CS-626 Assignment 1

Pos-Tagging

Shubham Nemani, 203050011,
Anish M M, 203050066,
Pooja Verma, 203050072

December 1, 2020

1 Hidden Markov model (HMM)

1.1 Average Accuracies

Accuracy values were averaged over 5 folds.

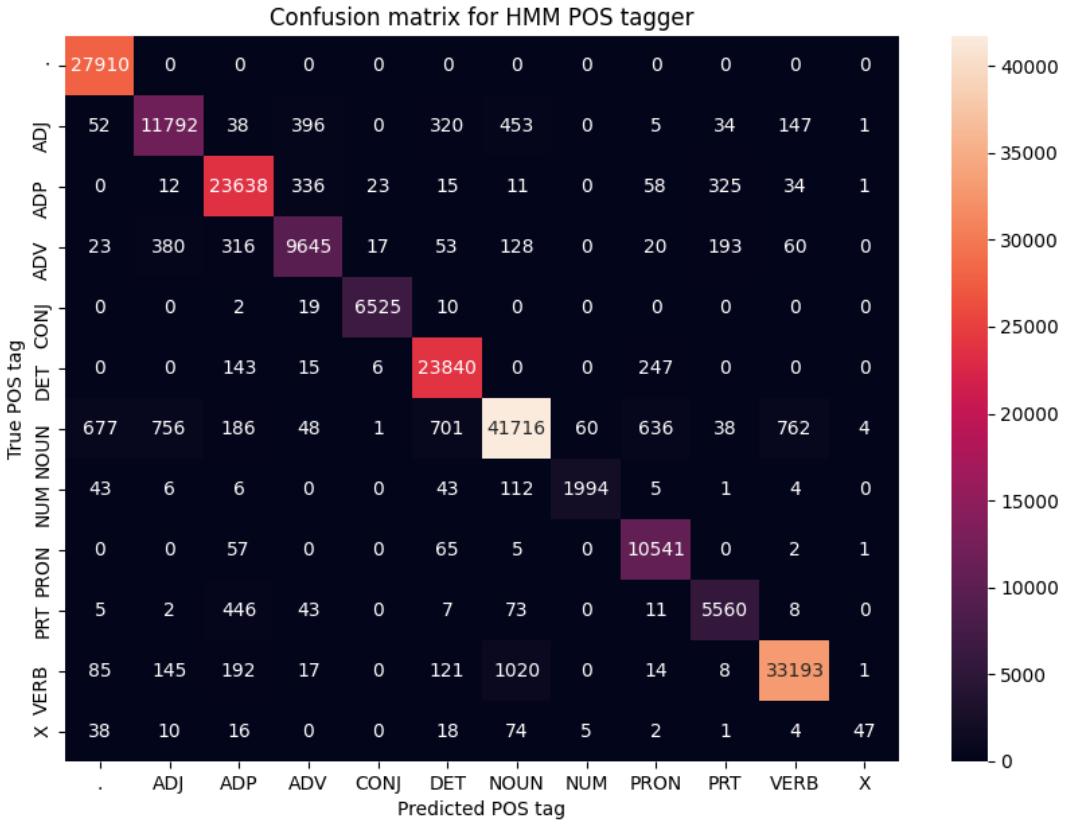
	Average accuracy (%)
Overall	94.63
.	99.53
ADJ	98.52
ADP	98.84
ADV	98.99
CONJ	99.96
DET	99.12
NOUN	96.94
NUM	99.83
PRON	99.51
PRT	99.46
VERB	98.65
X	99.90

Table 1: Tag wise metrics

1.2 Error Analysis

(The following discussion reports values for testing on the 4th fold after training on the remaining folds in the 5-fold split)

