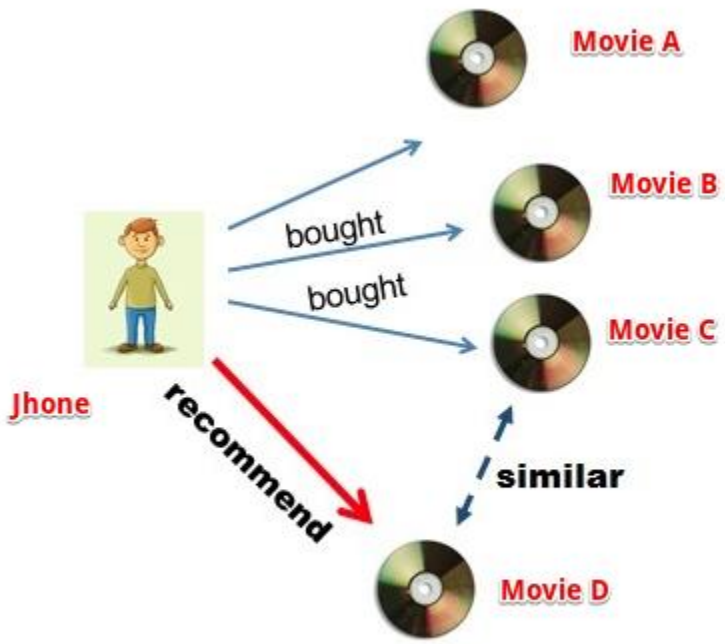
Item Based Collaborative Filtering

Item-item collaborative filtering is one kind of recommendation method which looks for similar items based on the items users have already liked or positively interacted with. It was developed by Amazon in 1998 and plays a great role in Amazon's success.



How IBCF works are that it suggests an item based on items the user has previously consumed. It looks for the items the user has consumed then it finds other items similar to consumed items and recommends accordingly.

Let's understand this with an example. Suppose our user Jhone wants to purchase a movie DVD. Our job is to recommend him a movie based on his past preferences. We will first search for movies that Jhone has watched or liked, let's call those movies 'A', 'B' and 'C'. Next, we will search for other movies similar to three movies. Suppose we found out that movie 'D' is highly similar to 'C', therefore, there is a highly likely chance that Jhone will also like movie 'D' because it is similar to one Jhone has already liked. Hence, we will suggest the movie 'D' to Jhone.

So at its core IBCF is all about finding items similar to the ones user has already liked. But how to find similar items? and what if there are multiple similar items in that case which item to suggest first? To understand this lets' first understand the intuition behind the process, this will assist us to comprehend the mathematics behind the IBCF recommendation filtering.

Finding Similarity Between Items

Suppose we have a table of users and their ratings for movies

User	1: Toy Story	2: Star Wars: Epi:	356: Forrest Gump	318: Shawshank Redemption,	593: Silence of the Lambs,	3578: Gladiator
755	2	5	2		4	4
5277	1			2	4	2
1577				5	2	
4388	2	3				1
1202		3	4	1	4	1
3823	3	4	4	4		
5448			3	1	1	4
5347	2				3	2
4117	4	1		4	2	4
2765	4	2		5	3	
5450	5	1	5			5
139	2	5	2			
1940	4			5	4	
3118	3		3		2	
4656	2	4			5	5

Let's pick two movies (Id 1: Toy Story and Id 2: Star Wars) for which we have to calculate the similarity i.e. how much these two movies are comparable to one another in terms of their likeness by users. To compute this we will:

First multiple ratings of both movies with each other and the sum of the result. Let's call this value 'A'.

User	1: Toy Story	2: Star Wars:	Product of Movie Ratings
755	2	5	10
5277	1		0
1577			0
4388	2	3	6
1202		3	0
3823	3	4	12
5448			0
5347	2		0
4117	4	1	4
2765	4	2	8
5450	5	1	5
139	2	5	10
1940	4		0
3118	3		0
4656	2	4	8
4796			0
6037	2		0
3048	4	5	20
4790	2	1	2
4489	2	2	4
		A =	89

Product of ratings

Sum of product of ratings

Second, we'll sum the squared movie ratings and then take the square root of them. So square all movie 1 ratings, sum them and then take the square root to get the final value (do the same for movie 2). Doing so will get us two values i.e. square root value of movie 1 and movie 2. Multiply both values. We call this final value 'B'

1. Square ratings

2. Sum of values from 1

User	1: Toy Story (	2: Star Wars:	Product of Mo	Square of Movie 1	Square of Movie 2
755	2	5	10	4	25
5277	1		0	1	0
1577			0	0	0
4388	2	3	6	4	9
1202		3	0	0	9
3823	3	4	12	9	16
5448			0	0	0
5347	2		0	4	0
4117	4	1	4	16	1
2765	4	2	8	16	4
5450	5	1	5	25	1
139	2	5	10	4	25
1940	4		0	16	0
3118	3		0	9	0
4656	2	4	8	4	16
4796			0	0	0
6037	2		0	4	0
3048	4	5	20	16	25
4790	2	1	2	4	1
4489	2	2	4	4	4
		A =	89	B =	137.9855065

Sum of Movie 1 Ratings	Sum of Movie 2 Ratings
140	136

Square Root of Sum of Movie 1 Ratings	Square Root of Sum of Movie 2 Ratings
11.83215957	11.66190379

3. Square root the result of 2

Product of values from 3

Third, divide A and B, this will get us a score that indicates how close Movie 1 and Movie 2 are to one another (link).

User	1: Toy Story (	2: Star Wars:	Product of Mov	Square of Movie 1	Square of Movie 2
755	2	5	10	4	25
5277	1		0	1	0
1577			0	0	0
4388	2	3	6	4	9
1202		3	0	0	9
3823	3	4	12	9	16
5448			0	0	0
5347	2		0	4	0
4117	4	1	4	16	1
2765	4	2	8	16	4
5450	5	1	5	25	1
139	2	5	10	4	25
1940	4		0	16	0
3118	3		0	9	0
4656	2	4	8	4	16
4796			0	0	0
6037	2		0	4	0
3048	4	5	20	16	25
4790	2	1	2	4	1
4489	2	2	4	4	4
		A =	89	B =	137.9855065

Sum of Movie 1 Ratings	Sum of Movie 2 Ratings
140	136

Square Root of Sum of Movie 1 Ratings	Square Root of Sum of Movie 2 Ratings
11.83215957	11.66190379

similarity between movie 1 and 2

Similarity (Movie 1 and 2) = 0.6449952772

Repeating the above process for all the movies will result in a table with similarities between each movie(in general we call them items).

Here is how the above process is depicted in mathematical form.

$$similarity(i,j) = \frac{\sum_u^U r_{(u,i)} r_{(u,j)}}{\sqrt{\sum_u^U r_{(u,i)}^2}\sqrt{\sum_u^U r_{(u,j)}^2}}$$

Adding some labels to the letters will ease in understanding of each part of the equation.