Computational Models of Discourse

Regina Barzilay

MIT

What is Discourse?

What is Discourse?

Landscape of Discourse Processing

- **Discourse Models:** cohesion-based, content-based, rhetorical, intentional
- **Applications:** anaphora resolution, segmentation, event ordering, summarization, natural language generation, dialogue systems
- **Methods:** supervised, unsupervised, reinforcement learniing

Discourse Exhibits Structure!

- Discourse can be partition into segments, which can be connected in a limited number of ways
- Speakers use linguistic devices to make this structure explicit
  cue phrases, intonation, gesture
- Listeners comprehend discourse by recognizing this structure
  - Kintsch, 1974: experiments with recall
  - Haviland&Clark, 1974: reading time for given/new information

Modeling Text Structure

Key Question: Can we identify consistent structural patterns in text?

> *"various types of [word] recurrence patterns seem to characterize various types of discourse"* (Harris, 1982)

Example

Stargazers Text(from Hearst, 1994)

- Intro - the search for life in space
- The moon's chemical composition
- How early proximity of the moon shaped it
- How the moon helped the life evolve on earth
- Improbability of the earth-moon system

Example

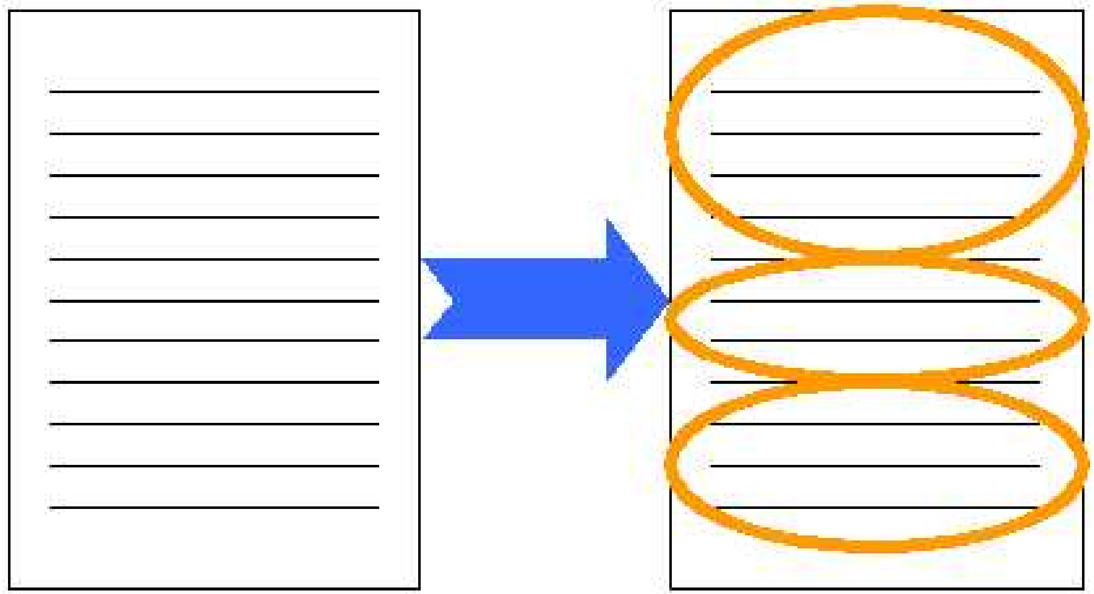
```
-------------------------------------------------------------------------------------------------------------+
Sentence:        05   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95|
-------------------------------------------------------------------------------------------------------------+
14       form    1     111 1   1                               1 1     1   1       1     1     1    1        |
 8  scientist                   11            1   1               1          1      1  1                     |
 5      space 11    1     1                                                            1                     |
25       star    1             1                                       11 22  111112  1 1  1   11 1111     1 |
 5     binary                                                           11  1          1                    1|
 4    trinary                                                            1   1         1                    1|
 8 astronomer 1                1                                        1 1           1   1    1  1          |
 7      orbit   1                    1                                     12    1 1                         |
 6       pull                          2     1 1                               1  1                          |
16     planet   1    1       11               1           1                21  11111                  1     1|
 7     galaxy    1                                                   1                1  11      1          1|
 4      lunar            1  1     1       1                                                                  |
19       life 1  1  1                              1    11 1  11  1      1                 1 1    1 111  1 1 |
27       moon        13  1111   1 1 22 21  21    21           11 1                                           |
 3       move                                         1   1   1                                              |
 7  continent                                         2 1 1 2 1                                              |
 3  shoreline                                               12                                               |
 6       time                      1               1  1  1     1                                          1  |
 3      water                                 11            1                                                |
 6        say                                1 1        1        11               1                          |
 3    species                                           1  1  1                                              |
-------------------------------------------------------------------------------------------------------------+
Sentence:        05   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95|
-------------------------------------------------------------------------------------------------------------+
```

Outline

- Text segmentation
- Coherence assessment

Text Segmentation

Goal: Partition a text into a sequence of topically coherent blocks

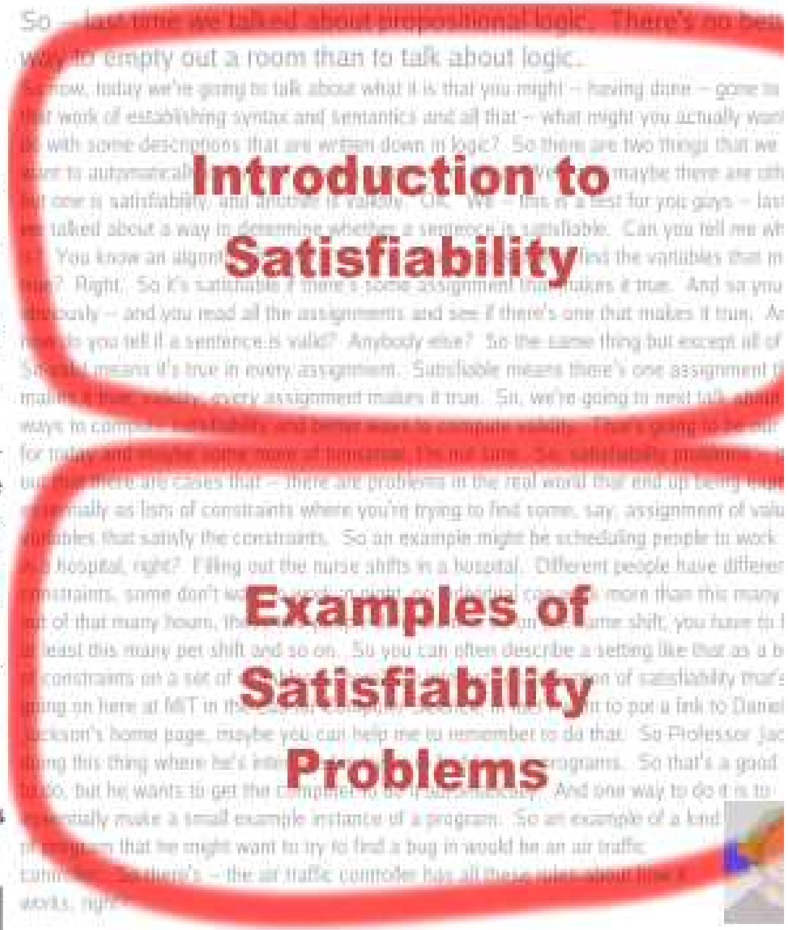


Applications: Information Retrieval, Summarization, Question Answering

Example

So -- last time we talked about propositional logic. There's no better way to empty out a room than to talk about logic.

