*SINDH MADRESSATUL ISLAM UNIVERSITY*

ASSIGNEMNT 01

Ahtesham ul haq CSC-18F-126

Ahsan Younus CSC-18F-130

Asharib ahmad CSC-18F-096

CS-8B

Natural Language Processing

Submitted To Syed M Hassan Zaidi

15th March 2022

***Select 100 sentence related to paitients i.e. paitients basic needs, you can take data from doctor and patients conversation too or pick from online. Compute unigram biagram and identity probability using python. Write a note that how this model can be work with machine learning or how we can train this using machine/ deep learning. Write some small introduction about nlp, basic need, problem statement, results required.***

all\_words="""Doctor: Hello! What can I do for you?

Patient: Good Morning Doctor. I don’t feel good.

Doctor: Come and sit here.

Doctor: Open your mouth.

Doctor: Since how long are you not feeling well?

Patient: Since yesterday.

Doctor: No problem. Did you have motions yesterday?

Patient: No Doctor. Not so freely.

Patient: Doctor I feel weak and do not feel like eating.

Doctor: Ok. And what else?

Patient: I feel like vomiting.

Doctor: Do you drink a lot of water?

Patient: No Doctor, I don’t have water too much.

Doctor: Did you took any medicine?

Patient: Yes Doctor, I took a Crocin.

Doctor: who asked you to take it?

Patient: No one Doctor. I took it myself.

Doctor: why did you take it?

Patient: Because I felt a headache.

Doctor: Nothing to be worried at.

Doctor: Do you need quick relief?

Patient: No Doctor. It is enough you give me medicines for now.

Conversation Between Doctor and Patient About Fever

Patient: Good Morning doctor.

Doctor: Good morning! You seem pale and your voice sounds different.

Patient: Yes doctor. I’m having a temperature and even a sore throat.

Doctor: Let me check.

Doctor: You have a moderate fever. Let me check your temperature.

(Doctor uses a thermometer to check the temperature of a patient by snapping around one of the fingers)

Patient: This thermometer is different from the last one you used.

Doctor: Yes, this is introduced newly by medical supplies companies. It’s much more comfortable and doesn’t need cleaning after every use.

Patient: That’s really great.

Doctor: Yes it is.

(Now he removed the thermometer)

Doctor: Temperature is not too high, around 99.8. Let me check your blood pressure as well.

(He then continues with measuring blood pressure.)

Doctor: Your blood pressure is fine.

(Now, he checks the throat.)

Doctor: It seems a bit scruffy. Not good.

Patient: Yes, it has been quite bad.

Doctor: Do you sweat and shiver?

Patient: Not sweating, but I feel slightly cold when I sit under a fan.

Doctor: Ok. You have few signs of malaria. I would recommend you undergo a blood test. Nothing to worry about. In most cases, the test comes out to be negative. It’s just careful, as there has been a spurt in malaria cases in the last month or so.

Doctor: I am writing three medicines and syrup. The number of dots in front of each medicine name tells you how many times in the day you have to take them. For example, the two dots here mean you’ve to take the medicine twice in the day, once in the morning and once after dinner.

Doctor: Do you have any other questions?

Patient: No, doctor. Thank you.

Doctor: So, what makes you come here?

Attendant: Doctor, my father has been diagnosed with unusual growth of cells in the colon, which other doctors so far have confirmed as a state of cancer.

Doctor: What tests have you conducted?

Attendant: On the recommendation of doctors in Delhi, we did blood tests, CT scan, and biopsy. We have taken three views on the reports and all have come up with opined colorectal cancer. Here are the reports.

Doctor: Yes, this seems cancerous. We need to admit him today itself and get a few tests done. And let’s meet tomorrow morning when you have the reports of these tests.

Attendant: OK. Thank you.

(The doctor formulates his observations on his letterhead and gives it over to the Patient. The person then admits the patient and visits the doctor again the next morning with new reports in hand.)

Doctor: The cancer is just one stage before it will expand to other parts. In this case, the best course of action is quick surgery to extract the affected part accompanied by radiation therapy.

Attendant: How many days of treatment will be needed?

Doctor: We can perform the operation in a couple of days. Consequently, we will keep the patient under observation for 4-5 days. Radiation therapy has become pretty regulated and therefore you can get it done in any tier-2 city, which will not only be useful to you but also decrease your cost.

Attendant: And for how long the patient will have to go through radiation therapy?

Doctor: The patient has to take one dose every two weeks for three months. You require to visit the hospital only on days you obtain therapy. Once radiation therapy is done, you should discuss an oncologist every three months in the beginning and every year later on to check for the remission of cancerous cells.

Attendant: OK. I’ve heard radiation therapy has side effects.

Doctor: Yes, radiation therapy has side effects. It can lead to hair loss, nausea, and loss of appetite.

