17 Solving two-choice questions makes AI clever
## - Deep pairwise comparison model for ASR hypothesis selection -
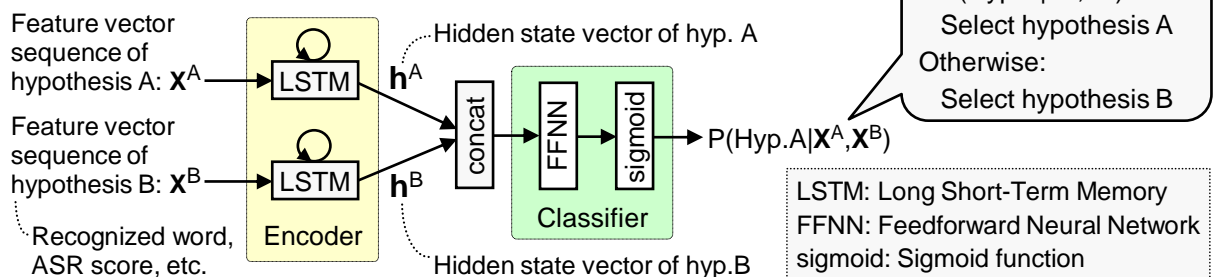
Abstract

We propose an AI system (a deep learning model) that solves two-choice questions. Our system selects the better answer or hypothesis among two hypothetical answers to a problem. By applying such a hypothesis selection mechanism to the multiple hypotheses of automatic speech recognition (ASR), we can find the best hypotheses and thus greatly improve the ASR performance. Our AI system is modeled based on state-of-the-art deep learning technology. By focusing on solving simple two-choice questions, our system can achieve high performance with a small model size. Our system outperforms the de facto standard model for ASR hypothesis selection using only one-tenth the number of parameters. Our system can also be applied to other tasks that output multiple hypotheses such as machine translation or text summarization. In addition, we often solve two-choice questions in our daily life and we expect that our AI system will also help us make better choices when we have to reach important decisions.

Hypothesis selection from multiple ASR hypotheses

Solving two-choice questions by AI
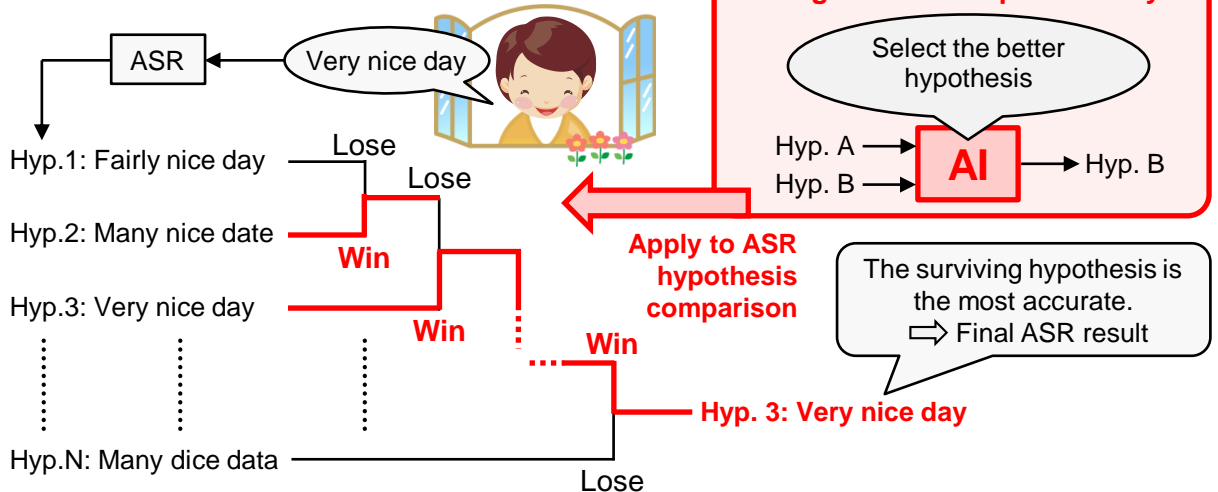
The most accurate hypothesis can be estimated by repeatedly comparing hypothesis using our AI.

Modeling based on state-of-the-art deep learning technology

Feature vector sequence of hypothesis A: $\mathbf{X}^A$

Feature vector sequence of hypothesis B: $\mathbf{X}^B$

Recognized word, ASR score, etc.

Hidden state vector of hyp. A

LSTM → $\mathbf{h}^A$

LSTM → $\mathbf{h}^B$

Encoder

concat → FFNN → sigmoid → $P(\text{Hyp.A}|\mathbf{X}^A,\mathbf{X}^B)$

Classifier

Hidden state vector of hyp.B

If $P(\text{Hyp.A}|\mathbf{X}^A,\mathbf{X}^B) \geqq 0.5$:
Select hypothesis A
Otherwise:
Select hypothesis B

LSTM: Long Short-Term Memory
FFNN: Feedforward Neural Network
sigmoid: Sigmoid function

References

[1] A. Ogawa, M. Delcroix, S. Karita, T. Nakatani, "Rescoring n-best speech recognition list based on one-on-one hypothesis comparison using encoder-classifier model," in *Proc. of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2018),* April 2018.

[2] A. Ogawa, M. Delcroix, S. Karita, T. Nakatani, "Rescoring N-best ASR hypotheses based on one-on-one hypothesis comparison using encoder-classier model," in *Proc. of 2018 Spring Meeting of the Acoustical Society of Japan,* 1-8-9, March 2018.

Contact

Atsunori Ogawa Signal Processing Research Group, Media Information Laboratory