IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 10, May 2025
Impact Factor: 7.67

Enhancing Marathi POS Tagging Accuracy Using Model Comparison and Hybrid Decision Strategy

Prof. P. S. Sajjanshetti[1], Pradnya Belhe[2], Amruta Barge[3] ,Yash Bari[4], Dnynesh Barnwal[5]
Asst.. Professor, Department of Computer Engineering[1]
Students, Department of Computer Engineering[2-5]
NBN Sinhgad Technical Institute Campus, Pune, India

Abstract: *Part-of-Speech (POS) tagging is an essential component in the field of Natural Language Processing (NLP), particularly for languages with rich morphology such as Marathi. Due to the language's intricate word structures and syntactic variability, achieving high accuracy in POS tagging presents considerable challenges. This study introduces a dual-model framework that integrates Hidden Markov Models (HMM) and Conditional Random Fields (CRF) to improve tagging performance. HMM captures sequential word patterns using probabilistic methods, while CRF leverages contextual relationships through discriminative learning. A dataset of 20,000 manually tagged Marathi words was used to train both models. Their results were analyzed at the sentence level using a hybrid conflict resolution technique based on rule-based inference and confidence metrics. Findings reveal that this integrative method surpasses the accuracy and reliability of standalone models, demonstrating the potential of model fusion to effectively manage the linguistic intricacies of Marathi. This work contributes to advancing NLP applications for under-resourced Indian languages.*

Keywords: Marathi POS Tagging, Hidden Markov Model, Conditional Random Fields, Hybrid Decision Strategy, Natural Language Processing, Low-Resource Languages, Sequence Labelling, Model Comparison, NLP in Indian Languages