**WHY?**
- do we need to care about it?

**SENTENCE ORDERING**
- is it needed?
- to do it?

**WHERE?**
- are the criteria used?

**HOW?**

**WHAT?**
How to fry an egg?

Mr. Hungry

Why?

Question Answering System for cooking purposes
Put oil in the pan
Warm it
Wait 3-5 min.
Move the egg to a plate
Put oil in the pan
Enjoy!

Put oil in the pan
Warm it
Crack the egg into the pan
Wait 3-5 min.
Move the egg to a plate
Enjoy!

Mr. Hungry
Bob, who graduated with a PhD from UMICH, has just finished his job interview, but he isn’t sure if he did well.

- Bob has just finished his job interview
- Bob graduated with a PhD from UMICH
- But he isn’t sure if he did well.
Columbia Summarizer

Human Judges (1)

10 Judged to be poorly ordered

Reorder Manually

30 document sets
DUC 2001 Dataset

<table>
<thead>
<tr>
<th>Document Set</th>
<th>Original Summary</th>
<th>Reordered</th>
</tr>
</thead>
<tbody>
<tr>
<td>D13</td>
<td>Bad</td>
<td>Bad</td>
</tr>
<tr>
<td>D19</td>
<td>Fair</td>
<td>Good</td>
</tr>
<tr>
<td>D24</td>
<td>Bad</td>
<td>Good</td>
</tr>
<tr>
<td>D31</td>
<td>Fair</td>
<td>Good</td>
</tr>
<tr>
<td>D32</td>
<td>Bad</td>
<td>Fair</td>
</tr>
<tr>
<td>D39</td>
<td>Bad</td>
<td>Bad</td>
</tr>
<tr>
<td>D45</td>
<td>Bad</td>
<td>Bad</td>
</tr>
<tr>
<td>D50</td>
<td>Bad</td>
<td>Good</td>
</tr>
<tr>
<td>D54</td>
<td>Bad</td>
<td>Fair</td>
</tr>
<tr>
<td>D56</td>
<td>Good</td>
<td>Good</td>
</tr>
</tbody>
</table>

No good ordering was possible

Need to care here:

- Question Answering
- Natural Language Generation
- Summarization
- Sentence Disaggregation
- ...

Where?
Why?
Where?
What?
How?
Need to care here:

- Question Answering
- Natural Language Generation
- Summarization
- Sentence Disaggregation
- ...
Majority Ordering
• Barzilay et al 2002

Chronological Ordering
• Barzilay et al 2002

Probabilistic Ordering
• Lapata 2003

Bottom-Up Approach
• Bollegala et al 2006

Sentence Adjacency
• Nie et al 2006

Cluster Adjacency
• Donghong & Yu 2008
Majority Ordering
  • Barzilay et al 2002
Chronological Ordering
  • Barzilay et al 2002
Probabilistic Ordering
  • Lapata 2003
Bottom-Up Approach
  • Bollegala et al 2006
Sentence Adjacency
  • Nie et al 2006
Cluster Adjacency
  • Donghong & Yu 2008
Cluster Source Sentences into Themes based on similarity to summary sentences

Cluster Source Sentences into *Themes* based on *similarity* to summary sentences.

Cluster Source Sentences into *Themes* based on *similarity* to summary sentences

Cluster Source Sentences into *Themes* based on *similarity* to summary sentences.

Cluster Source Sentences into *Themes* based on *similarity* to summary sentences

Cluster Source Sentences into *Themes* based on *similarity* to summary sentences.
Construct a Precedence Graph

Majority Ordering

Why?
Where?
What?
How?

Majority Ordering

Weight nodes (Themes) by their net weight

Majority Ordering

Order Summary Sentences Based on the Weights of their corresponding themes

**Majority Ordering**

**Limitations**

- Assumes that summary sentences belong to different themes.

<table>
<thead>
<tr>
<th></th>
<th>Poor</th>
<th>Fair</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority Ordering</td>
<td>3</td>
<td>14</td>
<td>8</td>
</tr>
</tbody>
</table>

The man accused of rebombing two Manhattan subways in 1994 was convicted Thursday after the jury rejected the notion that the drug Prozac led him to commit the crimes.

