Text Summarization

Regina Barzilay

MIT

December, 2005

What is summarization?

Identifies the most important points of a text and expresses them in a shorter document.

Summarization process:

- interpret the text;
- extract the relevant information (topics of the source);
- condense extracted information and create summary representation;
- present summary representation to reader in natural language.

Types of summaries

- Extracts are summaries consisting entirely of material copied from the input document (e.g., extract 25% of original document).
- Abstracts are summaries consisting of material that is not present in the input document.
- Indicative summaries help us decide whether to read the document or not.
- Informative summaries cover all the salient information in the source (replace the full document).

Extracts vs. Abstracts

The Gettysburg Address

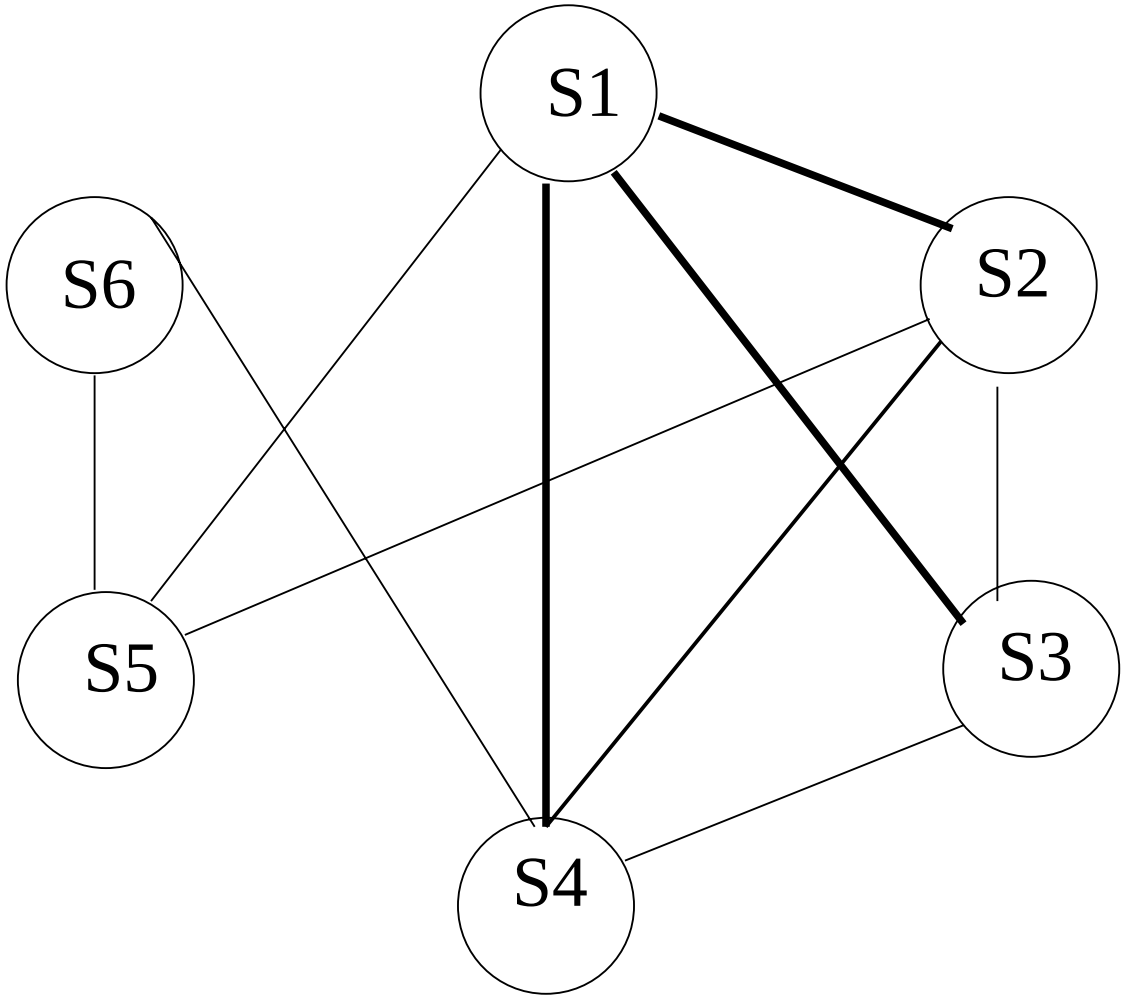
Four score seven years ago our fathers brought forth upon this continent a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. The brave men, living and dead, who struggled here, have consecrated it far above our poor power to add or detract.

The speech by Abraham Lincoln commemorates soldiers who laid down their lives in the Battle of Gettysburg. It reminds the troops that it is the future of freedom in America that they are fighting for.

