Impact Factor (SJIF): 4.542

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

Volume 4, Issue 4, April-2017

RECOMMENDATION SYSTEM USING CORRELATION COEFFICIENT

Dhruti Pandya[1], Prof. Amita shah[2]

1 Computer Engineering Department, LDCE, Ahmadabad, Gujarat, India
2 Assistant Professor in Computer Engineering Department, LDCE, Ahmadabad, Gujarat, India

***Abstract* — Nowadays, over 200 million of people connected with the World Wide Web. These people used to accomplish their daily essentiality through internet. However, for each item explosion of option available which create difficulty for individual user to choose most appropriate one. The aim of proposed method is to make recommendation of movie from the movie dataset based on the reviews given by the other users in accurate and fastest manner for that sentiment analysis and cosine similarity measure is used. At last precision, recall, F-measure and accuracy will be calculated.**

***Keywords* ---- Knowledge discovery, Cosine similarity, Recommendation, User reviews, Sentiment analysis**

I. INTRODUCTION

Today's internet driven world over 200 million of people use World Wide Web to fulfill their daily needs. Explosion of option available in the web which cause problem to choose most appropriate from the available ones. Recommendation System help users to find and select items from the huge number available on the web or in electronic information resources [1][2][3]. On the other hand, customer and merchants are interacting with each other in E-commerce to discuss reviews of product. Reviews are very important to recommend items to the active user. Hence, in this paper combine approach of cosine measure and sentiment analysis is used for recommendation of movies to the user. Organization of paper is as follow: II Related Work III comparison study of different similarity measure IV proposed approach V Data collection VI Implementation and results VII Conclusion and VIII References.

II. RELATED WORK

In this section, we present the work that is done under the Recommendation area.

Hui Li, Jiangtao Cui BingqingShen and Jianfeng Ma [4] work on novel program recommendation system namely KBridge. In this they use Tweets to recommend items to the user. They use Lingpipe toolkit for sentiment analysis through this user-item rating matrix is obtained. After that min-max frequent pattern mining algorithm is used to extract frequent pattern.

Ramkumar Tyagi and Anand Jawedkar [5] develop a new recommendation system which recommend items to the user using user's web history. They also used browsing path clustering algorithm to replace user similarity matrix with path similarity matrix using web log recommendation list is generated.

Gabriel Marques, Ana Respício, Ana Paula Afonso [6] work on mobile recommendation system which recommend restaurants to the individual or group of user. They developed a system which is called BomApetite System for individual and group recommendation. They use client-server model to develop application in which client runs on mobile device with Android os was developed in Java programming language through the IDE android studio and server consist of web services, recommendation system and database.

III. COMPARATIVE STUDY

	cosine	Jaccard	sorenson
Equation difference	number of common attributes is divided by the total number of possible attributes	number of common attributes is divided by the number of attributes that exists in at least one of the two objects.	Sorenson's original formula was intended to be applied to presence/absence data, and is $s = \frac{2N_t}{N_x + N_y}$

Application and Usage	identify plagiarism & used in the context of text mining for compare-ing documents or emails.	identify mirror sites & used for two types of binary cases: Symmetric and asymmetric	used in image segmentation
Other	whereas the cosine similarity has no such constructs	Jaccard Coefficient can be used as a dissimilarly or distance measure	

