Compression Strategies & Alternate Summarization

Systems and Applications
Ling 573
May 23, 3017

Roadmap

- Content Realization: Compression
  - Deep, Heuristic Approaches
  - Compression Integration
  - Compression Learning
- Alternate views of summarization
  - Dimensions of summarization redux
  - Abstractive summarization

Form	CLASSY	ISCI	UMd	SumBasic+	Cornell
Initial Adverbials	Y	M	Y	Y	Y
Initial Conj	Y		Y	Y	
Gerund Phr.	Y	M	M	Y	M
Rel clause appos	Y		M	Y	Y
Other adv	Y				
Numeric: ages,	Y				
Junk (byline, edit)	Y				Y
Attributives	Y	Y		Y	Y
Manner modifiers	M	Y	M		Y
Temporal modifiers	M	Y	Y		Y
POS: det, that, MD			Y		
XP over XP			Y		
PPs (w/, w/o constraint)			Y		
Preposed Adjuncts			Y		
SBARs			Y		M
Conjuncts			Y		
Content in parentheses		Y			Y

Deep, Minimal, Heuristic

- ICSI/UTD:
  - Use an Integer Linear Programming approach to solve
- Trimming:
  - Goal: Readability (not info squeezing)
  - Removes temporal expressions, manner modifiers, “said”
    - Why?: “next Thursday”
  - Methodology: Automatic SRL labeling over dependencies
    - SRL not perfect: How can we handle?
    - Restrict to high-confidence labels
- Improved ROUGE on (some) training data
  - Also improved linguistic quality scores

Example

A ban against bistros providing plastic bags free of charge will be lifted at the beginning of March.

A ban against bistros providing plastic bags free of charge will be lifted.

Deep, Extensive, Heuristic

- Both UMD & SumBasic+
  - Based on output of phrase structure parse
  - UMD: Originally designed for headline generation
  - Goal: Information squeezing, compress to add content
- Approach: (UMd)
  - Ordered cascade of increasingly aggressive rules
    - Subsumes many earlier compressions
    - Adds headline oriented rules (e.g. removing MD, DT)
    - Adds rules to drop large portions of structure
      - E.g. halves of AND/OR, wholescale SBAR/PP deletion

Integrating Compression & Selection

- Simplest strategy: (Classy, SumBasic+)
  - Deterministic, compressed sentence replaces original
- Multi-candidate approaches: (most others)
  - Generate sentences at multiple levels of compression
    - Possibly constrained by: compression ratio, minimum len
      - E.g. exclude: < 50% original, < 5 words (ICSI)
  - Add to original candidate sentences list
  - Select based on overall content selection procedure
    - Possibly include source sentence information
    - E.g. only include single candidate per original sentence

Multi-Candidate Selection

- (UMd, Zajic et al. 2007, etc)
- Sentences selected by tuned weighted sum of feats
  - Static:
    - Position of sentence in document
    - Relevance of sentence/document to query
    - Centrality of sentence/document to topic cluster
      - Computed as: IDF overlap or (average) Lucene similarity
    - # of compression rules applied
  - Dynamic:
    - Redundancy: $S = \Pi_{wi \text{ in } S} \lambda P(w|D) + (1 - \lambda) P(w|C)$
    - # of sentences already taken from same document
- Significantly better on ROUGE-1 than uncompressed
  - Grammaticality lousy (tuned on headlinese)

Learning Compression

- Cornell (Wang et al, 2013)
- Contrasted three main compression strategies
  - Rule-based
  - Sequence-based learning
  - Tree-based, learned models
- Resulting sentences selected by SVR model

Compression Corpus

- (Clark & Lapata, 2008)
- Manually created corpus:
  - Written: 82 newswire articles (BNC, ANT)
  - Spoken: 50 stories from HUB-5 broadcast news
- Annotators created compression sentence by sentence
  - Could mark as not compressable
- http://jamesclarke.net/research/resources/

Sequence-based Compression

- View as sequence labeling problem
  - Decision for each word in sentence: keep vs delete
  - Model: linear-chain CRF
    - Labels: B-retain, I-retain, O (token to be removed)
  - Features:
    - “Basic” features: word-based
    - Rule-based features: if fire, force to O
    - Dependency tree features: Relations, depth
    - Syntactic tree features: POS, labels, head, chunk
    - Semantic features: predicate, SRL
      - Include features for neighbors

Feature Set

- Detail:

Basic Features	**Syntactic Tree Features**
first 1/3/5 tokens (toks)?	POS tag
last 1/3/5 toks?	parent/grandparent label
first letter/all letters capitalized?	leftmost child of parent?
is negation?	second leftmost child of parent?
is stopword?	is headword?
Dependency Tree Features	in NP/VP/ADVP/ADJP chunk?
dependency relation (dep rel)	**Semantic Features**
parent/grandparent dep rel	is a predicate?
is the root?	semantic role label
has a depth larger than 3/5?	