He was found guilty of two counts of attempted murder, 14 counts of rst-degree assault and two counts of criminal possession of a weapon.

In December 1994, Leary ignited rebombs on two Manhattan subway trains. The second blast injured 50 people { 16 seriously, including Leary.

Leary wanted to extort money from the Transit Authority.

The defense argued that Leary was not responsible for his actions because of "toxic psychosis" caused by the Prozac.

A summary graded as Good

Hemingway, 69, died of natural causes in a Miami jail after being arrested for indecent exposure.

A book he wrote about his father, "Papa: A Personal Memoir," was published in 1976.

He was picked up last Wednesday after walking naked in Miami.

A transvestite who later had a sex-change operation, he suffered bouts of drinking, depression and drifting, according to acquaintances.

"It's not easy to be the son of a great man," Scott Donaldson, told Reuters.

At the time of his death, he lived in the Coconut Grove district where he was well-known to its Bohemian crowd.

He had been due to appear in court later that day on charges of indecent exposure and resisting arrest.

He sometimes went by the name of Gloria and wore women's clothes.

The cause of death was hypertension and cardiovascular disease.

Taken to the Miami-Dade Women's Detention Center, he was found dead in his cell early on Monday, spokeswoman Janelle Hall said.

**A summary graded as Poor**

Outline

Majority Ordering
• McKeown 2001; Barzilay et al 2002

Chronological Ordering
• Barzilay et al 2002

Probabilistic Ordering
• Lapata 2003

Bottom-Up Approach
• Bollegala et al 2006

Sentence Adjacency
• Nie et al 2006

Cluster Adjacency
• Donghong & Yu 2008
Cluster Source Sentences into *Themes* based on *similarity* to summary sentences

Cluster Source Sentences into Themes based on similarity to summary sentences

Chronological Ordering

Date themes by the earliest publication date of its sentences. Resolve ties according to the order in the source article.

Chronological Ordering

25 Summaries

Chronological Ordering

Human Judges

<table>
<thead>
<tr>
<th></th>
<th>Poor</th>
<th>Fair</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological Ordering</td>
<td>10</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>


Limitations

- Temporal information aren’t always available
- Not all information are time-related (e.g. background information)

<table>
<thead>
<tr>
<th></th>
<th>Poor</th>
<th>Fair</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological Order</td>
<td>10</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>
Augmented Chronological Ordering

Oct 5, 11:35 am
Oct 6, 6:11 am
Oct 5, 10:20 am
Oct 7, 3:45 am

Cluster Sentences into Themes

Why?
Where?
What?
How?

Augmented Chronological Ordering

Segment each input text based on word distribution and coreference analysis

Augmented Chronological Ordering

Compute Topical Relatedness

Topical Relatedness =

\[
\frac{\text{# of sentences that appear in the same Segment}}{\text{# of sentences that appear in the same text}}
\]

Augmented Chronological Ordering

Augmented Chronological Ordering

Oct 5, 11:35 am

TH1

Oct 6, 6:11 am

TH2

Oct 5, 10:20 am

TH3

Oct 7, 3:45 am

TH4

Remove the edges where: weight < Threshold

Augmented Chronological Ordering

Oct 5, 11:35 am
TH1

Oct 6, 6:11 am
TH2

Oct 5, 10:20 am
TH3

Oct 7, 3:45 am
TH4

Blocks of topical related themes

Augmented Chronological Ordering

Order sentences by blocks then by themes within the block.

Chronological Ordering

Limitations

- Depends on the availability of temporal information
- Depends on the quality of text segmentation

<table>
<thead>
<tr>
<th></th>
<th>Poor</th>
<th>Fair</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augmented Chronological Ordering</td>
<td>3</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

Outline

Majority Ordering
- McKeown 2001; Barzilay et al 2002

Chronological Ordering
- Barzilay et al 2002

Probabilistic Ordering
- Lapata 2003

Bottom-Up Approach
- Bollegala et al 2006

Sentence Adjacency
- Nie et al 2006

Cluster Adjacency
- Donghong & Yu 2008
Probabilistic Ordering

Training Phase
- Large Text Corpus
- Learn Ordering Constraints
- Trained Model

Testing Phase
- Unordered Summary
- Trained Model
- Ordered Summary

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
Choose the ordering that maximizes the probability

