Computational Models of Discourse

Regina Barzilay
EECS Department
MIT

November 2, 2004
What is Text?

A product of cohesive ties (cohesion)

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. The Athens seismological institute said the temblor’s epicenter was located 380 kilometers (238 miles) south of the capital. No injuries or damage were reported.
A product of structural relations (coherence)

$S_1$: A strong earthquake shook the Aegean Sea island of Crete on Sunday

$S_2$: but caused no injuries or damage.

$S_3$: The quake had a preliminary magnitude of 5.2
Content-based Structure

• Describe the strength and the impact of an earthquake
• Specify its magnitude
• Specify its location
• ...

Computational Models of Discourse
Rhetorical Structure

- Elaboration
- Contrast

S_1 \quad S_2 \quad S_3
Analogy with Syntax

Domain-independent Theory of Sentence Structure

- Fixed set of word categories (nouns, verbs, ...)
- Fixed set of relations (subject, object, ...)

P("A is sentence this weird")
Two Approaches to Text Structure

• Domain-dependent models
  – Content-based models
  – Rhetorical models

• Domain-independent models
  – Rhetorical Structure Theory
Motivation

- Summarization
  Extract a representative subsequence from a set of sentences

- Question-Answering
  Find an answer to a question in natural language

- Text Ordering
  Order a set of information-bearing items into a coherent text

- Machine Translation
  Find the best translation taking context into account
Today: Domain-Specific Models

• Rhetorical Models:
  – Argumentative Zoning of Scientific Articles (Teufel, 1999)

• Content-based Models:
  – Supervised (Duboue & McKeown, 2001)
  – Unsupervised (Barzilay & Lee, 2004)
Many of the recent advances in Question Answering have followed from the insight that systems can benefit from by exploiting the redundancy in large corpora.

Brill et al. (2001) describe using the vast amount of data available on the WWW to achieve impressive performance . . .

The Web, while nearly infinite in content, is not a complete repository of useful information . . .

In order to combat these inadequacies, we propose a strategy in which in information is extracted from . . .
Argumentative Zoning

BACKGROUND
Many of the recent advances in Question Answering have followed from the insight that systems can benefit from by exploiting the redundancy ...

OTHER WORK
Brill et al. (2001) describe using the vast amount of data available on the WWW to achieve impressive performance ...

WEAKNESS
The Web, while nearly infinite in content, is not a complete repository of useful information ...

OWN CONTRIBUTION
In order to combat these inadequacies, we propose a strategy in which in information is extracted from ...
Motivation

- Scientific articles exhibit (consistent across domains) similarity in structure
  - BACKGROUND
  - OWN CONTRIBUTION
  - RELATION TO OTHER WORK

- Automatic structure analysis can benefit:
  - Q&A
  - summarization
  - citation analysis
Approach

- Goal: Rhetorical segmentation with labeling

- Annotation Scheme:
  - Own work: aim, own, textual
  - Background
  - Other Work: contrast, basis, other

- Implementation: Classification
Examples

<table>
<thead>
<tr>
<th>Category</th>
<th>Realization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aim</strong></td>
<td>We have proposed a method of clustering words based on large corpus data</td>
</tr>
<tr>
<td><strong>Textual</strong></td>
<td>Section 2 describes three parsers which are ...</td>
</tr>
<tr>
<td><strong>Contrast</strong></td>
<td>However, no method for extracting the relationship from superficial linguistic expressions was described in their paper.</td>
</tr>
</tbody>
</table>
Kappa Statistics

(Siegal&Castellan, 1998; Carletta, 1999)
Kappa controls agreement $P(A)$ for chance agreement $P(E)$

$$K = \frac{P(A) - p(E)}{1 - p(E)}$$

Kappa from Argumentative Zoning:

- Stability: 0.83
- Reproducibility: 0.79
Features

- Position
- Verb Tense and Voice
- History
- Lexical Features ("other researchers claim that")
Results

- Classification accuracy is above 70%
- Zoning improves summarization
Supervised Content Modeling

(Duboue & McKeown, 2001)

- **Goal**: Find types of semantic information characteristic to a domain and ordering constraints on their presentation

- **Approach**: find patterns in a set of transcripts manually annotated with semantic units

- **Domain**: Patients records
He is a **58-year-old male**. History is significant for **Hodgkin’s disease**, treated with ... to his neck, back and chest. **Hyperspadias**, **BPH**, **hiatal hernia** and **proliferative lymph edema** in his right arm. No IV’s or blood pressure down in the left arm. Medications — **Inderal**, **Lopid**, **Pepcid**, **nitroglycerine** and **heparin**. **EKG** has PAC’s. ...
Semantic Sequence

age, gender, pmh, pmh, pmh, pmh, med-preop, med-preop, med-preop, med-preop, drip-preop, med-preop, ekg-preop, echo-preop, hct-preop, procedure, ...
Pattern Detection

