

A Movie Recommendation System: Using Content-based and Collaborative Information

¹Folord Suhas S, ²Naveen Ganesh, ³BELLO Mohamed, ⁴Laxmi B Rananavare

⁴Professor, ^{1,2,4}School of C&IT, REVA University

³Associate Professor, School of C&IT, REVA University

¹folordsuhas@gmail.com, ²ganeshnaveen9623@gmail.com, ³bellofad@gmail.com,

⁴laxmibranavare@reva.edu.in

Article Info

Volume 83

Page Number: 4280-4283

Publication Issue:

May - June 2020

Abstract

Since the late 20th century, the number of internet users has increased dramatically as has the number of web searches performed on a daily basis and the amount of information available to us. However, not all data that we get in search results are reliable or relevant which means that it may become more and more difficult to get satisfactory results from web searches. To solve this problem we use recommendation systems. Recommendation system shows us only the relevant information. Recommendation Engines can make various suggestions about artifacts to users. In our day-to-day lives, they may predict whether a user may like to buy a particular product online or if they are interested in a particular movie or are interested in listening to a particular song. To pick a movie, a user might search various websites to find a highly rated and well-reviewed movie which is very time-consuming. Here, we have created this recommendation system using the following methods. Content-based filtering takes keywords from the movie dataset and suggests it to a relevant user. Social based filtering takes reviews from multiple users and suggests it to another user. However, these methods do not use a significant amount of information available. This paper is an approach to a recommendation that is able to use both user ratings and other information available that will help in recommending movies to users. Our method uses these methods on a dataset containing more than 5000 different movies.

Index Terms: Movie recommendation system, Collaborative filtering, content-based filtering, Hybrid approach, Scalability

Article History

Article Received: 19 November 2019

Revised: 27 January 2020

Accepted: 24 February 2020

Publication: 12 May 2020

1. Introduction

Entertainment is very important these days, whether it is a busy day at work or just a bad day in general, entertainment such as movies, songs, etc. refresh us and get us ready for the next day. Movies play a huge part in entertaining us. There are many different types of movies such as kids movies, educational movies, comedy, animated movies, etc. Movies can be easily distinguished

by their genres. Another way to distinguish movies is by their year of release, language, cast information, etc.,

There are a number of ways in which we can search for a good movie. Recommendation systems have come a long way in this aspect. This system helps us save time and energy hence, they must be reliable and should be able to provide a movie that is exactly what we want or at least similar to what we want. In the last few years, with

represented as terms that usually occur in the document. User's profile is also described in the same way as analyzing the contents viewed previously by them.

C) Collaborative filtering

Collaborative filtering is a technique that can filter out items based on the likes of a similar group of users. Under collaborative filtering, we have,

It is an iterative supervised algorithm. Given a set of training examples, it divides them into one of the two given categories, shown in the figure below.



b) Ratings based filtering

This algorithm finds patterns that are consistent across the ratings of the other users. These patterns can be used on their own or in conjunction with social information to recommend content that a user may like.

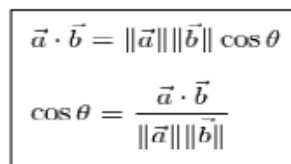
Year	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

Figure 3: Pivot table used by ratings based filtering

D) Cosine similarity measure

The “cosine similarity” measure for two vectors is calculated by taking a measure of the cosine of the angle between them. This checks for the orientation of the vectors, not their magnitude.

Formula for “cosine similarity”:



It is a criterion used to estimate how much identical objects (here movies) are to each other.

Euclidean's distance can measure how dissimilar users are to each other when they don't like similar movies, but cannot be implemented when users like similar movies. *Cosine Similarity* overcomes Euclidean's distance problem of measuring similar users when 2 users like the same movies and dislike the same movies.

privileged to have got all this support. We wish to acknowledge the help provided by our professors from the Computer science department of REVA University.

References

- [1] Base paper
reference:<https://ieeexplore.ieee.org/document/8058367/keywords#keywords>
- [2] Dataset reference:
<https://grouplens.org/datasets/movielens/>
- [3] Google News Personalization: Scalable Online Collaborative Filtering.
- [4] <https://www.kaggle.com/tmdb/tmdb-movie-metadata>
- [5] iopscience.iop.org/#
- [6] SajalHalder, A. M. Jehad Sarkar, Young-koo Lee “Movie recommendation system based on movie swarm” published in 2013 IEEE conference.
- [7] Ching-Seh Mike Wu, Deepti Garg, Unnathi Bhandary “Movie recommendation system using collaborative filtering” published in 2018.
- [8] “A Hybrid Approach using Collaborative filtering and Content-based Filtering for Recommender System” G Geetha, M Safa, C Fancy et al.
- [9] S.V.N. Vishwanathan, M. Narasimha Murty [2002], “SSVM: A Simple SVM Algorithm, Neural Networks, IJCNN '02. Proceedings of the International Joint Conference on Volume 3”
- [10] Eyrún A. Eyjólfsdóttir, Nan Li and GaurangiTilak [2008], “MovieGEN: A Movie Recommendation System”
- [11] ShiladSen, J. Ben Schafer, Jon Herlockerand DanFrankowski [2007] “Collaborative filtering recommender systems”
- [12] F. Kong, X. Sun, S. ye, [2005], “A comparison of several algorithms for collaborative filtering in startup stage”
- [13] J.L. Sanchez, F. Serradilla, E. Martinez, J. Bobadilla, [2008] “Choice of metrics used in collaborative filtering and their impact on recommender systems”
- [14] “Deep Matrix Factorization for Recommender Systems with Missing Data not at Random” Fei Zhang, Shiyu Peng and Jiaying Song.
- [15] Joseph Konstan, George Karypis, Badrul Sarwar, and John Riedl, (2001) “Item-Based Collaborative Filtering Recommendation Algorithms”.