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
Probabilistic Ordering

Choose the ordering that maximizes the probability

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
Assume that the probability of a sentence depends only on its previous sentence.
Probabilistic Ordering

\[ P(S_1, \ldots, S_n) = P(S_1) P(S_2 | S_1) P(S_3 | S_1, S_2) \ldots P(S_n | S_1 \ldots S_{n-1}) \]

But how can \( P(S_n | S_{n-1}) \) be estimated?

Assume that the probability of a sentence depends only on its previous sentence.
Proabilistic Ordering

\[
P(S_n | S_{n-1}) = P(\text{Features of } S_n | \text{Features of } S_{n-1})
\]

Assumption: features are independent

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
Probabilistic Ordering

Assumption: features are independent

But how the features can be used to calculate $P(S_n | S_{n-1})$?
Probabilistic Ordering

$P(S_3 | S_2) = P(h|e).P(h|f).P(h|g).P(i|e).P(i|f).P(i|g)$

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
Probabilistic Ordering

\[ P(S_3 | S_2 ) = P(h | e).P(h | f ).P(h | g).P(i | e).P(i | f ).P(i | g ) \]

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
Probabilistic Ordering

\[ P(f_1 \mid f_2) = \frac{\text{# times } f_1 \text{ is preceded by } f_2}{\text{# times } f_1 \text{ is preceded by any feature}} \]

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
Probabilistic Ordering

Why?
Where?
What?
How?

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
Laidlaw Transportation Ltd. said shareholders will be asked at its Dec, 7 annual meeting to approve a change of name to Laidlaw Inc.

Verb Features

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
Laidlaw Transportation Ltd. said shareholders will be asked at its Dec, 7 annual meeting to approve a change of name to Laidlaw Inc.

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
The company said its existing name hasn’t represented its businesses since the 1984 sale of its trucking operations.

Dependency Features

<table>
<thead>
<tr>
<th>Verb</th>
<th>Dependency Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>say V:subj:N company</td>
<td>represent V:subj:N name</td>
</tr>
<tr>
<td>represent V:have:have have</td>
<td>represent V:obj:N business</td>
</tr>
</tbody>
</table>
Probabilistic Ordering

Evaluation Metric

\[ T = 1 - \frac{2 \text{(number of inversions)}}{N(N - 1)/2} \]

N is the number of sentences

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
Probabilistic Ordering

25 Summaries

<table>
<thead>
<tr>
<th></th>
<th>T (avg)</th>
<th>StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Ordering</td>
<td>0.43</td>
<td>0.13</td>
</tr>
<tr>
<td>Probabilistic Ordering</td>
<td>0.56</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
• The large number of features may include many noisy features (words).

<table>
<thead>
<tr>
<th></th>
<th>Z (avg)</th>
<th>StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Ordering</td>
<td>0.43</td>
<td>0.13</td>
</tr>
<tr>
<td>Probabilistic Ordering</td>
<td>0.56</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Mirella Lapata, "Probabilistic Text Structuring: Experiments with Sentence Ordering", ACL, 2003
Outline

Majority Ordering
• McKeown 2001; Barzilay et al 2002

Chronological Ordering
• Barzilay et al 2002

Probabilistic Ordering
• Lapata 2003

Bottom-Up Approach
• Bollegala et al 2006

Sentence Adjacency
• Nie et al 2006

Cluster Adjacency
• Donghong & Yu 2008
Agglomerative Hierarchical Clustering

Why?

Where?

What?

How?


\[ f(X > Y) = \begin{cases} 
P & \text{(if X precedes Y)} \\
0 & \text{(if Y precedes X)} 
\end{cases} \]
Bottom-Up

\[
f_{chro} ( A > B ) = \begin{cases} 
1 & \text{If } \text{Time}(A) < \text{Time} (B) \\
1 & \text{If } \text{Doc}(A) = \text{Doc} (B) \& A \text{ precedes } B \\
0.5 & \text{If } \text{Time}(A) = \text{Time} (B) \& \text{Doc}(A) \neq \text{Doc} (B) \\
0 & \text{Otherwise} 
\end{cases}
\]

Chronology Criterion
$f_{\text{topic}}(A \succ B) = \frac{1}{|B|} \sum_{b \in B} \max_{a \in A} \text{sim}(a, b)$.  

