Summarization & Discourse

JSALT Summer Workshop: Summer School
June 24, 2015
Gina-Anne Levow
University of Washington
Roadmap

- Motivation & relation to discourse
- Framing the problem
- Structuring the solution:
  - Content selection
  - Information ordering
  - Content realization
- Evaluating summaries
- Conclusions & Future Directions
Motivation

- Information retrieval is very powerful
  - Search engines index and search enormous doc sets
  - Retrieve billions of documents in tenths of seconds

- But still limited!
  - Technically – keyword search (mostly)
  - Conceptually
    - User seeks information
      - Sometimes a web site or document
      - Sometimes the answer to a question
      - But, often a summary of document or document set
Why Summarization?

- Even web search relies on simple summarization
Why Summarization?

- Even web search relies on simple summarization
- Snippets!
  - Provide thumbnail summary of ranked document

---

**Caldera - Wikipedia, the free encyclopedia**

A caldera is a cauldron-like volcanic feature usually formed by the collapse of land following a volcanic eruption. They are sometimes confused with volcanic craters. The word comes from Spanish caldera, and this from Latin caldaria, meaning "cooking pot".

Volcanic crater - Yellowstone Caldera - Cauldron - Coatepeque Caldera

**How Volcanoes Work - Calderas**

When an erupting volcano empties a shallow-level magma chamber, the edifice of the volcano may collapse into the voided reservoir, thus forming...

**Caldera: Crater Formed by Volcanic Collapse or Explosion**

Calderas are some of the most spectacular features on Earth. They are large volcanic craters that form by two different methods: 1) an explosive volcanic eruption; or, 2) collapse of surface rock into an empty magma chamber.
Why Summarization?

- Complex questions go beyond factoids, infoboxes
- Require explanations, analysis
  - E.g. In children with an acute febrile illness, what is the efficacy of single-medication therapy with acetaminophen or ibuprofen in reducing fever?
- Highest search hit is manually created summary site
  - Umich medical
  - Vs 5 articles cited in creating
Ibuprofen is More Likely to Normalize Temperature than Acetaminophen, Though Both are Safe and Effective Antipyretics for Short-Term Use in Children

Question

- In children with an acute febrile illness, what is the efficacy of single-medication therapy with acetaminophen or ibuprofen compared with combination therapy combining the two medications in reducing fever while avoiding adverse effects?

Clinical Bottom Lines

1. Both acetaminophen and ibuprofen are effective antipyretics and are well-tolerated in short-term use in febrile children.
2. Ibuprofen is more effective at achieving temperature normalization than acetaminophen, though both effectively lower temperatures >1.5 C in most patients with standard dosing.
3. There is no data currently available comparing the efficacy and tolerability alternating regimens with ibuprofen and acetaminophen to single-drug regimens.¹

Summary of Key Evidence

1. 628 children aged 6 months to 6 years with initial temperature >38.5C were randomized to receive ibuprofen, acetaminophen, or dipyprone (banned in the US) in a 1:1:1 ratio. The study was double-blinded and multinational. There was no placebo arm.
Why Summarization?

- Complex questions go beyond factoids, infoboxes
- Require explanations, analysis
  - E.g. In children with an acute febrile illness, what is the efficacy of single-medication therapy with acetaminophen or ibuprofen in reducing fever?

  • Summ: Ibuprofen provided greater temperature decrement and longer duration of antipyresis than acetaminophen when the two drugs were administered in approximately equal doses. (PubMedID: 1621668)
Why Summarization?

- Huge scale, explosive growth in online content
  - 2-4K articles in PubMed daily, 41.7M articles/mo on WordPress alone (2014)
- How can we manage it?
  - Lots of aggregation sites
Why Summarization?

- Huge scale, explosive growth in online content
  - 2-4K articles in PubMed daily, 41.7M articles/mo on WordPress alone (2014)
  - How can we manage it?
    - Lots of aggregation sites

- Recordings of meetings, classes, MOOCs
  - Slow to access linearly, awkward to jump around
  - Structured summary can be useful
    - Outline of: how-tos, to-dos,
Why Summarization & Discourse?

