EE 6886: Topics in Signal Processing
-- Multimedia Security System

Lecture 8: Behavior and Speaker Identification

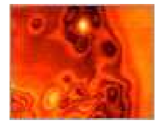
Ching-Yung Lin
Department of Electrical Engineering
Columbia University, New York, NY 10027

Course Outline

- Multimedia Security :
  - Multimedia Standards – Ubiquitous MM
  - Encryption and Key Management – Confidential MM
  - Watermarking – Uninfringible MM
  - Authentication – Trustworthy MM
- Security Applications of Multimedia:
  - Audio-Visual Person Identification – Access Control, Identifying Suspects
  - Surveillance Applications – Abnormality Detection
  - Media Sensor Networks – Event Understanding, Information Aggregation

Human – a complex multimodality subject/object

- "*Human and Social Dynamics (HSD)*" is identified as one of the five NSF key priorities among:
  - Nanoscale Science and Engineering
  - Biocomplexity in the Environment
  - Human and Social Dynamics
  - Mathematical Sciences
  - Cyberinfrastructure
  - (http://www.nsf.gov/news/priority_areas/)



E 6886 Topics in Signal Processing: Multimedia Security Systems

Person, Community, Society

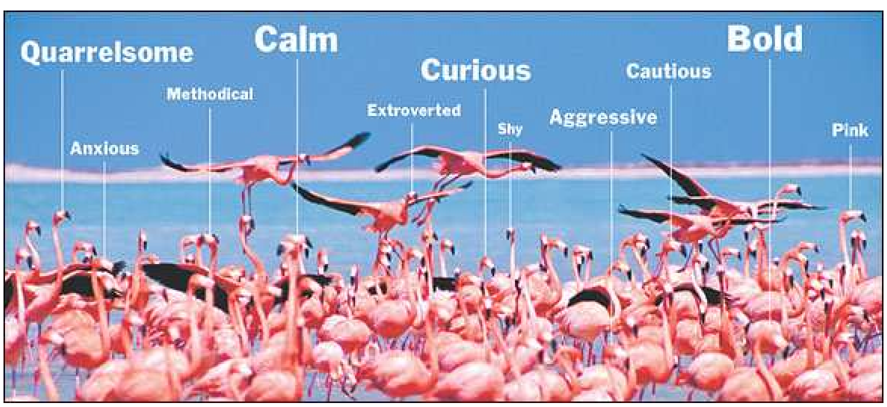


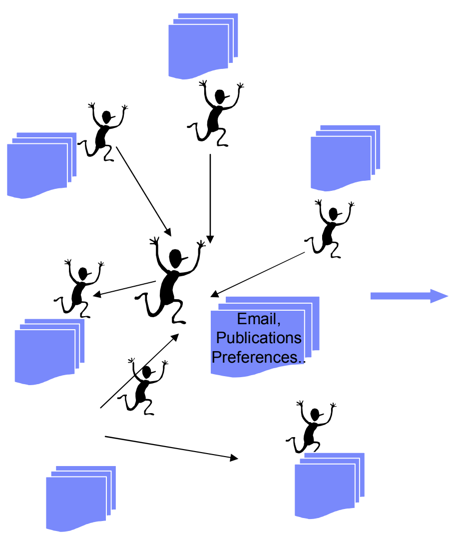
Photo Source: New York Times, 3/2/2005

Social Computing – when computer science meets sociology

- Computing-based Human Modeling (person)
  - Biometric-based modeling
  - Behavior modeling
  - Knowledge and Interest modeling
- Computing-based Society Modeling (person <-> person)
  - Community modeling
  - Social network modeling
- Computing-based Trust Management (person -> society, person -> person, person -> information)
  - Information trustworthiness
  - Human trustworthiness
  - System trustworthiness
- Computing-based Information Organization and Management (information -. person, information -> society)
  - Personalized information
  - Community-oriented information

Motivation – Latent Person Behavior Modeling, Analysis and Applications

The world consists of subjects, artifacts and relations over them



What are a person's
- roles in events?
- behavior evolutions?
-interests, tastes?
- ….

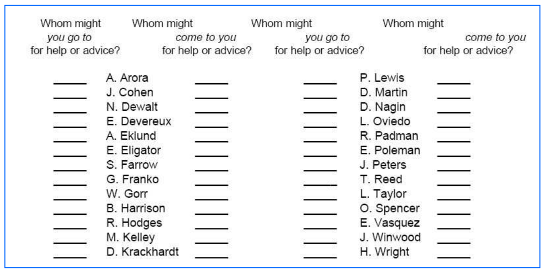
Building person models based on observations

Analyzing latent semantics based on person models and context Information

Applying on real-life problems such as:
Data mining; Information filtering;
Recommendation; Behavior Prediction;
Unusual activities detection;
Information flow modeling;
Personalized Services

Why is this topic interesting?

- Informal social network within formal organizations is a major factor affecting companies’ performances.
  - Krackhardt (2005) showed that companies with strong informal networks perform five or six times better than those with weak networks.
  - Since Weber (1920s) first studied modern bureaucracy structures, decades of related social scientific researches have been mainly relying on questionnaires and interviews to understand individual’s thoughts and behaviors.

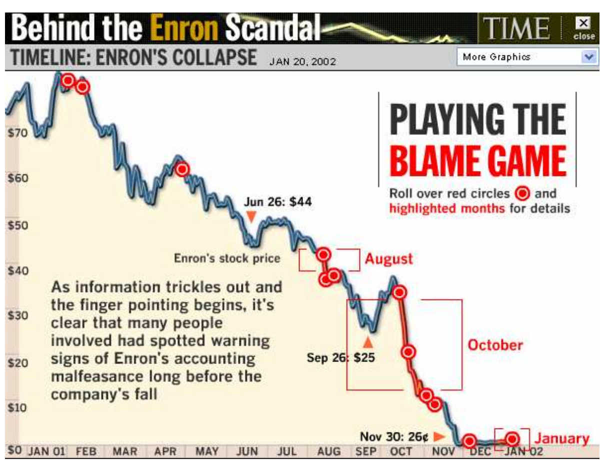


Sample questionnaire
(Prof. Krackhardt, CMU)

➔ Is it possible to ‘acquire’ social networks automatically?

➔ How about automatically building/updating ‘personal profiles’, ‘social capitals’?

Enron Corpus



Enron Corpus

- Preprocessing
  - Original messages – 517,431
  - Remove empty messages – 493,391 remain
    - 1999 – 11196
    - 2000 – 196157
    - 2001 – 272875
    - 2002 – 35922
  - Remove repeated messages – 166,653 remain
  - Only keep intra-communications among 149 users within Enron – 25,428 remain
  - Number of terms: 84649
  - Number of users: 149

Name	Email	Position
Robert Badeer	robert.badeer@enron.com	Director
Eric Bass	eric.bass@enron.com	Trader
Sally Beck	sally.beck@enron.com	Employee
Rick Buy	rick.buy@enron.com	Manager
David Delainey	david.delainey@enron.com	CEO
James Derrick	james.derrick@enron.com	In House Lawyer
Mark Haedicke	mark.haedicke@enron.com	Managing Director
Steven Kean	steven.kean@enron.com	Vice President
Louise Kitchen	louise.kitchen@enron.com	President
Phillip Allen	phillip.allen@enron.com	N/A

