KIT Dialogue System for NTCIR-13 STC Japanese Subtask

Team KIT16
Kyoto Institute of Technology

Hiroshi Nakatani, Shigenori Nishiumi
Takahiro Maeda, Masahiro Araki

Three Methods

- Retrieval-based method
  1. **TF-IDF / word2vec** method
     - similar word and related word
  2. **Clustering utterance with CRP** method
     - relationship between utterance

- Generation-based method
  3. **Seq2seq** method
     - It can generate replies repository does not have

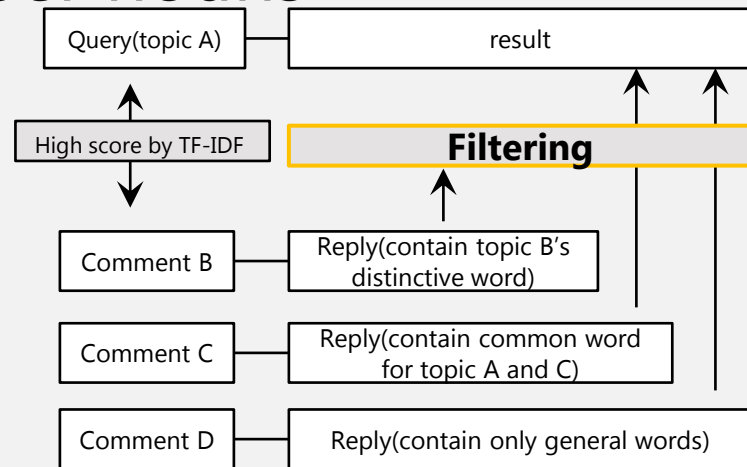# Retrieval-based method with TF-IDF and word2vec

- Flow of this method
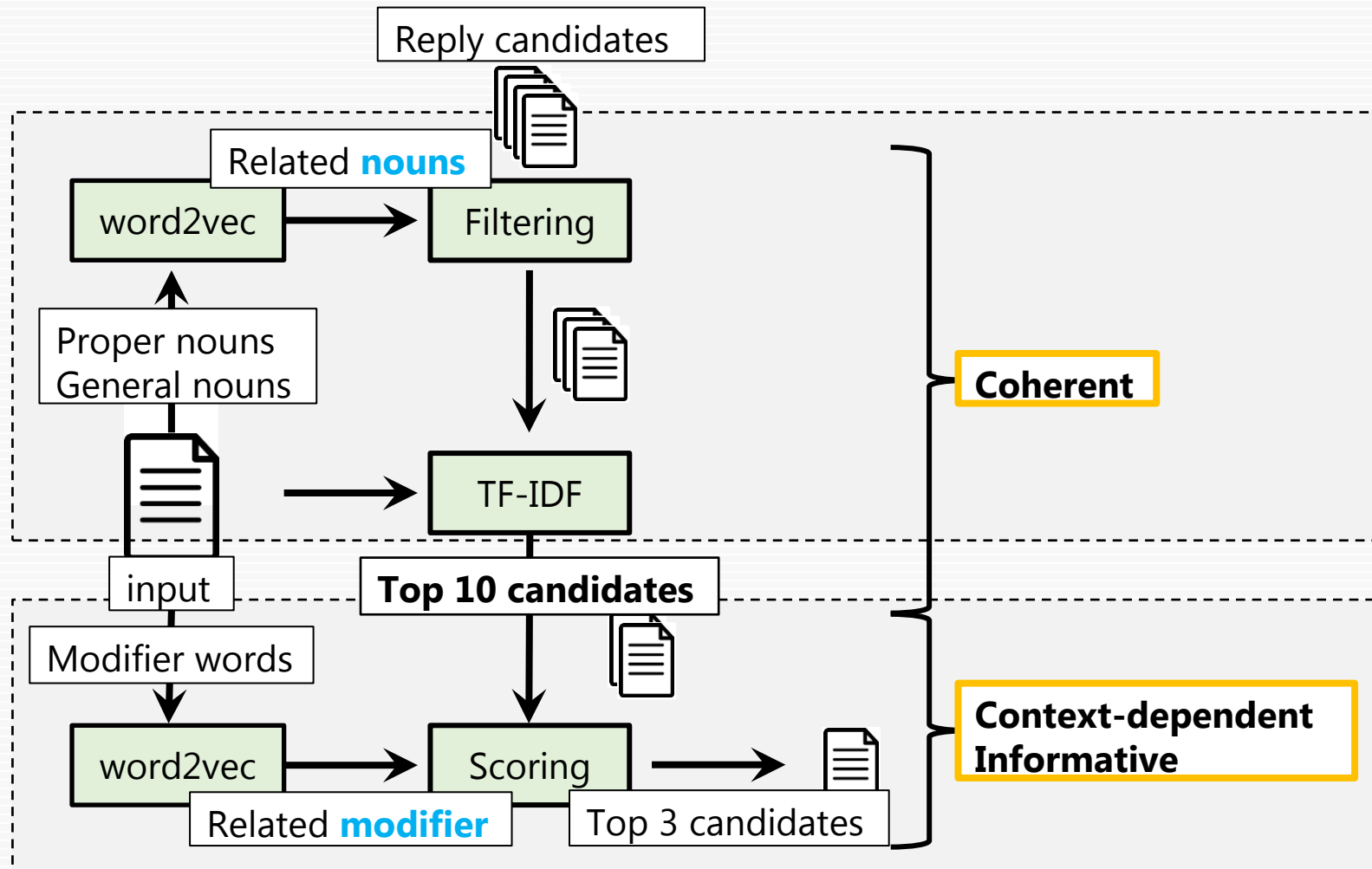
Coherence Evaluation

