Domain-Independent Abstract Generation for Focused Meeting Summarization

Lu Wang and Claire Cardie
Department of Computer Science
Cornell University
Introduction

• Meetings are a common way to collaborate, share information and exchange opinions.
Introduction

• Automatic meeting summarization can facilitate:
  • Quick access to the essential output
  • Digest large amount of information easily
  • Project management
Focused Meeting Summarization

- Generating summaries of the important outputs of a meeting
- Decisions
- Action Items
- Problems
- Progress
Focused Meeting Summarization

**Dialogue Acts:**

C: Looking at what we've got, we want an LCD display with a spinning wheel.

B: You have to have some push-buttons, don't you?

C: Just spinning and not scrolling, I would say.

B: I think the spinning wheel is definitely very now.

A: but since LCDs seems to be uh a definite yes,

C: We're having push-buttons on the outside

C: and then on the inside an LCD with spinning wheel,

**Decision Summary:**
The remote will have push buttons outside, and an LCD and spinning wheel inside.
Related Work

- Existing meeting summarization systems are mostly extractive
  - Maximum entropy, conditional random fields (CRFs), and support vector machines (SVMs) are investigated for extracting utterances to make up a generic meeting summary (Buist et al., 2004; Galley, 2006; Xie et al., 2008).

  - Supervised methods are utilized to identify key phrases for inclusion in the decision summary (Fernandez et al., 2008; Bui et al., 2009)

- The need for abstractive summarization
  - Murray et al. (2010) show that users demonstrate a strong preference for abstractive summaries
Goal

Dialogue Acts:
C: Looking at what we've got, we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling, I would say.
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
C: We're having push-buttons on the outside
C: and then on the inside an LCD with spinning wheel,

Final Summary:
The remote will have push buttons outside, and an LCD and spinning wheel inside.
**Contribution**

- We propose a fully automatic domain-independent abstract generation framework for focused meeting summarization.

- We rely on task-specific templates to guide abstract generation and present a novel template extraction algorithm based on Multiple Sequence Alignment.

- We use an Overgenerate-and-Rank strategy for abstract surface realization.
Related Work

• Beyond extractive meeting summarization
  • Murray et al. (2010a) present an abstraction system which first maps utterances to the summary types (e.g. decision), and then selects the ones covering most entities.

• Sentence compression is studied to drop redundant words in Liu and Liu (2009).
Related Work

- Concept-to-text generation
  - The generation process is usually decomposed into content selection (or text planning) and surface realization (Angeli et al., 2010; Konstas and Lapata, 2012).

  - Angeli et al. (2010) learn from structured database records and parallel textual descriptions, and then generate texts based on a series of decisions made to select the records, fields, and proper templates for rendering.
Dialogue Acts:
C: Looking at what we've got, we we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling, I would say.
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
Framework

Dialogue Acts:
C: Looking at what we've got, we we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling, I would say.
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
Dialogue Acts:
C: Looking at what we've got, we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling, I would say.
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
Dialogue Acts:
C: Looking at what we've got, we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling, I would say.
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,

Relation Extraction:
<want, an LCD display with a spinning wheel>
<have, some push-buttons>
<having, push-buttons on the outside>

Content Selection:

Template Filling:
<want, an LCD display with a spinning wheel>
- The team will want an LCD display with a spinning wheel.
- The team will work with an LCD display with a spinning wheel.

Statistical Ranking:
One-Best Abstract:
The group decide to use an LCD display with a spinning wheel.

Post-processing:
Final Summary:
The group decide to use an LCD display with a spinning wheel. There will be push-buttons on the outside.
Dialogue Acts:
C: Looking at what we've got, we we *want* [an LCD display with a spinning wheel].
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling, I would say.
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
C: We're having *push-buttons [on the outside]*
C: and then on the inside an LCD with spinning wheel,

Relation Instance:
<indicator, [argument]> pairs, where the *indicator* evokes a relation of interest and the *[argument]* is the target phrase containing the object.
Content Selection

Dialogue Acts:
C: Looking at what we've got, we *want* [an LCD display with a spinning wheel].
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling, I would say.
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
C: We're having *push-buttons* [on the outside]
C: and then on the inside an LCD with spinning wheel,

Relation Instance:

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

Sample Summary:
The group decided to use an LCD display with a spinning wheel. There will be *push-buttons* on the outside.
Content Selection via Relation Extraction

- Relation Instance Candidates:
  - **Indicator**: noun or verb
  - **Argument**: noun phrase (NP), prepositional phrase (PP) or adjectival phrase (ADJP).

