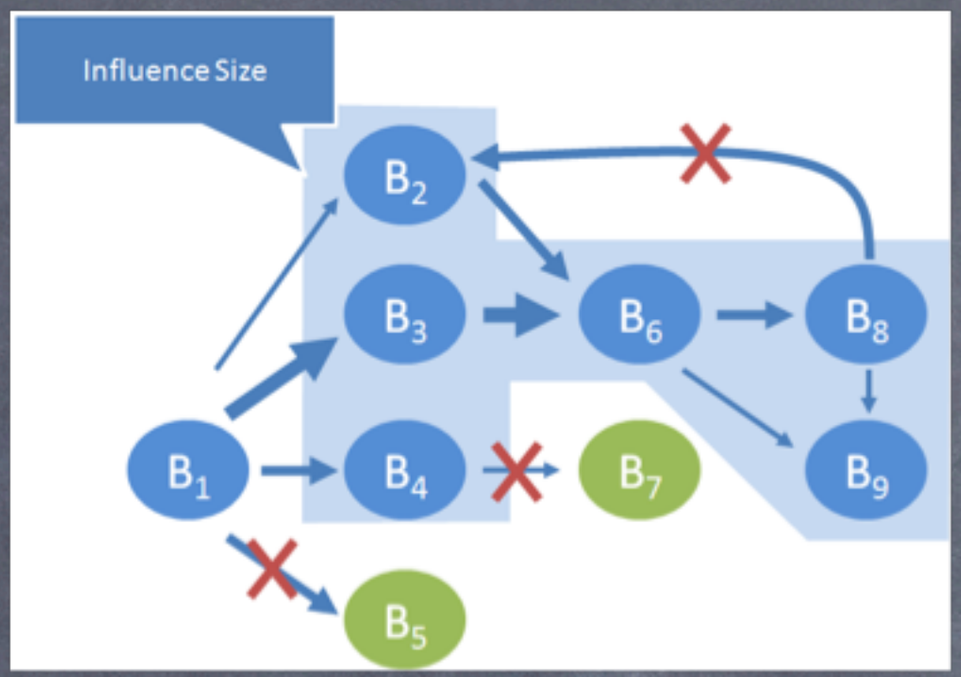
New Group1 Words	First appearance date on NYT	First issue activation date on blog corpus	Issue description
gustav	26- Aug	28-Aug	Hurricane name landed
lehman	-	10-Sep	A bankrupted company
eagles	-	09-Aug	-
$700	19-Sep	22-Sep	$700 Million dollar bailout plan
brothers	-	16-Aug	A bankrupted company
bailout	05-Aug	08-Sep	A heavily discussed government policy
ike	06-Sep	07-Sep	Hurricane name landed
gulf	27-Aug	03-Aug	Hurricane Gustav as it heads into the Gulf of Mexico
medal	-	12-Aug	Olympic medal
denver	25-Aug	13-Aug	2008 Democratic National Convention, in denver. Aug. 25 – 28.(found in NYT)

u+i lab
users and information

Example 3: Blog Influence Prediction

- What characteristics of a blog make it an influential blog?
- Previous approaches have focused mostly on network characteristics
- We combine network characteristics with content features to improve blog influence prediction

Definition of Influence



3 Levels of Influence:

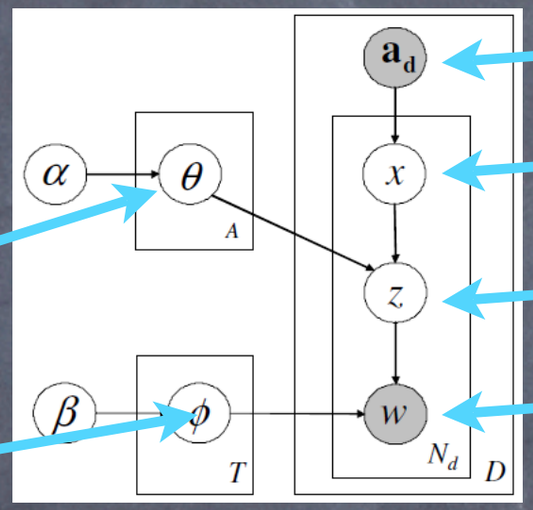
Tier 1: 1~6

Tier 2: 6~36

Tier 3: 36 ~ 223

Content Analysis

Author-Topic Model (a variant of LDA)



author distribution

author

topic distribution

topic

word

word distribution

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

"Arts"	
Suciu_D	0.01022
Naughton_J	0.00946
Levy_A	0.00706
DeWitt_D	0.00676
Wong_L	0.00672
Chakrabarti_K	0.00640
Ross_K	0.00606
Hellerstein_J	0.00586
Lenzerini_M	0.00538
Moerkotte_G	0.00530

Network Analysis

- Standard Network Metrics
  - Indegree
  - Outdegree
  - Totaldegree
  - Betweenness
  - Clustering Coefficient

Hybrid Learning Model

- Network + Topic Features all thrown into a machine learning algorithm (Support Vector Machine) for predicting blog influence level

u+i lab
users and information

Results

Learning Model	Multi-Class Classification	Tier 1	Tier 2	Tier 3
Topic	0.714	0.738	0.909	0.781
Network	0.809	0.861	0.947	0.809
Hybrid	0.882	0.909	0.939	0.916

Prediction Accuracies

Important Features

TABLE 5. Feature weights from a trained Network-Topic SVM model, Cosine Similarity Threshold = 0.5, 30% of blogs are used for training, and the other blogs are used for test. Weight on upper tier training instances is 8. Shadowed cells are features from the network analysis while the other cells are from the content analysis.

Rank by Absolute Value of Weights	SVM Classifier for Tier 1		SVM Classifier for Tier 2		SVM Classifier for Tier 3	
	Feature Weights	Feature Name	Feature Weights	Feature Name	Feature Weights	Feature Name
1	13.730	OutDegree	15.709	OutDegree	-5.301	Topic 47
2	4.066	TotalDegree	4.272	TotalDegree	-4.690	Topic 14
3	1.964	Betweenness	-3.928	Topic 9	4.114	Topic 15
4	1.802	Clustering Coefficient	-3.407	Topic 35	-3.989	Topic 43
5	1.650	Topic 8	3.152	Topic 13	3.881	Topic 41
6	-1.567	Topic 28	3.085	Topic 0	-3.812	Topic 38
7	1.555	Topic 19	-2.920	Topic 24	-3.783	Topic 42
8	1.383	InDegree	-2.816	Topic 27	-3.691	Topic 24
9	1.370	Topic 43	-2.813	Topic 16	-3.681	Topic 12
10	-1.325	Topic 23	2.610	Topic 19	-3.544	Topic 22

For Tier 1 (low influence blogs), network features are important.
For Tier 3 (high influence blogs), topic features are important.

u+i lab
users and information

Summary of Ex 1-3

- Example 1: Political Blog Classification
- Example 2: Topic Detection Using a Very Large Dataset
- Example 3: Blog Influence Prediction

u+i lab
users and information

Users & Information Lab

- Collaborators Welcome!

- Facebook: aoh@alum.mit.edu
- Twitter: aliceoh
- Email: alice.oh@cs.kaist.ac.kr
- Web: http://uilab.kaist.ac.kr

u+i lab
users and information