SIGDial 2012

Focused Meeting Summarization via Unsupervised Relation Extraction

Lu Wang and Claire Cardie
Department of Computer Science
Cornell University

"These lunch meetings are a lot more fun than conference calls."

Focused Meeting Summarization

- Usually, we are interested in generating summaries of the important outputs 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:***
  - Generating summaries for a particular aspect of a meeting rather than of the meeting as a whole.

Focused Summarization in Spoken Meetings

Decision-related Dialogue Acts (DRDA)
C: Say the standby button is quite kinda separate from all the other functions.
C: Maybe that could be a little apple.
C: It seems like you're gonna have rubber cases, as well as buttons.
A: Rubber buttons require rubber case.
A: You could have your company badge and logo.
A: I mean a lot of um computers for instance like like on the one you've got there, it actually has a sort of um stick on badge.
C: Shall we go for single curve, just to compromise?
B: We'll go for single curve, yeah.
C: And the rubber push buttons, rubber case.
D: And then are we going for sort of one button shaped like a fruit. <vocalsound> Or veg.
D: Could be a red apple, yeah.

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

Similarly, we can also have ***Action-related Dialogue Acts*** or ***Problem-related Dialogue Acts***.

Focused Summarization in Spoken Meetings

Decision-related Dialogue Acts (DRDA)
C: Say the standby button is quite kinda separate from all the other functions. (1)
C: Maybe that could be a little apple. (1)
C: It seems like you're gonna have rubber cases, as well as buttons. (2)
A: Rubber buttons require rubber case. (2)
A: You could have your company badge and logo. (3)
A: I mean a lot of um computers for instance like like on the one you've got there, it actually has a sort of um stick on badge. (3)
C: Shall we go for single curve, just to compromise? (2)
B: We'll go for single curve, yeah. (2)
C: And the rubber push buttons, rubber case. (2)
D: And then are we going for sort of one button shaped like a fruit. <vocalsound> Or veg. (1)
D: Could be a red apple, yeah. (1)

Decision Abstracts (Summary)
DECISION 1: The group decided to make the standby button in the shape of an apple.
DECISION 2: The remote will also feature a rubber case and rubber buttons, and a single-curved design.
DECISION 3: The remote will feature the company logo, possibly in a sticker form.

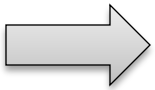
Figure 1: A clip of a meeting from the AMI meeting corpus (Carletta et al., 2005). A, B, C and D refer to distinct speakers; the numbers in parentheses indicate the associated meeting decision: DECISION 1, 2 or 3. Also shown is the gold-standard (manual) abstract (summary) for each decision.

