



Research Journal of Pharmaceutical, Biological and Chemical Sciences

A Survey on Tag Recommendation Techniques using Text Mining.

Vijay K *, and Monisha S.

Department of Information Technology, Faculty of Computing, Sathyabama University, Chennai, Tamil Nadu, India.

ABSTRACT

In the environment of World Wide Web (WWW) bookmark is used through Uniform Resource Identifier (URI) accessed and that is stored for the future references in many formats, the bookmarking is a centralized service from which the users can add, delete, view and organize more information. It also includes of tagging of the text, and also provides the benefits of book marking to the users. It can be classified into two types: 1.Public, 2.Private. Public bookmarking user can be viewed by all the users on the web whereas private bookmarking is only accessed by the particular person who has eligible to access the bookmarking. The tag recommendation is also a method used in bookmarking which is predicting the tags for a given user and item, based on the past user behavior. Parental control systems provide a form of authentication to prevent from unauthorized users from disabling it.

Keywords: Tag recommendation, Book marking, social networking, collaborative filtering, Term Frequency- Inverse Document Frequency.

**Corresponding author*

INTRODUCTION

Bookmarking is simply defined as a marker which is referred by the users to keep the readers place as a marker to enable the user to return back with ease. It provides centralized online services to the user to add bookmarks of web documents. At this bookmarking also provides a tag recommendation technique which is used to support the users during the tagging process to provide a facility called tagging which includes of tag recommender. Whenever web user finds any interesting information on the web page through any web browser then, it shows a top most recommendation tags to the searching results to web users, here the main aim of the recommender is to identify the exact keyword of a user and keeping it in mind whenever describing the webpage.

RELATED WORK

In this paper, it allows the user to share tagged information and also upload to use the resources in the form of bookmark and also in publications. It has changed from an individual bookmarking from an individual desktop to a conspired environment on web. It allows a tag recommendation that automatically suggests when a new user enters a tag to the related search. here it proposes a model that does not depend on the tagging history of a user or resource but it suggest the tags based on user submits the tag on system for first time. Multi label text classification is an increasingly required for modern applications and it is an independent classifier for each category ranking, threshold value. It also uses a technique of Term Frequency-Inverse Document Frequency (TFIDF), it is a numerical statistics which represents how important a word to a document. The main features of TFIDF are used as a weighting factor for mining the text and also retrieval of more relevant information. It provides more interoperability; this method may be used frequently, like kernel algorithms which cannot be up to mark to do with many numbers of samples. This is because of each individual and their learning through problem that describes in a specific identification mark in that dataset, with remaining contains in the rest, the complete dataset is used many times in test through number of classes. The constraint found in TFIDF is that, it clusters the documents that are keyword having more similarity. It is only useful for identifying the identical documents which consist of most recommended tags [1].

Herewith it represents a combined method of mitigation Sparsity problem in tag recommendation and it is mainly used for ranking candidate tags based on the users tag. It uses the collaborative filtering technique to enhance the performance. it is applied to many different kinds of data and this algorithm always requires the need of users active participation, users interest to the system and an algorithm that can able to match with the similar users interests. It has two methodologies one that is based on the users who share the same rating from the particular active web user, then it uses same ratings from other web user with same mind set. The benefits of this method is that explain ability of the result which is most important aspects of recommendation system, easy to develop and analyze, content independent of recommendation. The limitation of this method is that it affects the performance with large datasets [2].

Herewith produced new way of approach for combining item-based collaborative filtering with context based reasoning. It is mainly based on similarity between the items calculated using people's rating with selected items. First the system builds a stage of similarity between all the pairs of items and this helps to find correlation between ratings, Secondly it uses system to execute a recommendation tags. The features of this item based collaborating filter have less error compared with the user-user filtering method. The limitation is that the distance bound is often hard to find in item based collaborative filtering [3].

Since the online social network has been changed their way towards people to connect with each other to share information, they have introduced a fine grained attribute based reputation in terms of attributes and enable the users to rate each other's. It also provides the security analysis along with simulation result and privacy preservation. They have used zero knowledge proof technique for verification process for proving the statement it requires some secret information to prove to the verifier, it will not be able to prove the statement to anyone else. Since, they do not contain the secret information. It uses homographic way of encryption technique used, which will encrypt as cipher text content to plain text content, and also it allows chaining together of multiple services without knowing the data to each other services. The feature of this technique is computationally produces faster result and reduces the cost compared to other encryption techniques used. The limitations are that it is semantically unsecured and it will not provide verifiable computing [4].