So now, today we're going to talk about what it is that you might -- having done -- gone to all that work of establishing syntax and semantics and all that -- what might you actually want to do with some descriptions that are written down in logic? So there are two things that we might want to automatically determine about a sentence of logic. Well, and maybe there are others but one is satisfiability, and another is validity. OK. We -- this is a test for you guys -- last time we talked about a way to determine whether a sentence is satisfiable. Can you tell me what it is? You know an algorithm for this. Yes? It could be possible to find the variables that make it true? Right. So it's satisfiable if there's some assignment that makes it true. And so you could obviously -- and you read all the assignments and see if there's one that makes it true. And how do you tell if a sentence is valid? Anybody else? So the same thing but except all of them. So valid means it's true in every assignment. Satisfiable means there's one assignment that makes it true; validity, every assignment makes it true. So, we're going to next talk about better ways to compute satisfiability and better ways to compute validity. That's going to be our theme for today and maybe some more of tomorrow, I'm not sure. So, satisfiability problems -- it turns out that there are cases that -- there are problems in the real world that end up being expressed essentially as lists of constraints where you're trying to find some, say, assignment of values to variables that satisfy the constraints. So an example might be scheduling people to work shifts in a hospital, right? Filling out the nurse shifts in a hospital. Different people have different constraints, some don't want to work at night, no individual can work more than this many hours out of that many hours, these two people don't want to be on the same shift, you have to have at least this many per shift and so on. So you can often describe a setting like that as a bunch of constraints on a set of variables. There's an interesting application of satisfiability that's going on here at MIT in the Lab for Computer Science, in fact I want to put a link to Daniel Jackson's home page, maybe you can help me to remember to do that. So Professor Jackson's doing this thing where he's interested in trying to find bugs in programs. So that's a good thing to do, but he wants to get the computer to do it automatically. And one way to do it is to essentially make a small example instance of a program. So an example of a kind of program that he might want to try to find a bug in would be an air traffic controller. So there's -- the air traffic controller has all these rules about how it works, right?



Flow model of discourse

Chafe'76:

> "Our data ... suggest that as a speaker moves from focus to focus (or from thought to thought) there are certain points at which they may be a more or less radical change in space, time, character configuration, event structure, or even world ... At points where all these change in a maximal way, an episode boundary is strongly present."

Segmentation: Agreement

Percent agreement — ratio between observed agreements and possible agreements

A	B	C
−	−	−
−	−	−
+	−	−
−	+	+
−	−	−
+	+	+
−	−	−
−	−	−

$$\frac{22}{8 * 3} = 91\%$$

Results on Agreement

People can reliably predict segment boundaries!

Grosz&Hirschbergberg'92	newspaper text	74-95%
Hearst'93	expository text	80%
Passanneau&Litman'93	monologues	82-92%

Linguistic Basis: Lexical Cohesion

- Common assumption of unsupervised algorithms
  - Word repetition indicates topical cohesion [Halliday & Hasan,'76]
  - Variations in lexical distribution signal topic changes

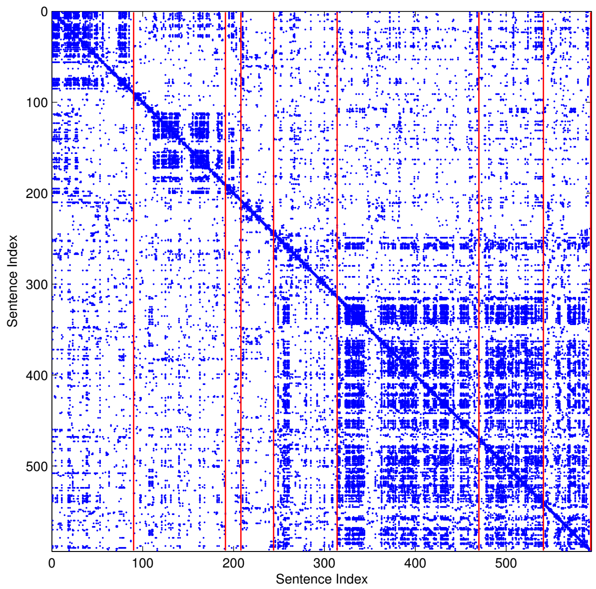
What is the instantaneous speed? Well, speed is not sign sensitive.

It's like a spacecraft in orbit or an elevator with a cut cable .

DotPlot Representation

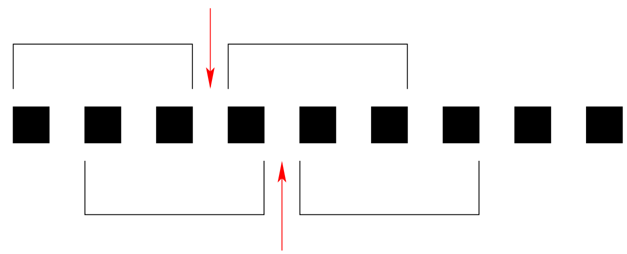
Key assumption: change in lexical distribution signals topic change (Hearst '94)

- Dotplot Representation: (i, j) – similarity between sentence i and sentence j



Segmentation Algorithm of Hearst

- Initial segmentation
  - Divide a text into equal blocks of k words
- Similarity Computation
  - compute similarity between m blocks on the right and the left of the candidate boundary



- Boundary Detection
  - place a boundary where similarity score reaches local minimum

Similarity Computation: Representation

Vector-Space Representation

SENTENCE$_1$: I like apples

SENTENCE$_2$: Apples are good for you

Vocabulary	Apples	Are	For	Good	I	Like	you
Sentence$_1$	1	0	0	0	1	1	0
Sentence$_2$	1	1	1	1	0	0	1

Similarity Computation: Cosine Measure

Cosine of angle between two vectors in n-dimensional space

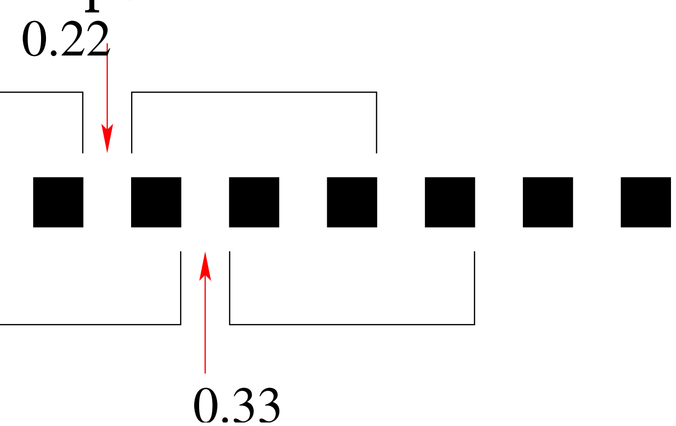
$$\text{sim}(b_1,b_2) = \frac{\sum_t w_{y,b_1} w_{t,b_2}}{\sqrt{\sum_t w^2_{t,b_1} \sum_{t=1}^{n} w^2_{t,b_2}}}$$

SENTENCE$_1$: 1 0 0 0 1 1 0

SENTENCE$_2$: 1 1 1 1 0 0 1

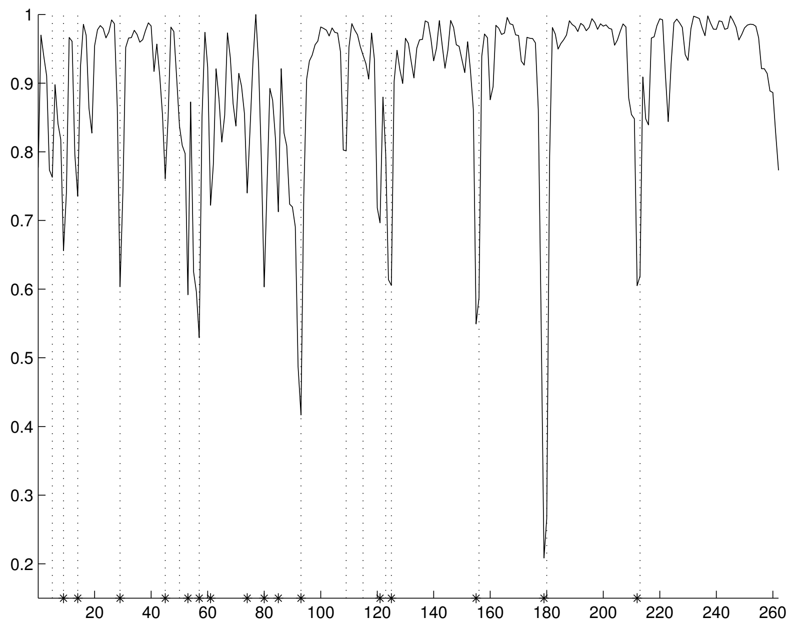
$$\text{sim}(S_1,S_2) = \frac{1*0+0*1+0*1+0*1+1*0+1*0+0*1}{\sqrt{(1^2+0^2+0^2+0^2+1^2+1^2+0^2)*(1^2+1^2+1^2+1^2+0^2+0^2+1^2)}} = 0.26$$

