Content Selection: Supervision & Discourse

Ling573
Systems & Applications
April 14, 2016

Roadmap

- Content selection
  - Supervised content selection
    - Analysis & Regression with rich features
    - “CLASSY”: HMM methods
  - Discourse structure
    - Models of discourse structure
    - Structure and relations for summarization

Supervised Word Selection

- RegSumm:
  - Improving the Estimation of Word Importance for News Multi-Document Summarization (Hong & Nenkova, '14)
- Key ideas:
  - Supervised method for word selection
  - Diverse, rich feature set: unsupervised measures, POS, NER, position, etc
  - Identification of common "important" words via side corpus of news articles and human summaries

Assessment: Words

- Select N highest ranked keywords via regression
- Compute F-measure over words in summaries
  - G_i: i = # of summaries in which word appears

G_i	#words	PROB	LLR	MRW	REGBASIC	**REGSUM**
G_1	80	43.6	37.9	38.9	39.9	45.7
G_1	100	44.3	38.7	39.2	41.0	46.5
G_1	120	44.6	38.5	39.2	40.9	46.4
G_2	30	47.8	44.0	42.4	47.4	50.2
G_2	35	47.1	43.3	42.1	47.0	49.5
G_2	40	46.5	42.4	41.8	46.4	49.2

Assessment: Summaries

- Compare summarization w/ROUGE-1,2,4

Basic Systems

State of The Art Systems

System	R-1	R-2	R-4
PROB	35.14	8.17	1.06
LLR	34.60	7.56	0.83
MRW	35.78	8.15	0.99
REGBASIC	37.56	9.28	1.49
KL	37.97	8.53	1.26
PEER-65	37.62	8.96	1.51
SUBMOD	39.18	9.35	1.39
DPP	39.79	9.62	1.57
REGSUM	**38.57**	**9.75**	**1.60**

CLASSY

- "Clustering, Linguistics and Statistics for Summarization Yield"
  - Conroy et al. 2000-2011
- Highlights:
  - High performing system
    - Often rank 1 in DUC/TAC, commonly used comparison
  - Topic signature-type system (LLR)
  - HMM-based content selection
  - Redundancy handling

Using LLR for Weighting

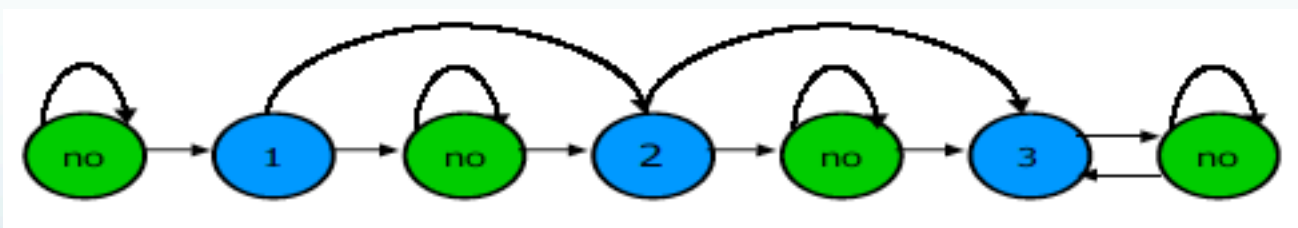
- Compute weight for all cluster terms
  - weight(w_i) = 1 if $-2\log\lambda > 10$, 0 o.w.
- Use that to compute sentence weights

$$weight(s_i) = \sum_{w \in s_i} \frac{weight(w)}{|\{w|w \in s_i\}|}$$

- How do we use the weights?
  - One option: directly rank sentences for extraction
- LLR-based systems historically perform well
  - Better than tf*idf generally

HMM Sentence Selection

- CLASSY strategy: Use LLR as feature in HMM
- How does HMM map to summarization?
  - Key idea:
    - Two classes of states: summary, non-summary
  - Feature(s)?: log(#sig+1) (tried: length, position,..)
    - Lower cased, white-space tokenized (a-z), stopped
  - Topology:



  - Select sentences with highest posterior (in "summary")

