14

Robust speech summarization against speech recognition errors

“Huh? What do you mean?” Summarize a long story short

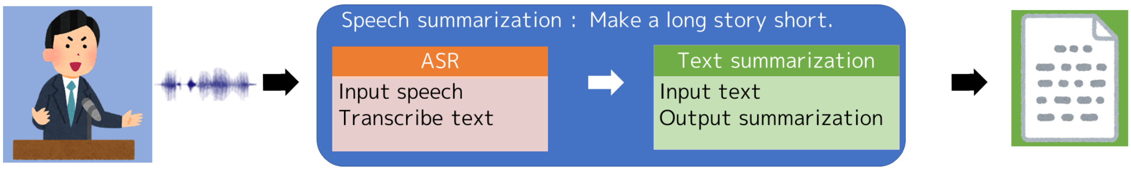
Abstract

Speech summarization aims at creating a summary from a long talk. It is an essential technology if we realize AI systems that can correctly understand human speech. One way to realize speech summarization is cascading automatic speech recognition (ASR) and text summarization. One issue of such approaches is that it is difficult to avoid ASR errors, which degrade the performance of summarization. To alleviate this problem, we propose a robust speech summarization against ASR errors. Our proposed system considers multiple ASR results and looks at the context and relationship between words to generate an accurate summary, even if each ASR result contains errors. The idea we proposed is general and can also be applied to other tasks such as speech translation. This research brings us one step closer to realizing machines that can deeply understand humans, by not only transcribing speech word-by-word but also accessing its meaning and intent.

Mishear but still understand correctly

*A part of this research is a collaboration with Carnegie Mellon University

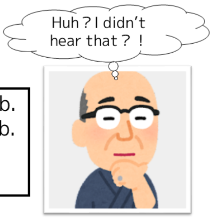
Speech summarization is a technology that summarizes the main points from a long speech, such as a lecture. It can be realized by combining automatic speech recognition (ASR) and text summarization.



Speech summarization： Make a long story short.

ASR
Input speech
Transcribe text

Text summarization
Input text
Output summarization



Problem: **ASR errors seriously affect text summarization!**

Announcements: We will reconsider the content of Salmon lunchbox. The revised result will be announced on the web.
ASR result: We will reconcile the content of Solomon lunchbox. The revised result will be a sold on the web.

Summary： **We will sell a Solomon lunchbox.**

Speech summarization robust against ASR errors

Difficult to achieve perfect ASR

- We exploit results from various ASR systems showing different error tendencies, and expect that the correct meaning can be extracted from the multiple ASR results

Summarize text without assuming that ASR is perfect

- We generate an accurate summary by combining the multiple recognition results
- We utilize a state-of-the-art natural language processing model (BERT*) to model word meanings and relationships.

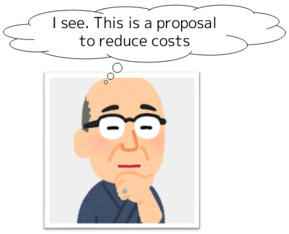
* Bidirectional Encoder Representations from Transformers

ASR A	We will reconcile the content of Solomon lunchbox. The review result will be a sold on the web.
ASR B	We will _________ the content of _______ lunchbox. The review result will be announced on the web.
ASR C	We will reconsider the content of Salmon ______. The review result will be announced on the web.

ASR hypothesis fusion considering context (BERT-based summarization model)

We find out the correct summary by comparing the meaning and content of each word in the ASR results A, B, and C, even if each recognition result individually is wrong

- “Salmon lunchbox” is more natural than “Solomon lunchbox”
- ”reconsider” is related to “revised result”
- The “revised result” should be “announced” not “sold.”



Summary： Changed the content of the Salmon lunchbox.

The proposed method **recovers up to 30% of performance degradation due to ASR errors!**

References

[1] T. Kano, A. Ogawa, M. Delcroix, S. Watanabe, “Attention-based multi-hypothesis fusion for speech summarization,” in *Proc. IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 487–494, 2021.

[2] T. Kano, A. Ogawa, M. Delcroix, S. Watanabe, “ASR hypothesis fusion using BERT for speech summarization,” in *Proc. The 2022 Spring Meeting of the Acoustical Society of Japan (ASJ)*, 2022.

[3] T. Kano, A. Ogawa, M. Delcroix, S. Watanabe, “Integrating multiple ASR systems into NLP backend with attention fusion,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2022.

Contact

Takatomo Kano / Signal Processing Research Group, Media Information Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp