Information Ordering

Ling573
Systems & Applications
April 20, 2017

Roadmap

- Information Ordering:
  - Basic approaches
    - Variants on chronological ordering
  - Ensembles for ordering

Basics

- Content selection:
  - Identified sentences or information units for summary
- Information ordering:
  - Linearize selected content into a smooth-flowing text
- 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?
  - 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

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 used in some eval. submissions

Improving Ordering

- Improve some set of chronology, cohesion, coherence
- Chronology, cohesion (Barzilay et al, '02)
- Key ideas:
  - Summarization and chronology over "themes"
  - Identifying cohesive blocks within articles
  - Combining constraints for cohesion within time structure

Importance of Ordering

- Analyzed DUC summaries scoring poor on ordering
- Manually reordered existing sentences to improve
- Human judges scored both sets:
  - Incomprehensible, Somewhat Comprehensible, Comp.
- Manually reorderings judged:
  - As good or better than originals
- Argues that people are sensitive to ordering, ordering can improve assessment

Framework

- Build on their existing systems (Multigen)
- Motivated by issues of similarity and difference
  - Managing redundancy and contradiction in docs
- Analysis groups sentences into “themes”
  - Text units from diff’t docs with repeated information
  - Roughly clusters of sentences with similar content
  - Intersection of their information is summarized
- Ordering is done on this selected content

Chronological Orderings I

- Two basic strategies explored:
  - CO:
    - Need to assign dates to **themes** for ordering
      - Theme sentences from multiple docs, lots of dup content
    - Temporal relation extraction is hard, try simple sub.
      - Doc publication date: what about duplicates?
    - **Theme** date: earlier pub date for theme sentence
  - Order **themes** by date
  - If different **themes** have same date?
    - Same article, so use article order
- Slightly more sophisticated than simplest model

Chronological Orderings II

- MO (Majority Ordering):
  - Alternative approachto ordering themes
    - Order the whole themes relative to each other
      - i.e. Th1 precedes Th2
    - How? If all sentences in Th1 before all sentences in Th2?
      - Easy: Th1 b/f Th2
      - If not? Majority rule
        - Problematic b/c not guaranteed transitive
    - Create an ordering by modified topological sort over graph
      - Nodes are themes:
        - Weight: sum of outgoing edges minus sum of incoming edges
      - Edges E(x,y): precedence, weighted by # texts
        - where sentences in x precede those in y

CO vs MO

- Neither of these is particularly good:

	Poor	Fair	Good
MO	3	14	8
CO	10	8	7

- MO works when presentation order consistent
  - When inconsistent, produces own brand new order
- CO problematic on:
  - Themes that aren't tied to document order
    - E.g. quotes about reactions to events
  - Multiple topics not constrained by chronology

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
- Perform topic segmentation on original texts
- Themes “related” if, when two themes appear in same text, they frequently appear in same 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.

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.

Deliverable #3

- Goals:
  - Focus on information ordering
    - Using one or more of:
      - Chronology, Cohesion, Coherence
  - Continue to improve content selection
    - Incorporate some guided/topic-orientation
- Same deliverable structure as D#2
  - Due in 3 weeks:
    - Code/results; Updated report

Notes

- Deliverable 2:
  - Code/results
  - Updated project report
  - Presentations next week:
    - Doodle poll will be sent after class
    - Please email me slide deck (or pointer) by noon
    - If planning to present remotely, contact me to check audio