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

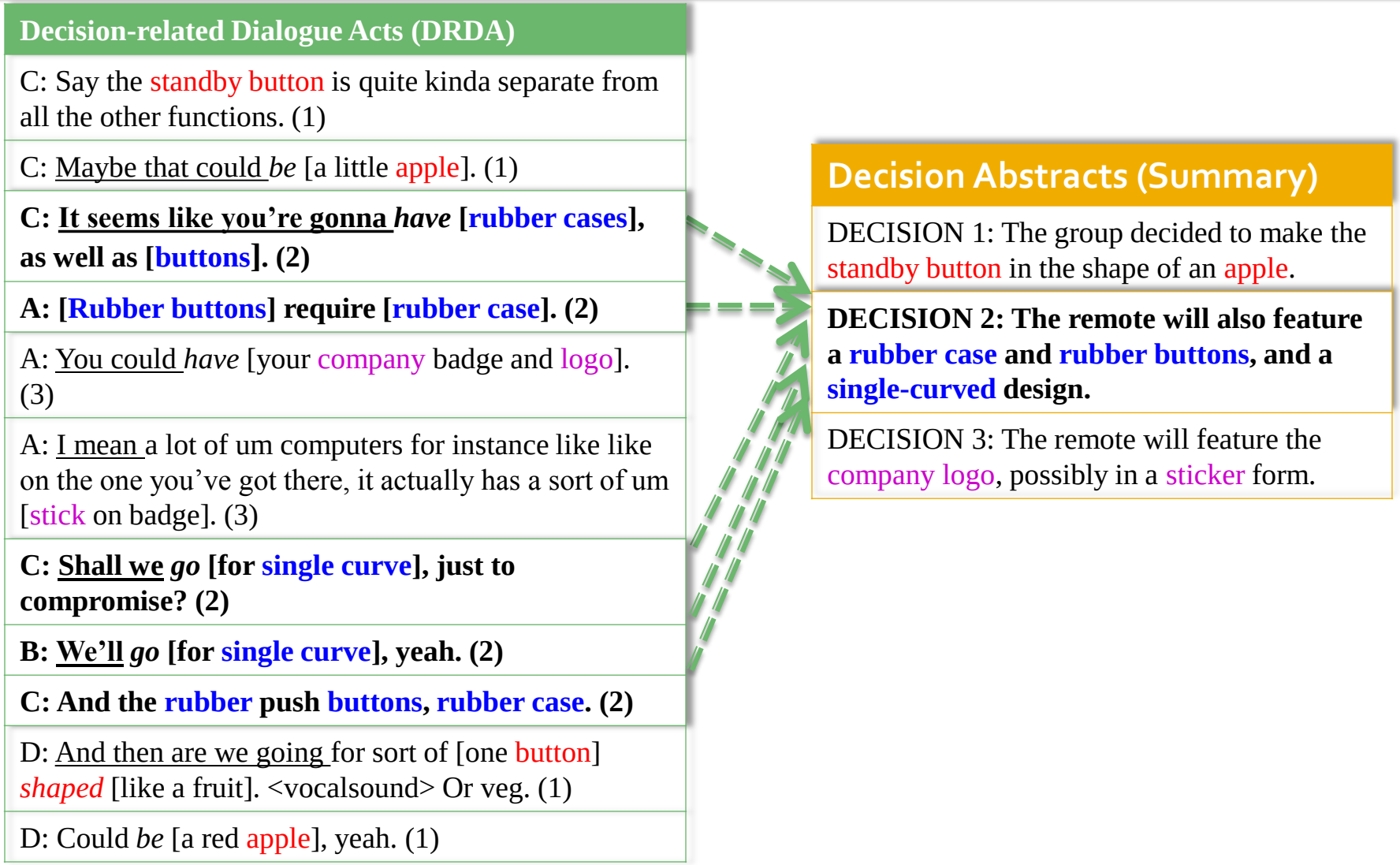
Decision-related Dialogue Acts (DRDA)
C: Say the standby button is quite kinda separate from all the other functions. (1)
C: Maybe that could be a little apple. (1)
C: It seems like you're gonna have rubber cases, as well as buttons. (2)
A: Rubber buttons require rubber case. (2)
A: You could have your company badge and logo. (3)
A: I mean a lot of um computers for instance like like on the one you've got there, it actually has a sort of um stick on badge. (3)
C: Shall we go for single curve, just to compromise? (2)
B: We'll go for single curve, yeah. (2)
C: And the rubber push buttons, rubber case. (2)
D: And then are we going for sort of one button shaped like a fruit. <vocalsound> Or veg. (1)
D: Could be a red apple, yeah. (1)



Output

Decision Abstracts (Summary)
DECISION 1: The group decided to make the standby button in the shape of an apple.
DECISION 2: The remote will also feature a rubber case and rubber buttons, and a single-curved design.
DECISION 3: The remote will feature the company logo, possibly in a sticker form.

Focused Summarization as Relation Extraction



Focused Summarization as Relation Extraction

Decision-related Dialogue Acts (DRDA)
C: Say the standby button is quite kinda separate from all the other functions. (1)
C: Maybe that could *be* [a little apple]. (1)
C: It seems like you're gonna *have* [rubber cases], as well as [buttons]. (2)
A: [Rubber buttons] require [rubber case]. (2)
A: You could *have* [your company badge and logo]. (3)
A: I mean a lot of um computers for instance like like on the one you've got there, it actually has a sort of um [stick on badge]. (3)
C: Shall we *go* [for single curve], just to compromise? (2)
B: We'll *go* [for single curve], yeah. (2)
C: And the rubber push buttons, rubber case. (2)
D: And then are we going for sort of [one button] *shaped* [like a fruit]. <vocalsound> Or veg. (1)
D: Could *be* [a red apple], yeah. (1)

Decision Abstracts (Summary)
DECISION 1: The group decided to make the standby button in the shape of an apple.
DECISION 2: The remote will also feature a rubber case and rubber buttons, and a single-curved design.
DECISION 3: The remote will feature the company logo, possibly in a sticker form.

Decision Cue:

DRDAs often begin with phrases that identify the utterance within the discourse as potentially introducing a decision, but do not themselves describe the decision.