Condensation genre

- headlines
- outlines
- minutes
- biographies
- abridgments
- sound bites
- movie summaries
- chronologies

Text as a Graph



Centrality-based Summarization(Radev)

- Assumption: The centrality of the node is an indication of its importance
- Representation: Connectivity matrix based on intra-sentence cosine similarity
- Extraction mechanism:
  - Compute PageRank score for every sentence u

$$PageRank(u) = \frac{(1-d)}{N} + d \sum_{v \in adj[u]} \frac{PageRank(v)}{deg(v)},$$

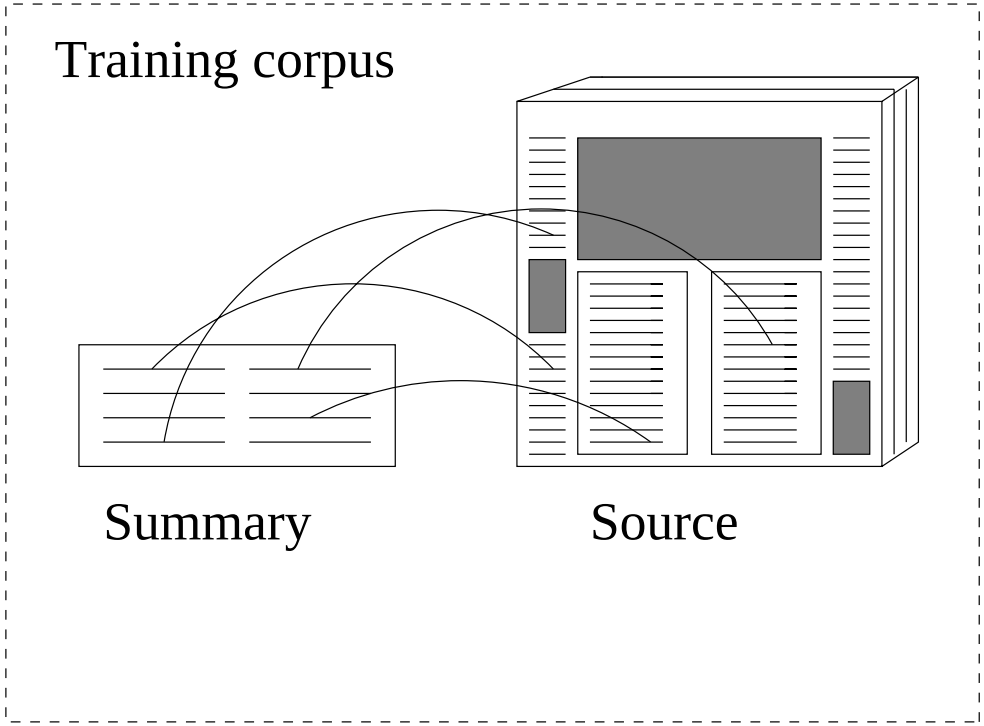
    where N is the number of nodes in the graph
  - Extract k sentences with the highest PageRanks score

Does it work?

- Evaluation: Comparison with human created summary
- Rouge Measure: Weighted n-gram overlap (similar to Bleu)

Method	**Rouge score**
Random	0.3261
Lead	0.3575
Degree	0.3595
PageRank	0.3666

Summarization as sentence extraction



- Given a corpus of documents and their summaries
- Label each sentence in the document as summary-worthy or not
- Learn which sentences are likely to be included in a summary
- Given an unseen (test document) classify sentences as summary-worthy or not

Summarization as sentence extraction

Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.

Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battle-field of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.

But, in a larger sense, we can not dedicate — we can not consecrate — we can not hallow — this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced.

red: not in summary, blue: in summary

Summarization as sentence extraction

- During training assign each sentence a score of importance/"extractworthiness".
- During testing extract sentences with highest score verbatim as extract.
- But how do we compute this score?
  - First match summary sentences against your document.
  - Then reduce sentences into important features.
  - Each sentence is represented as a vector of these features.

Summarization as sentence extraction

Sentence Length Cut-off Feature	true if sentence $>$ 5 words
Fixed-Phrase Feature	true if sentence contains
indicator	phrases: *this letter*,
	in conclusion
Paragraph Feature	initial, final, medial
Thematic Word Feature	true if sentence contains
	frequent words
UppercaseWord Feature	true if sentence contains proper
	names: *the American Society*
	for Testing and Materials

Summarization as sentence extraction

Training Data	Test Data
⇓	⇓
[1,0,INITIAL,1,0]	[0,0,MEDIAL,0,0]
[0,0,INITIAL,1,1]	[0,0,INITIAL,1,1]
[1,1,MEDIAL,0,0]	??
[1,1,MEDIAL,1,1]	
[0,0,MEDIAL,0,0]	
[1,1,INITIAL,1,1]	
[0,0,INITIAL,1,1]	

red: not in summary, blue: in summary

Combination of sentential features

Kupiec, Pedersen, Chen: A trainable document summariser, SIGIR 1995

$$P(s \in S|F_1, \ldots, F_k) = \frac{P(F_1,\ldots,F_k|s\in S)P(s\in S)}{P(F_1,\ldots,F_k)}$$

$$\approx \frac{P(s\in S)\prod_{j=1}^{k} P(F_j|s\in S)}{\prod_{j=1}^{k} P(F_j)}$$

$P(s \in S|F_1, \ldots, F_k)$**:** Probability that s from source text is in summary S, given feature values

$P(s \in S)$**:** Probability that s from source text is in summary S unconditionally

$P(F_j|\ s \in S)$**:** probability of feature-value pair occurring in sentence which is in the summary

$P(F_j)$**:** probability that feature-value pair F_j occurs unconditionally

Evaluation

- Corpus of 85 articles in 21 journals
- Baseline: select sentences from the beginning of a document
- Very high compression makes this task harder

Feature	Individual Sents Correct	Cumulative Sents Correct
Paragraph	163 (33%)	163 (33%)
Fixed Phrases	145 (29%)	209 (42%)
Length Cut-off	121 (24%)	217 (44%)
ThematicWord	101 (20%)	209 (42%)
Baseline	24%	

Content Models

Content models represent topics and their ordering in a domain text

Domain: newspaper articles on earthquakes

Topics: “strength,” “location,” “casualties,” . . .

