

# 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

## Introduction

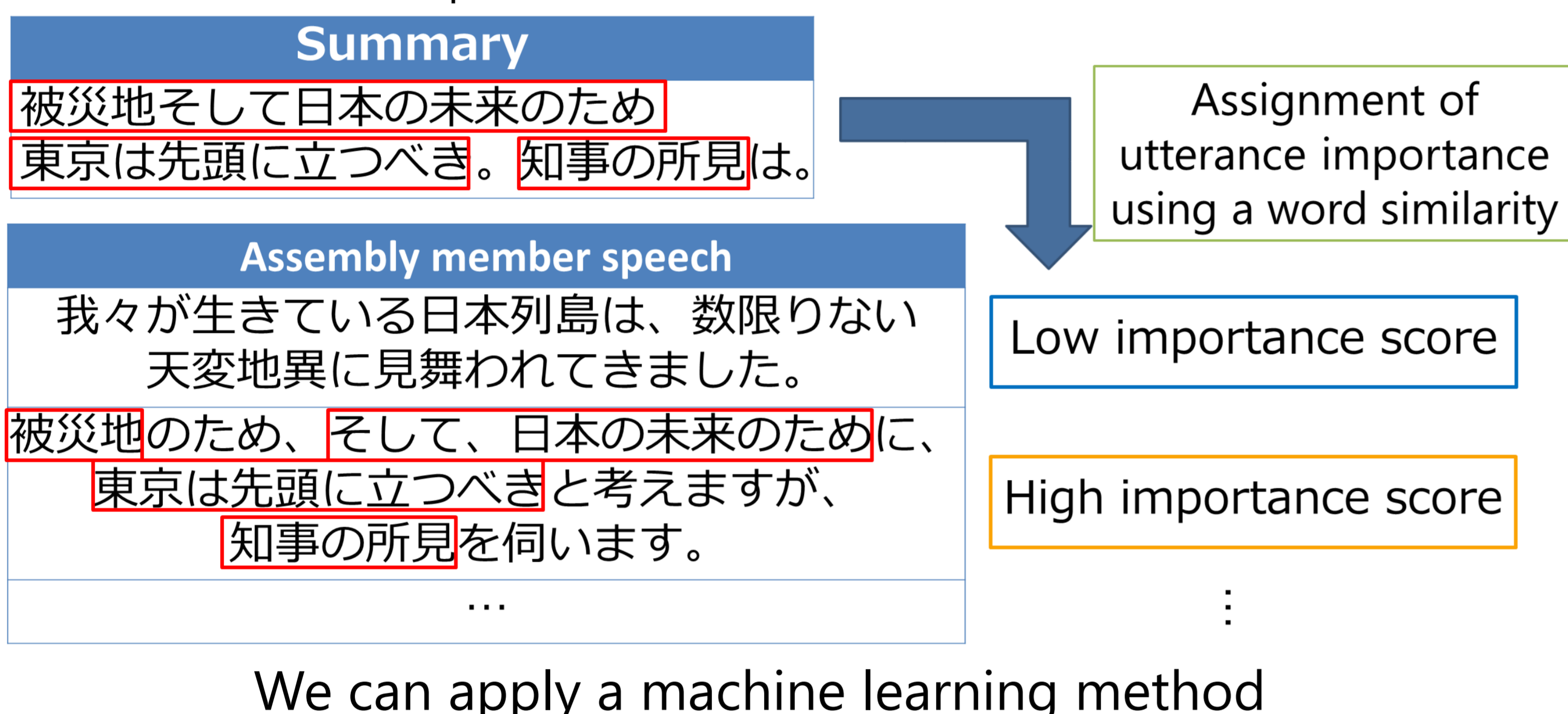
### Extractive Summarization System

- Extraction of a set of sentences in the source text
  - Preciseness is important for assembly minutes summarization
- Importance prediction using a machine learning
  - Supervised method usually shows better performance than unsupervised method

