Content Selection: Supervision & Discourse

Ling573
Systems & Applications
April 16, 2015

Roadmap

- Content selection
  - Supervised content selection
    - Analysis & Regression with rich features
  - 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

Basic Approach

- Learn keyword importance
  - Contrasts with unsupervised selection, learning sentences

  - Train regression over large number of possible features
    - Supervision over *words*
      - Did document word appear in summary or not?

  - Greedy sentence selection:
    - Highest scoring sentences: average word weight
    - Do not add if >= 0.5 cosine similarity w/any curr sents

Features I

- Unsupervised measures:
  - Used as binary features given some threshold

- Word probability: count(w)/N
  - Computed over input cluster

- Log likelihood ratio: Gigaword as background corpus

- Markov Random Walk (MRW):
  - Graphical model approach similar to LexRank
  - Nodes: words
  - Edges: # syntactic dependencies b/t wds in sentences
  - Weights via PageRank algorithm

Features II

- "Global" word importance:
  - Question: Are there words which are intrinsically likely to show up in (news) summaries?
  - Approach:
    - Build language models on NYT corpus of articles+summs
      - One model on articles, one model on summaries
      - Measures: $Pr_A(w)$, $Pr_A(w)$-$Pr_G(w)$, $Pr_A(w)/Pr_G(w)$
        - $KL(A||G) = Pr_A(w)*\ln (Pr_A(w)/Pr_G(w))$
        - $KL(G||A) = Pr_G(w)*\ln (Pr_G(w)/Pr_A(w))$
    - Binary features: top-k or bottom-k features

Features III

- Adaptations of common features:
  - Word position as proportion of document [0,1]
    - Earliest first, latest last, average, average first
  - Word type: POS, NER
    - Emphasizes NNS, NN, capitalization; ORG, PERS, LOC
  - MPQA and LIWC features:
    - MPQA: sentiment, subjectivity terms
      - Strong sentiment likely or not? NOT
    - LIWC: words for 64 categories: +: death, anger, money
      - Neg: pron, neg, fn words, swear, adverbs, etc

