Automatic documentation of results during online architectural meetings

Oleksandra Klymenko, 17.06.2019, Garching
Motivation

- **Communication** between team members is crucial for efficient software development work
- In distributed teams, **most of the meetings happen online**
- Capturing and explicitly documenting decisions **enables reasoning and decision support** [1]
- Manual effort, time and cost of explicit documentation is a **concern for practitioners**
- Previous research has mostly focused on detecting decisions in **issue management systems** and **source code commits**
- However, many decisions are implicitly made in **online meetings**
- The most frequent form of documentation of architectural design decisions is **meeting minutes** [2]
- **Virtual Online Assistant** can help to document, review and refer back to made decisions
<table>
<thead>
<tr>
<th>RQ 1</th>
<th>How are online meetings between software development professionals held in practice and how are they documented?</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ 2</td>
<td>What is the process of decision-making in distributed software development teams?</td>
</tr>
<tr>
<td>RQ 3</td>
<td>What are the requirements for a system that automatically documents online architectural meetings?</td>
</tr>
<tr>
<td>RQ 4</td>
<td>How to identify, extract and document design decisions in online architectural meetings?</td>
</tr>
</tbody>
</table>
Audio stream
Audio data
Speech to Text
Transcript
Decision Detection
PDF
Architectural meeting summary
Virtual Meeting Assistant
Reflection
Idea

17.06.2019 - Master's thesis of Oleksandra Klymenko - Final Presentation
Research Approach

1. Literature review

2. Design and conduct interviews (RQ1, RQ2, RQ3)
   - Transcribe interview recordings
   - Analyze feedback and elicit requirements

3. Data collection and analysis
   - Record and transcribe meetings
   - Label the data

4. Technical implementation (RQ4)
   - Rasa NLU

5. Evaluation
Interview phases

Questions about expertise (2 questions)
Questions about project organization (7 questions)
Questions about current challenges (5 questions)
Questions about personal opinion on the system (9 questions)

Planning

- Cooperation with UXD and RE departments
- Semi-structured interview
- 23 open questions
- 10 interviewees
- Mostly senior architects and product owners
- Ø 13 years of experience in IT industry
- Planned duration of the interview: 30 minutes, without interruptions
- Question catalog was not provided to the interviewees in advance or during the interview
Interview

Goal:
To understand participants’:
• Current challenges faced with Circuit
• View of an Assistive Bot during a Virtual meeting Scenario

Analysis:
• Creating transcripts
• Coding transcripts
• Consolidating the list of requirements

Results:
Obtained expert feedback concerning:
• Current online meeting process
• Decision-making process
• The proposed use cases
• Requirements for the bot
• Usefulness of an automatic summary
• Other ideas
Requirements for the system

**Things to include in the summary:**

- Action items / TODOs
  - Assigned tasks
- The person who has been assigned
  - Who brought something up
  - The person who made the decision
- Deadline
  - Open topics (things that need follow-up)
- Catch words
  - Come-in / Drop-out times
  - Information / News (e.g. news from management)
  - Participants’ telephone numbers

**Timing of bot’s questions:**

- Immediate notifications vs. waiting for the end of the call
- Optimal solution: combination of both approaches

**Other ideas**

- Anonymous notification when discussion goes off-topic
Implementation

Natural Language Understanding

• Rasa NLU
• spaCy pipeline
• ner_crf

Data Corpus

• Many options considered
• Only real internal meetings used
• 17 meeting recordings
• > 620 minutes of meetings
• 129 training examples in total

Concept Extraction

• Linked Data
• DBPedia
"intent": "decision",
"entities": [
{
"start": 113,
"end": 177,
"value": "we can experiment with the whole of the country credibility part",
"entity": "s"
},
{
"start": 314,
"end": 364,
"value": "we have to think of a strategy of integrating that",
"entity": "d"
}
],
"text": "And if we go to the next you did with the next big in here the first one I think to begin with I would still say we can experiment with the whole of the country credibility part as well with the existing dashboard as a service that is being developed in the sense that For me the complexity still seems to be that we have to think of a strategy of integrating that."
Demo
ASR Evaluation

• Word Error Rate (WER)

\[
WER = \frac{S+D+I}{N} = \frac{S+D+I}{S+D+C}
\]

• S is the number of substitutions
• D is the number of deletions
• I is the number of insertions
• C is the number of correctly transcribed words
• N is the total number of words originally spoken (N=S+D+C)

<table>
<thead>
<tr>
<th>ASR tool</th>
<th>Speechmatics v.2.0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substitutions (S)</td>
<td>208</td>
</tr>
<tr>
<td>Deletions (D)</td>
<td>110</td>
</tr>
<tr>
<td>Insertions (I)</td>
<td>27</td>
</tr>
<tr>
<td>Correct words (C)</td>
<td>1014</td>
</tr>
<tr>
<td>Total number of words (N)</td>
<td>1333</td>
</tr>
<tr>
<td>Word Error Rate (WER)</td>
<td>26%</td>
</tr>
</tbody>
</table>

Table 1: Evaluation of Automatic Speech Recognition
Model Evaluation

- 15 meetings that contained decisions
- 5-fold validation
- Precision, recall, F1-score

\[
R = \frac{TP}{TP+FN}, \quad P = \frac{TP}{TP+FP}, \quad F = 2 \times \frac{P \times R}{P + R}
\]

<table>
<thead>
<tr>
<th></th>
<th>Meeting 1</th>
<th>Meeting 2</th>
<th>Meeting 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>2</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>FP</td>
<td>13</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>FN</td>
<td>2</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>TN</td>
<td>91</td>
<td>202</td>
<td>121</td>
</tr>
<tr>
<td>Precision</td>
<td>0.133</td>
<td>0.207</td>
<td>0.267</td>
</tr>
<tr>
<td>Recall</td>
<td>0.5</td>
<td>0.353</td>
<td>0.5</td>
</tr>
<tr>
<td>F1 score</td>
<td>0.207</td>
<td>0.261</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Table 2: Evaluation of the decision detector model
Challenges and Limitations

• Data scarcity
• Quality of speech recognition
• Challenges of spoken language
• Uncertainty Expressions
  • "I think we should follow up with Martin"
  • "Maybe we should have another meeting on Wednesday"
• Identifying referring expressions
  • "We definitely have to implement it"
  • "I will ask her to document it"
• Identifying context and distinguishing decision types
  • "I will create the API"
  • "I will set up a meeting on Monday"
  • "I will share my screen"
Future work

- **Model Enhancement**
  - Solving Data Scarcity
  - Handling uncertainties
  - Resolving referring expressions

- **Virtual Assistant Development**
  - Add communication with participants
  - Implement “Suggestions and recommendations” use case
  - Implement Anonymous notifications
References


3. Icons: https://www.flaticon.com/
M.Sc.
Oleksandra Klymenko

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business Information Systems

Boltzmannstraße 3
85748 Garching bei München

alexandra.klymenko@tum.de
www.matthes.in.tum.de