AI HW11 Yu-Chieh Wang 1164132?

1.

3	W -1000	G +1000	P -1000
2	→ -1	↑	←
1	→ -1	↑	P -1000
	1	2	3

policy

3	-1000	+1000	-1000
2	808.1	899	808.1
1	726	808.1	-1000
	1	2	3

Utilities

a. Passive Reinforcement Learning - Utility

$U^{\pi}(s) = R(s) + \gamma \sum_{s'} P(s' \mid s, \pi(s)) U^{\pi}(s')$

$U([1,1]) = R([1,1]) + 0.9 \sum_{s'} P(s' \mid [1,1], \pi([1,1])) U(s')$

$= -1 + 0.9 [(1) U([2,1]) + (0) U([1,2]) + 0 U([1,1]) + 0 ([1,1])]$

$= -1 + 0.9 [1(0) + 0 + 0 + 0] = -1$

$U([2,1]) = R([2,1]) + 0.9 \sum_{s'} P(s' \mid [2,1], \pi([2,1])) U(s')$

$= -1 + 0.9 [(1) U([2,2])] = -1 + 0.9 [1(0)] = -1$

$U([2,2]) = R([2,2]) + 0.9 \sum_{s'} P(s' \mid [2,2], \pi([2,2])) U(s')$

$= -1 + 0.9 [(1) U([2,3])] = -1 + 0.9 (1(1000)) = 899$

$U([1,2]) = R([1,2]) + 0.9 \sum_{s'} P(s' \mid [1,2], \pi([1,2])) U(s')$

$= -1 + 0.9 [(1) U([2,2])] = -1 + 0.9 [(1)(899)] = 808.1$

$U([3,2]) = R([3,2]) + 0.9 \sum_{s'} P(s' \mid [3,2], \pi([3,2])) U(s')$

$= -1 + 0.9 [(1) U([2,2])] = -1 + 0.9 [(1)(899)] = 808.1$

Because there is only one direction affect each utility, we only need to do parts utilities which their next utilities change.

$U([1,1]) = R([1,1]) + 0.9 [(1) U([2,1])] = -1 + 0.9 [(1)(-1)] = -1.9$

$U([2,1]) = R([2,1]) + 0.9 [(1) U([2,2])] = -1 + 0.9 [(1)(899)] = 808$

$U([1,1]) = R([1,1]) + 0.9 [(1) U([2,1])] = -1 + 0.9 [(1)(808)] = 726$

Since $U([2,2])$ won't change, $U([2,1])$ won't change anymore.
Since $U([2,1])$ won't change, $U([1,1])$ won't change any more.
Because utility function only consider each utility's next utility.

b. Active Reinforcement learning — Temporal difference Q-learning

$$Q(s,a) \leftarrow Q(s,a) + \alpha(R(s) + \gamma \max_{a'} Q(s',a') - Q(s,a))$$

① $Q([1,1],R) \leftarrow Q([1,1],R) + 0.9[R([1,1]) + 0.9Q([2,1],U) - Q([1,1],R)]$

$= 0 + 0.9[(-1) + 0.9(0) - 0] = -0.9$

$Q([2,1],U) \leftarrow Q([2,1],U) + 0.9[R([2,1]) + 0.9Q([2,2],U) - Q([2,1],U)]$

$= 0 + 0.9[(-1) + 0.9(0) - 0] = -0.9$

$Q([2,2],U) \leftarrow Q([2,2],U) + 0.9[R([2,2]) + 0.9Q([2,3],\text{Terminal}) - Q([2,2],U)]$

$= 0 + 0.9[(-1) + 0.9((1000) - 0] = 809.1$

② $Q([1,1],R) \leftarrow Q([1,1],R) + 0.9[R([1,1] + 0.9Q([2,1],U) - Q([1,1],R)]$

$= -0.9 + 0.9[(-1) + 0.9(-0.9) - (-0.9)] = -0.9 - 0.819 = -1.719$

$Q([2,1],R) \leftarrow Q([2,1],U) + 0.9[R([2,1] + 0.9Q([2,2],U) - Q([2,1],U)]$

$= -0.9 + 0.9[(-1) + 0.9(809.1) - (-0.9)] = -0.9 + 655.281 = 654.381$

$Q([2,2],U) \leftarrow Q([2,2],U) + 0.9[R([2,2] + 0.9Q([2,3],\text{Terminal}) - Q([2,2],U)]$

$= 809.1 + 0.9[(-1) + 0.9(1000) - 809.1] = 809.1 + 80.91 = 890.01$

③ $Q([1,1],R) \leftarrow -1.719 + 0.9[(-1) + 0.9(654.381) - 1.719] = -1.719 + 527.601 = 525.882$