Output of Similarity computation:



Boundary Detection

- Boundaries correspond to local minima in the gap plot

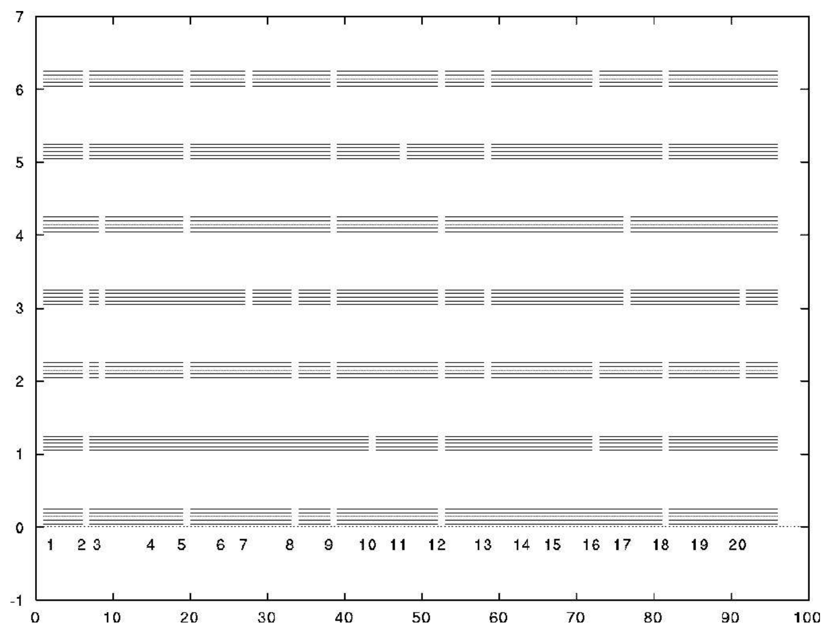


- Number of segments is based on the minima threshold ($s - \sigma/2$, where s and σ corresponds to average and standard deviation of local minima)

Segmentation Evaluation

Comparison with human-annotated segments(Hearst'94):

- 13 articles (1800 and 2500 words)
- 7 judges
- boundary if three judges agree on the same segmentation point

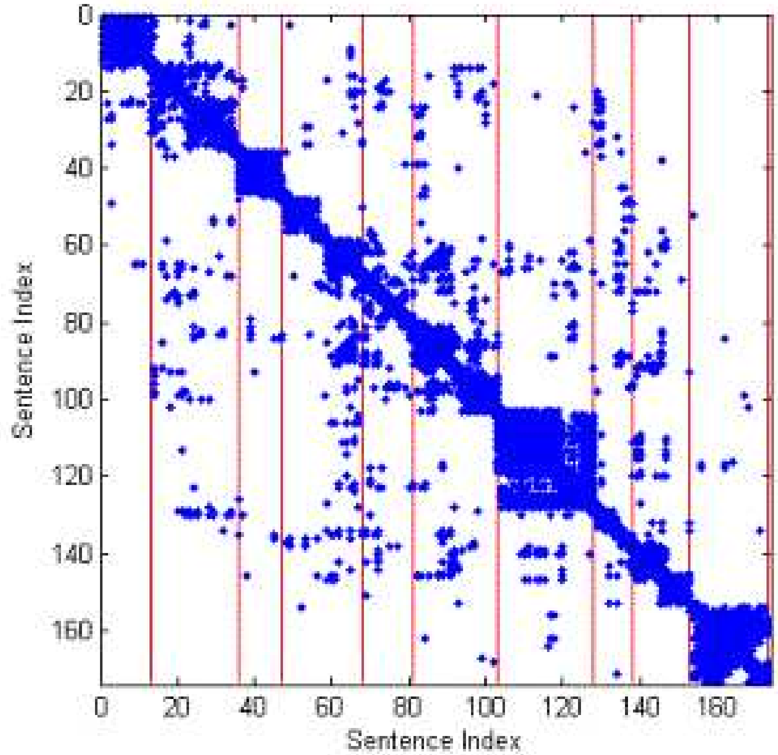


Evaluation Results

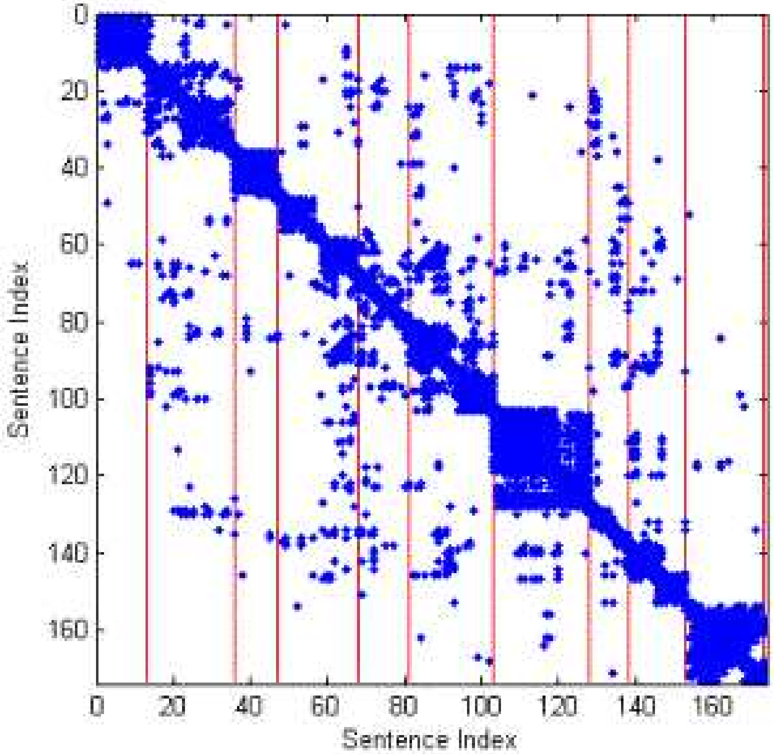
Methods	Precision	Recall
Random Baseline 33%	0.44	0.37
Random Baseline 41%	0.43	0.42
Original method+thesaurus-based similarity	0.64	0.58
Original method	0.66	0.61
Judges	0.81	0.71

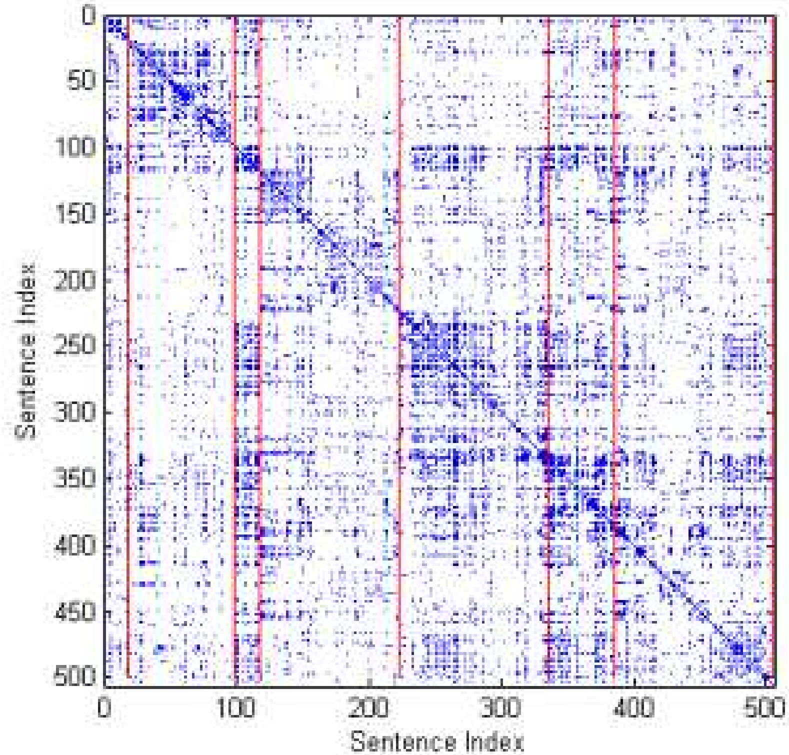
Synthetic Text Dotplot

- Broadcast News, synthetic document collections
  - Exhibit sharp segment transitions