- Constraints:

```

```
Content Selection via Relation Extraction

- Binary classifier based on Support Vector Machines
  - Summary-worthy vs. not summary-worthy

- Sample Features:

<table>
<thead>
<tr>
<th>Basic Features</th>
<th>Discourse Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF/IDF/TFIDF</td>
<td>Main speaker or not?</td>
</tr>
<tr>
<td>Constituent tag</td>
<td>Shared by adjacency pair?</td>
</tr>
<tr>
<td>Dependency relation</td>
<td>Target DA is a positive response?</td>
</tr>
</tbody>
</table>
Framework

Dialogue Acts:
C: Looking at what we've got, we we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling, I would say.
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,

Relation Extraction

Relation Instances:
<want, an LCD display with a spinning wheel>
<have, some push-buttons>
<having, push-buttons on the outside>
… (other possibilities)

Content Selection

Template Filling

Template Filling

<want, an LCD display with a spinning wheel>
• The team will want an LCD display with a spinning wheel.
• The team will work with an LCD display with a spinning wheel.
… (other possibilities)

Statistical Ranking

One-Best Abstract:
The group decide to use an LCD display with a spinning wheel.

Surface Realization

Post-processing

Final Summary:
The group decide to use an LCD display with a spinning wheel. There will be push-buttons on the outside.
Extracting Summary Templates

- Multiple Sequence Alignment (MSA)
  - MSA is commonly used in bioinformatics to identify equivalent fragments of DNAs (Durbin et al., 1998) and has also been employed for learning paraphrases (Barzilay and Lee, 2003)
Multiple Sequence Alignment (MSA)

\[
\begin{align*}
S1 &= T A T A G C T \\
S2 &= T T A G C T \\
S3 &= T A T C G C \\
S4 &= A C A C C T
\end{align*}
\]

<table>
<thead>
<tr>
<th></th>
<th>S1'</th>
<th>T</th>
<th>A</th>
<th>T</th>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2'</td>
<td>T</td>
<td>-</td>
<td>T</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>S3'</td>
<td>T</td>
<td>-</td>
<td>-</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>S4'</td>
<td>-</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td></td>
</tr>
</tbody>
</table>
Multiple Sequence Alignment (MSA)

\[
S1 = \text{T A T A G C T} \\
S2 = \text{T T A G C T} \\
S3 = \text{T A T C G C} \\
S4 = \text{A C A C C T} \\
\]

Scorer(x₁, x₂) = \[
\begin{cases} 
1, & \text{if } x₁ = x₂ \\
0, & \text{if } x₁ = “-” \text{ or } x₂ = “-” \\
-1, & \text{if } x₁ \neq x₂
\end{cases}
\]

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>A</th>
<th>T</th>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>S₁’</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>S₂’</td>
<td>T</td>
<td>-</td>
<td>T</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>S₃’</td>
<td>T</td>
<td>-</td>
<td>-</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>-</td>
</tr>
<tr>
<td>S₄’</td>
<td>-</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>T</td>
</tr>
</tbody>
</table>
Extracting Summary Templates

• We first cluster the abstracts in the training data according to their lexical and structural similarity.
  • We first replace all appearances of dates, numbers, and proper names with generic semantic labels.
  • We further replace sequences that appear in both the abstract and supporting dialogue acts by a label indicating its phrase type.
  • Following Barzilay and Lee (2003), we approach the sentence clustering task by hierarchical clustering with a similarity metric based on word n-gram overlap.
Extracting Summary Templates

1) The group were not sure whether to include a recharger for the remote.
2) The group were not sure whether to use plastic and rubber or titanium for the case.
3) The group were not sure whether the remote control should include functions for controlling video.
4) They were not sure how much a recharger would cost to make.

1) The group were not sure whether to [include]_{VP} [a recharger for the remote]_{NP}.
2) The group were not sure whether to use [plastic and rubber or titanium for the case]_{NP}.
3) The group were not sure whether [the remote control]_{NP} should include [functions for controlling video]_{NP}.
4) They were not sure how much [a recharger]_{NP} would cost to make.

