

AKBL at the NTCIR-17 QA Lab-PoliInfo-4 in **Answer Verification Subtask**

Tasuku Shin, **Haruki Ishikawa**, Yuki Gato, Eiji Kuramoto, Tomoyoshi Akiba
Toyohashi University of Technology (team: AKBL)

Dec. 15, 2023. NTCIR-17

Outline

1. Task Background
2. Data Augmentation
3. Text Edit Operations
4. Experiment
5. Result

Task Background

We fine-tuned the pre-trained BERT with distributed data.

The accuracy of the model was not very high.

	accuracy	F-measure
dry-run	<u>0.7848</u>	0.7606
formal-run	<u>0.7900</u>	0.8073

Because of the scarcity of the training data?

Task Background

There is an overview of the training dataset.

	# samples
True	415
False (Fake)	315

} the size is small !

Obtaining fake summary texts is challenging...
They are not commonly found in actual documents.

⇒ **Augmenting pseudo-fake data automatically**

Data Augmentation(1/2)

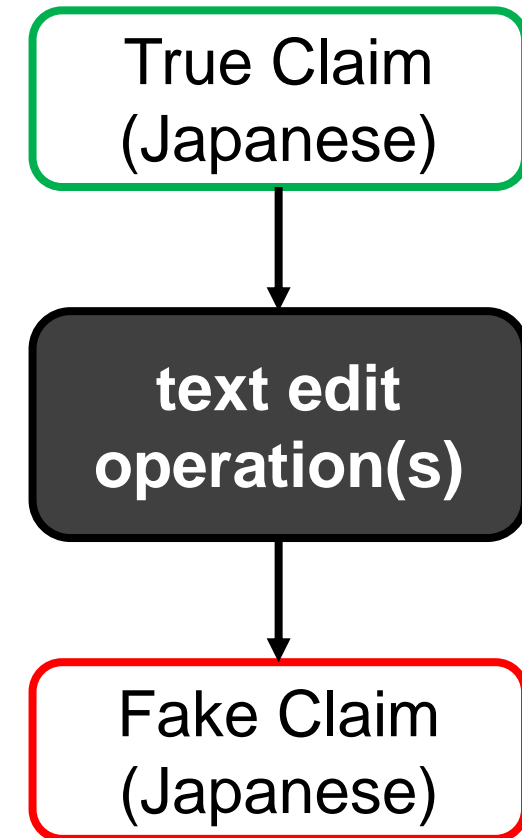
◆ Direct Manipulation

Generate a fake Japanese by operating on the true Japanese directly.



It may generate unnatural sentences when replacing one phrase with another forcefully.

This method is simple but forceful.

