Abstract

In the current world of web, social networking sites are becoming an inseparable part of the user’s life. The users almost every day on social networking sites are keen to find the similar users for their current requirements. So many of the systems recommend individuals based on their profile taste which is created during registration, but this actually fails to make individuals to meet their current requirements. To resolve this problem a system is needed that can fetch users current interest. The system is expected to be capable enough for reading all the content available on web page. After reading all the data, extracting the useful information and summarizing it in such a way so that it can help us for maintaining a list of user’s preferences. On the basis of such a list of user’s priority the system should recommend the other user of common interest. Already a lot of research work is available on the same topic. Many systems already exist that are performs the task of recommendation. This paper is a survey on such techniques that can help for the recommendations of the current data posted on the web. The techniques are like web crawler, fuzzy logic, different recommendation approaches like collaborative filtering approach and content based approach. The task is to analyze these techniques well and to conclude an approach that can give us the same task of recommendation with some better efficiency.
1. Introduction

Many of the proposed systems use predefined datasets to provide the recommendation; this actually yields a poor expectation on user’s requirement. Some other system uses datasets like Movie lens and Douban for their experiments which are indeed static [1]. So in a solution to that a system that can work in live data that is present on the web pages can be used. This kind of data can be used for the recommendation that will use the current data. For fetching up the data from the web pages an effective web crawler is needed that can extract the entire sub URLs that are present in the any given URL. For the same purpose different crawling algorithms are available. Many authors have introduced various different web crawlers that are using different crawling techniques. After the data received from web crawler some kind of recommendation approach will be applied to it. That approach can be any of the two i.e. collaborative filtering or content based filtering. Collaborative filtering approach of recommendation first analyze all the data regarding users interest and predicts the user likes on the basis of similarity of other users. Another approach is content based approach of recommendation system which is completely based on the users liking as well as preferences. In content based approach the recommendation is not at all based on any others liking.

Fuzzy logic is one more important technique. Fuzzy logic is nothing but a kind of uncertainty. This thing helps in removing the stop words, identifying the top words, stemming etc. [2] After that using fuzzy logic concept the statements can be weighted. And the highly weighted concept can be recommended. Many authors have already been worked on these techniques. In this paper a detailed survey there on what work had been done in each of different techniques by different authors. After the analysis of the work done yet on the approaches, an approach can be proposed that can have the overall performance of recommendation better than existing one.

2. Web Crawling

Web Crawler is such a program that creates a graph structure of web. It fetches the URL from any of the web page recursively [3]. The crawler starts its functioning from seed page and start fetching external links to fetch the rest of the web pages.

A. Frontier

Frontier is basically a to-do list which contains the URL of unvisited page by the crawler [3]. This maintains a queue of URL that system has not yet traversed.

![Flow of a sequential crawler](image_url)
B. Web Crawler Strategies

There are some basic algorithm that are used by web crawler [4]-

1) **Breadth First Search Algorithm**- This algorithm starts searching at the root node and afterwards searches on the same level all the neighboring nodes.

2) **Depth First Search Algorithm**- This algorithm starts searching at the root node and then traverses till child node first till depth.

3) **Page Rank Algorithm**- This algorithm counts the citations or back links to a given page.

4) **Genetic Algorithm**- By using fitness function find the fittest offspring that is obtained by crossing over of some best individuals.

5) **Naïve Bayes Classification Algorithm**- This is the algorithm of probabilistic learning and classification. It works on the assumption that one feature is independent of another.

Vladislav implements a distributed web crawler which runs smoothly on a network of workstations. This crawler can scale to hundreds of pages per second and also resilient against system crash [5]. Some special forum crawlers are also introduced that basically retrieve the content from the web pages with minimum overhead. The crawler i.e. Forum Crawler under Supervision (FoCUS) already used for same purpose [6]. For obtaining the complete friend list of user on online social network (OSN), an effective Facebook Crawler is introduced to overcome the shortcomings of crawlers that can get at most 400 friends [7].

![Figure 2.2: A model of multithreaded crawler](image-url)

The sequential crawlers take a lot of time for crawling process. For some span of time in sequential crawler either CPU is idle or network interface is idle. For overcome this shortcoming highly efficient multithreaded crawler [3] is used that gives comparatively better performance in the terms.
of speed. In this every thread runs and locks the frontier for picking up next URL. This operation can be implemented recursively for better performance.

C. COLLABORATIVE FILTERING

This is one of the approaches for design of recommender system. This approach basically analyses the large amount of information regarding users behavior and activities etc. and predicts the user likes on the basis of similarity of other users. Dang Cai introduced a very efficient way to reduce the work done by recommender system. To execute the Multiclass Co-Clustering (MCoC) and finding the effective solution by analyzing the correlated subgroups makes collaborative filtering more effective [8].

