Probabilistic Approaches for Modeling Text Structure and their Application to Text-to-Text Generation

Regina Barzilay
(MIT)
joint work with Branavan, Harr Chen, Mirella Lapata, Lillian Lee, Christina Sauper
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.
SCIgen: An Automatic CS Paper Generator

- An output of a system that automatically generates scientific papers (Stribling et al., 2005):

  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.

- The paper was accepted to a conference (not ACL!)
Less Exotic Example

• An output of a state-of-the-art multidocument summarization system

Newspapers reported Wednesday that three top Libyan officials have been tried and jailed in the Lockerbie case. The sanctions were imposed to force Libyan leader Moammar Gadhafi to turn the men over. *Louis Farrakhan congratulated Gadhafi on his recovery from a hip injury.*

• A system does not have any means to determine that it generates an incoherent text
Macular degeneration is a condition affecting the tissues lying under the retina, while a macular hole involves damage from within the eye, at the junction between the vitreous and the retina itself. There is no relationship between the two diseases. Depending upon the degree of attachment or traction between the vitreous and the retina, there may be risk of developing a macular hole in the retina itself. There is no relationship upon the degree of attachment or traction between the vitreous and the retina itself. Traction between retina, there may be risk of developing a macular hole in the retina itself.
Macular degeneration is a condition affecting the tissues lying under the retina, while a macular hole involves damage from within the eye, at the junction between the vitreous and the retina itself. There is no relationship between the two diseases. Depending upon the degree of attachment or traction between the vitreous and the retina, there may be risk of developing a macular hole in the retina itself. There is no relationship upon the degree of attachment or traction between the vitreous and the retina itself.
Linguistic Foundations: Discourse Theory

- Centering Theory (Grosz, Joshi & Weinstein, 1983)
- Rhetorical Structure Theory (Mann & Thompson, 1988)
- Content Schema (Bartlett, 1932)
- ...

Current Solutions

• **Work hard:** manually encode all the constraints
  – Expensive
  – Not scalable

• **Ignore coherence**, and hope for the best
  – Incoherent output

**Our goal:** linguistically-motivated text models without manual crafting
The Goal

Our goal: linguistically-motivated text models without manual crafting

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

• Application of text models
• Content Models (NAACL 2004, NAACL 2009)
• Coherence Models (ACL 2005, CL 2008)
Generating Coherent Text (1)

- Model assesses the likelihood that the input is well-formed
- It ranks candidate outputs based on the model scores
- Analogous to the use of language model in ASR and MT
• Specifies patterns in topic occurrence and their ordering
• Each topic is represented as a distribution over words
Generating Wikipedia Articles

• Induce a content model from existing Wikipedia articles in the target domain:
  – Query for Internet search (e.g., “Macular Hole” treatment)
  – Extractor of relevant content (trained on existing articles)
Generating Wikipedia Articles (2)

- Retrieve web pages using specified queries

Search: "macular hole" causes

Macular Hole - Causes, Symptoms & Treatment
Macular Hole Facts plus the Latest News on Macular Hole Treatments - HealthNewsflash ... A macular hole is just that: a hole in the macula. What causes a ... [http://healthnewsflash.com/conditions/macular_hole.htm](http://healthnewsflash.com/conditions/macular_hole.htm) - Cached

Facts About Macular Hole [NEI Health Information]
Publication about macular hole, a small break in the macula, which is located in the center of the eye's retina. ... What causes a macular hole? ... [http://www.nei.nih.gov/health/macularhole](http://www.nei.nih.gov/health/macularhole) - Cached

- Jointly select and order excerpts

What causes a macular hole?
As we age, the vitreous slowly shrinks and pulls away from the retinal surface. If the vitreous is firmly attached to the retina when it pulls away, it can tear the retina and create a macular hole. Natural fluids fill the area where the vitreous has contracted.

Detection and Diagnosis
The severity of the symptoms is dependent on whether the hole is partial or full-thickness. Visual acuity testing, Amsler grid, and ophthalmoscopy are all performed to evaluate the macula's health and function.

What are the treatment options?
Surgery is the only recommended treatment for a macular hole. Early diagnosis and treatment are absolutely essential to help reverse some of the vision loss caused by a macular hole.
Macular hole

From Wikipedia, the free encyclopedia

A macular hole is a small break in the macula, located in the center of the eye’s light-sensitive tissue called the retina.