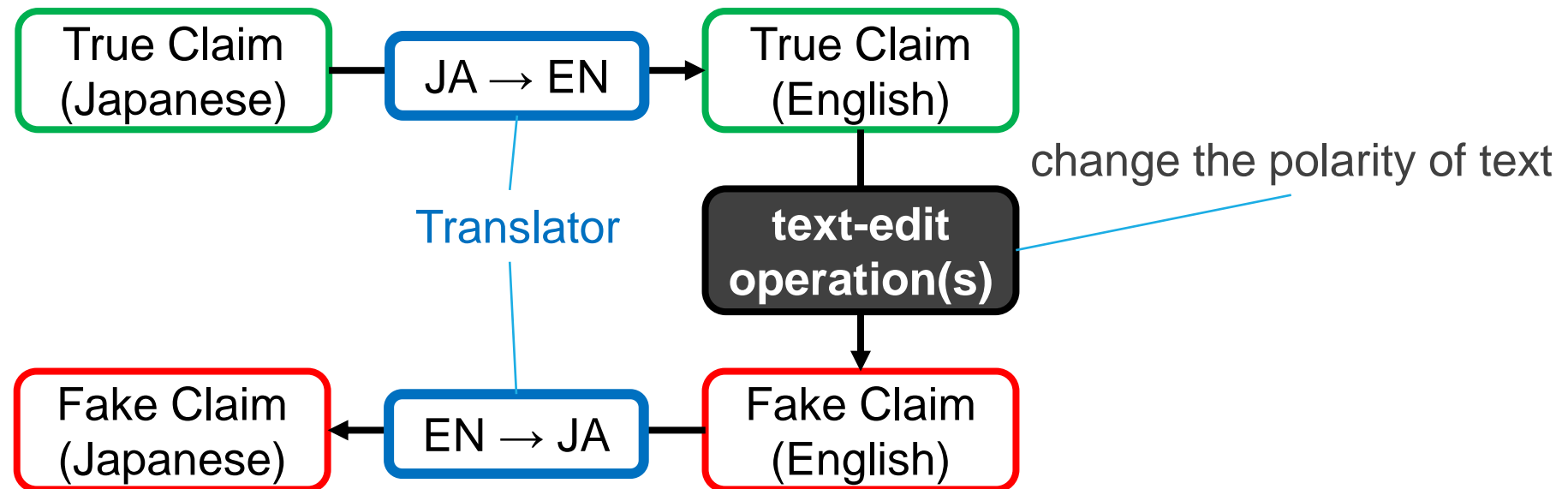


Data Augmentation(2/2)

◆ Round-Trip Manipulation (RTM)

Operate translated claims instead of directly operating.

- **Alleviate artifacts, More natural & fluent sentences**



Text Edit Operations (1/3)

◆ Insertion and Deletion of negation (NEG)

Original
claim

Only employees are eligible for special financial support.

**Add/remove "not" after verbs
(using the part-of-speech analysis)**

Fake
claim

*Only employees are **not** eligible for special financial ...*

Text Edit Operations (2/3)

◆ Conversion to antonyms (ANT)

Original
claim

Only employees are eligible for special financial support.

**Search and replace all adj. and adv.
to antonyms using WordNet**

Fake
claim

Only employees are ineligible for special nonfinancial support.

Text Edit Operations (3/3)

◆ Subject-Object Exchange (SOE)

Original
claim

Only employees are eligible for special financial support.

Identify and exchange the subject and object.

Fake
claim

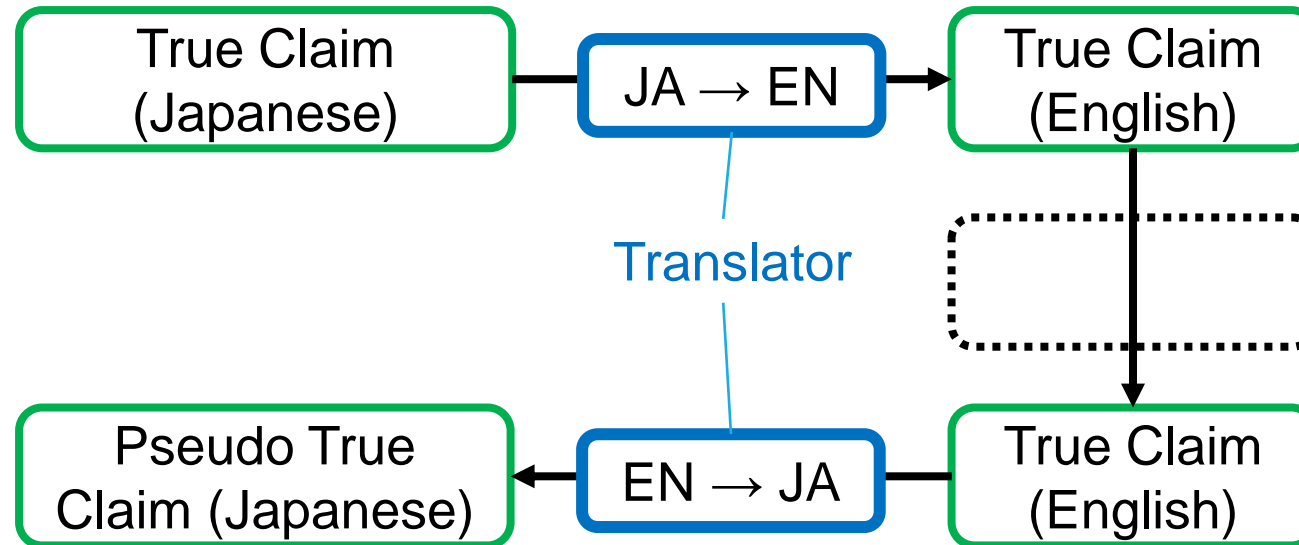
Only support are eligible for special financial employees.

