

CPE 372/641 Natural Language Processing

Homework 3: Sentence Classification

1. Source code that includes

1.1. My expanded training and testing sentence

Traning sentence

```
training_corpus = [  
    ('I am exhausted of this work.', 'Class_B'),  
    ("I can't cooperate with this", 'Class_B'),  
    ('He is my badest enemy!', 'Class_B'),  
    ('My management is poor.', 'Class_B'),  
    ('I love this burger.', 'Class_A'),  
    ('This is an brilliant place!', 'Class_A'),  
    ('I feel very good about these dates.', 'Class_A'),  
    ('This is my best work.', 'Class_A'),  
    ("What an awesome view", 'Class_A'),  
    ('I do not like this dish', 'Class_B'),  
    #new_A  
    ('I love to go shopping', 'Class_A'),  
    ('I am rich and famous.', 'Class_A'),  
    ('I could not be better', 'Class_A'),  
    ('She is my happiness', 'Class_A'),  
    ('Her smile makes me feel good', 'Class_A'),  
    #new_B  
    ('I can not go anywhere.', 'Class_B'),  
    ('I just got chewed out by my boss!', 'Class_B'),  
    ('Covid-19 make me feel bad', 'Class_B'),  
    ('I am done with you', 'Class_B'),  
    ('I do not trust him', 'Class_B')  
]
```

Testing sentence

```
test_corpus = [
    ("I am not feeling well today.", 'Class_B'),
    ("I feel brilliant!", 'Class_A'),
    ('Gary is a friend of mine.', 'Class_A'),
    ("I can't believe I'm doing this.", 'Class_B'),
    ('The date was good.', 'Class_A'),
    ('I do not enjoy my job', 'Class_B'),
    #new
    ('Covid-19 make me sad', 'Class_B'),
    ('I do not love him', 'Class_B'),
    ('This dish was terrible', 'Class_B'),
    ('Shopping mall is open now', 'Class_A'),
    ('I am out of money!', 'Class_B')
]
```

1.2.The results of classifying your testing sentences (5 old and 5 newly added)

Old sentences

```
test_corpus = [
    ("I am not feeling well today.", 'Class_B'),
    ("I feel brilliant!", 'Class_A'),
    ('Gary is a friend of mine.', 'Class_A'),
    ("I can't believe I'm doing this.", 'Class_B'),
    ('The date was good.', 'Class_A'),
    ('I do not enjoy my job', 'Class_B')
]

model = NBC(training_corpus)
print(model.classify("Their codes are amazing."))
print(model.classify("I don't like their computer."))
print(model.accuracy(test_corpus))
```

```
Their codes are amazing. Class_A
I don't like their computer. Class_B
0.8333333333333334
```

New sentences

```
test_corpus = [
    ("I am not feeling well today.", 'Class_B'),
    ("I feel brilliant!", 'Class_A'),
    ('Gary is a friend of mine.', 'Class_A'),
    ("I can't believe I'm doing this.", 'Class_B'),
    ('The date was good.', 'Class_A'),
    ('I do not enjoy my job', 'Class_B'),
    #new
    ('Covid-19 make me sad', 'Class_B'),
    ('I do not love him', 'Class_B'),
    ('This dish was terrible' , 'Class_B'),
    ('Shopping mall is open now', 'Class_A'),
    ('I am out of money!', 'Class_B')
]
```

```
Their codes are amazing. Class_A
I don't like their computer. Class_B
0.9090909090909091
```

2. Short write-up explains model accuracy and discussion

การเพิ่ม training sentences ด้วยประโยค A 5 ประโยค และ B 5 ประโยคทำให้ซึ่งหลังจากการทำการโมเดล เมื่อนำเอาประโยคเก่า(old sentence) ใน test_sentences ไปทดสอบทำให้ผลลัพธ์เป็นอย่างเดิมหรือ 0.8333 แต่เมื่อทำการเพิ่มประโยคเข้าไปใน test_sentences เป็น Class_B 4 ประโยค และ Class_A 1 ประโยค เมื่อนำไปทดสอบหาความแม่นยำจะได้ 0.909 เนื่องจากความถี่ของคำที่นำไปใช้ เช่น I am out of money!. ที่อยู่ใน Class_B ซึ่งใน Class_B ของ training sentence มีความถี่ของคำว่า I am ค่อนข้างมาก ทำให้เกิดความแม่นยำที่เพิ่มขึ้น

ดังนั้นการเพิ่ม Training_sentences เพียงอย่างเดียวแต่ไม่เพิ่ม Test_sentences จะทำให้ผลของค่าความแม่นยำเท่าเดิมเนื่องจากไม่ส่งผลต่อการหายคำใน naive bayes classifier

เมื่อทำการเพิ่ม Test_sentences จะสามารถเพิ่มความแม่นยำได้เนื่องจากความถี่ทั้งหมดของคำและประโยคที่เกิดขึ้น ทำให้นำไปทำนายความแม่นยำใน Training_sentences

ดังนั้นการเพิ่มความแม่นยำของโมเดลจะสามารถทำได้จากการเพิ่ม Test_sentences และ Training_sentences ที่เหมาะสมคืออยู่ในลักษณะ(class) ที่ถูกต้อง