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

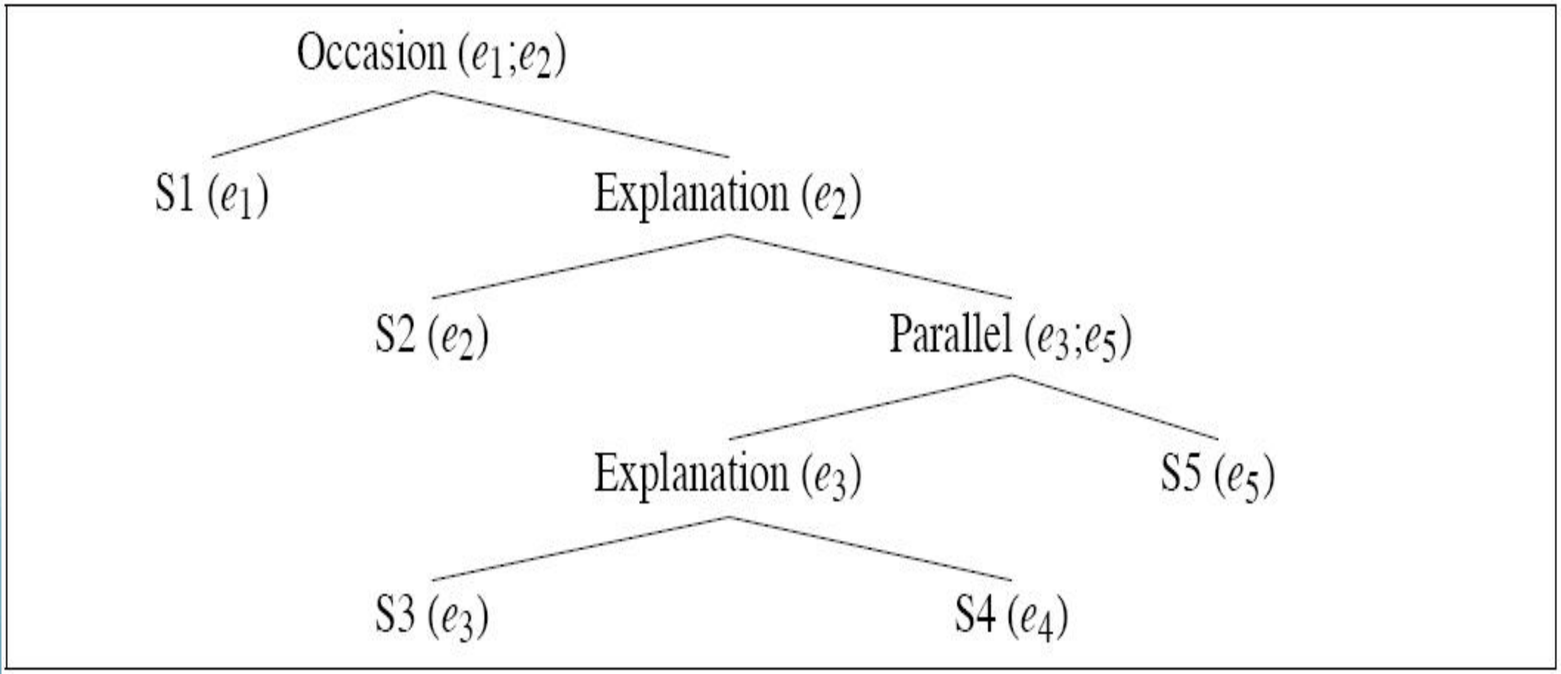
System	R-1	R-2	R-4
PROB	35.14	8.17	1.06
LLR	34.60	7.56	0.83
MRW	35.78	8.15	0.99
REGBASIC	37.56	9.28	1.49
KL	37.97	8.53	1.26
PEER-65	37.62	8.96	1.51
SUBMOD	39.18	9.35	1.39
DPP	39.79	9.62	1.57
REGSUM	**38.57**	**9.75**	**1.60**

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

Coherence Analysis

S1: John went to the bank to deposit his paycheck.
S2: He then took a train to Bill's car dealership.
S3: He needed to buy a car.
S4: The company he works now isn't near any public transportation.
S5: He also wanted to talk to Bill about their softball league.



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
- Schemas:
  - Grammar of legal relations between text spans
  - Define possible RST text structures
    - Most common: N + S, others involve two or more nuclei
- Structures:
  - Using clause units, complete, connected, unique, adjacent

