DATA 643 Proj 3

Dan Fanelli

June 19, 2017

Matrix Factorization methods

The last.fm story continued. . .

Figure 1:

From previous projects, we:

- Got listen count data from last.fm
- Filtered out the bad rows (?????)
- Derived a 0 to 5 score based on the listen count
- Performed top-N suggestions using UBCF (User Based Collaborative Filtering)
- Performed top-N suggestions using IBCF (Item Based Collaborative Filtering)

Now, we:

- Query google for "top [music_genre] bands" for 5 specific genres
- Set user 1 to rate (only) google's top 5 bands for that genre a 5/5
- Use recommenderlab's "SVD" strategy to suggest bands when K=200, 100, 50 and 10
- Analyze the differences in K values, see if our suggestions match any of google's top 18

SVD: google “top classic rock bands”

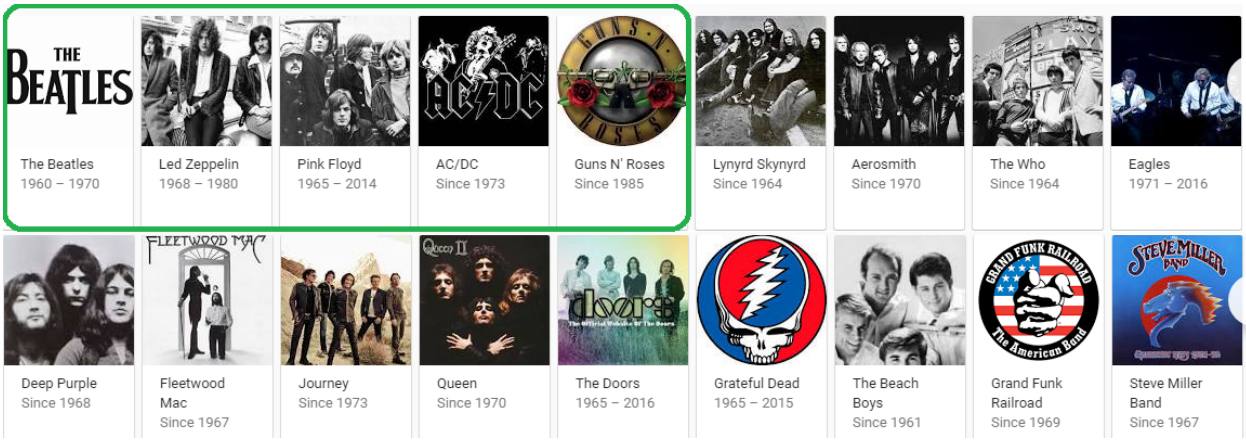

Figure 2:

If USER_1 likes the following artists:

userID	artistID	artist	listen_count
1	227	The Beatles	5
1	1412	Led Zeppelin	5
1	163	Pink Floyd	5
1	706	AC/DC	5
1	1249	Guns N” Roses	5

Then we suggest:

K	User	Rec_1	Rec_2	Rec_3	Rec_4	Rec_5	Rec_6
500	1	Johann Sebastian Bach	Johannes Brahms	Gramofocas	The Who	American Hi-Fi	Nightwish
200	1	Nightwish	The Who	Slipknot	Gramofocas	Queen	Johann Se
100	1	Slipknot	Nightwish	Queen	Avenged Sevenfold	Pearl Jam	Iron Maid
50	1	Metallica	Iron Maiden	Nirvana	Queen	Scorpions	RBD
10	1	Coldplay	Oasis	Muse	The Killers	Radiohead	Depeche M

SVD: google “top christian contemporary bands”

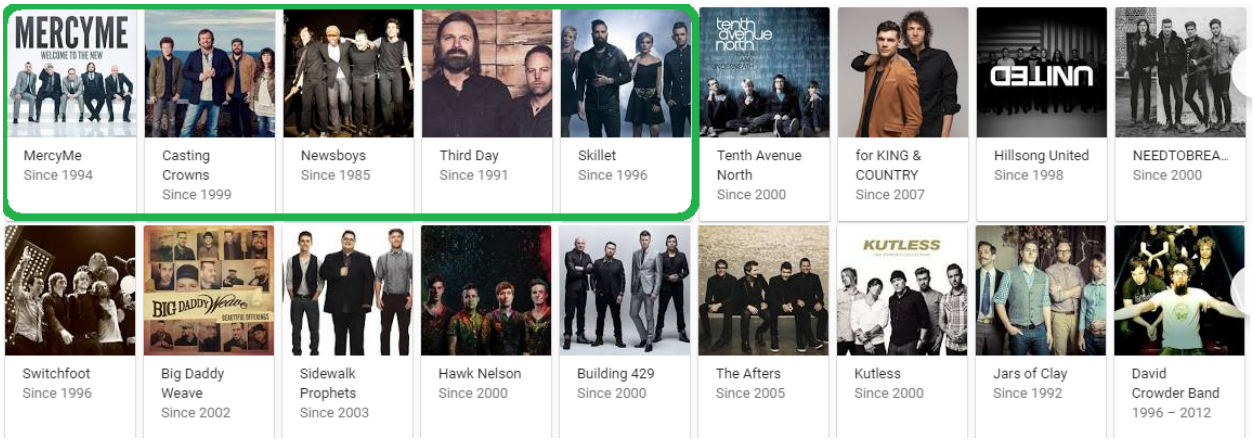

Figure 3:

If USER_1 likes the following artists:

userID	artistID	artist	listen_count
1	14042	MercyMe	5
1	7756	Casting Crowns	5
1	17875	Newsboys	5
1	7754	Third Day	5
1	371	Skillet	5

Then we suggest:

K	User	Rec_1	Rec_2	Rec_3	Rec_4	Rec_5	Rec_6
500	1	A Day to Remember	Low Shoulder	Switchfoot	Darin	The Fray	Marilyn
200	1	Adam Lambert	Marilyn Manson	Rammstein	The Ting Tings	BjÃ¶rk	Madonn
100	1	Katy Perry	Madonna	Marilyn Manson	Eminem	Rammstein	Beyonc
50	1	Marilyn Manson	Three Days Grace	Stone Sour	Trivium	Green Day	Good C
10	1	The Beatles	Nirvana	Muse	Red Hot Chili Peppers	Queen	AC/DC

SVD: google “top old school rap bands”

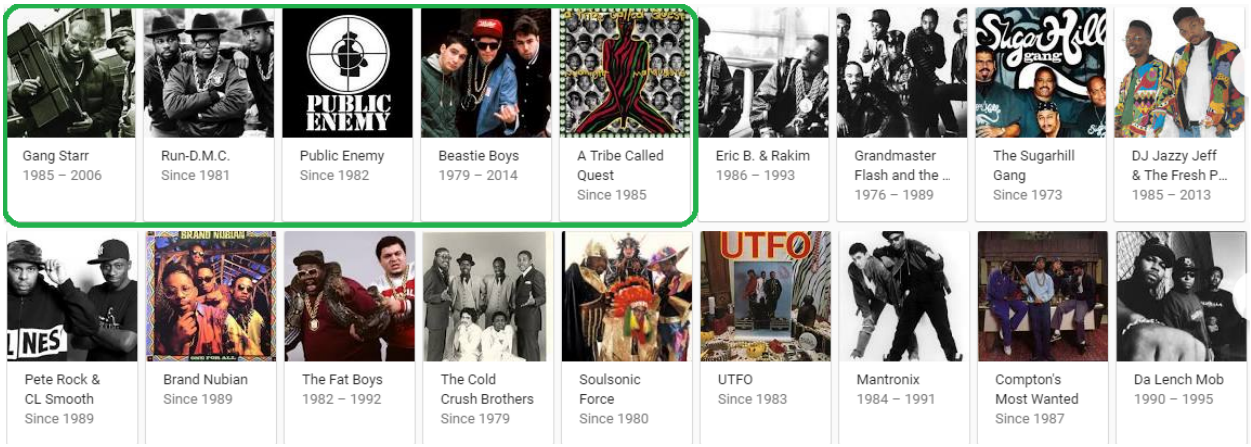

Figure 4:

If USER_1 likes the following artists:

userID	artistID	artist	listen_count
1	276	Gang Starr	5
1	9821	Run-D.M.C.	5
1	2698	Public Enemy	5
1	605	Beastie Boys	5
1	619	A Tribe Called Quest	5

Then we suggest:

K	User	Rec_1	Rec_2	Rec_3	Rec_4	Rec_5	Rec_6	
500	1	Big Punisher	2Pac	Mobb Deep	Cannibal Ox	OutKast	Common	
200	1	2Pac	Wu-Tang Clan	The Killers	Big Punisher	Method Man	Nirvana	
100	1	The Killers	Nirvana	Kanye West	Eminem	Nine Inch Nails	Wu-Tang Clan	
50	1	Eminem	The Smiths	Elliott Smith	The Killers	Kanye West	Lupe Fiasco	
10	1	The Beatles	Nirvana	Muse	Red Hot Chili Peppers	Queen	AC/DC	

SVD: google “top folk bands”