- "Maybe that could", "It seems like you're gonna "

Focused Summarization as Relation Extraction

Decision-related Dialogue Acts (DRDA)
C: Say the standby button is quite kinda separate from all the other functions. (1)
C: Maybe that could *be* [a little apple]. (1)
C: It seems like you're gonna *have* [rubber cases], as well as [buttons]. (2)
A: [Rubber buttons] require [rubber case]. (2)
A: You could *have* [your company badge and logo]. (3)
A: I mean a lot of um computers for instance like like on the one you've got there, it actually has a sort of um [stick on badge]. (3)
C: Shall we *go* [for single curve], just to compromise? (2)
B: We'll *go* [for single curve], yeah. (2)
C: And the rubber push buttons, rubber case. (2)
D: And then are we going for sort of [one button] *shaped* [like a fruit]. <vocalsound> Or veg. (1)
D: Could *be* [a red apple], yeah. (1)

Decision Abstracts (Summary)
DECISION 1: The group decided to make the standby button in the shape of an apple.
DECISION 2: The remote will also feature a rubber case and rubber buttons, and a single-curved design.
DECISION 3: The remote will feature the company logo, possibly in a sticker form.

Decision Cue:

DRDAs often begin with phrases that identify the utterance within the discourse as potentially introducing a decision, but do not themselves describe the decision.

- "Maybe that could", "It seems like you're gonna "

Decision Content:

The part contains the decision content and should be considered for incorporation into the focused summary.

- "go for single curve", "one button shaped like a fruit"

Focused Summarization as Relation Extraction

Decision-related Dialogue Acts (DRDA)
C: Say the standby button is quite kinda separate from all the other functions. (1)
C: Maybe that could *be* [a little apple]. (1)
C: It seems like you're gonna *have* [rubber cases], as well as [buttons]. (2)
A: [Rubber buttons] require [rubber case]. (2)
A: You could *have* [your company badge and logo]. (3)
A: I mean a lot of um computers for instance like like on the one you've got there, it actually has a sort of um [stick on badge]. (3)
C: Shall we *go* [for single curve], just to compromise? (2)
B: We'll *go* [for single curve], yeah. (2)
C: And the rubber push buttons, rubber case. (2)
D: And then are we going for sort of [one button] *shaped* [like a fruit]. <vocalsound> Or veg. (1)
D: Could *be* [a red apple], yeah. (1)

We aim to identify the ***<indicator, [argument]>*** pairs, where the ***indicator*** evokes a relation of interest and the **[argument]** is the target phrase containing the object.

Previous Work

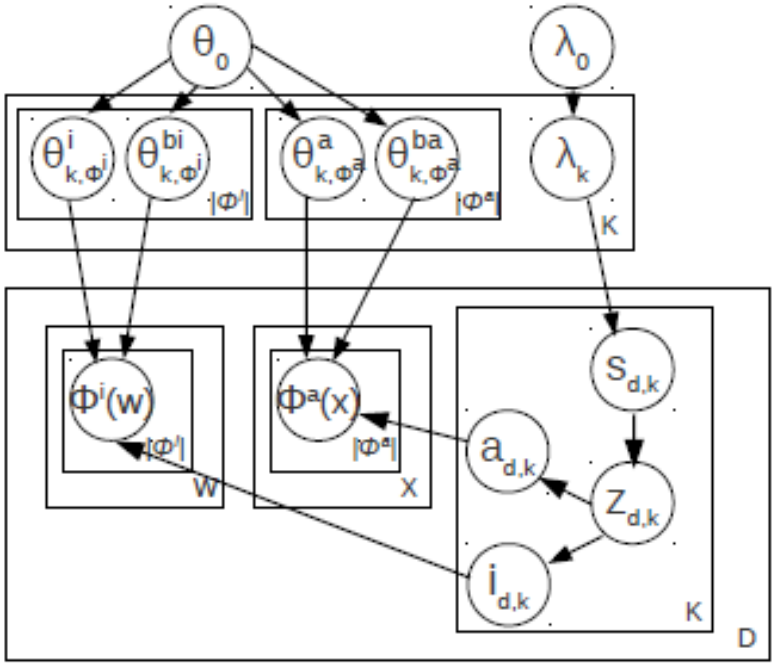
- SVM is used to rank candidate phrases for decision summaries. (Fernandez et al., 2008; Bui et al., 2009)
- Unsupervised and supervised approaches are explored in (Wang and Cardie, 2011) both on utterance-level and token-level decision summarization.
- (Hachey, 2009) uses relational representations to facilitate sentence-ranking for multi-document summarization.

The Contribution of this Work

A step towards abstracts	• As a step towards abstractive summarization, our system can identify those Content phrases that should comprise the abstract.
Information extraction task	• We hypothesize and show that existing methods for domain-specific relation extraction can be modified to identify salient relation instances for use in generating abstractive summaries.
Features and domain-independent constraints	• We define a new set of task-specific constraints and features.

