NLP Assignment 1

Enhancing Figurative Language Recognition Using POS Tagging

Name: Tohuto Sema
Roll No: 21CS10072

Introduction

This report presents the solutions implemented for NLP Assignment 1. The assignment involved building models to recognize figurative language in text using POS tagging and comparing performance across multiple approaches.

Task 1 — HMM POS Tagger

Methodology: Implemented a Hidden Markov Model (HMM) from scratch for Part-of-Speech tagging, with Viterbi decoding for sequence labeling. Trained and tested on the NLTK Penn Treebank dataset.

   Results: Accuracy achieved: `XX%` (replace with your result).

Task 2 — Baseline Figurative Language Classifier

Methodology: Built a baseline classifier using simple features (bag-of-words) without POS tags. Used Logistic Regression for classification.

   Results: Accuracy: `XX%`. Classification Report and Confusion Matrix are included in the notebook.

Task 3 — POS Feature-based Classifier

Methodology: Enhanced the baseline classifier by incorporating POS tag-based features. POS tags helped in capturing syntactic cues of figurative language.

   Results: Accuracy improved to: `XX%`. Classification Report and Confusion Matrix included in notebook.

Task 4 — Pre-trained Model (FLUTE / HuggingFace)

Methodology: Used a pre-trained transformer-based model from HuggingFace (FLUTE/DistilBERT). Fine-tuned the model on figurative language dataset for classification.

   Results: Accuracy: `XX%` (highest among all models).

Comparison of Models

- HMM POS Tagger — Accuracy: XX%

- Baseline Classifier — Accuracy: XX%

- POS Feature-based Classifier — Accuracy: XX%

- Pre-trained Transformer (FLUTE/HuggingFace) — Accuracy: XX%

Challenges Faced

- Implementing HMM with Viterbi decoding from scratch was error-prone.

- Figurative language is inherently ambiguous, making classification harder.

- Computational limits while fine-tuning pre-trained models.

Conclusion

In this assignment, multiple models were explored for figurative language recognition. Incorporating POS tagging improved classification performance over the baseline. The best results were achieved with pre-trained HuggingFace transformers, demonstrating the strength of contextual embeddings in figurative language tasks.