Matrix-based Selection

- Redundancy minimizing selection
- Create term x sentence matrix
  - If term in sentence, weight is nonzero
- Loop:
  - Select highest scoring sentence
    - Based on Euclidean norm
  - Subtract those components from remaining sentences
  - Until enough sentences
- Effect: selects highly ranked but different sentences
  - Relatively insensitive to weighting schemes

Combining Approaches

- Both HMM and Matrix method select sentences
- Can combine to further improve
- Approach:
  - Use HMM method to compute sentence scores
    - (e.g. rather than just weight based)
      - Incorporates context information, prior states
  - Loop:
    - Select highest scoring sentence
    - Update matrix scores
      - Exclude those with too low matrix scores
    - Until enough sentences are found

Other Linguistic Processing

- Sentence manipulation (before selection):
  - Remove uninteresting phrases based on POS tagging
    - Gerund clauses, restr. rel. appos, attrib, lead adverbs

- Coreference handling (Serif system)
  - Created coref chains initially
  - Replace all mentions with longest mention (# caps)
  - Used only for sentence selection

Outcomes

- HMM, Matrix: both effective, better combined

- Linguistic pre-processing improves
  - Best ROUGE-1,ROUGE-2 in DUC
- Coref handling improves:
  - Best ROUGE-3, ROUGE-4; 2nd ROUGE-2

Discourse Structure for Content Selection

Text Coherence

- Cohesion – repetition, etc – does not imply coherence
- Coherence relations:
  - Possible meaning relations between utts in discourse
  - Examples:
    - **Result:** Infer state of S_0 cause state in S_1
      - The Tin Woodman was caught in the rain. His joints rusted.
    - **Explanation**: Infer state in S_1 causes state in S_0
      - John hid Bill's car keys. He was drunk.
    - **Elaboration**: Infer same prop. from S_0 and S_1.
      - Dorothy was from Kansas. She lived in the great Kansas prairie.
  - Pair of locally coherent clauses: discourse segment

Rhetorical Structure Theory

- Mann & Thompson (1987)
- Goal: Identify hierarchical structure of text
  - Cover wide range of TEXT types
    - Language contrasts
  - Relational propositions (intentions)
- Derives from functional relations b/t clauses

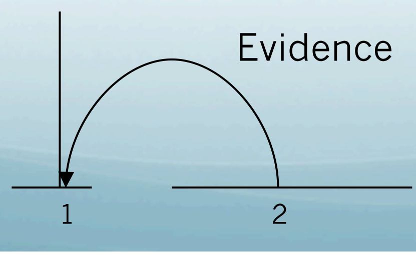
Components of RST

- Relations:
  - Hold b/t two text spans, nucleus and satellite
    - Nucleus core element, satellite peripheral
    - Constraints on each, between
    - Units: Elementary discourse units (EDUs), e.g. clauses

