Domain Mismatch Doesn't Always Prevent Cross-Lingual Transfer Learning

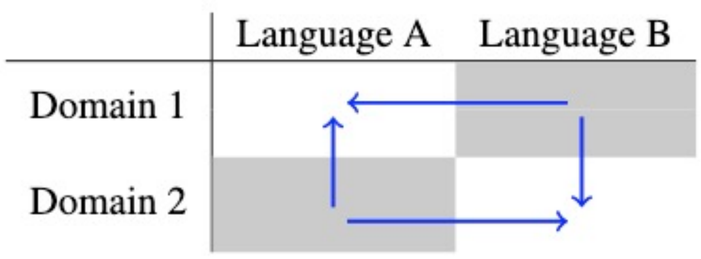
Daniel Edmiston, Phillip Keung, and Noah A. Smith

Introduction

- Many publications assert that domain mismatch prevents zero-shot cross-lingual transfer learning:

	Language Pair	Domain	Score	Δ
UBLI			*Accuracy@1*	
Søgaard et al. (2018)	En-Es	Euro/Euro	61.0	
		Euro/Wiki	0.1	–60.9
	En-Hu	Wiki/Wiki	6.7	
		Wiki/Euro	0.1	–6.6
UNMT			*BLEU*	
Marchisio et al. (2020)	Fr-En	UN/UN	27.6	
		UN/CC	3.3	–24.3
	Ru-En	UN/UN	23.7	
		UN/CC	0.7	–23.0

- **Our Contribution**: We show that previous work overstates the problem; careful joint-pretraining induces cross-lingual + cross-domain transfer



UNMT

- Unsupervised translation without parallel data in Lample et al. (2018)
- Strong UNMT results in domain-matched scenario disappear with domain mismatch
- Joint pre-training again recovers cross-lingual transfer in domain-mismatched scenario

Task		Monoling. Data	BLEU	Monoling. Data	BLEU
En UN → Fr UN	UNMT Baseline	En Wiki, Fr UN	3.99	En UN, Fr Wiki	3.75
	w/ Joint Pre-train.		25.45		12.76
	Δ		+21.46		+9.01
Fr UN → En UN	UNMT Baseline	Fr Wiki, En UN	3.99	Fr UN, En Wiki	3.70
	w/ Joint Pre-train.		20.91		14.19
	Δ		+16.92		+10.49
En UN → Ru UN	UNMT Baseline	En Wiki, Ru UN	1.50	En UN, Ru Wiki	0.79
	w/ Joint Pre-train.		10.83		4.52
	Δ		+9.33		+3.73
Ru UN → En UN	UNMT Baseline	Ru Wiki, En UN	1.27	Run UN, En Wiki	1.24
	w/ Joint Pre-train.		7.95		7.63
	Δ		+6.68		+6.39

- Gains more pronounced for closer language pairs
- Performance depends on whether monolingual data from same domain as task's target language

Bilingual Lexicon Induction

- MUSE learns a rotation matrix to align word embeddings across language pairs in an unsupervised way
- Mismatched domains cause cross-lingual alignment to fail
- Joint pre-training recovers losses:

Source Domain-Target Domain	Es-En	En-Es	Fr-En	En-Fr	Ru*-En	En-Ru*
Matched Domain						
Wiki - Wiki	81.8	82.5	81.3	82.2	59.5	64.0
UN - UN	68.7	70.8	74.2	75.2	55.2	56.7
Mismatched Domain						
Wiki - UN	0.1	0.2	35.2	33.8	16.6	0.1
Wiki - UN w/ Joint Pre-training	65.2	56.9	68.3	54.4	28.0	20.1
Δ	+65.1	+56.7	+33.1	+20.6	+11.4	+20.0

- Joint pre-training recovers cross-lingual transfer better than homograph anchors:

	Language Pair	Domain	Acc@1	Copying Baseline	Δ
Søgaard et al. (2018)	En-Es	Wiki/Euro	25.5	32.8	-7.3
	En-Fi		10.1	28.8	-18.7
	En-Hu		9.2	29.6	-20.4
Joint pre-training (this paper)	En-Es	Wiki/UN	56.9	32.8	+24.1
	Es-En		65.2	29.7	+35.5
	En-Fr		54.4	41.3	+13.1
	Fr-En		68.3	42.9	+25.4
	En-Ru		20.1	3.9	+16.2
	Ru-En		28.0	0.0	+28.0

Cross-lingual SWS

- Cosine similarity of cross-lingual word embeddings should be highly correlated with human judgment
- e.g., *leaf-Baum* are similar, *leaf-Hund* are not

	En-De	En-Es	En-Fa	En-It
Wiki-Wiki	0.45	0.54	0.25	0.49
UN-Wiki w/o Joint PT	0.02	-0.01	0.01	-0.05
UN-Wiki w/ Joint PT	0.43	0.47	0.23	0.45
Δ	+0.41	+0.48	+0.22	+0.50

- As before, joint pre-training induces cross-lingual + cross-domain transfer
- Joint pre-training also effective for distant language pairs (En-Fa)

Conclusion

- Matched domain corpora not required for cross-lingual transfer
- Joint pre-training sufficient to induce cross-lingual, cross-domain transfer

Corresponding authors: {danedm,keung}@amazon.com