International Journal of Engineering & Technology, 7 (4.5) (2018) 386-388

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Implementation of Recommender System for Web pages

Anuradha G
Department of Computer Science Engineering
VRSEC, Vijayawada
**Corresponding author E-mail:*anuradhagovada@gmail.com

Abstract

In Web based applications, Recommender systems have become fundamental information access system. They effectively prune large information spaces which provide appropriate decision making and suggestions, so that the users are directed towards those web pages that best meet their needs, preferences and interests. In web-based context, this can be achieved by basic rough set model and collaborative filtering techniques in decision making of the web pages. The recommender system can be implemented based on two types of techniques which are content based filtering and collaborative filtering. Content based filtering constructs a recommendation on basis of user's behavior (historical browsing information) and collaborative filtering uses group knowledge to form a recommendation based on like users. Sparsity, the cold-start problem, fraud are main challenges in recommender system.
This paper proposes recommendations to the user that varies from one user to another user which is based on the user's profile. (For example, for a search with same keyword a student will be suggested differently compared to a scholar or a teacher).
The abstract should state the purpose, approach, results and conclusions of the work. The author should assume that the reader has some knowledge of the subject but has not read the paper. Thus, the abstract should be intelligible and complete in it-self (no numerical references); it should not cite figures, tables, or sections of the paper. The abstract should be written using third person instead of first person.

Keywords*: Collaborative filtering, Content filtering, Sparsity, The cold-start problem, Fraud*

1. Introduction

The main goal of a recommender system is to provide meaningful recommendations to users for products or items that might interest them. Suggestions for books on amazon, or movies on netflix, are real world examples of the operation of industry-strength recommender systems. The recommendations are designed based on particular characteristics of the data and domains. For example, movie watchers on netflix frequently provide ratings on a scale of 1 (disliked) to 5 (liked). These types of data sources maintain quality of interactions between users and items. Additionally, the system may have access to user-specific and item-specific profile attributes such as demographics and product descriptions respectively. Recommender systems used these records to identify well-matched pairs. The techniques that are used to implement these recommender systems like Collaborative Filtering systems[1][2] analyze historical interactions alone, while Content-based filtering systems are based on profile attributes and Hybrid techniques attempt to combine both of these designs. Recommender systems for real world problems are an active research in present days [3].

Recommendations are main resources in human decision. Based on recommendations user can choose their product/item[6]. With burgeoning consumerism buoyed by the emergence of the web, buyers are being presented with an increasing range of choices while sellers are being faced with the challenge of personalizing their advertising efforts. In parallel, it has become common for enterprises to collect large volumes of transactional data that allows for deeper analysis of how customer interacts with the space of product offerings. The term "collaborative filtering" was introduced in the context of the first commercial recommender system, called Tapestry, which was designed to recommend documents drawn from newsgroups to a collection of users. The motivation was to leverage social collaboration in order to prevent users from getting inundated by a large volume of streaming documents. In content filtering, recommendations are not "collaborative" in the sense that suggestions made to a user do not explicitly utilize information across the entire user-base [7][8][12][13]. Some early successes of collaborative filtering on related domains included the Group Lens system. Concurrently, several efforts are attempted to combine content-based methods with collaborative filtering and to incorporate additional domain knowledge in the architecture of recommender systems [4] [5].

2. Problem Statement

In the existing System, a user is recommended the web pages in accordance to the common page ranking or other algorithms. The recommendations for a particular search are naturally common to all the users, i.e., the differences between the users behavior towards the web pages is not considered**.** The Technique "Collaborative Filtering Technique" is unanimously being used in the Commercial Web Sites like EBAY, AMAZON etc. where the users are recommended the products by using this technique[9][10].

