SMU
SINGAPORE MANAGEMENT UNIVERSITY
School of Computing and Information Systems

Translate-Train Embracing Translationese Artifacts

Sicheng Yu[1], Qianru Sun[1], Hao Zhang[2,3], Jing Jiang[1],

[1]Singapore Management University, Singapore
[2]Nanyang Technological University, Singapore
[3] Agency for Science, Technology and Research, Singapore

ACL 2022

Translate-train

Augmenting training data with translated text

Training data: English $\xrightarrow{\text{machine translation}}$ Target languages

Testing data: Target languages

Core idea: mitigate the gap of unseen target languages.

Will translate-train bring other undesirable effect?

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

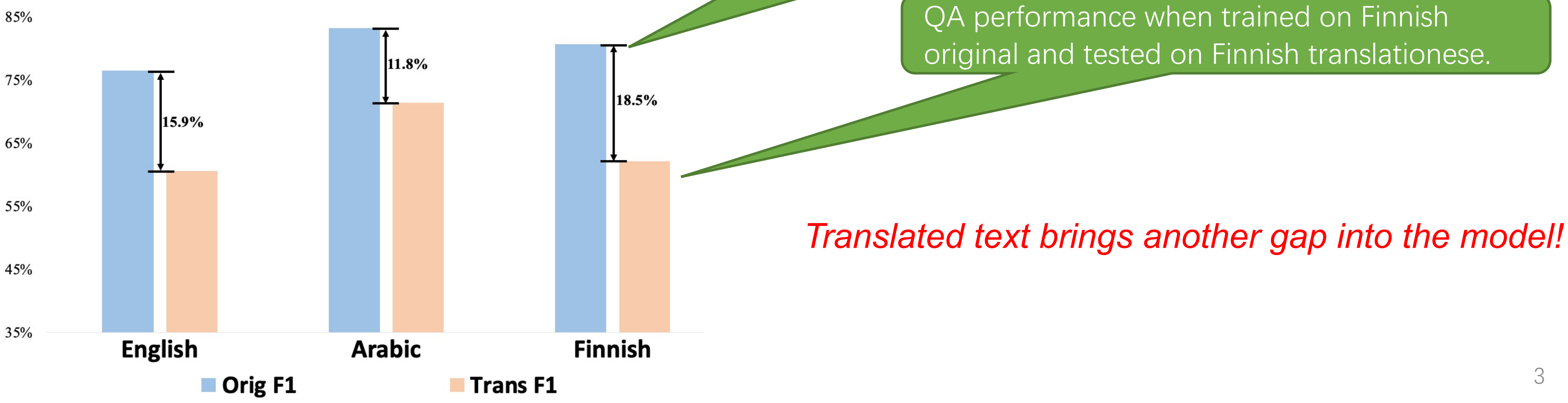
Translate-train

Original text v.s. Translated text

Original text: text directly written by humans --- denoted as originals

Translated text: text translated by humans or machine translators --- denoted as translationese

Exploration on TyDiQA:



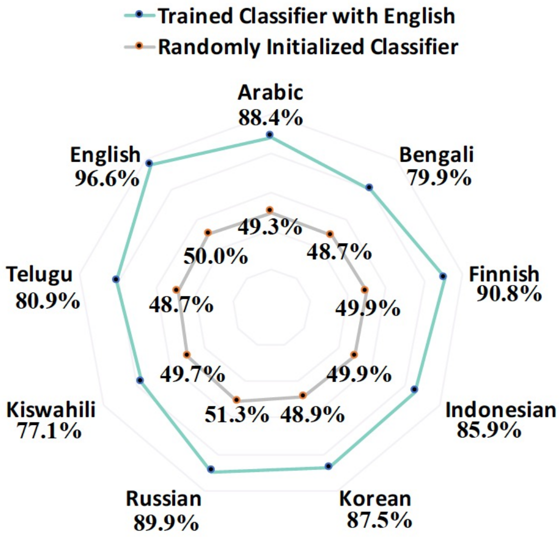
Translated text brings another gap into the model!

Translate-train

Hypothesis for originals-translationese gap

Training data: originals of English, translationese of target languages
Testing data in reality: originals of target languages

Hypothesis: whether the originals-translationese gap in English can be generalized to other languages?



Observations:

1. the patterns of translationese artifacts can be potentially learned to some extent.

2. model can likely transfer the learned patterns across different languages.

