International Journal of Advanced Technology in Engineering and Science www.ijates.com
Volume No.02, Issue No. 12, December 2014 ISSN (online): 2348 – 7550

A FRAMEWORK FOR IMPROVING ENDORSEMENTS BASED ON RANKING FUNCTIONS

D Meena[1], T Lavanya[2]

[1] Pursuing M.Tech (CSE) , [2] Assistant Professor , Nalanda Institute Of Technology(NIT), Siddharth Nagar, Guntur , Andhra Pradesh, (India)

ABSTRACT

Endorsement Systems have grown rapidly these days and became more important to individual users and organizations because of the customized endorsements. This paper helps to know more about endorsements provided based on the ranking function that can be utilized matching to the planned scenario. With this approach we can suggest the way to enhance the endorsement technique based on the market needs. The ranking based approach employed here is to give recommendations that will have substantially higher cumulative diversity across all other users.

Keywords: Endorsement Systems, Customized Endorsements, Ranking Function, Cumulative Diversity.

I INTRODUCTION

This paper purely deals with data mining architecture. Data mining is a software or tool with which a user can fetch the data from a very large set of database. Consider, searching a book named DEF from room which has some thousands of books inside; no doubt the book can be traced from that room by consuming some time but the important factor is time, tracing it in the less time such that time can be utilized in some other works rather than searching the same book. To do that in efficient and smart manner we need to employ some logic with which we can get the best results. Similar kind of concept can be applied in the computer technology i.e. searching an item with the best algorithm suited that will save the user time. Apart from searching it would be better if the ranking mechanism is also employed with it.

Ranking functionality helps the user to get the data that is mostly viewed by the users. It actually saves the user time during the search process as the data is kept separate which is most frequently utilized by many users.

Data Mining is as we all know that fetching the data after creating the data sets. Data sets are those which are created based on the user requests and based on the accessing.

Data Mining is a process which involves many small manipulations that gives out an effective result and this technology is fastest growing because of the growth in the technology and the user stipulated time spent on data accessing.

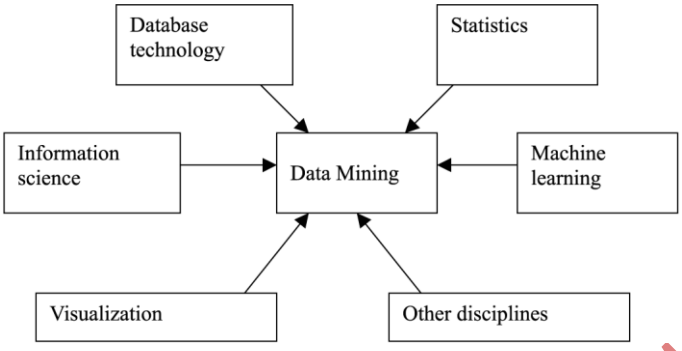

Fig 1: Data Mining Process

Now we need to know about the Ranking architecture and how it is useful in real time applications. Rank is given to those data which is viewed more number of times by users. In search type applications, two concepts i.e. ranking function and data mining play a vital role for giving the best output to the user as per their expectation. When a user searches for a specific data, it is then mined and from next time if the request comes for the same data based on the rank assigned data is fetched and given to the user. We can manipulate the result as per our requirement. Cumulative Diversity refers to gathering of the information from different units and displaying it in one place which will be helpful for further calculations. This process when it is done using the endorsement into consideration then the output will be even more perfect i.e. as per the expectations of the user. Process being implemented here in this paper is; we are first going to gather the information based on the endorsements and then as per the ranking assigned to it we will display it to the user from where the content can be easily taken.

This paper primarily focuses on the content available in the search engine related with the user requested search. Consider the example, in the present system when a user is searching for the data we get number of links linking to that data but there is no assurance that the data will be available in the first link, users need to search from the first link and if the data not available then the second link would be cross checked and the process goes on till the user achieves required data. In this example we consider that there are totally 17 links displayed for the user request but from which user has used first 5 links and signed out from that application. In this manner if some 10 users visit and take the data from this search engine, for an instance assume only initial links would be checked by users and not the below links so this is a drawback in the existing system because when user is not reaching all the links he/she will not get an idea about the data present in that link. The solution to this problem is designed in this proposed paper with efficient algorithms to give good results to users in less time and also utilize all the links for possible outcomes.

