Tutorial 05   N-grams and Bayes' Rule

Introduction

In this tutorial, you will gain better understanding about N-grams and Bayes' Rule and how they can be used for action Prediction. Refer to Lecture 03 handout for completing this tutorial.

In this tutorial, suppose there is a game that includes a speech recognizer for these action words: "Go", "Stop", "Jump", and "Turn". Denote these actions by their initial characters, i.e., G (Go); S (Stop); J (Jump); T (turn). During a gameplay, a player's actions have been recorded as a sequence of characters stored in the file `string.txt` (downloaded from Canvas). This text file contains a string of 200 characters formed by G, S, J or T representing the corresponding words Go, Stop, Jump and Turn recognized by the speech recognizer.

You should write a program to analyze this string in order to work out the following tasks.

Task 1.   N-grams

a) Write a program to compute the frequencies of the N-grams statistics when N=2 (also known as bi-grams). What would be the next word predicted using the bi-grams?
b) Write a program to compute the frequencies of the N-grams statistics when N=3 (also known as tri-grams). What would be the next word predicted using the tri-grams?

Note: For simplicity, you can just copy the string from the text file and paste it in your code for defining it as a string.

Task 2.   Prior

Write a program to estimate the prior of each word, i.e., P[G], P[S], P[J], P[T] from the observed relative frequency. What would be the next word predicted using only the prior?

Task 3.   Bayes' Rule

Suppose you also timed how long it took the player to say each word and you discovered that the time taken was a good indicator of which word was spoken. Here are the conditional probabilities, given the word the player spoke, that it took them more than half a second:

Word	P(duration>0.5s | word)
Go	0.1
Stop	0.05
Jump	0.3
Turn	0.4

With the prior determined in Task 2, using Bayes' Rule, what is the most likely word the player spoke, given that the word took 0.9 seconds to say it (thus duration>0.5s)?

Task 4.   Complete the Canvas Quiz

Complete the quiz "Tutorial 05" on the Canvas course page (Assignments > Tutorial 05) before the posted deadline.