2.1 Proposed System

We propose an automatic method to recommend web pages for the users by implementing the proposed technique, "Collaborative Filtering Technique" for the web pages. The system uses the profile data of the users and tries to study the behavior of the users. In this , the recommendations are to be different for different users because the behaviors of the users vary from person to person.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A. Feasibility Study

A feasibility study is an evolution of a proposal designed to determine the difficulty in carrying out a designated out a designated task. Generally, a feasibility study recedes technical development and project implementation. In other words, a feasibility study is an evolution or analysis of the potential impact of a proposed project.

B.Technical Feasibility

Evaluating the technical feasibility is the trickiest part of the feasibility study. This is because, at this point in time, not too many detailed design of the system making it difficult to access issues have to be considered while doing a technical an analysis. Understand the different technologies involved in the proposed system. Before commencing the project, we have to be very clear about what are the technologies those are required for the development of new system .Find out whether the organization currently possesses the required technologies.

C .Social Feasibility

The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

3. Implementation

The proposed algorithm implemented in three phases:

Phase1: Webpage Collection
Phase2: Implementation of Collaborative Filtering Technique
Phase3: Personalization Diagnosis

Phase1: Webpage Collection

5000 urls are collected and used for observing users information. Describing and organizing this vast amount of content is essential for realizing the web as an effective information resource. If the results returned by the search engine have been classified into a specified category, the user scan and choose the interesting category to continue browsing. Traditionally, text classification [11] is performed manually by domain experts. However, human classification is unlikely to keep pace with the rate of growth of the web. Hence, as the web continues to increase, the importance of automatic webpage classification becomes obvious. In addition, automatic classification is much cheaper and faster than human classification.

Phase2: Implementation of Collaborative Filtering Technique

Collaborative Filtering (CF) systems work by collecting user feedback in the form of ratings for items in a given domain and exploiting similarities in rating behaviour amongst several users in determining how to recommend an item. CF methods can be further sub-divided into neighbourhood-based and model-based approaches. Neighbourhood-based methods are also commonly referred to as memory based approaches. As one of the most successful approaches to building recommender systems, collaborative filtering (CF) uses the known preferences of a group of users to make recommendations or predictions of the unknown preferences for other users. We then present three main categories of CF techniques: memory-based, model based, and hybrid CF algorithms (that combine CF with other recommendation techniques), with examples for representative algorithms of each category, and analysis of their predictive performance and their ability to address the challenges.

Phase3: Personalization Diagnosis

Personalization methodology is used to categorize the users into various categories and the web pages are categorized into various categories based on their usage. The web pages are then suggested to the users based on their category which was to be taken from the registration details of the user.

$$\underset{h=1,\ldots,b}{Max} \sum_{i=1}^{n} Pr(r_a(j) = h | r_i(j) = y)$$

H is possible matches
N is number of users
$r_a(j)$ is of active user
$r_i(j)$ is of existing user

Registration Page where the user details are collected by the system is shown in fig1



Fig1: Shows user registration form

Users without registering themselves can't get any personalized recommendations. No recommendations are provided initially for a new user as his behavior can't be analyzed. This can be shown in fig2.



Fig2: Initially no recommendation for new user

The home page of the recommender system looks as below. The above page loads when user gets registered into the system and then is logged in. Now the user has to type the keyword for which he is expecting the search results in the search box. Now the user has to click on the button "Go Search" shown in Fig3.



Fig 3: Recommended Web links

The system is proposed to take the category of the user in the registration page itself and it is saved in the database. This category is used to match the category of the web link. This column can be shown in Fig4

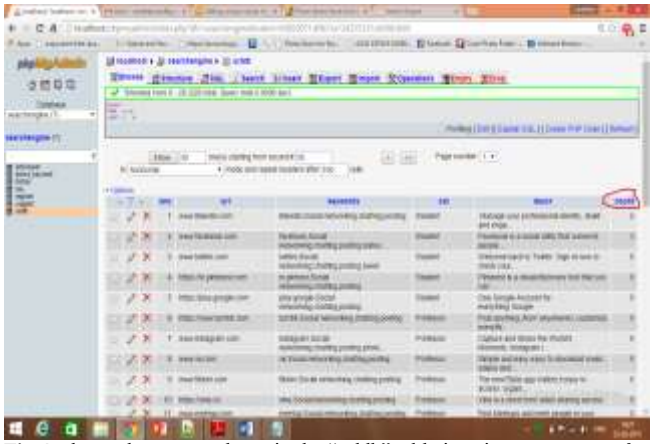


Fig 4: shows the count column in the "urldb" table is going to represent the number of times a URL is accessed by any user