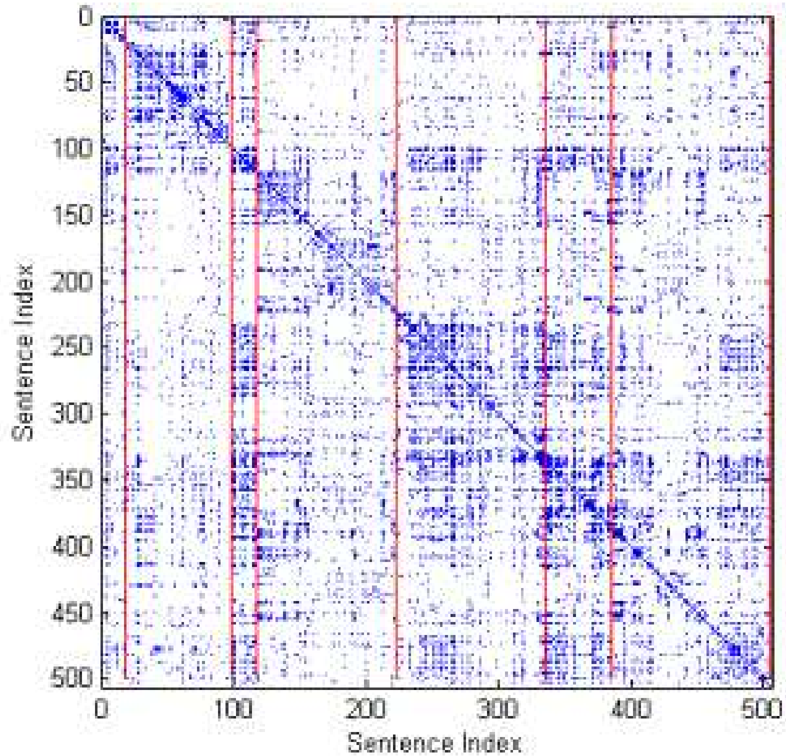
Synthetic vs. Real Data





Physics Lecture Dotplot

- Spoken Lecture Data
  - Exhibit very subtle topical transitions



Motion → Instantaneous Velocity → Average Acceleration → Numerical Example

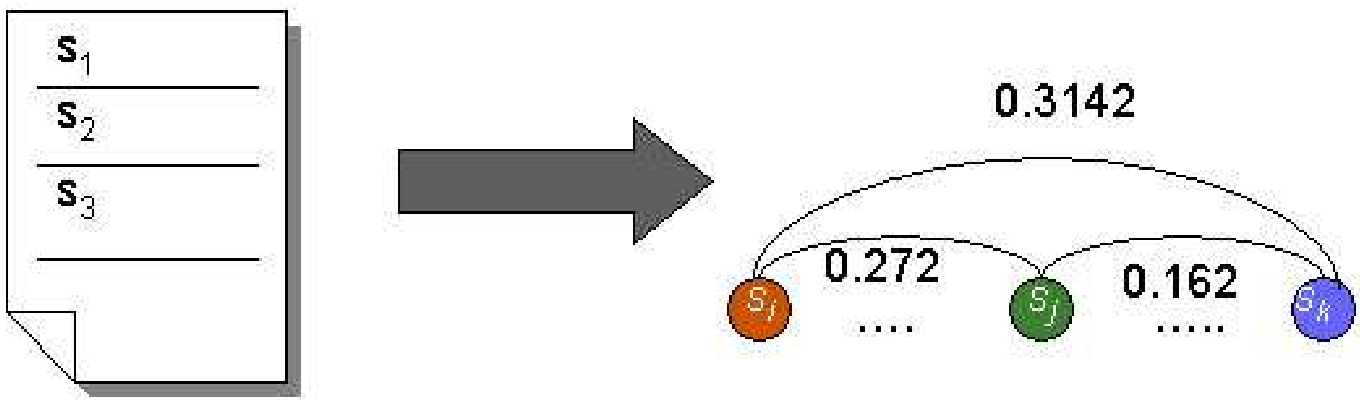
Minimum Cut Segmentation

- New graph-theoretic formalization of the segmentation objective
  - jointly maximizes within-cluster similarity and minimizes between-cluster similarity
  - Incorporates long-range lexical dependencies
- Exact, fast decoding using dynamic programming

Key Strength: Can detect subtle topic changes

Graph Based Representation

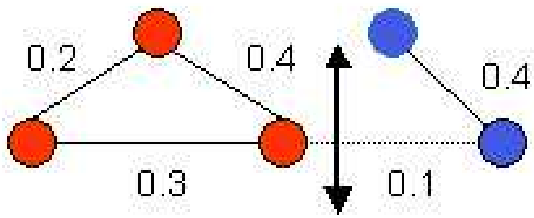
- Let $G(V,E)$ be a weighted, undirected, fully-connected graph
- Graph nodes represent textual units (e.g. sentences)
- Edge weights $w(s_i, s_j)$ indicate pairwise sentence similarity



Graph Cut Definitions

- Graph Cut - partitioning of the graph into two disjoint sets of nodes *A*,*B*
- **Between-segment similarity** (Cut Value) - sum of the edge weights between A,B
- **Within-segment similarity** (Volume) - sum of the edge weights for nodes in A
- Normalized Cut Value [Shi & Malik '00] :

$$Ncut(A,B) = \frac{cut(A,B)}{vol(A)} + \frac{cut(A,B)}{vol(B)}$$



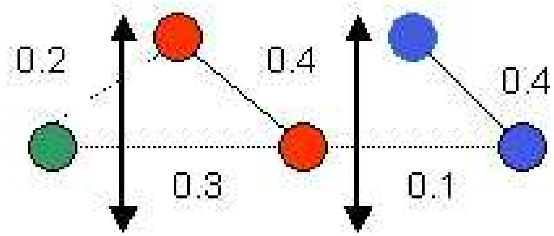
Normalized Cut Value = 0.6

Multi-way Graph Cuts

- K-way Graph Cut: partitioning of the graph into K disjoint sets, $A_1, \ldots A_k$
- K-way Normalized Cut Value:

$$Ncut_k(A_1, \ldots A_k) = \frac{cut(A_1, V - A_1)}{vol(A_1)} + \ldots + \frac{cut(A_k, V - A_k)}{vol(A_k)}$$

- 3-way Cut Example:



Normalized Cut Value = 1.8

Optimization Objective

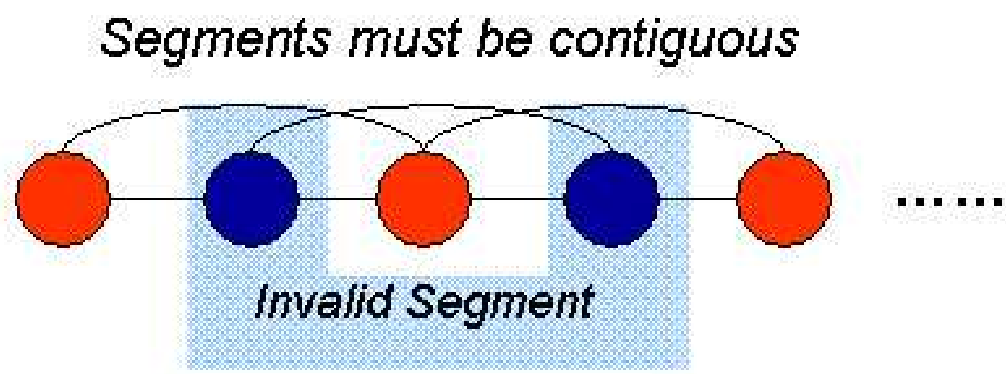
- For given k, we seek the k-way cut that minimizes the normalized cut value:

$$\min_{A_1,\ldots A_k} \frac{cut(A_1, V - A_1)}{vol(A_1)} + \ldots + \frac{cut(A_k, V - A_k)}{vol(A_k)}$$