Order: “casualties” prior to “rescue efforts”

Learning Content Structure

- Our goal: learn content structure from un-annotated texts via analysis of word distribution patterns

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

- The success of the distributional approach depends on the existence of recurrent patterns.
  - Linguistics: domain-specific texts tend to exhibit high similarity (Wray, 2002)
  - Cognitive psychology: formulaic text structure facilitates readers' comprehension(Bartlett, 1932)

Patterns in Content Organization

TOKYO (AP) A moderately strong earthquake rattled northern Japan early Wednesday, the Central Meteorological Agency said. There were no immediate reports of casualties or damage. The quake struck at 6:06 am (2106 GMT) 60 kilometers (36 miles) beneath the Pacific Ocean near the northern tip of the main island of Honshu. . . .

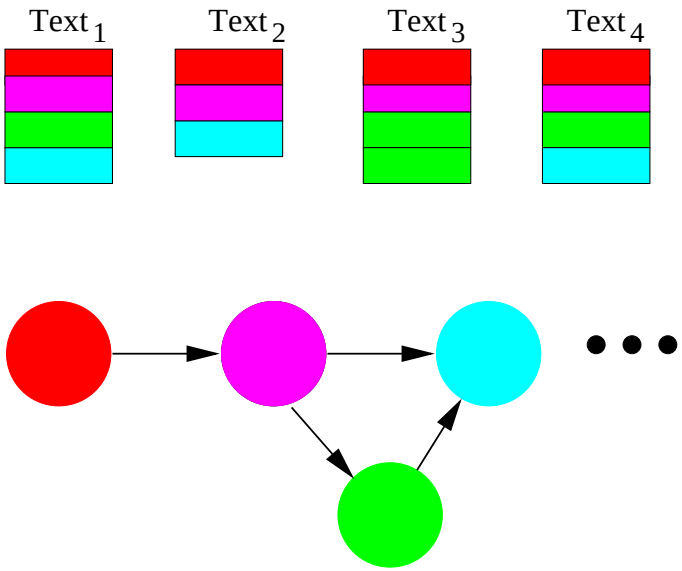
ATHENS, Greece (AP) A strong earthquake shook the Aegean Sea island of Crete on Sunday but caused no injuries or damage. The quake had a preliminary magnitude of 5.2 and occurred at 5:28 am (0328 GMT) on the sea floor 70 kilometers (44 miles) south of the Cretan port of Chania. . . .

Patterns in Content Organization

TOKYO (AP) A moderately strong earthquake rattled northern Japan early Wednesday, the Central Meteorological Agency said. There were no immediate reports of casualties or damage. The quake struck at 6:06 am (2106 GMT) 60 kilometers (36 miles) beneath the Pacific Ocean near the northern tip of the main island of Honshu. . . .

ATHENS, Greece (AP) A strong earthquake shook the Aegean Sea island of Crete on Sunday but caused no injuries or damage. The quake had a preliminary magnitude of 5.2 and occurred at 5:28 am (0328 GMT) on the sea floor 70 kilometers (44 miles) south of the Cretan port of Chania. . . .

