School of Computer Science Engineering and Technology

Course- B. Tech
Course Code- CSET346

Year- 2022
Date: 01-11-2022

Type- Elective
Course Name: Natural language processing
Semester- odd
Batch- ALL

Lab Assignment 09 – Language Model

CO-Mapping

Exp. No.	Name	CO1	CO2	CO3
09	Language Model	✓	✓	✓

Objective: The main motivation to make student be capable of implementing language model using the concept of NLP.

Introduction:
Language Modeling (LM) is one of the most important parts of modern Natural Language Processing (NLP). There are many sorts of applications for Language Modeling, like: Machine Translation, Spell Correction Speech Recognition, Summarization, Question Answering, Sentiment analysis etc.

The probabilistic language modelling
The probability of a sentence S (as a sequence of words wi) is : P(S)=P(w1,w2,w3,…,wn). Now it is important to find the probability of upcoming word. It is an everyday task made e.g. while typing your mobile keyboard. We will settle the conditional probability of w4 depending on all previous words.
For a 4 word sentence this conditional probability is:
P(S)=P(w4|w1,w2,w3)

There are two important techniques:

- Statistical machine translation
- Neural machine translation

Today's lab is dedicated for creating language model for your mother tongue (selected Indian language). In previous labs, you already work on the corpus for your selected Indian languages.
A sample code is available for LM for English.
Go through the following link and create similar LM for your selected language.

https://github.com/IBM/deep-learning-language-model