➔ Need of training data containing sentence importance information

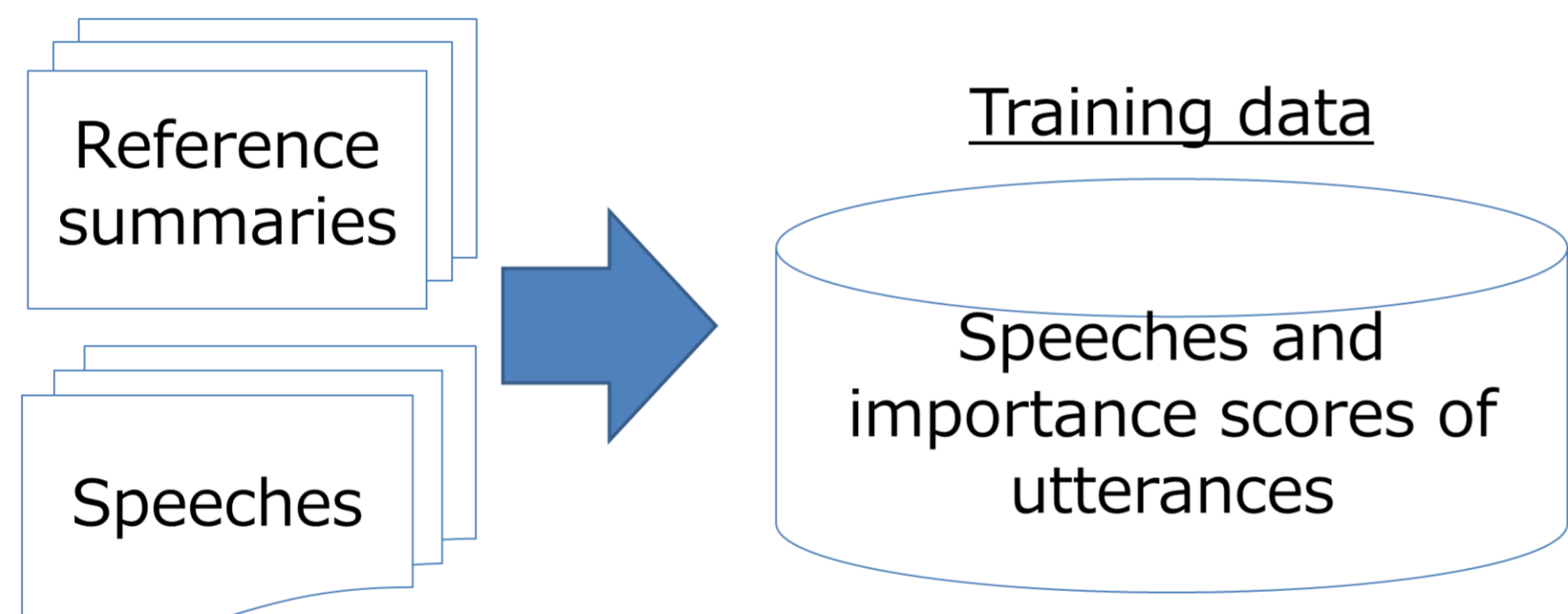
### Objective

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

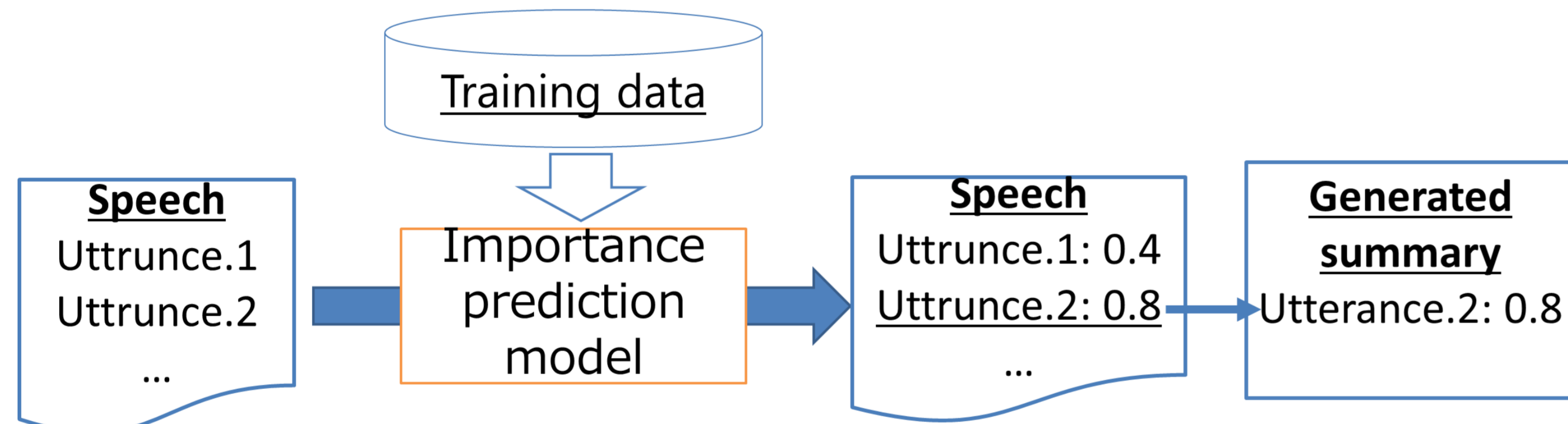


## Outline

- Training data construction

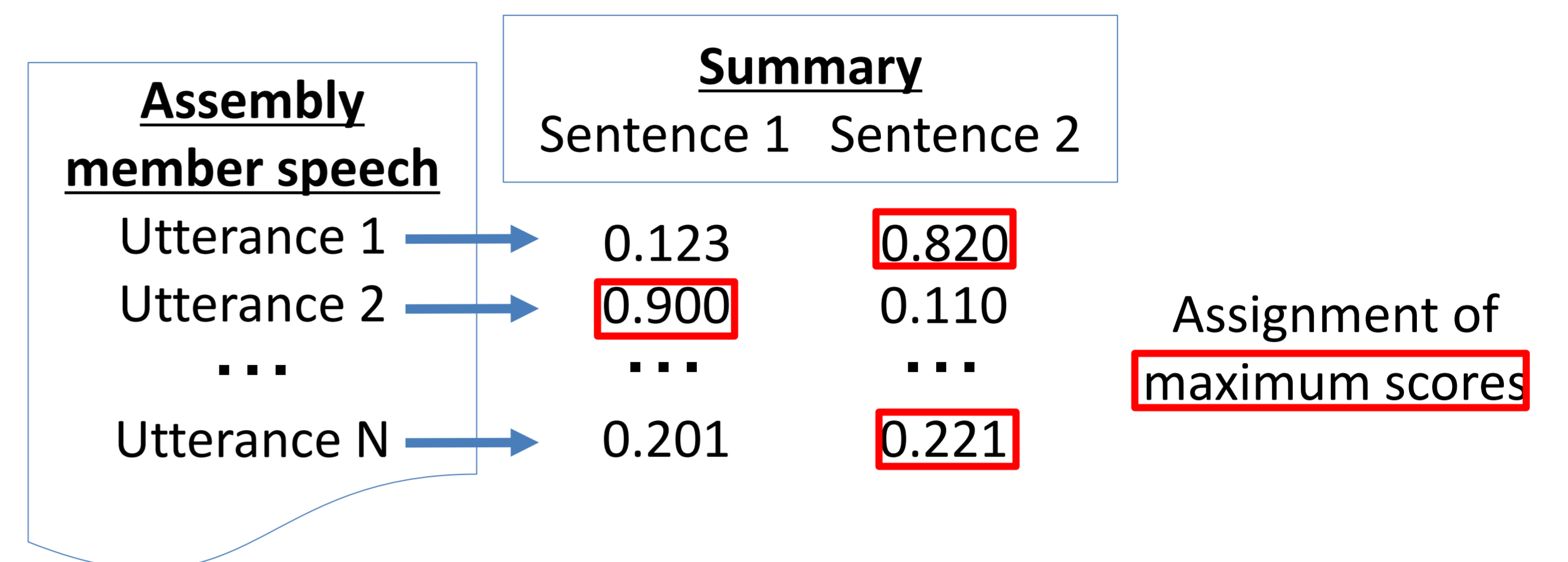


- Sentence extraction with trained model



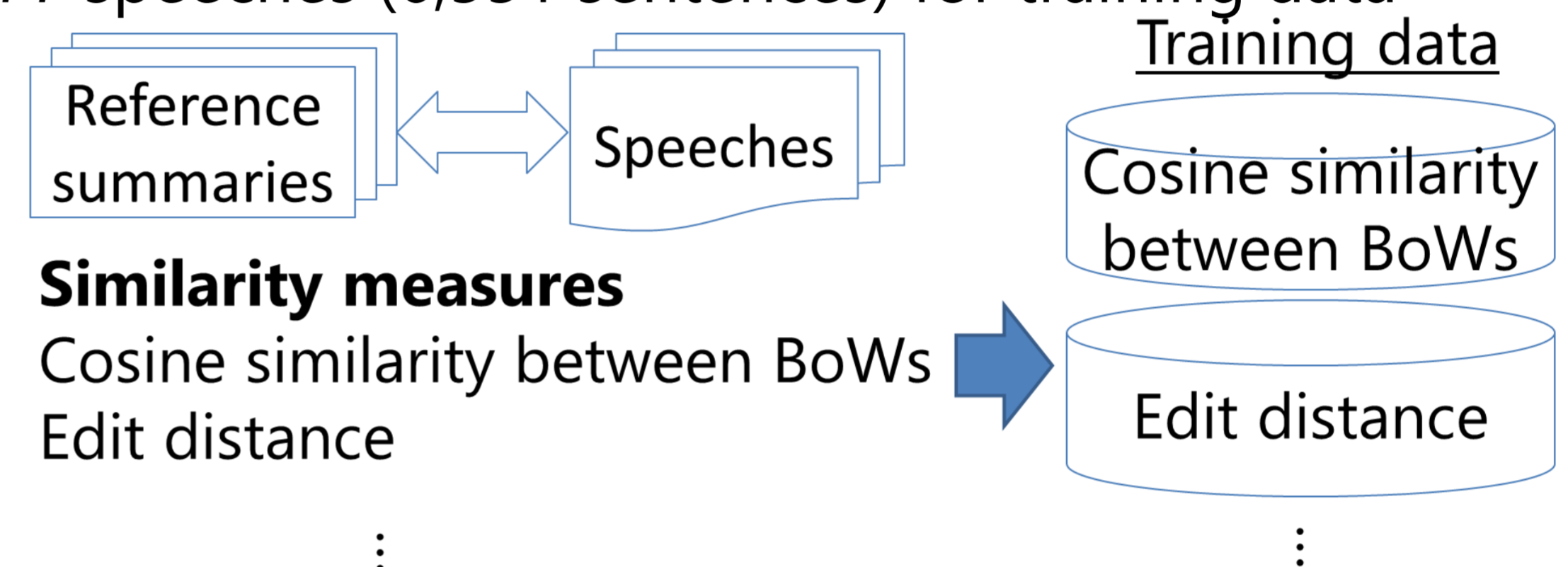
## Training Data Construction

- Assignment of importance scores of utterances using a word similarity

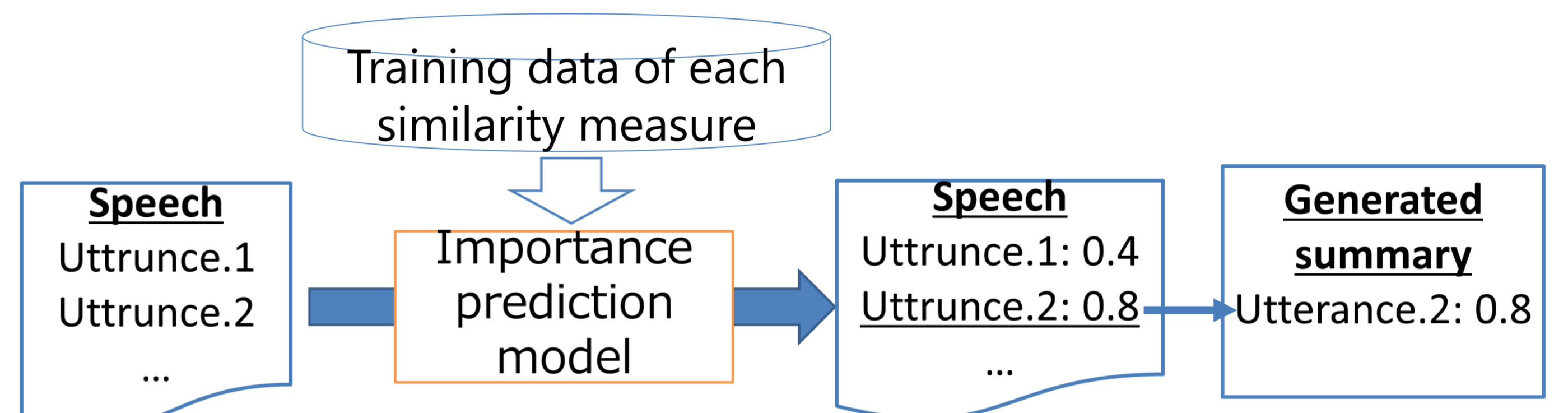


### Evaluation of Similarity Measures

- There are many way to calculate similarity, such as cosine similarity and edit distance etc.