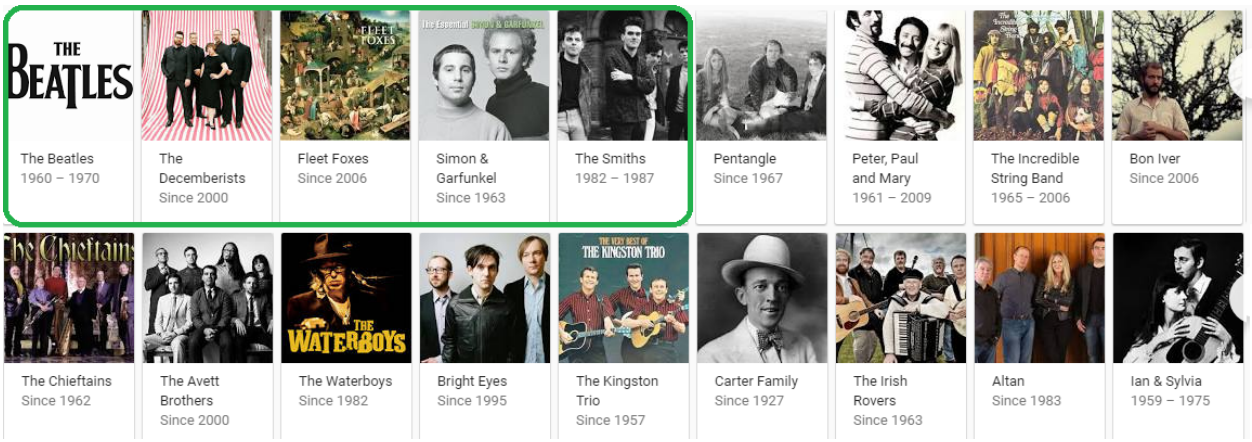

Figure 5:

If USER_1 likes the following artists:

userID	artistID	artist	listen_count
1	227	The Beatles	5
1	211	The Decemberists	5
1	1633	Fleet Foxes	5
1	2265	Simon & Garfunkel	5
1	429	The Smiths	5

Then we suggest:

K	User	Rec_1	Rec_2	Rec_3	Rec_4	Rec_5	Rec_6
500	1	Blondie	Modest Mouse	Pavement	Antony and the Johnsons	Metric	Ride
200	1	Modest Mouse	The Clash	Blondie	Pixies	Pavement	Pulp
100	1	The Clash	Pixies	Modest Mouse	Bruce Springsteen	The White Stripes	Simple
50	1	Muse	The Killers	Arctic Monkeys	RBD	The White Stripes	AnahÃ
10	1	Coldplay	Oasis	Muse	The Killers	Radiohead	Depeche

SVD: google “top funk bands”

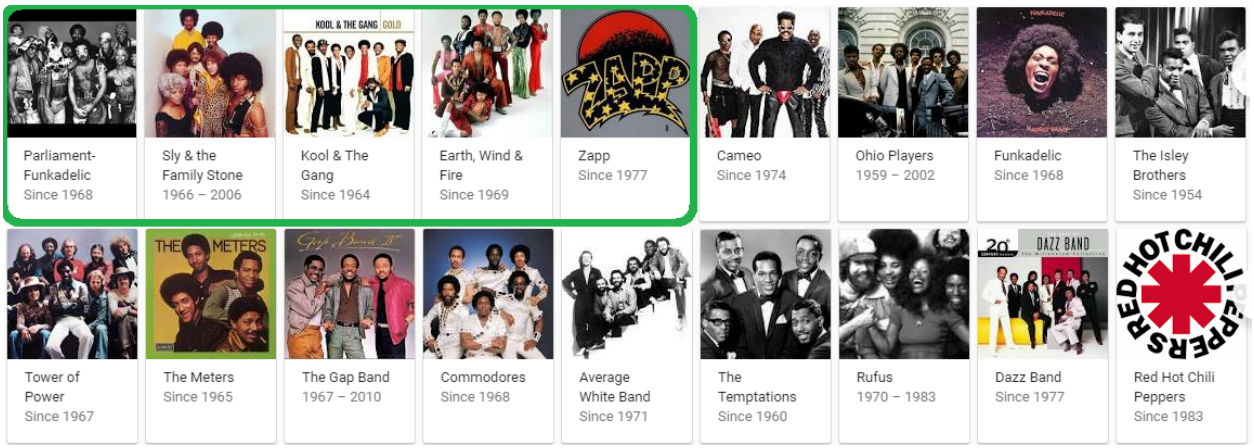

Figure 6:

If USER_1 likes the following artists:

userID	artistID	artist	listen_count
1	3776	Funkadelic	5
1	3778	Sly & The Family Stone	5
1	8998	Kool & The Gang	5
1	987	Earth, Wind & Fire	5
1	8993	Cameo	5

Then we suggest:

K	User	Rec_1	Rec_2	Rec_3	Rec_4	Rec_5	Rec_6
500	1	Madlib	Isaac Hayes	Onra	Gil Scott-Heron	Juanes	Reik
200	1	Garbage	t.A.T.u.	ZoÃ©	The Rolling Stones	Kasabian	Ellie Goul
100	1	Florence + the Machine	Maroon 5	Sufjan Stevens	Ellie Goulding	Out of Ether	Michael C
50	1	The Beatles	Aphex Twin	New Order	Green Day	Out of Ether	David Bow
10	1	The Beatles	Nirvana	Muse	Red Hot Chili Peppers	Queen	AC/DC

Conclusions

- The value of K makes an exponential difference in ***accuracy***
- The value of K makes an exponential difference in ***performance***
- Matrix Factorization allows for analysis of a much larger data set than without
- Run-time was not an issue for smaller K's, but unfortunately, as the predictions started getting better and better with larger K values, the performance again became a major issue - even though SVD was being used for dimensionality reduction
- In general, the large K's seemed to have some good suggestions, but not good enough to use to actually choose content. It's probably a good idea to be using an ensemble of strategies rather than just this one.