- Summarization is fundamentally a discourse task
  - Operates on extended spans of text/speech

- Produces a coherent, cohesive (compressed) span

- Informed by discourse structure, relations, coreference
  - Highlight salient information

- Organize output for coherence, cohesion
Structuring the Summarization Task

- Summarization Task: (Mani and Mayberry 1999)
  - Process of distilling the most important information from a text to produce an abridged version for a particular task and user
Structuring the Summarization Task

- Summarization Task: (Mani and Mayberry 1999)
  - Process of distilling the most important information from a text to produce an abridged version for a particular task and user

- Main components:
  - Content selection
  - Information ordering
  - Sentence realization
Dimensions of Summarization

- Rich problem domain:
  - Tasks and Systems vary on:
    - Use purpose
    - Audience
  - Derivation
  - Coverage
  - Reduction
  - Units
  - Input/Output form factors
Dimensions of Summarization

- Rich problem domain:
  - Tasks and Systems vary on:
    - Use purpose: What is the goal of the summary?
    - Audience
  - Derivation
  - Coverage
  - Reduction
  - Units
  - Input/Output form factors
Dimensions of Summarization

- Rich problem domain:
  - Tasks and Systems vary on:
    - Use purpose: What is the goal of the summary?
    - Audience: Who is reading the summary?
  - Derivation
  - Coverage
  - Reduction
  - Units
  - Input/Output form factors
Dimensions of Summarization

- Rich problem domain:
  - Tasks and Systems vary on:
    - Use purpose: What is the goal of the summary?
    - Audience: Who is reading the summary? Expert/novice
  
  - Derivation
  - Coverage
  - Reduction
  - Units
  - Input/Output form factors
Dimensions of Summarization

- Rich problem domain:
  - Tasks and Systems vary on:
    - Use purpose: What is the goal of the summary?
    - Audience: Who is reading the summary? Expert/novice

- Derivation: How is the summary formed? Extract/Abstract
  - Coverage
  - Reduction
  - Units
  - Input/Output form factors
Extract vs Abstract

Extract from the Gettysburg Address:

Four score and 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 can long endure. We are met on a great battlefield of that war. We have come to dedicate a portion of that field. But the brave men, living and dead, who struggled here, have consecrated it far above our poor power to add or detract. From these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion — that government of the people, by the people for the people shall not perish from the earth.

Abstract of the Gettysburg Address:

This 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.

Figure 23.13 An extract versus an abstract from the Gettysburg Address (abstract from Mani (2001)).
Dimensions of Summarization

- Rich problem domain:
  - Tasks and Systems vary on:
    - Use purpose: What is the goal of the summary?
    - Audience: Who is reading the summary? Expert/novice
  - Derivation: How is the summary formed? Extract/Abstract
  - Coverage: What parts are summarized? General/Focussed
  - Reduction
  - Units
  - Input/Output form factors
Dimensions of Summarization

- Rich problem domain:
  - Tasks and Systems vary on:
    - Use purpose: What is the goal of the summary?
    - Audience: Who is reading the summary? Expert/novice
  - Derivation: How is the summary formed? Extract/Abstract
  - Coverage: What parts are summarized? General/Focussed
  - Reduction: How much compression?
  - Units
  - Input/Output form factors
Dimensions of Summarization

- Rich problem domain:
  - Tasks and Systems vary on:
    - Use purpose: What is the goal of the summary?
    - Audience: Who is reading the summary? Expert/novice
  - Derivation: How is the summary formed? Extract/Abstract
  - Coverage: What parts are summarized? General/Focused
  - Reduction: How much compression?
  - Units: Single vs Multi-document
  - Input/Output form factors
Dimensions of Summarization

- Input/Output form factors:
  - Language: Evaluations include:
    - English, Arabic, Chinese, Japanese, multilingual
  - Register: Formality, style
  - Genre: e.g. News, sports, medical, technical,....
  - Structure: forms, tables, lists, web pages
  - Medium: text, speech, video, tables
  - Subject
Typical Research Paradigm

- Use purpose: Reflective, informational
- Audience: Generic, expert
- Derivation: Predominantly extractive
- Coverage: Mix of Generic and Focused
- Reduction: Typically 100-250 words
- Units: Multi-document
- Input/Output form factors:
  - Mostly English newswire text summaries
General Architecture
General Strategy

- Given a document (or set of documents):
  - Select the key content from the text
  - Determine the order to present that information
  - Perform clean-up or rephrasing to create coherent output
  - Evaluate the resulting summary
General Strategy

