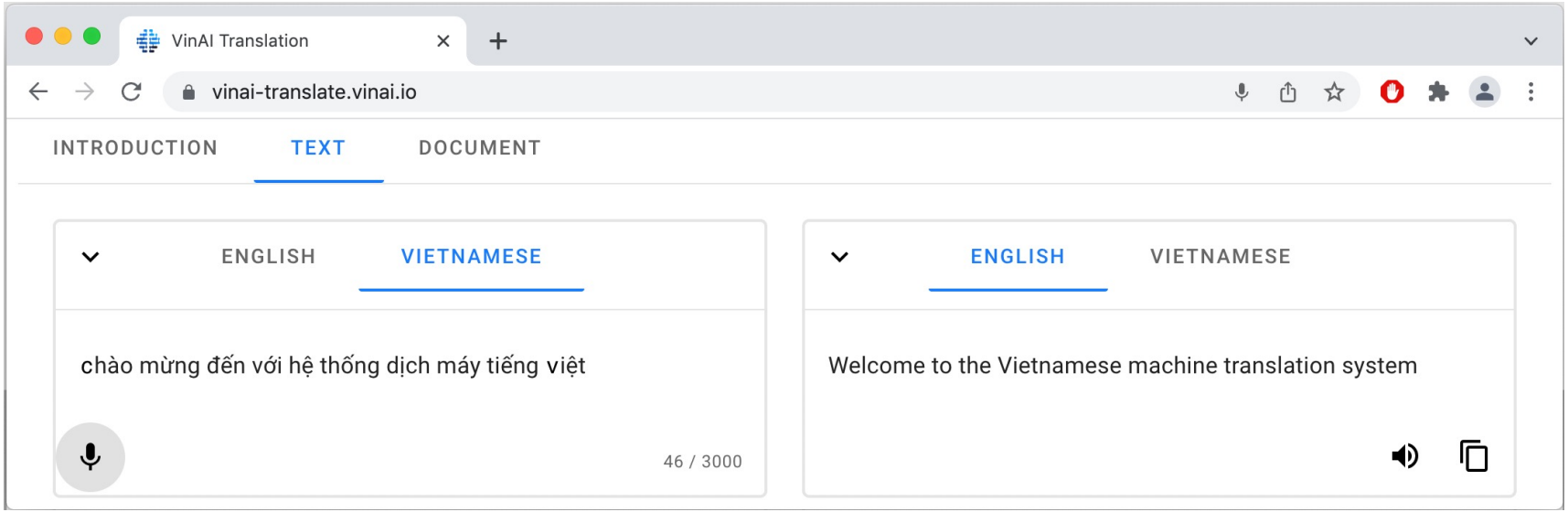
A Vietnamese-English Neural Machine Translation System