Contents
1 Diagnosis
2 Causes
3 Symptoms
4 References
5 External links

Diagnosis

Macular degeneration is a condition affecting the tissues lying under the retina, while a macular hole involves damage from within the eye, at the junction between the vitreous and the retina itself. There is no relationship between the two diseases. Depending upon the degree of attachment or traction between the vitreous and the retina, there may be risk of developing a macular hole in the other eye. Your eye care provider can determine the status of the vitreous jelly and its degree of traction on the retinal surface in the uninvolved eye. In those cases where the vitreous has already become separated from the retinal surface, there is very little chance of developing a macular hole in the other eye. On the other hand, when the vitreous remains adherent and pulling on the macular region in both eyes, then there may be a greater risk of developing a hole in the second eye. In very rare instances, trauma or other conditions lead to the development of a macular hole. In the vast majority of cases, however, macular holes develop spontaneously. As a result, there is no known way to prevent their development through any nutritional or chemical means, nor is there any way to know who is at risk for developing a hole prior to its appearance in one or both eyes. [1]

Causes

The eye contains a jelly-like substance called the vitreous. Shrinkage of the vitreous usually causes the hole. As a person ages, the vitreous becomes thicker and stronger and begins to pull away from the retina. If the vitreous is firmly attached to the retina when it pulls away, a hole can result. [2]

Symptoms

However, if the vitreous is firmly attached to the retina when it pulls away, it can tear the retina and create a macular hole. Also, once the vitreous has pulled away from the surface of the retina, some of the fibers can remain on the retinal surface and can contract. This increases tension on the retina and can lead to a macular hole. In either case, the fluid that has replaced the shrunken vitreous can then seep through the hole onto the macula, blurring and distorting central vision. [3]
Readers’ Reaction
More Automatically-Generated Wikipedia Articles

http://en.wikipedia.org/wiki/Felty_syndrome
http://en.wikipedia.org/wiki/Cogan_syndrome
http://en.wikipedia.org/wiki/Ancylostomiasis
http://en.wikipedia.org/wiki/Amyoplasia
http://en.wikipedia.org/wiki/Hemorrhagic_cystitis
http://en.wikipedia.org/wiki/Heterotopic_ossification
http://en.wikipedia.org/wiki/Hypophosphatasia
http://en.wikipedia.org/wiki/Vestibular_neuronitis
http://en.wikipedia.org/wiki/3-M_syndrome
http://en.wikipedia.org/wiki/Macular_hole
Ordering Task 1

**Input:** $N$ alternative text realizations

**Task:** Find the most coherent alternative
Ordering Task 2

Input: $N$ information-bearing items

Task: Organize input items into a text based on their semantic relations
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.
Outline

• Applications of Text Models
• Coherence Models (ACL 2005, CL 2008)
• Content Models (NAACL 2004, NAACL 2009)
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

- 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.
Computing Content Model

Implementation: Hidden Markov Model

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

\[ p_{si}(w'|w) \overset{\text{def}}{=} \frac{f_{ci}(ww') + \delta_1}{f_{ci}(w) + \delta_1|V|}, \]

• Estimation for the “insertion” state:

\[ p_{sm}(w'|w) \overset{\text{def}}{=} \frac{1 - \max_{i<m} p_{si}(w'|w)}{\sum_{u \in V} (1 - \max_{i<m} p_{si}(u|w))}. \]
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: 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)
Evaluation Results: Ranking

Average number of sentences per document:  10

<table>
<thead>
<tr>
<th>Model</th>
<th>Disasters</th>
<th>NTSB Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Models (HMM)</td>
<td>88.0</td>
<td>75.8</td>
</tr>
<tr>
<td>Random</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>
### Evaluation Results: Ordering

Average number of sentences per document: 10

<table>
<thead>
<tr>
<th>Model</th>
<th>Disasters</th>
<th>NTSB Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Models (HMM)</td>
<td>81.0</td>
<td>44.0</td>
</tr>
<tr>
<td>Random</td>
<td>&lt;1.0</td>
<td>&lt;1.0</td>
</tr>
</tbody>
</table>
The Need for Global Structure

• **Challenge:** preserving topic contiguity
  – Topics do not repeat in disconnected sections of a document

• **Solution:** model topic structure as *permutation*
The Need for Global Structure

• *Challenge*: model shared structure across documents
Modeling Predominant Permutations

Concentrate probability mass around this predominant permutation
The Generalized Mallows Model over Permutations

- Probability distribution over permutations
- Biases toward low Kendall tau distance to canonical ordering
Learning Framework

• Training
  – *Given*: Unannotated collection of documents
  – *Goal*: Estimate model parameters
    • *Predominant ordering, topic language models, etc.*

• Testing
  – *Given*: Bag of text fragments
  – *Goal*: Find maximum likelihood ordering
Generative Process