- Given a document (or set of documents):
  - Select the key content from the text
  - Determine the order to present that information
  - Perform clean-up or rephrasing to create coherent output
  - Evaluate the resulting summary
- Systems vary in structure, complexity, information
More specific strategy

- For single document, extractive summarization:
  - Segment the text into sentences
  - Identify the most prominent sentences
  - Pick an order to present them
  - Do any necessary processing to improve coherence
More specific strategy

- For single document, extractive summarization:
  - Segment the text into sentences
  - Identify the most prominent sentences
  - Pick an order to present them
    - Maybe trivial, i.e. document order
  - Do any necessary processing to improve coherence
    - Shorten sentences, fix coref, etc
Content Selection

- Goal: Identify most important/relevant information
- Common perspective:
  - View as binary classification: important vs not
    - For each unit (e.g. sentence in the extractive case)
  - Can be unsupervised or supervised
- What makes a sentence (for simplicity) extract-worthy?
Cues to Saliency

- Approaches significantly differ in terms of cues
- Word-based unsupervised techniques
- Graph-based centrality
- Discourse-based:
  - Discourse saliency $\Rightarrow$ extract-worthiness
- Supervised learning with many features:
  - Cues include position, cue phrases, word salience, ..
Unsupervised Word-based Techniques

Key idea:

- Compute **topic signature** of words characterizing topic
  - Includes words with weight above some threshold
- Select content/sentences with highest weight
Unsupervised Word-based Techniques

- Key idea:
  - Compute **topic signature** of words characterizing topic
    - Includes words w/ **weight** above some threshold
  - Select content/sentences with highest weight

- Centroid-based models:
  - Signature terms= pseudo-sentence “centroid” of all sentences in topic. Select closest.

- What weights?
Unsupervised Word-based Techniques

- **Key idea:**
  - Compute **topic signature** of words characterizing topic
    - Includes words w/ **weight** above some threshold
  - Select content/sentences with highest weight

- **Centroid-based models:**
  - Signature terms = pseudo-sentence “centroid” of all sentences in topic. Select closest.

- **What weights?**
  - $\text{Tf*idf}$: term frequency $\times$ inverse document frequency
  - Log likelihood ratio (LLR)

- **Sentence weight:**

\[
\text{weight}(s_i) = \sum_{w \in s_i} \frac{\text{weight}(w)}{|\{w|w \in s_i\}|}
\]
Log Likelihood Ratio

- \( k_1 \) = count of w in topic cluster
- \( k_2 \) = count of w in background corpus
- \( n_1 \) = # features in topic cluster; \( n_2 \) = # in background
- \( p_1 = k_1 / n_1 \); \( p_2 = k_2 / n_2 \); \( p = (k_1 + k_2) / (n_1 + n_2) \)

- \( L(p, k, n) = p^k (1 - p)^{n-k} \)

\[
-2 \log \lambda = 2[\log L(p_1, k_1, n_1) + \log L(p_2, k_2, n_2) - \log L(p, k_1, n_1) - \log L(p, k_2, n_2)]
\]
Graph-based Centrality

- Graph-based approach:
  - Sentences (or other units) in cluster link to each other
  - Salient if similar to many others
    - More central or relevant to the cluster
    - Low similarity with most others, not central

- Graph:
  - Nodes: sentences
  - Edges: measure of similarity between sentences

- How do we compute similarity b/t nodes?
  - Here: tf*idf (could use other schemes); thresholded
LexRank

- LexRank idea:
  - Node can have high(er) score via high scoring neighbors
    - Same idea as PageRank, Hubs & Authorities
      - Page ranked high b/c pointed to by high ranking pages

- Can compute iteratively to convergence via:
  
  \[ p(u) = \frac{d}{N} + (1-d) \sum_{v \in \text{adj}(u)} \frac{p(v)}{\text{deg}(v)} \]
  
  - (power method computes eigenvector)

- Iteratively select highest scoring sentences
Discourse Structure for Content Selection

- Intuition:
  - Discourse structure, relations highlight salient content
    - Louis et al, 2010

