Information Ordering

Ling 573
Systems and Applications
May 3, 2016

Roadmap

- Ordering models:
  - Chronology and topic structure
  - Mixture of experts
    - Preference ranking:
      - Chronology, topic similarity, succession/precedence
  - Entity-based cohesion
    - Entity transitions
    - Coreference, syntax, and salience

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.
- Manual 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: earliest 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 approach to 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.

Integrating Ordering Preferences

- Learning Ordering Preferences
  - (Bollegala et al, 2012)
- Key idea:
  - Information ordering involves multiple influences
    - Can be viewed as soft preferences
  - Combine via multiple experts:
    - Chronology
    - Sequence probability
    - Topicality
    - Precedence/Succession

Basic Framework

- Combination of experts
- Build one expert for each of diff't preferences
  - Take a pair of sentences (a,b) and partial summary
    - Score > 0.5 if prefer a before b
    - Score < 0.5 if prefer b before a
- Learn weights for linear combination
- Use greedy algorithm to produce final order

Chronology Expert

- Implements the simple chronology model
  - If sentences from two different docs w/diff't times
    - Order by document timestamp
  - If sentences from same document
    - Order by document order
  - Otherwise, no preference

Topicality Expert

- Same motivation as Barzilay 2002
- Example:
  - The earthquake crushed cars, damaged hundreds of houses, and terrified people for hundreds of kilometers around.
  - A major earthquake measuring 7.7 on the Richter scale rocked north Chile Wednesday.
  - Authorities said two women, one aged 88 and the other 54, died when they were crushed under the collapsing walls.
- 2 > 1 > 3

Topicality Expert

- Idea: Prefer sentence about the “current” topic
- Implementation:?
  - Prefer sentence with highest similarity to sentence in summary so far
  - Similarity computation:?
    - Cosine similarity b/t current & summary sentence
    - Stopwords removed; nouns, verbs lemmatized; binary

Precedence/Succession Experts

- Idea: Does current sentence look like blocks preceding/following current summary sentences in their original documents?
- Implementation:
  - For each summary sentence, compute similarity of current sentence w/most similar pre/post in original doc
    - Similarity?: cosine
- $PREF_{pre}(u,v,Q)$= 0.5 if [Q=null] or [pre(u)=pre(v)]
- 1.0 if [Q!=null] and [pre(u)>pre(v)]
- 0 otherwise
  - Symmetrically for post

Sketch

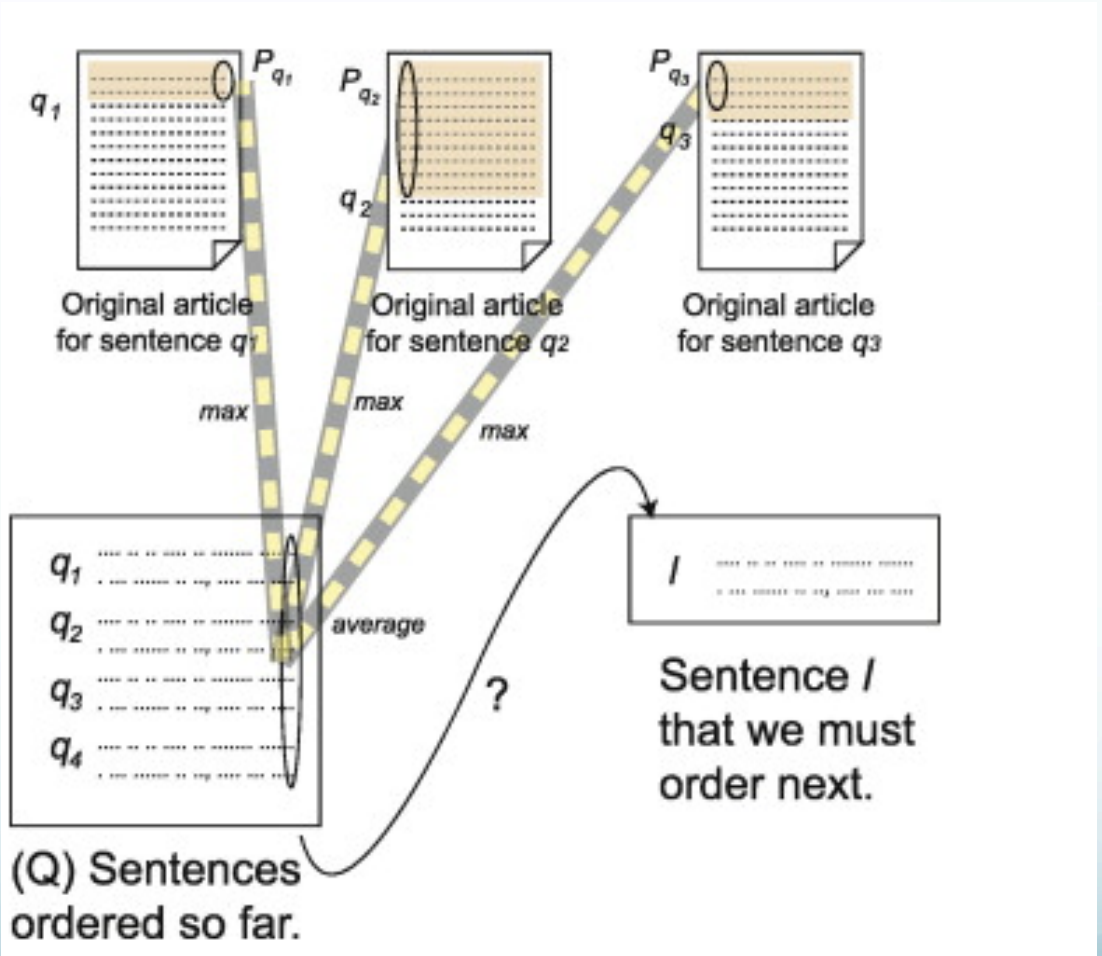

Fig. 4.

Precedence expert.

Probabilistic Sequence

- Intuition:
  - Probability of summary is the probability of sequence of sentences in it, assumed Markov
  - $P(\text{summary}) = \Pi P(S_i | S_{i-1})$
- Issue:
  - Sparsity: will we actually see identical pairs in training?
- Repeatedly backoff:
  - To N, V pairs in ordered sentences
  - To backoff smoothing + Katz

Results & Weights

- Trained weighting using a boosting method
- Combined:
  - Learning approach significantly outperforms random, prob
  - Somewhat better that raw chronology

Expert	Weight
Succession	0.44
Chronology	0.33
Precedence	0.20
Topic	0.016
Prob. Seq.	0.00004

Observations

- Nice ideas:
  - Combining multiple sources of ordering preference
  - Weight-based integration
- Issues:
  - Sparseness everywhere
    - Ubiquitous word-level cosine similarity
    - Probabilistic models
  - Score handling