

AI VIET NAM – COURSE 2024

Numpy - Day 21

Ngày 6 tháng 7 năm 2024

Ngày thực hiện:	30/06/2024
Người thực hiện:	Đinh Thị Tâm
Nguồn:	AIO2024 - Day 21
Nguồn dữ liệu (nếu có):	Link Sources code
Từ khóa:	Numpy- Term frequency and Invert Document Frequency
Người tóm tắt:	Đinh Thị Tâm

1. Bài tập:

(a) Code

```
1 import numpy as np
2 import math
3 # B c 1: T o t p t i l i u m u
4 documents = ["T i t h c h c A I", "A I l t r t u n h n t o ",
5             "A G I l s i u t r t u n h n t o "]
6
7 # B c 2: T i n x l í - t c h t v t n h t n s
8
9
10 def compute_tf(doc=documents):
11     var_tmp = [x.strip().split() for x in doc]
12     result = []
13     for d in var_tmp:
14         d_hist = []
15         length = len(d)
16         for x in d:
17             d_hist.append(d.count(x)/length)
18         result.append(d_hist)
19     # print(result)
20     return result
21
22 ## Your code here ##
23 # B c 3: T n h t o n I D F
24
25
26 def compute_idf(docs=documents):
27     # tong so tai lieu D
28     n_doc = len(docs)
29     var_tmp = [x.strip().split() for x in docs]
30     result = []
31     for d in var_tmp:
32         d_hist = []
33         for x in d:
```

```

34         x_counter = 0
35         for t in var_tmp:
36             if (t.count(x) > 0):
37                 x_counter += 1
38             d_hist.append(np.log(n_doc/(1+x_counter)))
39         result.append(d_hist)
40     return (result)
41 # Your code here ##
42 # B   c   4: T   nh   to   n   TF-IDF
43
44
45 def compute_tf_idf(tf, idf):
46     tf_idf = []
47     length = len(tf)
48     for k in range(length):
49         var_tmp = np.array(tf[k])*np.array(idf[k])
50         tf_idf.append(var_tmp)
51     return tf_idf
52
53
54 def print_result(docs, mt_tf_idf):
55     var_tmp = [x.strip().split() for x in docs]
56     n_docs = len(var_tmp)
57     for k in range(n_docs):
58         print('-'*50)
59         print('Doc 1:', var_tmp[k])
60         lenght = len(var_tmp[k])
61         for i in range(lenght):
62             print(f'{var_tmp[k][i]}: {mt_tf_idf[k][i]:.4f}')
63
64
65 if __name__ == "__main__":
66     tf = compute_tf()
67     idf = compute_idf()
68     tf_idf = compute_tf_idf(tf, idf)
69     print_result(documents, tf_idf)

```

(b) Kết quả thực thi chương trình

```

-----
Doc 1: ['Tôi', 'thích', 'học', 'AI']
Tôi: 0.1014
thích: 0.1014
học: 0.1014
AI: 0.0000
-----
Doc 1: ['AI', 'là', 'trí', 'tuệ', 'nhân', 'tạo']
AI: 0.0000
là: 0.0000
trí: 0.0000
tuệ: 0.0000
nhân: 0.0000
tạo: 0.0000
-----
Doc 1: ['AGI', 'là', 'siêu', 'trí', 'tuệ', 'nhân', 'tạo']
AGI: 0.0579
là: 0.0000

```

siêu: 0.0579

trí: 0.0000

tuệ: 0.0000

nhân: 0.0000

tạo: 0.0000