- With this objective, we jointly
  - minimize the Cut Value ~ similarity between segments
  - maximize the Volume ~ similarity within segments

Linearity Constraint

- Without further constraints, this optimization is *NP*-complete [Papadimitriou '00]
- However, the segmentation problem imposes a natural **linearity** constraint on the form of the solution:



- Reformulation: $\min_{A_1,\ldots A_k} Ncut_k(A_1,\ldots A_k)$
  $\text{s.t. linearity constraint}$

Dynamic Programming Solution

Exact solution can be found using dynamic programming in *O(kn²)* time:

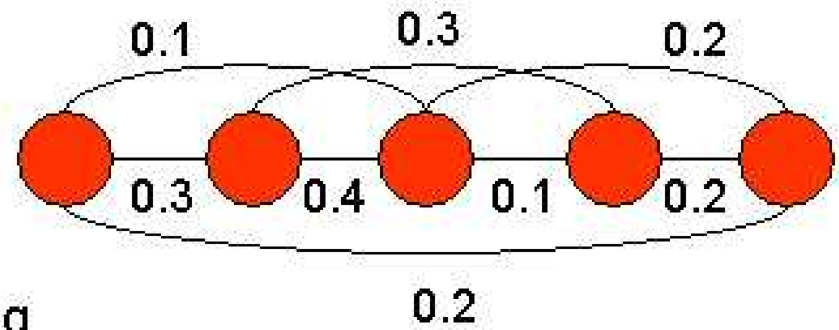
- $C[i,m]$: Minimum normalized cut of the segmentation of the first m sentences into i segments

- $C[i,m]$ can be computed recursively by choosing the best sentence *j* prior to *m* to begin the *i*th segment:

$$C\left[i,m\right] = \min_{j<m} \left[C\left[i-1,j\right] + \frac{cut\left[A_{j,m}, V - A_{j,m}\right]}{vol\left[A_{j,m}\right]} \right]$$

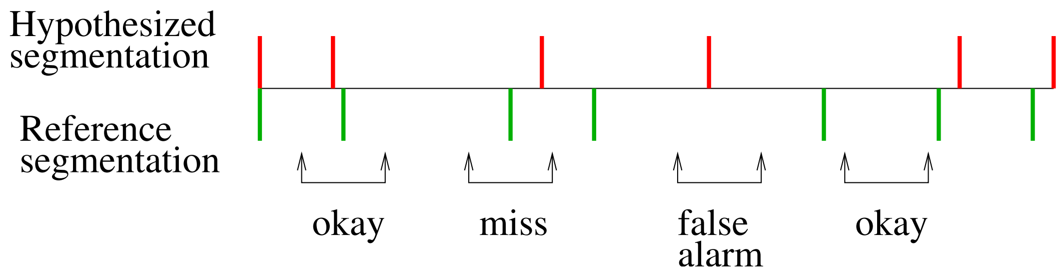
Graph Construction

- Node representation
  - Fixed blocks of text
- Topology
  - Fully-connected Graph
- Edge Weights
  - Weighted Cosine Similarity
  - Word Occurrence Smoothing



$$\tilde{s}_i = \sum_{j=i}^{i+k} e^{-\alpha(j-i)} s_j,$$

Evaluation Metric: P_k Measure



P_k: Probability that a randomly chosen pair of words k words apart is inconsistently classified (Beeferman '99)

- Set k to half of average segment length
- At each location, determine whether the two ends of the probe are in the same or different location. Increase a counter if the algorithm's segmentation disagree
- Normalize the count between 0 and 1 based on the number of measurements taken

Notes on P_k measure

- $P_k \in [0, 1]$, the lower the better
- Random segmentation: $P_k \approx 0.5$
- On synthetic corpus: $P_k \in [0.05, 0.2]$
- On real segmentation tasks: $P_k \in [0.2, 0.4]$

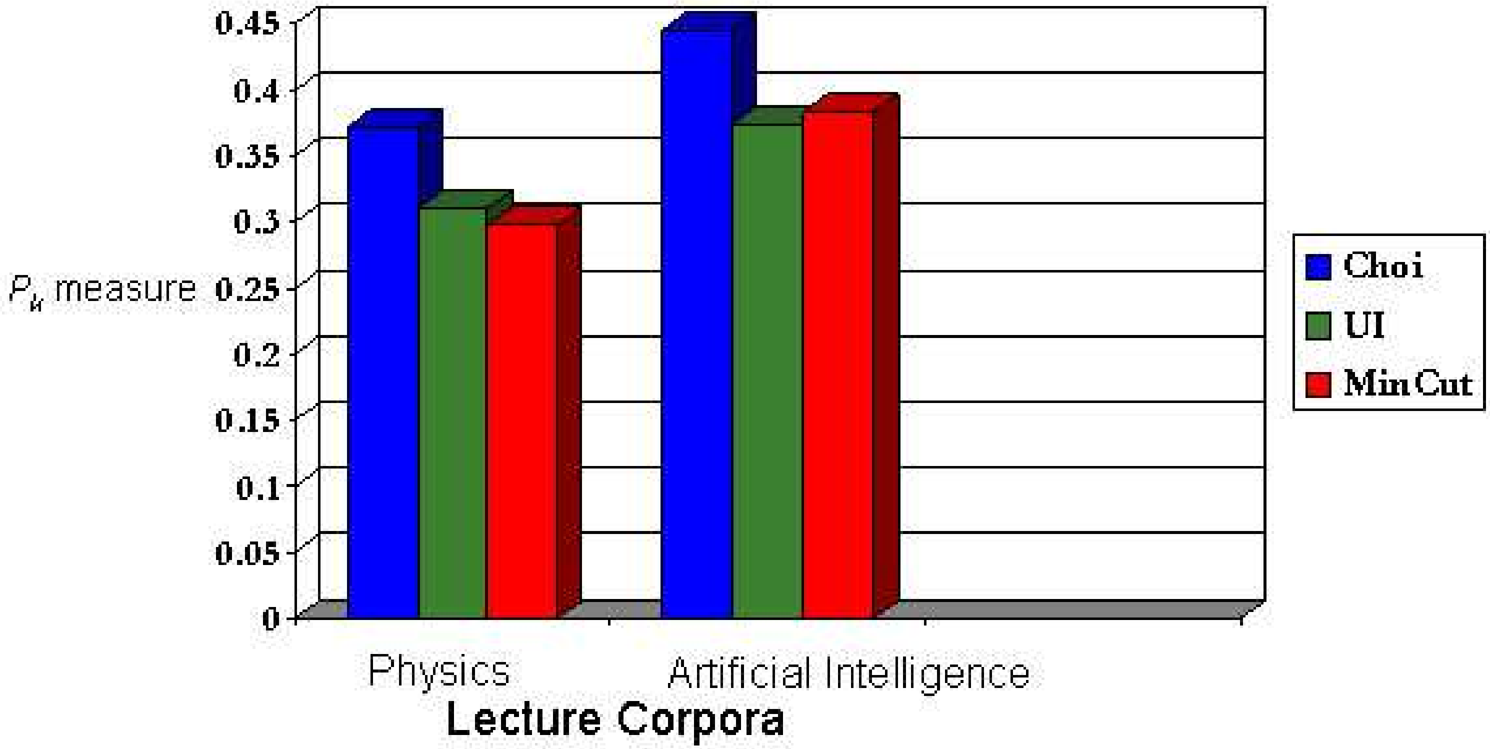
Experiments

- Data: MIT Physics and AI Lecture Corpus
  - Verbose and colloquial language
  - Subtle topic transitions
  - Automatic Speech Recognition Error

- Baselines: State-of-the-art unsupervised segmentation systems
  - Utiyama & Isahara (UI) 2001 - language modeling approach
  - Choi 2000 - local similarity-based approach

