International Journal of Advanced Technology in Engineering and Science www.ijates.com
Volume No 03, Special Issue No. 01, April 2015 ISSN (online): 2348 – 7550

USER PROFILING USING RECOMMENDATION SYSTEMS

P. Poornima[1], Dr. K. Kavitha[2]

[1]M.Phil Scholar, Computer Science, Mother Teresa Women's University, (India)

[2]Assistant Professor, Dept. of Computer Science, Mother Teresa Women's University, (India)

ABSTRACT

Web Mining is an upcoming trend which serves as per the user requirements. User Profiling is the booming area in research. Web Usage mining is to discover interesting user navigation patterns and can be applied to many real world problems, such as improving website/pages, making additional customer behavior studies etc. Identification of user's preferences using their profile can be used to add more suggestions to them and to improve the online shopping business. This can save a lot of effort on the user's part in browsing various advertisements regarding their preferences. This proposed system uses information such as profile info, websites they have visited so far, preferences given to the advertisements are collected from log file then process the selected info using recommendation system techniques to improvise the users recommendation as per the results.

Keywords: User profiling, Recommendation systems, Social networks, Big Data.

I. INTRODUCTION

Data Mining is the process of extracting knowledge from the large amount of data like data bases, data warehouse, data mart etc. This includes the interestingness of the user patterns. Mining the information is the concept. Data mining can be subdivided into many types like Text mining, Web Mining and Image mining.

Raw Data ⟶ Interesting Patterns ⟶ Knowledge

Now in each and every field, data mining has entered. It can be viewed as decision making systems. In business it is used to compare the current and previous year process by summarization and other techniques. In financial systems, it used to predict the upcoming status of the stock exchange or nifty changes. Data mining can also be used for security process. In banking sector, it is used for detecting the fraudulent activities while transacting large amount of money. It varies according the application used.

In recent years graphs of computer usage and World Wide Web are in increasing level. People who are using these facilities cannot be accountable. As it increases periodically user requirements also increasing. Users want the Wide World should shrink into a tab as he/she requires way. So there should be a system that analysis the user's expectation & preferences explicitly from user profiles or implicitly from user log files. Here the difficulty is automatic extraction of the useful information. Numerous web pages and sites are created day by day using HTML and other web applications. It is necessary to create a system using Data mining concepts in web technology. That is called Web data mining system. It has many sub divisions. The one is User Profiling.

II. RELATED WORK

Pooja Mehta *et al.* (2012) has described as Web mining is the application of data mining techniques to extract knowledge from Web data including web documents, hyperlinks between documents, usage logs of web sites, etc.[5]. Data may consist of text, images, audio, video, or structured records such as lists and tables.

Ahmad *et al.(*2013) has described as with the widespread of social media websites in the internet, and the huge number of users participating and generating infinite number of contents in these websites, the need for personalization increases dramatically to become a necessity. One of the major issues in personalization is building users' profiles, which depend on many elements; such as the used data, the application domain they aim to serve, the representation method and the construction methodology[6]. This survey aims to discuss the available user modeling techniques for social media websites, and to highlight the weakness and strength of these methods and to provide a vision for future work in user modeling in social media websites.

Pasquale Lops *et al*(2011) has described that recommendation system based on the entities provided by the LinkedIn profiles such as "Interest", "Groups", "Associations', etc..In the network, vectors of adjacent users are added to the user's vector. The similarity between the paper's vector and user's vector is then calculated to recommend the appropriate users using recommendation engine.[2]

Burke (2002) has described the interests of researchers using probabilistic topic modeling. To analyze the terms in large bodies of text and how they are interconnected, the above method relies on statistical methods.[1]

Fabian Abel *et al.*(2011) has described about approach. The approach where Twitter data along with the URLs content from the tweet of the user is used instead of LinkedIn data. Additionally, entity recognition is used to enrich the user model. In vector space, the user model is again represented as the articles are recommended to the user. [3]

Qi Gao, et al.(2011) has described that semantic user modeling based on Twitter posts. We introduce and analyze methods for linking Twitter posts with related news articles in order to contextualize Twitter activities. Then propose and compare strategies that exploit the semantics extracted from both tweets and related news articles to represent individual Twitter activities in a semantically meaningful way. A large-scale evaluation validates the benefits of our approach and shows that our methods relate tweets to news articles with high precision and coverage, enrich the semantics of tweets clearly and have strong impact on the construction of semantic user profiles for the Social Web.[4]

III. USER PROFILING

Web user profiling is the process of acquiring values of different user properties that makes the user model. To mine the user's interest from their historical data, considerable efforts are being made. User's interest can be represented by creating a list of relevant keywords but this is not sufficient for modeling and understanding user's behaviors. To provide high quality web service a complete user profile along with the educational background, experience and contact information is an important. These types of systems are developed from the concept of Web Mining.

This user profiling helps the online advertising firms to target more customers based in the current position rather than focusing only on their interest. Formerly, user profiling is considered as an engineering issue and was done manually or conducted separately in an ad-hoc manner. In web based social networking site such as Face book and MySpace, certain user profiles would be incomplete, just because they do not wish to fill those details. There may also exist in some profiles with inconsistent and irrelevant details.

User profiling can also be done by using the list of keyword generated using statistical methods. For example, discovering most often used words from the user entered information. However, in this composition little semantic information like affiliation and location are ignored.

Works are been conducted to automate the process of building the used profile by data extraction technology. The proposed work use predefined rules or specific machine learning models to extract the various types of profile data in a distributed pattern. The profile information in user's related documents are retrieved from web page.