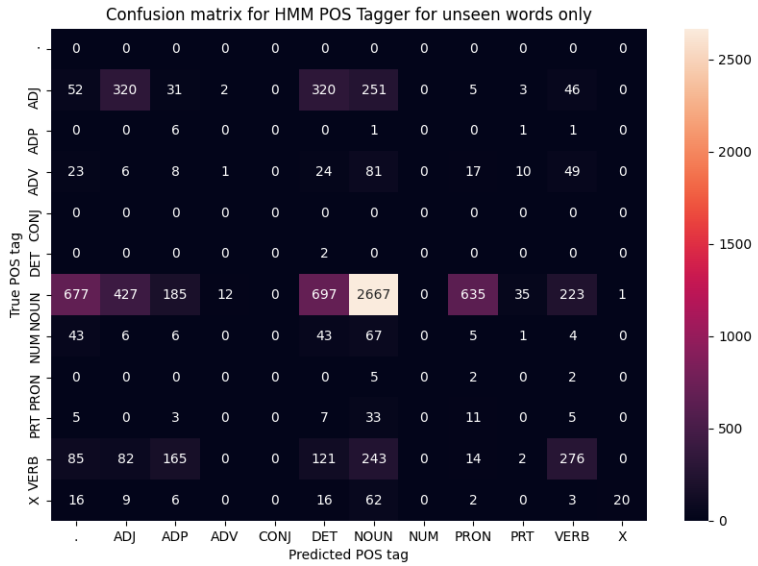
- **Tag X:** Foreign words which should be tagged as X get misclassified quite often (78% of the time). This is because $P(X \mid T_1)$ and $P(T_2 \mid X)$ are lesser than $P(T \mid T_1)$ and $P(T_2 \mid T)$ respectively for most tags T, T_1, and T_2. When the word is unseen, owing to our strategy of same default emission probability from each tag, the above transition probabilities will decide the prediction choice. In addition, in the



train set, many words which appear in X also appear in other classes (e.g. Service, Theatre, say, may, etc). In these cases also, the above transition probabilities will have a major say in the prediction decision.

- **Tags misclassified as . :** NOUN was misclassified as "." 677 times in total. 100% of these were unseen words. Similarly, unseen words also constitute 100% of VERB, PRT, NUM, ADV and ADJ misclassified as "." .

- **Unseen words:** Similarly, Unseen words constitute 56.5% of NOUN misclassified as ADJ, 99.5% of NOUN misclassified as ADP, 25% of NOUN misclassified as ADV, 99.4% of NOUN misclassified as DET, 99.84% of NOUN misclassified as PRON, 92.11% of NOUN misclassified as PRT, 29.3% of NOUN misclassified as VERB, 25% of NOUN misclassified as X. *This indicates that using a default probability is not a very efficient way of handling unseen words. Better smoothing mechanisms may help here.*



- **VERB misclassified as NOUN:** When VERBs occur as last word in sentence, they tend to get misclassified as NOUNs. This is especially so in the case of words which can serve both as NOUN and VERB. This can be explained by the fact that transition probability of verb to · is much lower than transition probability from noun to · (i.e., sentences end in NOUN much more frequently than in VERB), but for these words, emission probability will be relatively high for both noun and verb classes. These errors constitute 10.6% of all VERB misclassified as NOUN.

  $P(\cdot \mid VERB) = 0.07841466301443982$

  $P(\cdot \mid NOUN) = 0.283017056685444$

  Example sentence 1:

  ... I will do my best to SUPPLY . (VERB suppy tagged as NOUN)

  $P(suppy \mid VERB) = 0.00022304229693012693$

  $P(supply \mid NOUN) = 0.00022176516373661255$

  Example sentence 2:

  ... that can accelerate growth in human UNDERSTANDING . (VERB understanding tagged as NOUN)

  $P(understanding \mid VERB) = 0.00018924800951647134$

  $P(understanding \mid NOUN) = 0.0003348219138768464$

  Along with unseen words, we have accounted for about 34.42% of errors in this category.

- **NOUN misclassified as VERB:** This is the second largest error category (in number). About 20% of this error category consists of possible gerunds (words ending with *-ing*). This happens since many such words appear both as the gerund form of verbs and as nouns in the training corpus. Along with unseen words, we have thus accounted for about 50% of this error category.

- **NOUN misclassified as ADJ:** One component here is compound nouns formed by two NOUNs getting tagged as ADJ NOUN. One reason for this could be the higher transition prob from ADJ to NOUN than from NOUN to NOUN ($P(NOUN \mid ADJ) = 0.66$, $P(NOUN \mid NOUN) = 0.15$). 70% of errors in this category have the sequence NOUN NOUN misclassified as ADJ NOUN. Examples include "Kimbell-Diamond doctrine", "trasferor corporation", "Aerospace Industries", "serial number", "TR list", etc.

- The conj tag has highest accuracy with 99.96% because
  - the class is closed with just 26 members.
  - most member words have fixed tags
  - the number of occurences of this class is low and the negative class influences per-tag accuracy more.
  - no unseen words

- Per-tag accuracy is not a very useful measure in this case because the negative class is always much bigger than the positive class and therefore, the accuracy value is always high. Noun has comparatively lower accuracy since noun is the biggest class in corpus and so the influence of the negative class is lower.