- Discourse structure models (from this morning):
  - RST: hierarchical, tree-based model of discourse
    - Relations between nucleus and satellite EDUs
  - PDTB: Linear discourse structure
    - Explicit and implicit relations b/t inter-, intra-sentential args
Discourse Structure Example

1. [Mr. Watkins said] 2. [volume on Interprovincial’s system is down about 2% since January] 3. [and is expected to fall further,] 4. [making expansion unnecessary until perhaps the mid-1990s.]
Discourse Structure Features

- Satellite penalty:
  - For each EDU: # of satellite nodes b/t it and root
    - 1 satellite in tree: (1), one step to root: penalty = 1

- Promotion set:
  - Nuclear units at some level of tree
    - At leaves, EDUs are themselves nuclear

- Depth score:
  - Distance from lowest tree level to EDUs highest rank
    - 2,3,4: score = 4; 1: score = 3

- Promotion score:
  - # of levels span is promoted:
    - 1: score = 0; 4: score = 2; 2,3: score = 3
PDTB “Semantic” Features

- Capture specific relations on spans
- Binary features over tuple of:
  - Implicit vs Explicit
- Name of relation that holds
  - Top-level or second level
- If relation is between sentences,
  - Indicate whether Arg1 or Arg2
- E.g. “contains Arg1 of Implicit Restatement relation”
- Also, # of relations, distance b/t args w/in sentence
Example 1

- In addition, its machines are easier to operate, so customers require less assistance from software.

- Is there an explicit discourse marker?
Example 1

- In addition, its machines are easier to operate, so customers require less assistance from software.

- Is there an explicit discourse marker?
  - Yes, ‘so’

- Discourse relation?
Example 1

- In addition, its machines are easier to operate, so customers require less assistance from software.

- Is there an explicit discourse marker?
  - Yes, ‘so’

- Discourse relation?
  - ‘Contingency’
Example II

- (1) Wednesday’s dominant issue was Yasuda & Marine Insurance, which continued to surge on rumors of speculative buying. (2) It ended the day up 80 yen to 1880 yen.

- Is there a discourse marker?
Example II

- (1) Wednesday’s dominant issue was Yasuda & Marine Insurance, which continued to surge on rumors of speculative buying. (2) It ended the day up 80 yen to 1880 yen.

- Is there a discourse marker?
  - No

- Is there a relation?
Example II

- (1) Wednesday’s dominant issue was Yasuda & Marine Insurance, which continued to surge on rumors of speculative buying. (2) It ended the day up 80 yen to 1880 yen.

- Is there a discourse marker?
  - No

- Is there a relation?
  - Implicit (by definition)

- What relation?
Example II

- (1) Wednesday’s dominant issue was Yasuda & Marine Insurance, which continued to surge on rumors of speculative buying. (2) It ended the day up 80 yen to 1880 yen.

- Is there a discourse marker?
  - No

- Is there a relation?
  - Implicit (by definition)

- What relation?
  - Expansion (or more specifically (level 2) restatement)

- What Args?
Example II

- (1) Wednesday’s dominant issue was Yasuda & Marine Insurance, which continued to surge on rumors of speculative buying. (2) It ended the day up 80 yen to 1880 yen.

- Is there a discourse marker?
  - No

- Is there a relation?
  - Implicit (by definition)

- What relation?
  - Expansion (or more specifically (level 2) restatement)

- What Args? (1) is Arg1; (2) is Arg2 (by definition)
Significant Features

- Associated with summary sentences
  - Structure: depth score, promotion score

- Semantic: Arg1 of Explicit Expansion, Implicit Contingency, Implicit Expansion, distance to arg
Significant Features

- Associated with non-summary sentences
  - Structural: satellite penalty

- Semantic: Explicit expansion, explicit contingency, Arg2 of implicit temporal, implicit contingency, ...
  - # shared relations

- Discourse structure, relations can improve summary
  - In conjunction with other non-discourse information
Multi-document Example

Eight killed in Finnish school massacre TUUSULA, Finland, Nov 7, 2007 AFP

Finnish school shooter carried 500 bullets: police Vantaa, Finland, Nov 8, 2007 AFP

Student kills eight in Finnish school shooting HELSINKI, Nov 7, 2007 AFP

Finnish PM extends condolences to school shooting victims HELSINKI, Nov. 7 Xinhua

At least 7 killed in Finland school shooting TUUSULA, Finland 2007-11-07 15:14:41 UTC

