Controlling the Voice of a Sentence in Japanese-to-English Neural Machine Translation

Hayahide Yamagishi, Shin Kanouchi, Takayuki Sato and Mamoru Komachi Tokyo Metropolitan University

Abstract

We must consider the difference in expression between languages in MT. For example, the **active/passive voice** may change in Japanese-English translation. MT systems should consider the information structure to improve the coherence of the output. Sennrich et al. (NAACL, 2016) attempted to control the honorific in English-German NMT. Similar to Sennrich et al., we report on our attempt to control the voice of a sentence generated by an encoder-decoder.

Verb in the training data	# Active	# Passive	# Total
show	21,703	11,441	32,144
describe	12,300	17,414	29,774
introduce	6,030	9,167	15,197
examine	3,795	11,100	14,895
detect	468	2,858	3,326

Automatic Labeling of the Voice

① Recognizing the voice of the target (English) sentence.
② Adding a special token, <Active> or <Passive>, as a word to the end of the source sentence.

The design and characters of the circuit were explained.

① Is the root the verb in the past participle form?
AND
Is there a be-verb in the children of the root?

Yes → <Passive>
No → <Active>

回路 の 設計 と 特性 を 解説 した 。<Passive>

② Training an attentional encoder-decoder model [Bahdanau+ 2015] with a labeled Japanese sentence.

[Test] 回路 の 設計 と 特性 を 解説 した 。<Active>

Positive result — We explained the design … .

Negative result — The design of … were explained.

Settings

Corpus: ASPEC (Asian Scientific Paper Expert Corpus)

- 827,503 sentences, obtained by eliminating sentences with more than 40 words in the first 1 million sentences.
- Word2Vec (gensim) was trained with all 3.0M sentences of ASPEC.

Tools

- MeCab (the tool of Japanese Morphological Analysis)
- Cabocha (the tool of Japanese Dependency Structure Analysis)
    - Both of them used the Dictionary IPADIC ver. 2.7.0
- Stanford Parser 3.5.2

Hyper-Parameters of Encoder-Decoder

- Vocabulary: 30000, epoch: 10
- Embed size and Hidden size: 512, Batch size: 128
- Optimizer: Adagrad (Learning rate: 0.01)

Experiments

① Train the attentional encoder-decoder by the labelled data.
② Add the label to the end of sentence of the test data.
③ Check the voice of output sentence.

Testing the following four patterns of labeling the voice features.

a. **ALL_ACTIVE**: Controlling all target sentences to the active voice.
b. **ALL_PASSIVE**: Controlling all target sentences to the passive voice.
c. **REFERENCE**: Controlling each target sentence to the same voice as that of the reference sentence.
d. **PREDICT**: Controlling each target sentence to the predicted voice.

- Adding the majority label of the voice distribution in the training set.
- It was submitted to WAT 2016

We checked the voice of 200 generated sentence manually.
We calculated the BLEU score with the test data of all 1812 sentences.
The accuracy was calculated as the agreement between the label and the voice of the generated sentence.

How to PREDICT (d.)

回路 の 設計 と 特性 を 解説 した 。

Is the root in the voice Distribution DB of the training data?

Yes → Choosing the majority of the Voice Distribution DB.
No → <Active>

Voice Distribution DB
…
解説した (*explained*)
<Active>: 88, <Passive>: 117
…

[PREDICT] 回路 の 設計 と 特性 を 解説 した 。<Passive>

Result of Experiments	Active	Passive	Error	Accuracy	BLEU
Reference	100	100	0	-	-
Baseline (No Labels)	74	117	9	-	20.53
a. ALL_ACTIVE	151	36	13	75.5%	19.63
b. ALL_PASSIVE	17	175	8	87.5%	19.93
c. REFERENCE	97	94	9	89.5%	**21.26**
d. PREDICT (Compared to Ref.) (Compared to Label)	72	121	7	69.5% 87.5%	20.42

Discussion

1. There were many sentences that persisted the "be-verb + verb in past participle form" structure.
2. In the case that the root verb in the target should be an intransitive verb, it exchanged like "do ⋧ be found to do" or "can be done ⋧ is able to be done".
3. The result of voice controlling tended to fail sometimes if we input the verb that had the skewed voice distribution.
4. PREDICT failed to predict the voice of the reference, especially with high-frequency verbs.

Future work

- PREDICT resulted in decrease in the BLEU, so we want to think about another method how to predict.
- We will study how the non-root verb must be treated in order to obtain the consistency of the document expression.

Positive examples	[P1] The voice of reference is "Active."	[P2] The voice of reference is "Passive."
Input	熱戻り反応の機構を議論した。	リサイクルに関する最近の話題を紹介した。
Reference	This paper discusses the mechanism of the heat return reaction.	Recent topics on recycling are introduced.
ACTIVE	We discuss the mechanism of the thermal return reaction.	This paper introduces recent topics on recycling.
PASSIVE	The mechanism of the thermal return reaction is discussed.	Recent topics on recycling are introduced.

Negative examples	[N1] The voice of the target is controlled, but meaning is different.	[N2] The voice of the target is not controlled.
Input	自己組織化構造に分子の形と分子間相互作用が大きく影響する。	その結果、THZ波はSTJでのトンネリング電流信号として検出できる。
Reference	Molecular shape and intermolecular interaction influence self-assembled structures greatly.	Consequently, the THZ waves can be detected as tunneling current signals at STJ.
ACTIVE	The molecular structure and molecular interaction greatly affect the self-organization structure.	As a result, the THZ wave can be detected as a current current signals in the <unk>.
PASSIVE	The molecular structure and molecular interaction are greatly affected by the self-organization structure.	As a result, the THZ wave can be detected as a current current signals in the <unk>.