$Q([2,1],U) \leftarrow 654.381 + 0.9[(-1) + 0.9(890.01) - 654.381] = 654.381 + 131.065 = 785.446$

$Q([2,2],U) \leftarrow 890.01 + 0.9[(-1) + 0.9(1000) - 890.01] = 890.01 + 8.091 = 898.101$

④ $Q([1,1],R) \leftarrow 525.882 + 0.9[(-1) + 0.9(785.446) - 525.882] = 525.882 + 162.017 = 687.899$

$Q([2,1],U) \leftarrow 785.446 + 0.9[(-1) + 0.9(898.101) - 785.446] = 785.446 + 19.660 = 805.106$

$Q([2,2],U) \leftarrow 898.101 + 0.9[(-1) + 0.9(1000) - 898.101] = 898.101 + 0.809 = 898.910$

⑤ $Q([1,1],R) \leftarrow 687.899 + 0.9[(-1) + 0.9(805.106) - 687.899] = 687.899 + 32.126 = 720.025$

$Q([2,1],U) \leftarrow 805.106 + 0.9[(-1) + 0.9(898.910) - 805.106] = 805.106 + 2.621 = 807.727$

$Q([2,2],U) \leftarrow 898.910 + 0.9[(-1) + 0.9(1000) - 898.910] = 898.910 + 0.081 = 898.991$

⑥ $Q([1,1],R) \leftarrow 720.025 + 0.9[(-1) + 0.9(807.727) - 720.025] = 720.025 + 5.336 = 725.361$

$Q([2,1],U) \leftarrow 807.727 + 0.9[(-1) + 0.9(898.991) - 807.727] = 807.727 + 0.328 = 808.055$

$Q([2,2],U) \leftarrow 898.991 + 0.9[(-1) + 0.9((1000) - 898.991] = 898.991 + 0.008 = 898.999$

⑦ $Q([1,1],R) \leftarrow 725.361 + 0.9[(-1) + 0.9(808.055) - 725.361] = 725.361 + 0.799 = 726.160$

$Q([2,1],U) \leftarrow 808.055 + 0.9[(-1) + 0.9(898.999) - 808.055] = 808.055 + 0.039 = 808.094$

$Q([2,2],U) \leftarrow 898.999 + 0.9[(-1) + 0.9(1000) - 898.999] = 898.999 + 0.000 = 898.999$ ← converge

1b. ⑧ $Q([1,1],R) \leftarrow 726.160 + 0.9[(-1) + 0.9(808.094) - 726.160] = 726.160 + 0.112 = 726.272$

$Q([2,1],U) \leftarrow 808.094 + 0.9[(-1) + 0.9(898.999) - 808.094] = 808.094 + 0.004 = 808.098$

$Q([2,2],U) \leftarrow 898.999 + 0.9[(-1) + 0.9(1000) - 898.999] = 898.999 + 0 = 898.999$

⑨ $Q([1,1],R) \leftarrow 726.272 + 0.9[(-1) + 0.9(808.098) - 726.272] = 726.272 + 0.014 = 726.286$

$Q([2,1],U) \leftarrow 808.098 + 0.9[(-1) + 0.9(898.999) - 808.098] = 808.098 + 0 = 808.098$ ← converge

$Q([2,2],U) \leftarrow 898.999 + 0.9[(-1) + 0.9(1000) - 898.999] = 898.999 + 0 = 898.999$

⑩ $Q([1,1],R) \leftarrow 726.286 + 0.9[(-1) + 0.9(808.098) - 726.286] = 726.286 + 0.001 = 726.287$

$Q([2,1],U) \leftarrow 808.098 + 0.9[(-1) + 0.9(898.999) - 808.098] = 808.098 + 0 = 808.098$

$Q([2,2],U) \leftarrow 898.999 + 0.9[(-1) + 0.9(1000) - 898.999] = 898.999 + 0 = 898.999$

2. Bigram model.

a. the player is next to the gold.