8 dead in Finland school shooting, TUUSULA, Finland 2007-11-07 16:47:30 UTC

Grief and disbelief in Finland after school massacre TUUSULA, Finland, Nov 8, 2007 AFP
Multi-document Challenges

- Key issue
Multi-document Challenges

- Key issue: redundancy
- General idea:
  - Add salient content that is least similar to that already there
Multi-document Challenges

• Key issue: redundancy
  • General idea:
    • Add salient content that is least similar to that already there

• One technique: Maximal Margin Relevance
  • Carbonell and Goldstein (1998)
  • Penalize sentence based on similarity to current summary
  • MMR Penalty Score(s) = \( \lambda \max_{s_i \text{in summary}} \text{sim}(s,s_i) \)
  • Alternatives: clustering, optimization, etc
Information Ordering

- Goal: Determine presentation order for salient content
- Factors:
Information Ordering

- Goal: Determine presentation order for salient content

- Factors:
  - Semantics
    - Chronology: respect sequential flow of content (esp. events)
  - Discourse
    - Cohesion: Adjacent sentences talk about same thing
    - Coherence: Adjacent sentences naturally related (PDTB)
Single vs Multi-Document

- Strategy for single-document summarization?
Single vs Multi-Document

- Strategy for single-document summarization?
  - Just keep original order
  - Chronology? Cohesion? Coherence?

- Multi-document
Single vs Multi-Document

- Strategy for single-document summarization?
  - Just keep original order
  - Chronology? Cohesion? Coherence?

- Multi-document
  - “Original order” can be problematic
  - Chronology?
Single vs Multi-Document

- Strategy for single-document summarization?
  - Just keep original order
  - Chronology? Cohesion? Coherence?

- Multi-document
  - “Original order” can be problematic
  - Chronology?
    - Publication order vs document-internal order
    - Differences in document ordering of information
Single vs Multi-Document

- Strategy for single-document summarization?
  - Just keep original order
  - Chronology? Cohesion? Coherence?

- Multi-document
  - “Original order” can be problematic
  - Chronology?
    - Publication order vs document-internal order
    - Differences in document ordering of information
  - Cohesion?
  - Coherence?
Single vs Multi-Document

- Strategy for single-document summarization?
  - Just keep original order
  - Chronology? Ok
  - Cohesion? Ok
  - Coherence? Iffy

- Multi-document
  - “Original order” can be problematic
  - Chronology?
    - Publication order vs document-internal order
    - Differences in document ordering of information
  - Cohesion? Probably poor
  - Coherence? Probably poor
A Bad Example

- 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.

- “He had a difficult life.”

- 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.
A Basic Approach

- Publication chronology:
- Given a set of ranked extracted sentences

Order by:
- Across articles
  - By publication date
- Within articles
  - By original sentence ordering

Clearly not ideal, but works pretty well
New Approach

- Experiments on sentence ordering by subjects
  - Many possible orderings but far from random
    - Blocks of sentences group together (cohere)
New Approach

- Experiments on sentence ordering by subjects
  - Many possible orderings but far from random
    - Blocks of sentences group together (cohere)

- Combine chronology with cohesion
  - Order chronologically, but group similar themes (sent clusters)

- Perform topic segmentation on original texts

- Themes “related” if, when two themes appear in same text, they frequently appear in same topic segment (threshold)

- Order over groups of themes by CO,
  - Then order within groups by CO

- Significantly better!
Before and After

Thousands of people have attended a ceremony in Nairobi commemorating the first anniversary of the deadly bombings attacks against U.S. Embassies in Kenya and Tanzania. Saudi dissident Osama bin Laden, accused of masterminding the attacks, and nine others are still at large. President Clinton said, "The intended victims of this vicious crime stood for everything that is right about our country and the world".

U.S. federal prosecutors have charged 17 people in the bombings. Albright said that the mourning continues.

Kenyans are observing a national day of mourning in honor of the 215 people who died there.
Before and After

Thousands of people have attended a ceremony in Nairobi commemorating the first anniversary of the deadly bombings attacks against U.S. Embassies in Kenya and Tanzania. Saudi dissident Osama bin Laden, accused of masterminding the attacks, and nine others are still at large. President Clinton said, "The intended victims of this vicious crime stood for everything that is right about our country and the world".