- ## Filtering based on proper nouns

  - Proper nouns by word2vec based on input proper nouns.

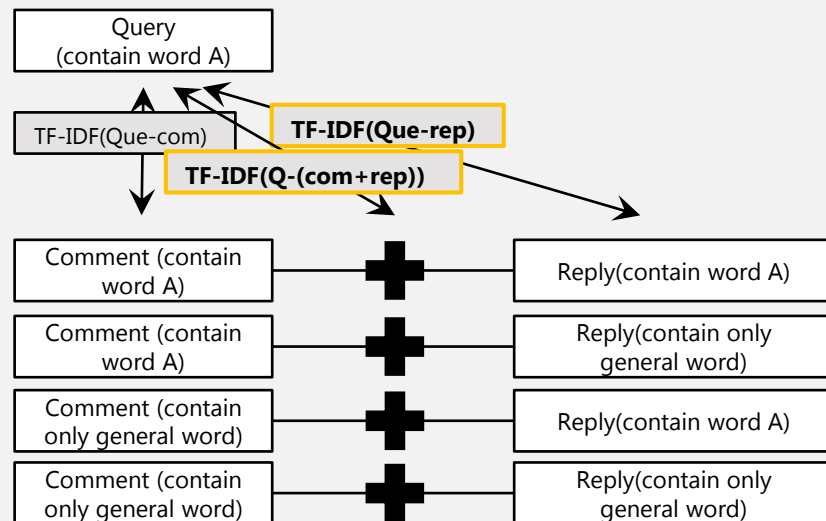 - If other proper nouns are in reply, then remove it.

Query(topic A)		result

High score by TF-IDF　　**Filtering**

Comment B	Reply(contain topic B's distinctive word)
Comment C	Reply(contain common word for topic A and C)
Comment D	Reply(contain only general words)

- ## Scoring using TF-IDF

  Calculating cosine distance between
  ①[input data] and [reply],
  ②[input data] and [comment + reply]

