IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Book Recommendation System using Multiple User Opinion

Jaya Chauhan[1*], Pragya Shukla[2], Nilima Karankar[3]

[1] Dept. of IT, Institute of Engineering and Technology, Devi Ahilya Vishwavidhyalaya,Indore, India
[2] Dept. of CE, Institute of Engineering and Technology, Devi Ahilya Vishwavidhyalaya, Indore, India
[3] Dept. of IT, Institute of Engineering and Technology, Devi Ahilya Vishwavidhyalaya, Indore, India

[*]*Corresponding Author: jayasinghchauhan006@gmail.com, Tel.: +918602751001*

Available online at: www.ijcseonline.org

Accepted: 19/July/2018, Published: 31/July/2018

***Abstract*—** Popularly used recommendation for product is very important, also creating association among them serves with best recommendation. In this work FP Growth is used which calculates scores for final recommendation. Different approaches are used for dealing with different keywords to suggest recommendation. Efficient recommendation can be provided by creating relationships between associated books. Recommendation system recommend items and products depending on user's interest and preferences. The proposed recommendation system for books is based on calculating scores and frequency by using opinion mining and sentimental analysis.

Keywords- Opinion Mining, FP Growth, Recommendation System, Sentimental Analysis.

I. INTRODUCTION

Recommendation system is most importantly used for the purpose of better suggestion. User interest is must here for the purpose of recommendation. Recommendation is done in various fields like recommending books, news, products, friends etc. The work describes recommending books. Book Recommendation is implemented here on the basis of reviews and ratings(feedback and comment) of user. Then after preferred information is stored in user profile fro the purpose of future recommendation. Book details are stored on the basis of searched books and purchased books. Previous transaction details are also checked and display interested books on the basis of user's interest. List of Books can be searched through content-based filtering because it uses content of books and ratings. Recommendation system also evaluates system for quality and rating dependent. Efficient recommendation system can be provided by creating relationship between associated books. It is helpful system and as it is for books, then it is helpful for students and academics.

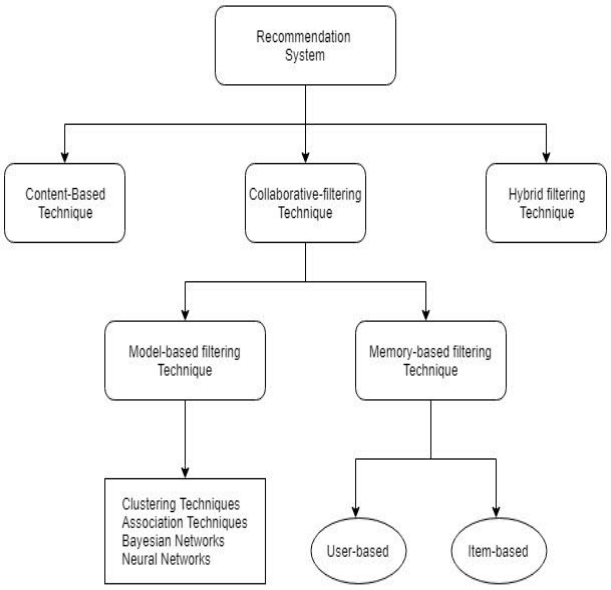

Figure 1: Book Recommendation System

Section I contain the introduction of Book Recommendation System, Section II contains the related work of various authors in book recommendation, Section III contains some measures of proposed methodology, Section IV contain the architecture and essential steps in recommendation, Section V describes the results and discussion, and Section VI concludes research work with future directions.

II. RELATED WORK

A.K. Singh et al. in [1] described about frequent itemset occurrence using association rule. High counts are important in database because of decrease in size of database. Some of the rules are not compulsory and they can be removed and these can minimize the database with generating new norms and are called as attractive count.

S. Rao. Et al. in [2] proposed several kinds of approaches with using association rules and consuming memory and different set of rules. Traditional algorithm for Apriori comes up with some issues and these issues should be handled using data mining and association rule. Iteration issues can also be handled using association rule with effective mining by extracting patterns.

O.R. Zane et al. in [3] introduces about the improvement of materials because of the requirement of learner. These types of issues can be classified with solutions over the internet.

