UBC Natural Language Processing Lab | Department of Linguistics | School of Information

One Model to Pronounce Them All: Multilingual Grapheme-to-Phoneme Conversion With a Transformer Ensemble

Kaili Vesik, Muhammad Abdul-Mageed, Miikka Silfverberg

SIGMORPHON 2020 Shared Task 1

Introduction

- Grapheme-to-phoneme (G2P) conversion is an important component of both speech recognition and synthesis.
- SIGMORPHON 2020 Shared Task 1 involves G2P for 15 languages, with the goal of converting written symbols to pronunciation symbols for words in any of the 15 languages.
- Our approach is inspired by earlier work on multilingual machine translation tasks; e.g., Johnson et al. (2017), Dong et al. (2015), Firat et al. (2016).

Contributions

- Develop a single model to perform G2P conversion on multiple languages.
- Combine several approaches to compensate for limited training data.
- Achieve best performance on Icelandic test data for Shared Task 1.

Shared Task Data

- Data provided by task organizers is extracted from Wiktionary.
- Languages: Adyghe (ady), Armenian (arm), Bulgarian (bul), Dutch (dut), French (fre), Georgian (geo), Modern Greek (gre), Hindi (hin), Hungarian (hun), Icelandic (ice), Japanese hiragana (jpn), Korean (kor), Lithuanian (lit), Romanian (rum), Vietnamese (vie).
- Per language: 4,050 gold labeled grapheme-phoneme pairs, split into a training set (3,600) and a development set (450). Blind test data consists of 450 sources.
- See sample pairs in Table 1.

Models

1. Fully Supervised Multilingual Model

- Transformer implemented using OpenNMT toolkit, with hyper-parameters following those adopted by Vaswani et al. (2017).
- Trained on data from all 15 languages, with language code token prepended to each grapheme sequence source.
- At inference we use an ensemble consisting of 4 training checkpoints from each of the models generated by 3 different random seeds.

2. Self-Trained Multilingual Model

- Employ self-training approach in order to augment training data.
- 1 million words from 12 of 15 languages sourced from Wikipedia articles, and duplicates removed.
- 35,418 words selected for self-training whose predicted targets have NLL > 0.2.
- Combine selected data with original data and re-train models using same hyper-parameters.

Language	Source	Target (IPA)
Alphabet:		
arm	ահեղ	ɑ h ɛ ʁ
	լյարժեք	l j ɑ ɾ ʒ ɛ kʰ
fre	front	f ʁ ɔ̃
	vêtu	v e t y
Alphasyllabary:		
hin	दिखावा	d̪ ɪ kʰ ɑː ʋ ɑː
	हटना	ɦ ə ʈ n ɑː
kor	개벽	k e̞ b j ʌ̹ k̚
	오빠	o̞ p̤ a̠
Syllabary:		
jpn	いなり	i n a̠ ɾʲ i
	やせん	j a̠ s ẽ̞ ɴ

Table 1: *Sample pairs from training data*

Results

- Average word error rates (WER) and phoneme error rates (PER) are lower than task baselines from provided monolingual models.
- Test set results (published by organizers) in Table 2.
- Our self-trained model achieved the lowest WER for Icelandic, of all submitted models.

	Best baseline		Multilingual		Self-trained	
Lang	WER	PER	WER	PER	WER	PER
ady	28.00	6.49	28.44	6.46	29.11	6.46
arm	14.22	3.29	13.11	2.98	12.89	3.07
bul	31.11	5.94	27.11	5.91	30.89	6.92
dut	15.78	2.89	15.78	2.98	16.89	3.07
fre	6.22	1.32	5.33	1.24	5.78	1.36
geo	26.44	5.14	26.00	5.25	26.67	5.23
gre	18.89	3.06	16.67	2.68	15.78	2.60
hin	6.67	1.47	6.44	1.58	6.67	1.66
hun	5.33	1.18	4.67	1.05	4.22	0.98
ice	10.00	2.21	9.56	2.11	9.11	1.83
jpn	7.33	1.79	6.00	1.44	6.00	1.40
kor	43.78	16.78	32.22	8.54	32.44	8.86
lit	19.11	3.55	19.33	3.63	20.00	3.68
rum	10.67	2.53	9.33	1.96	10.44	2.23
vie	4.67	1.52	4.89	1.66	4.00	1.28
avg			14.99	3.30	15.39	3.37

Table 2: *Blind test set results for best organizer-provided baseline models, as compare to our fully-supervised multilingual and self-trained multilingual models.*

Conclusions & Future Work

- An ensemble of multilingual transformers demonstrates success on multilingual G2P conversion.
- Due to time constraints, only a portion of available Wikipedia data was used for self-training. We did not see the results we had hoped for, but our future work will involve scaling up this augmented training data.

THE UNIVERSITY OF BRITISH COLUMBIA

Acknowledgements

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), the Social Sciences Research Council of Canada (SSHRC), and Compute Canada (www.computecanada.ca).