Sequence Labeling, POS Tagging

Dr. Sambit Praharaj
Assistant Professor (II)

Sequence Labeling – Introduction

- Sequence labeling assigns a label to each element in a sequence
- In NLP: sequence = words, labels = linguistic tags
- Context and order of words are crucial

Why Sequence Labeling Matters

- Words are not independent
- Label of a word depends on neighboring words
- Captures structure in natural language

Example

- Sentence: 'Ratan Tata founded Tata Group'
- Each word gets a label
- Labels depend on surrounding words

Sequence Labeling vs Classification

- Classification predicts one label for whole input
- Sequence labeling predicts label per token
- Sequence labeling uses contextual dependency

Common Sequence Labeling Tasks

- POS Tagging – grammatical roles
- Named Entity Recognition – real-world entities
- Chunking – phrase detection
- Semantic Role Labeling – predicate-argument structure

BIO Tagging Scheme

- B-X: Beginning of entity X
- I-X: Inside entity X
- O: Outside any entity
- Defines clear entity boundaries

POS Tagging as Sequence Labeling

- Assigns POS tag to each word
- First and simplest sequence labeling task
- Foundation for higher-level tasks

How POS Tagging Helps Sequence Labeling

- Provides syntactic structure
- Reduces ambiguity for later tasks
- Constrains possible labels

POS Tags Helping NER – Example

- Proper nouns (NNP) indicate entity candidates
- Prepositions signal entity boundaries
- Verb tags separate entities

HMM (Hidden Markov Model) for Sequence Labeling

- Hidden states = labels
- Observations = words
- Uses transition and emission probabilities
- Goal: maximize P(Tags | Words)

Numerical HMM Example – Sentence

- Sentence: 'Dogs bark'
- Goal: find best POS tag sequence
- Possible tags: Dogs → NNS, bark → NN / VBP

Step 1: Transition Probabilities

- P(NNS | START) = 0.6
- P(VBP | NNS) = 0.75
- P(NN | NNS) = 0.25

Step 2: Emission Probabilities

- P(Dogs | NNS) = 0.5
- P(bark | VBP) = 0.33
- P(bark | NN) = 0.10

Step 3: Compute Path 1

- Path: START → NNS → VBP
- $0.6 \times 0.5 \times 0.75 \times 0.33$
- Total ≈ 0.074

Step 4: Compute Path 2

- Path: START → NNS → NN
- $0.6 \times 0.5 \times 0.25 \times 0.10$
- Total ≈ 0.0075

Step 5: HMM Decision

- Compare path probabilities
- Path 1 > Path 2
- Final tags: Dogs/NNS bark/VBP

Viterbi Interpretation

- Evaluates all tag paths efficiently
- Keeps only highest probability paths
- Backtracks to get best sequence

Why HMM Works (and Fails)

- Works well with limited ambiguity
- Fails with long-range dependencies
- Motivation for Neural models

HMM Summary

- Transition = P(tag_i | tag_{i−1})
- Emission = P(word_i | tag_i)
- HMM selects tag sequence with maximum probability

NER as Sequence Labeling

- NER assigns entity labels to each word
- Uses BIO scheme
- More complex than POS tagging

Challenges in Sequence Labeling

- Ambiguity of words
- Long-range dependencies
- Maintaining label consistency
- Data sparsity

Neural Sequence Labeling Models

- BiLSTM models left and right context
- Transformers capture long-range dependencies
- Often combined with CRF

End-to-End NLP Pipeline

- Tokenization
- POS Tagging
- Sequence Labeling (NER, Chunking, SRL)
- Information Extraction / NLP Applications

Key Takeaways

- Sequence labeling assigns label to each token
- Context and order are essential
- POS tagging is the gateway sequence labeling task