Collected information about the emails

ID	Subject	Time	From	To
31265382.1075858640461	FW: California gas intrastate matte	2001-07-10T19:32:29.	k..allen@enron.com	matt.smith@enron.com,
14873812.1075858640483	FW: West Power Strategy Briefing	2001-07-11T12:56:41.	k..allen@enron.com	keith.holst@enron.com, mike
3650242.1075858640506.		2001-07-11T15:25:40.	k..allen@enron.com	barry.tycholiz@enron.com,
483924.1075858640549.Ja	FW: Party	2001-07-12T12:04:29.	k..allen@enron.com	s..shively@enron.com,
13141541.1075858640571		2001-07-12T19:55:20.	k..allen@enron.com	michael.l.brunner@rssmb.com
14620083.1075858640594	CA Instrate Gas matters	2001-07-13T13:44:25.	k..allen@enron.com	leslie.lawner@enron.com,
634023.1075858640616.Ja	FW: CA Instrate Gas matters	2001-07-13T13:45:39.	k..allen@enron.com	mike.grigsby@enron.com,
19282752.1075858640638	Analyst/Associate Program: 2 Minu	2001-07-13T19:47:41.	k..allen@enron.com	ramabile@execlead.com,
16515849.1075858640659	FW: American Express Letter	2001-07-16T13:20:48.	k..allen@enron.com	johnny.ross@enron.com,
19618808.1075858640681	Party	2001-07-16T13:22:52.	k..allen@enron.com	richard.toubia@truequote.com
27461841.1075858640705	FW: Party	2001-07-16T13:44:37.	k..allen@enron.com	s..shively@enron.com,

What happened?– collect the ground truth

- Summarize important events from different timelines
- The events with most occurrences from multiple media's timelines
  - 14 August 2001 -- Jeffrey Skilling resigns after just six months; Mr Lay returns to day-to-day management of the company.
  - 20 August 2001 -- Mr Lay exercises Enron share options worth $519,000.
  - 12 October 2001 -- Accounting firm Andersen begins destroying documents relating to the Enron audits.The destruction continues until November when the company receives a subpoena from the Securities and Exchange Commission.
  - 16 October 2001-- Enron reports losses of $638m run up between July and September and announces a $1.2 billion reduction in shareholder equity. The reduction in company value relates to partnerships set up and run by chief financial officer Andrew Fastow.
  - …

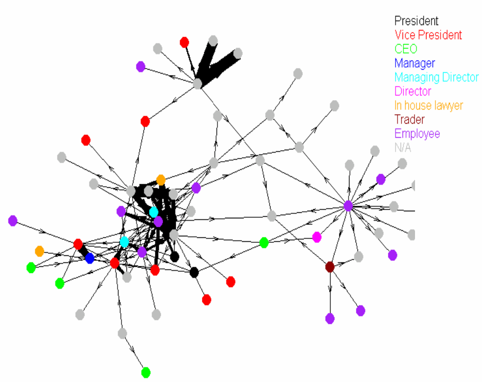
Social Network Analysis:

- Potential of social network analysis
  - every person in the world is only six edges away from every other, if an edge between *i* and *j* means "*i* knows *j*" [Milgram 1967]
- Static social network analysis
  - In social network analysis: Exponential Random Graph Models [Wasserman and Pattison, 1996]
  - In information mining area:
    - Mine social relationships from email logs by using a set of heuristic graph algorithms [Schwartz and Wood 1993]
    - Mine a social network from a wide variety of publicly-available online information to help individuals find experts who could answer their questions [Kautz *et al.* 1997]
    - *M*ine communities from the Web (defined as sets of sites that have more links to each other than to non-members) [Flake *et al. 2002*]
    - Use a betweeness centrality algorithm for the automatic identification of communities of practice from email logs within an organization [Joshua T. *et al. 2003]*
    - The Google search engine and HITS algorithm [Page *et al.* et al. 1998] [Kleinberg 1998]
- Dynamical social networks
  - In social network analysis: Dynamic actor-oriented social network [Snijder 2002]
    - Changes in the network are modeled as the stochastic result of network effects (density, reciprocity, etc.)
    - Network evolution is modeled by continuous time Markov chain models
  - In information mining area:
    - link prediction problem - Infer which new interactions among its members are likely to occur in the near future [Liben-Nowell 2003]
    - Track changes in large-scale data by periodically creating an agglomerative clustering and examining the evolution of clusters over time [Kubica *et al. 2002*]

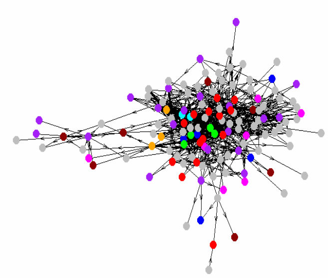
➔ However, all of them are only based on pure network properties, without knowing what people are talking about and why they have close relationship

Dynamic social networks

Email contacts within Enron: 1999



Email contacts within Enron: 2000



People with top 10 centralities in Enron

Centrality: Actor has high involvement in many relations, regardless of send/receive directionality (volume of activity)

Centrality	1999		2000		2001		2002	
	Name	Position	Name	Position	Name	Position	Name	Position
1	Mark_Taylor	Employee	David_Delainey	CEO	Steven_Kean	Vice_President	Kevin_Presto	Vice_President
2	Tana_Jones	N/A	Steven_Kean	Vice_President	John_Lavorato	CEO	Louise_Kitchen	President
3	Sara_Shackleton	N/A	John_Lavorato	CEO	Jeff_Dasovich	Employee	John_Lavorato	CEO
4	Richard_Sanders	Vice_President	Vince_Kaminski	Manager	Vince_Kaminski	Manager	Hunter_Shively	Vice_President
5	Elizabeth_Sager	Employee	Jeff_Skilling	CEO	Louise_Kitchen	President	James_Steffes	Vice_President
6	Mark_Haedicke	Managing_Director	Mike_McConnell	N/A	David_Delainey	CEO	Greg_Whalley	President
7	John_Hodge	Managing_Director	Greg_Whalley	President	Greg_Whalley	President	Fletcher_Sturm	Vice_President
8	Steven_Kean	Vice_President	Sally_Beck	Employee	Mark_Haedicke	Managing_Director	Doug_Gilbert-Smith	N/A
9	Dan_Hyvl	Employee	Jeffrey_A_Shankman	N/A	Phillip_Allen	N/A	Dana_Davis	N/A
10	Carol_Clair	In_house_lawyer	John_Arnold	Vice_President	Mary_Hain	In_house_lawyer	Mark_Haedicke	Managing_Director

People with top-10 prestige in Enron

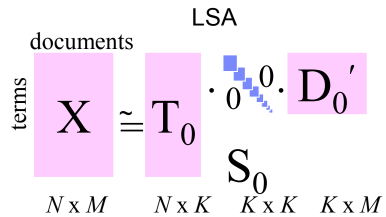
Prestige: Actor is the recipient of many directed ties

Prestige	1999		2000		2001		2002	
	Name	Position	Name	Position	Name	Position	Name	Position
1	John_Hodge	Managing_Director	John_Lavorato	CEO	John_Lavorato	CEO	Darron_Giron	N/A
2	Steven_Kean	Vice_President	Greg_Whalley	President	Louise_Kitchen	President	Phillip_Love	N/A
3	Vince_Kaminski	Manager	David_Delainey	CEO	Phillip_Allen	N/A	Kam_Keiser	Employee
4	Mark_Haedicke	Managing_Director	Steven_Kean	Vice_President	Greg_Whalley	President	Errol_McLaughlin	N/A
5	Elizabeth_Sager	Employee	Vince_Kaminski	Manager	Kevin_Presto	Vice_President	Stacey_White	N/A
6	Richard_Sanders	Vice_President	Rick_Buy	Manager	Barry_Tycholiz	Vice_President	Fletcher_Sturm	Vice_President
7	Kevin_Presto	Vice_President	Kevin_Presto	Vice_President	Steven_Kean	Vice_President	NA	NA
8	Mark_Taylor	Employee	Jeffrey_A_Shankman	N/A	Mike_Grigsby	N/A	NA	NA
9	Michelle_Cash	N/A	Phillip_Allen	N/A	David_Delainey	CEO	NA	NA
10	Stacy_Dickson	Employee	Jeff_Skilling	CEO	Hunter_Shively	Vice_President	NA	NA