Analogous to motif detection

$T_1: \underline{A} \underline{B} \underline{C} \underline{D} \ F \ A \ A \ B \ F \ D$

$T_2: F \ C \ A \ B \ D \ D \ F \ F$

- Scanning
- Generalizing
- Filtering
### Example of Learned Pattern

<table>
<thead>
<tr>
<th>intraop-problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>operation</td>
</tr>
<tr>
<td>drip</td>
</tr>
<tr>
<td>intraop-problems</td>
</tr>
<tr>
<td>total-meds-anesthetics</td>
</tr>
</tbody>
</table>

*Computational Models of Discourse*
Evaluation

Pattern confidence: 84.62%

Constraint accuracy: 89.45%
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)
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. . .
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. . . .
Implementation: Hidden Markov Model

- States represent topics
- State-transitions represent ordering constraints
Model Induction

Computational Models of Discourse
Model Induction

Computational Models of Discourse
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:

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

- Estimation for the “insertion” state:

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

- Five automatically constructed domain-specific collections
- For each domain 100 training, 100 testing, 20 development

<table>
<thead>
<tr>
<th>Domain</th>
<th>Average Length</th>
<th>Vocabulary Size</th>
<th>Token/type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earthquake</td>
<td>10.4</td>
<td>1182</td>
<td>13.158</td>
</tr>
<tr>
<td>Clashes</td>
<td>14</td>
<td>1302</td>
<td>4.464</td>
</tr>
<tr>
<td>Drugs</td>
<td>10.3</td>
<td>1566</td>
<td>4.098</td>
</tr>
<tr>
<td>Finance</td>
<td>13.7</td>
<td>1378</td>
<td>12.821</td>
</tr>
<tr>
<td>Accidents</td>
<td>11.5</td>
<td>2003</td>
<td>5.556</td>
</tr>
</tbody>
</table>

*Computational Models of Discourse*
Applications of Content Models

- Information ordering
- Summarization
Information Ordering

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

- 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.
Evaluation: Ordering

• Baselines:
  – “Straw” baseline: Bigram Language model
    * represent a sentence using lexico-syntactic features
    * compute pairwise ordering preferences
    * find optimally global order

• Metrics: Prediction Accuracy, Rank, $\tau$
## Results: Ordering

<table>
<thead>
<tr>
<th>Domain</th>
<th>Algorithm</th>
<th>Prediction Accuracy</th>
<th>Rank</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earthquake</td>
<td>Content</td>
<td>72%</td>
<td>2.67</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Lapata '03</td>
<td>24%</td>
<td>(N/A)</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Bigram</td>
<td>4%</td>
<td>485.16</td>
<td>0.27</td>
</tr>
<tr>
<td>Clashes</td>
<td>Content</td>
<td>48%</td>
<td>3.05</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>Lapata '03</td>
<td>27%</td>
<td>(N/A)</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>Bigram</td>
<td>12%</td>
<td>635.15</td>
<td>0.25</td>
</tr>
<tr>
<td>Drugs</td>
<td>Content</td>
<td>38%</td>
<td>15.38</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>Lapata '03</td>
<td>27%</td>
<td>(N/A)</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Bigram</td>
<td>11%</td>
<td>712.03</td>
<td>0.24</td>
</tr>
<tr>
<td>Finance</td>
<td>Content</td>
<td>96%</td>
<td>0.05</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>Lapata '03</td>
<td>17%</td>
<td>(N/A)</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>Bigram</td>
<td>66%</td>
<td>7.44</td>
<td>0.74</td>
</tr>
<tr>
<td>Accidents</td>
<td>Content</td>
<td>41%</td>
<td>10.96</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>Lapata '03</td>
<td>10%</td>
<td>(N/A)</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Bigram</td>
<td>2%</td>
<td>973.75</td>
<td>0.19</td>
</tr>
</tbody>
</table>
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 its 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.
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 Zihuataneco. 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.
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

<table>
<thead>
<tr>
<th>Summarizer</th>
<th>Extraction accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content-based</td>
<td>88%</td>
</tr>
<tr>
<td>Sentence classifier (words + location)</td>
<td>76%</td>
</tr>
</tbody>
</table>
Here is a right order

(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.

Computational Models of Discourse 48/49
Learning Curves for Summarization

- Content-model
- Word+loc
- Lead

Computational Models of Discourse