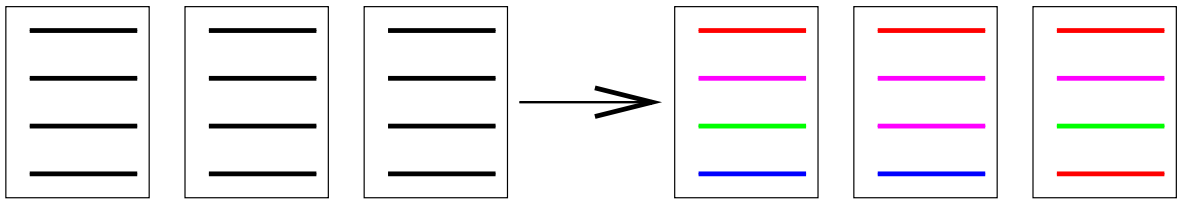
Computing Content Model



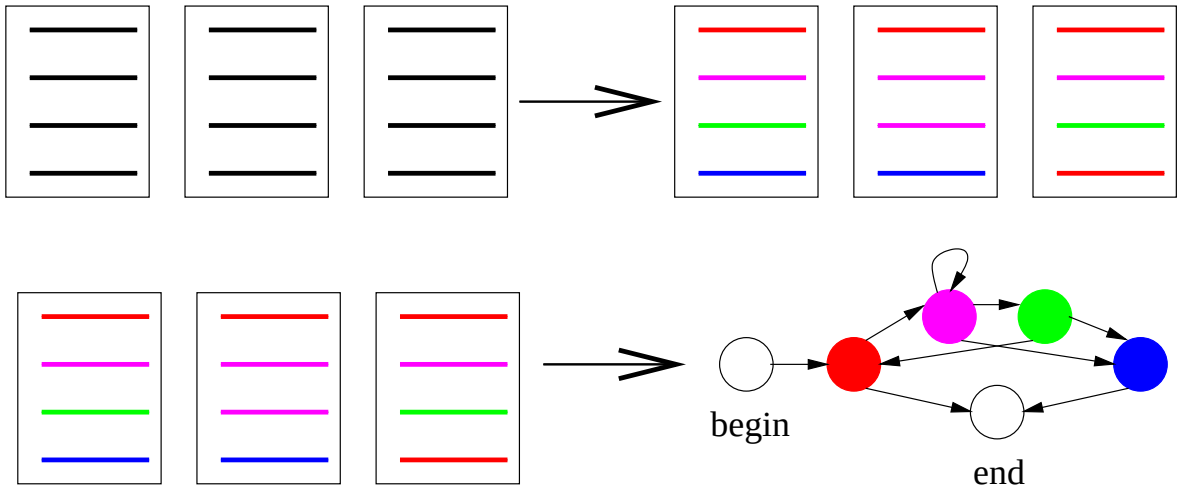
Implementation: Hidden Markov Model

- States represent topics
- State-transitions represent ordering constraints

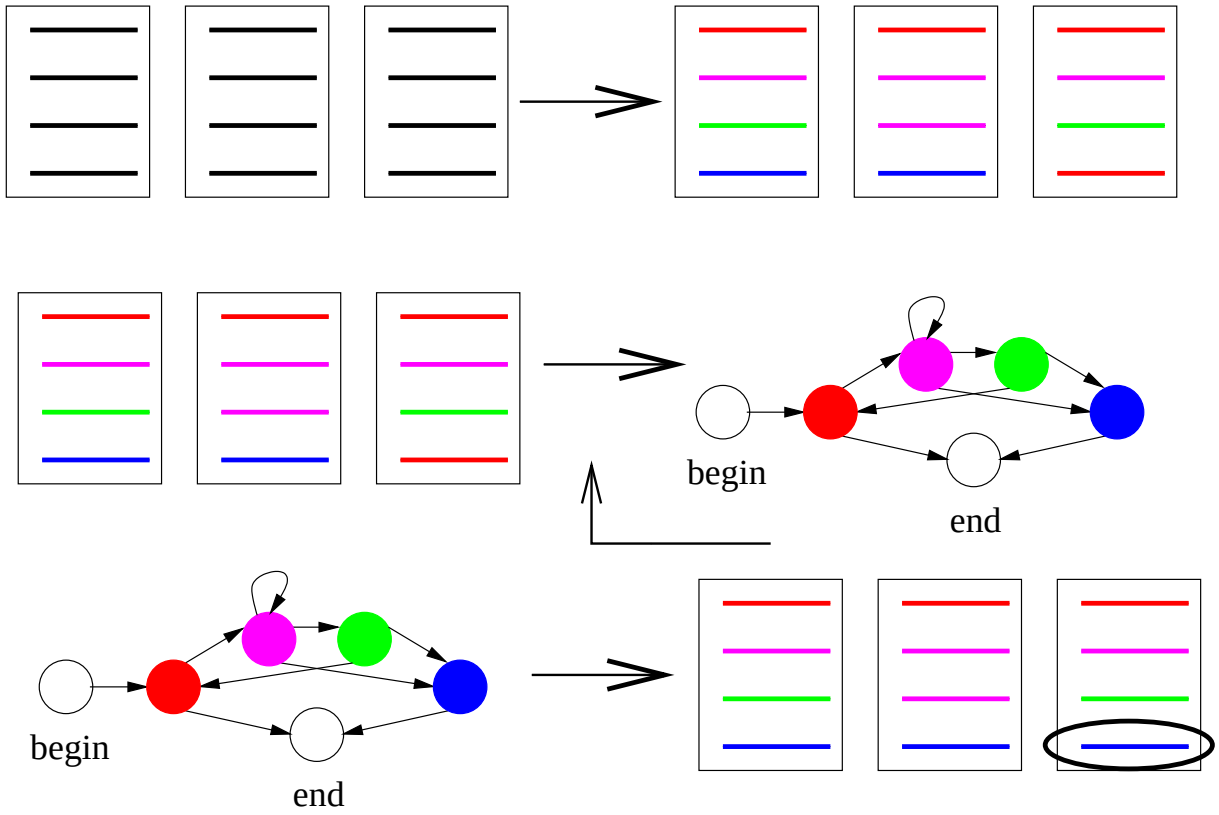
Model Induction



Model Induction



Model Induction



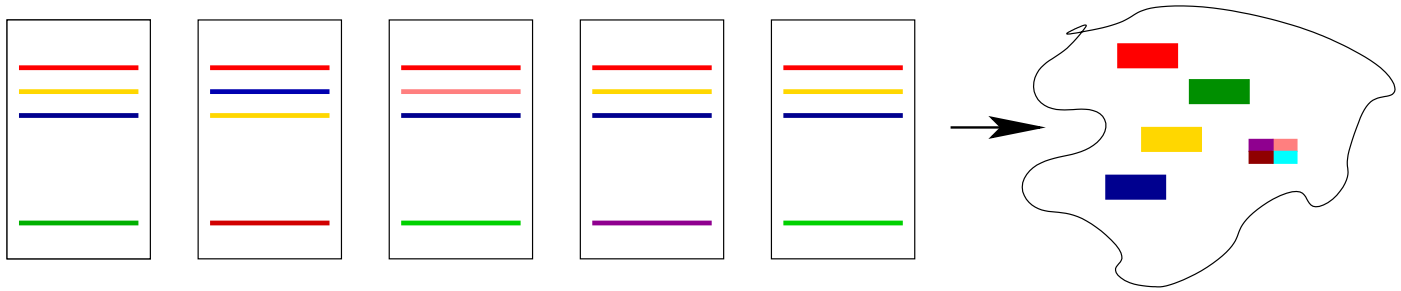
Initial Topic Induction

Agglomerative clustering with cosine similarity measure

The Athens seismological institute said the temblor's epicenter was located 380 kilometers (238 miles) south of the capital.
Seismologists in Pakistan's Northwest Frontier Province said the temblor's epicenter was about 250 kilometers (155 miles) north of the provincial capital Peshawar.
The temblor was centered 60 kilometers (35 miles) northwest of the provincial capital of Kunming, about 2,200 kilometers (1,300 miles) southwest of Beijing, a bureau seismologist said.

From Clusters to States

- Each large cluster constitutes a state
- Agglomerate small clusters into an *"insert"* state



Estimating Emission Probabilities

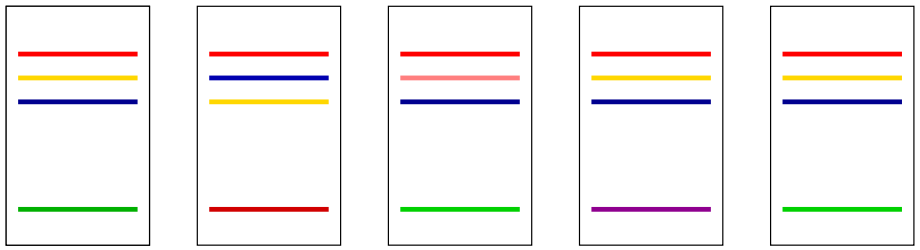
- Estimation for a "normal" state:

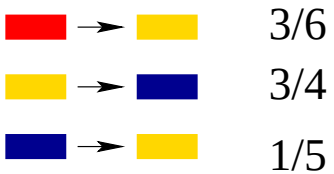
$$p_{s_i}(w'|w) \stackrel{def}{=} \frac{f_{c_i}(ww') + \delta_1}{f_{c_i}(w) + \delta_1|V|},$$

- Estimation for the "insertion" state:

$$p_{s_m}(w'|w) \stackrel{def}{=} \frac{1 - \max_{i<m} p_{s_i}(w'|w)}{\sum_{u \in V}(1 - \max_{i<m} p_{s_i}(u|w))}.$$

Estimating Transition Probabilities





$$p(s_j|s_i) = \frac{g(c_i, c_j) + \delta_2}{g(c_i) + \delta_2 m}$$

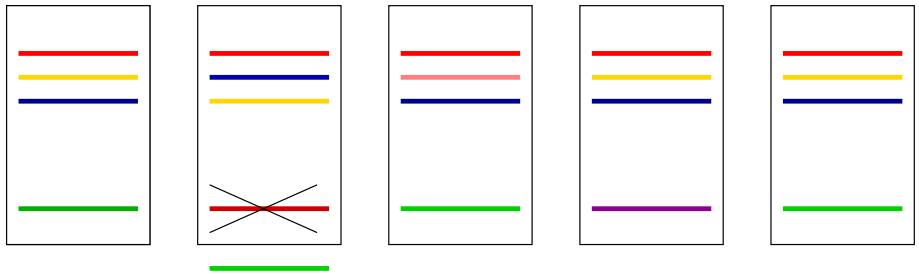
$g(c_i, c_j)$ is a number of adjacent sentences (c_i, c_j)

$g(c_i)$ is a number of sentences in c_i

Viterbi re-estimation

Goal: incorporate ordering information

- Decode the training data with Viterbi decoding



- Use the new clustering as the input to the parameter estimation procedure

Applications of Content Models

- Information ordering
- Summarization

Information Ordering

- Motivation: summarization, natural language generation, question-answering

Figures removed for copyright reasons.

- Evaluation: select the original order across n permutations of text sentences

Application: 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.

Application: 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.