R. Srikant et al. in[4] proposed Apriori hybrid algorithm which overcome issue of traditional Apriori algorithm. Dependency of this algorithm is on the basis of product availability in database and transaction size. This algorithm deals with some essential features. Traditional algorithm is also given with equal priority to deal with available items in database.

P. Devika et al. in[5] proposed an approach using FP Intersect algorithm. Here, user comments and feedback are the opinions on which recommendation is based. Ratings are extracted here on the basis of reviews and comments. Polarity is defined on the basis of comments and extracted ratings with which score is calculated. Apriori algorithm with association rule is generated to overcome the limitations of work which decreases no. of scans.

The study concludes that there is a strong requirement to develop popularity and collection of significant information tools on the basis of user recommendation. Existing work considers only small datasets which is the issue faced in related work. Book ratings are also not considered which arises the issue of not proper recommendation. They also haven't considered users preferences and interest and only aimed sentiment analysis for the purpose of recommendation.
Traditional approach does not work better with this huge data due to facing high latency issue because no. of scanning in large database is high. Modification of existing work is considered in this approach with accompanying limitations and evaluating solutions of the arrived issue.

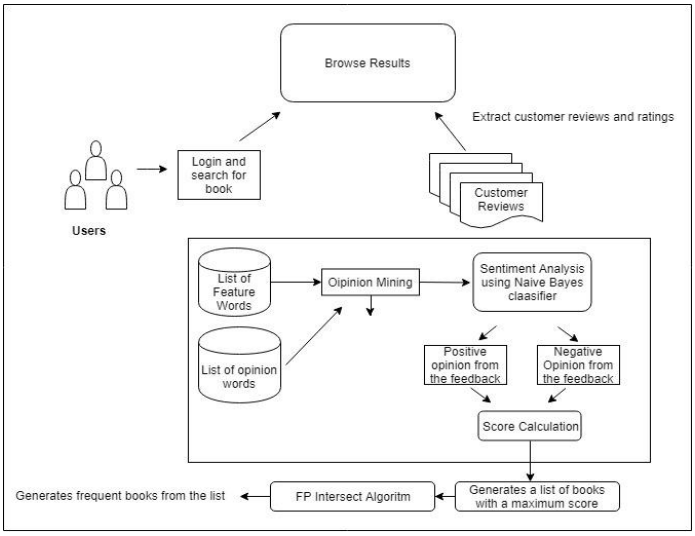

FIGURE 2 : EXISTING WORK[1]

III. METHODOLOGY

Proposed solution works on described methodology and analyzing it depending on the interest basis of user.

- Gathering information from Amazon dataset and the collected information are book datasets consisting of product id, its review and name of the book. 5MB - 5GB datasets are collected.
- Descriptive content are extracted from reviews with using sentimental analysis.
- Solution proposed classifies the problem of existing work that is the book rating which is not considered in existing work. Proposed work is implemented on Big data with using function of MapReduce on the premises of review and rating with the starting of work.
- In future it can be explored more by combining sentimental analysis with content opinion for getting better recommendation.
- Polarity is checked and analyzed using ambiguous word with positive, negative and neutral reviews of the context of user.
- Books are categorized and analyzed on the basis of interest with increasing accuracy in an effective way. Our work is calculating score with the modification to evaluate enhancement in accuracy by comparing proposed work with existing work.
- This work also calculates frequency of items based on the iems reviwed by the user.

IV. METHODOLOGY

System architecture explains that books are identified on the basis of review and rating of user with the preferences and interest of user to increase popularity of product. A

significant aspect with similarity of word in an documents is based on relevancy of word. Similar words which are mostly used, read and preferred by user are the most frequently appeared words which are used.

Complete work is concluded using Book Recommendation System with proposed steps and is cited below:

Step 1: Data is collected from Amazon Dataset on the basis of rating and ranking of product and also collected on the basis of user interest.

Step 2: Data cleaning is performed in review based text sentences which involves cleaning up of incomplete transactions (transaction in which rating is not mentioned properly and reviews are half generated). Also a list is prepared based on the no. of items a particular user has reviewed.