Table I comparison between different measure

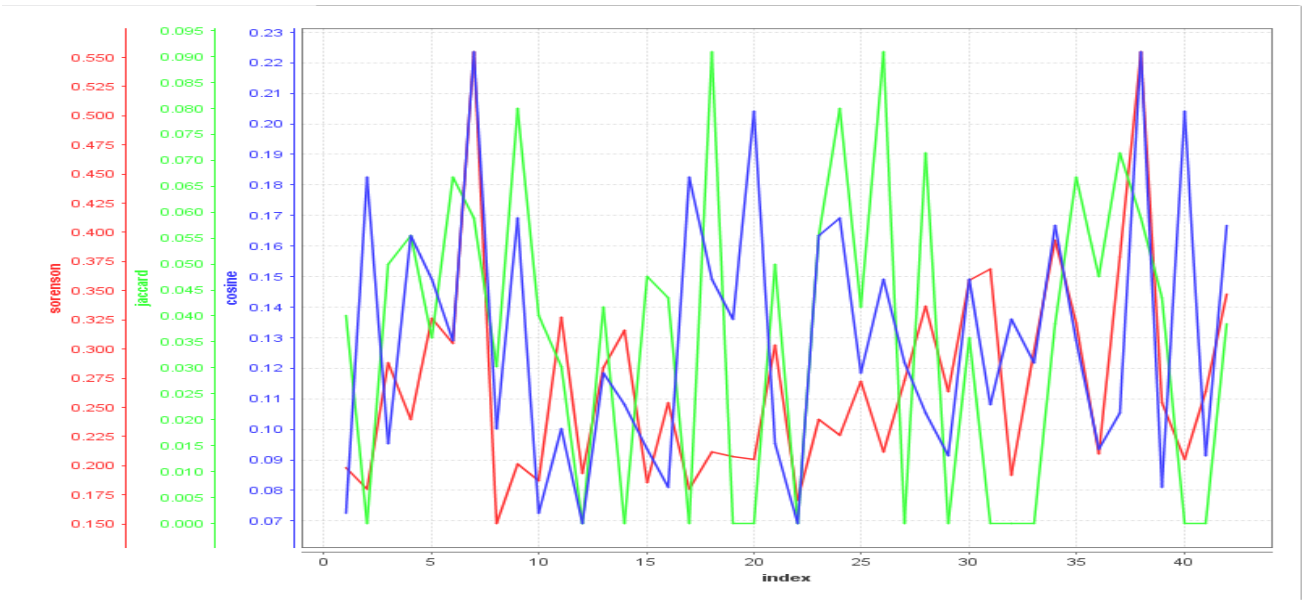


Fig. 1

IV. PROPOSED WORK

Following steps needs to be considered in our work:

Step 1:

Input to the system is Review Dataset. I have crawled reviews of movie using import.io API from rotten tomatoes. Dataset contains filed like movie_id, movie_name, user_id, review.

Step 2:
Preprocessing is done on review dataset.

- Splitting: Input format is paragraph of text. Splitting generates list of words.
  e.g.: [['this', 'is', 'a', 'sentence'], ['this', 'is', 'another', 'one']]
- Stemming: It converts word into its grammatical root form. Stemming technique converts word like “teach”, ”teacher”, ”teaching”, ”taught”, ”teaches” to root word “teach”.
- POS tagging: The Part-Of-Speech of a word is a linguistic category that is defined by its syntactic or morphological behavior. Noun, verb, adjective, adverb, pronoun, preposition, conjunction, and interjection are POS common categories.

Step 3:
Calculating Polarity for each review using Harward dictionary.

Step 4:
Group reviews according to positive, negative, neutral polarity

Step 5:
Finding cosine similarity if the active user's review is in positive polarity cluster then compare it with reviews within that cluster same as for negative and neutral review. Cosine gives value between 0 to 1 values. Values near to 0 indicate less similarity and near to 1 indicates high similarity.

Step 6:
Sorting similarity values in descending order.

Step 7:
Based on highest similarity N items chosen for recommendation.