The urls are suggested to the user based on the search and the count of the URL in the "urldb". This count is used to define the priority of the URL. The most priority link is displayed first and the descendant follows the top. This can be shown in Fig5



Fig5: Final Recommendations

The Recommendations now are suggested to the users by following the Collaborative Filtering Technique, using "Hybrid Methodology", where the user is recommended with the URL's based on his previous searches and the searches of other users with similar behavior. The numbers in the Fig5 are displayed to show the percentage of match the active user's urlset got matched with the other users in the database. This is depicted as above.

4. Conclusion

The main aim of the paper is to implement collaborative filtering technique for recommending suggestions for the users. Collaborative Filtering Technique is widely used in the implementation of commercial web site like "Amazon" etc., where the users are recommended with certain products likely to be required by the user based on his previous searches and search history. Similarly, we aimed to implement this collaborative filtering technique for the recommendations of web pages in this recommendation system.
The resultant recommendations of this system are probably of most common searches of the active user with another user with most common nature, i.e., the user with maximum matching of links. This System, hence gives probably the best suggestions to the users.

References

[1] K. Miyahara and M. J. Pazzani, "Collaborative filtering with the simple Bayesian classifier," in Proceedings of the 6th Pacific Rim International Conference on Artificial Intelligence, pp. 679–689, 2000.

[2] D. Billsus and M. Pazzani, "Learning collaborative information filters," in Proceedings of the 15th International Conference on Machine Learning (ICML '98), 1998.

[3] B. M. Kim and Q. Li, "Probabilistic model estimation for collaborative filtering based on items attributes," in Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI '04), pp. 185–191, Beijing, China, September 2004.

[4] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in Proceedings of the 10th International Conference on World Wide Web (WWW '01), pp. 285–295, May 2001.

[5] K. Miyahara and M. J. Pazzani, "Improvement of collaborative filtering with the simple Bayesian classifier," Information Processing Society of Japan, vol. 43, no. 11, 2002.

[6] M Prem Melville,, Vikas Sindhwani, "Recommender System" Encyclopedia in Machine Learning",, Springer Science and bussiness media, pp. 829–838,2011.

[7] Balabanovic, M. Andashoham,. FAB: Content-based collaborative recommendation. Commun. ACM 40, 3 (Mar.) y. 1997.

[8] Basu, c., hirsh, h., and cohen, Recommendation as classification: Using social and content-based information in recommendation. In Proceedings of the 1998 Workshop on Recommender Systems. AAAI Press, Reston, Va. 11–15, 1998

[9] BEEFERMAN, D. AND BERGER, "Agglomerative clustering of a search engine query log", In Proceedings of ACM SIGKDD International Conference. ACM, New York, 407–415, 2000

[10] KITTS, B., FREED, D., AND VRIEZE "Cross-sell: A fast promotion-tunable customer–item recommendation method based on conditional independent probabilities", In Proceedings of ACM SIGKDD International Conference. , ACM, New York, 437–446, 2000

[11] Lewis. D. Evaluating Text Categorization. In Proceedings of the Speech and Natural Language Workshop. Asilomar, CA., 1991

[12] L.M. de Campos, J.M. Fernández-Luna, J.F. Huete, M.A. Rueda-Morales, Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks.

[13] Balabanovic, M.; and Shoham¸ContentBased, Collaborative Recommendation. Communications of the ACM Vol. 40, No. 3. March, 1997.