Rule-Based Features

For each rule in Table 2 , we construct a corresponding feature to indicate whether the token is identified by the rule.

Tree-based Compression

- Given a phrase-structure parse tree,
  - Determine if each node is: removed, retained, or partial

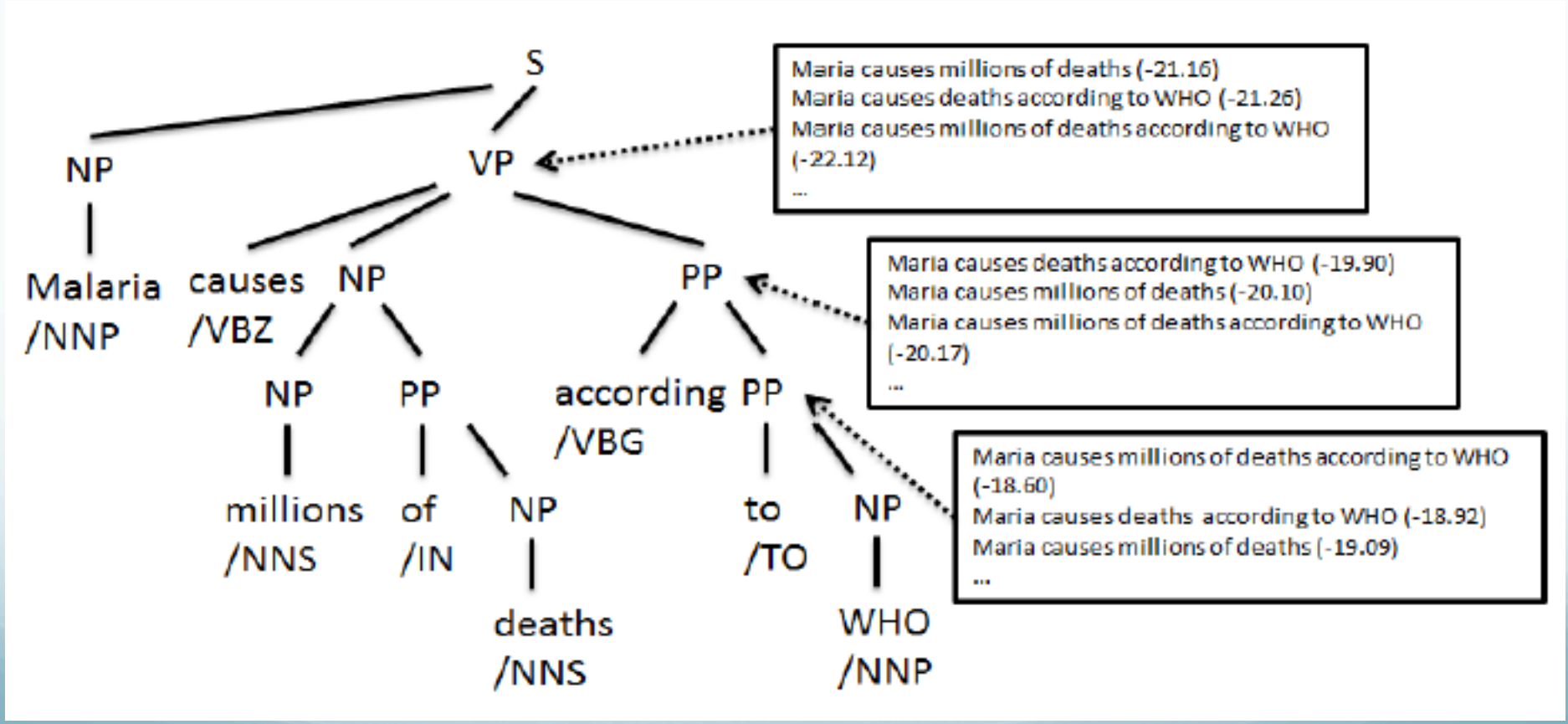
Tree-based Compression

- Given a phrase-structure parse tree,
  - Determine if each node is: removed, retained, or partial
- Issues:
  - # possible compressions exponential
  - Need some local way of scoring a node
  - Need some way of ensuring consistency
  - Need to ensure grammaticality

Tree-based Compression

- Given a phrase-structure parse tree,
  - Determine if each node is: removed, retained, or partial
- Issues & Solutions:
  - # possible compressions exponential
    - Order parse tree nodes (here post-order)
    - Do beam search over candidate labelings
  - Need some local way of scoring a node
    - Use MaxEnt to compute probability of label
  - Need some way of ensuring consistency
    - Restrict candidate labels based on context
  - Need to ensure grammaticality
    - Rerank resulting sentences using n-gram LM