Mallows Distribution
Generative Process

Mallows Distribution

Topic Permutation
Generative Process

Mallows Distribution

Topic Permutation

Topic Count Distribution
Generative Process

Mallows Distribution

Topic Permutation

Topic Count Distribution

Topic Counts
Generative Process

Mallows Distribution

Topic Permutation

Document Topic Assignments

Topic Count Distribution

Topic Counts
Generative Process

Mallows Distribution

Distribution Topic Counts

Generative Process

Topic Permutation

Document Topic Assignments

Topics Counts

Topic Count Distribution


Training Details

• Hidden variables:
  – Predominant ordering, Mallows parameters
  – Topic counts, language models

• Estimated from unannotated documents using Bayesian inference techniques
  – Collapsed Gibbs sampling

• Mallows distribution admits computationally tractable inference procedures
Predicting Document Ordering

• Find best topic for each paragraph independently
  – Computationally simple, because marginalization over all orderings is not necessary!
Predicting Document Ordering

• Find best topic for each paragraph independently
  – Computationally simple, because marginalization over all orderings is not necessary!
• Sort paragraphs according to topic number
  – Topic assignments determine best ordering
Evaluation Results: Ordering

Average number of sentences per document: 10

<table>
<thead>
<tr>
<th>Model</th>
<th>Disasters</th>
<th>NTSB Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Models (HMM)</td>
<td>81.0</td>
<td>44.0</td>
</tr>
<tr>
<td>Content Models (GMM)</td>
<td>83.0</td>
<td>57.0</td>
</tr>
<tr>
<td>Random</td>
<td>&lt;1.0</td>
<td>&lt;1.0</td>
</tr>
</tbody>
</table>

See NAACL’09 for evaluation on segmentation and semantic labeling
Outline

• Applications of Text Models
• Coherence Models (ACL 2005, CL 2008)
• Content Models (NAACL 2004, NAACL 2009)
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.
Our 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
Our Inspiration: 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
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
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]_s_, was arrested in [London]_x_ on [October 14th]_x_ 1998.

2. [Pinochet]_s_, 82, was recovering from [surgery]_x_.

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

4. [Pinochet]_s_ was charged with murdering [thousands]_o_ including many [Spaniards]_o_.

5. [He]_s_ is awaiting [a hearing]_o_ [his fate]_x_ in [the balance]_x_.

6. [American scholars]_s_ applauded the [arrest]_o_.

Notation: _s_=subjects, _o_=object, _x_=other
Input Text with Coreference Information

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

2. [Pinochet]$_s$, 82, was recovering from [surgery]$_x$.
3. [The arrest]$_s$ was in [response]$_x$ to [an extradition warrant]$_x$ served by [a Spanish judge]$_s$.
4. [Pinochet]$_s$ was charged with murdering [thousands]$_o$, including many [Spaniards]$_o$.
5. [He]$_s$ is awaiting [a hearing]$_o$, [his fate]$_x$ in [the balance]$_x$.
## Output Entity Grid

<table>
<thead>
<tr>
<th></th>
<th>Pinochet</th>
<th>London</th>
<th>October</th>
<th>Surgery</th>
<th>Arrest</th>
<th>Extradition</th>
<th>Warrant</th>
<th>Judge</th>
<th>Thousands</th>
<th>Spaniards</th>
<th>Hearing</th>
<th>Fate</th>
<th>Balance</th>
<th>Scholars</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>s</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>s</td>
<td>-</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>s</td>
<td>x</td>
<td>x</td>
<td>s</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>s</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>o</td>
<td>o</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>s</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>o</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>o</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>s</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6</td>
</tr>
</tbody>
</table>
## Comparing Grids

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>S</th>
<th>S</th>
<th>S</th>
<th>X</th>
<th>X</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>S</th>
<th>S</th>
<th>X</th>
<th>X</th>
<th>X</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 'S' represents a starting point.
- 'X' represents a target point.
- 'O' represents an obstacle.
Text Encoding as Feature Vector

| $d_{i1}$ | 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)
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)
## Evaluation Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Disasters</th>
<th>NTSB reports</th>
<th>Summary Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid</td>
<td>87.2</td>
<td>90.4</td>
<td>80.0</td>
</tr>
<tr>
<td>Content Model (HMM)</td>
<td>88.0</td>
<td>75.8</td>
<td>N/A</td>
</tr>
<tr>
<td>Random</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>
Follow-Ups

• Elsner, Austerweil, and Charniak (NAACL 2007): combining content and coherence models is effective
  – **Brown Coherence Toolkit**: software for coherence models and test applications

• Pitler and Nenkova (EMNLP 2008): grid-based features help to predict readability of human-authored texts
Conclusions

• Practical Benefits: The proposed models can be easily incorporated in text-to-text generation

• Main Findings: The proposed document models can be successfully applied to modeling text structure

• Future Research: Design statistical models that match representational power of traditional discourse theories
References


• Regina Barzilay, Lillian Lee "Catching the Drift: Probabilistic Content Models, with Applications to Generation and Summarization", NAACL-HLT, 2004