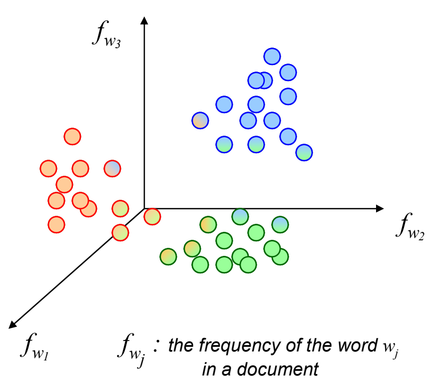
Most of them have relatively high position in Enron, which reveal the roles in the social network actually are almost corresponding to the roles in the real life

Content Analysis Prior Art I -- Latent Semantic Analysis

- Latent Semantic Analysis (LSA) [Landauer, Dumais 1997]
  - Descriptions:
    - Capture the semantic concepts of documents by mapping words into the latent semantic space which captures the possible synonym and polysemy of words
    - Training based on different level of documents. Experiments show the synergy of the # of training documents and the psychological studies of students at 4th, 10th, and college level. Used as an alternative to TOEFL test.
  - Based on truncated SVD of document-term matrix: optimal least-square projection to reduce dimensionality
  - Capture the concepts instead of words
    - Synonym
    - Polysemy



Traditional Content Clustering



f_{w_j} : *the frequency of the word w_j in a document*

Clustering:

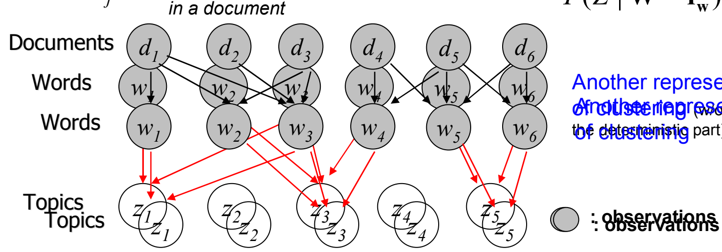
Partition the feature space into *segments* based on training documents. Each segment represents a topic / category. (← Topic Detection)

Hard clustering: e.g., K-mean clustering

$$d = \{f_{w_1}, f_{w_2}, \ldots, f_{w_N}\} \rightarrow z$$

Soft clustering: e.g., Fuzzy C-mean clustering

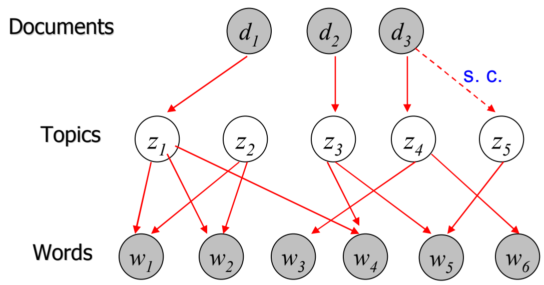
$$P(Z \mid \mathbf{W} = \mathbf{f_w})$$



Another representation of clustering

Another representation of clustering (w/o showing the deterministic part)

Content Clustering based on Bayesian Network



$h(D)$ $\qquad$ $P(Z \mid D)$

hard clustering $\qquad$ soft clustering

$P(W \mid Z)$

: observations

Bayesian Network:

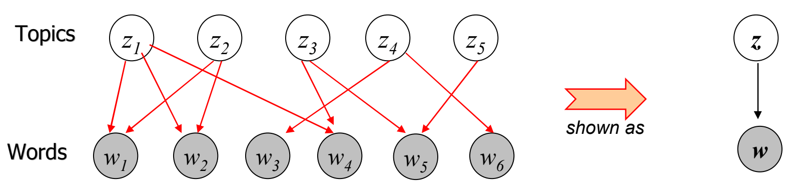
- Causality Network – models the causal relationship of attributes / nodes
- Allows hidden / latent nodes

Hard clustering:

$$h(D = d) = \arg\max_{z} P(\mathbf{W} = \mathbf{f}_{\mathbf{w}} \mid Z)$$ <= MLE

$$P(W \mid Z) = \frac{P(Z \mid W)P(W)}{P(Z)}$$ <= Bayes Theorem

Content Clustering based on Bayesian Network – Hard Clustering



: observations

$$P(W \mid Z) = \frac{P(Z \mid W)P(W)}{P(Z)} = \frac{P(Z \mid W)P(W)}{\int P(Z \mid W)dW}$$

N: the number of words
(The number of topics (M) are pre-determined)

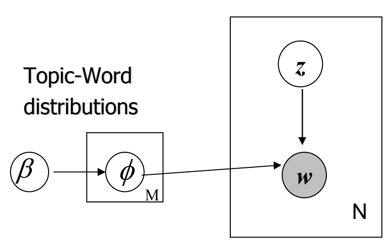
Major Solution 1 -- Dirichlet Process:

- Models P(W | Z) as mixtures of Dirichlet probabilities
- Before training, the prior of P(W|Z) can be a easy Dirichlet (uniform distribution). After training, P(W|Z) will still be Dirichlet. (← *The reason of using Dirichlet*)

Major Solution 2 -- Gibbs Sampling:

- A Markov chain Monte Carlo (MCMC) method for integration of large samples → calculate P(Z)

Topic-Word distributions

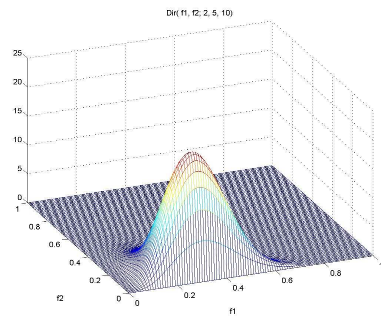


Latent Dirichlet Allocation (LDA) (Blei 2003)

Comparison of Dirichlet Distribution with Gaussian Mixture Models

- Dirichlet Distribution: $\rho(f_1, f_2, ..., f_{r-1}; a_1, a_2, ..., a_r) = \frac{\Gamma(N)}{\prod_{k=1}^{r} \Gamma(a_k)} f_1^{a_1 - 1} f_2^{a_2 - 1} ... f_r^{a_r - 1}$

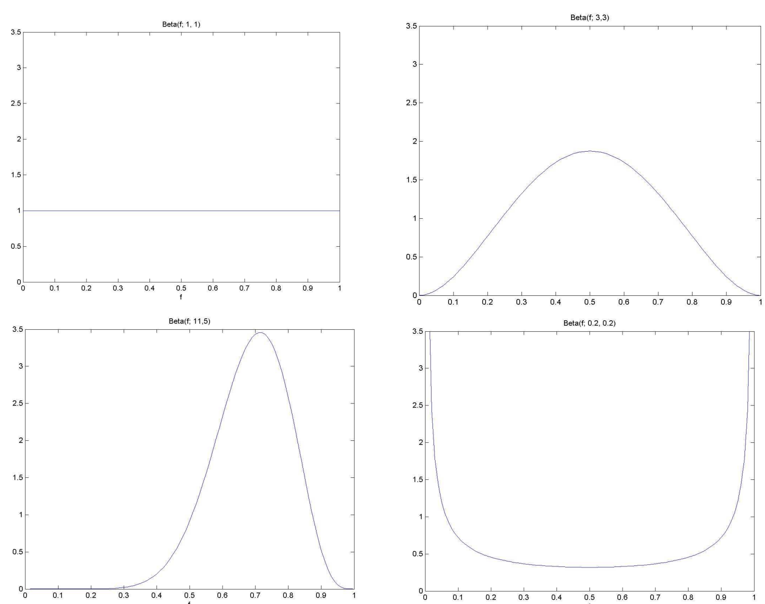
  $0 \le f_k \le 1 \qquad \sum_{k=1}^{r} f_k = 1$



- Multivariate Gaussian:

$$\rho(f_1, f_2, ..., f_{r-1}; \mu_1, \sigma_1, \mu_2, \sigma_2, ..., \mu_{r-1}, \sigma_{r-1}) = \frac{1}{(2\pi)^{r-1} \prod_{k=1}^{r-1} \sigma_k} e^{-\frac{(f_1 - \mu_1)^2}{\sigma_1^2}} e^{-\frac{(f_2 - \mu_2)^2}{\sigma_2^2}} ... e^{-\frac{(f_{r-1} - \mu_{r-1})^2}{\sigma_{r-1}^2}}$$

Beyond Gaussian: Examples of Dirichlet Distribution

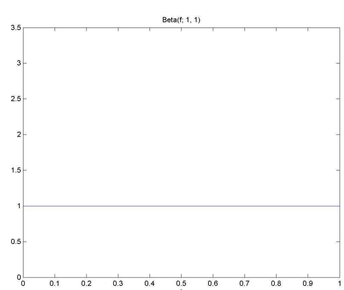


Importance of Dirichlet Distribution

- In 1982, Sandy Zabell proved that, if we make certain assumptions about an individual's beliefs, then that individual must use the Dirichlet density function to quantify any prior beliefs about a relative frequency.