Tree Compression Hypotheses



Features

- Basic features:
  - Analogous to those for sequence labeling
- Enhancements:
  - Context features: decisions about child, sibling nodes
  - Head-driven search:
    - Reorder so head nodes at each level checked first
      - Why? If head is dropped, shouldn't keep rest
      - Revise context features

Summarization Features

- (aka MULTI in paper)
- Calculated based on current decoded word sequence W
- Linear combination of:
  - Score under MaxEnt
  - Query relevance:
    - Proportion of overlapping words with query
  - Importance: Average sumbasic score over W
  - Language model probability
  - Redundancy: 1 --- proportion of words overlapping summ

Summarization Results

	DUC 2006			DUC 2007		
System	**C Rate**	**R-2**	**R-SU4**	**C Rate**	**R-2**	**R-SU4**
Best DUC system	–	9.56	15.53	–	12.62	17.90
Davis et al. (2012)	–	10.2	15.2	–	12.8	17.5
SVR	100%	7.78	13.02	100%	9.53	14.69
LambdaMART	100%	9.84	14.63	100%	12.34	15.62
Rule-based	78.99%	10.62 ∗†	15.73 †	78.11%	13.18†	18.15†
Sequence	76.34%	10.49 †	15.60 †	77.20%	13.25†	18.23†
Tree (BASIC + $Score_{Basic}$)	70.48%	10.49 †	15.86 †	69.27%	13.00†	18.29†
Tree (CONTEXT + $Score_{Basic}$)	65.21%	10.55 ∗†	16.10 †	63.44%	12.75	18.07†
Tree (HEAD + $Score_{Basic}$)	66.70%	10.66 ∗†	16.18 †	65.05%	12.93	18.15†
Tree (HEAD + MULTI)	70.20%	**11.02** ∗†	**16.25** †	73.40%	**13.49**†	**18.46**†

Discussion

- Best system incorporates:
  - Tree structure
  - Machine learning
  - Summarization features
- Rule-based approach surprisingly competitive
  - Though less aggressive in terms of compression
- Learning based approaches enabled by sentence compression corpus

General Discussion

- Broad range of approaches:
  - Informed by similar linguistic constraints
  - Implemented in different ways:
    - Heuristic vs Learned
    - Surface patterns vs parse trees vs SRL
  - Even with linguistic constraints
    - Often negatively impact linguistic quality
    - Key issue: errors in linguistic analysis
      - POS taggers → Parsers → SRL, etc

Alternate Views of Summarization

Dimensions of TAC Summarization

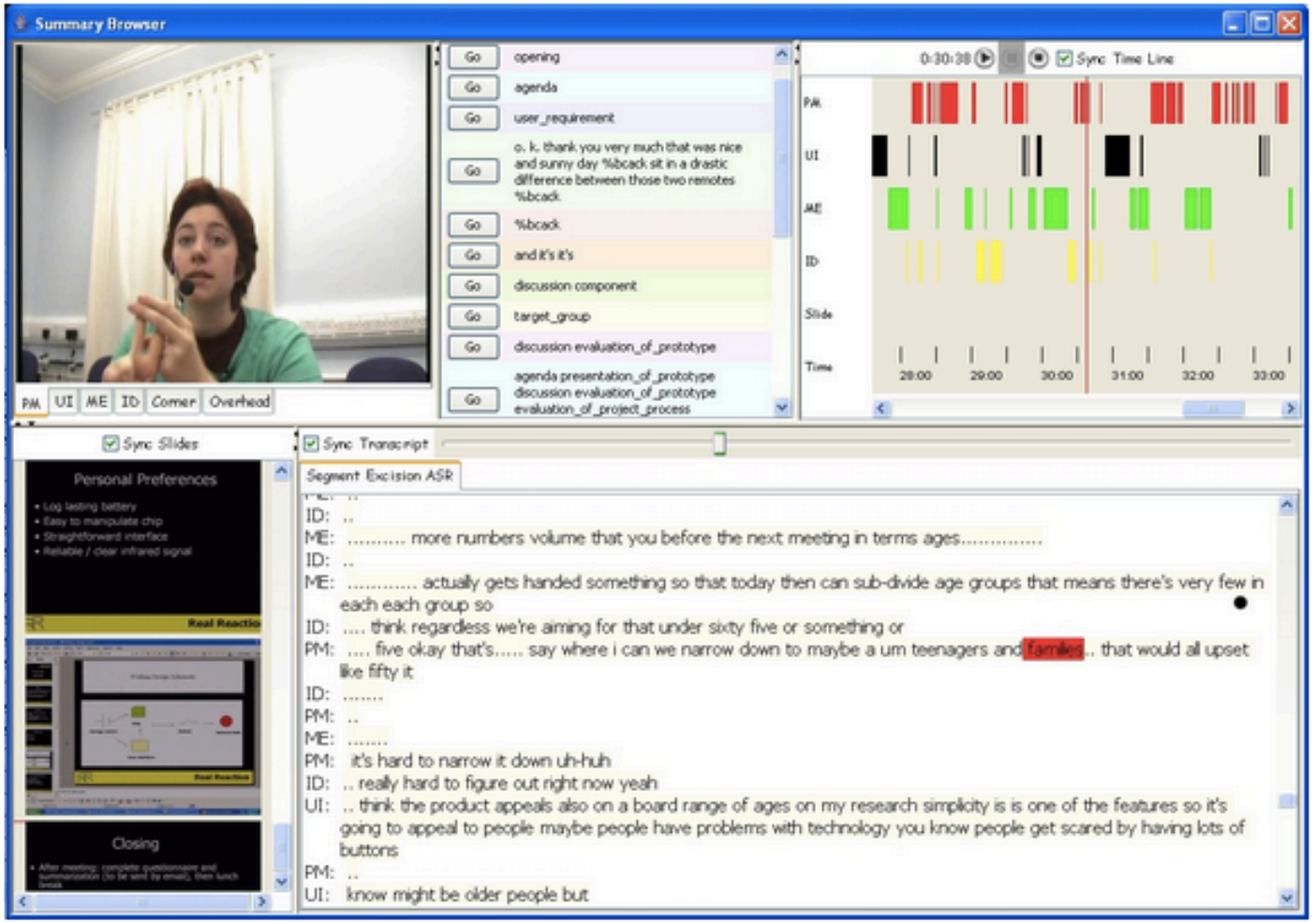
- Use purpose: Reflective summaries
- Audience: Analysts
- Derivation (extactive vs abstractive): Largely extractive
- Coverage (generic vs focused): “Guided”
- Units (single vs multi): Multi-document
- Reduction: 100 words
- Input/Output form factors (language, genre, register, form)
  - English, newswire, paragraph text