Table 1

U1	B1 B4 B6
U2	B2 B7
U3	B3 B5 B8

Step 3: At Lemmatization:

- Based on the review text obtained from data cleaning phase now review text sentences is passed to lemmatizer.
- Now the lemmatizer will split the given text into tokens and return the output in the following form:("annotators", tokenize, ssplit, pos, lemma).
- Now annotation class will annotate each text sentence by sentence and fetch tokens (words). Stop will be removed from the list of lemmas and final output will be list of lemmatized words.

Step 4: At Sentiwordnet:

- Now the list of lemmatized words is passed to sentiwordnet for score calculation.It will iterate each lemmas word by word and trim the spaces by # and add wordtypemarker(pos).Now it will calculate synset score as

synset score as score = PosS – NegS .....(1)

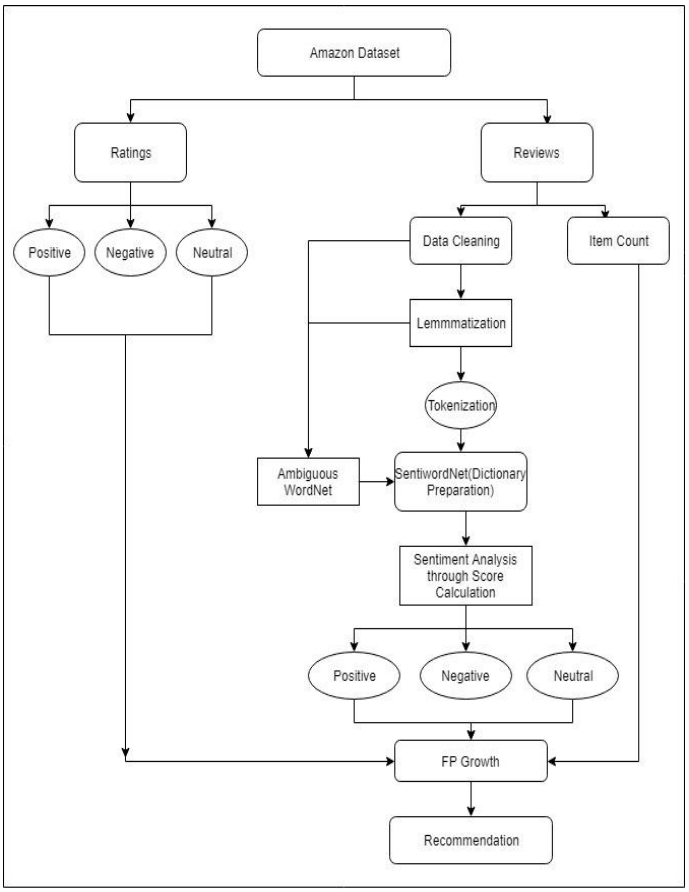

Figure 3: System Architecture

Eg: able:- positive value=0.125 negative value=0 (values are obtained from sentiwordnet.txt).

able = 0.125-0

able = 0.125(Positive)

- Now it will split the term based on hash value and add wodtypemarker and sysnset rank in the form

synTerm = synTermAndRank + "#" + wordTypeMarker;

.....(2)

Eg: able = 1+#+a, therefore able1#a

- Temporary dictionary will mainatin the record in the form: synterm,synTermRank, synsetScore.

Eg: (able,1,0.125)

- Average weight calculation is done if more than two word of same type occurs in the lemmatized list.

Eg: able {1#0.125}
{2#0.386}
{3#0.54}

Score = 1/2*First + 1/3*Second + 1/4*third... (3)

Score = 1/1*0.125+1/2*0.386+1/3*0.54 = 0.498

And calculates the sum as

Sum = 1/1 + 1/2 + 1/3.....
.....(4)
Sum=1/1+1/2+1/3 = 1.84
Therefore average score will be
score = score / sum;
.....(5)
score = 0498/1.84 = 0.270

- Now the final output that we obtained is in the form dictionary.put(word, score) i.e. dicitonary(able,1.25).

Step5: At Ambiguity Wordnet:

- Ambiguity is checked for lemmmatized wods that creates ambiguity in the text sentences. The words having same score and same wordtypemarker creates ambiguity.
- Input to the ambiguity word list list (synTermAndRank + "#" + wordTypeMarker, Score), The Ambiguity wordnet will obtain corresponding review text and perform ambiguity check. It pairs the words in the text phrase and determines the score of each text phrase and words having highest value is taken into account. Final score of words is then returned to the sentiwordnet for further classification purpose.