  Score = ①×②

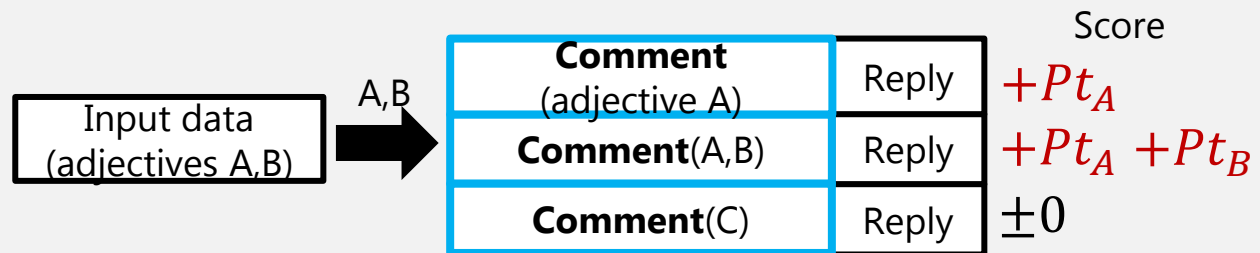Query (contain word A)

TF-IDF(Que-com)　TF-IDF(Que-rep)

TF-IDF(Q-(com+rep))

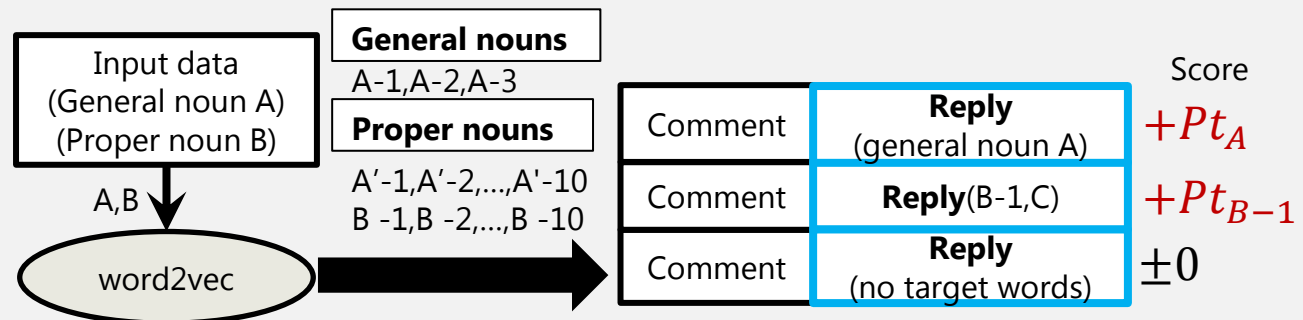| Comment (contain word A)	+	Reply(contain word A)
Comment (contain word A)	+	Reply(contain only general word)
Comment (contain only general word)	+	Reply(contain word A)
Comment (contain only general word)	+	Reply(contain only general word)

Context-dependence and Informativeness Evaluation

- ● Scoring using adjectives and adverbs
  - • If a comment includes same adjectives (or adverbs) as input data, then the score is increased

		Score
Comment (adjective A)	Reply	$+Pt_A$
Comment(A,B)	Reply	$+Pt_A + Pt_B$
Comment(C)	Reply	± 0

Input data (adjectives A,B) →A,B→

- ● Scoring using nouns
  - • If a reply includes nouns (obtained by word2vec based on input data), then the score is increased

Input data (General noun A) (Proper noun B)

General nouns A-1,A-2,A-3

Proper nouns A'-1,A'-2,...,A'-10 B-1,B-2,...,B-10

→A,B→ word2vec

		Score
Comment	**Reply** (general noun A)	$+Pt_A$
Comment	**Reply**(B-1,C)	$+Pt_{B-1}$
Comment	**Reply** (no target words)	± 0

