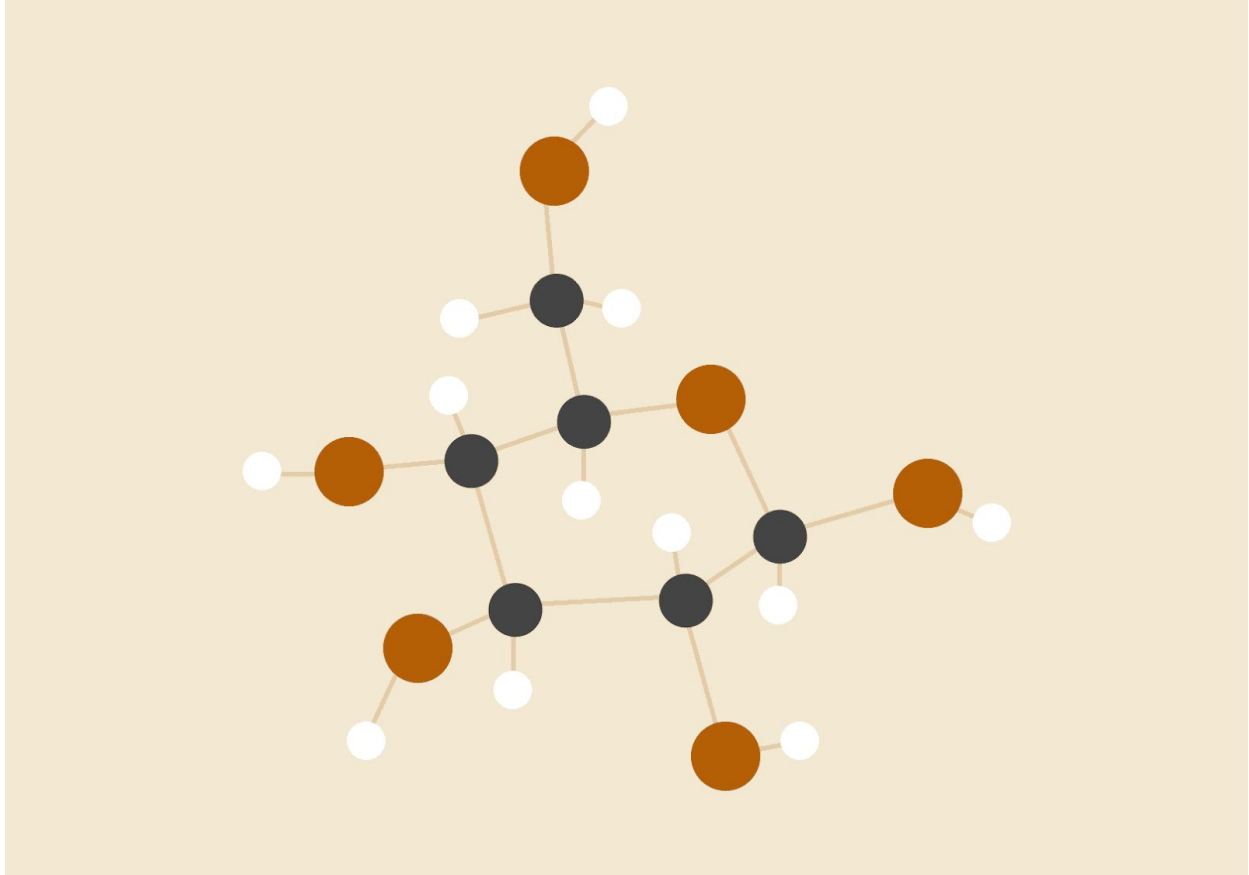


ASSIGNMENT 3

IBM MODEL



Vijitha Gunta 2015B3A70491H

Shivani Reddy 2015B3A70531H

Shubham Bharadwaj 2016A7PS0120H

INTRODUCTION

In this assignment we have implemented IBM Models 1, 2, EM algorithm and extracted phrases from the required dataset to calculate the respective phrase scores and rank them in the order of descending probability. These modules are typically used in Machine Translation.

IBM MODEL-1

IBM Model 1 is weak in terms of conducting reordering or adding and dropping words. In most cases, words that follow each other in one language would have a different order after translation, but IBM Model 1 treats all kinds of reordering as equally possible.

EM Algo

EM algo solves the maximization problem of determining the translation probs and the alignment probs which are dependent on each other but neither of which we have to begin with. Hence we start with assuming equal prob of translating a foreign word to an english word and then calc alignment probs based on that and then use those alignment probs to then recompute translation probs. This is thus an iterative process.

Packages Used

1. NLTK
2. json
3. itertools
4. numpy
5. copy

RESULTS

- Our alignment probs were similar to that obtained by NLTK inbuilt function. (Hence in the 3rd q we used the inbuilt func directly.)
- We also tested the model on our own created German dataset and got ideal results.