



Restricted Translation Task at WAT'22

Task Organizers

Akiko Eriguchi (Microsoft),
[Kaori Abe](#) (Tohoku University),
and Yusuke Oda (Inspired Cognition)

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 a 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 in En-Ja/Ja-En

- 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

....
micromachine
....
Zn substitution
....

...
マイクロマシン
....
Zn置換効果
...

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

Building Restricted Vocabulary in Zh-Ja/Ja-Zh

- Automatically extract term candidates for each language with termextracter*
- Manually check alignments between the extracted terms in two steps
 - Asked one bilingual speaker to make alignments
 - Asked another two bilingual annotators to give translation 0–100 scores

.... 鼓风机 克罗恩氏病 ブロウ クローン病 94 88
--------------------------------------	------------------------------------	----------------------------------

Examples of annotated vocabularies & scores in Zh/Ja

	Zh-Ja	Ja-Zh
Dev	1.2	1.2
Devtest	1.5	1.5
Test	1.4	1.4

Table 1: Avg # of annotated phrases per sentence

* <http://gensen.dl.itc.u-tokyo.ac.jp/pytermextract/>

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:
 - Team TMU systems submitted for English-Japanese/Japanese-English tasks
- approach: soft lexical constraint method (LeCA) + post-editing method (LevT)

Fast decoding speed
but imperfect constraint terms

Compensate for the rest terms

Presentation of system description papers regarding RT task:

- TMU NMT System with Automatic Post-Editing by Multi-Source Levenshtein Transformer for the Restricted Translation Task of WAT 2022, *Seiichiro Kondo and Mamoru Komachi*

Results and Human Evaluation

En-Ja Outputs	final	HE (Accuracy)	HE (src-based DA)
LeCA+LevT (<i>ensemble</i>)	52.7	4.18	76.4*
LeCA+LevT	50.5	4.19	76.6*
LeCA only	37.6	4.24	74.9
Human Reference	—	—	76.6

Ja-En Outputs	final	HE (Accuracy)	HE (src-based DA)
LeCA+LevT (<i>ensemble</i>)	40.8	4.31	74.1*
LeCA+LevT	38.1	4.22	72.0
LeCA only	23.0	4.14	73.3
Human Reference	—	—	74.7

- “LeCA+LevT (ensemble)” won the best in both our metric & human evals
- The score of our eval. Metric (final) is reverse compared with human evals (Acc.) in En-Ja
 - Some undergeneration *other than target terms* in LeCA+LevT (ensemble) output
 - Domain knowledge of annotators?
- Accuracy: WAT official 5-scales evaluation ([Nakazawa et al. 2022](#))
- Src-based direct assessment (DA) ([Cettolo et al, 2017](#))

Summary

Restricted Translation task: “Translate sentences, given a target vocabulary list”

- Add new direction@WAT’22: Zh-Ja/Ja-En (but no submission in this year)
 - We look forward to your participation in next year!
- Discuss how MT systems can achieve consistent translation

New Approach at RT task

- Soft lexical constraint method + post-editing method

Results and findings at RT task

- Team “TMU” model performs well in both En-Ja/Ja-En directions
- Weakness of our Metric : Should capture undergeneration other than terms
- Still challenging problem : Trade-off between complete vocabulary constraints and translation quality