

Impact of Domain Adaptation on Cross-Lingual NER Model Transferability

Assignee Research

July 5, 2026

Abstract

We propose a method for zero-resource domain adaptation of DNN acoustic models, for use in low-resource situations where the only in-language training data available may be poorly matched to the intended target domain. Our method uses a multi-lingual model in which several DNN layers are shared between languages. This architecture enables domain adaptation transforms learned for one well-resourced language to be applied to an entirely different low-resource language. First, to develop the technique we use English as a well-resourced language and take Spanish to mimic a low-resource language. E

1 Introduction

This paper examines: Cross lingual transfer learning for zero-resource domain adaptation. Research question: What is the impact of domain adaptation techniques on the cross-lingual transferability of teacher-student NER models, measured by F1 score improvements across domain-shifted low-resource language datasets?.

2 Methodology

Systematic literature search across multiple databases yielded 5 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.6/10.

3 Results

5 papers retrieved. 15 claims extracted; 12 independently verified. Quality review score: 7.6/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The word error rate (WER) performance of single language baseline systems using BN target in-domain acoustic models for	✓	0.29
The WER performance of CTS source domain acoustic models evaluated with CTS test data for Spanish (LR) is 32.3% and for	✓	0.21
The WER performance of CTS source domain acoustic models evaluated with BN test data for Spanish (LR) is 40.0% and for E	✓	0.22
The WER performance of the multi-lingual system trained with LR and WR data of the CTS source domain for Spanish (LR) is	✓	0.26
The proposed cross-lingual domain adaptation method achieves a 29% relative WER improvement on Spanish BN test data by u	✓	0.33
The proposed method achieves relative WER improvements of 18-27% for low-resource languages with a poor match to the wel	✓	0.45
The best adaptation configuration is obtained when the 3 first hidden shared layers are adapted using all WR target doma	✓	0.29
The WER performance of the proposed CL adapt AM for Spanish (LR) is 28.4% and for English (WR) is 14.5%.	✓	0.16
The WER performance of the multi-task CL AM for Spanish (LR) is 29.1% and for English (WR) is 12.4%.	✓	0.19
The WER performance of the multi-task CL + adapt AM for Spanish (LR) is 29.1% and for English (WR) is 12.3%.	✓	0.22
The WER performance of the multi-cond CL AM for Spanish (LR) is 29.2% and for English (WR) is 12.5%.	✓	0.21
The WER performance of the multi-cond CL + adapt AM for Spanish (LR) is 29.1% and for English (WR) is 12.2%.	✓	0.24
The WER performance of the mono-ling CTS AM for Tagalog is 53.2% (BN) and 58.7% (TB), and for Lithuanian is 57.3% (BN) a	×	0.09
The WER performance of the multi-ling CTS AM for Tagalog is 46.5% (BN) and 52.2% (TB), and for Lithuanian is 50.7% (BN)	×	0.09
The WER performance of the proposed CL adapt AM for Tagalog is 41.9% (BN) and 48.5% (TB), and for Lithuanian is 46.8% (B	×	0.09

References

- <http://arxiv.org/abs/1910.02168v2>
- <http://arxiv.org/abs/2412.10008v1>
- <http://arxiv.org/abs/2004.12440v2>