Visualizing LSA

Demo

- Create a fake TFIDF matrix with a strong concept
- Plot the matrix on a two term axis
- Perform SVD decomposition
- Plot the new axes
- Reduce the dimensionality of SVD
- Plot the new axes
Using MATLAB For LSA

Demo

- Show “M”

```matlab
>> M = inv(Sk)*Uk';

M =

Columns 1 through 10
-0.0248  -0.0299  -0.0257  0.0000  -0.0006  -0.0005  -0.0018  0.0000  -0.0230  -0.0017
-0.0000  -0.0000  0.0000  -0.0280  0.0000  0.0000  0.0000  -0.0280  -0.0000  -0.0000
 0.0006  0.0008  0.0005  0.0000  -0.0263  -0.0305  -0.0021  0.0000  0.0006  -0.0024

Columns 11 through 14
-0.0004  -0.0005  0.0000  0.0000
-0.0000  0.0000  -0.0280  -0.0259
-0.0241  -0.0290  0.0000  0.0000
```
Demo

- Demonstrate what SVD is capturing
- 1st concept (1st row of M)

First concept is selecting for Wikipedia?

First concept is selecting for baseball?
Using MATLAB For LSA

Demo

- Demonstrate what SVD is capturing
- 2nd concept (2nd row of M)

First concept is selecting for wikipedia?

Second concept is selecting for UCI?
Demo

- Demonstrate what SVD is capturing
- 3rd concept (3rd row of M)

First concept is selecting for Wikipedia?

Third concept is selecting for coffee?
• **Execute a query “coffee stores”**

```matlab
q = [ 0 0 0 0 (1+log2(1))*log2(c/df(5)) 0 0 0 0 (1+log(1))*log2(c/df(12)) 0 0 ]

Columns 1 through 10

0     0     0     0     1.5850     0     0     0     0     0

Columns 11 through 14

0     2.5850    0     0

qc=M*q'

qc =

-0.0023
  0.0000
-0.0000
-0.1166
```
Using MATLAB For LSA

Demo

- Execute a query

\[
sim(q, d_i) = \frac{\vec{V}(q) \cdot \vec{V}(d_i)}{|\vec{V}(q)| |\vec{V}(d_i)|}
\]

>> Cc = inv(Sk)*Uk'*tfidf

Cc =

-0.9894  -0.1434  -0.0117    0.0000  -0.0020  -0.0189
  0.0000    0.0000    0.0000  -1.0000  -0.0000  -0.0000
  0.0227  -0.0084  -0.2332    0.0000  -0.1147  -0.9653

>> sim = qc'*Cc;
>> sim = sim ./ [norm(Cc(:,1)) norm(Cc(:,2)) norm(Cc(:,3)) norm(Cc(:,4)) norm(Cc(:,5)) norm(Cc(:,6))]

sim =

-0.0004  0.0091  0.1166  -0.0000  0.1166  0.1166
Demo

• Execute a query “coffee stores”

• Answer:
  • starbucks (0.1166)
  • wiki:starbucks(0.1166)
  • wiki:coffee (0.1166)
  • wiki:bat (0.0091)
  • djp3 paper (0.0)
  • wiki:baseballBat (-0.0004)
Using MATLAB For LSA

Demo

- Execute a query “baseball bat”

```matlab
g = [(1+log2(1))*log2(c/df(1)) (1+log2(1))*log2(c/df(2)) 0 0 0 0 0 0 0 0 0 0]
g = 
Columns 1 through 10
2.5850 2.5850 0 0 0 0 0 0 0 0
Columns 11 through 14
0 0 0 0

gc = M'g'
gc =
-0.1413
-0.0000
0.0037
```
Using MATLAB For LSA

Demo

- Execute a query

\[
sim(q, d_i) = \frac{\vec{V}(q) \cdot \vec{V}(d_i)}{|\vec{V}(q)||\vec{V}(d_i)|}
\]

```matlab
>> Cc = inv(Sk)*Uk'*tfidf
Cc =
     -0.9894   -0.1434   -0.0117     0.0000    -0.0020    -0.0189
     0.0000     0.0000     0.0000    -1.0000    -0.0000    -0.0000
     0.0227   -0.0084   -0.2332     0.0000    -0.1147    -0.9653

>> qc = gc'*Cc;
>> sim = sim ./ [norm(Cc(:,1)) norm(Cc(:,2)) norm(Cc(:,3)) norm(Cc(:,4)) norm(Cc(:,5)) norm(Cc(:,6))]

sim =
     0.1414     0.1408     0.0035     0.0000    -0.0012    -0.0009
```
Demo

• Execute a query “baseball bat”

• Answer:
  • wiki:baseballBat (0.1414)
  • wiki:bat (0.1408)
  • wiki:coffee (0.0035)
  • djp3 paper (0.000)
  • wiki:starbucks (-0.0009)
  • starbucks (-0.0012)