1) The group were not sure whether to \textit{VP} \textit{NP}.
2) The group were not sure whether to use \textit{NP}.
3) The group were not sure whether \textit{NP} should include \textit{NP}.
4) They were not sure how much \textit{NP} would cost to make.
1) The group were not sure whether to include a recharger for the remote.
2) The group were not sure whether to use plastic and rubber or titanium for the case.
3) The group were not sure whether the remote control should include functions for controlling video.
4) They were not sure how much a recharger would cost to make.

1) The group were not sure whether to [include]VP [a recharger for the remote]NP.
2) The group were not sure whether to use [plastic and rubber or titanium for the case]NP.
3) The group were not sure whether [the remote control]NP should include [functions for controlling video]NP.
4) They were not sure how much [a recharger]NP would cost to make.

1) The group were not sure whether to VP NP.
2) The group were not sure whether to use NP.
3) The group were not sure whether NP should include NP.
4) They were not sure how much NP would cost to make.
Extracting Summary Templates

1) The group were not sure whether to include a recharger for the remote.
2) The group were not sure whether to use plastic and rubber or titanium for the case.
3) The group were not sure whether the remote control should include functions for controlling video.
4) They were not sure how much a recharger would cost to make.

1) The group were not sure whether to include [a recharger for the remote].
2) The group were not sure whether to use [plastic and rubber or titanium for the case].
3) The group were not sure whether [the remote control] should include [functions for controlling video].
4) They were not sure how much [a recharger] would cost to make.

1) The group were not sure whether to [include] VP NP.
2) The group were not sure whether to use [plastic and rubber or titanium for the case] NP.
3) The group were not sure whether [the remote control] NP should include [functions for controlling video] NP.
4) They were not sure how much [a recharger] NP would cost to make.
Extracting Summary Templates

• Multiple Sequence Alignment
  • Computing an optimal MSA is NP-complete (Wang and Jiang, 1994).
  • We implement an approximate algorithm (Needleman and Wunsch, 1970) that iteratively aligns two sequences each time and treats the resulting alignment as a new sequence.
1) The group were not sure whether to use VP NP.
2) The group were not sure whether to use NP.
3) The group were not sure whether NP should include NP.
4) They were not sure how much NP would cost to make.

The backbone nodes shared by at least 50% sentences are shaded.
Extracting Summary Templates

**Template Examples:**

T1: *The group were not sure whether to* VP NP.

T2: *The group were not sure whether* NP VP VP NP.

T3: *NP were not sure* WHADJP WHADJP NP VP VP VP VP VP.
**Dialogue Acts:**
C: Looking at what we've got, we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling, I would say.
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
Surface Realization

• An Overgenerate-and-Rank Approach
  • Overgenerate-and-Rank strategy has been used for sentence planning (Walker et al., 2001) and questions generation (Heilman and Smith, 2010).

Template Filling

<want, an LCD display with a spinning wheel>
• The team will want an LCD display with a spinning wheel.
• The team will work with an LCD display with a spinning wheel.
  … (other possibilities)

Statistical Ranking: 
One-Best Abstract: 
The group decide to use an LCD display with a spinning wheel.

Post-processing:
Final Summary:
The group decide to use an LCD display with a spinning wheel. There will be push-buttons on the outside.
Template Filling

- The templates are represented by their parse tree.

- We then fill the slots with relation instances.
Template Filling

- Constituent-level transformation of the relation instances
Template Filling

Template:

NP

The /DT

team /NN

will /MD

S \n
VP \n
NP

Relation instance:

Indicator: [VB want]
Argument: [NP an LCD display [PP with a spinning wheel ] ]
Full- Constituent Mapping

Template:

S

NP

The team

VP

will [VB want]

[NP an LCD display [PP with a spinning wheel ] ]

Relation instance:

Indicator: [VB want]
Argument: [NP an LCD display [PP with a spinning wheel ] ]
Template:

```
S
   \  \  \\
NP    VP
       / \  \  \\
  The team will [VP want] [NP an LCD display]
```

Relation instance:

Indicator: [VB want]
Argument: [NP an LCD display [PP with a spinning wheel]]
Removal

Template:

```
S
   VP
      NP
         The /DT
         team /NN
         will /MD
         [NP an LCD display [PP with a spinning wheel ] ]
```

Relation instance:

Indicator: [VB want]
Argument: [NP an LCD display [PP with a spinning wheel ] ]
Template Filling

**Templates:**

T1: *The team will* VP NP.
T2: *The team will work with* NP.