II BACKGROUND

Here in this section, we are going to discuss about the work analysis that helps to understand the process of Endorsement Diversity. Recommendation is basically classified into three categories based on the approach of recommendation i.e. Collaborative, Content based and Hybrid based. Content based technique refers to the items which were preferred by the users in the former analysis. Collaborative filtering refers to the search which is performed based on the data that was liked by neighbors i.e. past user, and Hybrid refers to combination of both content based and collaborative. Endorsements can also be classified based on the memory-based or and

model based approaches. Technique used most commonly for recommendation purpose is model based approach.

Model based approach involves two steps i.e. in the first step based on the request from the user it will diversify the content, and in the second while displaying to user it will take the ranking into consideration.

The first step involves the heuristic process i.e. the previous data which was liked by the neighbor will be displayed considering the ranks and in the second step for displaying the data which is not viewed by users and which does not have any ranking. With the implementation of the above two steps we can achieve an application that will give output to the user from where effective data can be accessed in short span of time.

We now need to know the similarity between the users for a particular item and it can be calculated using the below equation,

$$\text{sim}(u,u') = \frac{\sum_{i \in I_{(u,u')}} R(u,i) - R(u'-i)}{\sqrt{\sum_{i \in I(u,u')} R(u,i)^2} \sqrt{\sum_{i \in I(u,u')} R(u',i)^2}}$$

In the above equation, u is the actual user, u' refers to other users, R(u,i) is the rating given by the user to an item and similarly R(u',i) is rating given by other users to that item.

R*(u,i) is calculated as adjusted weighted sum of all known ratings R(u',i). The equation used for the manipulation is,

$$R^*(u,i) = R(u) + \frac{\sum_{u' \in N(u)} sim(u,u').(R(u',i) - \overline{R(u')})}{\sum_{u' \in N(u)} |sim(u,u')|}$$

The above mentioned process is for getting the diversity for the endorsements but we may not get the effective output as the ranking function is not associated with it and to do so we need to include the ranking algorithm for assigning the rank.

2.1 Standard Ranking Approach

As already mentioned that without the ranking function diversity does not have that noticeable significance because the user will check with the same data again and again this may give rise to a situation where user searching for data from the link and that link in actual hardly contains any data regarding the request. For the above problem when we include ranking functionality then based upon the user request we will display the content which has got more number of visits by the data users and this can in turn be treated for the ranking purpose.

The logic behind the rank to the particular link when selected by user is, increment the counter and then assigning it to a variable and this value should get incremented as and when the users checking this link. With this process, a number keeps on incrementing for the link associated with it but we must know the accurate result i.e. a specific link visited by how many number of users and is calculated as below,

For a particular link L(i) there are n users who visited m times and the rating given by individual user is r(i). So the average rating for this link can be given as

$$\mathbf{R(i)} = \frac{sum(r(i))}{n}$$

With the above generated values, we get an average rating for a link and this type of rating is calculated for all other links which will be used by users and maintained in a database for further manipulations.

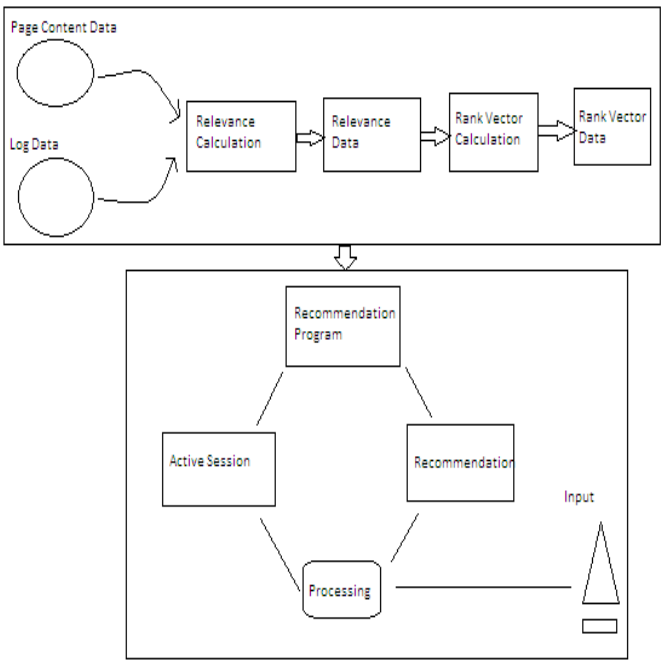

Fig 2: Architecture Flow

Once the data is ready, the time during displaying the links to the users, admin will take care i.e. to display the links that are useful to the end user and from which the user can extract the data for requested request. Output is based on the average ranking value obtained from the above calculation process and for processing we will use the sql query. This will check the condition on the rating column and will display the links satisfying the condition specified by admin.