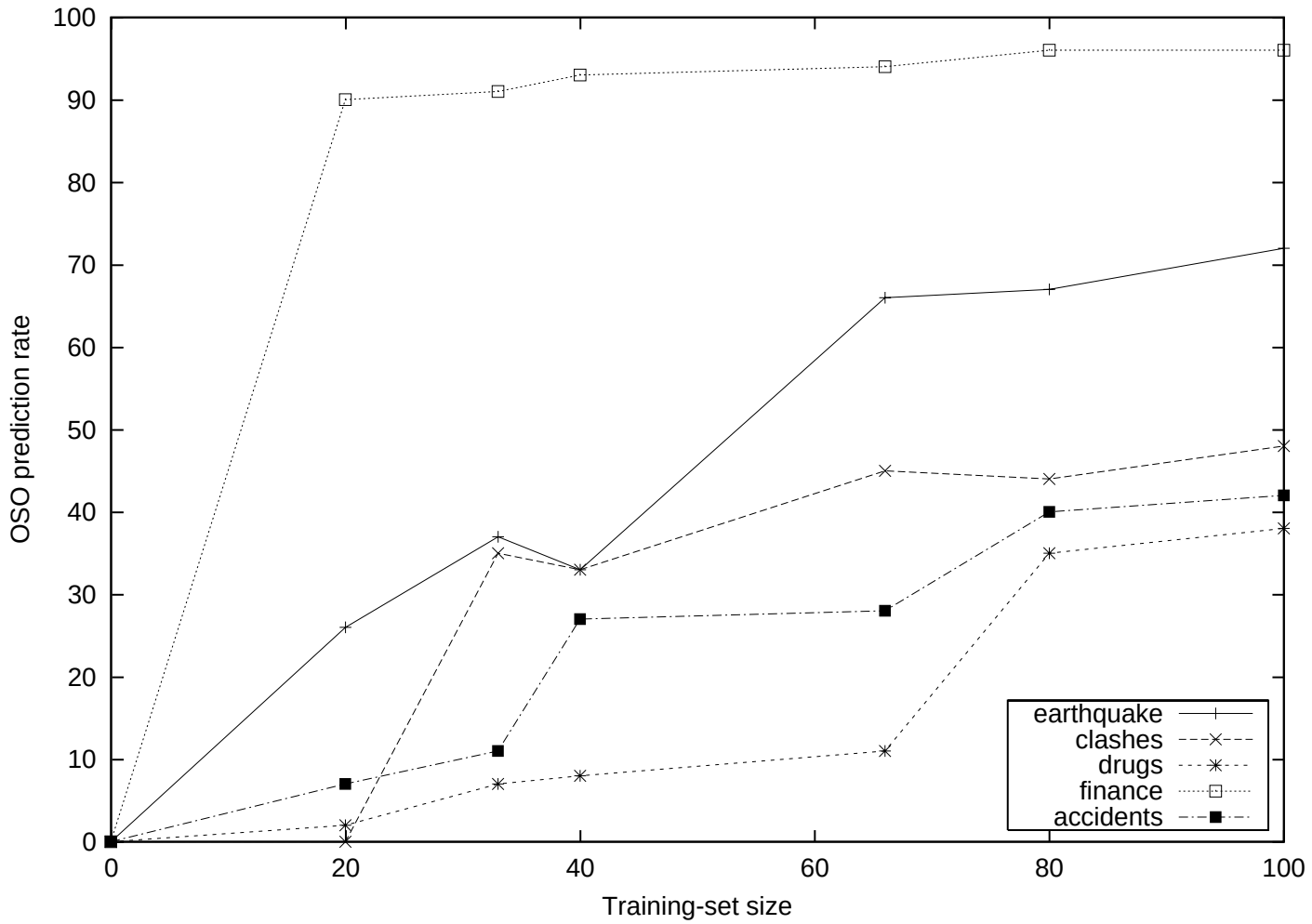
(a) During a third practice forced landing, with the landing gear extended, the CFI took over the controls.

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

Results: Ordering

Algorithm	Earthquake	Clashes	Drugs	Finance	Accidents
COntent Model	72	48	38	96	41

Learning Curves for Ordering



Summarization Task

MEXICO CITY (AP) A strong earthquake shook central Mexico Saturday, sending panicked tourists running from an airport terminal and shaking buildings in the capital.
There were no immediate reports of serious injuries.
The quake had a preliminary magnitude of 6.3 and it epicenter was in Guerrero state, 290 kilometers (165 miles) southwest of Mexico City, said Russ Needham of the U.S. Geological Survey's Earthquake Information Center in Golden, Colo.
Part of the roof of an airport terminal in the beach resort of Zahuatanejo collapsed and its windows shattered, sending scores of tourists running outside.
Power and telephone service were briefly interrupted in the town, about 340 kilometers (200 miles) southwest of Mexico City.
A fence was toppled in a poor neighborhood in Zihuatanejo.
The Red Cross said at least 10 people suffered from nervous disorders caused by the quake.
The quake started around 10:20 am and was felt for more than a minute in Mexico City, a metropolis of about 21 million people.
Buildings along Reforma Avenue, the main east-west thoroughfare, shook wildly.
"I was so scared.
Everything just began shaking," said Sonia Arizpe, a Mexico City street vendor whose aluminum cart started rolling away during the temblor.
But Francisco Lopez, a visiting Los Angeles businessman, said it could have been much worse.
"I've been through plenty of quakes in L.A. and this was no big deal."
The quake briefly knocked out electricity to some areas of the capital.
Windows cracked and broke in some high-rise buildings, and fire trucks cruised the streets in search of possible gas leaks.
Large sections of downtown Mexico City were devastated by a 8.1 magnitude quake in 1985.
At least 9,500 people were killed.