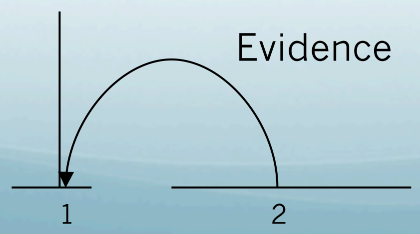
RST Relations

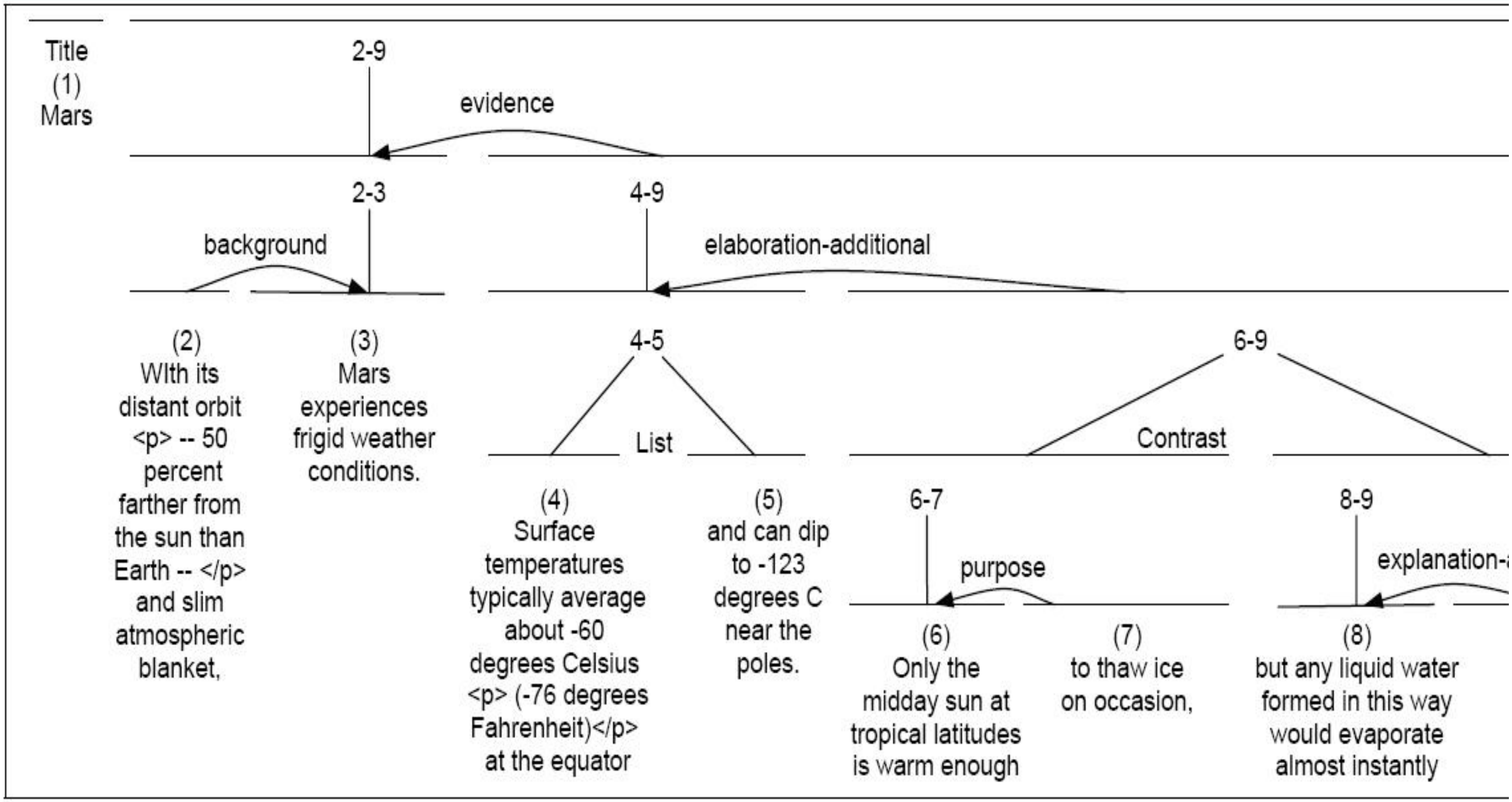
- Core of RST
  - RST analysis requires building tree of relations
  - Circumstance, Solutionhood, Elaboration. Background, Enablement, Motivation, Evidence, Justify, Vol. Cause, Non-Vol. Cause, Vol. Result, Non-Vol. Result, Purpose, Antithesis, Concession, Condition, Otherwise, Interpretation, Evaluation, Restatement, Summary, Sequence, Contrast
- Captured in:
  - RST treebank: corpus of WSJ articles with analysis
  - RST parsers: Marcu, Peng and Hirst 2014

RST Relations

- Evidence
  - Effect: Evidence (Satellite) increases R's belief in Nucleus
    - 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




Title
(1)
Mars
2-9
evidence
2-3
4-9
background
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
explanation-
(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