Dat Quoc Nguyen
VinAI Research

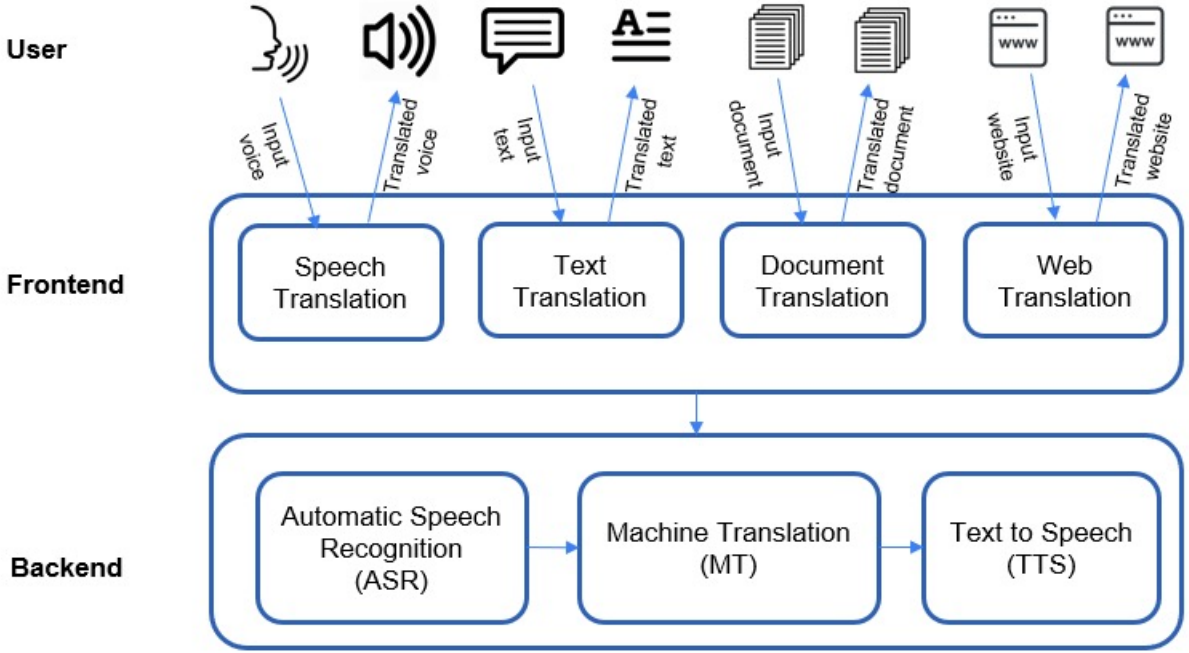
30/07/2022

Introduction



- Thien Hai Nguyen, Tuan-Duy H. Nguyen, Duy Phung, Duy Tran-Cong Nguyen, Hieu Minh Tran, Manh Luong, Tin Duy Vo, Hung Hai Bui, Dinh Phung and Dat Quoc Nguyen. **2022**. A Vietnamese-English Neural Machine Translation System. In *Proceedings of INTERSPEECH 2022: Show and Tell Demonstrations*.
- Long Doan*, Linh The Nguyen*, Nguyen Luong Tran*, Thai Hoang and Dat Quoc Nguyen. **2021**. PhoMT: A High-Quality and Large-Scale Benchmark Dataset for Vietnamese-English Machine Translation. In *Proceedings of EMNLP 2021*.

Introduction



- Employ modern ASR, MT and TTS approaches to build an application that helps translate speech and text from Vietnamese to English and vice versa at a high-level quality

Automatic Speech Recognition

- For Vietnamese
  - Train Conformer-CTC using an in-house 5700-hour dataset augmented by noise injection and intensity adjustment approaches
  - Obtain the word error rate (WER) at about 1.4% on an internal test set
- For English
  - Train Conformer-CTC on the Librispeech training set and obtain WER at 1.8% on the Librispeech test-clean set
- For inference in each language: Incorporate a 6-gram Byte-Pair-Encoding-based language model into the decoder to enhance the ASR performance

Machine Translation

- Approach: *Fine-tune the pre-trained Seq2Seq model mBART on a large-scale parallel dataset*
- Construct PhoMT, a high-quality and large-scale Vietnamese-English parallel dataset
    1. Collect parallel document pairs from 6 domains: TED Talks, Wikihow, MediaWiki, OpenSubtitles, News, Blog
    2. Pre-processing:
        - Manually inspect and remove low-quality document pairs from OpenSubtitles domain
        - Filter English paragraphs inside Vietnamese documents (and vice versa)
        - Perform sentence segmentation using VnCoreNLP and Stanford CoreNLP

Machine Translation

- Construct PhoMT, a high-quality and large-scale Vietnamese-English parallel dataset
  3. Align parallel sentence pairs:
     - Translate English source sentences into Vietnamese using Google Translate
     - Align between translated source sentences and target sentences using 3 toolkits: Hunalign, Gargantua, Bleualign
     - Select pairs that are aligned by at least 2/3 toolkits
  4. Post-processing:
     - Split the dataset into train/validation/test sets
     - Manually inspect validation and test sets and remove misaligned and low-quality sentence pairs (0.8%)

Machine Translation

- Construct PhoMT, a high-quality and large-scale Vietnamese-English parallel dataset, consisting of 3.02M sentence pairs

