Restricted Translation Task at WAT'21

Task Organizers

Akiko Eriguchi (Microsoft),
Kaori Abe (Tohoku University),
and Yusuke Oda (Legal Force)

Consistency: Key to Accurate Translation

…シフトする帰還回路であり…物体の磁気特性の変化…で検出することができる。

(This is a feedback circuit .... can detect change of magnetic features of an object present ....)

This circuit is a feedback circuit .... can detect changes in the magnetic characteristics of the object in the center ...
(Online System A's Translation)

This circuit is a feedback circuit .... detect changes in the magnetic properties of objects in the center of ...
(Online System B's Translation)

Document-level translation often requires us to have consistent terminology related to technical terms, proper nouns and so on.

Task Details

- Redesign ASPEC scientific-domain translation task (Nakazawa et al., 2016)
- "Translate sentences, given a target vocabulary list" (**Restricted Vocabulary**)
  - Vocab. list containing scientific technical terms in target language
  - NOT accept other terms that are semantically similar to the specified ones

…シフトする帰還回路であり…物体の磁気特性の変化…で検出することができる。

["feedback circuit", "magnetic features", ...]

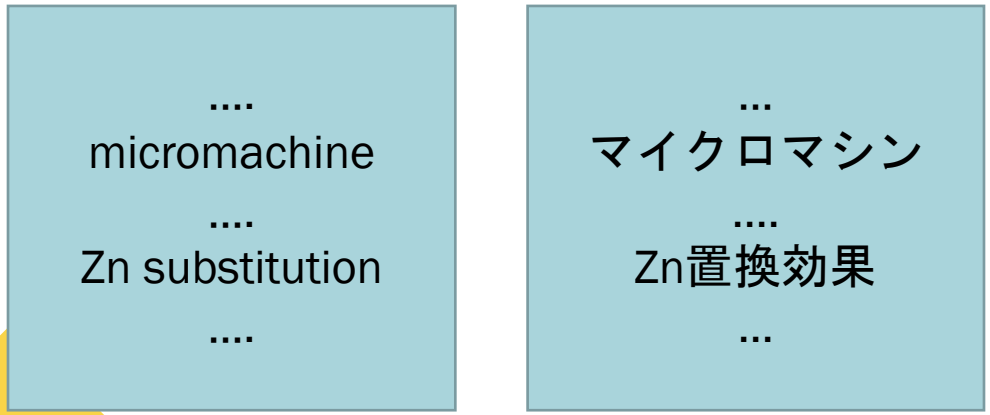
(Restricted Vocabulary List)

This circuit is a.... can detect changes in the magnetic characteristics of the object in ... ✗

This is a feedback circuit .... can detect change of magnetic features of an object present .... ✓

Building Restricted Vocabulary

- Asked 10 bilingual speakers to annotate technical terms/phrases in ASPEC corpus (Nakazawa et al., 2016)
- Created restricted vocabularies for ASPEC's En-Ja/Ja-En MT tasks



Examples of annotated vocabularies in En/Ja

	En-Ja	Ja-En
Dev	2.8	2.8
Devtest	3.2	3.2
Test	3.3	3.2

Table 1: Avg # of annotated phrases per sentence

Automatic Evaluation Metric for RT Task

- Re-rank submitted systems with our designed task-specific metric
  - **BLEU score**, to measure "translation accuracy"
  - **Exact match score** with restricted vocab, to measure ”consistency in translation"

…シフトする帰還回路であり…物体の磁気特性の変化…で検出することができる。

[“feedback circuit”, “magnetic features”, ….]

This circuit is a…. can detect changes in the magnetic characteristics of the object in …

✗

This is a feedback circuit …. can detect change of magnetic features of an object present ….

✓

Overview of Submitted Systems

- In Restricted Translation Task, we received:
  - 3 systems submitted for English-Japanese task
  - 4 systems submitted for Japanese-English task
- Major approaches: Neural MT systems + lexical constraints decoding

Presentation of system description papers regarding RT task:
- 8:10 ~ “NHK’s Soft Lexically Constrained Neural Machine Translation at WAT 2021”

- 8:20 ~ “Input Augmentation Improves Constrained Beam Search for Neural Machine Translation: NTT at WAT 2021”

- 8:30 ~ “NICT's Neural Machine Translation Systems for the WAT21 Restricted Translation Task”

- 8:40 ~ “Machine Translation with the Pre-specified Target-side Words Using a Semi-autoregressive Model”

Results and Human Evaluation

- Our designed eval. metric is better correlated with human eval scores
- Top-ranked team “NTT” won the best in human evals, too.
  (even higher than human ref.!)
- Achieving human parity?
  - Detected low-quality human ref data with ratio of (En-Ja, Ja-En) = (13%, 12%)

En-Ja		Human Eval.	
Team	final	src-based DA	src-based CA
NTT	**57.2**	**77.5**	**79.7**
NHK	33.9	74.1	77.2
NICTRB	28.8	73.6	77.1
(human ref.)	—	73.4	76.4

Ja-En		Human Eval.	
Team	final	src-based DA	src-based CA
NTT	**44.1**	**75.6**	**74.4**
NHK	37.5	73.9	73.5
NICTRB	31.8	72.1	71.8
TMU	22.6	50.2	48.3
(human ref.)	—	74.1	72.9

- Src-based direct assessment (DA) (Cettolo et al, 2017)
- Src-based contrastive assessment (CA) (Sakaguchi and Van Durme, 2018)

Summary

Restricted Translation task

- New MT task @WAT'21: "Translate sentences, given a target vocabulary list"
- Built restricted vocabulary from ASPEC for En-Ja/Ja-En tasks
- Design novel task-specific evaluation metric
- Discuss how MT systems can achieve consistent translation

Major Approaches at RT task

- Neural MT systems + lexical constraints

Results and findings at RT task

- Team "NTT" got won in both tasks, which human eval result also supported