Choosing Transfer Languages for Cross-Lingual Learning

Yu-Hsiang Lin, Chian-Yu Chen, Jean Lee, Zirui Li, Yuyan Zhang, Mengzhou Xia, Shruti Rijhwani, Junxian He, Zhisong Zhang, Xuezhe Ma, Antonios Anastasopoulos, Patrick Littell*, Graham Neubig
Language Technologies Institute, Carnegie Mellon University; *National Research Council, Canada

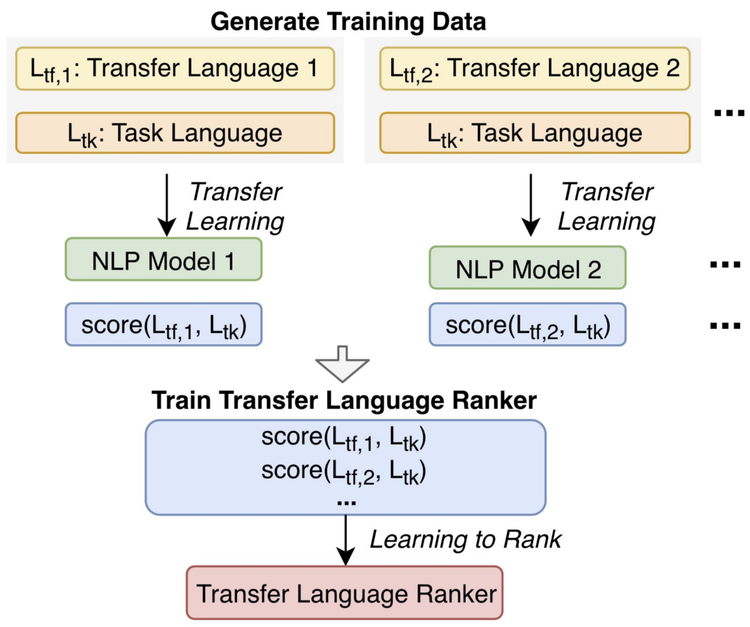
Carnegie Mellon University School of Computer Science

NRC · CNRC

Cross-lingual Transfer

- Better learning for low-resource ***task languages*** by training on related high-resource ***transfer language***
- **Question 1:** What transfer language do we choose?
  - e.g., when processing Azerbaijani, do we choose French, with lots of data, or Turkish with similar typology?
- **Question 2:** Can we do this process automatically?
  - Given a new task language, choose a transfer language without training.

Experimental Paradigm



1. Perform large-scale experiments, exhaustively transferring from/to many different languages.

v extra v \ trg >>	ara	aze	bel	ben	bos	bul	ces	cmn	dan	deu
ara	0.2838	0.0645	0.0987	0.0838	0.164	0.3575	0.2465	0.1742	0.3837	0.3211
aze	0.286	0.031	0.0324	0.03	0.0433	0.3637	0.247	0.1753	0.3673	0.3336
bel	0.2854	0.0314	0.0285	0.0317	0.0412	0.362	0.2456	0.1733	0.3699	0.3291
ben	0.2867	0.0314	0.0311	0.0279	0.0424	0.3634	0.2449	0.1703	0.3741	0.3296
bos	0.2878	0.0333	0.0314	0.0297	0.0293	0.3575	0.2522	0.1744	0.3715	0.3211
bul	0.28	0.0714	0.114	0.0947	0.1932	0.3578	0.256	0.1735	0.3946	0.3245
ces	0.2805	0.0667	0.1247	0.101	0.1881	0.364	0.2507	0.1757	0.3828	0.3251
cmn	0.2766	0.0643	0.0842	0.0886	0.1355	0.3549	0.2421	0.17	0.3692	0.3199
dan	0.2862	0.0639	0.1109	0.0909	0.1662	0.3617	0.2513	0.1755	0.3773	0.3343
deu	0.28	0.0668	0.1143	0.0978	0.1632	0.3601	0.2489	0.174	0.3951	0.3315
ell	0.2803	0.0668	0.1077	0.0845	0.1851	0.3568	0.2522	0.175	0.3888	0.3284
epo	0.2855	0.0303	0.0356	0.0351	0.0402	0.3597	0.249	0.1763	0.3698	0.3316

2. Learn a ranking model, ***LangRank***, to recommend transfer languages, from best to worst, for a new task language.
3. From analysis, gain insights into what makes a good transfer languages.