Augmentation of True Data

◆ simple Round-Trip Translation (RTT)

By extending only the fake claims, the bias is introduced.

- So, we also generate pseudo true claims.



Example of generated data

Method	Example of generated sentence	True/Fake
ORG	国は有効な対策打ち出していない。 The <u>State</u> has not put forward <u>effective measures</u> .	True
RTT	政府は有効な対策を打ち出せていない。 The State has not put forward effective measures.	True
NEG	政府は効果的な対策を打ち出して <u>います</u> 。 The State has put forward effective measures.	Fake
RTM	その <u>対策</u> は、効果的な <u>政府</u> には至っていない。 The <u>measures</u> has not put forward effective <u>State</u> .	Fake
ANT	政府は <u>効果のない</u> 対策は打ち出していない。 The State has not put forward <u>ineffective measures</u> .	Fake

Experiment

For an answer verification model,
pre-trained BERT published by Inui lab. at Tohoku Univ^[1].

For RTM, we used

- ◆ DeepL API [2]
 - Translation in two directions (JA↔EN)
- ◆ NLTK [3]
 - Part-of-speech analysis
 - Text edit operations (using WordNet)

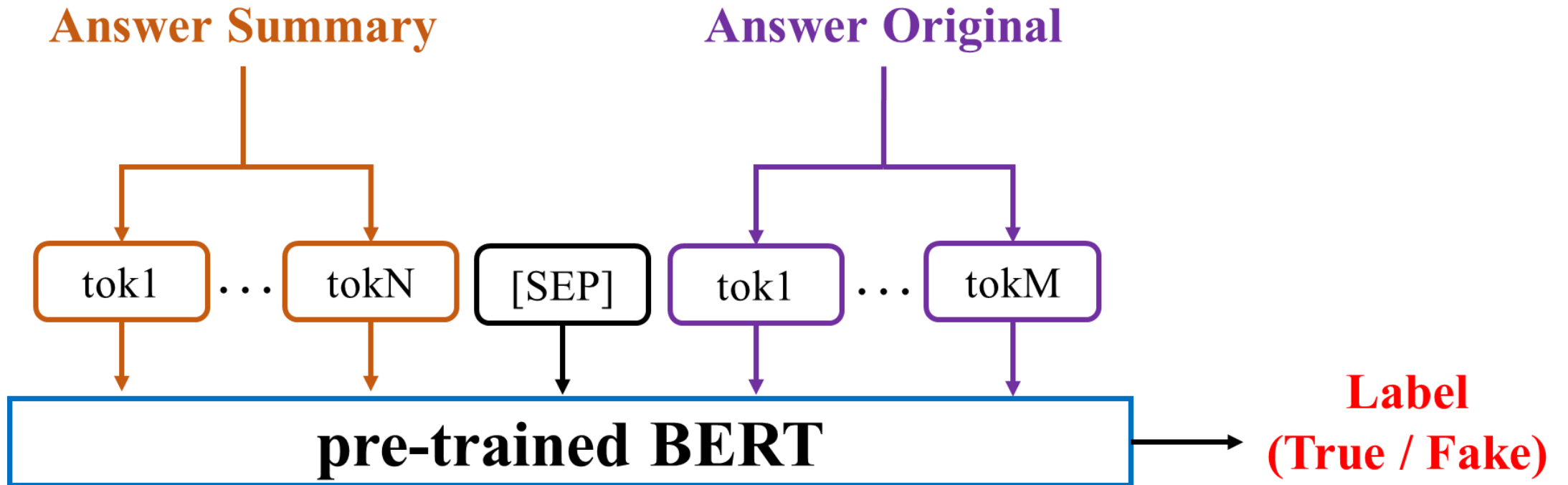
[1] <https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

[2] <https://www.deepl.com/pro-api>

[3] <https://www.nltk.org/>

Answer Verification model (Baseline)

We fine-tuned the pre-trained BERT with provided training data^[1].