Step 6: At Item Count:

- A table with transaction id's and item count is generated in item count phase. In this how many no. of items have occurred in the particular transaction is prepared in the form of table.

Step 7: At FP Growth:

- At last frequent pattern mining algorithm is applied on the table generated during itemcount phase. Consider FP only for the items reviewed by the user. The algorithm then extracts frequent patterns from the item list based on the user preferences and extracts corresponding associations and related confidence value.
- At computational mapper items are separated by their #value and related associativity is also generated.
- At computational reducer the list that has has been separated by # are assigned a value (Table 2). Now Frequent pattern algorithm finds the frequency of related and associated items and put them in the form from parent to child nodes(Table 3).
- Now the list has been Separated bye tab space and the frequency is prefixed with ; and 0. Confidence is now calculated and final list of books with related associativity is displayed.

Table 2

B1 B2 B3	4
B1 B2	10
B2 B3	8
B1 B3	7

Table 3

B1 B2 B3	4	B1 B2 B3 0 ;4
B1 B2	10	B1 B2 0;10
B2 B3	8	B2 B3 0;8
B1 B3	7	B1 B3 0;7

B1 B2 B3 = 4 BI B2 =10 B2 B3 = 8 B1 B3 = 7

(For B1 B2) Confidence = B1 B2 B3/ B1 B2

Confidence = 4/10 = 0.4 (6)
(For B2 B3) Confidence = B1 B2 B3/ B2 B3

Confidence = 4/8 = 0.2 (7)
(For B1 B3) Confidence = B1 B2 B3/ B1 B3
Confidence = 4/7 = 0.57 (8)

Step 8: At last displays final recommendation.

```
Input: FP-tree, using DB and a mininum
support threshold S.
Output: The complete set of frequent
patterns.

FPT_pattern_generation(fptree T, set
fpset) {
    IF T contains a single path P {
       FOR each combination (denoted as
b) of nodes in path P
           generate pattern b U fpset
with support =
                        minimum support
of nodes in b; }
    ELSE FOR (cilk_for) each a in the
head of T {
        generate pattern b = a U fpset
```


V. RESULTS AND DISCUSSION

Result Analysis of the complete work is described in this phase with evaluating sentiments either: positive or negative. On the basis of computation time result is evaluated.

Table 4: Existing Work

File Size (MB)	Existing Work (Single Node) in seconds
8.6	419
41	1720
99	4120
213	11369
519	22905

File Size (MB)	Existing Work (Single Node) in seconds
8.6	419
41	1720
99	4120
213	11369
519	22905

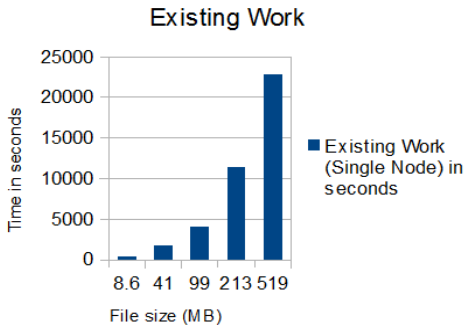

Figure 4: Graph representing Existing Work

Table 5: Proposed Work

File Size	Proposed Work (Single Node) in seconds
84.9 KB	2
853.3 KB	8
10.6 MB	62
112.9 MB	98
1.2 GB	3600

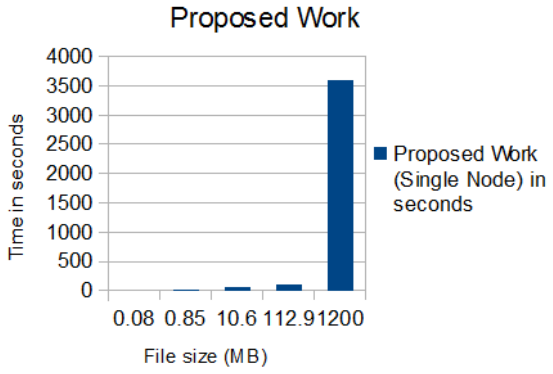

Figure 5: Graph representing Proposed Work

VI. CONCLUSION AND FUTURE WORK

Frequency pattern intersect algorithm is used in this approach. Positive, negative and neutral contextual rating and reviews are used for recommendation. Existing work is modified in our approach on the basis of user interest and frequency of appearance of content. Sentiments of user are explained on the basis of positive, negative and neutral opinion (review and rating). Performance is evaluated for increased accuracy. The Future work that can be performed is Proposed solution can also be evaluated for different category dataset i.e. Electronics. Proposed solution can also evaluate on multi node cluster.