Penn Discourse Treebank

- PDTB (Prasad et al, 2008)
  - "Theory-neutral" discourse model
  - No stipulation of overall structure, identifies local rels
- Two types of annotation:
  - Explicit: triggered by lexical markers ('but') b/t spans
    - Arg2: syntactically bound to discourse connective, ow Arg1
  - Implicit: Adjacent sentences assumed related
    - Arg1: first sentence in sequence
- Senses/Relations:
  - Comparison, Contingency, Expansion, Temporal
    - Broken down into finer-grained senses too

Discourse & Summarization

- Intuitively, discourse should be useful
  - Selection, ordering, realization
- Selection:
  - Sense: some relations more important
    - E.g. cause vs elaboration
  - Structure: some information more core
    - Nucleus vs satellite, promotion, centrality
- Compare these, contrast with lexical info
  - Louis et al, 2010

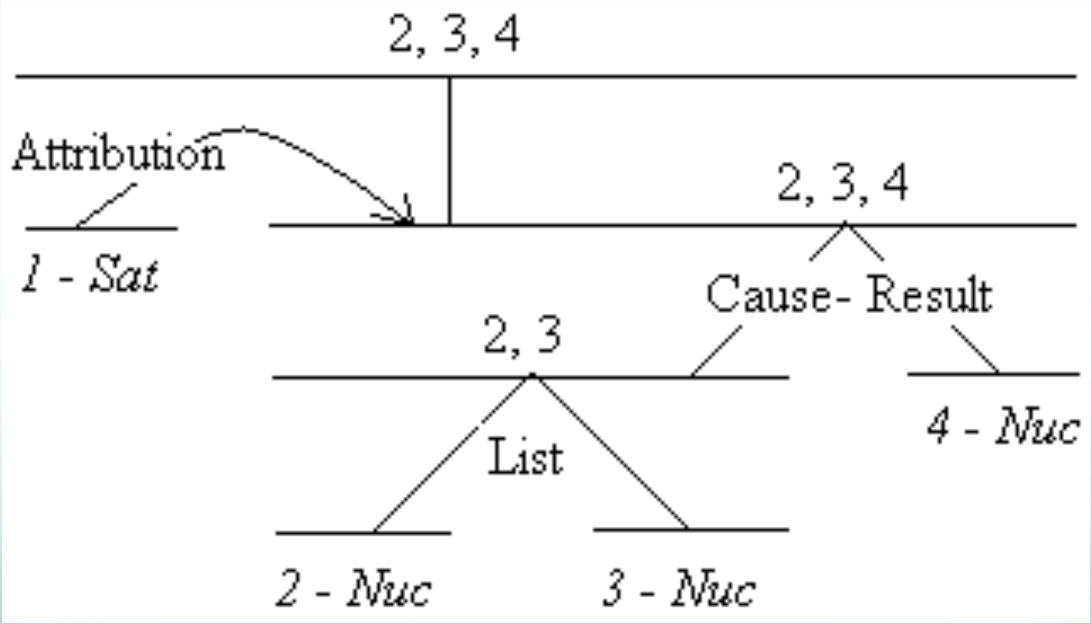
Framework

- Association with extractive summary sentences
  - Statistical analysis
    - Chi-squared (categorical), t-test (continuous)

- Classification:
  - Logistic regression
    - Different ensembles of features
  - Classification F-measure
  - ROUGE over summary sentences

Discourse Structure Example

- 1. [Mr. Watkins said] 2. [volume on Interprovincial's system is down about 2% since January] 3. [and is expected to fall further,] 4. [making expansion unnecessary until perhaps the mid-1990s.]



RST Parsing

- Learn and apply classifiers for
  - Segmentation and parsing of discourse

RST Parsing

- Learn and apply classifiers for
  - Segmentation and parsing of discourse
- Assign coherence relations between spans

RST Parsing

- Learn and apply classifiers for
  - Segmentation and parsing of discourse
- Assign coherence relations between spans
- Create a representation over whole text => parse
- Discourse structure
  - RST trees
    - Fine-grained, hierarchical structure
      - Clause-based units