Retrieval-based method with Clustering utterance with CRP

- **Dialogue-act** clustering
  By learning the tendency of dialogue-act in pairs in the repository, we estimate which dialogue-act should be used in response to a new given utterance.
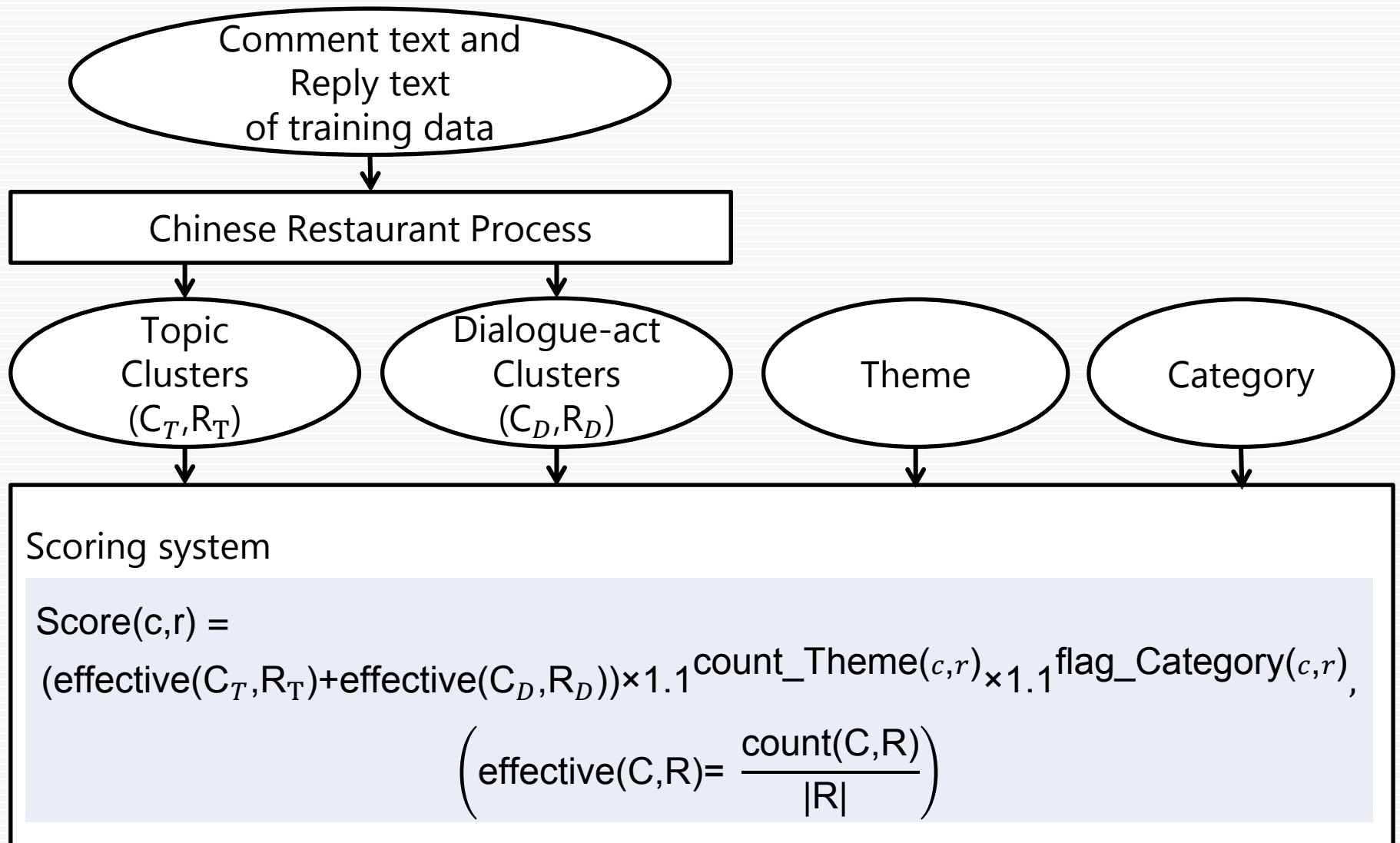
- **Topic** clustering
  By retrieving reply that is same cluster to the input comment, the reply is same topic to the comment.

By using these clustering, a natural response is selected.

