

GRADUAL LANGUAGE MODEL ADAPTATION USING FINE-GRAINED TYPOLOGY

MARCELL RICHARD FEKETE JOHANNES BJERVA SIGTYP - 6TH MAY 2023



CARLSBERGFONDET

THE PROBLEM

- Universally multilingual NLP is still out of reach
- Data-hungry training methods struggle to cover languages of smaller communities (Erzya, Faroese)
- Language model adaptation techniques exploit cross-lingual transfer
 - Fully fine-tuning model weights (Hu et al., 2020)
 - Inject extra weights (language adapters) (Pfeiffer et al., 2020, Üstün et al., 2022)
 - Retrain the lexical layer (Artetxe et al., 2020)



CROSS-LINGUAL TRANSFER

- Approaches often do not consider more intricate language relations than in-family membership
- Typological information like similarity may aid cross-lingual transfer (Pires et al., 2019; Üstün et al., 2022; de Vries et al., 2021)

Can we improve transfer by **exploiting typological similarity** for language model adaptation?

	Ordinals (53A)	Imperative (70A)	Imperative- Hortative (72A)
Faroese	7	1	4
Icelandic	7	2	4
Swedish	7	4	4
Estonian	7	1	1
Finnish	7	1	1
Hungarian	7	5	1



GRADUAL ADAPTATION

- Instead of direct adaptation from source to target language
- Carry out language model adaptation in stages using intermediate languages
 - Rank candidate languages using some typological similarity measure
 - 2) Maintain **typological similarity between languages** used in the adaptation stages

More data-efficient language model adaptation





MEASURING TYPOLOGY

- Using hand-crafted resources like WALS (Dryer and Haspelmath, 2013)
 - Certain feature values are missing
 - Values are coarse-grained (Ponti et al., 2019)
- Alternative: data-driven measures like counts of UD dependency links ("structural similarity", Bjerva et al., 2019)
 - More fine-grained information
 - More precise ranking of languages

af	1.00											
da	0.94	1.00										
	0.99	0.94	1.00									
en	0.96	0.99	0.96	1.00								
Q	0.71	0.77	0.66	0.75	1.00							
de	0.73	0.71	0.68	0.70	0.91	1.00						
got	0.71	0.79	0.66	0.76	0.99	0.91	1.00					
<u>.</u>	0.94	0.92	0.95	0.93	0.79	0.80	0.78	1.00				
spu	0.98	0.95	1.00	0.98	0.68	0.68	0.68	0.95	1.00			
ou	0.92	0.99	0.91	0.98	0.81	0.73	0.83	0.90	0.93	1.00		
SV	0.93	1.00	0.92	0.98	0.80	0.72	0.82	0.90	0.94	1.00	1.00	
SW	0.57	0.49	0.52	0.50	0.79	0.87	0.77	0.68	0.51	0.51	0.51	1.00
	af	da	nl	en	fo	de	dot	is	nds	no	sv	sw/



OUTCOMES & FUTURE WORK

- More data-efficient language model adaptation efforts
- Assess correlation between typological similarity measures and transfer
- Extend coverage and experiment with several language model adaptation techniques











GRADUAL LANGUAGE MODEL ADAPTATION USING FINE-**GRAINED TYPOLOGY**

MARCELL RICHARD FEKETE

mrfe@cs.aau.dk

direct adaptation



CARLSBERGFOND





gradual adaptation



0.99 0.91 1.00

1.00 0.98 0.68 0.68 0.68 0.95 1.00

0.81 0.73 0.83

0.90 0.93 1.00

0.99 0.94 1.00 0.96 0.99 0.96 1.00

0.68

0.95 0.93

0.91 0.98



AALBORG UNIVERSITET

ENSEMBLING ADAPTERS

- Pooling language adapters to encourage cross-lingual transfer
 - Wang et al. (2021); Chronopoulou et al. (2023)
- Weighting based on typological similarity
 - 1) Select an appropriate ranking function to assign α weights
 - 2) Select an appropriate **pooling function**





LEXICAL LAYER ADAPTATION

- Retraining the lexical layer of pretrained language models improves performance on the target language (Rust et al., 2021)
- Lexically overlapping tokens may catalyse transfer (Pfeiffer et al., 2021)
- Improve transfer by maintaining a high degree of vocabulary overlap between adaptation stages



REFERENCES

- Artetxe et al. 2020. Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond.
- Bjerva et al. 2019. What Do Language Representations Really Represent?
- Dryer and Haspelmath. 2013. The World Atlas of Linguistic Structures.
- Hu et al. 2020. XTREME: A Massively Multilingual Multi-task Benchmark for Evaluating Cross-lingual Generalisation.
- Joshi et al. 2020. The State and Fate of Linguistic Diversity and Inclusion in the NLP World.
- Pfeiffer et al. 2020. MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer.
- Pires et al. 2019. How Multilingual is Multilingual BERT?
- Ponti et al. 2019. Modeling Language Variation and Universals: A Survey on Typological Linguistics for Natural Language Processing.
- Üstün et al. 2022. UDapter: Typology-based Language Adapters for Multilingual Dependency Parsing and Sequence Labeling.
- de Vries et al. 2021. Adapting Monolingual Models: Data can be Scarce when Language Similarity is High.