U.S. federal prosecutors have charged 17 people in the bombings. Albright said that the mourning continues.

Kenyans are observing a national day of mourning in honor of the 215 people who died there.

Thousands of people have attended a ceremony in Nairobi commemorating the first anniversary of the deadly bombings attacks against U.S. Embassies in Kenya and Tanzania. Kenyans are observing a national day of mourning in honor of the 215 people who died there.

Saudi dissident Osama bin Laden, accused of masterminding the attacks, and nine others are still at large. U.S. federal prosecutors have charged 17 people in the bombings.

President Clinton said, "The intended victims of this vicious crime stood for everything that is right about our country and the world". Albright said that the mourning continues.
Entity-Centric Cohesion

- Continuing to talk about same thing(s) lends cohesion to discourse

- Incorporated variously in discourse models
  - Lexical chains: Link mentions across sentences
    - Fewer lexical chains crossing $\rightarrow$ shift in topic
  - Salience hierarchies, information structure
    - Subject $>$ Object $>$ Indirect $>$ Oblique $>$ ....
  - Centering model of coreference
    - Combines grammatical role preference with
    - Preference for types of reference/focus transitions
Entity-Based Ordering

- Idea:
  - Leverage patterns of entity (re)mentions
Entity-Based Ordering

- Idea:
  - Leverage patterns of entity (re)mentions

- Intuition:
  - Captures local relations b/t sentences, entities
  - Models cohesion of evolving story
Entity-Based Ordering

- Idea:
  - Leverage patterns of entity (re)mentions

- Intuition:
  - Captures local relations b/t sentences, entities
  - Models cohesion of evolving story

- Pros:
  - Largely delexicalized
    - Less sensitive to domain/topic than other models
  - Can exploit state-of-the-art syntax, coreference tools
Entity Grid

- Need compact representation of:
  - Mentions, grammatical roles, transitions
    - Across sentences

- Entity grid model:
  - Rows:
Entity Grid

- Need compact representation of:
  - Mentions, grammatical roles, transitions
    - Across sentences

- Entity grid model:
  - Rows: sentences
  - Columns:
Entity Grid

- Need compact representation of:
  - Mentions, grammatical roles, transitions
    - Across sentences

- Entity grid model:
  - Rows: sentences
  - Columns: entities
  - Values: grammatical role of mention in sentence
    - Roles: (S)ubject, (O)bject, X (other), __ (no mention)
1 [The Justice Department]_s is conducting an [anti-trust trial]_o against [Microsoft Corp.]_x with [evidence]_x that [the company]_s is increasingly attempting to crush [competitors]_o.
2 [Microsoft]_o is accused of trying to forcefully buy into [markets]_x where [its own products]_s are not competitive enough to unseat [established brands]_o.
3 [The case]_s revolves around [evidence]_o of [Microsoft]_s aggressively pressuring [Netscape]_o into merging [browser software]_o.
4 [Microsoft]_s claims [its tactics]_s are commonplace and good economically.
5 [The government]_s may file [a civil suit]_o ruling that [conspiracy]_s to curb [competition]_o through [collusion]_x is [a violation of the Sherman Act]_o.
6 [Microsoft]_s continues to show [increased earnings]_o despite [the trial]_x.
Grids → Features

- Intuitions:
  - Some columns dense: focus of text (e.g. MS)
    - Likely to take certain roles: e.g. S, O
  - Others sparse: likely other roles (x)
  - Local transitions reflect structure, topic shifts
Intuitions:
- Some columns dense: focus of text (e.g. MS)
  - Likely to take certain roles: e.g. S, O
- Others sparse: likely other roles (x)
- Local transitions reflect structure, topic shifts

Local entity transitions: \{s,o,x,\}_n
- Continuous column subsequences (role n-grams?)
- Compute probability of sequence over grid:
  - # occurrences of that type/# of occurrences of that len
Vector Representation

- Document vector:
  - Length
Vector Representation

- Document vector:
  - Length: # of transition types
  - Values:
Vector Representation & Learning

- **Document vector:**
  - Length: # of transition types
  - Values: Probabilities of each transition type

