Alternate Views of Summarization

Ling 573
Systems and Applications
May 26, 2015

Roadmap

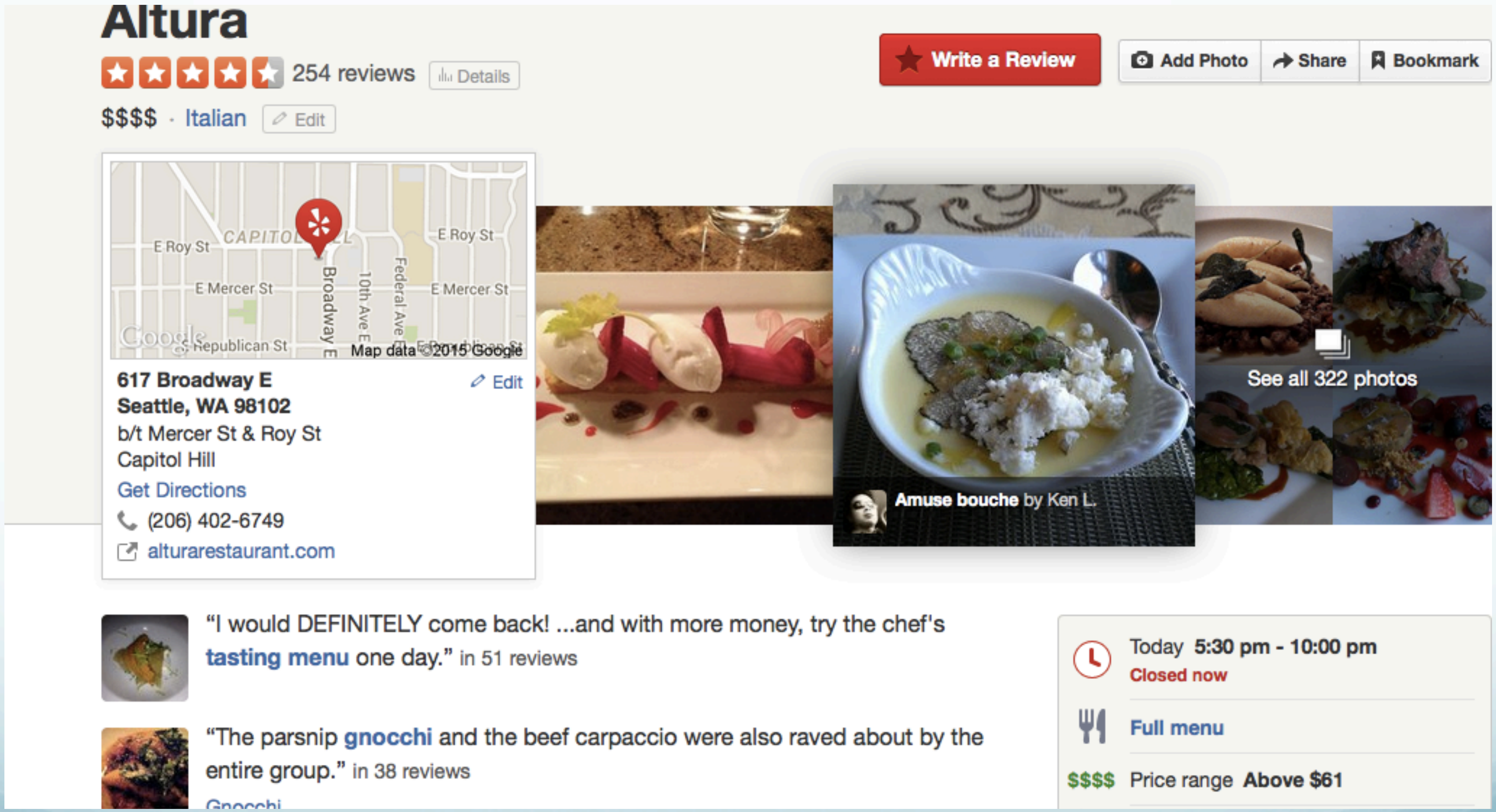
- Alternate views of summarization:
  - Dimensions of the TAC model
  - Other methods, goals, data
    - Abstractive summarization
    - Summarizing reviews
    - Summarizing speech

