SIGDial 2012

# Unsupervised Topic Modeling Approaches to Decision Summarization in Spoken Meetings

Lu Wang and Claire Cardie
Department of Computer Science
Cornell University


Google Docs


skype

# Focused Meeting Summarization

- Meetings play an integral role in most of our daily lives — they let us share information and collaborate with others to solve a problem, to generate ideas, and to weigh options.

- Usually, we are interested in generating summaries of a particular aspect of a meeting.
  - **Decisions** — "*The remote will feature speech recognition.*"
  - **Action items** — "*The Marketing Expert will prepare a prototype evaluation.*"
  - **Problems** — "*Where to place the company slogan on the remote.*"

# Focused Meeting Summarization

- Meetings play an integral role in most of our daily lives — they let us share information and collaborate with others to solve a problem, to generate ideas, and to weigh options.

- Usually, we are interested in generating summaries of a particular aspect of a meeting.
  - ***Decisions*** — *"The remote will feature speech recognition."*
  - **Action items** — *"The Marketing Expert will prepare a prototype evaluation."*
  - **Problems** — *"Where to place the company slogan on the remote."*

# Decision Summarization in Spoken Meetings

| Decision-related Dialogue Acts (DRDA) |
| --- |
| A:We decided our target group is the focus on who can afford it , |
| B:Uh I'm kinda liking the idea of latex , if if spongy is the in thing . |
| B:what I've seen , just not related to this , but of latex cases before , is that [vocalsound] there's uh like a hard plastic inside , and it's just covered with the latex. |
| C:Um [disfmarker] And I think if we wanna keep our costs down , we should just go for pushbuttons , |
| D:but if it's gonna be in a latex type thing and that's gonna look cool , then that's probably gonna have a bigger impact than the scroll wheel . |
| A:we're gonna go with um type pushbuttons , |
| A:So we're gonna have like a menu button , |
| C:uh volume , favourite channels , uh and menu . |
| A:Pre-set channels |

***Decision-related Dialogue Acts (DRDA):***
The utterances support one or multiple decisions in the meeting. They usually contain the decision content.

# Decision Summarization in Spoken Meetings

| Decision-related Dialogue Acts (DRDA) |
| --- |
| A:We decided our target group is the focus on who can afford it , (1) |
| **B:Uh I'm kinda liking the idea of latex , if if spongy is the in thing . (2)** |
| **B:what I've seen , just not related to this , but of latex cases before , is that [vocalsound] there's uh like a hard plastic inside , and it's just covered with the latex . (2)** |
| C:Um [disfmarker] And I think if we wanna keep our costs down , we should just go for pushbuttons , (3) |
| **D:but if it's gonna be in a latex type thing and that's gonna look cool , then that's probably gonna have a bigger impact than the scroll wheel . (2)** |
| A:we're gonna go with um type pushbuttons , (3) |
| A:So we're gonna have like a menu button , (4) |
| C:uh volume , favourite channels , uh and menu . (4) |
| A:Pre-set channels (4) |

| Decision Abstracts (Summary) |
| --- |
| DECISION 1: The target group comprises of individuals who can afford the product. |
| **DECISION 2: The remote will have a latex case.** |
| DECISION 3: The remote will have pushbuttons. |
| DECISION 4: The remote will have a power button, volume buttons, channel preset buttons, and a menu button. |

Figure 1: A clip of a meeting from the AMI meeting corpus (Carletta et al., 2005). A, B, C and D refer to distinct speakers; the numbers in parentheses indicate the associated meeting decision: DECISION 1, 2, 3 or 4. Also shown is the gold-standard (manual) abstract (summary) for each decision.