Attendant: Thank you, doctor. We will get him admitted today.

Conversation About Stomach Pain With Doctor

Doctor: Yes, tell me. What problems you have?

Patient: I am suffering from stomachache and motions since last night. I have also vomited a few times last night.

Doctor: What did you have yesterday?

Patient: I had some snacks on the roadside stalls. It could be because of it.

Doctor: It is possible that you had contaminated food. Because of diarrhea, you have lost plenty of body fluids. You require to be hydrated. Drink enough water regularly, at least 10-12 glasses. Mix some Glucon-D powder or Electoral in water and have it. Fruit juice is also fine. Avoid caffeine, dairy products, and solid foods at least till evening. And get plenty of rest.

Patient: Any medicines, doctor.

Doctor: Yes, I am prescribing a few to control diarrhea.

Patient: Thank you, doctor.

"""

**Bigram:**

import nltk

from nltk.corpus import reuters

from nltk import bigrams, trigrams

from collections import Counter, defaultdict

# Create a placeholder for model

model = defaultdict(lambda: defaultdict(lambda: 0))

# sents(fileids="docName")

# Count frequency of co-occurance

for sentence in nltk.sent\_tokenize(all\_words.lower()):#nltk\_tokens = nltk.sent\_tokenize(sentence\_data)

sentence = nltk.word\_tokenize(sentence)

# print(sentence)

for w1, w2, w3 in trigrams(sentence, pad\_right=True, pad\_left=True):

model[(w1, w2)][w3] += 1

# Let's transform the counts to probabilities

for w1\_w2 in model:

total\_count = float(sum(model[w1\_w2].values()))

for w3 in model[w1\_w2]:

model[w1\_w2][w3] /= total\_count

**probability:**

**dict(model[("do","you")])**

**unigram:**

# Count frequency of co-occurance

for sentence in nltk.sent\_tokenize(all\_words.lower()):#nltk\_tokens = nltk.sent\_tokenize(sentence\_data)

sentence = nltk.word\_tokenize(sentence)

# print(sentence)

for w1, w2 in bigrams(sentence, pad\_right=True, pad\_left=True):

model[(w1)][w2] += 1

# Let's transform the counts to probabilities

for w1 in model:

total\_count = float(sum(model[w1].values()))

for w2 in model[w1]:

model[w1][w2] /= total\_count

**probability:**

dict(model[(":")])

This model can work using ML using NLTK utils tools.

In natural language processing n-gram is a contiguous sequence of n items generated from a given sample of text where the items can be characters or words and n can be any numbers like 1,2,3, etc.

1. To generate 1-grams we pass the value of n=1 in ngrams function of NLTK. But first, we split the sentence into tokens and then pass these tokens to ngrams function.
2. For generating 2-grams we pass the value of n=2 in ngrams function of NLTK. But first, we split the sentence into tokens and then pass these tokens to ngrams function.
3. In case of 3-grams, we pass the value of n=3 in ngrams function of NLTK. But first, we split the sentence into tokens and then pass these tokens to ngrams function.

***Use of n-grams in NLP:***

* N-Grams are useful to create features from text corpus for machine learning algorithms like SVM, Naive Bayes, etc.
* N-Grams are useful for creating capabilities like autocorrect, autocompletion of sentences, text summarization, speech recognition, etc.

**NLP:**

Natural Language Processing (NLP) is the sub-field of Computer Science especially Artificial Intelligence (AI) that is concerned about enabling computers to understand and process human language. Technically, the main task of NLP would be to program computers for analyzing and processing huge amount of natural language data.

There are two main phases to natural language processing: data preprocessing and algorithm development.

Data preprocessing involves preparing and "cleaning" text data for machines to be able to analyze it. preprocessing puts data in workable form and highlights features in the text that an algorithm can work with. There are several ways this can be done, including:

* [**Tokenization**](https://www.techtarget.com/searchsecurity/definition/tokenization)**.**
* [**Stop word**](https://whatis.techtarget.com/definition/stop-word)**removal**
* [**Lemmatization**](https://www.techtarget.com/searchenterpriseai/definition/lemmatization)**and stemming**
* **Part-of-speech tagging**

Once the data has been preprocessed, an algorithm is developed to process it. There are many different natural language processing algorithms, but two main types are commonly used:

* **Rules-based system**
* **Machine learning-based system**

**BASIC NEED:**

NLP is important because it helps resolve ambiguity in language and adds useful numeric structure to the data for many downstream applications, such as speech recognition or text analytics.

**PROBLEM STATEMENT:**

The main challenge of NLP is the understanding and modeling of elements within a variable context. In a natural language, words are unique but can have different meanings depending on the context resulting in ambiguity on the lexical, syntactic, and semantic levels.