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

Review Summaries



Review Summary Dimensions

- Use purpose: Product selection, comparison
- Audience: Ordinary people/customers
- Derivation (extactive vs abstractive): Extractive+
- Coverage (generic vs focused): Aspect-oriented
- Units (single vs multi): Multi-document
- Reduction: Varies
- Input/Output form factors (language, genre, register, form)
  - ??, user reviews, less formal, pros & cons, tables, etc

Meeting Summaries

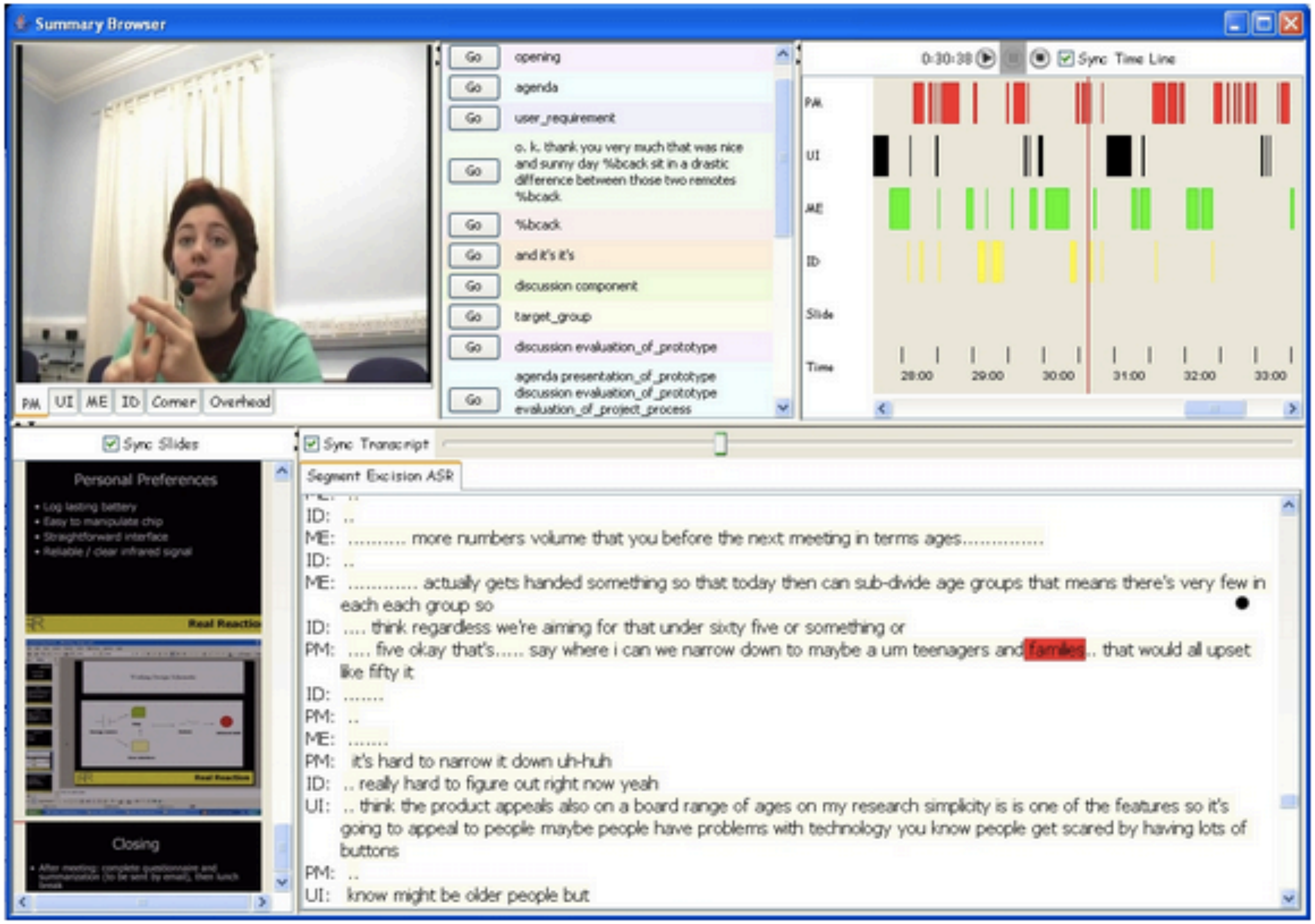
- What do you want out of a summary?

- 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

Example



Summary Browser
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
PM
UI
ME
ID
Slide
Time
28:00
29:00
30:00
31:00
32:00
33:00
PM UI ME ID Corner Overhead
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

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...