# The Problem

- **Given a set of Decision-related Dialogue Acts (DRDAs), our system will output the summary for each decision made during the meeting.**

**Input**

| Decision-related Dialogue Acts (DRDA) |
|---|
| A:We decided our target group is the focus on who can afford it , (1) |
| **B:Uh I'm kinda liking the idea of latex , if if spongy is the in thing . (2)** |
| **B:what I've seen , just not related to this , but of latex cases before , is that [vocalsound] there's uh like a hard plastic inside , and it's just covered with the latex . (2)** |
| C:Um [disfmarker] And I think if we wanna keep our costs down , we should just go for pushbuttons , (3) |
| **D:but if it's gonna be in a latex type thing and that's gonna look cool , then that's probably gonna have a bigger impact than the scroll wheel . (2)** |
| A:we're gonna go with um type pushbuttons , (3) |
| A:So we're gonna have like a menu button , (4) |
| C:uh volume , favourite channels , uh and menu . (4) |
| A:Pre-set channels (4) |

**Output**

| Decision Abstracts (Summary) |
|---|
| DECISION 1: The target group comprises of individuals who can afford the product. |
| **DECISION 2: The remote will have a latex case.** |
| DECISION 3: The remote will have pushbuttons. |
| DECISION 4: The remote will have a power button, volume buttons, channel preset buttons, and a menu button. |

# The Challenges

| Decision-related Dialogue Acts (DRDA) |
| --- |
| A:We decided our target group is the focus on who can afford it , (1) |
| B:Uh I'm kinda liking the idea of latex , if if spongy is the in thing . (2) |
| B:what I've seen , just not related to this , but of latex cases before , is that [vocalsound] there's uh like a hard plastic inside , and it's just covered with the latex . (2) |
| C:Um [disfmarker] And I think if we wanna keep our costs down , we should just go for pushbuttons , (3) |
| D:but if it's gonna be in a latex type thing and that's gonna look cool , then that's probably gonna have a bigger impact than the scroll wheel . (2) |
| A:we're gonna go with um type pushbuttons , (3) |
| **A:So we're gonna have like a menu button , (4)** |
| **C:uh volume , favourite channels , uh and menu . (4)** |
| **A:Pre-set channels (4)** |

**Mostly, no single DRDA corresponds all that well with its decision abstract.**

- The three DRDAs in bold make the decision, "**The remote will have a power button, volume buttons, channel preset buttons, and a menu button.**" However, none of them cover all of the information.

# The Challenges

| Decision-related Dialogue Acts (DRDA) |
| --- |
| A:We decided our target group is the focus on who can afford it , (1) |
| B:Uh I'm kinda liking the idea of latex , if if spongy is the in thing . (2) |
| B:what I've seen , just not related to this , but of latex cases before , is that [vocalsound] there's uh like a hard plastic inside , and it's just covered with the latex . (2) |
| C:Um [disfmarker] And I think if we wanna keep our costs down , we should just go for pushbuttons , (3) |
| **D:but if it's gonna be in a latex type thing and that's gonna look cool , then that's probably gonna have a bigger impact than the scroll wheel . (2)** |
| A:we're gonna go with um type pushbuttons , (3) |
| A:So we're gonna have like a menu button , (4) |
| C:uh volume , favourite channels , uh and menu . (4) |
| A:Pre-set channels (4) |

**It is not easy to identify the core topic when multiple topics are discussed.**

- Besides "latex", the highlighted DRDA also mentions "bigger impact" and "the scroll wheel", which are not specifically relevant for the corresponding decision, i.e., "The remote will have a latex case."

# The Challenges

| Decision-related Dialogue Acts (DRDA) |
| --- |
| A:We decided our target group is the focus on who can afford it , (1) |
| B:Uh I'm kinda liking the idea of latex , if if spongy is the in thing . (2) |
| B:what I've seen , just not related to this , but of latex cases before , is that [vocalsound] there's uh like a hard plastic inside , and it's just covered with the latex . (2) |
| C:Um [disfmarker] And I think if we wanna keep our costs down , we should just go for pushbuttons , (3) |
| D:but if it's gonna be in a latex type thing and that's gonna look cool , then that's probably gonna have a bigger impact than the scroll wheel . (2) |
| A:we're gonna go with um type pushbuttons , (3) |
| **A:So we're gonna have like a menu button , (4)** |
| **C:uh volume , favourite channels , uh and menu . (4)** |
| **A:Pre-set channels (4)** |

**Complementary knowledge is needed from the context.**

- For DECISION 4 – " **The remote will have a power button, volume buttons, channel preset buttons, and a menu button.**", the "power button" is not specified in any of the listed DRDAs supporting it. However, it appears in one of the context utterance.