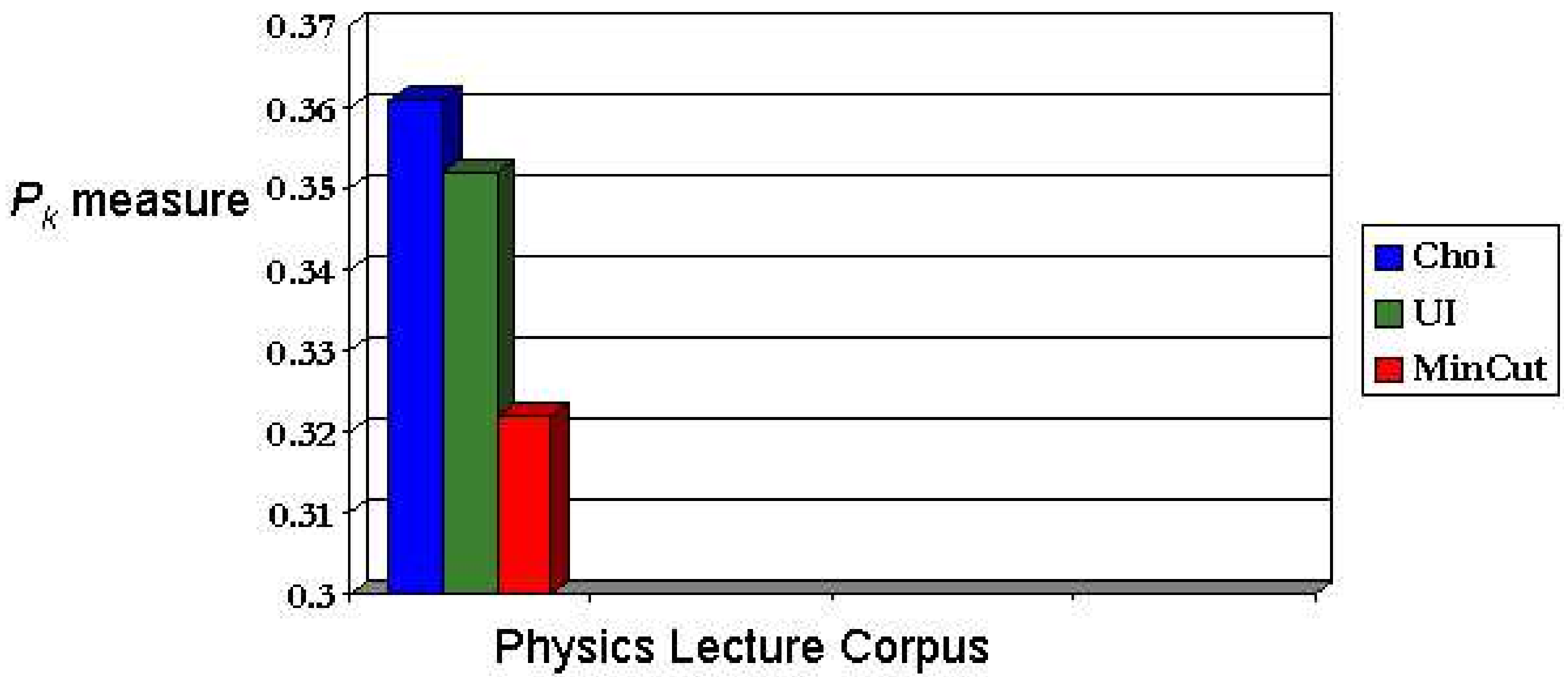
To control for segmentation granularity, the target number of segments for the baselines and our system is fixed

Results: Manually Transcribed Data



Human Agreement: $P_k \in [0.22; 0.42]$

Results: ASR Data (WER = 20%)



Impact of Long-range Dependencies

- Experiment: remove edges between nodes separated by a specified cutoff

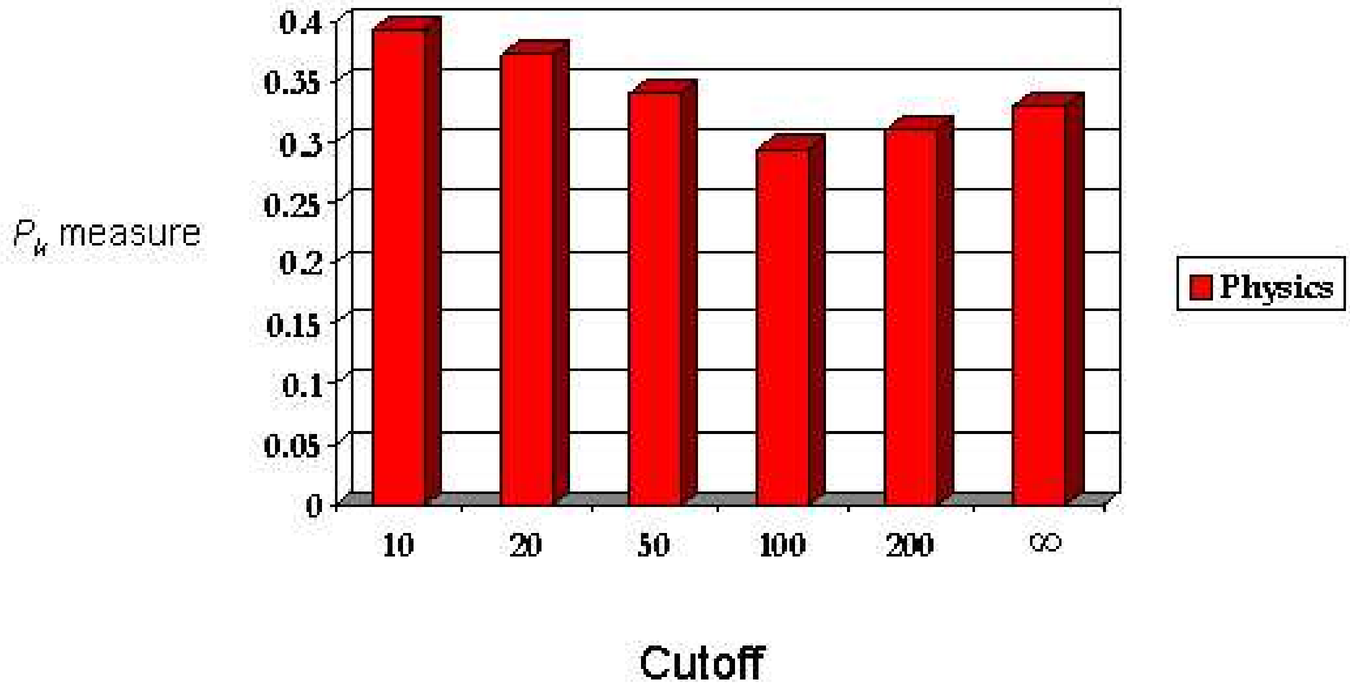
Edges connecting node X

Cutoff = 3

Cutoff = 2

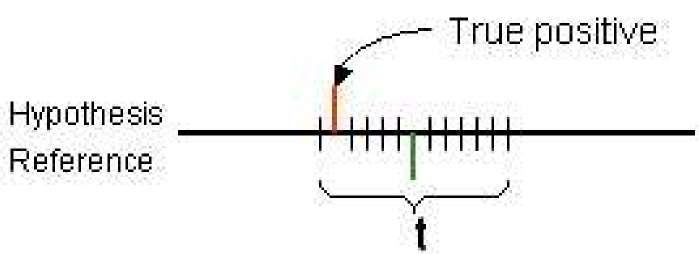
Impact of Long-range Dependencies

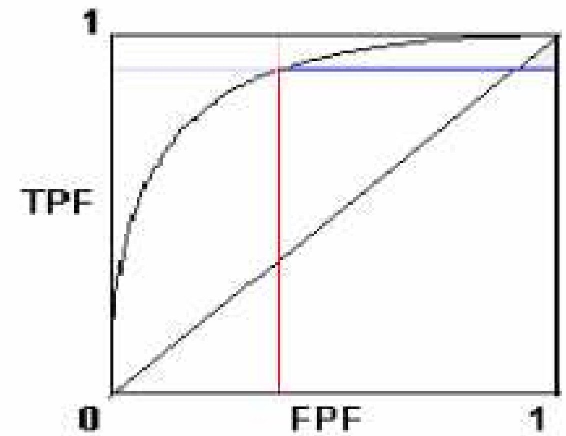
- Long-range dependencies improve performance



