

CPE 372/641 Natural Language Processing

Homework 3: Sentence Classification

1. Source code that includes

- Expanded training and testing sentence

Training 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  
    ('I like this book', 'Class_A'),  
    ('I have a bad day', 'Class_B'),  
    ('Why bad things always happen to me.', 'Class_B'),  
    ('My pets are so lovely.', 'Class_A'),  
    ('Those neighbors are so annoying.', 'Class_B'),  
    ('What a nice day.', 'Class_A'),  
    ('I hate rain.', 'Class_B'),  
    ('My family will be proud of my good grade', 'Class_A'),  
    ('I do not like this place', 'Class_B'),  
    ('I can not wait some of great things that will happen in the next day', 'Class_A'),  
]
```

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
    ('We are best buddies', 'Class_A'),
    ('I think I got some bad lucks.', 'Class_B'),
    ('Such an awful day', 'Class_B'),
    ('Love this town.', 'Class_A'),
    ('I got bad a feeling about this', 'Class_B'),
]

```

- The results of classifying testing sentences

Old test sentence

```

-----
Test sentence
-----
('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')
-----
Classify
-----
Their codes are amazing. -> Class_B
I don't like their computer. -> Class_B
Model accuracy ->
0.8333333333333334

```

New test sentence

```

-----
Test sentence
-----
('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')
('We are best buddies.', 'Class_A')
('I think I got some bad lucks.', 'Class_B')
('Such an awful day.', 'Class_B')
('Love this town.', 'Class_A')
('I got bad a feeling about this.', 'Class_B')
-----
Classify
-----
Their codes are amazing. -> Class_B
I don't like their computer. -> Class_B
Model accuracy ->
0.72727272727273

```

2. Short write-up that explains model accuracy and discussion

ในส่วนของความแม่นยำของตัวโมเดลนั้น ก่อนที่จะทำการเพิ่ม Training sentence ตัวโมเดลมีความแม่นยำที่ 0.833 ตามโจทย์ และหลังจากที่ทำการเพิ่ม Training sentence ด้วยประโยคประเภท A 5 ประโยคและประโยคประเภท B 5 ประโยค แล้วทำการเทรนนิ่งด้วย Training sentence อันเก่าได้ผลความแม่นยำเท่าเดิม หรือ 0.83 แต่ตัวโมเดลทำนายประโยคได้ผลลัพธ์ใหม่ โดยตัวโมเดลทำนายประโยค “Their codes are amazing.” จากประเภท A กลายเป็นประเภท B แทน ซึ่งสาเหตุที่เป็นเช่นนั้น เพราะ จากแนวความคิดของ Naïve Bayes Classifier ซึ่งใช้หลักความถี่ของคำจากประเภทต่างๆ รวมกันทุกคำในประโยค โดยความน่าจะเป็นของประเภทใดมากกว่า แสดงว่าประโยคนั้นน่าจะอยู่ในประเภทดังกล่าว ซึ่งเมื่อเทียบกับผลลัพธ์ที่เกิดขึ้นนั้น คำที่น่าจะมีผลกับผลลัพธ์ที่ออกมา คือ คำว่า “are” เนื่องจากใน Training sentence มีประโยคที่ใช้คำว่า “are” อยู่ได้แก่ “My pets are so lovely.” จากประเภท A และ “Those neighbors are so annoying.” จากประเภท B แต่ตัวของประเภท

B ส่งผลให้ความน่าจะเป็นที่ประโยคที่ใช้ “are” เป็น Class B มากกว่า เพราะฉะนั้น โมเดลจึงทำนายให้ “Their codes are amazing ” อยู่ใน Class B

ความแม่นยำของตัวโมเดลก่อนที่จะทำการเพิ่ม Testing sentence นั้น สาเหตุที่การเพิ่ม Training sentence ไม่ส่งผลต่อความแม่นยำของโมเดลนั้น เป็นเพราะ Training sentence ที่เพิ่มเข้าไปอาจจะไม่ได้ไปเพิ่มความน่าจะเป็นใน Naïve Bayes Classifier หรืออาจจะเพิ่มน้อยจนไม่ส่งผลต่อการทำนายเป็นอีกประเภทหนึ่งของ Testing sentence ทำให้ค่าความแม่นยำของโมเดลต่อ Testing sentence เท่าเดิมก่อนการเพิ่ม Training sentence แต่ก็อาจจะส่งผลต่อการทำนายประโยคอื่นๆ นอกเหนือจาก Testing sentence เห็นได้จากการทำนายประโยค “Their codes are amazing”

หลังจากที่ทำการเพิ่ม Testing sentence เข้าไปเป็นจำนวน 5 ประโยค แบ่งออกเป็นประเภท A 2 ประโยคและประเภท B 3 ประโยค นำโมเดลมาทำนายด้วยประโยคเดิม ผลลัพธ์ที่ได้คือ ประโยคทั้ง 2 ถูกทำนายให้อยู่ในประเภท B เช่นเดิม ในส่วนของความแม่นยำนั้นลดลงไปอยู่ที่ 0.727 ซึ่งสาเหตุที่ลดลงเป็นเพราะการเพิ่มขึ้นของ Training sentence ซึ่งตัวของโมเดลนั้นไม่รู้จักข้อมูลเหล่านี้ และไม่เคยพบกรณีใหม่ที่เข้ามา ทำให้ทำนายผลของประโยคที่เข้ามาใหม่พลาด ทำให้ความแม่นยำของโมเดลลดลง