2 Bi-directional Long Short-Term Memory (Bi-LSTM)

2.1 Average Accuracies

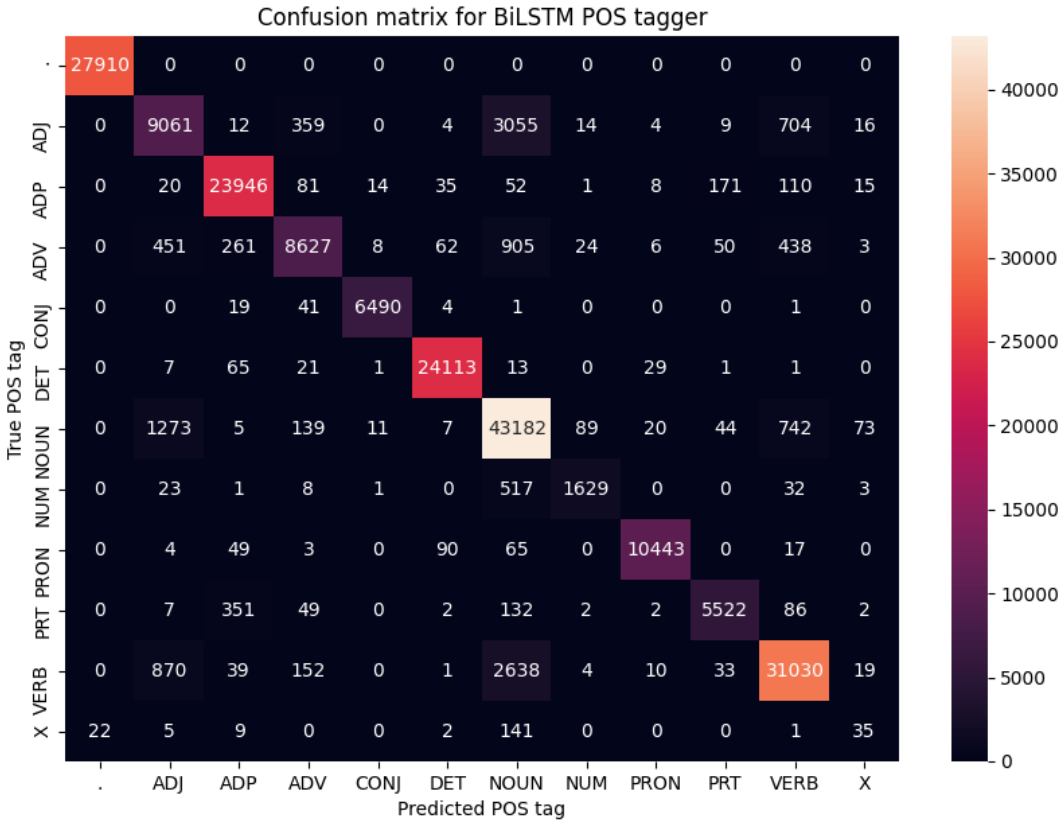
Accuracy values were averaged over 5 folds.

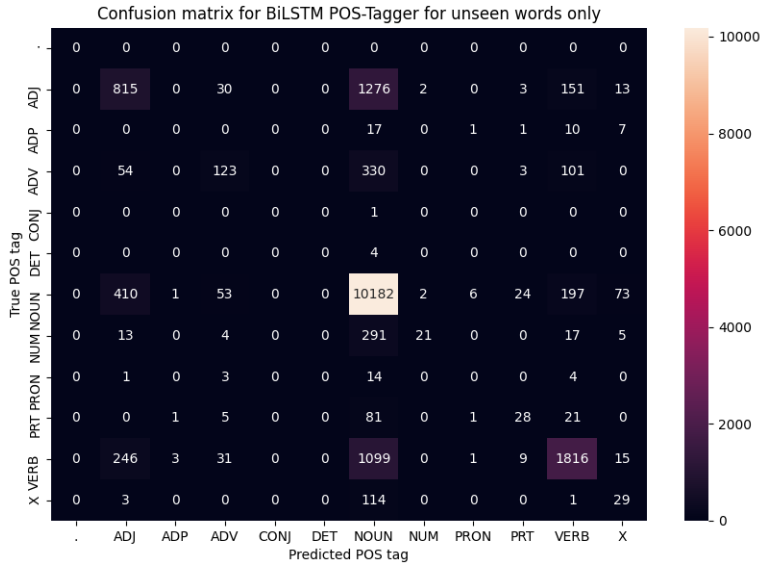
	Average accuracy (%)
Overall	92.29
.	99.996
ADJ	96.29
ADP	99.37
ADV	98.48
CONJ	99.94
DET	99.83
NOUN	94.81
NUM	99.55
PRON	99.84
PRT	99.58
VERB	97.07
X	99.83

