Big Data Analytics

ESSEC

Olga Klopp

Solution Home work 7 (Recommendation Systems)

1. **Exercise 9.2.1 MMDS book:** Three computers, A, B, and C, have the numerical features listed below:

Feature	A	B	C
Processor Speed	3.06	2.68	2.92
Disk Size	500	320	640
Main-Memory Size	6	4	6

We may imagine these values as defining a vector for each computer; for instance, A's vector is $[3.06, 500, 6]$. We can compute the cosine distance between any two of the vectors, but if we do not scale the components, then the disk size will dominate and make differences in the other components essentially invisible. Let us use 1 as the scale factor for processor speed, α for the disk size, and β for the main memory size.

Solution:

(a) In terms of α and β, compute the cosines of the angles between the vectors for each pair of the three computers. For example for A and B we have

$$\cos(A,B) = \frac{3.06 \times 2.68 + \alpha^2 \times 500 \times 320 + \beta^2 \times 24}{\sqrt{3.06^2 + \alpha^2 \times 500^2 + 36\beta^2}\sqrt{2.68^2 + \alpha^2 \times 320^2 + 16\beta^2}}$$

and

$$\cos(A,B) = \frac{8.2008 + 20^4\alpha^2 + 24\beta^2}{\sqrt{9.3636 + 25 \times 10^4 \times \alpha^2 + 36\beta^2}\sqrt{7.1824 + 102400 \times \alpha^2 + 16\beta^2}}$$

(b) For $\alpha = \beta = 1$ we have

$$\cos(A,B) = \frac{8.2008 + 20^4 + 24}{\sqrt{9.3636 + 25 \times 10^4 + 36}\sqrt{7.1824 + 102400 + 16}}$$

and

$$\cos(A,B) = \frac{160032, 2008}{500.045361542 \times 320.03622045} = 0.999997333$$

(c) For $\alpha = 0.01$ and $\beta = 0.5$ we have

$$\cos(A,B) = \frac{8.2008 + 0.01^2 \times 20^4 + 0.5^2 \times 24}{\sqrt{9.3636 + 0.01^2 \times 500^2 + 36 \times 0.5^2}\sqrt{7.1824 + 0.01^2 \times 320^2 + 16 \times 0.5^2}}$$

and

$$\cos(A,B) = \frac{30.2008}{6.585104403 \times 4.62843386} = 0.990881501$$

(d) A certain user has rated the three computers as follows: A: 4 stars, B: 2 stars, C: 5 stars.

- We have the average rating equal to $11/3$, so the normalized ratings are:

$$(1/3, -5/3, 4/3)$$

- User Profile for
  - For Disk Size: $(500 \times 1/3) + (320 \times -5/3) + 640 \times 4/3 = 486.667$
  - For Processor Speed: $(3.06 \times (1/3)) + (2.68 \times -5/3) + (2.92 \times 4/3) = 0.4467$
  - For Main-memory size: $(6 \times 1/3) + (4 \times -5/3) + (6 \times 4/3) = 3.333$

  So, user profile is $\frac{1}{3}(486.667, 0.4467, 3.33) = (162.2, 0.1489, 1.11)$.

2. **Exercise 9.3.1 MMDS book:** The following Figure is a utility matrix, representing the ratings, on a 1–5 star scale, of eight items, a through h, by three users A, B, and C.

	a	b	c	d	e	f	g	h
A	4	5		5	1		3	2
B		3	4	3	1	2	1	
C	2		1	3		4	5	3

Compute the following from the data of this matrix.

(a) Treating the utility matrix as boolean, compute the Jaccard distance between each pair of users.

(b) Repeat Part (a), but use the cosine distance.

(c) Treat ratings of 3, 4, and 5 as 1 and 1, 2, and blank as 0. Compute the Jaccard distance between each pair of users.

(d) Repeat Part (c), but use the cosine distance.

(e) Normalize the matrix by subtracting from each nonblank entry the average value for its user.

(f) Using the normalized matrix from Part (e), compute the cosine distance between each pair of users.

Solution:

(a) For example $Sim(A, B) = 4/8 = 1/2$.

(b) $\cos(A, B) = \dfrac{1 \times 0 + 1 + 1 \times 0 + 1 + 1 + 1 \times 0 + 1 + 1 \times 0}{\sqrt{6}\sqrt{6}} = \dfrac{4}{6} = \dfrac{2}{3}$

(c) $Sim(A, B) = 2/5$

(d) $\cos(A, B) = \dfrac{2}{\sqrt{4}\sqrt{3}} = \dfrac{1}{\sqrt{3}}$

(e)

	a	b	c	d	e	f	g	h
A	2/3	5/3		5/3	-7/3		-1/3	-4/3
B		2/3	5/3	2/3	-4/3	-1/3	-4/3	
C	-1		-2	0		1	2	0

(f) For example,

$$\cos(A, B) = \frac{9 \times (2 \times 5/3 \times 2/3 + 7/3 \times 4/3 + 1/3 \times 4/3)}{\sqrt{4 + 25 + 25 + 49 + 1 + 16}\sqrt{4 + 25 + 4 + 16 + 1 + 16}}$$

and

$$\cos(A, B) = \frac{52}{\sqrt{120}\sqrt{66}} \approx 0.5843$$