$= P(\text{player} \mid \text{the}) \cdot P(\text{is} \mid \text{player}) \cdot P(\text{next} \mid \text{is}) P(\text{to} \mid \text{next}) P(\text{the} \mid \text{to}) P(\text{gold} \mid \text{the})$

$= \frac{2000}{2000+2000} \cdot \frac{1000}{1000} \cdot \frac{3000}{3000} \cdot \frac{4000}{4000} \cdot \frac{6000}{6000+5000} \cdot \frac{2000}{2000+2000}$

$= \frac{1}{2}(1)(1)(1)\left(\frac{6}{11}\right)\left(\frac{1}{2}\right) = \frac{3}{22} = 0.1363$

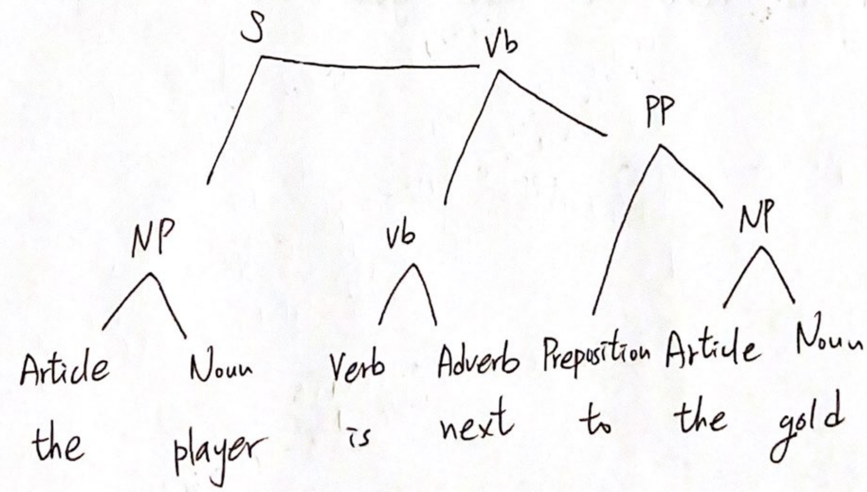
b. the player is next to a pit

$= P(\text{player} \mid \text{the}) P(\text{is} \mid \text{player}) P(\text{next} \mid \text{is}) P(\text{to} \mid \text{next}) P(\text{a} \mid \text{to}) P(\text{pit} \mid \text{a})$

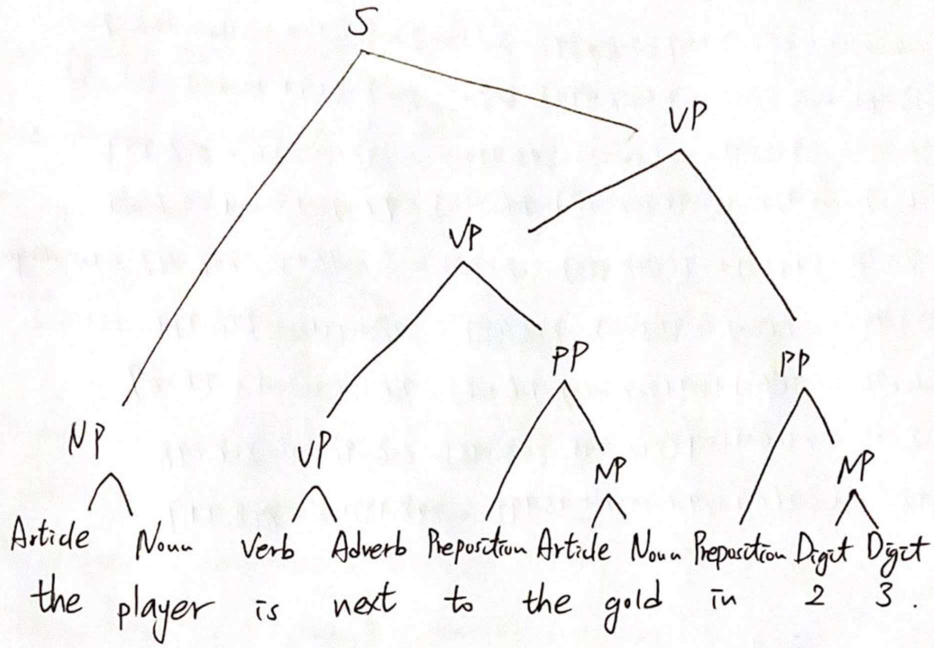
$= \frac{2000}{2000+2000} \cdot \frac{1000}{1000} \cdot \frac{3000}{3000} \cdot \frac{4000}{4000} \cdot \frac{5000}{6000+5000} \cdot \frac{1000}{1000}$

