Distributed Representations for Croos-lingual Lexicon Induction

Abstract

write me

1 Introduction

Motivate low-resource MT, cite our EACL work.

Say that inducing translation lexicons is an important part.

Say that the standard approach (starting tithe Rapp 95) uses vector space contextual similarity and a bilingual dictionary to project between languages.

Say that there are two problems with those methods: (1) defending the feature representations and a a metric to measure similarity is largely a heuristic task, (2) representations are quite large (on the order of vocal size).

One alternative proposed in (**?**) addresses these issues by inducing the *same* embedding for words in both languages, so that words which are semantically similar are “near” one another in the

2 Previous Work

write me

2.1 Model

write me

2.2 Experiments

Table 1 shows performance on the lexicon induction task. The alignment dictionary score is the performance of the dictionary derived from the intersection alignments over the training data alone, which is used as supervision to both the old contextual scorer and the distributed representations learner. The fact that the accuracy using the alignment based dictionary alone is so low speaks to how noisy the alignments are and how limited the training data is. The old contextual score uses the same dictionary based on the intersection alignments over the training data for each language to project context vectors. The distributed representations use an interaction matrix defined also by the intersection alignments over the training data for each language. Both models use the same tokenization of all of the monolingual data that we have available for each language, which is taken from web crawls and Wikipedia. Evaluation is over *all word types* in the development set for each language.

	Top-1	Top-10	Top-100
Tamil			
Intersection Train Dict	6.70	9.58	9.60
Old-Contextual	2.32	8.38	25.44
Distrib Rep L2 Dist	15.50	17.77	20.44
Distrib Rep Learn Dist			
Bengali			
Intersection Train Dict	8.60	11.39	11.39
Old-Contextual	3.91	12.39	30.53
Distrib Rep L2 Dist	24.01	25.86	28.01
Distrib Rep Learn Dist			
Hindi			
Intersection Train Dict	13.51	18.38	18.38
Old-Contextual	5.22	14.72	34.31
Distrib Rep L2 Dist	33.93	37.64	42.00
Distrib Rep Learn Dist			

Table 1: Comparison of performance of old definition of contextual similarity with new distributed representations model