Some model considers that user’s similarity can be there in basically two ways that they are having similar taste or might be having similar attractiveness for the users who contact them. So the neighbor based collaborative filtering algorithm name as SocialCollab [9] is introduced for doing recommendations based on above two factors. Some systems also uses latent factor model for recommendation rather than collaborative filtering method. Predicting the latent factor from audio has already introduced [10].

D. CONTENT BASED FILTERING:

Another approach of designing recommender system is content based filtering. This method is completely based on user profile and preferences. Michael implemented the content based approached that is used in various domains like Television programs, restaurants, item for sale etc [11].

![Content Boosted Collaborative Filtering System Overview](image)

Figure 2.3: Content Boosted Collaborative Filtering System Overview

One more new approach i.e. Content-Boosted Collaborative Filtering is introduced (as shown in figure) that is basically a hybrid approach providing the advantages of both the approaches. It uses content-based predictor and also providing some personalized suggestion by using collaborative filtering [12].

E. FUZZY LOGIC

Fuzzy Logic is the logic of approximate reasoning as well as imprecision [13]. Fuzzy logic helps in Text Summarization which is basically having two approaches – abstraction and extraction. The approach of extraction for text summarization helps in sentence selection on the basis of sentence weighting [2]. The first step that it uses is identification of important features.
appropriate document removing of stop words, stemming, segmentation etc is needed. The basic aim of using fuzzy logic in text summarization is to improve the summary that is created by statistical method.

### Table 2.1: Comparison Of Different Technique

<table>
<thead>
<tr>
<th>Title</th>
<th>Author</th>
<th>Methodology</th>
<th>Drawback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personalized Recommendation Combining User Interest and Social Circle</td>
<td>Xueming Qian, Member, IEEE, He Feng, Guoshuai Zhao, Tao Mei, Senior Member, IEEE</td>
<td>Combining social network factors: personal interest, interpersonal interest similarity, and interpersonal influence for recommendation</td>
<td>Does not provide fresh recommendation</td>
</tr>
<tr>
<td>An Exploration of Improving Collaborative Recommender Systems via User-Item Subgroups</td>
<td>Bin Xu, Jiajun Bu, Chun Chen, Deng Cai</td>
<td>To formulate the Multiclass Co-Clustering (MCoC) and finding the effective solution by analyzing the correlated subgroups makes collaborative filtering more effective.</td>
<td>The method of finding the subgroups for recommendation is not at all efficient</td>
</tr>
<tr>
<td>Content-based Recommendation Systems</td>
<td>Michael J. Pazzani and Daniel Billsus</td>
<td>Uses content based approach that performs the recommendation on the basis of user’s profile.</td>
<td>Can’t give the proper recommendation if the proper information is not there to distinguish the user’s likes and dislikes.</td>
</tr>
<tr>
<td>Content-Boosted Collaborative Filtering for Improved Recommendations</td>
<td>Prem Melville, Raymond J. Mooney and Ramadass Nagarajan</td>
<td>Uses a content-based predictor which enhance existing user data, and provides personalized suggestions by using Collaborative Filtering.</td>
<td>Difficult to implement and performance improved by only 4%</td>
</tr>
</tbody>
</table>

### 3. Conclusions

This paper is the survey on various techniques that can be helpful for personalized recommendation. By the survey, the results can be drawn as recommendation of live data available on web page gives us more fresh and appropriate recommendation rather than the recommendation done on the previously available data set.

For getting fresh information content based filtering approach of recommendation system is comparatively better than collaborative filtering as user preferences changes time to time. So the upcoming systems can show the great improvement in the performance if it uses content based approach with the use of fuzzy logic and effective web crawler to work on live data that is present on web pages for recommendation.

### References

[1] Xueming Qian, He Feng, Guoshuai Zhao, Tao Mei “Personalized Recommendation Combining User Interest and Social Circle” IEEE Transactions Knowledge And Data Engineering Vol:26 No:7 Year 2014

Shruti Agrawal, V. D. Thombre: A Survey On Recommendations By Collecting The Users Interest From Social Networking Web Pages


[9] Xiongcai Cai, Michael Bain, Alfred Krzywicki, Wayne Wobcke, Yang Sok Kim, Paul Compton, and Ashesh Mahidadia “Collaborative Filtering for People to People Recommendation in Social Networks” School of Computer Science and Engineering, University of New South Wales, Sydney NSW 2052, Australia, 2010