  - Selection of the most suitable measure on this task
- Given corpus: 529 speeches (7,226 sentences)
  - 477 speeches (6,551 sentences) for training data



- Evaluation of generated summaries of 52 speeches (675 sentences)



- The result of the evaluation

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 the similarity measures	<b>0.349</b>

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

## Evaluation on Formal Run

### Our method

- w/ sentence compression
  - We applied sentence compression on the basis of simple rule
- w/o sentence compression

### ROUGE Scores

		Recall				F-measure			
		N1	N2	N3	N4	N1	N2	N3	N4
Surface form	w/o sentence compression	<b>0.440</b>	<b>0.185</b>	<b>0.121</b>	<b>0.085</b>	<b>0.357</b>	0.147	0.096	0.067
	w/ sentence compression	0.390	0.174	0.113	0.078	0.343	<b>0.154</b>	<b>0.101</b>	<b>0.069</b>
	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 participants

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

### Participants Assessment

	Content		Formed	Total
	X=0	X=2		
w/o sentence compression	<b>0.856</b>	<b>1.134</b>	<b>1.732</b>	<b>0.912</b>
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 outperformed OtherSysAve on all scores
- The formedness score of the method with sentence compression was **lower** than OtherSysAve

- Our method was effective
- The improvement of the sentence compression step is important future work