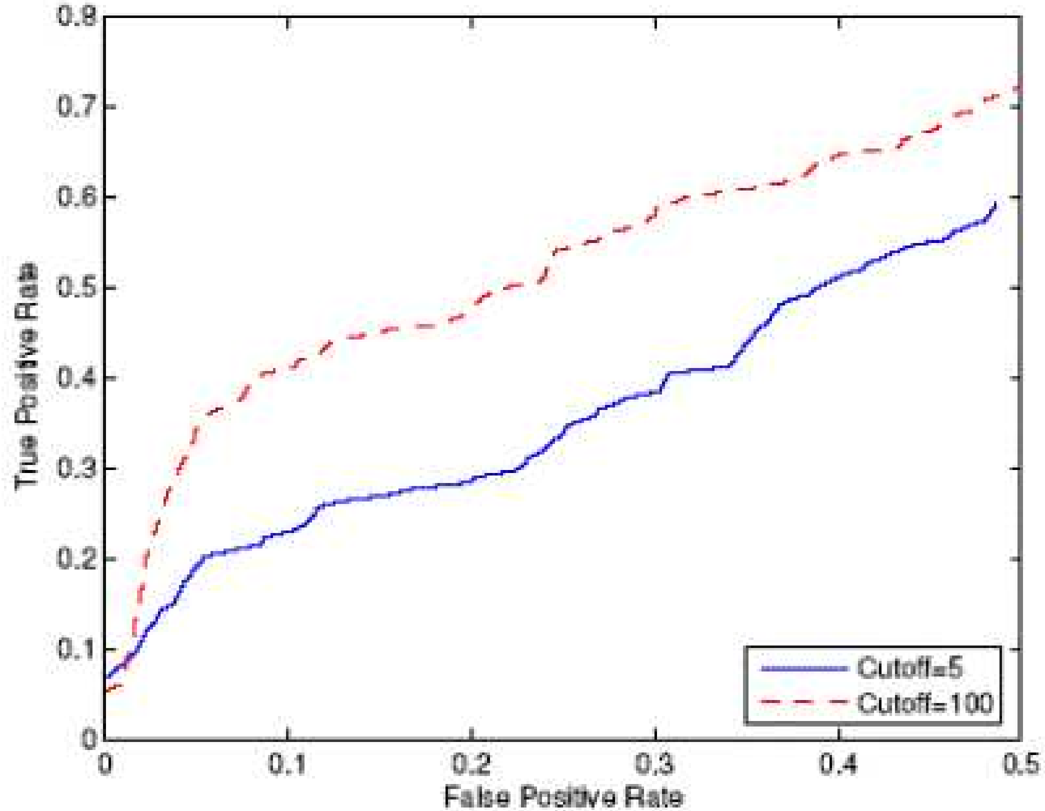
Evaluation Metrics - ROC

- Receiver Operating Characteristic Curve – represents tradeoff between true positives and false positives
- In segmentation, a true positive is a hypothesized boundary that occurs within a threshold *t* of the true boundary
- By varying *t*, we obtain points along the ROC curve





ROC Plot: Physics Lecture Data



Mincut ROC Plot with edge cutoffs at 5 and 100 sentences

Outline

- Text segmentation
- Coherence assessment

Modeling Coherence

Active networks and virtual machines have a long history of collaborating in this manner. The basic tenet of this solution is the refinement of Scheme. The disadvantage of this type of approach, however, is that public-private key pair and red-black trees are rarely incompatible.

- **Coherence** is a property of well-written texts that makes them easier to read and understand than a sequence of randomly strung sentences
- **Local coherence** captures text organization at the level of sentence-to-sentence transitions

Centering Theory

Grosz&Joshi&Weinstein,1983; Strube&Hahn,1999; Poesio&Stevenson&Di Eugenio&Hitzeman,2004

- Constraints on the entity distribution in a coherent text
  - Focus is the most salient entity in a discourse segment
  - Transition between adjacent sentences is characterized in terms of focus switch
- Constraints on linguistic realization of focus
  - Focus is more likely to be realized as subject or object
  - Focus is more likely to be referred to with anaphoric expression

Phenomena to be Explained

Johh went to his favorite music store to buy a piano.

He had frequented the store for many years.

He was excited that he could finally buy a piano.

He arrived just as the store was closing for the day.

John went to his favorite music store to buy a piano.

It was a store John had frequented for many years.

He was excited that he could finally buy a piano.

It was closing just as John arrived.

Analysis

- The same content, different realization
- Variation in coherence arises from choice of syntactic expressions and syntactic forms

Another Example

John really goofs sometimes.
Yesterday was a beautiful day and he was excited about trying out his new sailboat.
He wanted Tony to join him on a sailing trip.
He called him at 6am.
He was sick and furious at being woken up so early.

Centering Theory: Basics

- Unit of analysis: centers
- "Affiliation" of a center: utterance (U) and discourse segment (DS)
- Function of a center: to link between a given utterance and other utterances in discourse

Center Typology

- Types:
  - Forward-looking Centers C_f (U, DS)
  - Backward-looking Centers C_b (U, DS)
- Connection: C_b (U_n) connects with one of C_f (U_{n-1})

Constraints on Distribution of Centers

- C_f is determined only by U;
- C_f are partially ordered in terms of salience
- The most highly ranked element of C_f (U_{n-1}) is realized as C_b (U_n)
- Syntax plays role in ambiguity resolution: subj $>$ ind obj $>$ obj $>$ $others$
- Types of transitions: center continuation, center retaining, center shifting

Center Continuation

Continuation of the center from one utterance not only to the next, but also to subsequent utterances

- $C_b(U_{n+1})=C_b(U_n)$
- $C_b(U_{n+1})$ is the most highly ranked element of $C_f(U_{n+1})$ (thus, likely to be $C_b(U_{n+2})$)

Center Retaining

Retention of the center from one utterance to the next

- $C_b(U_{n+1}) = C_b(U_n)$
- $C_b(U_{n+1})$ is not the most highly ranked element of $C_f(U_{n+1})$ (thus, unlikely to be $C_b(U_{n+2})$)

Center Shifting

Shifting the center, if it is neither retained no continued

- $\mathrm{C}_b(\mathrm{U}_{n+1}) <> \mathrm{C}_b(\mathrm{U}_n)$

Coherent Discourse

Coherence is established via center continuation

John went to his favorite music store to buy a piano.	John went to his favorite music store to buy a piano.
He had frequented the store for many years.	It was a store John had frequented for many years.
He was excited that he could finally buy a piano.	He was excited that he could finally buy a piano.
He arrived just as the store was closing for the day.	It was closing just as John arrived.

Application to Essay Grading

(Miltsakaki&Kukich'00)

- Framework: GMAT e-rater
- Implementation: manual annotation of coreference information
- Grading: based on ratio of shifts
- Data: GMAT essays

Study results

- Correlation between shifts and low grades (established using t-test)
- Improvement of score prediction in 57%

Statistical Approach

Key Premise: the distribution of entities in locally coherent discourse exhibits certain regularities

- Abstract a text into an entity-based representation that encodes syntactic and distributional information
- Learn properties of coherent texts, given a training set of coherent and incoherent texts

Text Representation

- Entity Grid — a two-dimensional array that captures the distribution of discourse entities across text sentences
- Discourse Entity — a class of coreferent noun phrases

Input Text

1 Former Chilean dictator Augusto Pinochet, was arrested in London on October 14th, 1998.

2 Pinochet, 82, was recovering from surgery.

3 The arrest was in response to an extradition warrant served by a Spanish judge.

4 Pinochet was charged with murdering thousands, including many Spaniards.

5 He is awaiting a hearing, his fate in the balance.

6 American scholars applauded the arrest.

Input Text with Syntactic Annotation

Use Collins' parser(1997):

1. [Former Chilean dictator Augusto Pinochet]$_{\mathbf{S}}$, was arrested in [London]$_{\mathbf{X}}$ on [October 14th]$_{\mathbf{X}}$ 1998.
2. [Pinochet]$_{\mathbf{S}}$, 82, was recovering from [surgery]$_{\mathbf{X}}$.
3. [The arrest]$_{\mathbf{S}}$ was in [response]$_{\mathbf{X}}$ to [an extradition warrant]$_{\mathbf{X}}$ served by [a Spanish judge]$_{\mathbf{S}}$.
4. [Pinochet]$_{\mathbf{S}}$ was charged with murdering [thousands]$_{\mathbf{O}}$, including many [Spaniards]$_{\mathbf{O}}$.
5. [He]$_{\mathbf{S}}$ is awaiting [a hearing]$_{\mathbf{O}}$, [his fate]$_{\mathbf{X}}$ in [the balance]$_{\mathbf{X}}$.
6. [American scholars]$_{\mathbf{S}}$ applauded the [arrest]$_{\mathbf{O}}$.