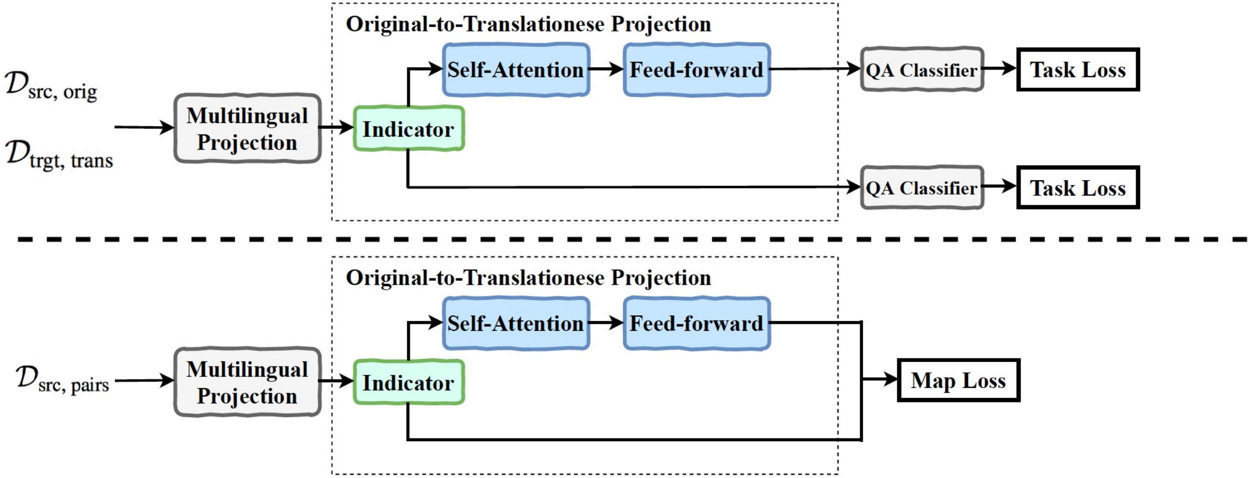
SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Our solution: TEA

Key idea of TEA (Translate-train Embracing Artifacts):

Learn the mapping function between originals and translationese on English and directly apply it on target languages.

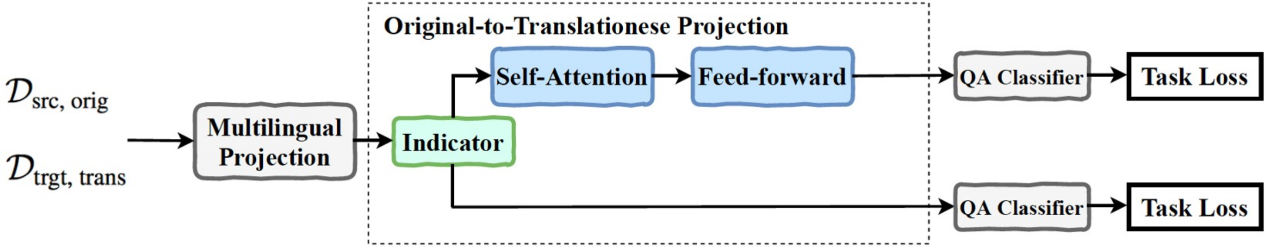
Training framework



SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Our solution: TEA

Training framework - Task loss

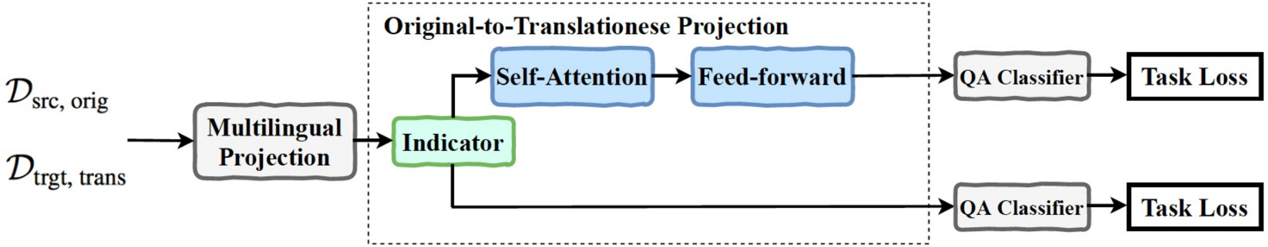


- Inputs: originals of source language (English) and translationese of target languages which is translated from English.
- Multilingual Projection (MP): XLM-R.
- Original-to-Translationese Projection (OTP): mapping the original domain to translationese domain if the inputs are originals which contains one layer of transformer structure.