<table>
<thead>
<tr>
<th></th>
<th>s</th>
<th>s</th>
<th>s</th>
<th>s</th>
<th>o</th>
<th>o</th>
<th>o</th>
<th>o</th>
<th>x</th>
<th>x</th>
<th>o</th>
<th>o</th>
<th>x</th>
<th>--</th>
<th>--</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_1$</td>
<td>.01</td>
<td>.01</td>
<td>0</td>
<td>.08</td>
<td>.01</td>
<td>0</td>
<td>0</td>
<td>.09</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.03</td>
<td>.05</td>
<td>.07</td>
</tr>
<tr>
<td>$d_2$</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.07</td>
<td>0</td>
<td>.02</td>
<td>.14</td>
<td>.14</td>
<td>.06</td>
<td>.04</td>
<td>.03</td>
<td>.07</td>
<td>.01</td>
<td>.36</td>
<td></td>
</tr>
<tr>
<td>$d_3$</td>
<td>.02</td>
<td>0</td>
<td>0</td>
<td>.03</td>
<td>.09</td>
<td>0</td>
<td>.09</td>
<td>.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.05</td>
<td>.03</td>
<td>.07</td>
</tr>
</tbody>
</table>
Vector Representation & Learning

- Document vector:
  - Length: # of transition types
  - Values: Probabilities of each transition type

Used as input to SVMRank
- Yields substantial improvements in ordering
  - Even with imperfect coreference
Content Realization

- Goal: Create a fluent, readable, compact output
Content Realization

- Goal: Create a fluent, readable, compact output
- Abstractive approaches range from templates to full NLG
Content Realization

- Goal: Create a fluent, readable, compact output
- Abstractive approaches range from templates to full NLG
- Extractive approaches focus on:
Content Realization

- Goal: Create a fluent, readable, compact output
- Abstractive approaches range from templates to full NLG
- Extractive approaches focus on:
  - Sentence simplification/compression:
    - Manipulation of parse tree to remove unneeded info
    - Rule-based, machine-learned
Content Realization

- Goal: Create a fluent, readable, compact output
- Abstractive approaches range from templates to full NLG
- Extractive approaches focus on:
  - Sentence simplification/compression:
    - Manipulation of parse tree to remove unneeded info
    - Rule-based, machine-learned
  - Reference presentation and ordering:
    - Based on saliency hierarchy of mentions
Examples

- Compression:
  - When it arrives sometime next year in new TV sets, the V-chip will give parents a new and potentially revolutionary device to block out programs they don’t want their children to see.
Examples

• Compression:
  • When it arrives sometime next year in new TV sets, the V-chip will give parents a new and potentially revolutionary device to block out programs they don’t want their children to see.
Examples

• Compression:
  • When it arrives sometime next year in new TV sets, the V-chip will give parents a new and potentially revolutionary device to block out programs they don’t want their children to see.

• Coreference:
  • Advisers do not blame O’Neill, but they recognize a shakeup would help indicate Bush was working to improve matters. U.S. President George W. Bush pushed out Treasury Secretary Paul O’Neill and ...
Examples

- Compression:
  - When it arrives sometime next year in new TV sets, the V-chip will give parents a new and potentially revolutionary device to block out programs they don’t want their children to see.

- Coreference:
  - Advisers do not blame Treasury Secretary Paul O’Neill, but they recognize a shakeup would help indicate U.S. President George W. Bush was working to improve matters. Bush pushed out O’Neill and ...
Evaluation

- **Extrinsic evaluations:**
  - Does the summary allow users to perform some task?
    - As well as full docs? Faster?
  - **Examples:**
    - Relevance assessment: relevant or not?
  - **Time-limited fact-gathering:**
    - Answer questions about news event
      - Compare with full doc, human summary, auto summary

- Hard to frame in general, though
Intrinsic Evaluation

- Need basic comparison to simple, naïve approach
- Baselines:
Intrinsic Evaluation

- Need basic comparison to simple, naïve approach
- Baselines:
  - Random baseline:
    - Select N random sentences
Intrinsic Evaluation

- Need basic comparison to simple, naïve approach

- Baselines:
  - Random baseline:
    - Select N random sentences

- Leading sentences:
  - Select N leading sentences
  - For news, surprisingly hard to beat
    - (For reviews, last N sentences better.)
Intrinsic Manual Evaluation

• Readability:
  • How fluent, coherent, grammatical is the summary?
  • Attend to:
    • Grammaticality, non-redundancy, referential clarity, focus, and structure and coherence.
Intrinsic Manual Evaluation

- Readability:
  - How fluent, coherent, grammatical is the summary?
  - Attend to:
    - Grammaticality, non-redundancy, referential clarity, focus, and structure and coherence.