Clustering

	Topic Clustering	Dialogue-act Clustering
Feature	BoW (noun, adjective, verb)	BoW (symbol, emotion, filler, adverb, particle, conjunction, adnominal)
Method	CRP	CRP
Number of Clusters	722	74

Scoring

Comment text and
Reply text
of training data

Chinese Restaurant Process

Topic
Clusters
(C_T, R_T)

Dialogue-act
Clusters
(C_D, R_D)

Theme

Category

Scoring system

Score(c,r) =
$$(\text{effective}(C_T,R_T)+\text{effective}(C_D,R_D))\times 1.1^{\text{count\_Theme}(c,r)}\times 1.1^{\text{flag\_Category}(c,r)},$$

$$\left(\text{effective}(C,R)= \frac{\text{count}(C,R)}{|R|}\right)$$

Generation-based method with seq2seq model

We add the salient **keyword of the news** to the comment.
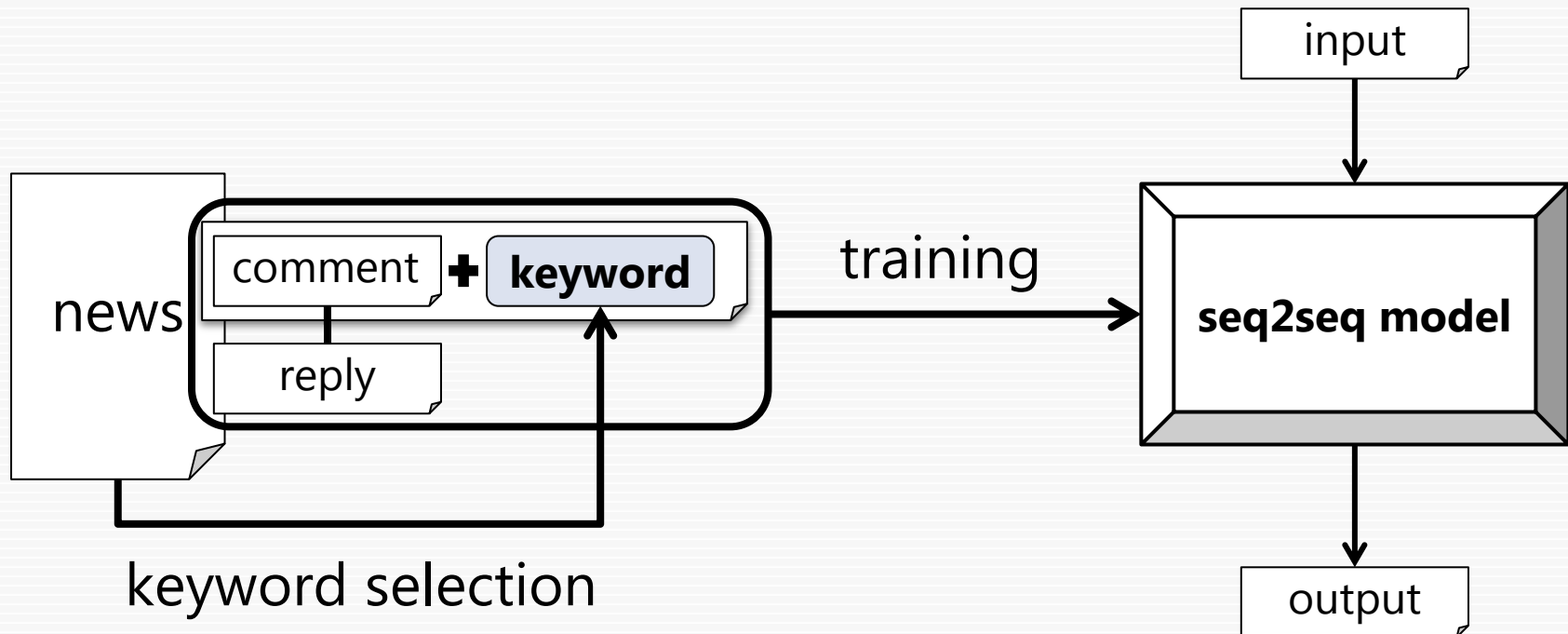
Keyword selection

- Keyword is a word that express the news.