VII. REFERENCES

[1] P Devika, R C Jisha and G P Sajeev, "A Novel Approach for Book Recommendation Systems," Published in International Conference on Computational Intelligence and Computing Research, 2016 IEEE.

[2] A.K. Singh, A. Kumar, and A. K. Maurya, "Association rule mining for web usage data to improve websites," in Advances in Engineering and Technology Research (ICAETR), 2014 International Conference on. IEEE, 2014, pp. 1–6.

[3] S. Rao and P. Gupta, "Implementing improved algorithm over apriori data mining association rule algorithm 1," 2012.

[4] O. R. Za ıane, "Building a recommender agent for e-learning systems," in Computers in Education, 2002. Proceedings. International Conference on. IEEE, 2002, pp. 55–59.

[5] R. Agrawal, R. Srikant et al., "Fast algorithms for mining association rules," in Proc. 20th int. conf. very large data bases, VLDB, vol. 1215, 1994, pp. 487–499.

[6] P. Nagarnaik and A. Thomas, "Survey on recommendation system methods," in Electronics and Communication Systems (ICECS), 2015 2nd International Conference on. IEEE, 2015, pp. 1496–1501.

[7] J. Yang, Z. Li, W. Xiang, and L. Xiao, "An improved apriori algorithm based on features," in Computational Intelligence and Security (CIS), 2013 9th International Conference on. IEEE, 2013, pp. 125–128.
[8] Y. W. Lo and V. Potdar, "A review of opinion mining and sentiment classification framework in social networks," in 2009 3rd IEEE International Conference on Digital Ecosystems and Technologies. Ieee, 2009, pp. 396–401.
[9] P. Jomsri, "Book recommendation system for digital library based on user profiles by using association rule," in Innovative Computing Technology (INTECH), 2014 Fourth International Conference on. IEEE, 2014, pp. 130–134.
[10] M. Al-Maolegi and B. Arkok, "An improved apriori algorithm for association rules," arXiv preprint arXiv:1403.3948, 2014.
[11] International Journal of Scientific Research in Computer Sciences and Engineering (ISSN: 2320-7639).

Authors Profile

Jaya Chauhan was born in Madhya Pradesh in 1994. She graduated in Information Technology from Rajiv Gandhi Proudyogiki Vishwavidyalaya, University of Bhopal, Madhya Pradesh, India in 2016. She is currently pursuing Master of Engineering from Institute of Engineering and Technology, DAVV INDORE.

Dr. Pragya Shukla pursed BE in Computer Science from Govt Engg. College Bhopal in 1996, ME from Devi Ahilya Vishwavidhyalaya in 2004 and Ph.D in Computer Engineering Department, IET, DAVV, Indore in 2013. She is a member of IEEE and ACM since 2011. She is life member of CSI, India. She has published more than 25 research papers in various reputed Journals and conferences. Her main research work focuses on Artificial Intelligence, Big Data Analytics, Data Mining, Soft Computing. She has 17 years of teaching experience.

Nilima Karankar pursed Bachelor of Engineering from Mandsaur Institute of Engineering, Mandsaur in year 2006 and Master of Engineering from Institute of Engineering and Technology, DAVV INDORE in 2015. She is currently working as Assistant Professor in Department of Computer Engineering, Institute of Engineering and Technology, DAVV INDORE since 2006. She is a member of ACM since 2011. She has published 7 research papers in reputed international journals and it's also available online. Her main research work focuses on Cryptography Algorithms, Network Security, Cloud Security and Privacy,Artificial Intelligence based education. She has 11 years of teaching experience.

