ALIGNMEET: A Comprehensive Tool for Meeting Annotation, Alignment, and Evaluation

Peter Polák, Muskaan Singh, Anna Nedoluzhko, Ondřej Bojar
<surname>@ufal.mff.cuni.cz

Charles University, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Prague, Czech Republic

Introduction

Summarization is a challenging problem and more so for spontaneous discussions. ALIGNMEET is a comprehensive tool for meeting annotation, alignment, and evaluation. It provides fast annotation while mitigating the risk of introducing errors. The evaluation mode enables a comprehensive quality evaluation of meeting minutes.

What is an ALIGNMENT?

An n-to-1 mapping capturing which dialogue acts are associated with a particular summary point.

Why do we need the ALIGNMENT?

• Visual aid for annotators.
• Annotation error check (no left-out transcript).
• Data for the training of automatic minuting.
• More reliable minutes evaluation (via aggregating meeting-level scores across hunks of the alignment).

Evaluation

• Annotator evaluates on a summary-point-level rather than on document-level.
• More fine-grained and less mentally challenging.
• Adequacy, grammaticality and fluency.
• We aggregate the metrics using plain average to get document-level scores, and we compute coverage of the transcript by aligned summary points.
• We additionally ask annotators to provide a doc-level adequacy.

User-friendly tool for creating and curating meeting transcripts and summaries, aligning between them, and evaluating the summaries.

Evaluation mode

Why n-to-1 alignment?

• Aligning multiple summary points to a single transcript line would further increase the difficulty.
• “Summary point fragmentation” – the annotator might address the same information in separate summary points.

Can we align everything?

• Some parts of the transcript cannot be aligned to a summary point.
• They can be manually flagged as belonging to organizational or small talk.

Pilot study

<table>
<thead>
<tr>
<th>Annotator</th>
<th>English</th>
<th>Czech</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>E2</td>
<td>23</td>
<td>14</td>
</tr>
<tr>
<td>E3</td>
<td>21</td>
<td>24</td>
</tr>
<tr>
<td>C1</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>C2</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>C3</td>
<td>21</td>
<td>24</td>
</tr>
</tbody>
</table>

#Summary points 378 378 282 282 282
#Alignments 378 378 282 282 282
Avg. adequacy 3.71 3.71 3.67 4.93 4.67
Avg. adequacy 3.66 3.31 3.26 4.13 4.67
Avg. adequacy 3.31 3.17 3.67 4.93 4.67
Avg. adequacy 3.66 3.31 3.26 4.13 4.67
Avg. adequacy 3.31 3.17 3.67 4.93 4.67

This work has received support from the project “Grant Schemes at CU” (reg. no. CZ.02.2.69/0.0/0.0/19_073/00016935), the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement No 825460 (ELITR), and 19-26934X (NEUREM3) of the Czech Science Foundation, and partially supported by SVV project number 260 575.