

JOKE RECOMMENDER

~ If you think an apple a day keeps the doctor away, try a joke!

Northwestern

ANJALI VERMA

MOTIVATION FOR APP

The sound of roaring laughter is far more contagious than any cough, snuffle, or sneeze. When laughter is shared, it binds people together and increases happiness

The motivation for this application lies in engaging people in humor with jokes and content that match their interests. Help them discover jokes they might not have found otherwise and ultimately spread happiness by sharing jokes

The Joke Recommender app enables users to rank jokes on a visual analog scale and produces updated recommendations based on their inputs

A man joins a big corporate empire as a trainee. On his very first day of work, he dials the pantry and shouts into the phone:

"Get me a coffee, quickly!"

The voice from the other side responds,

"You fool, you've dialed the wrong extension! Do you know who you're talking to, dumbbo?"
"It's the CEO of the company"

The trainee shouts back,

"And do YOU know who YOU are talking to, you fool?!"

the CEO indignantly replies

No!

"Good!" replied the trainee, and puts down the phone.'

DATA

Comprises of two tables :

- **Jokes** ~ Complete textual jokes mapped to each joke id
- **Ratings** ~ User ratings corresponding to each joke

Source :

http://eigentaste.berkeley.edu/dataset/jester_dataset_2.zip

Dataset had about 1.7 million rows (ratings records) and 3 columns namely User_ID, JokeID and Rating with **150 unique jokes**

Sampled about 700,000 records based on the top 20 most rated jokes (**20 unique jokes** and **59,132 unique users**)

Joke ratings range from -10 to 10 with median rating being 2.2

- No missing values or inconsistencies
- No duplicate records
- Transformed data into user-item interaction matrix
- Sparsity for the interaction matrix is 0.59

Jokes Table

JokeID		Joke
0	1	A man visits the doctor. The doctor says, "I h...
1	2	This couple had an excellent relationship goin...
2	3	Q. What's 200 feet long and has 4 teeth? A. Th...
3	4	Q. What's the difference between a man and a t...
4	5	Q. What's O. J. Simpson's web address? A. Slas...

Ratings Table

	User_ID	JokeID	Rating
1	1	7	-9.281
2	1	8	-9.281
3	1	13	-6.781
4	1	15	0.875
5	1	16	-9.656

User Joke Interaction

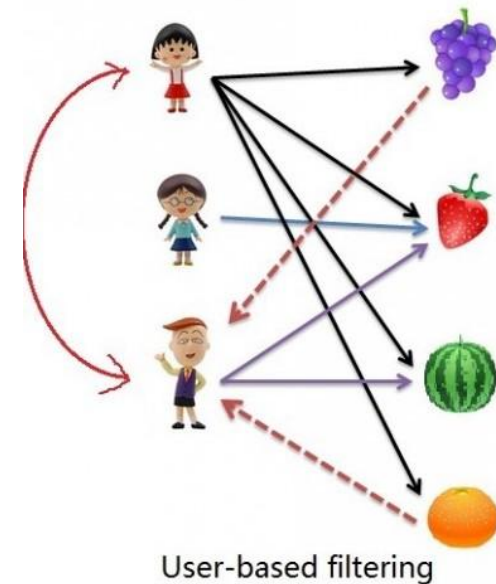
	7	8	13	15	16	17	18
0	-9.281	-9.281	-6.781	0.875	-9.656	-9.031	-7.469
1	9.938	9.531	9.938	0.406	3.719	9.656	-2.688
2	-9.844	-7.219	-2.031	-9.938	-9.969	-9.875	-9.812
3	-4.500	-4.906	0.000	0.000	0.000	0.000	0.000
4	4.750	-5.906	-0.406	-4.031	3.875	6.219	5.656

MODEL

RECOMMENDATION TECHNIQUE

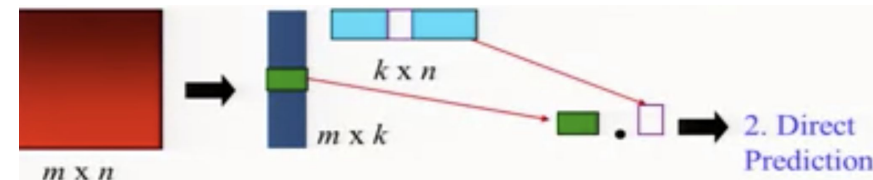
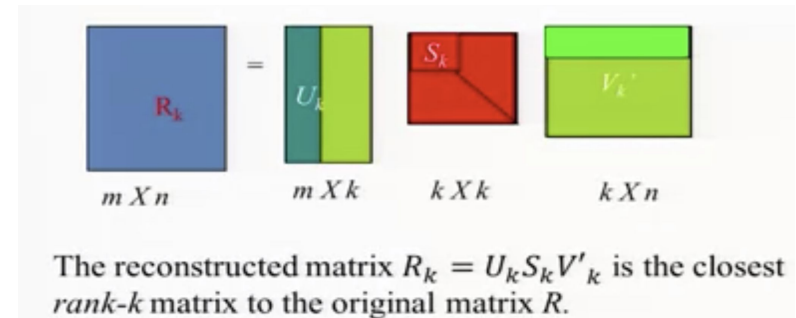
~ *Collaborative Filtering*

Making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating).