Use Dirichlet Distribution to model prior and posterior beliefs (I)

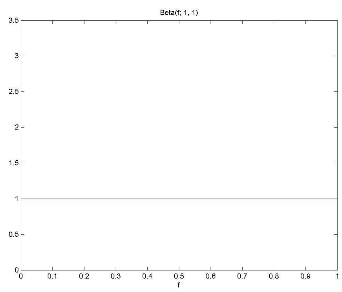
- Prior beliefs: $\rho(f) = beta(f; a, b)$
- E.g.: *fair* coin?
  - Flipping a coin, what's the probability of getting 'head'.



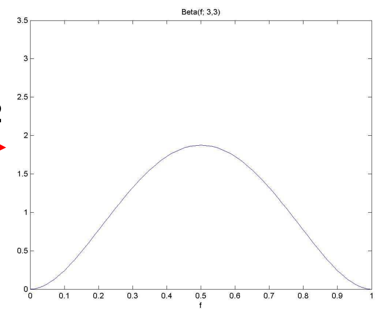
beta(1,1): Never tossed that coin.
No prior knowledge

Use Dirichlet Distribution to model prior and posterior beliefs (II)

- Prior beliefs: $\rho(f) = beta(f; a, b)$
- Posterior beliefs: $\rho(f \mid d) = beta(f; a+s, b+t)$



d: 2, 2

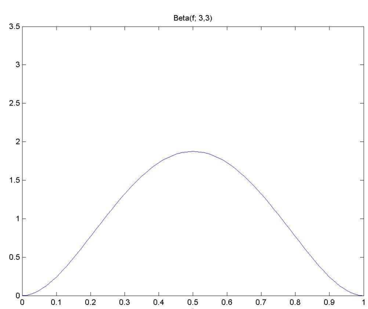


beta(1,1): No prior knowledge

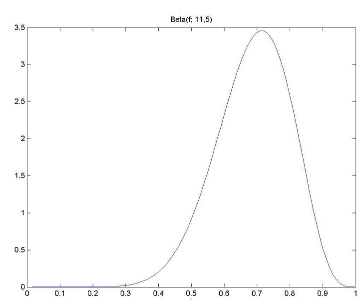
beta(3,3): posterior knowledge after 2 head trials (*s*) and 2 tail trials (*t*) -- this coin may be fair

Use Dirichlet Distribution to model prior and posterior beliefs (III)

- Prior beliefs: $\rho(f) = beta(f; a, b)$
- Posterior beliefs: $\rho(f \mid d) = beta(f; a+s, b+t)$



d: 8, 2

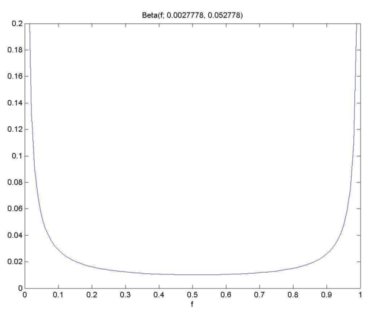


beta(3,3): prior knowledge -- this coin may be fair

beta(11,5): posterior belief -- the coin may not be fair after tossing 8 heads and 2 tails

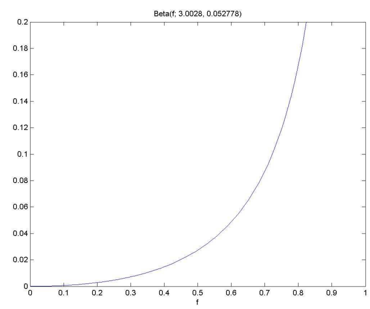
Use Dirichlet Distribution to model prior and posterior beliefs

- Another Example:
  - Our belief that an event, which in general has 5% occurrence possibility, may be true in our case.



beta(1/360,19/360): prior knowledge that an 5% chance-event may be true

d: 3,0



beta(1/360+3,19/360): posterior belief that an 5% chance-event may be true

Gibbs Sampling (I)

- Given a PDF, $\rho(X)$, usually it's difficult to get an integration, $\int f(X) \cdot \rho(X) dX$.
- We can use Gibbs Sampling to simulate the *samples* $X_s = \{X_{s,1}, X_{s,2}, \ldots, X_{s,T}\}$ that follows the PDF. Thus, we can do an integration by summing $f(X_s)$ from those samples.

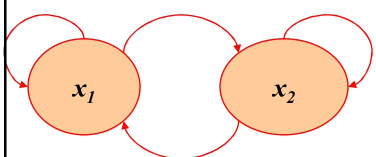
Gibbs Sampling (II)

- Gibbs Sampling ➔ Find a Markov Chain $\rho_s(X_i \mid X_{i-1})$, s.t., the outcome of samples follow the PDF $\rho(X)$

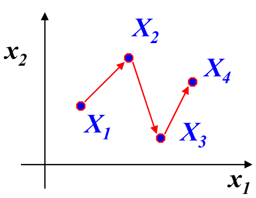
Example: 2-dimensional X vector: $X = (x_1, x_2)$

$\rho_s(X_i \mid X_{i-1})$ **is determined by** $\rho_{1,s}(x_{1,i} \mid x_{2,i-1})$, $\rho_{2,s}(x_{2,i} \mid x_{1,i})$

Derived from the PDF



⇨ $\rho_s(X_i \mid X_{i-1})$ ⇨



Some Insight on BN-based Content Clustering

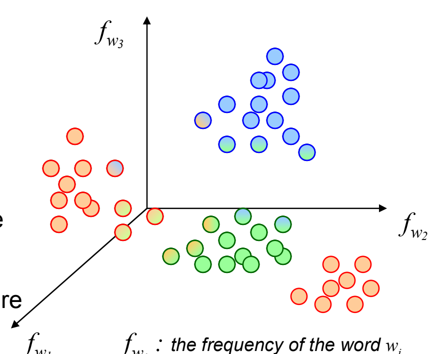
Bayesian Network:

• Models the *practical* causal relationships..

Content Clustering:

• Because documents and words are dependent,

➔ only close documents in the feature space can be clustered together as one topic.



f_{w_j} : the frequency of the word w_j in a document

⇒Incorporating human factors can possibly *link* multiple clusters together.

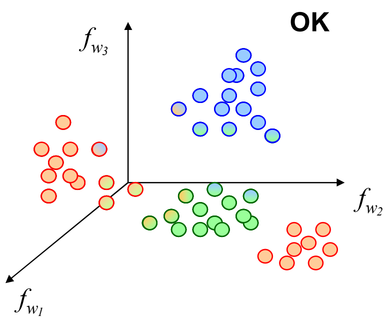
Why We Need Simultaneous Multimodality Clustering?

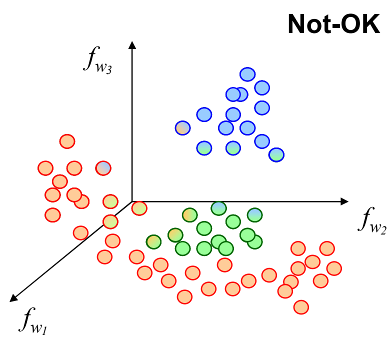
Multiple-Step Clustering:

• e.g., Naïve way to combine content filtering and collaborative filtering

➔ Independently cluster first. Combine later.

