# FIT5217 (T3, 2022) - Assignment 1

**Part 1**

### Question 1.1

We can add the start tags ‘<s>’ and end tags ‘</s>’.

Unigrams:

[('<s>',), ('JINMING',), ('READ',), ('A',), ('DIFFERENT',), ('BOOK',), ('</s>',)]

[('<s>',), ('JINMING',), ('READ',), ('A',), ('DIFFERENT',), ('BOOK',), ('</s>',)]

[('<s>',), ('SHE',), ('READ',), ('A',), ('BOOK',), ('BY',), ('LIU',), ('</s>',)]

Bigrams:

[('<s>', 'YI'), ('YI', 'READ'), ('READ', 'BALL'), ('BALL', 'LIGHTNING'), ('LIGHTNING', '</s>')]

[('<s>', 'JINMING'), ('JINMING', 'READ'), ('READ', 'A'), ('A', 'DIFFERENT'), ('DIFFERENT', 'BOOK'), ('BOOK', '</s>')]

[('<s>', 'SHE'), ('SHE', 'READ'), ('READ', 'A'), ('A', 'BOOK'), ('BOOK', 'BY'), ('BY', 'LIU'), ('LIU', '</s>')]

Trigrams:

[('<s>', 'YI', 'READ'), ('YI', 'READ', 'BALL'), ('READ', 'BALL', 'LIGHTNING'), ('BALL', 'LIGHTNING', '</s>')]

[('<s>', 'JINMING', 'READ'), ('JINMING', 'READ', 'A'), ('READ', 'A', 'DIFFERENT'), ('A', 'DIFFERENT', 'BOOK'), ('DIFFERENT', 'BOOK', '</s>')]

[('<s>', 'SHE', 'READ'), ('SHE', 'READ', 'A'), ('READ', 'A', 'BOOK'), ('A', 'BOOK', 'BY'), ('BOOK', 'BY', 'LIU'), ('BY', 'LIU', '</s>')]

### Question 1.2

P(YI READ A BOOK) = P(YI | <s>) \* P(READ | YI) \* P(A| READ) \* P(BOOK| A) \* P(</s>| BOOK)

P(YI | <s>) = 1/3

P(READ | YI) = 1/1

P(A| READ) = 2/3

P(BOOK| A) = 1/2

P(</s>| BOOK) = 1/2

P(YI READ A BOOK.) = 1/3 \* 1 \* 2/3 \* 1/2 \* 1/2 = 0.056

P(LIU READ A BOOK) = P(LIU | <s>) \* P(READ | LIU) \* P(A| READ) \* P(BOOK| A) \* P(</s>| BOOK)

P(LIU | <s>) = 0

P(LIU READ A BOOK) = 0

### Question 1.3

The number of verbs is 13.

P(YI READ A BOOK) = P(YI | <s>) \* P(READ | YI) \* P(A| READ) \* P(BOOK| A) \* P(</s>| BOOK)

P(YI | <s>) = (1+1)/(3+13)

P(READ | YI) =(1+1)/(1+13)

P(A| READ) = (2+1)/(3+13)

P(BOOK| A) = (1+1)/(2+13)

P(</s>| BOOK) = (1+1) / (2+13)

P(YI READ A BOOK) = 2/16 \* 2/14 \* 3/16\* 2/15\* 2/15 =5.95e-05

P(LIU READ A BOOK) = P(LIU | <s>) \* P(READ | LIU) \* P(A| READ) \* P(BOOK| A) \* P(</s>| BOOK)

P(LIU | <s>) = 1/(3+13) = 1/16

P(READ | LIU) = 1/14

P(LIU READ A BOOK) = 1/16 \* 1/14 \* 3/16 \* 2/15\* 2/15 = 1.488e-5

**Part 2**

### Question 2.1

Given the training data, we could calculate the matrices A, B and in following formula:

### Question 2.2

![D:\my_file\QQ_file\285379133\Image\Group2\[D\Z]\[DZ](5G@MQ7)79DZU`9[`IX.jpg](data:image/jpeg;base64,)

### Question 2.3

, ,

The most state sequence is NV.