Summarization Task

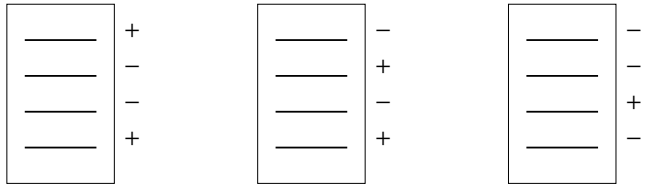
MEXICO CITY (AP) A strong earthquake shook central Mexico Saturday, sending panicked tourists running from an airport terminal and shaking buildings in the capital.
There were no immediate reports of serious injuries.
The quake had a preliminary magnitude of 6.3 and it epicenter was in Guerrero state, 290 kilometers (165 miles) southwest of Mexico City, said Russ Needham of the U.S. Geological Survey's Earthquake Information Center in Golden, Colo.
Part of the roof of an airport terminal in the beach resort of Zahuatanejo collapsed and its windows shattered, sending scores of tourists running outside.
Power and telephone service were briefly interrupted in the town, about 340 kilometers (200 miles) southwest of Mexico City.
A fence was toppled in a poor neighborhood in Zihuatanejo.
The Red Cross said at least 10 people suffered from nervous disorders caused by the quake.
The quake started around 10:20 am and was felt for more than a minute in Mexico City, a metropolis of about 21 million people.
Buildings along Reforma Avenue, the main east-west thoroughfare, shook wildly.
"I was so scared.
Everything just began shaking," said Sonia Arizpe, a Mexico City street vendor whose aluminum cart started rolling away during the temblor.
But Francisco Lopez, a visiting Los Angeles businessman, said it could have been much worse.
"I've been through plenty of quakes in L.A. and this was no big deal."
The quake briefly knocked out electricity to some areas of the capital.
Windows cracked and broke in some high-rise buildings, and fire trucks cruised the streets in search of possible gas leaks.
Large sections of downtown Mexico City were devastated by a 8.1 magnitude quake in 1985.
At least 9,500 people were killed.

Summarization Task

The quake started around 10:20 am and was felt for more than a minute in Mexico City, a metropolis of about 21 million people. There were no immediate reports of serious injuries. Buildings along Reforma Avenue, the main east-west thoroughfare, shook wildly.

Summarization: Algorithm

Supervised learning approach

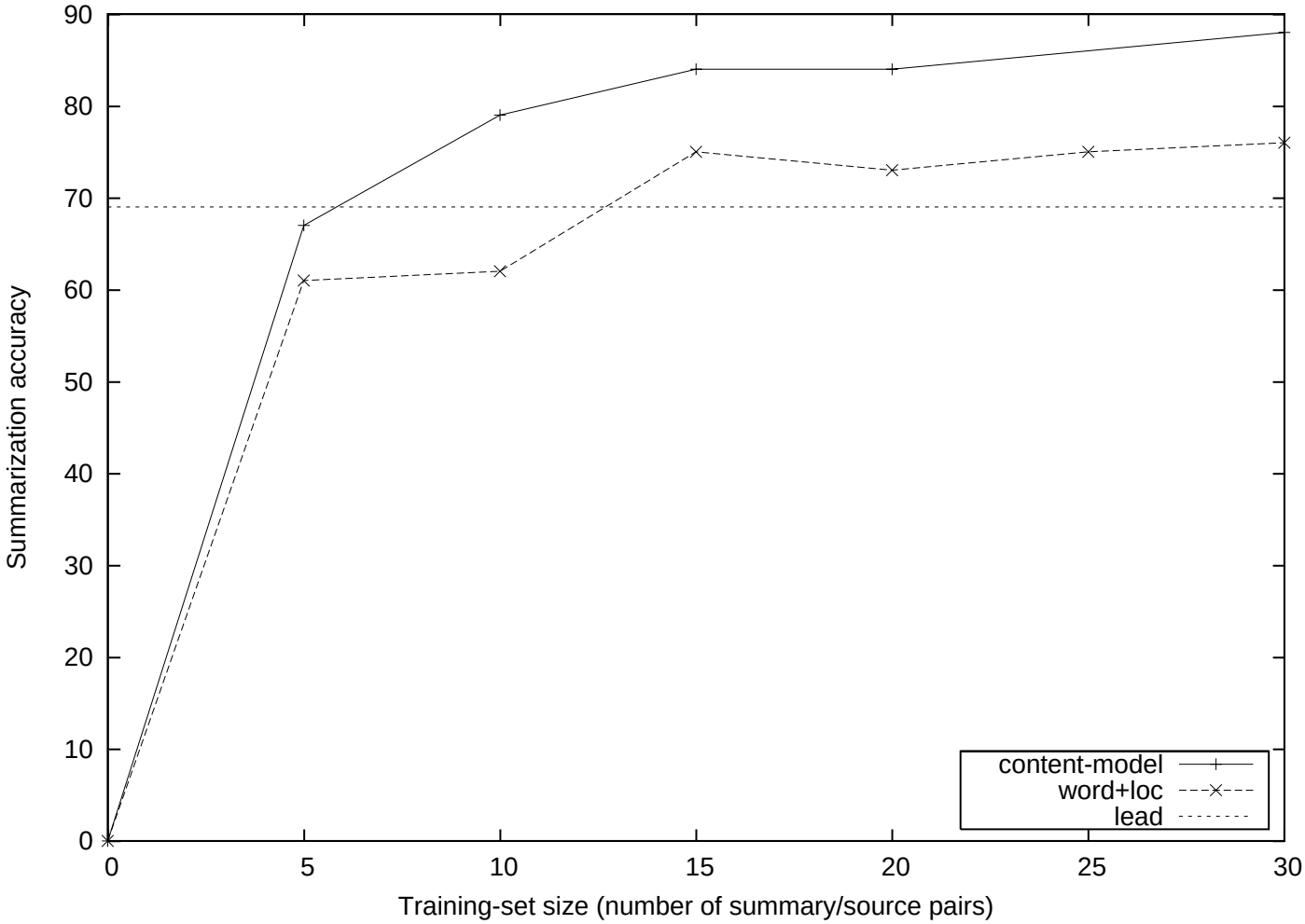


- Traditional (local) approach: look at lexical features
- Our approach: look at structural features