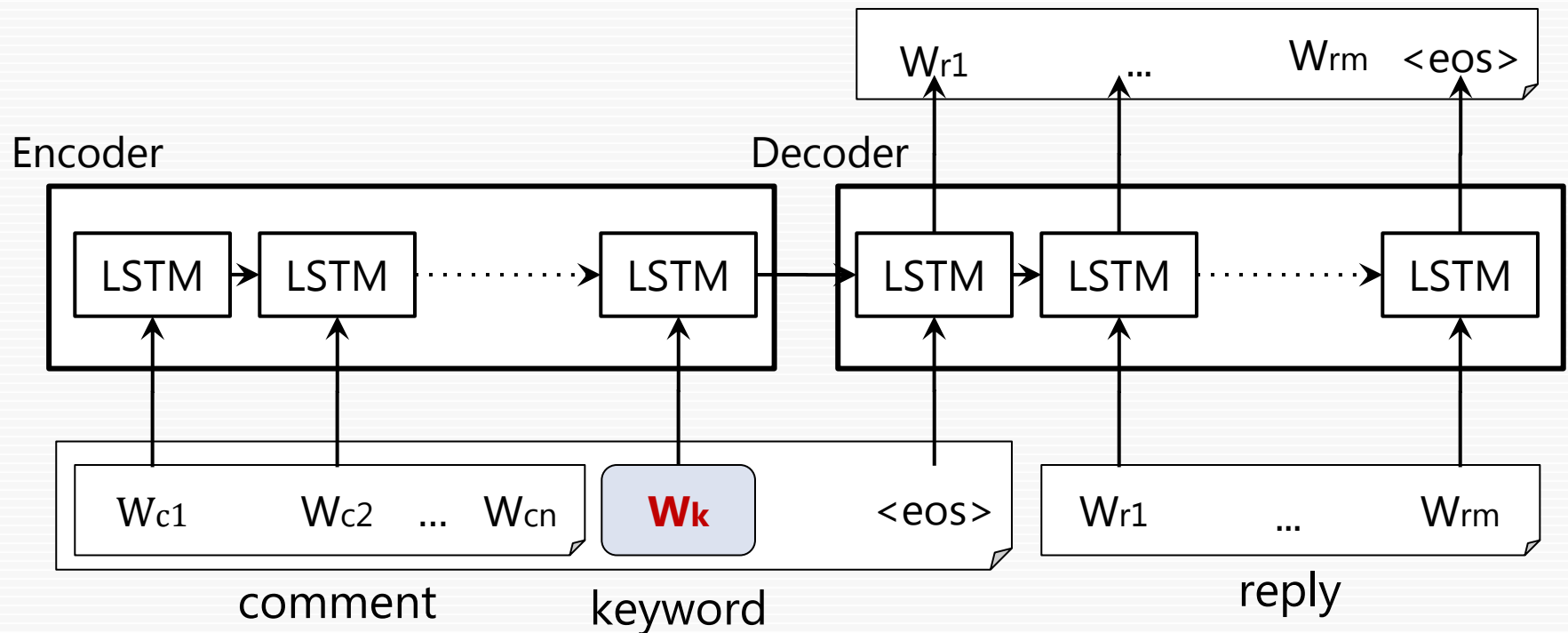- Keyword is a **proper noun** that first comes in titles in Yahoo! Topics.

Keyword examples
- *America*
- *YouTuber*
- *Hakone Ekiden*
- *SMAP×SMAP*

etc.

seq2seq model training

Keyword is added to the end of comment.

Evaluation Results

Retrieval-based method with TF-IDF and word2vec achieved the best results in our systems.

method	Mean AccL2@1 (Rule-1)	Mean AccL1,L2@1 (Rule-1)	Mean AccL1,L2@1 (Rule-2)
GOLD	0.4720	0.8980	0.8660
TF-IDF/ word2vec	0.1800	0.8240	0.6320
CRP	0.0860	0.4660	0.3840
seq2seq	0.0960	0.6320	0.4680

Conclusion

We addressed on the STC Japanese Subtask

from three approaches:

1. **TF-IDF** / **word2vec** method
2. **Clustering utterance with CRP** method
3. **seq2seq** model method

As a result, Retrieval-based method with

TF-IDF and word2vec showed the best results.

The major cause of the result is that

TF-IDF can select similar replies to the comment.