With the above process we can say that, half of the work is completed i.e. displaying the links which have got some rating and if the process is stopped here then it means that there is no proper solution given to the proposed work. To overcome the problem of the existing system, admin will work on the data sets which do not have any rating given by users. Admin will process two different queries, first one is for the above mentioned process and second is to fetch the links from non rated data and final result will then display the values which is the combination of both the results generated by these two queries and will display to the user in the random manner. With this process we can assure that the data present in the database will be used to provide information to users and also the client utilizing these results will be satisfied based on the content received. The main logic

behind displaying the random values is with taking the help of Random class i.e. available from the utility package.

Random r = new Random();

int i=r.nextInt(5);

In the above two steps written, it means that Random class from utility package is used and it is generating a random number between the range 0 and 5.

With the above class we can achieve effective output which itself is a new implementation for the existing problem. In designed application when a user is trying to fetch the information about a particular data there will be two sets of data that would be displayed to the user out of which first set contains the rated data and second set contains non-rated data.

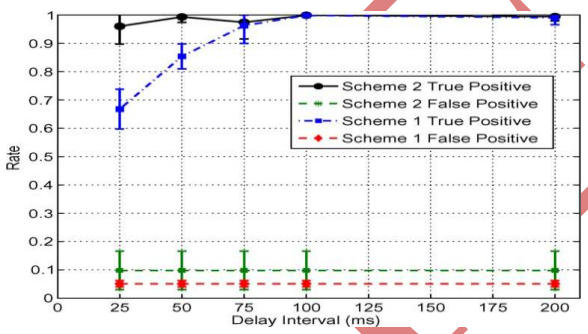

Fig 3- Detection rate versus delay interval

Both these outputs fetched from SQL will be merged and then displayed to the user, but the user will not know anything about the process running in the background.

Ranking function embedded in the application will need to work always in the background i.e. it can be called as Recommendation Re-Ranking. The process which takes place here is that all the ratings assigned previously will become secondary values and when the search action is again performed the rank is changed. This in turn will affect the ranking function and it automatically changes rating for that particular link. Accordingly the output which has to be shown to the end user changes but the implementation of the algorithm will not change.

III CONCLUSION

Ranking Recommendation Model is proposed in this paper which help users get their data with more effective data and in less time. User can save search time with the help of this application and can fetch large amount of data from various domains that are clubbed in a single application. In extension to the search algorithms we added the ranking algorithms that give flexibility to the developers as they can also work with other rating prediction algorithms. Our proposed work is designed with recommendation diversity using ranking based technique for the end users.

REFERENCES

[1] J. Delgado and N. Ishii, "Memory-Based Weighted-Majority Prediction for Recommender Systems,"
[2] D. Fleder and K. Hosanagar, "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity"
[3] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions,"
[4] R. M. Bell, Y. Koren, and C. Volinsky, "The Bellkor 2008 solution to the Netflix Prize.," J. Bennett, and S. Lanning, "The Netflix Prize,"
[5] D. Billsus and M. Pazzani, "Learning Collaborative Information Filters,"
[6] K. Bradley and B. Smyth, "Improving Recommendation Diversity,"
[7] S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering,"
[8] E. Brynjolfsson, Y J. Hu, and, D. Simester, "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales"
[9] M. Balabanovic and Y. Shoham, "Fab: Content-Based, Collaborative Recommendation,"
[10] R. Bell and Y. Koren, and C. Volinsky, "The BellKor solution to the Netflix Prize,"
[11] E. Brynjolfsson, Y. Hu, and M.D. Smith, "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers,"
[12] J. Carbonell and J. Goldstein, "The user of MMR, diversity-based re-ranking for reordering documents and producing summaries,"

AUTHORS PROFILE

Daida Meena is currently pursuing M.Tech in the Department of Computer Science & Engineering, from Nalanda Institute of Technology (NIT), siddharth Nagar, Kantepudi(V), Sattenapalli (M), Guntur (D), Andhra Pradesh , Affiliated to JNTU-KAKINADA.

Thota Lavanya (M.Tech) working as Assistant Professor at Nalanda Institute of Technology (NIT), siddharth Nagar, Kantepudi(V), Sattenapalli (M), Guntur (D), Andhra Pradesh , Affiliated to JNTU-KAKINADA.