ALGORITHM ~ *Singular Value Decomposition*

- New user's rating for a joke taken as input in the form: (eg. if user gives joke at index 7, a rating of 5.
[0,0,0,0,0,0,5,0,0,0,0,0,0,0,0,0,0,0])
- New user rating appended to existing interaction matrix
- New matrix with appended user (dimensions $m \times n$) decomposed into 3 simpler matrices U (dimension $m \times k$), $Sigma(S)$ (dimension $k \times k$), and V (dimension $k \times n$)
- SVD predicts joke ratings for the new user and the joke with the highest rating other than the joke the user has already seen, gets recommended



MODEL

SUCCESS CRITERION ~

Evaluation Metric

- Split data into train and test sets - 80/20
- Predicted joke ratings for each user in the test set
- k is a tuning parameter. For k ranging from 1 to 19 :
- Calculated Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) using:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad RMSE = \sqrt{MSE}$$

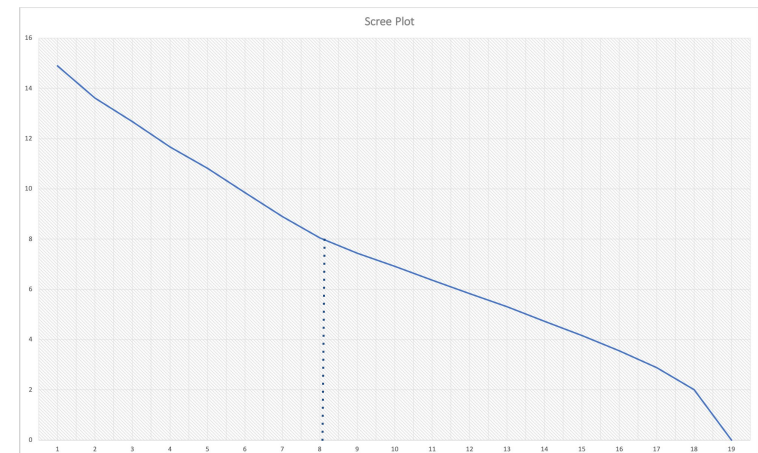
where y_i = actual rating

\hat{y} = predicted rating

n = number of test users

- Plot the RMSE values against k values to find optimal k as 8 with RMSE of 8.05 such that original matrix is approximated with lowest possible dimensions

k	MSE	k	RMSE
1	221.94	1	14.90
2	185.26	2	13.61
3	160.66	3	12.68
4	135.97	4	11.66
5	116.88	5	10.81
6	97.03	6	9.85
7	79.09	7	8.89
8	64.75	8	8.05
9	55.34	9	7.44
10	47.77	10	6.91
11	40.51	11	6.37
12	34.07	12	5.84
13	28.16	13	5.31
14	22.31	14	4.72
15	17.33	15	4.16
16	12.66	16	3.56
17	8.26	17	2.87
18	4.00	18	2.00
19	0.00	19	0.00

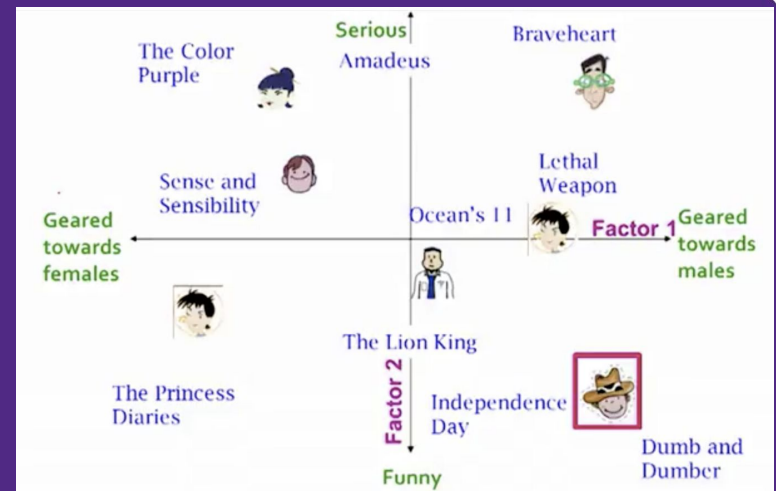


INTERESTING INSIGHT

SVD - Behind the scenes

- Decreases the dimension of the interaction matrix by extracting its latent factors.
- Maps each user and each item into a latent space with dimension k .
- Better understand the relationship between users and items as they become directly comparable

**Latent factor - broad idea which describes a property or concept that a user or an item has*



THANK YOU!

Questions?

Northwestern

ANJALI VERMA
anjaliverma2020@u.northwestern.edu
Master of Science in Analytics
Northwestern University