Domain	Total		Training			Validation			Test		
	#doc	#pair	#pair	#en/s	#vi/s	#pair	#en/s	#vi/s	#pair	#en/s	#vi/s
News	2559	41504	40990	24.4	32.0	257	22.3	30.3	257	26.8	34.5
Blogspot	1071	93956	92545	25.0	34.6	597	26.4	37.8	814	23.7	31.5
TED-Talks	3123	320802	316808	19.8	23.8	1994	20.0	24.6	2000	22.0	27.9
MediaWiki	38969	496799	490505	26.0	32.8	3024	25.3	32.3	3270	27.0	33.7
WikiHow	6616	513837	507379	18.9	22.4	3212	17.9	21.5	3246	17.5	21.5
OpenSub	3312	1548971	1529772	9.7	11.1	9635	9.5	10.7	9564	10.0	11.4
All	55650	3015869	2977999	15.7	19.0	18719	15.3	18.7	19151	16.2	19.8

Machine Translation

- Fine-tune mBART on the PhoMT training set of ~3M pairs for English-to-Vietnamese
- From each English-Vietnamese sentence pair in “noisy” datasets CCAligned and WikiMatrix
  - Employ the fine-tuned model to translate the English sentence into Vietnamese
  - Select pairs with a BLEU score between the Vietnamese-translated variant and the Vietnamese target sentence ranging from 0.15 to 0.95, resulting in 6M pairs
- A collection of 3M + 6M = 9M “high-quality” sentence pairs
- Simulate the ASR output: Lowercase and remove punctuations from the source sentences while keeping the target sentences intact, obtaining 9M pairs for each translation direction
- Fine-tune mBART for each translation direction using 9M + 9M = 18M sentence pairs

Text-to-Speech

- Convert the translated text into phonemes based on their pronunciation and text normalization rules
- Predict mel-spectrogram from input phonemes
  - Employ & modify Glow-TTS for Vietnamese, using a Vietnamese phoneme dictionary
  - Employ Tacotron2 for English
- Use HiFi-GAN to generate efficient and high-fidelity speech synthesis from the predicted mel-spectrogram

(PhoMT) Text Translation Results

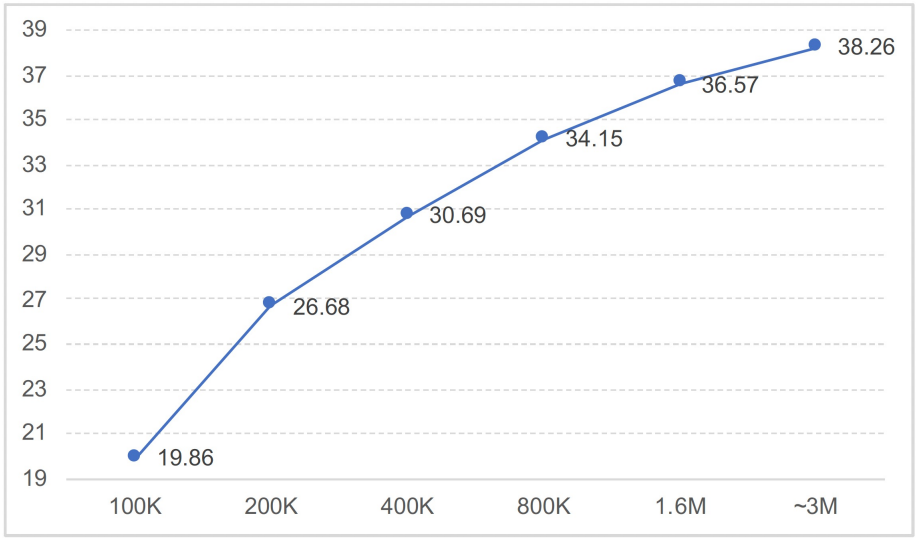
- Experimental results on the PhoMT validation and test sets while using the PhoMT training set of 2.97M pairs for training