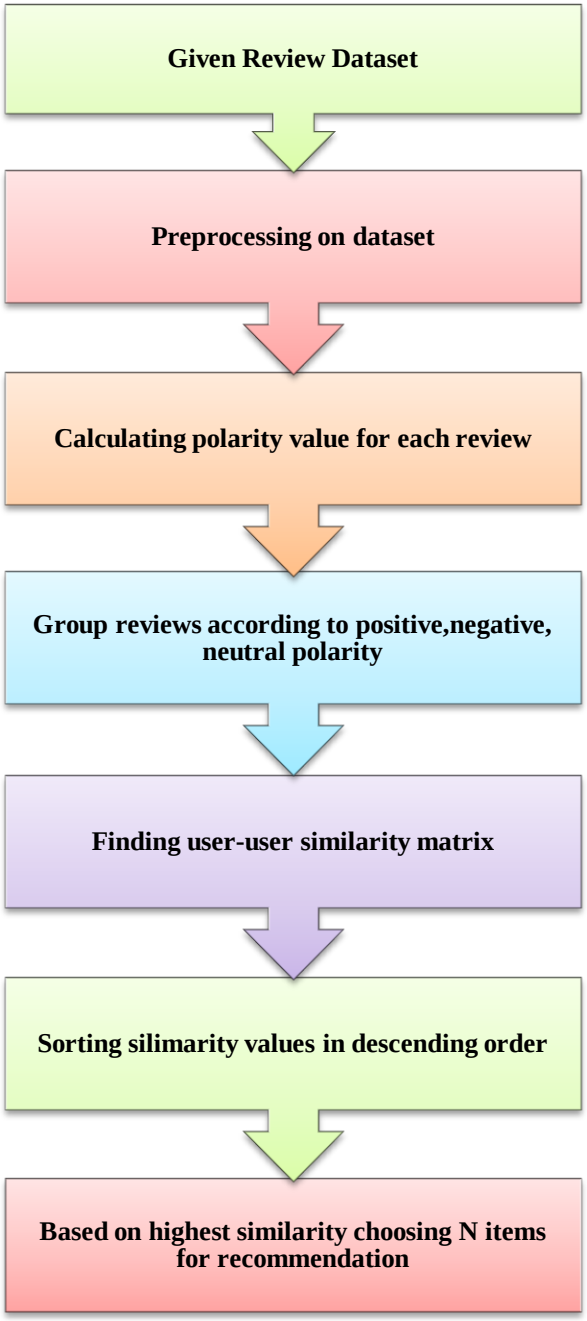
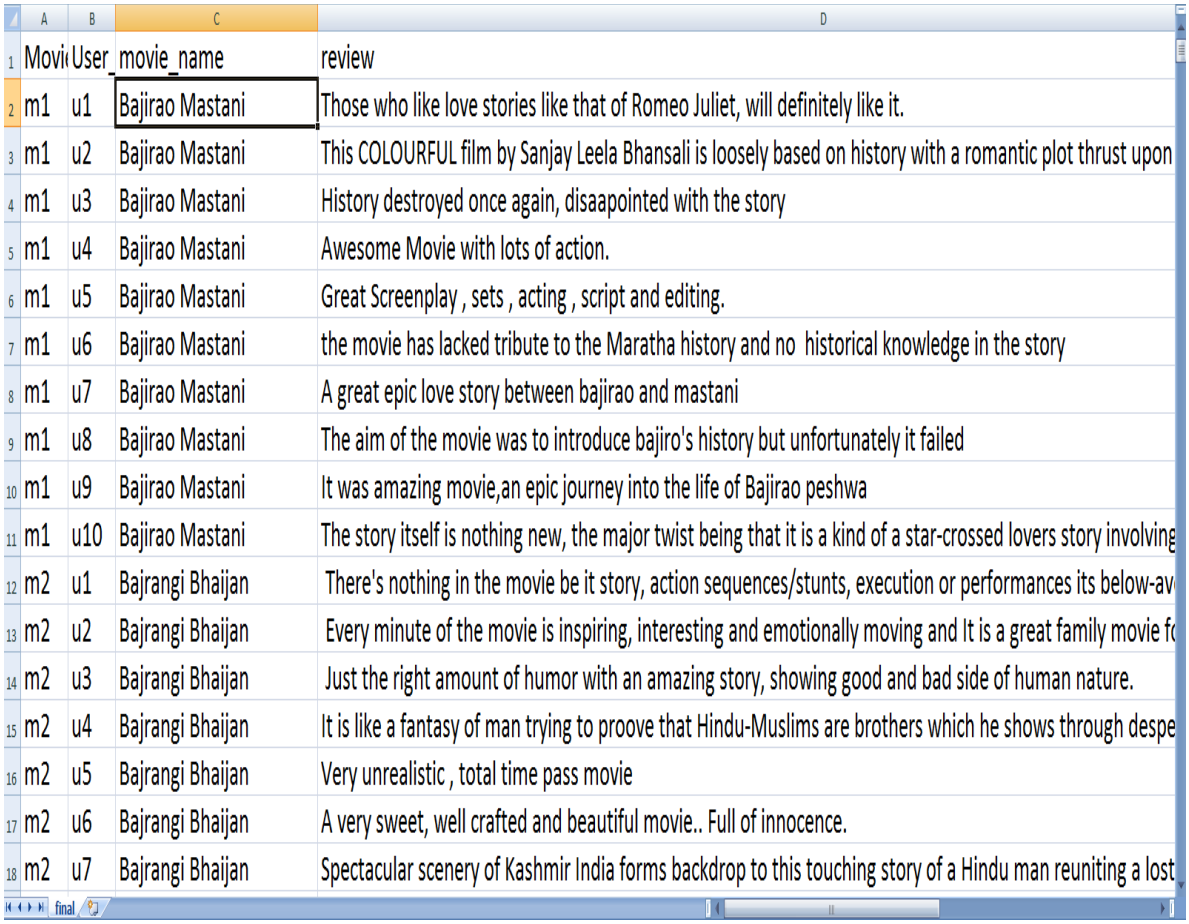


Fig-2

V. DATA COLLECTION

Reviews about different movies crawled from different websites like rottontomotoes, IMDB using import.io API. We have collected over 1000 reviews for more than 100 movies, average 8 reviews per movies.