# Previous Work

- Utterance-Level meeting summarization
  - Supervised Learning
    - Maximum entropy, conditional random fields (CRFs), and support vector machines (SVMs) are investigated. (Buist et al., 2004; Galley, 2006; Xie et al., 2008)
  - Unsupervised Learning
    - Maximal marginal relevance (MMR) is studied in (Zechner, 2002) and (Xie and Liu,2010).
    - A concept-based global optimization framework by using integer linear programming (ILP) is proposed by (Gillick et al. (2009)).
    - Document-level topic models are applied for speech summarization. (Kong and shan Leek, 2006; Chen and Chen, 2008; Hazen, 2011).
- Token-level or phrase-level decision summarization
  - SVM is used to rank candidate phrases for inclusion of the decision summaries. (Fernandez et al., 2008) and (Bui et al., 2009)
  - SVM and CRF are explored in (Wang and Cardie, 2011).

# The Contribution of this Work

| | |
|---|---|
| Token-level summarization framework | • As a step towards creating the abstractive summaries, we propose a token-level rather than sentence-level framework. |
| Unsupervised topic modeling approaches | • We explore and evaluate topic modeling approaches to discover the topic structures of the utterances. |
| Exploration of context information | • We investigate the role of context in our token-level summarization framework. |

# *DomSum*:
# A Token-level Summarization Framework

- ***DomSum*** utilizes latent topic structure in utterances to extract words from the **Dom**inant Topic and form **Sum**maries.
- Input:
  - Clusters of dialogue acts
  - Topic structure for each DA: $P(T_i|DA)$
  - Word Distribution for each Topic: $P(w_k|T_i)$
- Algorithm
  - Step one: choose the dominant topic, i.e., $DomTopic = max_{T_i} P(T_i|DA)$
  - Step two: collect the words with a high joint probability with the dominant topic
    - $WordTopic = max_{T_i} P(w_k|T_i)P(T_i|DA)$
    - The word is collected , if WordTopic equals DomTopic

# Using *DomSum* to Leverage Context Information

- Two types of "Context"
  - Adjacent Dialogue Acts – immediately preceding and succeeding DAs of each DRDA
  - Similar Dialogue Acts – the DAs having top TF-IDF similarities with each DRDA

# Topic Modeling Approaches

- Local LDA **(LocalLDA**) (Brody and Elhadad, 2010)
  - It uses almost the same probabilistic generative model as Latent Dirichlet Allocation (LDA) (Blei et al., 2003), except that it treats each sentence as a separate document.
- Multi-grain LDA (**MG-LDA**) (Titov and McDonald, 2008)
  - It can model both the meeting specific topics (e.g. the design of a remote control) and various concrete aspects (e.g. the cost or the functionality).
- Segmented Topic Model (**STM**) (Du et al., 2010)
  - It jointly models document- and sentence-level latent topics using a two-parameter Poisson Dirichlet Process (PDP).

# Experimental Setup

- Dataset
  - AMI meeting corpus
  - For 129 scenario-driven meetings, a short abstract is manually constructed to summarize each decision discussed in the meeting.
  - Gold standard summaries are human-written abstracts.
- System Input
  - True Clusterings of DRDAs
  - System Clusterings of DRDAs
    - We use an existing hierarchical agglomerative clustering algorithm from (Wang and Cardie, 2011).
- Evaluation Metrics
  - ROUGE