Table 2: Tag wise metrics

2.2 Error Analysis

- VERB got misclassified as NOUN 2638 times. But, unlike HMM, only 8.76% of these had VERB as the last word. This illustrates that error region of BiLSTM is different from that of HMM.
- NOUN got misclassified as VERB 742 times. Possible gerunds (words ending with *-ing*) constitute only 2.6% of this error category (as opposed to about 20% for HMM). This again illustrates that the error regions are different than that of HMM.
- **NOUN misclassified as ADJ:** 71.88% of this error category constitutes of bi-grams with true labels (NOUN NOUN) getting misclassified as (ADJ NOUN). HMM also had 70% of such cases in this category. This tells us that the network could possibly have encoded the information that *"the word preceding a NOUN is more likely to be an ADJ than a NOUN."*





- Unseen words constitute 41.66% of VERBs misclassified as NOUNs, 41.77% of ADJs misclassified as NOUNs, 32.22% of NOUNs misclassified as ADJs, 26.55% of NOUNs misclassified as VERBs, 8.36% of ADJs misclassified as ADVs, etc. Similar values can be computed for other class pairs using the above two confusion matrices.

- In hmm we saw that a lot of unseen words were misclassified as ·, the class of punctuations. This is a rather stupid mistake since the class · doesn't contain any real words. However, we see that BiLSTM doesn't misclassify any unseen word as ·, although it has some trouble with the class X being misclassified as ·. This is a much more acceptable behaviour.

3 SVM

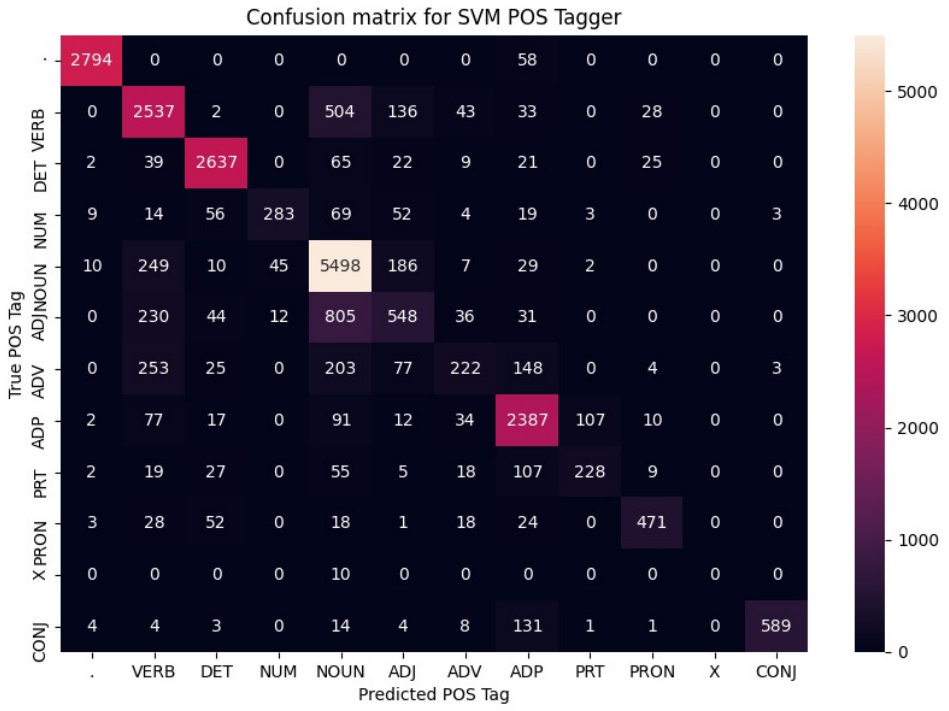
3.1 Accuracies

	Average accuracy (%)
Overall	80.03
·	99.60
ADJ	92.73
ADP	95.82
ADV	96.09
CONJ	99.23
DET	98.16
NOUN	89.57
NUM	98.74
PRON	99.03
PRT	98.44
VERB	97.70
X	99.96

Table 3: Tag wise metrics

3.2 Error Analysis

- We can observe in the confusion matrix that quite a number of verbs are tagged as nouns and verbs are labelled as nouns due to dual nature of some words which can act as both verbs and nouns.
- 'X' couldn't be classified as 'X' even once.But 'X' was predicted as 'NOUN' mostly.



- Prefix and suffix play an important role as features, hugely impacting the accuracy of the model.
- Exceptions in using suffixes words with -ing suffix are classified as VERB but words like - 'meeting' should be NOUN but misclassified.
- Accuracy of PRT = 0.48
  This is mainly because of contracted forms like : you're , we're , you'll , misclassified as NOUN , VERB.