... 

**Relation Instances:**

<want, an LCD display with a spinning wheel>

• *The team will* want an LCD display with a spinning wheel.
• *The team will work with* an LCD display with a spinning wheel.

... (other possibilities)
Statistical Ranking

- Support Vector Regression (SVR) (Smola and Scholkopf, 2004) is utilized to rank the generated abstracts.

- Sample Features

<table>
<thead>
<tr>
<th>Basic Features</th>
<th>Structure Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of words</td>
<td>Constituent tag for ind/arg</td>
</tr>
<tr>
<td>Number of new nouns</td>
<td>Dependency relation for ind/arg</td>
</tr>
<tr>
<td>Ind/Arg has content words?</td>
<td>The relation matches the template?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Realization Features</th>
<th>Language Model Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract has verb?</td>
<td>log $p_{LM}(\text{abstract})$</td>
</tr>
<tr>
<td>Abstract starts with verb?</td>
<td>Bigram probability of ind/arg</td>
</tr>
<tr>
<td>Abstract has adjacent verb/adj?</td>
<td>Trigram probability of ind/arg</td>
</tr>
</tbody>
</table>
Redundancy Handling

• Post-selection aims to maximize the information coverage and minimize the redundancy of the summary.

• We use a greedy algorithm (Lin and Bilmes, 2010) to select a subset of the generated abstracts.
Experimental Setup

- Two disparate corpora
  - The AMI meeting corpus (McCowan et al., 2005)
    - 139 scenario-driven meetings
    - Contains abstracts for decisions, action items, and problems
  - The ICSI meeting corpus (Janin et al., 2003)
    - 75 naturally occurring meetings
    - Contains abstracts for decisions, progress, and problems

- System Input
  - True/System clusterings of summary related DAs
Content Selection Results

• Evaluation metric
  • ROUGE (Lin and Hovy, 2003) -- computes the ngram overlapping between the system summaries and the reference summaries.

  • It has been widely used for text and speech summarization (Dang, 2005; Xie et al., 2008).

• Comparison:
  • Baselines: Longest DA, Centroid DA
  • Supervised Learning: Utterance-level and token-level extractive summarization with SVMs (Xie et al., 2008; Sandu et al., 2010; Fernandez et al., 2008)
Content Selection Results

Content Selection for Action Item (AMI)

ROUGE-SU4 F1

Number of Meetings

Oracle
Our System
Centroid DA
Longest DA
SVM-DA
SVM-TOKEN
Content Selection Results

Content Selection for Decision (AMI)

ROUGE-SU4 F1

Number of Meetings
Content Selection Results

Content Selection for Decision (ICSI)

- **Oracle**
- **Our System**
- **Centroid DA**
- **Longest DA**
- **SVM-DA**
- **SVM-TOKEN**

**ROUGE-SU4 F1**

**Number of Meetings**

- 8
- 15
- 23
- 30
- 38
- 45
Abstract Generation Results

- Evaluation metrics
  - BLEU (Papineni et al., 2002) – computes the precision of unigrams and bigrams with a brevity penalty
    - has been used to evaluate a variety of language generation systems (Angeli et al., 2010; Konstas and Lapata, 2012).

- Human evaluation
  - Fluency -- is the text grammatical?
  - Semantic correctness -- does the summary convey the gist of the DAs in the cluster?
  - Overall quality in content, conciseness and grammaticality.
Abstract Generation Results

Full System for Action Item (AMI)

Number of Meetings

BLEU

18 35 53 70 87 104

Oracle
Our System
SVM-DA
Centroid DA
Longest DA
Abstract Generation Results

Full System for Decision (AMI)

- Oracle
- Our System
- SVM-DA
- Centroid DA
- Longest DA

BLEU vs. Number of Meetings
Abstract Generation Results

Full System for Decision (ICSI)