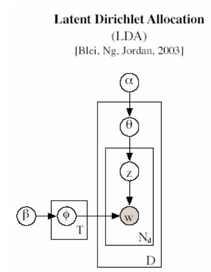
➔Simultaneous Multimodality Clustering is important.

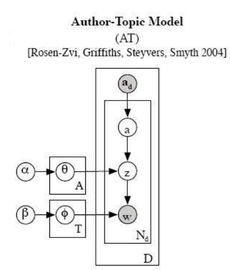




Content Analysis Prior Art II – Probabilistic LSA (pLSA), Latent Dirichlet Allocation (LDA) and AT Model

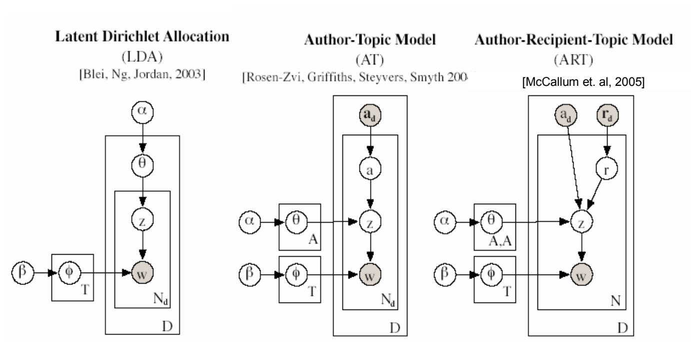
- Probabilistic LSA [Hofmann 1999]
  - Statistical view of LSA
- Latent Dirichlet Allocation (LDA) [Blei et al 2003]
  - A generative model which Introduce a Dirichlet prior for the class modeling at pLSA
- Author-Topic model [Rosen-Zvi et al 2004]
  - Try to recognize which part of the document is contributed by which co-author. A document with multiple authors is a mixture of the distributions associated with authors.
  - Each author is associated with a multinomial distribution over topic; Each topic is associated with a multinomial distribution over words





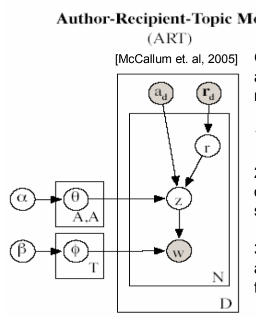
Content Analysis Prior Art III – Author-Receipt-Topic (ART) Model

- Author-Receipt-Topic model [McCallum et al. April 2005]
  - Given the sender and the set of receivers of an email, find senders have similar role in events



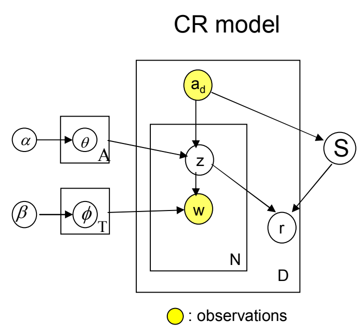
Novel Content-Time-Relation Algorithm -- I

- Content-Relation (CR) [Song, Lin, Tseng, Sun, KDD-submission Feb. 2005]:
  - Content topic classification
  - Integrate social network model
  - *➔Combine content and social relation information with Dirichlet allocations, a causal Bayesian network and an Exponential Random Graph Social Network Model.*



Given the sender and the set of receivers of an email:

1. Pick a receiver
2. Get the probability of a topic given the sender and receiver
3. Get the probability a word given the topic



Given the sender of an email:

1. Get the probability of a topic given the sender
2. Get the probability of the receiver given the sender and the topic
3. Get the probability of a word given the topic

a: sender/author, z: topic, S: social network (Exponential Random Graph Model / p* model), D: document/emailr: receivers, w: content words, N: Word set, T: Topic

Novel Content-Time-Relation Algorithm -- II

- Content-Time-Relation (CTR) [Song, Lin, Tseng, Sun, KDD-submission Feb. 2005]:
  - Incremental Latent Dirichlet Allocations
  - Capture evolutionary information
  - Integrate social network model

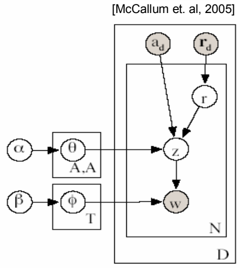
  → *Combine content, time and social relation information with Dirichlet allocations, a causal Bayesian network and an Exponential Random Graph Social Network Model.*

  - *Besides, for the time windowing, one can use Poisson distribution to replace the Dirichlet allocation, where* $\varphi = \gamma$ ^ $|t - t_0|$.

Author-Recipient-Topic Model

(ART)

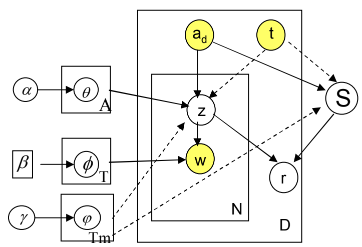
[McCallum et. al, 2005]



Given the sender and the set of receivers of an email:

1. Pick a receiver
2. Get the probability of a topic given the sender and receiver
3. Get the probability a word given the topic

CTR model



: observations

Given the sender and the time of an email:

1. Get the probability of a topic given the sender
2. Get the probability of the receiver given the sender and the topic
3. Get the probability of a word given the topic

Boxes represent replication.

Topic Analysis Results - Hot and cold topics in Enron Email Corpus

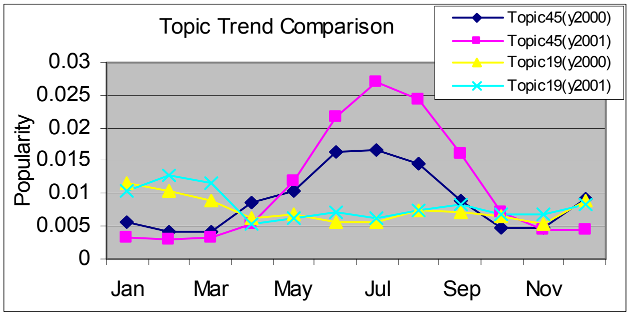
Table 1. Hot Topics

Meeting	Deal	Petroleum	Texas	Document
meeting plan conference balance presentation discussion	deal desk book bill group explore	Petroleum research dear photo Enron station	Houston Texas Enron north America street	letter draft attach comment review mark

Table 2. Cold Topics

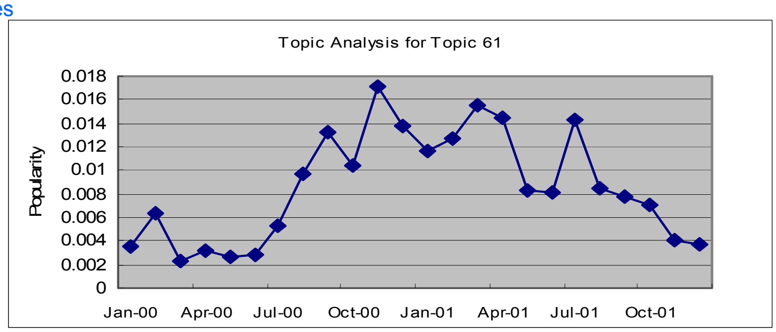
Trade	Stock	Network	Project	Market
trade London bank name Mexico conserve	Stock earn company share price new	network world user save secure system	Court state India server project govern	call market week trade description respond

Corporate Topic Trend Analysis Example: Yearly repeating events



Topic 45, which is talking about a schedule issue, reaches a peak during June to September. For topic 19, it is talking about a meeting issue. The trend repeats year to year.

