Comparison of Inference Efficiency Between Task-Specific Pre-Trained Models and Zero-Shot Cross-Lingual Retrieval Models on

Assignee Research

July 13, 2026

Abstract

Using task-specific pre-training and leveraging cross-lingual transfer are two of the most popular ways to handle code-switched data. In this paper, we aim to compare the effects of both for the task of sentiment analysis. We work with two Dravidian Code-Switched languages - Tamil-Engish and Malayalam-English and four different BERT based models. We compare the effects of task-specific pre-training and cross-lingual transfer and find that task-specific pre-training results in superior zero-shot and supervised performance when compared to performance achieved by leveraging cross-lingual transfe

1 Introduction

This paper examines: Task-Specific Pre-Training and Cross Lingual Transfer for Code-Switched Data. Research question: How does the inference efficiency of task-specific pre-trained models compare to zero-shot cross-lingual retrieval models when evaluated on code-switched data using the GLUE benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.5/10.

3 Results

12 papers retrieved. 15 claims extracted; 10 independently verified. Quality review score: 7.5/10.

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 Malayalam-English and Tamil-English datasets are used for a three-class sentiment classification problem (positive,	✓	0.15
Previous work by Chakravarthi et al. (2020c) treated the Malayalam-English and Tamil-English sentiment analysis problem	✓	0.24
The best performing systems for the Tamil-English and Malayalam-English shared tasks were built on top of BERT variants.	✓	0.23
Chakravarthi et al. (2020b) and (2020a) provided baseline results using algorithms including Support Vector Machines, De	✓	0.26
The TweetEval sentiment classifier was trained on a dataset of English Tweets.	×	0.13
The Tamil-English and Malayalam-English datasets were created by scraping Youtube comments.	✓	0.16
mBERT and XLM-RoBERTa models are trained on more than 100 languages.	✓	0.18
The Tamil-English dataset contains 10,559 Positive, 2,037 Negative, and 850 Neutral samples.	×	0.10
The Malayalam-English dataset contains 2,811 Positive, 738 Negative, and 1,903 Neutral samples.	×	0.09
The Hinglish (Sentimix) dataset contains 6,616 Positive, 5,892 Negative, and 7,492 Neutral samples.	×	0.11
The uncased-base XLM-RoBERTa model was trained on 2.5TB of webcrawled data.	✓	0.28
The TweetEval sentiment classification model was trained on a dataset of 60M English Tweets.	✓	0.19
The underlying RoBERTa model used in TweetEval has 160M parameters.	×	0.13
The study evaluates results based on weighted average scores of precision, recall, and F1.	✓	0.27
The weighted average metric calculates scores for each class and averages them based on the number of samples in each cl	✓	0.18

References

- http://arxiv.org/abs/2102.12407v1
- http://arxiv.org/abs/2109.07622v1
- http://arxiv.org/abs/1905.03197v3