Abstractive Summarization

- Basic components:
  - Content selection
  - Information ordering
  - Content realization
    - Comparable to extractive summarization
- Fundamental differences:
  - What do the processes operate on?
    - Extractive? Sentences (or subspans)
    - Abstractive? Major question
      - Need some notion of concepts, relations in text

Levels of Representation

- How can we represent concepts, relations from text?
  - Ideally, abstract away from surface sentences
- Build on some deep NLP representation:
  - Dependency trees: (Cheung & Penn, 2014)
  - Discourse parse trees: (Gerani et al, 2014)
  - Logical Forms
  - Abstract Meaning Representation (AMR): (Liu et al, 2015)

Representations

- Different levels of representation:
  - Syntax, Semantics, Discourse
- All embed:
  - Some nodes/substructure capturing concepts
  - Some arcs, etc capturing relations
  - In some sort of graph representation (maybe a tree)
- What's the right level of representation??

Typical Approach

- Parse original documents to deep representation
- Manipulate resulting graph for content selection
  - Splice dependency trees, remove nucleus nodes, etc
- Generate based on resulting revised graph
- All rely on parsing/generation to/from representation

AMR

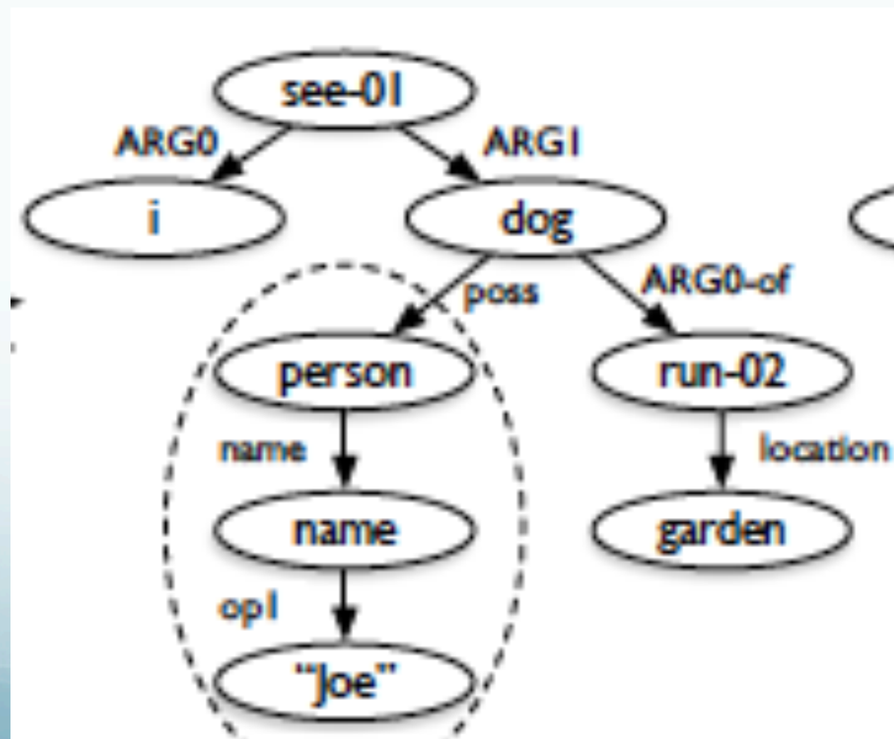
- "Abstract Meaning Representation"
  - Sentence-level semantic representation

  - Nodes: Concepts:
    - English words, PropBank predicates, or keywords ('person')

  - Edges: Relations:
    - PropBank thematic roles (ARG0-ARG5)
    - Others including 'location', 'name', 'time', etc...
    - ~100 in total

AMR 2

- AMR Bank: ~20K annotated sentences
- JAMR parser: 63% F-measure
  - Alignments b/t word spans & graph fragments
- Example: “I saw Joe’s dog, which was running in the garden.”