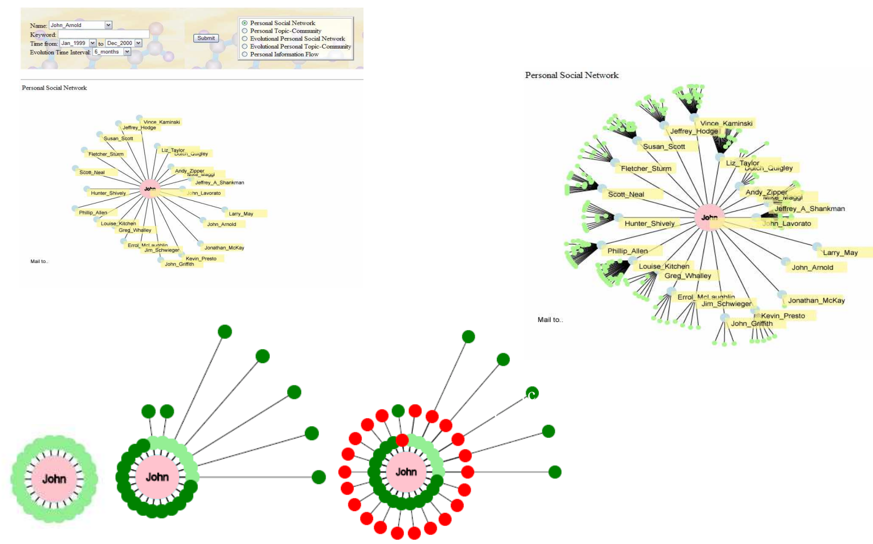
Topic Detection and Key People Detection of "California Power" Match Their Real-Life Roles



Key Words	power 0.089361 California 0.088160 electrical 0.087345 price 0.055940 energy 0.048817 generator 0.035345 market 0.033314 until 0.030681
Key People	Jeff_Dasovich 0.249863 James_Steffes 0.139212 Richard_Shapiro 0.096179 Mary_Hain 0.078131 Richard_Sanders 0.052866 Steven_Kean 0.044745 Vince_Kaminski 0.035953

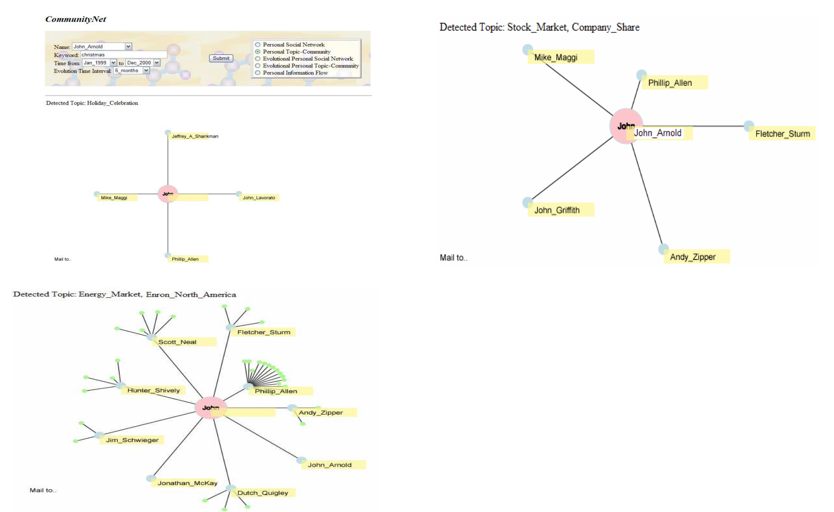
Event "California Energy Crisis" occurred at exactly this time period. Key people are active in this event except Vince_Kaminski …

Personal Social Network - who a user contacts with during a time period



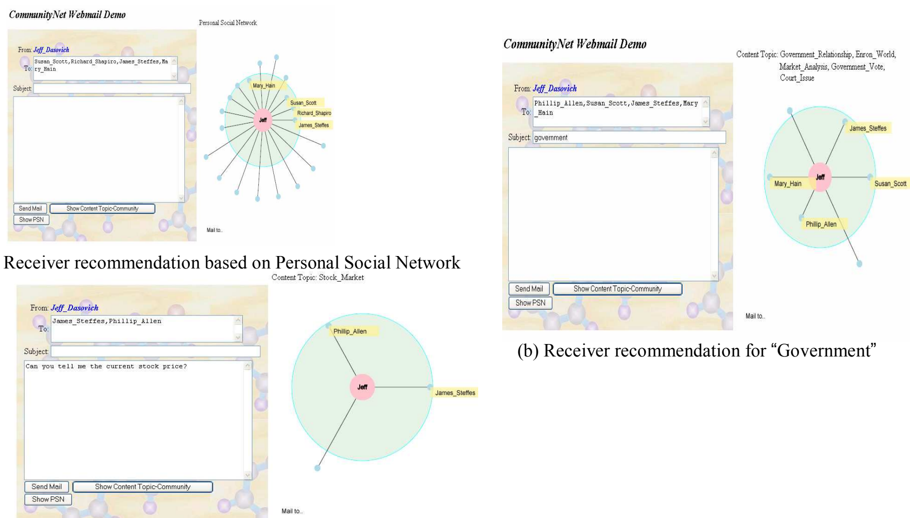
(b1) Jan-'99 to Dec-'99 (b2) Jan-'00 to Jun-'00 (b3) Jul-'00 to Dec-'00
(b) Evolutionary Personal Social Network

Personal Topic-Community Networks- who a user will contact with under different topics



(b) Personal Topic-Community Information flow for keyword "energy" with two-level personal Topic-Community network

Personal Social Capital Management -Receiver Recommendation Demo



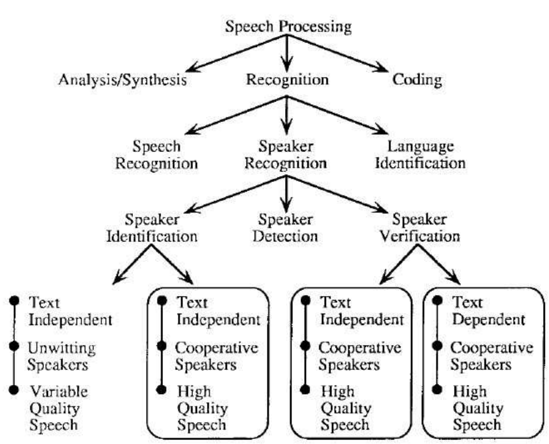
(a) Receiver recommendation based on Personal Social Network

(b) Receiver recommendation for "Government"

(c) Receiver recommendation for "Can you tell me the current stock price?"

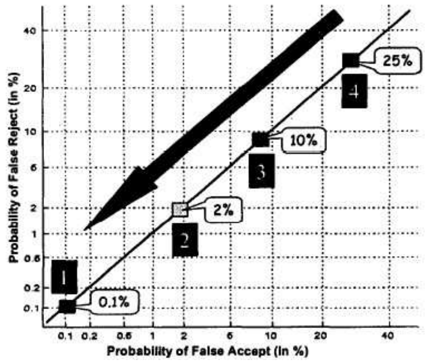
Speaker Recognition

- Speaker Verification
- Speaker Identification



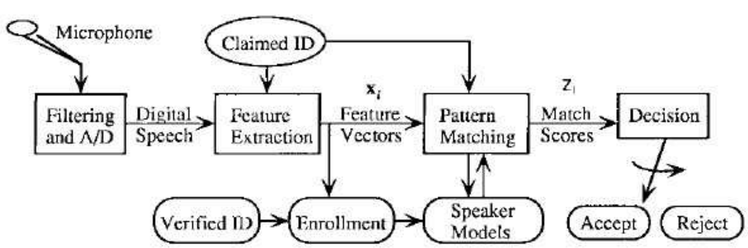
Usage Environment and Performance of State-of-the-art Systems

1. Text-depending using combinations lock phrases (e.g., 35-41-89). Clean data recorded using a single handset over multiple sessions. Used about 3 min of training data and 2 s test data. (0.1% - 1%).
2. Text-dependent using 10 digit strings. Telephone data using multiple handsets with multiple sessions. Two strings training data and single string verification. (1% - 5%)
3. Text-independent using conversational speech. Telephone data using multiple handsets with multiple sessions. Two minutes training data and 30 s test data. (7% - 15%).
4. Text-independent using read sentences. Very noisy radio data using multiple military radios and microphones with multiple sessions. Thirty sec training and 15 s testing. (20% - 35%).