Results: Summarization

Summarizer	Extraction accuracy
Content-based	**88%**
Sentence classifier (words + location)	76%

Learning Curves for Summarization



Sentence Compression

- Sentence compression can be viewed as producing a summary of a single sentence
- A compressed sentence should:
  - Use less words than the original sentence
  - Preserve the most important information
  - Remain grammatical

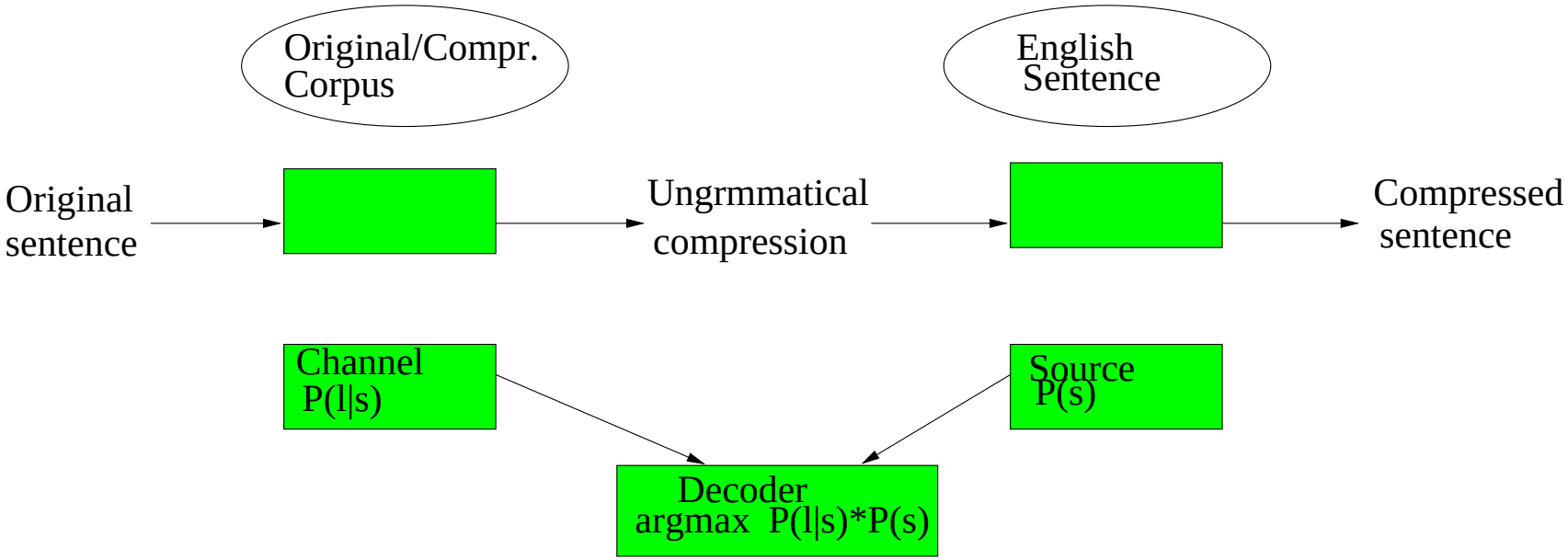
Sentence Compression

- Sentence compression can involve:
  - word deletion
  - word reordering
  - word substitution
  - word insertion
- Simplified formulation: given an input sentence of words $w_1, \ldots, w_n$ a compression is formed by dropping any subset of these words

Sentence Compression: Example

Prime Minister Tony Blair insisted the case for holding terrorism suspects without trial was “absolutely compelling” as the government published new legislation allowing detention for 90 days without charge.
Tony Blair insisted the case for holding terrorism suspects without trial was “compelling”

Noisy-Channel Model



Sentence Compression

- Source Model: A good compression is one that looks grammatical (bigram score) and has a normal looking parse tree (PCFG score). Scores are estimated from WSJ and Penn Treebank.
- Channel Model: Responsible for preserving important information. Estimated from a parallel corpus of original/compressed sentence pairs.
- Decoder: Uses a packed forest representation and tree extractor

Output Examples

Beyond the basic level, the operations of the three products vary widely.

The operations of the three products very widely.

Arborscan is reliable and worked accurately in testing, but it produces very large dxf files.

Arborscan is reliable and worked accurately in testing very large dxf files.

Evaluation Results

Baseline: compression with highest word-bigram score

	Baseline	Noisy-channel	Humans
Compression	63.7%	70.37%	53.33%
Grammaticality	1.78%	4.34%	4.92%
Importance	2.17%	3.54%	4.24%