Notation: **S**=subjects, **O**=object, **X**=other

Input Text with Coreference Information

Use noun-phrase coreference tool (Ng and Cardie, 2002):

1. [Former Chilean dictator Augusto Pinochet]$_{\mathbf{S}}$, was arrested in [London]$_{\mathbf{X}}$ on [October 14]$_{\mathbf{X}}$ 1998.
2. [Pinochet]$_{\mathbf{S}}$, 82, was recovering from [surgery]$_{\mathbf{X}}$.
3. [The arrest]$_{\mathbf{S}}$ was in [response]$_{\mathbf{X}}$ to [an extradition warrant]$_{\mathbf{X}}$ served by [a Spanish judge]$_{\mathbf{S}}$.
4. [Pinochet]$_{\mathbf{S}}$ was charged with murdering [thousands]$_{\mathbf{O}}$, including many [Spaniards]$_{\mathbf{O}}$.
5. [He]$_{\mathbf{S}}$ is awaiting [a hearing]$_{\mathbf{O}}$, [his fate]$_{\mathbf{X}}$ in [the balance]$_{\mathbf{X}}$.
6. [American scholars]$_{\mathbf{S}}$ applauded the [arrest]$_{\mathbf{O}}$.

Output Entity Grid

	Pinochet	London	October	Surgery	Arrest	Extradition	Warrant	Judge	Thousands	Spaniards	Hearing	Fate	Balance	Scholars	
1	s	x	x	–	–	–	–	–	–	–	–	–	–	–	1
2	s	–	–	x	–	–	–	–	–	–	–	–	–	–	2
3	–	–	–	–	s	x	x	s	–	–	–	–	–	–	3
4	s	–	–	–	–	–	–	–	o	o	–	–	–	–	4
5	s	–	–	–	–	–	–	–	–	–	o	x	x	–	5
6	–	–	–	–	o	–	–	–	–	–	–	–	–	s	6

Comparing Grids

S	S	S	X	X	–	–	–	–	–	–	–	–	–	–
–	–	S	–	–	X	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	S	X	X	O	–	–	–	–	–
–	–	S	–	–	–	–	–	–	–	O	O	–	–	–
–	–	S	–	–	–	–	–	–	–	–	–	O	X	X
–	–	–	–	–	–	O	–	–	–	–	–	–	–	–

S	S	X	X	X	–	–	–	–	–	–	–	–	–	X
–	–	X	–	–	X	–	–	–	–	–	–	–	–	X
–	–	X	–	–	–	–	X	X	O	–	–	–	–	X
–	–	X	–	–	–	–	–	–	–	O	O	–	–	X
–	–	X	–	–	–	–	–	–	–	–	–	O	X	X
–	–	X	–	–	–	O	–	–	–	–	–	–	–	X

Coherence Assessment

- Text is encoded as a distribution over entity transition types
- Entity transition type — $\{\text{S}, \text{O}, \text{X}, \text{–}\}^n$

	S S	S O	S X	S –	O S	O O	O X	O –	X S	X O	X X	X –	– S	– O	– X	– –
d_{i1}	0	0	0	.03	0	0	0	.02	.07	0	0	.12	.02	.02	.05	.25
d_{i2}	.02	0	0	.03	0	0	0	.06	0	0	0	.05	.03	.07	.07	.29

How to select relevant transition types?:

- Use all the unigrams, bigrams, . . . over $\{\text{S}, \text{O}, \text{X}, \text{–}\}$
- Do feature selection

Text Encoding as Feature Vector

	S S	S O	S X	S –	O S	O O	O X	O –	X S	X O	X X	X –	– S	– O	– X	– –
d_{i1}	0	0	0	.03	0	0	0	.02	.07	0	0	.12	.02	.02	.05	.25
d_{i2}	.02	0	0	.03	0	0	0	.06	0	0	0	.05	.03	.07	.07	.29

Each grid rendering x_{ij} of a document d_i is represented by a feature vector:

$$\Phi(x_{ij}) = (p_1(x_{ij}), p_2(x_{ij}), \ldots, p_m(x_{ij}))$$

where m is the number of all predefined entity transitions, and $p_t(x_{ij})$ the probability of transition t in the grid x_{ij}

Learning a Ranking Function

- **Training Set**
  Ordered pairs (x_{ij}, x_{ik}), where x_{ij} and x_{ik} are renderings of the same document d_i, and x_{ij} exhibits a higher degree of coherence than x_{ik}

- **Training Procedure**
  - Goal: Find a parameter vector $\vec{w}$ that yields a "ranking score" function $\vec{w} \cdot \Phi(x_{ij})$ satisfying:

    $$\vec{w} \cdot (\Phi(x_{ij}) - \Phi(x_{ik})) > 0$$
    $$\forall (x_{ij}, x_{ik}) \text{ in training set}$$

  - Method: Constraint optimization problem solved using the search technique described in Joachims (2002)

Evaluation: Information Ordering

- Goal: recover the most coherent sentence ordering
- Basic set-up:
  - Input: a pair of a source document and a permutation of its sentences

  - Task: find a source document via coherence ranking
- Data: Training 4000 pairs, Testing 4000 pairs (Natural disasters and Transportation Safety Reports)

Information Ordering

(a) During a third practice forced landing, with the landing gear extended, the CFI took over the controls.

(b) The certified flight instructor (CFI) and the private pilot, her husband, had flown a previous flight that day and practiced maneuvers at altitude.

(c) The private pilot performed two practice power off landings from the downwind to runway 18.

(d) When the airplane developed a high sink rate during the turn to final, the CFI realized that the airplane was low and slow.

(e) After a refueling stop, they departed for another training flight.

Information Ordering

(b) The certified flight instructor (CFI) and the private pilot, her husband, had flown a previous flight that day and practiced maneuvers at altitude.

(e) After a refueling stop, they departed for another training flight.

(c) The private pilot performed two practice power off landings from the downwind to runway 18.

(a) During a third practice forced landing, with the landing gear extended, the CFI took over the controls.

(d) When the airplane developed a high sink rate during the turn to final, the CFI realized that the airplane was low and slow.

Evaluation: Summarization

- Goal: select the most coherent summary among several alternatives
- Basic set-up:
  - Input: a pair of system summaries
  - Task: predict the ranking provided by human
- Data: 96 summary pairs for training, 32 pairs for testing (from DUC 2003)

Baseline: LSA

Coherence Metric: Average distance between adjacent sentences measured by cosine (Foltz et al. 1998)

- Shown to correlate with human judgments
- Fully automatic
- Orthogonal to ours (lexicalized)

Evaluation Results

Tasks:

- O_1=ordering(Disasters)
- O_2=ordering(Reports)
- S=summary ranking

Model	O_1	O_2	S
Grid	87.3	90.4	81.3
LSA	72.1	72.1	25.0

Varying Linguistic Complexity

What is the effect of syntactic knowledge?

- Reduce alphabet to { **X,–** }

Model	O_1	O_2	S
+Syntax	87.3	90.4	68.8
-Syntax	86.9	88.3	62.5

Conclusions

- Word distribution patterns strongly correlate with discourse patterns within a text
- Distributional-level approaches can be successfully applied to text-level relations