The Model

- The probabilistic model for relation discovery



θ_k: **feature** distributions
λ_k: location distributions
ϕ^i: features for indicator
ϕ^a: features for argument

- Global preferences for the relation instances, such as the syntactic structure of the expressions or discourse behavior, are enforced by **constraints** and implemented by *posterior regularization*.

Features

Basic Features
unigram (stemmed)
part-of-speech (POS)
constituent label
dependency label
Meeting Features
Dialogue Act (DA) type
speaker role
topic
Structural Features
Adjacency Pairs
Semantic Features
first Synset of head word with the given POS
first hypernym path for the first synset of head word
Other Features (only for Argument)
number of words (without stopwords)
has capitalized word or not
has proper noun or not

Features

- Decision Cue relations

Decision Cue Relations	Relation Instances
Group Wrap-up / Recap	we have, we are, we say, we want
Personal Explanation	I mean, I think, I guess, I (would) say
Suggestion	do we, we (could/should) do
Final Decision	it is (gonna), it will, we will

Features

- Decision Cue relations

Decision Cue Relations	Relation Instances
Group Wrap-up / Recap	we have, we are, we say, we want
Personal Explanation	I mean, I think, I guess, I (would) say
Suggestion	do we, we (could/should) do
Final Decision	it is (gonna), it will, we will

- New discourse features

Discourse Features
clause position (first, second, other)
position to the first decision cue relation if any (before, after)

Constraints

- Syntactic Constraints
  - At least 80% of the induced relation instances are expected to match one of the our syntactic patterns.
    - The indicator is a verb and the argument is a noun phrase. The headword of the argument is the object of the indicator or the subject of the indicator.
    - The indicator is a verb and the argument is a prepositional phrase or a clause starting with "to". The indicator and the argument have the same parent in the constituent parsing tree.
    - The indicator is a noun and is the headword of a noun phrase, and the argument is a prepositional phrase. The noun phrase with the indicator as its headword and the argument have the same parent in the constituent parsing tree.

Constraints

- Prevalence Constraints
  - The prevalence constraint is enforced on the number of times a relation is instantiated.
- Occurrence Constraints
  - The diversity of relation types is enforced through occurrence constraints.
- Discourse Constraints
  - The discourse constraint captures the insight that the final decision is always made at the end of the decision-related discussion.

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

- Comparison
  - Baseline 1: Longest DA
  - Baseline 2: Prototype DA
  - Supervised methods: Conditional Random Fields, Support Vector Machines
  - Generic Relation Extraction (GRE) by (Hachey, 2009)

Experimental Results

	ROUGE-1			ROUGE-2	ROUGE-SU4
	PREC	REC	F1	F1	F1
Longest DA	34.06	31.28	32.61	12.03	13.58
Prototype DA	40.72	28.21	33.32	12.18	13.46
GRE (5 topics)	38.51	30.66	34.13	11.44	13.54
CRF	53.95	26.57	35.61	11.52	14.07
SVM	42.30	41.49	40.87	12.91	16.29
Our Method	37.94	**37.03**	**37.47**	**12.20**	**14.59**
Oracle	100.00	***45.05***	62.12	33.27	34.89

Sample System Output

DRDA (1): Uh the batteries, uh we also thought about that already,
DRDA (2): uh will be chargeable with uh uh an option for a mount station
DRDA (3): Maybe it's better to to include rechargeable batteries
DRDA (4): We already decided that on the previous meeting.
DRDA (5): which you can recharge through the docking station.
DRDA (6): normal plain batteries you can buy at the supermarket or retail shop. Yeah.

Decision Abstract: The remote will use rechargeable batteries which recharge in a docking station.

Longest DA: normal plain batteries you can buy at the supermarket or retail shop. Yeah.
GRE: normal plain batteries you can buy at the supermarket or retail shop. Yeah. which you can recharge through the docking station.
SVM: batteries include rechargeable batteries decided recharge docking station
Our Method: <option, for a mount station>, <include, rechargeable batteries>, <decided, that on the previous meeting>, <recharge, through the docking station>, <buy, normal plain batteries>

Conclusion

- We present a novel unsupervised framework for focused meeting summarization based on information extraction approach.
- Our method is shown to outperform unsupervised utterance-level extractive summarization baselines as well as an existing generic relation extraction-based summarization method.
- Our approach also produces summaries competitive with those generated by supervised methods in terms of the standard ROUGE score.
- The output of our system can be used as the input for the NLG system to generate abstractive summaries.

Thank you!