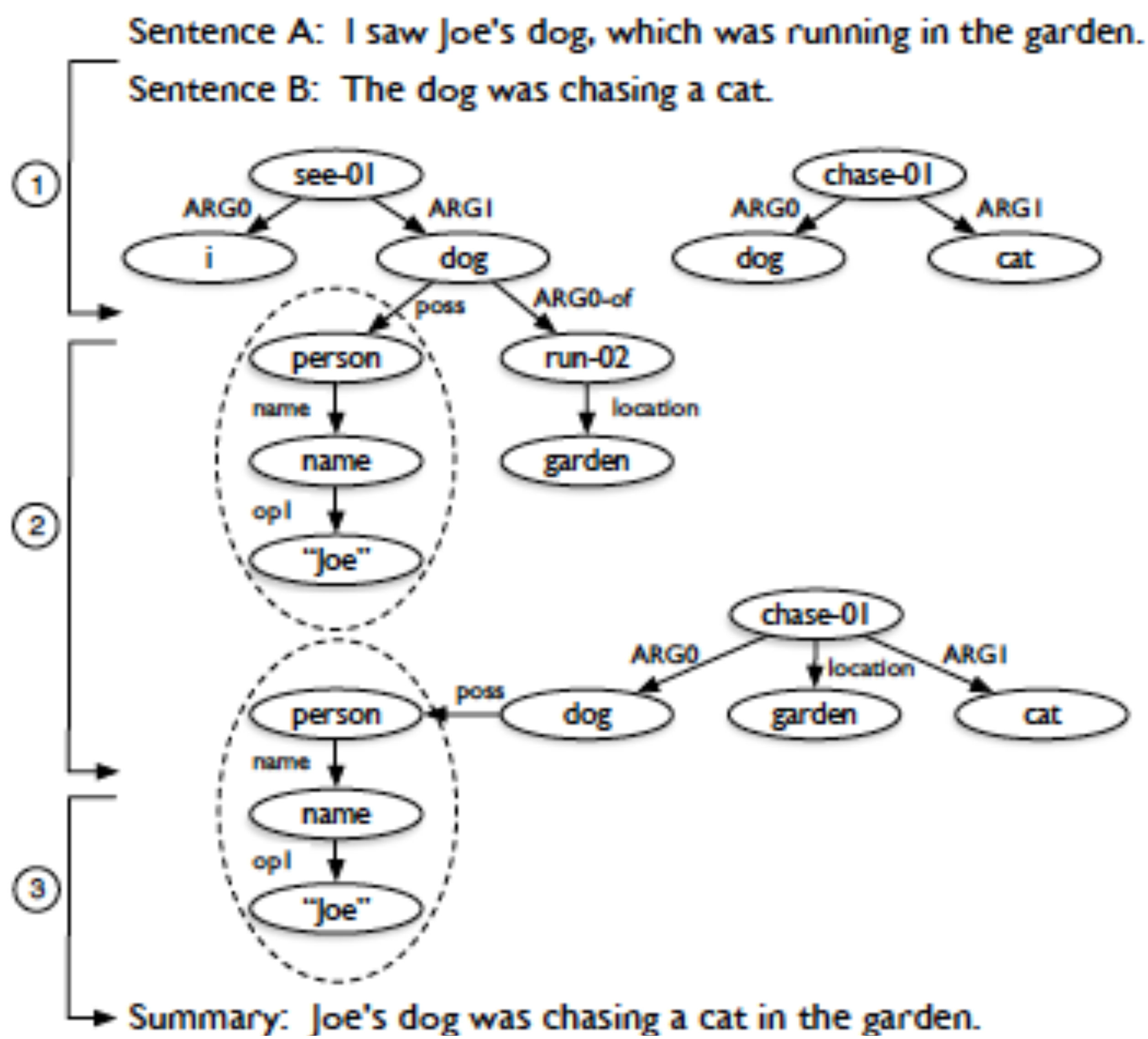


AMR-Based Summarization

- Use JAMR to parse input sentences to AMR

- Perform "concept merging" to link coref nodes
- Join sentence AMRs to dummy ROOT
- Create other connections as needed
- Select subset of nodes for inclusion in summary
- *Generate surface realization of AMR (future work)

Toy Example



Sentence A: I saw Joe's dog, which was running in the garden.
Sentence B: The dog was chasing a cat.
1
see-01
ARG0
ARG1
i
dog
chase-01
ARG0
ARG1
dog
cat
poss
ARG0-of
2
person
run-02
name
location
name
garden
op1
"Joe"
chase-01
ARG0
location
ARG1
poss
person
dog
garden
cat
name
name
3
op1
"Joe"
Summary: Joe's dog was chasing a cat in the garden.