Other Types of Summaries

Meeting Summaries

- What do you want out of a summary?

Example


Summary Browser
PM
UI
ME
ID
Corner
Overhead
Go
opening
agenda
user_requirement
o. k. thank you very much that was nice and sunny day %bcack sit in a drastic difference between those two remotes %bcack
%bcack
and it's it's
discussion component
target_group
discussion evaluation_of_prototype
agenda presentation_of_prototype discussion evaluation_of_prototype evaluation_of_project_process
Sync Time Line
0:30:38
28:00
29:00
30:00
31:00
32:00
33:00
Sync Slides
Personal Preferences
Closing
Sync Transcript
Segment Excision ASR
ID: ..
ME: ..... more numbers volume that you before the next meeting in terms ages.....
ID: ..
ME: ..... actually gets handed something so that today then can sub-divide age groups that means there's very few in each each group so
ID: .... think regardless we're aiming for that under sixty five or something or
PM: .... five okay that's..... say where i can we narrow down to maybe a um teenagers and families.. that would all upset like fifty it
ID: .....
PM: ..
ME: .....
PM: it's hard to narrow it down uh-huh
ID: .. really hard to figure out right now yeah
UI: .. think the product appeals also on a board range of ages on my research simplicity is is one of the features so it's going to appeal to people maybe people have problems with technology you know people get scared by having lots of buttons
PM: ..
UI: know might be older people but

Meeting Summaries

- What do you want out of a summary?

- Minutes?
- Agenda-based?
- To-do list
- Points of (Dis)agreement

Dimensions of Meeting Summaries

- Use purpose: Catch up on missed meetings
- Audience: Ordinary attendees
- Derivation (extactive vs abstractive): Extractive or Abstr.
- Coverage (generic vs focused): User-based?
- Units (single vs multi): Single event
- Reduction: ?
- Input/Output form factors (language, genre, register, form)
  - English, speech+, lists/bullets/todos

Examples

- Decision summary:
  - 1. The remote will resemble the potato prototype
  - 2. There will be no feature to help find the remote when it is misplaced;
  - instead the remote will be in a bright colour to address this issue.
  - 3. The corporate logo will be on the remote.
  - 4. One of the colours for the remote will contain the corporate colours.
  - 5. The remote will have six buttons.
  - 6. The buttons will all be one colour.
  - 7. The case will be single curve.
  - 8. The case will be made of rubber.
  - 9. The case will have a special colour.

Examples

- Action items:
  - They will receive specific instructions for the next meeting by email.
  - They will fill out the questionnaire.

Examples

- Abstractive summary:
  - When this functional design meeting opens the project manager tells the group about the project restrictions he received from management by email. The marketing expert is first to present, summarizing user requirements data from a questionnaire given to 100 respondents. The marketing expert explains various user preferences and complaints about remotes as well as different interests among age groups. He prefers that they aim users from ages 16-45, improve the most-used functions, and make a placeholder for the remote...