KitAi-PI: Summarization System for NTCIR-14 QA Lab-PoliInfo

Satoshi Hiai, Yuka Otani, Takashi Yamamura and Kazutaka Shimada

Department of Artificial Intelligence, Kyushu Institute of Technology

Contents

Contents

Introduction – Assembly Minutes Summarization

- Two types of summarization methods
  - Abstractive: Use of expressions not contained in the source text
  - Extractive: Use of expressions in the source text
- Assembly minutes corpus
  - A summary consists of expressions contained in a speech

Summary
被災地そして日本の未来のため東京は先頭に立つべき。知事の所見は。

The same expressions

Assembly member speech	
U.1	我々が生きている日本列島は、数限りない天変地異に見舞われてきました。
U.2	被災地のため、そして、日本の未来のために、東京は先頭に立つべきと考えますが、知事の所見を伺います。
	…

Summary generation with an extractive approach

Introduction – Extractive Summarization

- Extraction of a set of important utterances
  - Supervised method usually shows better performance than unsupervised method
  - Use of a machine learning method
    - Construction of importance prediction model
- Problem
  - Given assembly minutes data do not contain importance information for each utterance

Objective

- Automatic training data construction
  - Hypothesis
    - An utterance with high similarity to a sentence in a summary is more important

Summary
被災地そして日本の未来のため 東京は先頭に立つべき。知事の所見は。

Assembly member speech	
U.1	我々が生きている日本列島は、数限りない天変地異に見舞われてきました。
U.2	被災地のため、そして、日本の未来のために、東京は先頭に立つべきと考えますが、知事の所見を伺います。
	…

No importance information for utterances

Objective

- <u>Automatic training data construction</u>
  - Hypothesis
    - An utterance with high similarity to a sentence in a summary is more important

Summary
被災地そして日本の未来のため東京は先頭に立つべき。知事の所見は。

Assembly member speech	
U.1	我々が生きている日本列島は、数限りない天変地異に見舞われてきました。
U.2	被災地のため、そして、日本の未来のために、東京は先頭に立つべきと考えますが、知事の所見を伺います。
	…