RST Relations

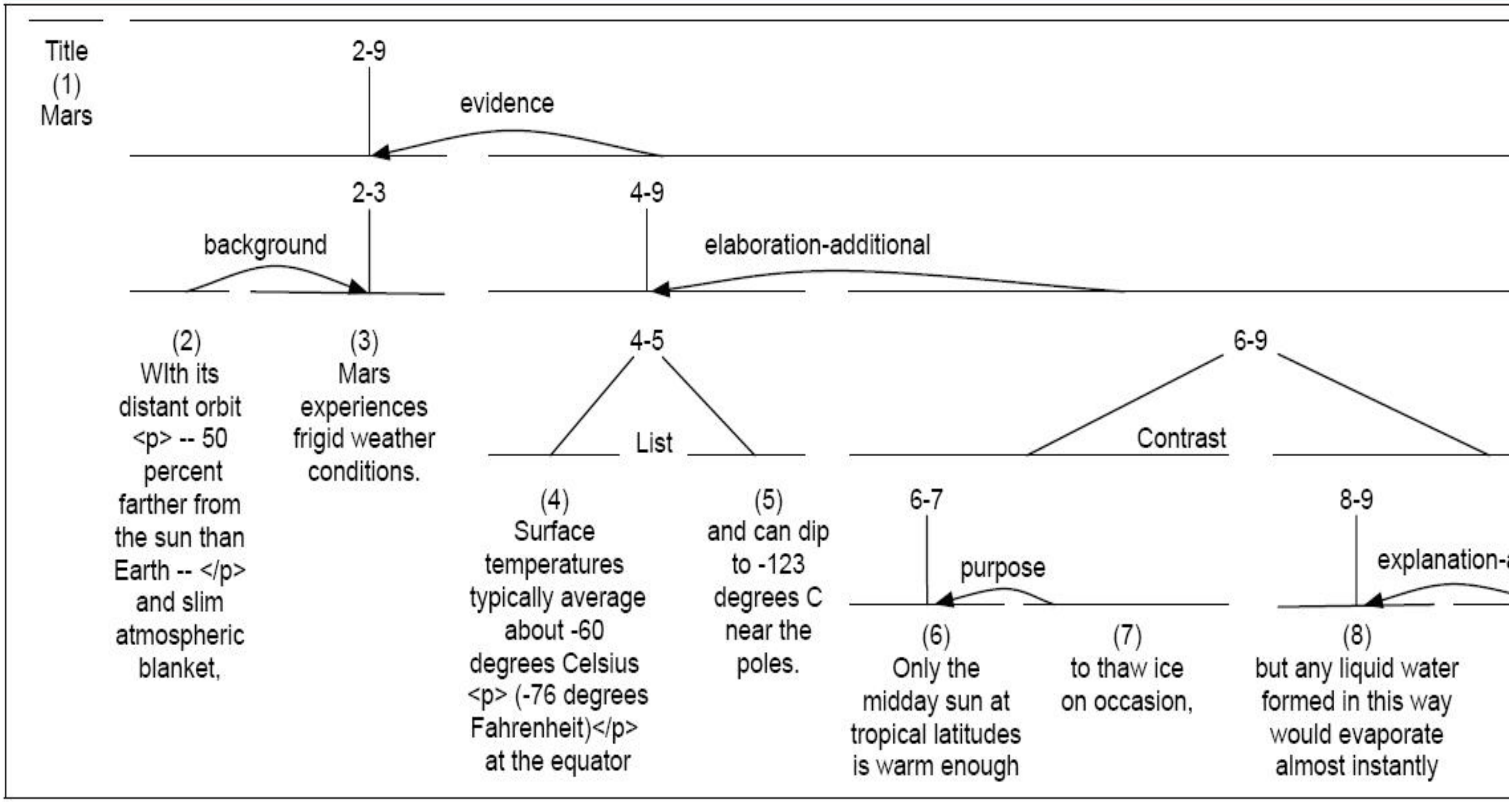
- Evidence
  - The program really works. (N)
  - I entered all my info and it matched my results. (S)

Relation Name:	Evidence
Constraints on N:	R might not believe N to a degree satisfactory to W
Constraints on S:	R believes S or will find it credible
Constraints on N+S:	R's comprehending S increases R's belief of N
Effects:	R's belief of N is increased



RST Relations

- Core of RST
  - RST analysis requires building tree of relations
  - Relations include:
    - Circumstance, Solutionhood, Elaboration. Background, Enablement, Motivation, Evidence, etc
- Captured in:
  - RST treebank: corpus of WSJ articles with analysis
  - RST parsers: Marcu, Peng and Hirst 2014


Title
(1)
Mars
2-9
evidence
2-3
background
4-9
elaboration-additional
(2)
WIth its distant orbit <p> -- 50 percent farther from the sun than Earth -- </p> and slim atmospheric blanket,
(3)
Mars experiences frigid weather conditions.
4-5
List
6-9
Contrast
(4)
Surface temperatures typically average about -60 degrees Celsius <p> (-76 degrees Fahrenheit)</p> at the equator
(5)
and can dip to -123 degrees C near the poles.
6-7
purpose
8-9
(6)
Only the midday sun at tropical latitudes is warm enough
(7)
to thaw ice on occasion,
(8)
but any liquid water formed in this way would evaporate almost instantly

GraphBank

- Alternative discourse structure model
  - Wolf & Gibson, 2005
- Key difference:
  - Analysis of text need not be tree-structure, like RST
  - Can be arbitrary graph, allowing crossing dependency
- Similar relations among spans (clauses)
  - Slightly different inventory