Paper ID: 2512057

Paper Title: BiLSTM-based Language Identification in Code-Switched Social Media Text

Summary

This paper investigates word-level language identification in code-switched social media text for English–Hindi and English–Spanish language pairs. A hierarchical BiLSTM architecture is proposed, integrating character-level representations with word-level embeddings to handle noisy spelling, transliteration, and contextual ambiguity. Experimental results on benchmark Twitter datasets demonstrate that the proposed model significantly outperforms SVM and CRF baselines, achieving an accuracy of 96.8% and strong per-class F1 scores.

Similarity Score Analysis

The Turnitin plagiarism report indicates an overall similarity score of 5%. The matched content primarily consists of standard background descriptions and commonly used terminology in code-switching and language identification research. No missing citations, quotation issues, or integrity flags were reported.

Strengths

• Addresses a practically important NLP task in multilingual social media
• Effective combination of character-level and word-level modeling
• Strong experimental results with clear improvements over baselines
• Use of benchmark datasets supports reproducibility
• Detailed error analysis highlighting linguistic challenges

Weaknesses (Minor)

• Architecture represents an incremental extension of standard BiLSTM models
• Evaluation limited to two language pairs
• No comparison with recent transformer-based approaches
• Unresolved figure reference in the methodology section
• Computational efficiency and latency are not analyzed
• Dependence on large labeled datasets limits low-resource applicability

Required Minor Changes

• Clarify figure references and resolve missing placeholders
• Add discussion comparing results with transformer-based LID models
• Briefly analyze computational complexity and inference latency
• Discuss limitations related to data requirements and low-resource settings
• Improve consistency of formatting and presentation

Final Recommendation

ACCEPT WITH MINOR REVISIONS