Speaker Verification Process

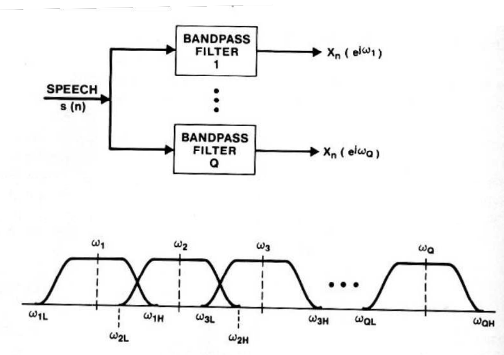
- Feature Extraction
- Classification Model



Common Audio Features

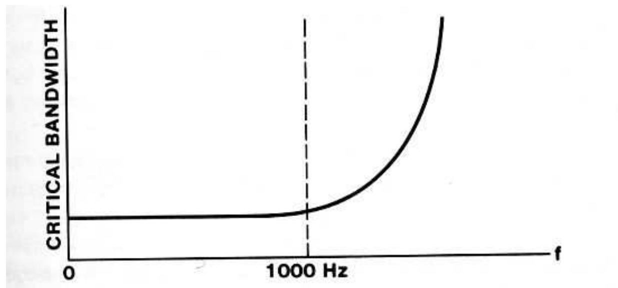
- Linear Predictive Coefficient (LPC)
- Linear Predictive Cepstral Coefficient (LPCC)
- Short-Time-Filter-Bank / Short-Time Fourier Transform
- Mel-Frequency Cepstral Coefficients (MFCC)

Bank-of-filters Analysis Model

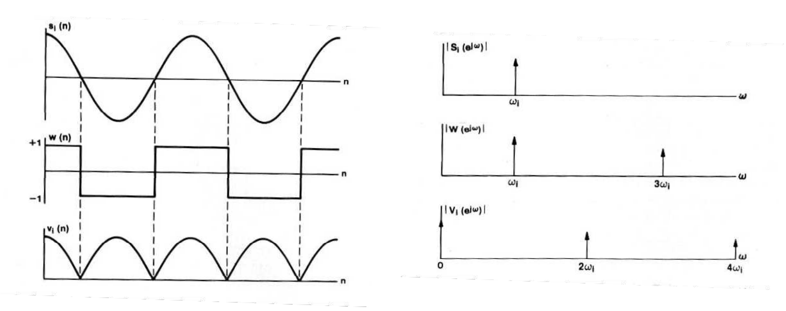


Human Voice Threshold

- Range of human voices

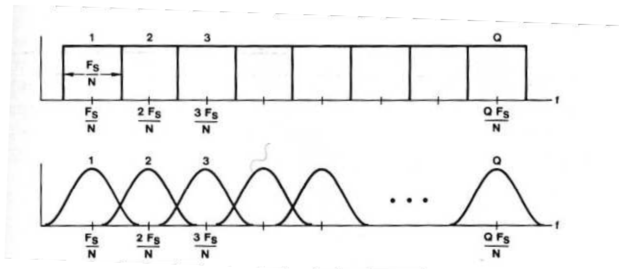


Typical waveforms and spectra for analysis of a pure sinusoid in the filter-bank model



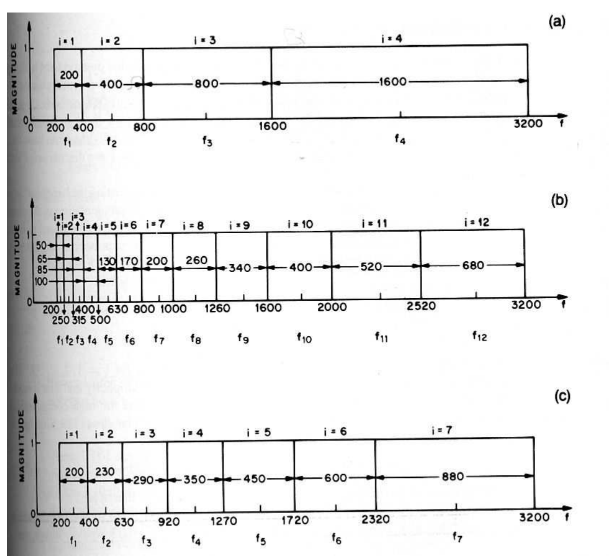
Short-Time Analysis

- Filter bank



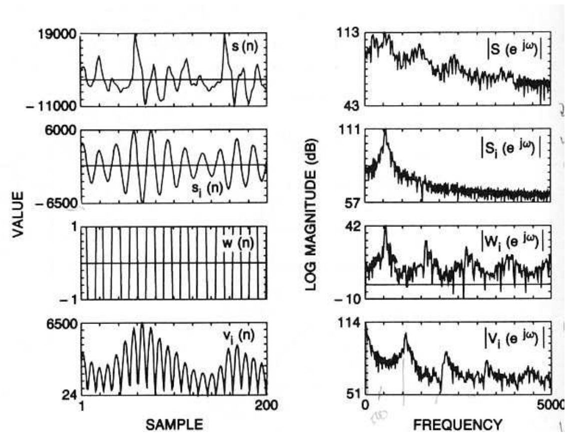
Filter Bank

- Analysis using different time scale
  - (a) a 4-channel octave band-filter bank
  - (b) a 12-channel third-octave band filter bank
  - (c) a 7-channel critical band scale filter bank
- These filter banks covering the telephone bandwidth range (200 – 3200 Hz).



Filter Bank Analysis

- Typical waveforms and spectra of a voice speech signal in the bank-of-filters analysis model



MFCC

The upper and lower passband frequencies of each filter are the center frequencies of the adjacent filters. The frequency range of the filters was chosen to cover the telephone bandwidth. The central frequencies of the 17 filters are (in Hz): 300, 400, 500, 600, 700, 800, 900, 1000, 1149, 1320, 1516, 1741, 2000, 2297, 2639, 3031 and 3482. Since the filter bank coefficients are highly correlated, a discrete cosine transform is used to de-correlate them:

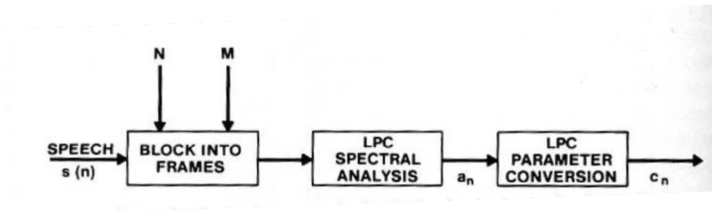
$$c_i = \frac{1}{N_F}\sum_{j=1}^{N_F} f_j \cos\{\frac{\pi i}{N_F}(j-0.5)\} \quad i = 0, 1, ..., N_F - 1 \qquad (1)$$

where N_F is the number of filters and f_j are the log filter bank energies. The MFCC feature vector is made up of $\{c_i, i = 1, 2, ..., N_F - 1\}$, ie. c_0 is omitted. c_0 represents the average value of the spectrum and hence is susceptible to varying background noise.