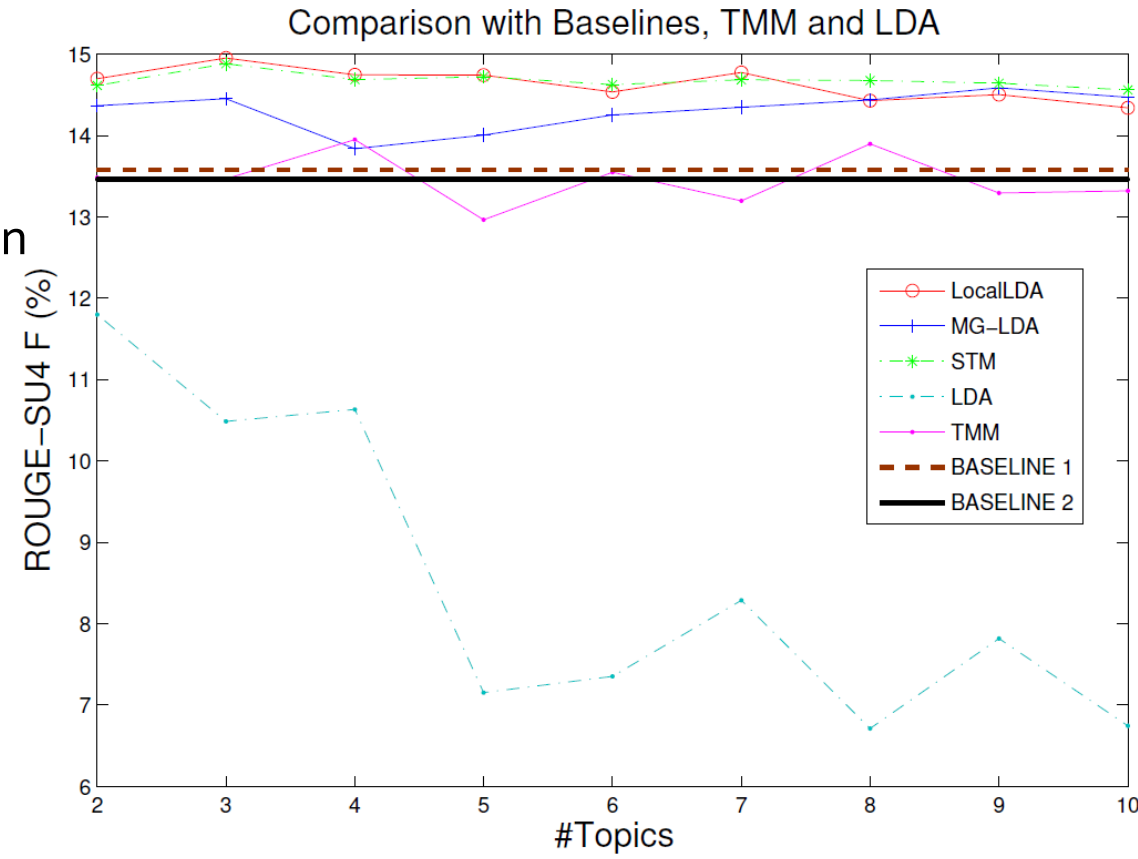
# Experimental Results

- How does the token-level summarization framework perform with fine-grained topic models?

Comparison:

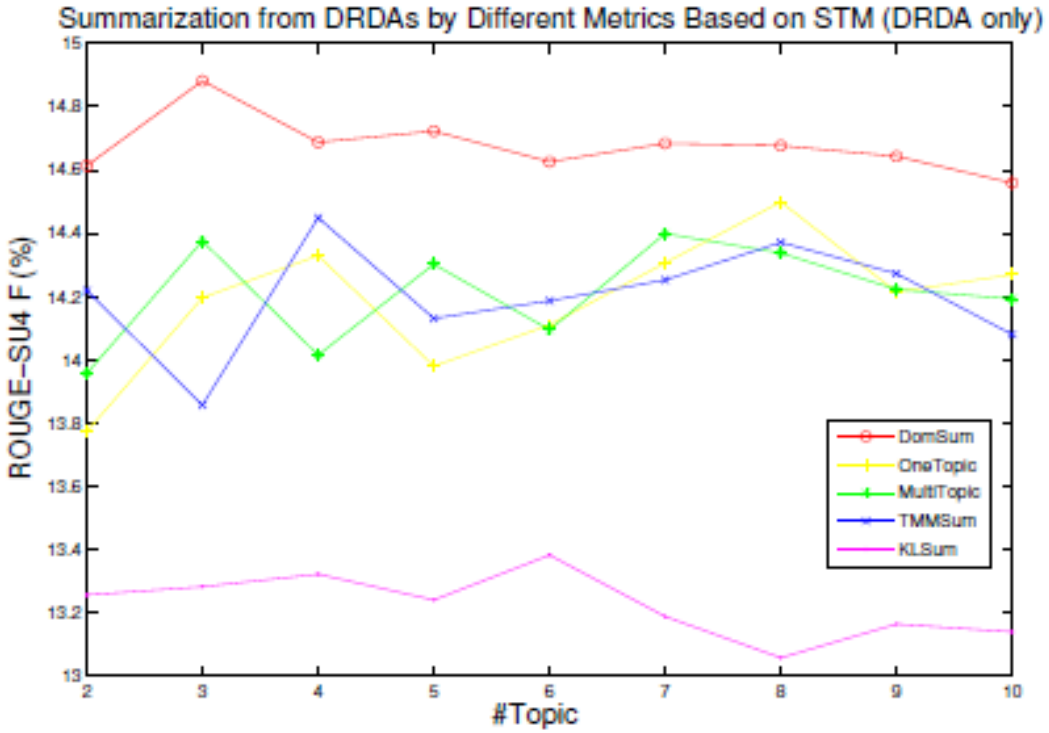
- Baseline 1: Longest DA
- Baseline 2: Prototype DA
- LDA: Latent Dirichlet Allocation
- TMM: Topical Mixture Model

(Chen and Chen, 2008)