Dataset

		Q&A pairs	true	fakes
	ORG	730	415	315
	RTT	415	415	0
PoliInfo-4	NEG	333	0	333
	SOE	265	0	265
	ANT	312	0	312
PoliInfo-3 (only true)		351	351	0
GDADC (selected)		785 (50)	0	785 (50)

RTM

Result of Formal-run (1/2)

No.	Model	trues	fakes	Acc.	F-measure
1	baseline (ORG)	415	314	0.79	0.8073
2	ORG / NEG	415	647	0.81	0.8403
3	ORG / SOE	415	579	0.82	0.8421
4	ORG / ANT	415	626	0.84	0.8620
5	ORG + RTT / NEG + SOE + ANT	830	1224	0.80	0.8148
6	PoliInfo-3 + ORG + RTT / NEG + SOE + ANT	1186	1224	0.88	0.8966

Result of Formal-run (1/2)

No.	Model	true	fake	Acc.	F-measure	
1	baseline (ORG)	415	314	0.79	0.8073	
2	<p style="text-align: center; color: red; font-weight: bold;">All models improved in accuracy compared to Baseline!</p> <p style="text-align: center; color: red; font-weight: bold;">best accuracy = 0.88</p>			0.81	0.8403	
3				0.82	0.8421	
4				0.84	0.8620	
5		ORG + RTT / NEG + SOE + ANT	850	1224	0.80	0.8148
6		PolInfo-3 + ORG + RTT / NEG + SOE + ANT	1186	1224	0.88	0.8966

Result of Formal-run (2/2)

Model	trues	fakes	Acc.	F-measure
Baseline (ORG)	415	314	0.79	0.8108
Proposed (PolilInfo-3 + ORG + RTT / NEG + SOE + ANT)	1186	1224	0.88	0.8966
Proposed + GDADC	1186	1999	<u>0.84</u>	<u>0.8620</u>
Proposed + GDADC (selected)	1186	1274	<u>0.84</u>	<u>0.8621</u>

Fine-tuning with GDADC did not yield an improvement in accuracy.

As mismatch between formal-run and human-created GDADC?

Thank you for listening.

All Result

	trues	fakes	accuracy	f1	ID
baseline(ORG)	415	314	0.79	0.8073	ID68
ORG+NEG	415	647	0.81	0.8403	
ORG+SOE	415	579	0.82	0.8421	
ORG+ANT	415	626	0.84	0.8621	
ORG+NEG+ANT	415	959	0.83	0.8468	
ORG+NEG+SOE	415	912	0.87	0.8889	
ORG+SOE+ANT	415	891	0.84	0.8621	
ORG+NEG+SOE+ANT	415	1224	0.84	0.8689	
ORG+RTT+NEG+ANT	830	959	0.79	0.8142	
ORG+RTT+NEG+SOE	830	912	0.85	0.8673	
ORG+RTT+SOE+ANT	830	891	0.81	0.8257	
ORG+RTT+NEG+SOE+ANT	830	1224	0.80	0.8148	
PolilInfo-3 +ORG+NEG+SOE+ANT	771	1224	0.85	0.8780	ID233
PolilInfo-3 +ORG+RTT+NEG+SOE+ANT	1186	1224	0.88	0.8966	ID70
PolilInfo-3 + ORG+RTT+NEG+SOE+ANT +GDADC	1186	1999	0.84	0.8620	
PolilInfo-3 +ORG+RTT+NEG+SOE+ANT +GDADC(selected)	1186	1274	0.84	0.8621	ID138

DM vs RTM (Pre-Experiment)

We created two models that were retrained with DM / RTM and tested these models with human-created test data.

Model	Precision	Recall	F-measure	Accuracy
Baseline	0.800	0.067	0.123	0.535
DM (ORG + NEG)	1.000	0.600	0.750	0.800
RTM (ORG + NEG)	0.927	0.850	0.887	0.892