Linear Predictive Coding (LPC) Analysis Model

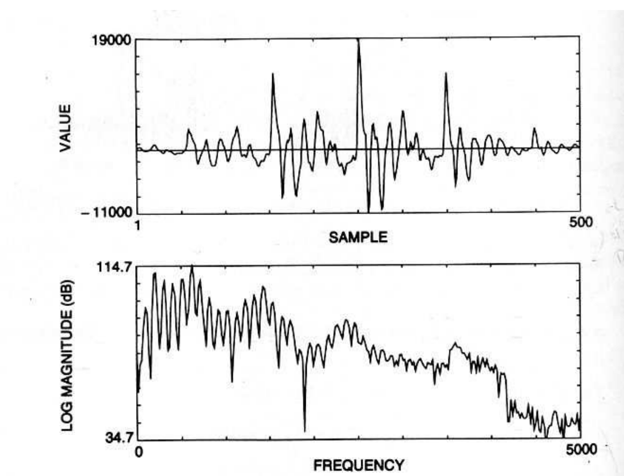
- LPC model:

  $A(z) = 1 + a_1z^{-1}+z_2z^{-2}+ \ldots + a_pz^{-p}$



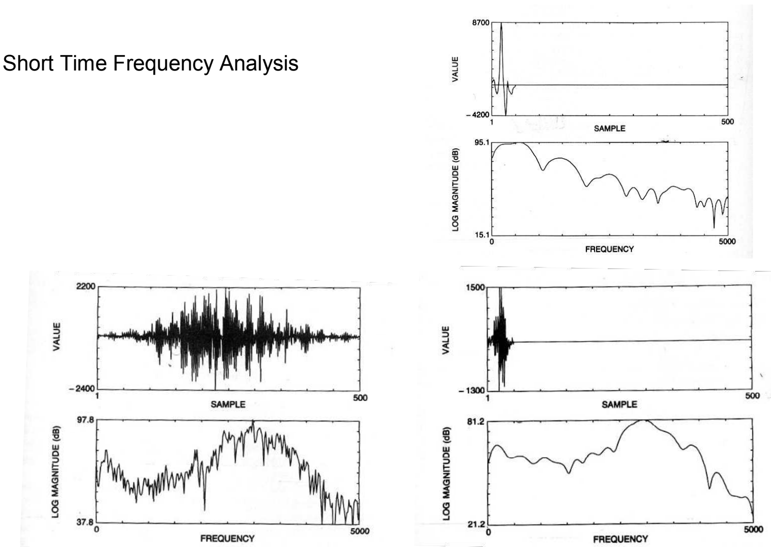
Example of the speech signal

- Frequency Transfor



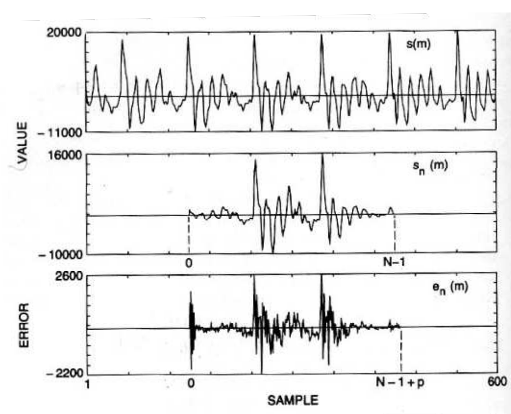
Example of Signal Analysis (II)

- Short Time Frequency Analysis



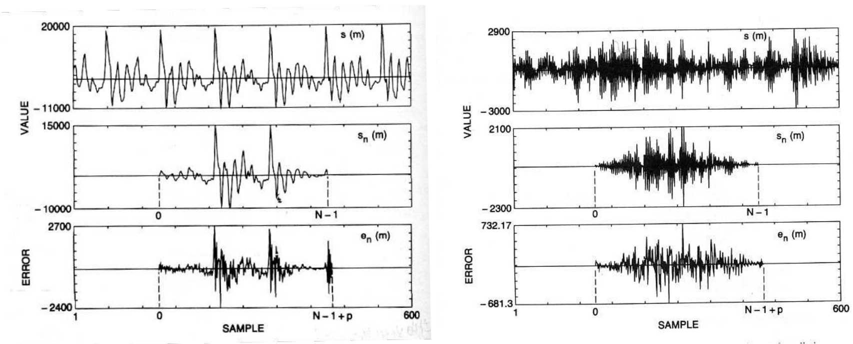
Example of Frequency Analysis (III)

- Illustration of speech sample, weighted speech section and prediction error



Example of Frequency Analysis (IV)

- Illustration of speech sample, weighted speech section and prediction error

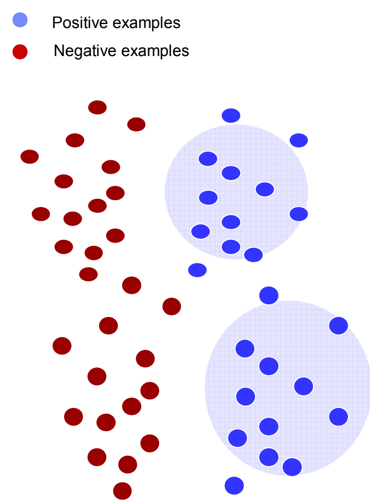


Linear Predictive Cepstral Coefficient

- Finding LPCC coefficients

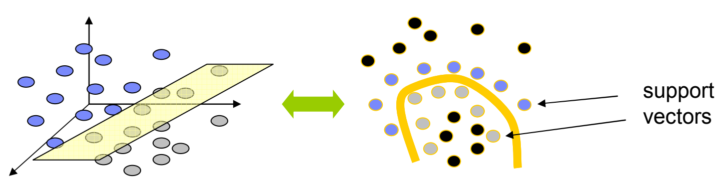
$$LPCC_i = -LPC_i + \frac{1}{i}\sum_{k=1}^{i-1}(i-k)LPC_k LPCC_{i-1}$$

Gaussian Mixture Models



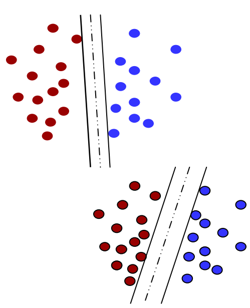
- Positive Examples are modeled as a summation of multiple Gaussians distributed in the feature space
- Pros:
  - The size of model parameters are very small: mean, std at each feature dim.
- Cons:
  - Models may vary depending on the initial condition

Support Vector Machine



$$f(x) = \sum_{i=0}^{N_s} \alpha_i y_i K(x, s_i) + b$$

$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2}$$



- Support Vector Machine
  - Largest margin hyperplane in the projected feature space
  - With good kernel choices, all operations can be done in low-dimensional input feature space
  - We use Radial Basis Functions as our kernels
  - Sequential Minimal Optimization = currently, fastest known algorithm for finding hyperplane

References

- X. Song, C.-Y. Lin, B. L. Tseng and M.-T. Sun, ""**Modeling and Predicting Personal Information Dissemination Behavior**," *ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, Chicago, August 2005.
- R. E. Neapolitan, "Learning Bayesian Networks," Prentice Hall, 2004.
- J. Campbell, "Speaker Recognition: A Tutorial", Proc. of the IEEE, Vo. 85, No. 9, Sep. 1997.
- D. A. Reynolds, "An Overview of Automatic Speaker Recognition Technology," ICASSP 2002
- L. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition", Prentice-Hall 1993 (Chapter 3)
- S.Y. Kung, Mak, and S. H. Lin, "Biometric Authentication", Prentice-Hall, 2005. (Chapter 9)

Homework Assignment #3 (due April 5)

1. [20 pts] Record your voice by reading the material provided by TA. Submit these audio clips to TA by this Thursday (March 24). These dataset will include two sets – training and testing. Two kinds of testing sets will be provided – the same document or different document.
2. If you are going to work on this homework by yourself:
   1. [50 pts] Implement LPCC or MFCC coefficient extractor. Apply it on the training and the testing data. Train your own speaker model and anti-you speaker model.
   2. [30 pts] Use the Nearest Neighborhood Method based on the Hausdoff distance to do the speaker verification. Illustrate the system performance.
3. If there are two people working on this homework together:
   1. [50 pts] Implement LPCC and MFCC extractors. Apply them on the training and the testing data. Train both team members' models. For each of the member, there should be your own model and an auti-you speaker model.
   2. [30 pts] The same as 2-2.
4. If three people are going to work together:
   1. [40 pts] The same as 3-1. Train models of these member.
   2. [20 pts] The same as 3-2.
   3. [20 pts] Use the Gaussian Mixture Models for training and testing.
5. [Extra Bonus 10 points]
   1. Repeat the processes. Conduct experiments on the speaker identification.
6. [Extract Bonus 30 points]
   1. For single person or two people: apply Gaussian Mixture Models on testing and training.
   2. For three people: apply Support Vector Machines for training and testing.