- **Oracle**
- **Our System**
- **SVM-DA**
- **Centroid DA**
- **Longest DA**

Bleu vs. Number of Meetings
<table>
<thead>
<tr>
<th>System</th>
<th>AMI Decision</th>
<th>ICSI Decision</th>
<th>AMI Problem</th>
<th>ICSI Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R-SU4</td>
<td>BLEU</td>
<td>R-SU4</td>
<td>BLEU</td>
</tr>
<tr>
<td>SVM-DA (In-Domain)</td>
<td>4.7</td>
<td>9.7</td>
<td>4.5</td>
<td>5.7</td>
</tr>
<tr>
<td>Our system (In-Domain)</td>
<td>6.2</td>
<td>11.6</td>
<td>7.1</td>
<td>10.0</td>
</tr>
<tr>
<td>Our system (Out-of-Domain)</td>
<td>6.1</td>
<td>10.3</td>
<td>6.4</td>
<td>7.8</td>
</tr>
<tr>
<td>Oracle</td>
<td>12.0</td>
<td>22.8</td>
<td>14.9</td>
<td>20.2</td>
</tr>
<tr>
<td>System</td>
<td>AMI Decision</td>
<td>ICSI Decision</td>
<td>AMI Problem</td>
<td>ICSI Problem</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------</td>
<td>---------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td></td>
<td>R-SU4</td>
<td>BLEU</td>
<td>R-SU4</td>
<td>BLEU</td>
</tr>
<tr>
<td>SVM-DA (In-Domain)</td>
<td>4.7</td>
<td>9.7</td>
<td>4.5</td>
<td>5.7</td>
</tr>
<tr>
<td>Our system (In-Domain)</td>
<td>6.2</td>
<td>11.6</td>
<td>7.1</td>
<td>10.0</td>
</tr>
<tr>
<td>Our system (Out-of-Domain)</td>
<td>6.1</td>
<td>10.3</td>
<td>6.4</td>
<td>7.8</td>
</tr>
<tr>
<td>Oracle</td>
<td>12.0</td>
<td>22.8</td>
<td>14.9</td>
<td>20.2</td>
</tr>
</tbody>
</table>
## Human Evaluation

<table>
<thead>
<tr>
<th>System</th>
<th>Fluency</th>
<th>Semantic Correctness</th>
<th>Avg. Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Our system (In-Domain)</td>
<td>3.67</td>
<td>0.85</td>
<td>3.27</td>
</tr>
<tr>
<td>Our system (Out-of-Domain)</td>
<td>3.58</td>
<td>0.90</td>
<td>3.25</td>
</tr>
<tr>
<td>SVM-DA (In-Domain)</td>
<td>3.36</td>
<td>0.84</td>
<td>3.44</td>
</tr>
</tbody>
</table>
## Human Evaluation

<table>
<thead>
<tr>
<th>System</th>
<th>Fluency</th>
<th>Semantic Correctness</th>
<th>Avg. Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Our system (In-Domain)</td>
<td>3.67</td>
<td>0.85</td>
<td>3.27</td>
</tr>
<tr>
<td>Our system (Out-of-Domain)</td>
<td>3.58</td>
<td>0.90</td>
<td>3.25</td>
</tr>
<tr>
<td>SVM-DA (In-Domain)</td>
<td>3.36</td>
<td>0.84</td>
<td>3.44</td>
</tr>
</tbody>
</table>

Judges selected our system as the best system in 62.3% scenarios (In-Domain: 35.6%, Out-of-Domain: 26.7%)
Sample Summary

Dialogue Acts:
C: Looking at what we've got, we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling, I would say.
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
C: We're having push-buttons on the outside
C: and then on the inside an LCD with spinning wheel,

Decision Summaries:
Human: The remote will have push buttons outside, and an LCD and spinning wheel inside.

Our System (In-Domain): The group decided to use an LCD display with a spinning wheel. There will be push-buttons on the outside.

Our System (Out-of-Domain): LCD display is going to be with a spinning wheel. It is necessary having push-buttons on the outside.
Conclusion

• We presented a domain-independent abstract generation framework for focused meeting summarization.

• Experimental results on two disparate meeting corpora show that our system can uniformly outperform the state-of-the-art supervised extraction-based systems in both automatic and manual evaluation.

• Our system also exhibits an ability to train on out-of-domain data to generate abstracts for a new target domain.
Thank you!