$= \frac{1}{2}(1)(1)(1)\left(\frac{5}{11}\right)(1) = \frac{5}{22} = 0.2272$

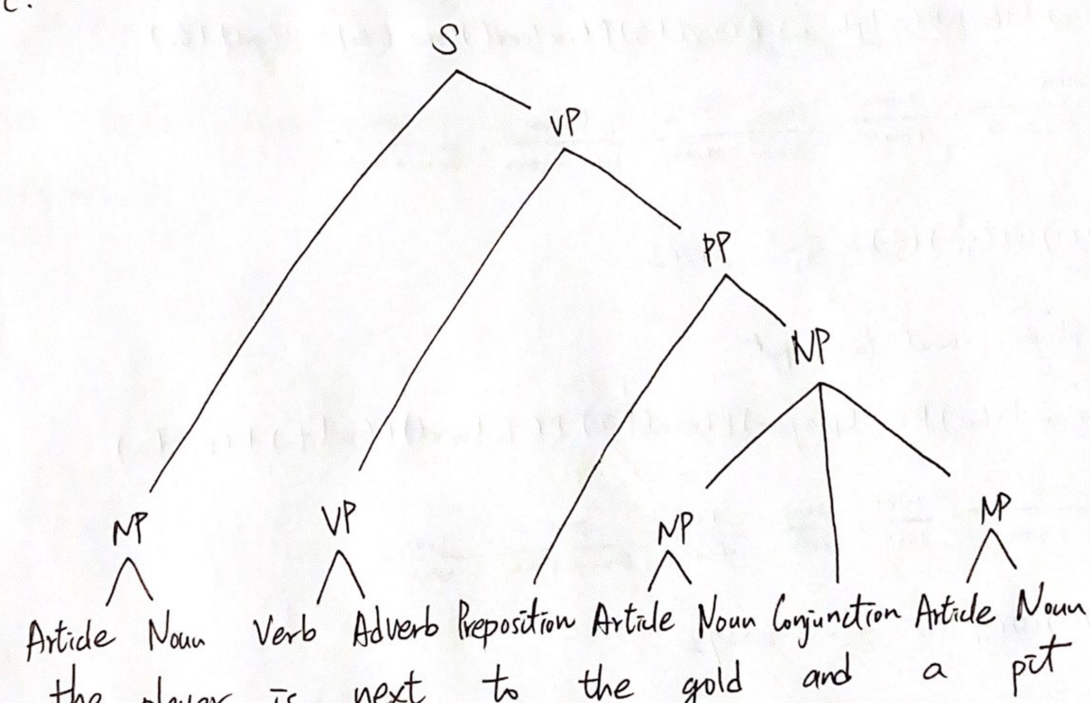
3. a.



3b.


S
VP
VP
PP
PP
NP
VP
NP
NP
Article Noun Verb Adverb Preposition Article Noun Preposition Digit Digit
the player is next to the gold in 2 3.

c.


S
VP
PP
NP
NP
VP
NP
NP
Article Noun Verb Adverb Preposition Article Noun Conjunction Article Noun
the player is next to the gold and a pit

4. Bigram model from ngrams.

a. the player is next to the gold.

$P(\text{player} \mid \text{the}) P(\text{is} \mid \text{player}) P(\text{next} \mid \text{is}) P(\text{to} \mid \text{next}) P(\text{the} \mid \text{to}) P(\text{gold} \mid \text{the})$.

$$= \frac{298493}{644556257} \cdot \frac{58891}{1003165} \cdot \frac{32245}{151715611} \cdot \frac{612358}{4739378} \cdot \frac{34979789}{345046584} \cdot \frac{174684}{644556257}$$

$= (0.000463)(0.058705)(0.000212)(0.129206)(0.101377)(0.000271)$

$= 2.045424e{-14}$

b. the player is next to a pit

$P(\text{player} \mid \text{the}) P(\text{is} \mid \text{player}) P(\text{next} \mid \text{is}) P(\text{to} \mid \text{next}) P(\text{a} \mid \text{to}) P(\text{pit} \mid \text{a})$

$$= \frac{298493}{644556257} \cdot \frac{58891}{1003165} \cdot \frac{32245}{151715611} \cdot \frac{612358}{4739378} \cdot \frac{8309391}{345046584} \cdot \frac{16816}{271737789}$$

$= (0.000463)(0.058705)(0.000212)(0.129206)(0.024081)(0.000061)$

$= 1.093651e{-15}$

AI_ngrams.py

```python
import pandas as pd

data = pd.read_csv('ngrams_words_2.txt', header = None, delimiter="\t")
data.columns = ["a", "b", "c", "d", "e"]
num = 1000000
sum = 0
target = 0
word1 = 'a'
word2 = 'pit'

for i in range(0, num):
    if str.lower(str(data['b'][i])) == word1:
        sum += data['a'][i]
        #print(data['c'][i])
        if str.lower(str(data['c'][i])) == word2:
            target += data['a'][i]

    if i % 10000 == 0:
        print(i, '/', num)
print("sum = ", sum)
print("target =", target)
print("Pro = ", (target/sum))
```