Homework 7

Patrick Grasso

December 19, 2016

Contents

1 Chapter 4 Exercises

1.1 1. Validating PageRank

```
# Patrick Grasso
# PageRank algorithm

# T_j   = docs linking to page A
# C(T)  = links in page T
# PR(A) = d + (1-d)*sum(j; PR(T_j) / C(T_j))

iters   = 25

#           A  B  C
graph   = [[0, 1, 1], # A
           [1, 0, 0], # B
           [1, 0, 0]] # C

ranks   =  [1, 1, 1]

d = 0.1

def prnt(i, arr):
    print(("{:2}, " + ", ".join(["{:.2f}"]*len(ranks))).format(i, *arr))

def PR(page):
    term = lambda other: ranks[other] * graph[page][other] / sum(graph[other])
    return d + (1-d)*sum(map(term, range(len(ranks))))
```

```
for i in range(iters + 1):
    prnt(i, ranks)
    for page in range(len(ranks)):
        ranks[page] = PR(page)
```

Output from `validation.py`

```
 0, 1.00, 1.00, 1.00
 1, 1.90, 0.96, 0.96
 2, 1.82, 0.92, 0.92
 3, 1.75, 0.89, 0.89
 4, 1.70, 0.87, 0.87
 5, 1.66, 0.85, 0.85
 6, 1.62, 0.83, 0.83
 7, 1.59, 0.82, 0.82
 8, 1.57, 0.81, 0.81
 9, 1.55, 0.80, 0.80
10, 1.54, 0.79, 0.79
11, 1.53, 0.79, 0.79
12, 1.52, 0.78, 0.78
13, 1.51, 0.78, 0.78
14, 1.50, 0.78, 0.78
15, 1.50, 0.77, 0.77
16, 1.49, 0.77, 0.77
17, 1.49, 0.77, 0.77
18, 1.49, 0.77, 0.77
19, 1.48, 0.77, 0.77
20, 1.48, 0.77, 0.77
21, 1.48, 0.77, 0.77
22, 1.48, 0.77, 0.77
23, 1.48, 0.77, 0.77
24, 1.48, 0.76, 0.76
25, 1.48, 0.76, 0.76
```

1.2 4. PageRank and internet surfing

PageRank is related to the behavior of internet surfing in that it measures pages' popularity based on how many references to the page it finds, similar to how an internet surfer might discover documents by clicking through links on various pages. Statistically, the more links to a page that exist, the higher the probability that an internet surfer will click on a link to that page (generally).

1.3 5. Most common text

I would guess that the most common anchor text is the title of whichever website the anchor points to, which is not necessarily helpful for a page ranking algorithm that relies on such text. There are many instances where people will simply link the name or title of something, say polar bears, which can be inferred from the title of the page being linked to.

2 Abiguous Words

Ambiguous sentences:

1. Hand me the spoon, please
2. He put his hand on top of his head

The ambiguous word here is "hand". In the first sentence, it is meant to mean "pass", as in "pass the bread". In the second sentence, "hand" refers to a human hand, the part of the body attached to the arm.

```
$ python ambiguity.py

Ambiguous Word:   hand

Sentence:     hand me the spoon, please
True Defn:    pass.v.05 - place into the hands or custody of
Predict Defn: pass.v.05 - place into the hands or custody of
Similarity:   1.0

Sentence:     he put his hand on top of his head
True Defn:    hand.n.01 - the (prehensile) extremity of the superior limb
Predict Defn: hand.n.08 - a rotating pointer on the face of a timepiece
Similarity:   0.2222222222222222
```

The first sense of "hand" is matched perfectly. However, the second sentence is identified with a clock hand. Given the structure of the sentence, this somewhat makes sense (if some of the words were changed, it might make sense; at least it is the correct part of speech). However, the use of possessive pronouns like "his" clearly rules this out as a sensical match.