Assignment of utterance importance using a word similarity

Low importance score

High importance score

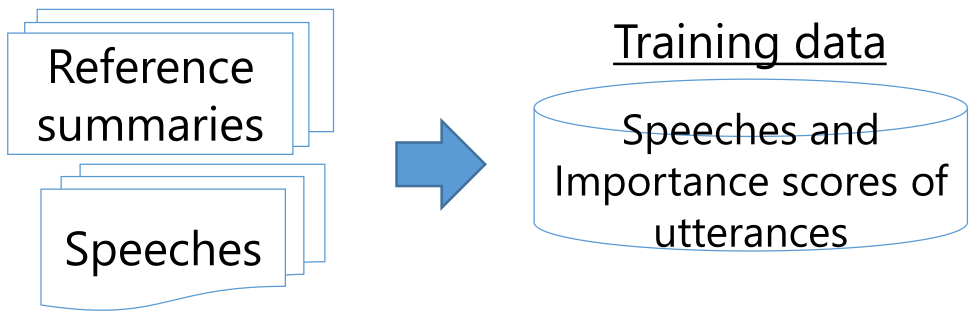
⋮

We can apply a machine learning method

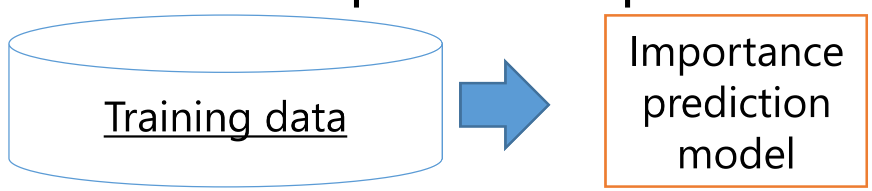
Contents

Outline of Our System

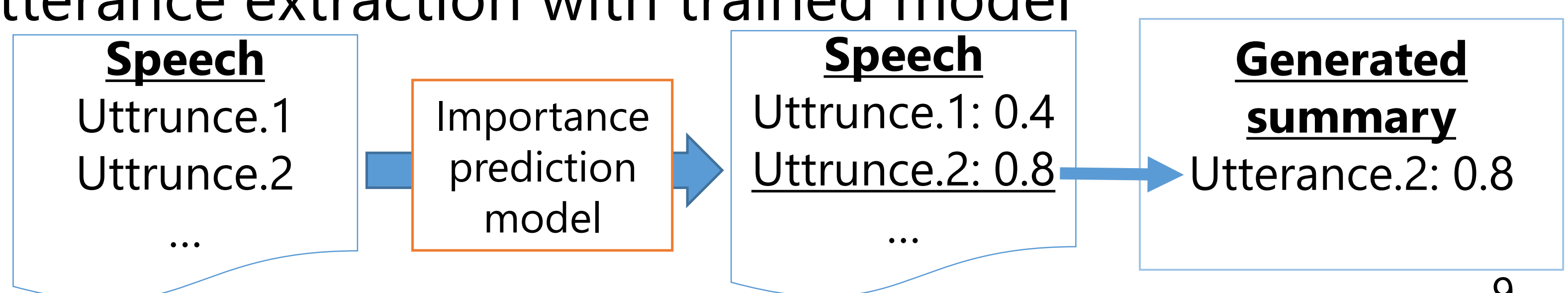
- Training data construction



- Training utterance importance prediction model



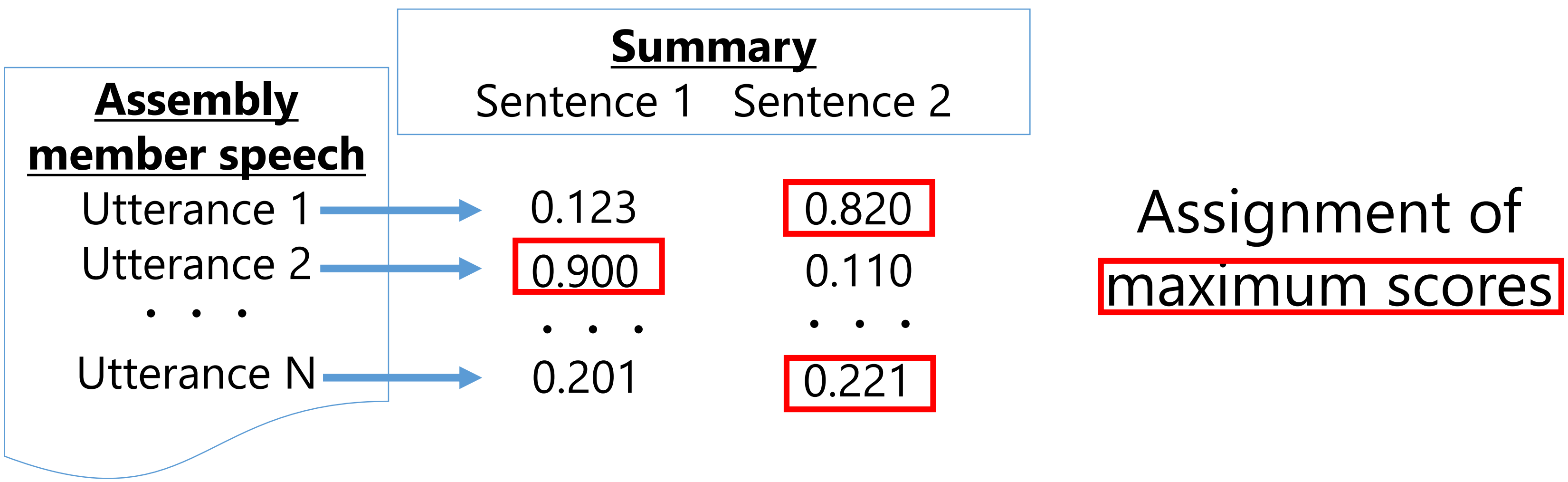
- Utterance extraction with trained model



Contents

Training Data Construction – Assignment of Importance Scores

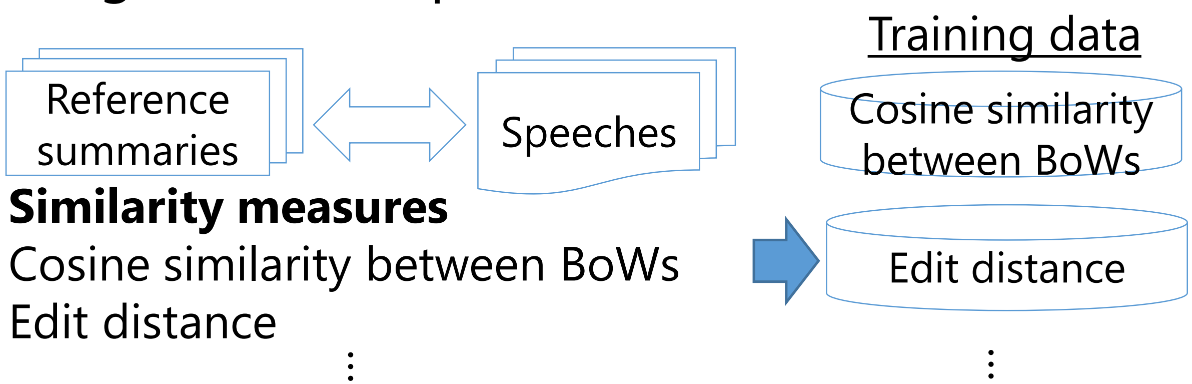
- Automatic assignment of an importance score to each utterance using a word similarity
  - We regard a word similarity as an importance score



