

Submission Assignment #1

Instructor: Burcu Can*Name:* Metin DEMİR, *Netid:* 21526902

1 Introduction

Language Modeling is one of the fundamental concepts of Natural Language Processing (NLP). In this assignment, I will build some of the basic language models and generate sentences with these models. The models concept will unigram, bigram and trigram. As training set I will use Facebook children stories dataset.

2 Approach

First all things, I have to read my dataset then put them into an array. After that create a dictionary for Unigram, key is the word and value of this key is probability of it. For generate sentences I generated a random float between 0 and 1 then get word which close probability of this value. After all return this sentences. For Bigram and Trigram I created a class which store first words, second word, number of match, probability and smooth probability. We calculate all this feature I updated after all these. My all model is ready.

3 Result