	A	B	C	D
1	Movi	User_	movie_name	review
2	m1	u1	Bajirao Mastani	Those who like love stories like that of Romeo Juliet, will definitely like it.
3	m1	u2	Bajirao Mastani	This COLOURFUL film by Sanjay Leela Bhansali is loosely based on history with a romantic plot thrust upon
4	m1	u3	Bajirao Mastani	History destroyed once again, disaapointed with the story
5	m1	u4	Bajirao Mastani	Awesome Movie with lots of action.
6	m1	u5	Bajirao Mastani	Great Screenplay , sets , acting , script and editing.
7	m1	u6	Bajirao Mastani	the movie has lacked tribute to the Maratha history and no historical knowledge in the story
8	m1	u7	Bajirao Mastani	A great epic love story between bajirao and mastani
9	m1	u8	Bajirao Mastani	The aim of the movie was to introduce bajiro's history but unfortunately it failed
10	m1	u9	Bajirao Mastani	It was amazing movie,an epic journey into the life of Bajirao peshwa
11	m1	u10	Bajirao Mastani	The story itself is nothing new, the major twist being that it is a kind of a star-crossed lovers story involving
12	m2	u1	Bajrangi Bhaijan	There's nothing in the movie be it story, action sequences/stunts, execution or performances its below-av
13	m2	u2	Bajrangi Bhaijan	Every minute of the movie is inspiring, interesting and emotionally moving and It is a great family movie f
14	m2	u3	Bajrangi Bhaijan	Just the right amount of humor with an amazing story, showing good and bad side of human nature.
15	m2	u4	Bajrangi Bhaijan	It is like a fantasy of man trying to proove that Hindu-Muslims are brothers which he shows through despe
16	m2	u5	Bajrangi Bhaijan	Very unrealistic , total time pass movie
17	m2	u6	Bajrangi Bhaijan	A very sweet, well crafted and beautiful movie.. Full of innocence.
18	m2	u7	Bajrangi Bhaijan	Spectacular scenery of Kashmir India forms backdrop to this touching story of a Hindu man reuniting a lost

final

Fig. -3

VI. IMPLEMENTATION AND RESULT

We have used python and natural language toolkit for pre-processing, sentiment analysis and similarity finding. Desktop computers with 2.0 GHz Intel(R) Core(TM) 2 Duo processor and 2 GB RAM having Windows 8 (32 bit) Operating system is required for the working of above proposed work.

A. Result Analysis

In this section, we have discuss precision, recall, accuracy, F-measure based on contegency table.

Classified label	**Correct label**	
	Positive	Negative
Positive	TP(true positive)	FP(false positive)
Negative	FN(false negative)	TN(true negative)

Table 2 Contegency table

- **Accuracy [7]:** It is the proportion of correctly classified reviews to the total number of reviews.

  $$\textbf{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}$$

- **Precision [7]:** Precision is used to measure exactness. Precision is the number of review correctly ladled as positive on the total number of review that are classified as positive.

  $$\textbf{Precision}=\frac{TP}{TP+FP}$$

- **Recall [7]:** Recall is a measure of completeness. Recall is the no of review that correctly classified as positive divided by the total number of review that are positive.

  $$\textbf{Recall}=\frac{TP}{TP+FN}$$

- **F-Measure [7]:** F-Measure gives the score that balance between precision and recall. It combines them into one score for easier usage.

  $$\textbf{F-Measure}=\frac{2*Precision*Recall}{Precision+Recall}$$

Accuracy	71.28%
Precision	0.7019
Recall	0.7449
F-Measure	0.7228

Table 3 Performance Matrix

VII. CONCLUSION

This paper introduced combination of two concepts to recommend items to the active user. Using sentiment analysis technique we try to minimize execution time in similarity calculation.

VIII. REFERENCES

[1]Burke , R. , "The Wasabi personal shopper: A case-based recommender system"In Proceedings of the Sixteenth National Conference on Artificial Intelligence , AAAI Press, Orlando, FL, 1999 , 844-849.
[2] Rensnick , P. , and Varain , H. , "Recommender systems" Communication of the ACM, 1997, 40, 56-58.
[3]Burke ,R. ,Hammond ,K. ,Andyoung ,B. , "Knowledge-based navigation of complex information space" In Proceedings of the Thirteenth National Conference on Artificial Intelligence, AAAI Press, Portland, OR, 1996,462-468.
[4] Hui Li, Jiangtao Cui, Bingqing Shen, Jianfeng Ma, "An intelligent movie recommendation system through group-level sentiment analysis in microblogs", Elsevier, 2015.
[5] Ramkumar Tyagi , Aanand Jawedkar , "An Advanced recommendation system for E-commerce user" , CDAN ,2016.
[6] Gabriel Marques, Ana Respício, Ana Paula Afonso , " A Mobile recommendation system supporting group collaborative decision making" , Elsevier , 2016, 560-567.
[7] Farhan Hassan Khan , Usman Qmar, M.younus Javed , " Sentiview: A Visual sentiment analysis framework" i-society , 2014.