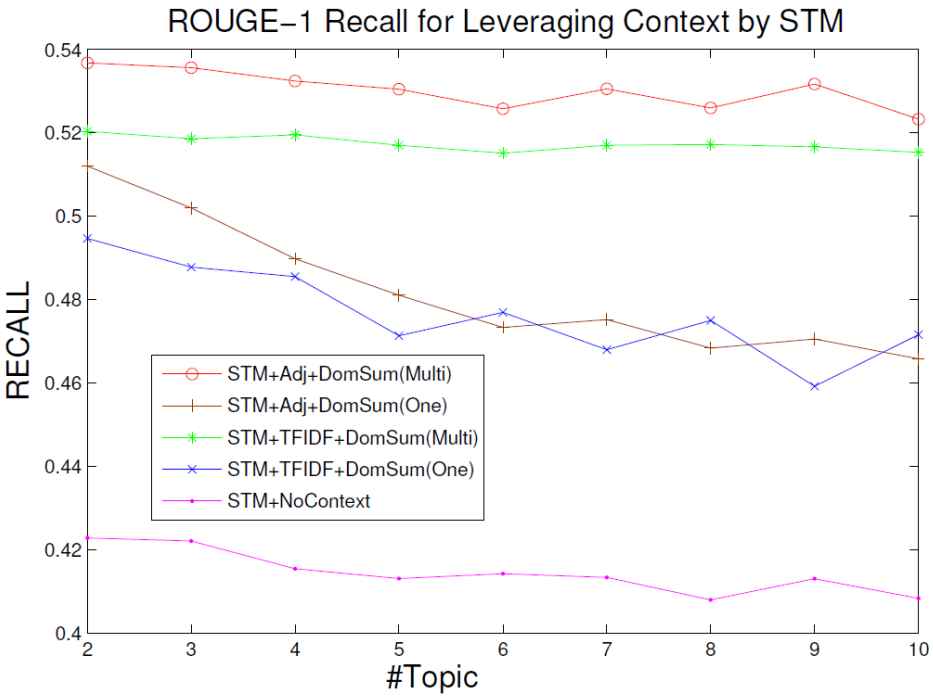
# Experimental Results

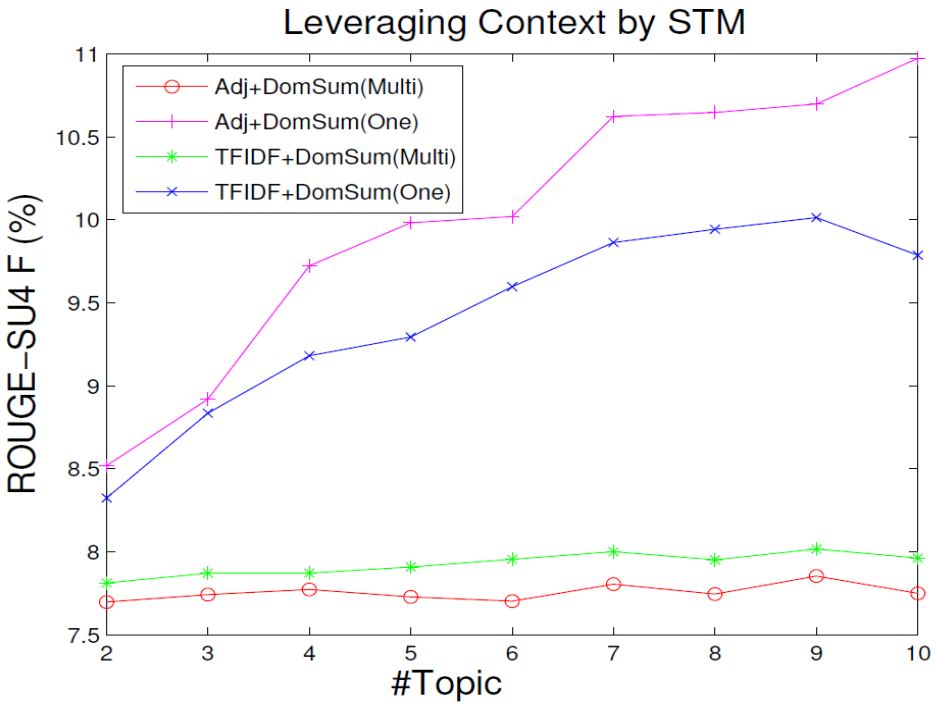
- Can the proposed token-level summarization framework better identify important words and remove redundancies than utterance selection methods?



# Experimental Results

- Can context information help with improve the summary? And which way is better for leveraging context information?



# Experimental Results

- How do our approach perform when compared with supervised learning approaches?

| | ROUGE-1 | | | ROUGE-2 | ROUGE-SU4 |
|---|---|---|---|---|---|
| | PREC | REC | F1 | F1 | F1 |
| **CRF** | 52.89 | 26.77 | 35.53 | 11.48 | 14.03 |
| **SVM** | 43.24 | 37.92 | 40.39 | 12.78 | 16.24 |
| **STM** | 34.06 | 41.30 | **37.32** | **12.42** | **14.82** |
| **Oracle** | 100.00 | ***45.05*** | 62.12 | 33.27 | 34.89 |

# Sample System Output

| |
|---|
| **DRDA (1)**: I think if we can if we can include them at not too much extra cost, then I'd put them in, |
| **DRDA (2)**: Uh um we we're definitely going in for voice recognition as well as LCDs, mm. |
| **DRDA (3)**: So we've basically worked out that we're going with a simple battery, |
| **context DA (1)**:So it's advanced integrated circuits? |
| **context DA (2)**:the advanced chip |
| **context DA (3)**: and a curved on one side case which is folded in on itself , um made out of rubber |

**Decision Abstract:** It will have voice recognition, use a simple battery, and contain an advanced chip.

| |
|---|
| **Longest DA**: Uh um we we're definitely going in for voice recognition as well as LCDs, mm. |
| **TMM**: I think if we can if we can include them at not too much extra cost, then I'd put them in, |
| **SVM**: cost, voice recognition, simple battery |
| **STM**: extra cost, definitely going voice recognition LCDs, simple battery |
| **STM + context**: cost, company, advanced integrated circuits, going voice recognition, simple battery, advanced chip, curved case rubber |

# Conclusion

- We propose a token-level summarization framework based on topic models.
- We show that modeling topic structure at the utterance-level is better at identifying summary-worthy words and phrases than document-level models.
- Context information can be leveraged to improve the summary.

# Thank you!