### Question 2.4

Assumption 1: A certain state of HMM at any time t only depends on its previous state, which has nothing to do with the state and observation of other times, and has nothing to do with time t.

Assumption 2: The observation at any time only depends on the state of the Markov chain at that time, and is independent of other observations and states. Without such two assumptions, all theorems of HMM are invalid.

**Part 3**

### Question 3.1

1. False
2. True
3. True
4. True
5. False

**Part 4**

### Question 4.1

Original Grammar:

S -> NP ADVP VP

NP -> PRP JJ NN

NP -> DT NN

ADVP -> RB

VP -> VBZ NP

Lexicon:

PRP -> My

JJ -> silver

NN -> cat | dog

RB -> never

VBZ -> likes

DT -> the

CNF Grammar:

S -> X1 VP

X1 -> NP AD

NP -> X2 NN

NP -> DT NN

X2 -> PRP JJ

ADVP -> never

VP -> VBZ NP

**Part 5**

### Question 5.1

Grammar:

[0.25] S -> NP VP

[0.5] S -> NP ADVP VP

[0.25] S -> VP

[0.5] NP -> PRP NN

[0.16] NP -> PRP JJ NN

[0.16] NP -> NN

[0.16] NP -< DT NN

[0.25] VP -> VBD NP

[0.25] VP -> VBG NP

[0.25] VP -> VBZ NP

[0.25] VP -> VBZ S

[1] ADVP -> RB

### Question 5.2

CNF Grammar:

[0.25] S -> NP VP

[0.5] S -> X1 VP

[0.25] S -> VP

[1] X1 -> NP ADVP

[0.5] NP -> PRP NN

[0.16] NP -> X2 NN

[0.16] NP -< DT NN

[0.085] NP -> dog

[0.056] NP -> cat

[0.029] NP -> sausage

[1] X2 -> PRP JJ

[0.25] VP -> VBD NP

[0.25] VP -> VBG NP

[0.25] VP -> VBZ NP

[0.25] VP -> VBZ S

[0.5] ADVP -> never

[0.5] ADVP -> also

Lexicon:

[1] PRP -> My

[0.5] NN -> dog

[0.33] NN -> cat

[0.17] NN -> sausage

[1] VBD -> chased

[1] VBZ -> likes

[1] VBG -> eating

[1] DT -> the

[0.5] RB -> also

[0.5] RB -> never

[1] JJ -> silver

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| The | dog | chased | my | silver | cat |
| DT: 1 | NP: 0.16\*1\*0.5 = 0.08 | None | None | None | S:0.25\* 0.08\*0.0132 = 0.000264 |
|  | NN: 0.5,  NP: 0.085 | None | None | None | S: 0.25\*0.085\*0.0132 = 0.000281 |
|  |  | VBD: 1 | None | None | VP: 0.25\*1\* 0.0528 =0.0132 |
|  |  |  | PRP: 1 | X2: 1 \* 1\* 1 =1 | NP: 0.16\*1\*0.33 =0.0528 |
|  |  |  |  | JJ: 1 | None |
|  |  |  |  |  | NN: 0.33,  NP: 0.056 |

P(The dog chased my silver cat | ) = 0.000264

**Part 6**

Because n-gram language models are only conditional on the local window of linear word level context, they are poor models of long-term syntactic dependence. The author proposes a generated syntax language model, which conditionally constrains the local context tree elements in the parsing tree and fallback to smaller tree elements as needed. By applying the same smoothing technique to n-gram models, they can simply train their models and speed up the training of 1billion data tokens on a machine in a few hours. In addition, their method is based on the success of the n-gram language model and scores all overlapping contexts. Their model is superior to the generated baseline in several evaluation indicators, and achieves the same performance as the most advanced discriminant classifier specially trained for several types of negative data.