Graph Creation

- Concept merging:
  - Constrained
    - Applied to Named entities & dates
    - Treat fragment as unary entity
    - Merge with identical nodes
      - Barak Obama = Barak Obama; Barak Obama ≠ Obama
    - Replace multiple edges b/t two nodes with unlabeled edge
- Compare to gold standard AMR edges for articles
  - “proxy report”
  - Cover 2/3 – 3/4

Content Selection

- Formulated as subgraph selection
  - Modeled as ILP
- Maximize the graph score (over edges, nodes)
  - Subject to:
    - Graph validity: edges must include endpoint nodes
    - Graph connectivity
    - Tree structure (one incoming edge/node)
    - Compression constraint (size of graph in edges)
- Features: Concept/label, frequency, depth, position,
  - Span, NE?, Date?

Evaluation

- Compare to gold-standard "proxy report"
  - Single-document summaries
  - Subgraph overlap with AMR
  - ROUGE-1 w/text
    - Generating most freq subspans associated w/fragments
- Results:
  - ROUGE-1: P: 0.5; R: 0.4; F: 0.44
    - Similar for manual AMR and auto parse
  - Oracle: P: 0.85; R: 0.44; F: 0.58
    - Based on similar bag-of-phrase generation from gold AMR

Summary

- Interesting strategy based on semantic represent'n
  - Builds on graph structure over deep model
  - Promising strategy
- Limitations:
  - Single-document
    - Does extension to multi-doc make sense?
  - Literal matching:
    - Reference, lexical content
  - Generation

Sentiment Summarization

- Classic approach: (Hu and Liu, 2004)
- Summarization of product reviews (e.g. Amazon)
  - Identify product features mentioned in reviews
  - Identify polarity of sentences about those features
  - For each product,
    - For each feature,
      - For each polarity: provide illustrative examples

Example Summary

- Feature: picture
  - Positive: 12
    - Overall this is a good camera with a really good picture clarity.
    - The pictures are absolutely amazing - the camera captures the minutest of details.
    - After nearly 800 pictures I have found that this camera takes incredible pictures.
    - …
- Negative: 2
  - The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
  - Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.

Learning Sentiment Summarization

- Classic approach is heuristic:
  - May not scale, etc.

- What do users want?

  - Which example sentences should be selected?

    - Strongest sentiment?

    - Most diverse sentiments?

    - Broadest feature coverage?

Review Summarization Factors

- Posed as optimizing score for given length summary
  - Using a sentence extractive strategy
- Key factors:
  - Sentence sentiment score
  - Sentiment mismatch: b/t summary and product rating
  - Diversity:
    - Measure of how well diff't "aspects" of product covered
    - Related to both quality of coverage, importance of aspect

Review Summarization Models

- Sentiment Match (SM): Neg(Mismatch)
  - Prefer summaries w/sentiment matching product
  - Issue?
    - Neutral rating ➔ neutral summary sentences
  - Approach: Force system to select stronger sents first
- Sentiment Match + Aspect Coverage (SMAC):
  - Linear combination of:
    - Sentiment intensity, mismatch, & diversity
  - Issue?
    - Optimizes overall sentiment match, but not per-aspect

Review Summarization Models

- Sentiment-Aspect Match (SAM):
  - Maximize coverage of aspects
    - *consistent* with per-aspect sentiment
  - Computed using probabilistic model
  - Minimize KL-divergence b/t summary, orig documents

Human Evaluation

- Pairwise preference tests for different summaries
  - Side-by-side, along with overall product rating
  - 1-4 symmetric preference
- Also collected comments that justify rating
- Usually some preference, but not significant
  - Except between SAM (better) and SMAC
  - And SMAC significantly better than LEAD baseline
    - (70% vs 25%)

Qualitative Comments

- Preferred:
  - Summaries with list (pro vs con)
- Disliked:
  - Summary sentences w/o sentiment
  - Non-specific sentences
  - Inconsistency b/t overall rating and summary
- Preferences differed depending on overall rating
  - Prefer SMAC for neutral vs SAM for extremes
    - (SAM excludes low polarity sentences)

Conclusions

- Ultimately, trained meta-classifier to pick model
  - Improved prediction of user preferences
- Similarities and contrasts w/TAC:
  - Similarities:
    - Diversity ~ Non-redundancy
    - Product aspects ~ Topic aspects: coverage, importance
  - Differences:
    - Strongly task/user oriented
    - Sentiment focused (overall, per-sentence)
    - Presentation preference: lists vs narratives