Model	Validation set				Test set					
	En-to-Vi		Vi-to-En		En-to-Vi			Vi-to-En		
	TER↓	BLEU↑	TER↓	BLEU↑	TER↓	BLEU↑	Human↑	TER↓	BLEU↑	Human↑
Google Translate	45.86	40.10	44.69	36.89	46.52	39.86	23/100	45.86	35.76	10/100
Bing Translator	45.36	40.82	45.32	36.61	46.04	40.37	14/100	46.09	35.74	15/100
Transformer-base	42.77	43.01	43.42	38.26	43.79	42.12	13/100	44.28	37.19	13/100
Transformer-big	42.13	43.75	43.08	39.04	43.04	42.94	18/100	44.06	37.83	28/100
mBART	**41.56**	**44.32**	**41.44**	**40.88**	**42.57**	**43.46**	**32/100**	**42.54**	**39.78**	**34/100**

- mBART achieves the best performances, in both translation directions and on all metrics
- Neural MT baselines outperform automatic translation engines

(PhoMT) Text Translation Results

- BLEU scores of Transformer-base on the Vi- to-En validation set when varying training sizes on PhoMT



- Sample a set of 1.55M non-duplicate Vietnamese-English sentence pairs from OPUS's OpenSubtitles, which has the same size as the PhoMT's OpenSubtitles training subset:
  - OPUS's OpenSubtitles: 29.72 BLEU
  - PhoMT's OpenSubtitles: 31.11 BLEU

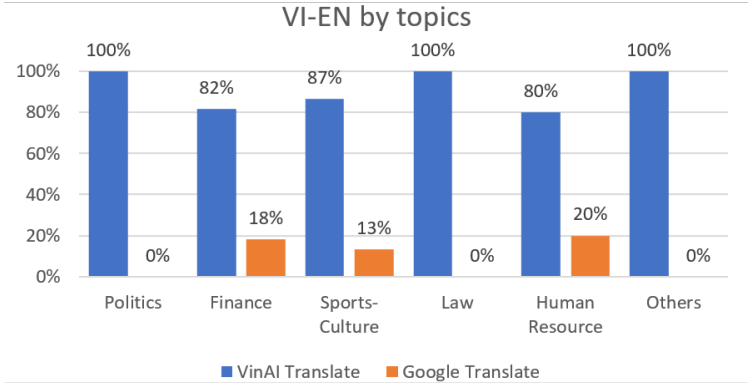
Our curation effort paid off!

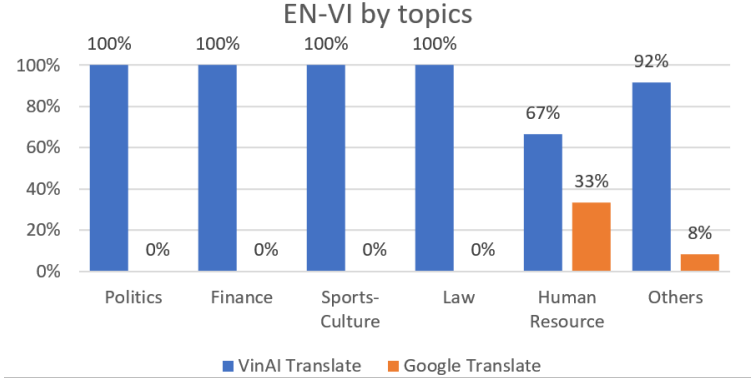
(VinAI Translate) Text Translation Results

- Automatic evaluation results

Model	Validation set		Test set	
	EN-VI	VI-EN	EN-VI	VI-EN
Google Translate	40.10	36.89	39.86	35.76
PhoMT [1]	44.32	40.88	43.46	39.78
VinAI Translate	**45.31**	**41.41**	**44.29**	**40.42**

- Human evaluation results





Demonstration Video

Conclusion

- Introduce VinAI Translate to translate speech and text between Vietnamese and English
- Automatic and human evaluation results show that the system obtains state-of-the-art performances
- Publicly release the pre-trained text translation models at: https://github.com/VinAIResearch/VinAI_Translate
- The system is available at: https://vinai-translate.vinai.io

Thank you!

@VinAI

https://www.vinai.io/