Natural Language Understanding (CSL7640)
Assignment 1

Submitted By:

Pooja Naveen Poonia (M24CSA020)

Pranjal Malik (M24CSA021)

Shivani Tiwari (M24CSA029)

Suvigya Sharma (M24CSA033)

Submitted To:

Dr. Asif Ekbal

POS TAGGING USING HMM WITH VITERBI ALGORITHM

Colab Link

1. **Introduction** The goal was to develop a Part-of-Speech (POS) Tagger using different configurations of Hidden Markov Models (HMMs). We trained and evaluated two configurations:

   (a) 36-tag Model. Uses the full Penn Treebank POS tag set
   (b) 4-tag Model. Collapses POS tags into four categories:
      - N (Nouns): NN, NNS, NNP, NNPS
      - V (Verbs): VB, VBD, VBG, VBN, VBP, VBZ
      - A (Adjectives & Adverbs): JJ, JJR, JJS, RB, RBR, RBS
      - O (Other): All other tags (e.g., IN, PRP, CC)

2. **HMM Configurations Evaluated** We trained and evaluated three different HMM configurations for both 36/4 tag sets:

   (a) First-Order HMM (Order = 1). Assumes that the probability of a tag depends only on the immediate previous tag.
   (b) Second-Order HMM (Order = 2). Assumes that the probability of a tag depends on the two preceding tags.
   (c) First-Order HMM with Previous Word Context. Assumes that the probability of a word depends on the current tag as well as the previous word.

3. **Results Summary**

Configuration	36-Tag Accuracy	4-Tag Accuracy
First-Order HMM	87.67%	89.43%
Second-Order HMM	77.28%	83.83%
First-Order HMM with Previous Word	84.41%	89.77%

4. Detailed Analysis

(a) First-Order HMM

36-Tag Accuracy: 87.67%

4-Tag Accuracy: 89.43%

Observations:

- This configuration achieved the highest accuracy for the 36-tag model.
- The 4-tag model outperformed the 36-tag model, likely due to the simplification of categories, reducing the likelihood of misclassification.

(b) Second-Order HMM

36-Tag Accuracy: 77.28%

4-Tag Accuracy: 83.83%

Observations:

- The accuracy dropped significantly compared to the First-Order HMM.
- The additional dependency on the second previous tag increased sparsity, making probability estimates less reliable.
- The 4-tag model again performed better, showing more stability in the simplified version.

(c) First-Order HMM with Previous Word Context

36-Tag Accuracy: 84.41%

4-Tag Accuracy: 89.77%

Observations:

- The 4-tag model achieved the highest accuracy across all configurations.
- The 36-tag model performed slightly worse than the standard First-Order HMM.
- Incorporating the previous word helped improve accuracy in some cases, but it added complexity that made the model less stable for finer-grained POS tags.

5. **Comparison of 36-Tag and 4-Tag Models**

Model	Best Accuracy	Best Configuration
36-Tag Model	87.67%	First-Order HMM (Order = 1)
4-Tag Model	89.77%	First-Order HMM with Previous Word

- The 4-tag model consistently outperformed the 36-tag model across all configurations.
- The 36-tag model suffered from misclassification errors among similar tags, whereas the 4-tag model benefited from reduced complexity.
- If the goal is generalized POS tagging, the 4-tag model is simpler and more robust. However, for detailed linguistic analysis, the 36-tag model is necessary.

6. **Tag-wise Accuracy Analysis**

(a) <u>36-Tag Model Observations</u>

- High Accuracy Tags:
    - TO (100%), WRB (100%), CC (99.07%), DT (99.30%) – These are function words that are easier to classify.
    - Punctuation symbols (`-LRB-`, `-RRB-`, `#`) were tagged with 100% accuracy.
- Low Accuracy Tags:
    - UH (0%), LS (0%) – These are very rare in the dataset.
    - NNPS (50%), RBR (35.7%), PDT (50%) – More difficult to classify due to similar context usage.

(b) <u>4-Tag Model Observations</u>

- High Accuracy Tags:
    - O (96.99%) – Function words and prepositions were easy to classify.
    - N (90.07%) – Nouns were well recognized.

- Low Accuracy Tags:
  - A (76.38%) – Adjectives and adverbs had the lowest accuracy, as they were collapsed into a single category, causing confusion.

7. Conclusion

- First-Order HMM performed best for the 36-tag model.
- Second-Order HMM performed the worst due to increased data sparsity.
- Using the previous word slightly improved accuracy for the 4-tag model but was not beneficial for the 36-tag model.
- Collapsing tags into four broad categories helped improve accuracy across all configurations.
- The 4-tag model is more effective for general POS tagging, while the 36-tag model provides more linguistic granularity but at the cost of accuracy.

8. Recommendations

- Use Smoothing Techniques: Laplace or Good-Turing smoothing could improve transition probability estimates, especially for the second-order model.
- Try Neural POS Tagging: A BiLSTM-based tagger could be tested for comparison.
- Experiment with More POS Tag Collapsing Strategies: Exploring different collapses (e.g., function vs. content words) could provide further insights.