- Overall responsiveness:
  - How well does the summary cover relevant content without undesirable redundancy? Is it well-formed?
Intrinsic Manual Evaluation

- **Readability:**
  - How fluent, coherent, grammatical is the summary?
  - Attend to:
    - Grammaticality, non-redundancy, referential clarity, focus, and structure and coherence.

- **Overall responsiveness:**
  - How well does the summary cover relevant content without undesirable redundancy? Is it well-formed?

- Both scored on 1 to 5 scale
Charles Carl Roberts IV may have planned to molest the girls at the Amish school, but police have no evidence that he actually did. Charles Carl Roberts IV entered the West Nickel Mines Amish School in Lancaster County and shot 10 girls, killing five. The suspect apparently called his wife from a cell phone shortly before the shooting began, saying he was “acting out in revenge for something that happened 20 years ago,” Miller said. The gunman, a local truck driver Charles Roberts, was apparently acting in “revenge” for an incident that happened to him 20 years ago.
Intrinsic Evaluation

- Most common automatic method: ROUGE
  - “Recall-Oriented Understudy for Gisting Evaluation”
  - Inspired by BLEU (MT)
Intrinsic Evaluation

- Most common automatic method: ROUGE
  - “Recall-Oriented Understudy for Gisting Evaluation”
  - Inspired by BLEU (MT)
  - Computes overlap b/t auto and human summaries
  - E.g. ROUGE-2: bigram overlap
Intrinsic Evaluation

- Most common automatic method: ROUGE
- “Recall-Oriented Understudy for Gisting Evaluation”
- Inspired by BLEU (MT)
- Computes overlap b/t auto and human summaries
- E.g. ROUGE-2: bigram overlap

\[
\text{ROUGE}_2 = \frac{\sum_{S \in \{\text{ReferenceSummaries}\}} \sum_{\text{bigram} \in S} \text{count}_{\text{match}}(\text{bigram})}{\sum_{S \in \{\text{ReferenceSummaries}\}} \sum_{\text{bigram} \in S} \text{count}(\text{bigram})}
\]

- Also, ROUGE-L (longest seq), ROUGE-S (skipgrams)
ROUGE

• Pros:
ROUGE

- Pros:
  - Automatic evaluation allows tuning
    - Given set of reference summaries
  - Simple measure

- Cons:
ROUGE

• Pros:
  • Automatic evaluation allows tuning
    • Given set of reference summaries
  • Simple measure

• Cons:
  • Even human summaries highly variable, disagreement
  • Poor handling of coherence
  • Okay for extractive, highly problematic for abstractive
Conclusions

- Summarization:
  - Broad range of applications
    - Differ across dimensions
  - Dominated by extractive methods over newswire data

- Draws on wide range of:
  - Shallow, deep NLP methods
  - Machine learning models

- Many remaining challenges, opportunities
Future Directions

- Beyond English newswire
  - Other modalities: Speech, video
  - Other domains: Reviews, meetings, lectures
  - Other languages: Multi-lingual summarization
  - Some common strategies, new challenges

- Beyond extractive summarization:
  - Most current approaches extraction
  - Abstractive summarization exciting direction
    - Highly challenging
    - New approaches using different concept representations
      - Dependency trees, semantic representations (AMR)
LexRank

- LexRank idea:
  - Node can have high(er) score via high scoring neighbors
    - Same idea as PageRank, Hubs & Authorities
      - Page ranked high b/c pointed to by high ranking pages
  - Think of matrix $X$ as transition matrix of Markov chain
    - i.e. $X(i,j)$ is probability of transition from state $i$ to $j$
  - Will converge to a stationary distribution ($r$)
    - Probability of ending up in each state via random walk
  - Can compute iteratively to convergence via:
    $$p(u) = \frac{d}{N} + (1 - d) \sum_{v \in \text{adj}(u)} \frac{p(v)}{\text{deg}(v)}$$
    - (power method computes eigenvector)