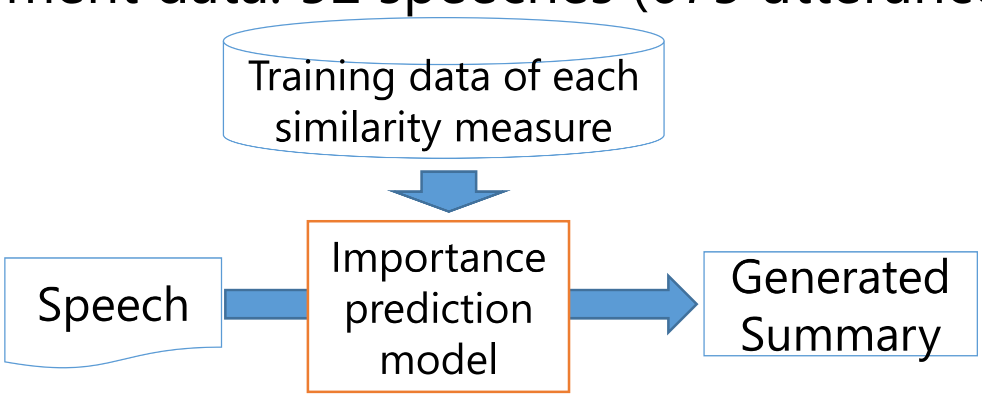
- Evaluation of similarity measures
  - e. g. cosine similarity, edit distance, ...

Training Data Construction – Evaluation of Similarity Measures

- Given corpus: 529 speeches (7,226 utterances)
  - Training data: 477 speeches (6,551 utterances)



  - Development data: 52 speeches (675 utterances)



Training Data Construction – Similarity Measures

- Cosine similarity between bag-of-words
- Edit distance
  - We adopt 1 – (the distances) as the similarity measure
- ROUGE-1 similarity score
  - We use word unigram overlap
- Cosine similarity between sentence embeddings
  - Two methods to generate sentence embeddings
    - Average of word embeddings generated with word2vec
    - Sentence embedding generated with doc2vec
- Average of all the similarity measures

Training Data Construction – Result of Similarity Measures Evaluation

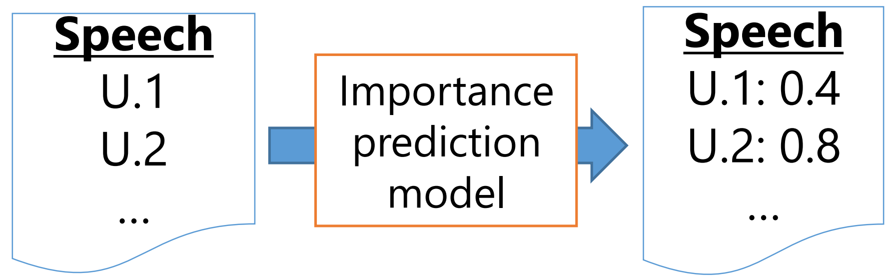
- Evaluation of generated summaries

Similarity measure	Rouge N1
Cosine similarity between bag-of-words	0.333
Edit distance	0.338
ROUGE-1 similarity score	0.341
Cosine similarity between sentence embedding (Word2vec)	0.306
Cosine similarity between sentence embedding (Doc2vec)	0.316
Average of all of the similarity measures	**0.349**

Average of all the similarity measures
is adopted on the formal run

Contents

Settings for Formal Run



- Importance prediction model
  - Features
    - BoW, sentence position in the speech, speaker of the speech
  - Support vector regression (SVR)
- Our methods for the formal run
  - w/ sentence compression
    - We applied a sentence compression on the basis of simple rules

The first content word

~~このため、~~関係機関と連携し、狭隘道路における
消火栓等の整備を促進~~してまいります。~~

The last verbal noun

  - w/o sentence compression

Result on Formal Run – ROUGE Scores

		Recall				F-measure			
		N1	N2	N3	N4	N1	N2	N3	N4
Surface form	w/o sentence compression	**0.440**	**0.185**	**0.121**	**0.085**	**0.357**	0.147	0.096	0.067
	w/ sentence compression	0.390	0.174	0.113	0.078	0.343	**0.154**	**0.101**	**0.069**
	OtherSysAve	0.282	0.096	0.058	0.038	0.272	0.088	0.051	0.033

OtherSysAve: the average scores of all the submitted runs of all the participants

- Our methods outperformed OtherSysAve on all the scores
- F-measure of Rouge N4 of the method with sentence compression was **the best score**
  - It can generate summaries containing important phrases

Result on Formal Run – Participants Assessment

- Quality question scores

	Content		Formed	Total
	X=0	X=2		
w/o sentence compression	**0.856**	**1.134**	**1.732**	**0.912**
w/ sentence compression	0.788	1.035	1.308	0.667
OtherSysAve	0.423	0.603	1.655	0.435

- The method w/o the sentence compression step outperformed OtherSysAve on all the scores
- The formedness score of the method with sentence compression was **lower** than OtherSysAve

The improvement of the sentence compression step is important future work

Summary

- KitAi-PI: extractive summarization system
  - Automatic training data construction
  - Applying the supervised machine learning method
- The formal run result showed the effectiveness of our method
  - Summaries containing important phrases but ill-formed ones
- The improvement of the sentence compression step is important future work