Structure-dependent Summarization of Spoken Lectures

Narine Kokhlikyan, Ying (Joy) Zhang, Alex Waibel
narine.kokhlikyan@sv.cmu.edu, joy.zhang@sv.cmu.edu, waibel@cs.cmu.edu

Missed the lecture? I can summarize all you need to know, in 5 minutes

Why summarize lectures?

- Retrieval of required information becomes a challenging task
- A large number of recorded lectures

<table>
<thead>
<tr>
<th>Sources</th>
<th>Number of Lectures</th>
</tr>
</thead>
<tbody>
<tr>
<td>TED</td>
<td>900+</td>
</tr>
<tr>
<td>MIT</td>
<td>500+</td>
</tr>
<tr>
<td>Oxford</td>
<td>300+</td>
</tr>
<tr>
<td>30+ best colleges</td>
<td>19000+</td>
</tr>
</tbody>
</table>

What makes speech summarization difficult?

Let's try to summarize the following:

Hmmm... So, so welcome to today's lecture hmm ... well we have two things on our our agenda today well So let's ... well

Speech summarization deals with:
- Non-gramatical sentences
- Speech disfluencies
- Redundent information
- ASR error rate
...

Detecting hierarchical structure

Introduction
- "Welcome to today's lecture."
- "My name is Alex Waibel."
  - "today we want to get busy with the second lecture of machine learning"

Content
- "we want to get familiar with one of the most popular approaches, the so-called back-propagation algorithm."

Conclusion
- "Today we learned about neuronal networks, back-propagation algo..."
- "we come to the end today."
- "I hope you will have a beautiful day"

Summary

Hidden Markov Model

- Learning segmentation and summarization parameters
  - Baum-Welch Algorithm
- Applying learned parameter to segment and summarize unseen lectures
  - Viterbi Algorithm

Improvement of summarization quality

Segmentation improves the performance of summarization system

<table>
<thead>
<tr>
<th>Average F Score</th>
<th>80% Summary</th>
<th>Non-Segmented</th>
<th>Segmented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Extraction</td>
<td>87.48</td>
<td>86.23</td>
<td>88.67</td>
</tr>
<tr>
<td>Sentence Extraction</td>
<td>86.35</td>
<td>87.35</td>
<td>88.75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average F Score</th>
<th>20% Summary</th>
<th>Non-Segmented</th>
<th>Segmented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Extraction</td>
<td>35.29</td>
<td>35.3</td>
<td>35.3</td>
</tr>
<tr>
<td>Sentence Extraction</td>
<td>64.75</td>
<td>71.32</td>
<td></td>
</tr>
</tbody>
</table>

Evaluation metric -> ROUGE

Summarization methods

Sentence extraction vs. word selection

- Extractive Summarizer
  - Extracts informative sentences based on a model trained on lectures
  - Model parameters are estimated using lexical, structural, contextual features

Humans prefer sentence extraction

References

[1] ROUGE - Recall-Oriented Understudy of Gisting Evaluation
http://www.berouge.com/

Jiaying Jian Zhang, et al., 2010

[3] Speech-To-Text and Speech-to-Speech Summarization of Spontaneous Speech
Sadaki Furui, Chiori Hori, et al. 2004