CS 221 Fall 2014 Homework [3]

SUNet ID: [xunwang]

Name: Xun Wang

Collaborators: Joe Fan, Vinod Kumar

By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

Problem 1.

(a) Below is an example that the greedy algorithm fails to find the lowest-cost segmentation of the input:

input string = "saturn"

Use bi-gram model:

c(-begin-, sat) = 1

c(-begin-, saturn) = 2.

c(sat, urn) = 10.

greedy $\Longrightarrow$ 1+10 = 11

dp $\Longrightarrow$ 2

Therefore, the greedy algorithm here does not give the optimal segmentation that DP gives. This is because the greedy algorithm tends to favor the leftmost one in case of a tie.

(b) See submission.py.

Problem 2: Vowel insertion.

a). The example below shows that the greedy algorithm does not give optimal insertion:

input: bty nd bst.

the optimal insertion: beauty and beast.

while greedy gives beaty and best.

greedy algorithm: it from left to right, repeatedly pick the immediate-best vowel insertion, thus favor the shorter word.

b). See submission.py

Problem 3: putting it together.

a) state: a tuple of (index of the character being explored, previously filled word). (state[0], state[1])

action: a string of the previously filled word of the new state.

initial state: (0, wordseg_util.SENTENCE_BEGIN)

goal test: if state[0] == len(query word).

b). see submission.py.

c). $u(w)$ = minimum of b(all words in corpus, w)

$h_u(s)$ = minimum of $u(a)$ where 'a' is a possible action from state S.

According to the equation below: (referring to the lecture slides page-30)

h is heuristic if

$$\begin{cases} Cost'(s,a) = Cost(s,a) + h(Succ(s,a)) - h(s) \geq 0 \\ h(s_{goal}) = 0 \end{cases}$$

$\Rightarrow cost(s,a) + h(succ(s,a)) - h(s) \geq 0$.

$h(succ(s,a))$ is min unigram, which then is the min bigram, therefore $\neq 0$.

$\Rightarrow h(Succ(s,a)) \neq 0$.

Comparing $cost(s,a)$ and $h(s)$, $h(s)$ is the min unigram, which then is the min bi-gram.

$\Rightarrow cost(s,a) \geq h(s)$

$cost(s,a) - h(s) \geq 0$

$(cost(s,a) - h(s) + h(s) \neq 0$ — condition 1.

The heuristic would return 0 if there are no possible action from a state.

Thus $h(s_{goal}) = 0$ — condition 2.

Thus, the heuristic is consistent.

3/3.