Human-centric Artificial Intelligence

2020/2021

Worksheet #9:
Recommender Systems (RecSys)

Luís Macedo
Amílcar Cardoso, Hugo Oliveira

9.1 Topics

- Recommender Systems
  - Collaborative-based RecSys
  - Content-based RecSys
  - Hybrid RecSys
  - Evaluation of RecSys

9.2 Pre-class Work

- Readings:
  - Ricci et al. [2015], Ricci et al. [2011]
  - Informing, suggesting, and helping other artificial or human agents making decisions (Russell and Norvig [2010] (ch. 17))
- Tutorials, video-classes:
  - https://www.youtube.com/watch?v=9AP-DgFBNP4
  - https://towardsdatascience.com/how-youtube-recommends-videos-b6e003a5ab
  - https://www.kdnuggets.com/?s=recommender+systems
  - https://www.kdnuggets.com/?s=recommender+systems
  - https://www.kdnuggets.com/2019/07/building-recommender-system-part-2.html
  - https://www.kdnuggets.com/2018/02/recommender-engine.html
  - https://www.kdnuggets.com/2015/12/cartoon-surprise-data-science-recomme.html
  - https://www.kdnuggets.com/2019/04/k-means-clustering-unsupervised-learn.html
  - https://www.kdnuggets.com/2019/10/youtube-recommending-next-video.html

9.3 Theoretic-Practical Exercises

Problem 9.1

Define and give some examples of Recommender Systems.

Problem 9.2

What are the benefits of Recommender Systems?

Problem 9.3

What are the main categories of Recommender Systems? Present their differences, Pros and Cons.

Problem 9.4

Consider the following data set:

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

1. Compute the similarity between Alice and each one of the other users, using Pearson Correlation:

$$sim(a,b) = \frac{\sum_{p\in P}(r_{a,p}-\overline{r}_a)(r_{b,p}-\overline{r}_b)}{\sqrt{\sum_{p\in P}(r_{a,p}-\overline{r}_a)^2}\sqrt{\sum_{p\in P}(r_{b,p}-\overline{r}_b)^2}}$$

Solution

We start by computing the average of ratings of each user:

$\overline{r}_{Alice} = (5+3+4+4)/4 = 4$

$\overline{r}_{U1} = (3+1+2+3)/4 = 2.4$

$\overline{r}_{U2} = (4+3+4+3)/4 = 3.5$

$\overline{r}_{U3} = (3+3+1+5)/4 = 3$

$\overline{r}_{U4} = (1+5+5+2)/4 = 3.25$

Now we can use this average ratings to compute the Pearson Correlation between two users. In this case, we are going to compute the similarity between Alice (denote for short by A) and each one of the other users:

$$sim(A,U1) = \frac{\sum_{p\in P}(r_{A,p}-\overline{r}_A)(r_{U1,p}-\overline{r}_{U1})}{\sqrt{\sum_{p\in P}(r_{A,p}-\overline{r}_A)^2(r_{U1,p}-\overline{r}_{U1})^2}}$$

$$= \frac{(5-4)(3-2.4)+(3-4)(1-2.4)+(4-4)(2-2.4)+(4-4)(3-2.4)}{\sqrt{(5-4)^2+(3-4)^2+(4-4)^2+(4-4)^2}\sqrt{(3-2.4)^2+(1-2.4)^2+(2-2.4)^2+(3-2.4)^2}} = 0.85$$

Using a similar procedure, we can compute the rest of the similarities:

$sim(A,U2) = 0.70$

$sim(A, U3) = 0$

$sim(A, U4) = -0.79$

2. What is the neighbor set of Alice if we confine the set to the two most similar ones?

   Solution From the previous computations, we can restrict the neighbor set to:

   $N_{Alice} = \{U1, U2\}$

3. Generate a prediction for the rating of Alice for item 5, based on the neighbor's ratings and using the prediction function:

   $Pred(\overline{r}_{a,p}) = \overline{r}_a + K \sum_{b \in N_a} sim(a, b)(r_{b,p} - \overline{r}_b)$

   where

   $k = \frac{1}{\sum_{b \in N_a} sim(a,b)}$

   Solution

   $Pred(\overline{r}_{a,p}) = 4 + \frac{1}{0.85+0.70}[0.85(3 - 2.4) + 0.70(5 - 3.8)] = 4.87$

Problem 9.5

Consider the following data set:

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	1	?
User1	3	1	2	5	5
User2	4	3	3	3	2
User3	3	3	1	5	4
User4	1	5	5	2	1

1. Compute the similarity between Alice and each one of the other users, using the following distance function (Manhatten Distance):

   $d(x, y) = \sum_{p \in P} \|r_{x,p} - r_{y,p}\|$

   Solution $d(A, U1) = \|5 - 3\| + \|3 - 1\| + \|4 - 2\| + \|1 - 5\| = 10$

   Using a similar procedure, we can compute the rest of the distances:

$d(A, U2) = 4$

$d(A, U3) = 9$

$d(A, U4) = 8$

Normalizing (the maximum distance is 4x5=20):
$d(A, U1) = 10/20$

$d(A, U2) = 4/20$

$d(A, U3) = 9/20$

$d(A, U4) = 8/20$

The similarities are:
$sim(A, U1) = 1 - 10/20 = 10/20$

$sim(A, U2) = 1 - 4/20 = 16/20$

$sim(A, U3) = 1 - 9/20 = 11/20$

$sim(A, U4) = 1 - 8/20 = 12/20$

2. What is the neighbor set of Alice if we confine the set to the users with (non-normalized) distance less than or equal to 8?

   Solution From the previous computations, we can restrict the neighbor set to:
   $N_{Alice} = \{U2, U4\}$

3. Generate a prediction for the rating of Alice for item 5, based on the neighbor's ratings and using the prediction function:

   $Pred(\overline{r}_{a,p}) = \overline{r}_a + K \sum_{b \in N_a} sim(a, b)(r_{b,p} - \overline{r}_b)$

   where

   $k = \frac{1}{\sum_{b \in N_a} sim(a,b)}$

   Solution $Pred(\overline{r}_{Alice,Item5}) = \frac{13}{4} + \frac{1}{\frac{16}{20} + \frac{12}{20}} \left[\frac{16}{20}(2 - 3.25) + \frac{12}{20}(1 - 3.25)\right]$

Problem 9.6 Repeat the previous exercise, considering other similarity metrics such as the Cosine Similarity, the adjusted Cosine Similarity, the Euclidean Distance, Minskowski Distance, Hamming Distance.

Problem 9.7 The previous exercises rely on user-based collaborative filtering. Repeat now the previous exercises, using item-based collaborative filtering.

Problem 9.8 Consider the following data:

	Item1	Item2	Item3	Item4	Item5
Alice	5		4	1	
User1	3	1	4	5	5
User2	4		3	3	2
User3	3	3		5	4
User4	1	5	5		1

1. Compute the similarity between Alice and each one of the other users, using Jaccard Index:
   $J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A|+|B|-|A \cap B|}$
2. What is the neighbor set of Alice if we confine the set to the two most similar ones?
3. Generate a prediction for the rating of Alice for item 5, based on the neighbor's ratings and using the prediction function:
   $Pred(\overline{r}_{a,p}) = \overline{r}_a + K \sum_{b \in N_a} sim(a,b)(r_{b,p} - \overline{r}_b)$
   where
   $k = \frac{1}{\sum_{b \in N_a} sim(a,b)}$

Bibliography

Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor. *Recommender systems handbook*. Springer, New York; London, 2011.

Francesco Ricci, Lior Rokach, and Bracha Shapira. *Recommender Systems Handbook*. Springer Publishing Company, Incorporated, 2nd edition, 2015. ISBN 1489976361.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition, 2010.