

# TMUOU Submission for WMT20 Quality Estimation Shared Task

**Akifumi Nakamachi**  
Osaka University  
akifumi.nakamachi@ist.osaka-u.ac.jp

**Tomoyuki Kajiwara**  
Osaka University  
kajiwara@ids.osaka-u.ac.jp

**Hiroki Shimanaka**  
Tokyo Metropolitan University  
shimanaka-hiroki@ed.tmu.ac.jp

**Mamoru Komachi**  
Tokyo Metropolitan University  
komachi@tmu.ac.jp

## Introduction

- Quality Estimation (QE): estimate translation quality without reference sentences
- Ensembled XLM-RoBERTa regressors with language tags for QE shared task
- Ranked in the 3rd in Pearson correlation and 2nd in MAE and RMSE

## Result and Conclusion

	en-de	en-zh	et-en	ne-en	ro-en	si-en	ru-en	avg.
① CLS	0.455	0.490	0.747	0.742	0.860	0.646	0.693	0.662
② CLS + LangTag	0.461	0.440	0.738	0.751	0.873	0.658	0.689	0.659
③ CLS + Avg	0.419	0.465	0.744	0.763	0.874	0.648	0.701	0.652
④ CLS + Avg + LangTag	0.410	0.465	0.764	0.769	0.885	0.646	0.699	0.663
Ensemble	0.485	0.506	0.783	0.801	0.897	0.691	0.726	0.698

Table 1. Pearson's correlation on the development sets

- XLMRoBERTa based regressor are highly correlated to human annotation
- To-En pairs are relatively high correlation compared to From-En pairs
- Ensemble of different architecture increase the prediction performance

## Conclusion

- Apply XLM-RoBERTa to QE
- In the official evaluation, we ranked 3rd in Pearson and 2nd in MAE and RMSE

Team	MAE	RMSE	Avg Pearson
Bergamot-LATTE	0.408	0.527	0.718
TMUOU	0.418	0.543	0.686
IST and Unbabel	0.433	0.569	0.673
TransQuest	0.480	0.596	0.722
WL Research	0.538	0.683	0.546
IST and Unbabel	0.547	0.719	0.583
baseline	0.788	0.999	0.376
Bergamot-LATTE	0.895	1.062	0.489
nc	0.918	1.141	0.462

Table 2. Test result from official announce

## Methods

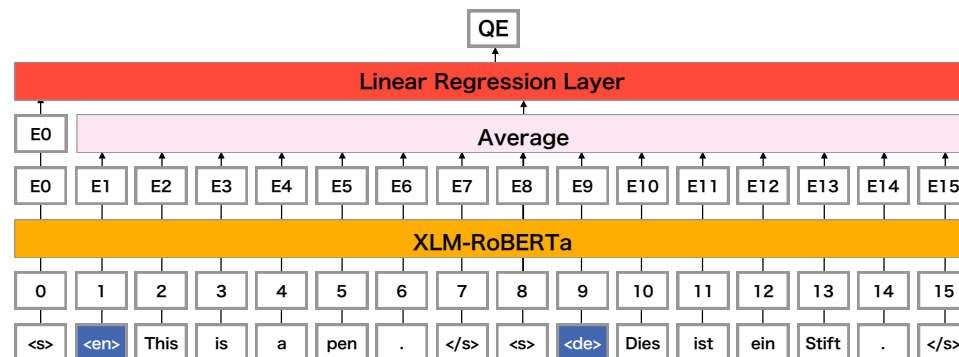


Fig 1. Overview of Regression Model

## XLM-Roberta Regressor (fig1)

- Add language tags to input
- Average encoder vectors

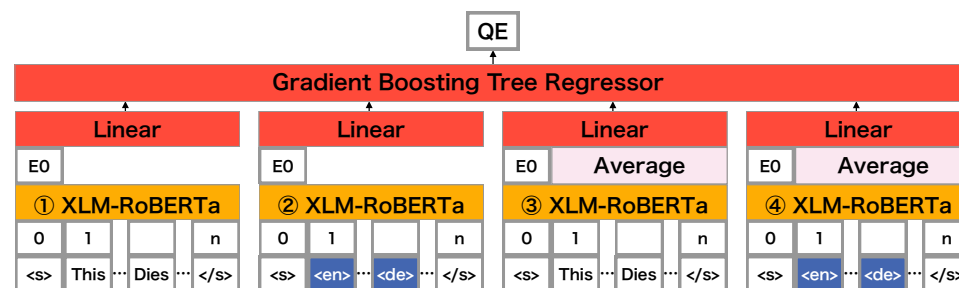


Fig 2. Overview of Ensemble Model

## Emsemble model (fig2)

- Combine different regressor architectures
- Emsemble prediction score and sentence generation probability
- Emsemble scores by boosting tree