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Our solution: TEA

Training framework - Task loss

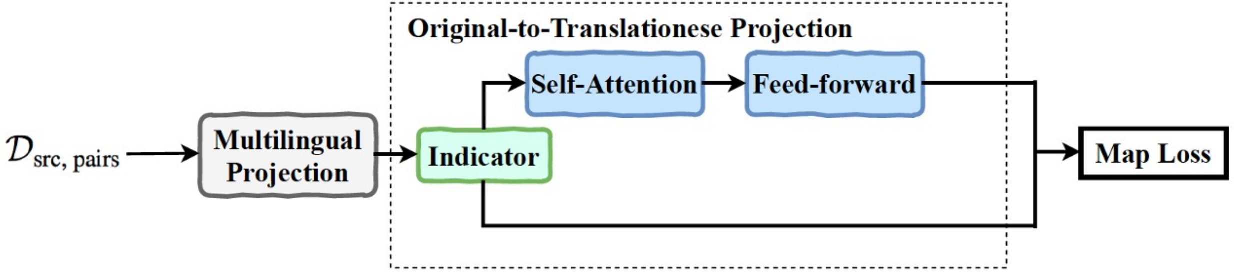


- QA Classifier (QA): classification layer for TyDiQA task.
- Task Loss: Cross-entropy loss.

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Our solution: TEA

Training framework - Map loss



- Inputs: originals of source language (English) and translationese of source language generated by back-translation.
- Map loss: maximize the cosine similarity of the representation between originals-translationese pair.

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Experiments

Main results

Method	D	ar	bn	fi	id	ko	ru	sw	te	*med*	*all-in-one*	*avg*
STT	✘	40.4/67.6	47.8/64.0	53.2/70.5	61.9/77.4	10.9/31.9	42.1/67.0	48.1/66.1	43.6/70.1	45.7/67.3	45.2/67.2	43.5/64.3
FILTER	✘	50.8/72.8	56.6/70.5	57.2/73.3	59.8/76.8	12.3/33.1	46.6/68.9	65.7/77.4	50.4/69.9	53.7/71.7	51.6/70.3	49.9/67.8
STT*	✘	58.0/76.6	54.6/70.2	59.0/74.8	64.7/80.2	48.0/61.6	49.5/71.2	58.7/74.6	57.0/76.2	57.5/74.7	56.8/74.4	56.2/73.2
TAG*	✔	56.9/76.4	55.5/70.0	59.4/75.2	64.4/79.6	48.6/61.7	49.1/70.4	60.7/76.0	57.8/76.4	57.4/75.5	56.9/74.5	56.5/73.2
TST*	✔	58.4/75.5	60.2/72.2	58.3/74.4	65.5/78.9	49.3/62.6	49.0/69.7	63.5/76.7	56.2/76.1	58.3/75.0	57.3/74.1	57.6/73.3
GRL*	✔	57.6/75.6	58.4/72.6	59.7/74.8	65.3/79.9	49.6/62.2	49.1/70.4	62.9/76.9	58.2/77.0	58.3/75.2	57.6/74.6	57.6/73.7
TEA*	✔	56.5/76.1	60.2/74.9	60.9/76.5	63.6/79.3	48.6/61.4	51.5/72.0	66.7/78.9	60.7/78.7	**60.5/76.3**	**58.6/75.6**	**58.6/74.7**

Baselines

- STT: Standard translate-train
- FILTER : SOTA Translate-train method
- TAG: Adding tag to denote originals or translationese
- TST: Adding another round of training only on originals
- GRL: Gradient reversal layer

Methods considering the gap between translationese and originals perform better.

TEA surpasses strong baselines.

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Experiments

Ablation study

Observation

- The improvement of our method is not caused by additional parameters or data.
- TOP still mitigates the artifacts, but OTP obtaining better performance.
- Our loss function and architecture are more effective.

Settings	**EM**	**F1**
STT	56.2	73.2
(1) STT+$\mathcal{X}_{\text{src, trans}}$	56.6	73.2
(2) STT+params	56.3	73.5
(3) TOP	57.9	74.1
(4) MLP in OTP	56.7	73.3
(5) MSE loss	58.0	73.9
Full method	**58.6**	**74.7**

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Conclusion

- We expose the drawback caused by translationese in translate-train and demonstrate that the pattern of translationese is transferrable.

- We propose a simple mapping method learned on English to mitigate the translationese artifacts.

- Our method outperforms translate-train baselines and SOTA translationese mitigation methods designed for machine translation.

Thank You!