**CS 5316 – Natural Language Processing**

**Quiz 2 Solution**

**(Time limit: 12 minutes)**

1. (6 points) Find the minimum edit distance (Levenshtein) and show the corresponding alignment between COMIC and CONVICT. Use the following grid to show your computation.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***X*** | C | O | N | V | I | C | T |
| ***X*** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| C | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| O | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 |
| M | 3 | 2 | 1 | 2 | 3 | 4 | 5 | 6 |
| I | 4 | 3 | 2 | 3 | 4 | 3 | 4 | 5 |
| C | 5 | 4 | 3 | 4 | 5 | 4 | 3 | 4 |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

1. (4 points) Given the following sentences: (a) estimate P(Chinese | food) using add-one smoothing, and (b) list the bigrams that occur 2 times and what is Nc (Good-Turing).

I love Chinese food. I love Italian cuisine. I like Chinese food.

(a)

V: I, love, Chinese, food, Italian, cuisine, like

P(Chinese | food) = (c(food, Chinese)+1)/(c(food) + |V|) = (0+1)/(2+7) = 1/9

(b)

Bigrams occurring twice: I love, Chinese food

N2 = 2

**OR**

(a)

V: I, love, Chinese, food, Italian, cuisine, like, .

P(Chinese | food) = (c(food, Chinese)+1)/(c(food) + |V|) = (0+1)/(2+8) = 1/10

(b)

Bigrams occurring twice: I love, Chinese food, food .

N2 = 3