Intelligent Data Mining - Exercise 7

Michael Debono Mrđen

12 January 2018

1   Assignment 1: Content-based recommendation

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.

a. In terms of α and β, compute the cosines of the angles between the vectors for each pair of the three computers.

- $A = [3.06, 500\alpha, 6\beta]$
- $B = [2.68, 320\alpha, 4\beta]$
- $C = [2.92, 640\alpha, 6\beta]$

$$\cos(AB) = \frac{(3.06)(2.68) + (500\alpha)(320\alpha) + (6\beta)(4\beta)}{\sqrt{3.06^2 + (500\alpha)^2 + (6\beta)^2}\sqrt{2.68^2 + (320\alpha)^2 + (4\beta)^2}}$$

$$\cos(AC) = \frac{(3.06)(2.92) + (500\alpha)(640\alpha) + (6\beta)(6\beta)}{\sqrt{3.06^2 + (500\alpha)^2 + (6\beta)^2}\sqrt{2.92^2 + (640\alpha)^2 + (6\beta)^2}}$$

$$\cos(CB) = \frac{(2.92)(2.68) + (640\alpha)(320\alpha) + (6\beta)(4\beta)}{\sqrt{2.92^2 + (640\alpha)^2 + (6\beta)^2}\sqrt{2.68^2 + (320\alpha)^2 + (4\beta)^2}}$$

b. What are the angles between the vectors if $\alpha = \beta = 1$?

	cos	$\angle$
AB	0.999997	0.1403°
AC	0.9999953431	0.1749°
CB	0.9999878534	0.2824°

c. What are the angles between the vectors if $\alpha = 0.01$ and $\beta = 0.5$?

	cos	$\angle$
AB	0.990882	7.743°
AC	0.991555	7.451°
CB	0.969178	14.26°

d. A certain user has rated the three computers as follows: A: 4 stars, B: 2 stars, C: 5 stars. Normalize the ratings for this user.

Average rating:
$$\frac{4 + 2 + 5}{3} = \frac{11}{3}$$

Normalized ratings:

A	B	C
1/3	-5/3	4/3

e. Compute a user profile for the user, with components for processor speed, disk size, and main memory size.

- Processor Speed $= [3.06, 2.68, 2.92]$
- Disk Size $= [500, 320, 640]$
- Main-Memory Size $= [6, 4, 6]$

Processor Speed component:
$$\frac{(1/3)(3.06) - (5/3)(2.68) + (4/3)(2.92)}{3} = 0.1489$$

Disk Size component:
$$\frac{(1/3)(500) - (5/3)(320) + (4/3)(640)}{3} = 162.2$$

Main-Memory Size component:
$$\frac{(1/3)(6) - (5/3)(4) + (4/3)(6)}{3} = 1.111$$

User profile:
$$[0.1489, 162.2, 1.111]$$

2 Assignment 2: Singular Value Decomposition

Use the SVD form Fig. 11.7. Suppose Leslie assigns rating 3 to Alien and rating 4 to Titanic, giving us a representation of Leslie in "movie space" of [0,3,0,0,4]. Find the representation of Leslie in concept space. What does that representation predict about how well Leslie would like the other movies appearing in our example data?

M' is M including Leslie as the last row.

$$M' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 3 & 0 & 0 & 4 \end{bmatrix}$$

U' is the concept space for the rows of M' and is U with the addition of the eigenvector introduced by Leslie.

$U'_{Leslie} = [x, y]$

This shows that Leslie would like romance movies slightly more than science fiction movies(, maybe).

3 Assignment 3: Singular Value Decomposition with Mahout

Use the data "netflix_data.txt" uploaded in StudIP which contains real Netflix user ratings of 100 different movies. Pick 3 different users of your choice and print out the movie titles of the first 4 recommendations of each of the three following Mahout recommenders:

- SVDRecommender using the SVDPlusPlusFactorizer

- SVDRecommender using the ALSWRFactorizer

- GenericUserBasedRecommender using and appropriate UserSimilarity and UserNeighborhood