**Topical-Relatedness Criterion**


\[
f_{\text{pre}}(A \succ B) = \frac{1}{|B|} \sum_{b \in B} \max_{a \in A, p \in P_b} \text{sim}(a, p)
\]

Precedence Criterion
Bottom-Up

Limitations

• Supervised Learning -> Domain Dependent

<table>
<thead>
<tr>
<th></th>
<th>T (avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom-up Ordering</td>
<td>0.603</td>
</tr>
<tr>
<td>Topical</td>
<td>0.414</td>
</tr>
<tr>
<td>Chronological</td>
<td>0.583</td>
</tr>
</tbody>
</table>

Outline

Majority Ordering
• McKeown 2001; Barzilay et al 2002

Chronological Ordering
• Barzilay et al 2002

Probabilistic Ordering
• Lapata 2003

Bottom-Up Approach
• Bollegala et al 2006

Sentence Adjacency
• Nie et al 2006

Cluster Adjacency
• Donghong & Nie 2008
Sentence Adjacency

Nie et al, 2006
Sentence Adjacency

\[ C_{i,j} = \frac{1}{K \times L} \sum_{k,l} C_f(f_{ik}, f_{jl}) \]

\[ C_f(f_i, f_j) = \frac{f(f_i, f_j)^2}{f(f_i)f(f_j)} \]

Nie et al, 2006
### Sentence Adjacency Ordering

157 Summaries  
DUC 2004

<table>
<thead>
<tr>
<th>Ordering</th>
<th>T (avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Ordering</td>
<td>-0.007</td>
</tr>
<tr>
<td>Majority Ordering</td>
<td>0.143</td>
</tr>
<tr>
<td>Probability ordering</td>
<td>0.144</td>
</tr>
<tr>
<td>Adjacency</td>
<td>0.276</td>
</tr>
</tbody>
</table>

Nie et al, 2006
Outline

Majority Ordering
- McKeown 2001; Barzilay et al 2002

Chronological Ordering
- Barzilay et al 2002

Probabilistic Ordering
- Lapata 2003

Bottom-Up Approach
- Bollegala et al 2006

Sentence Adjacency
- Nie et al 2006

Cluster Adjacency
- Donghong & Nie 2008
Cluster Adjacency

Summary Sentences (Labeled)

Source Sentences (Unlabeled)

Label Propagation

Clusters of related sentences

Cluster Adjacency (Same as sentence adjacency but on cluster level)

Cluster Adjacency

<table>
<thead>
<tr>
<th></th>
<th>T (avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Ordering</td>
<td>-0.007</td>
</tr>
<tr>
<td>Majority Ordering</td>
<td>0.143</td>
</tr>
<tr>
<td>Probability ordering</td>
<td>0.144</td>
</tr>
<tr>
<td>Sentence Adjacency</td>
<td>0.316</td>
</tr>
<tr>
<td>Cluster Adjacency</td>
<td>0.415</td>
</tr>
</tbody>
</table>

Conclusions

• **Majority ordering** performs poorly and requires that source documents agree on information order.

• **Chronological ordering** performs fairly well if temporal information are available.

• **Topical Relatedness feature** significantly improves chronological ordering but this highly depends on the accuracy of topic segmentation.

• **Probabilistic ordering** performs well but the large number of features used introduces a significant amount of noise.

• **Bottom-up ordering** combines several ordering features in a machine learning approach. It outperforms all the previous methods. However, being supervised makes it domain dependent.
Discussion

• What other criteria could be used?

• Are these methods applicable to other domains (other than multi-document summarization)?

• More accurate metrics and methods for automatic evaluation?
The power to question is the basis of all human progress.

-- Indira Gandhi

It's a very good question, very direct, and I'm not going to answer it.

-- George H.W. Bush