Information Ordering

Ling573
Systems & Applications
May 2, 2017

Roadmap

- Information ordering
  - Ensemble of experts
    - Integrating sources of evidence
  - Entity-based cohesion
    - Motivation
    - Defining the entity grid
    - Entity grid for information ordering

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

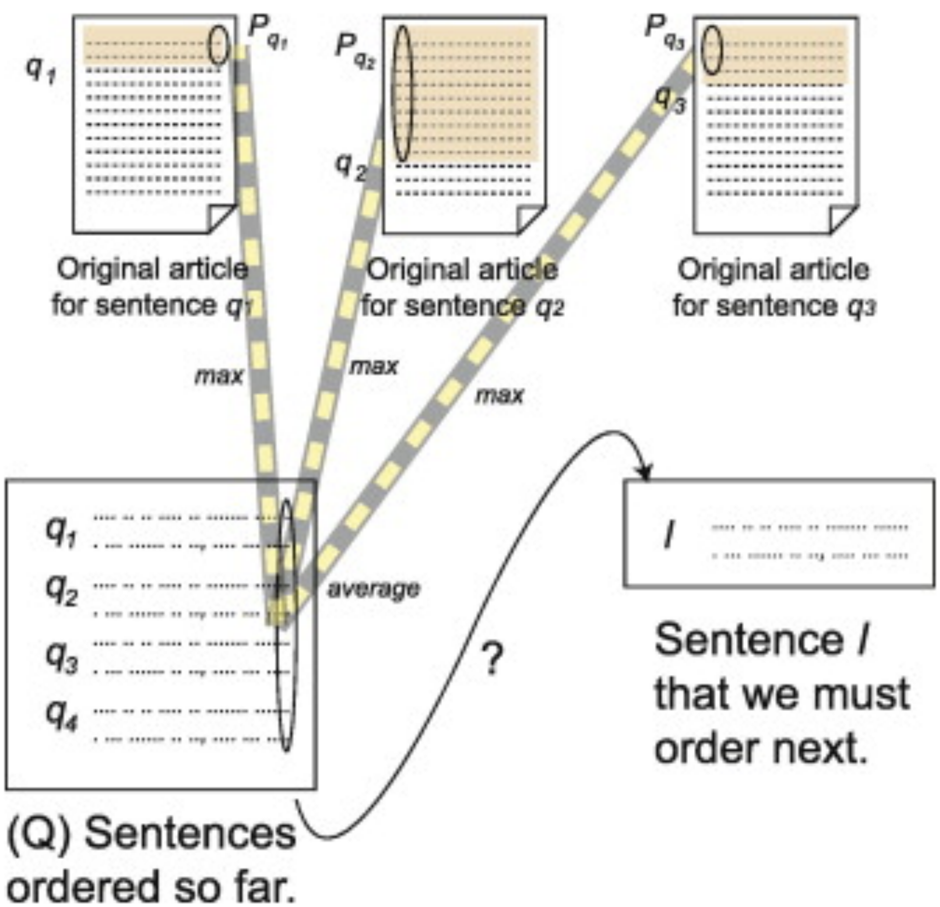

Fig. 4.

Precedence expert.

Probabilistic Sequence

- Intuition:
  - Probability of summary is the probability of sequence of sentences in it, assumed Markov
  - $P(\text{summary}) = \prod P(S_i | S_{I-1})$
- 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

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 → 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
- 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: sentences
  - Columns: entities
  - Values: grammatical role of mention in sentence
    - Roles: (S)ubject, (O)bject, X (other), __ (no mention)
    - Multiple mentions: Take highest

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	s	o	s	x	o	–	–	–	–	–	–	–	–	–	–	1
2	–	–	o	–	–	x	s	o	–	–	–	–	–	–	–	2
3	–	–	s	o	–	–	–	–	s	o	o	–	–	–	–	3
4	–	–	s	–	–	–	–	–	–	–	–	s	–	–	–	4
5	–	–	–	–	–	–	–	–	–	–	–	–	s	o	–	5
6	–	x	s	–	–	–	–	–	–	–	–	–	–	–	o	6

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
- 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: # of transition types
  - Values: Probabilities of each transition type

	S S	S O	S X	S –	O S	O O	O X	O –	X S	X O	X X	X –	– S	– O	– X	– –
d_1	.01	.01	0	.08	.01	0	0	.09	0	0	0	.03	.05	.07	.03	.59
d_2	.02	.01	.01	.02	0	.07	0	.02	.14	.14	.06	.04	.03	.07	0.1	.36
d_3	.02	0	0	.03	.09	0	.09	.06	0	0	0	.05	.03	.07	.17	.39

- Can vary by transition types:
  - E.g. most frequent; all transitions of some length, etc