1 *LangRank* Model for Choosing Transfer Languages

- Two Categories of Features
  - **Linguistic features** (dataset independent)
    - From URIEL Typological Database
    - Distance between languages deriving from geography, language family tree, phonological features, and syntactic structures
  - **Statistical features** (dataset dependent)
    - Dataset size
    - Word overlap: $\frac{\#\text{ common tokens}}{\sum_{corpora} \#\text{ tokens}}$
    - Type-token ratio (TTR): $\frac{\#\text{ unique tokens}}{\#\text{ tokens}}$

- **Ranking model**
  - Gradient boosted decision trees
  - Trained with LambdaRank, optimizing normalized discounted cumulative gain $\text{DCG}@p = \sum_{i=1}^{p} \frac{2^{\gamma_i} - 1}{\log_2(i+1)}$
  - Interpretable, reveal important linguistic or statistical features

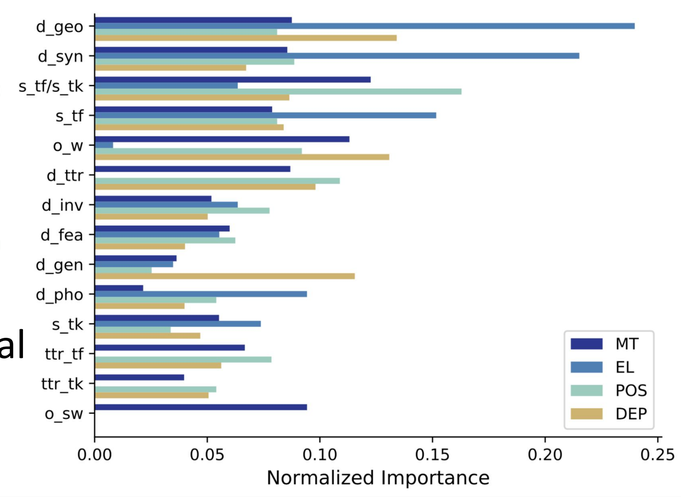
- **NLP tasks, dataset, # task languages/# transfer languages**
  - Machine translation (MT), TED talk corpus, 54/54
  - Entity linking (EL), Rijhwani 2019 corpus, 9/53
  - Part-of-speech tagging (POS), Universal Dependencies v2.2, 26/60
  - Dependency parsing (DEP), Universal Dependencies v2.2, 30/30

2 Ranking Evaluation in NDCG@3

Method	MT	EL	POS	DEP
Best baseline by linguistic feature	d_{gen} 24.2	d_{gen} 50.9	d_{geo} 15.7	d_{geo} 46.4
Best baseline by statistical feature	o_{sw} 29.2	o_w 30.7	o_w 13.4	o_w 52.3
LANGRANK	**53.7**	**63.0**	**28.9**	**65.0**

3 Feature Importance for Different NLP Tasks

- Most influential features
  - MT: dataset size, word overlap
  - EL: geographic, syntactic
  - POS: dataset size, TTR
  - DEP: geographic, word overlap
- Linguistic features most influential for EL as it has small dataset and lacks rich statistical features



4 Closer Look at Examples

Task Lang	LANG RANK	Best Dataset	Best URIEL	True Best
MT aze	tur (1) fas (3) hun (4)	o_w tur (1) hrv (5) ron (31)	d_{fea} ara (32) fas (3) sqi (22)	tur (1) kor (2) fas (3)
MT ben	hun (1) tur (2) fas (4)	o_w vie (3) ita (20) por (18)	d_{geo} mya (30) hin (27) mar (41)	hun (1) tur (2) vie (3)
EL tel	amh (6) orm (40) msa (7)	o_w amh (6) swa (32) jav (9)	d_{inv} pan (2) hin (1) ben (5)	hin (1) pan (2) mar (3)

Top-3 recommended transfer languages

Ground-truth ranking

- By considering ***all*** linguistic and statistical features jointly, ***LangRank*** outperforms strong baseline in task *aze*
- Even in hard case *ben* where baselines fail, ***LangRank*** remains good
- In small EL dataset, ***LangRank*** overfits and fails task *tel*, while linguistic features serve as good fallback

5 Best Performance Attainable in Top K Recommendations

- If we train models on the top K transfer languages suggested by the ranking model, and pick the best one, how good a model do we expect to get?