IV. CONTENTS IN USER PROFILES

4.1 User Info

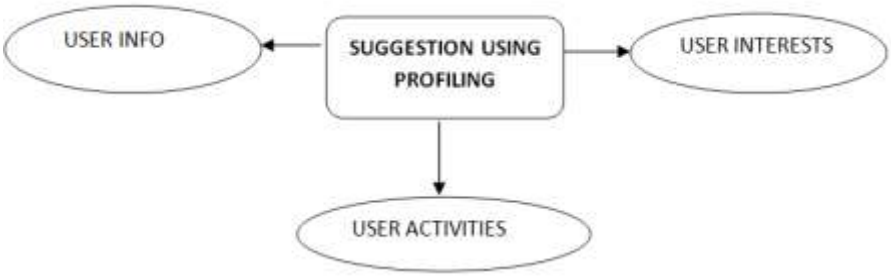
Information related to the users like name, hobbies and educational information etc., given by the user itself. These types of information are given by the user whenever they creating their account in social networks and other websites.

4.2 Interests

Information regarding the interests likes user preferences. Frequently visited website addresses. And the hobbies related to the user and his friends. For example items like sports, media, cinema. Likes and commands the users have given particular items which are calculated using recommendation systems.

4.3 User Activity

User activities are observed implicitly by using recommendation systems. Writing a blog every week, online tickets booking frequently like things says that the user is interested in writing stories and traveling . So books related that blog and advertisement related tour packages can be given as suggestion by predicting the user ctivity.



V. USER PROFILING TECHNIQUES

5.1 Recommendation Systems

Typically, Recommender system is used to compare the user profile information given directly by users with some reference characteristics, and also predicts the rating given to the new items like suggestions and friends. These characteristics may be associated with the information contained in the item (the content based approach) or the user's social environment (the collaborative filtering approach). Implicit data collection is important in this context.

Implicit data collection includes:

•Observing the items that a user views.

•Analyzing item/user viewing time.

•Keeping a record of the items that a user has purchased.

•Obtaining a list of items that a user has listened to or watched on their computer.

•Analyzing the user's social network and discovering similar likes and dislikes.

RA Compares the previous data to check whether similar and non similar data. Then suggestion items can then be calculated for the current user using the above checking process. For our recommender model, we simulate collection of user data from the first two sources: observing the items user views and the viewing time for that item. However, in this preliminary experiment, we assume that when the user inputs a query to our model, we compare the similarity between this query and previous users search information to select a suitable recommender model.

5.2 Collaborative Filtering Approach

The collaborative filtering approach makes automatic predictions about the interests of a user by collecting preference information from many other users. There are different types of collaborative filtering methods available: memory-based (measures the similarity between pairs of users to give the prediction, the Pearson scheme is a well-known memory-based scheme), model-based (finds patterns from the training data, such as SVD, Bayes methods) and the rating-based approach (predicts how a user would rate a item from other users rating, such as the in Discover website). In our investigation, we explore the rating-based collaborative filtering approach. In our work we chose the Weighted Slope One algorithm to compute predictions since it is efficient to query, reasonably accurate, and supports both online querying and dynamic updates, which makes it a good candidate for real-world systems. The Weighted Slope One algorithm comprises of the following two steps:

- Looking for other users who rate the same document as the current user.
- Using the ratings from these like-minded users to give predictions for our current user.

Rating-based collaborative filtering requires data pairs from a user: the document ID and its relevance weighting entered either explicitly or inferred from implicit data such as viewing time. The response of a RA to a query pair is an array of recommended pairs (document ID, rating) of documents based on the training data captured from previous users, which the current user has not rated yet.

5.3 Machine Learning

It is an important concept in artificial intelligence. It helps to take decisions to suggest the interested patters. Many algorithms are used in this machine learning like NLP (Natural Language Processing). For example unimportant messages and mails are automatically forwarded into spam without getting suggestion from the user. Here full observation of user activity is used. That is the messages or mail to which user has given more important like star and save options. These types of machine learning approaches identify the interestingness patters. Supervised and Unsupervised are two types in machine learning which includes neural networks, cluster analysis respectively.

VI. CONCLUSION

In this paper, Researcher discussed about the important concept of Web Mining. Most of the research work highly focused on User profiling. This user profiling is not only used in social networks and also in the websites where online shopping and purchasing is the main concept. By studying implicitly about the user and his profile,

the recommended items can easily be reached to the user. This will result in time saving, profit increasing and effective use of storage spacing. Recommendation systems are collectively used with Data mining concept in order to predict the interestingness patterns.

REFERENCES

[1]. Burke, R. Hybrid recommender systems: "User Modelling and User-Adapted Interaction" ,12(4):331–370, 2002

[2]. Pasquale Lops, Marco de Gemmis, Giovanni Semeraro, FedelucioNarducci, and CataldoMusto"Leveraging the linked in social network data for extracting content-based user profiles"ACM conference , pages 293–296. ACM, 2011.

[3]. Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. "Analyzing user modeling on twitter for personalized news recommendations."User Modeling, Adaption and Personalization, pages 1–12, 2011.

[4]. Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao "Semantic enrichment of twitter posts for user profile construction on the social web" Thesemanic web: research and applications - Volume Part II, ESWC'11, pages 375–389, Berlin,Heidelberg, 2011.

[5]. PoojaMehtaa, Brinda Parekh, KiritModi, and PareshSolanki, "Web Personalization Using Web Mining: Concept and Research Issue", International Journal of Information and Education Technology, 2012.

[6]. Ahmad Abdel-Hafez and YueXu. "A survey of user modelling in social media websites. Computer and Information Science", 6(4):p59, 2013.