Introduced a technique for generating a personalized tag recommendation for users of social bookmarking it recommends the tags from url's (Uniform Resources Locator) that are similar . They have used a content based method that is capable of recommending tags that are not used before in tagging. It retrieves the content from large collection of syntactical features. The content based similarity search analyzes the metadata such as tags, keywords and other associations. The benefits of this is tagging-based method is more over easy way to implement, since it does not require a separate index for the content fetched from the different urls (address).The limitation found, it is somewhat effectiveness at finding good recommendations than the content based method usage, and has much lower coverage for large dataset[5].

Herewith they have used a approach called Personal Learning Environment (PLE) concept which helps the scholar to take control and manage their own learning it also provides a support for their level of achievement and they have used tag recommendation with 16 types of tag recommendation algorithm with memory and model based it also helps to provide accuracy and the users satisfaction. Herewith, introduced a learning technique called as TEL(Technology Enhanced Learning) it allows in such a way that maximize learning with an environment provides particular type of more noise coarse in the concept of design work which helps the web users to provide a good amount of time, place and different types of learning styles. The features of this model the PLE (Personal Learning Environment) are not only to seek the information but it also used to share the information with other web users. The limitation of this model it attempts to provide an overview of recommender system for TEL settings, as well as highlights the particular and compared to recommender system [6].

In this paper they have explained about the feature changes from working on single personal computer to social bookmarking websites that can able to store, manage, access the information by the user in the website that is based on the users interest and recommendation searching is carried on. From these techniques it has an emerging technology called spamming which is identified as a problem over here. To overcome the language models of class stochastic n-gram model is used to measure the degree of surprise in some of the new spam of text. They have given a three strategies detection, demotion and prevention where prevention becomes a toughest problem to overcome in social bookmarking websites to restrict the access of certain implementation such as tagging or post. It also additional features in social bookmarking and generates a back link which is useful for the webmasters to attempt to improve the website. The limitation of this spamming in social bookmarking is the more often the page is submitted and tagged it has the better chances of being found and spammed [7].

In this paper they have given a co-occurrences of tags which helps the users to show off the popularity of a web that helps the spammers to get into the website of targeting the services which leads to misleading of tags by simply confusing the user, in order to solve this problem they have introduced a algorithm spam score propagating , it analysis in three way of directions as user, tag and web and it represents as a graph in that each node constitutes a user and it also assigns a value to identify whether the user is spammer or not, the values are propagated through graphs. The main features of this that is new service to generate misleading tags to increase the visibility of some useful resources. The limitation is that when compared to spam detection from other web application on detecting spam from collaborative tagging system is very limited [8].

DISCUSSION ABOUT TAG RECOMMENDATION

Once web users recommended the particular information on the web by using tagging the content, forth coming web users can be able access the information with very less time. Based on the content and it is shared with help of clusters, updating information also be comes under same cluster. So it easy to any new web users to follow the recommended tag added by the previous users. Topic searches are easier to the web users with this tagged content on the web. Eg user searches about the sports, current affairs, medical, and weather report and so on. We take with medical terms such as human brain, liver, heart, nervous etc. Brain related information has been tagged; other user can be identify the appropriate content and can proceed further.

CONCLUSIONS

Herewith this survey deals with which improves the tag recommendation and its efficiency accessing the websites. The website consists of lot of content and information in it. This contents consists of lot of

information in it, were tags are updated day by day in the websites. To improve the efficiency the following techniques are used Term Frequency- Inverse Document Frequency (TFIDF), and also collaborative filtering technique, combining item-based collaborative filtering is used. These techniques are used for improving better efficiency of tag recommendation .This allow organizing tag in more fast and effective manner. Collaborative tagging method which provides efficient way to analyze the tagging tagged content from the website.

FUTURE DIRECTION

Once the recommendation of tags obtained by the user and it can be forward to the further users. Other users need not search for a long time to get their apt content from the website. In recent day's user are accessing web pages through different types of devices.

ACKNOWLEDGEMENTS

We would like to thank Department of Information Technology, Sathyabama University, provided the lab facility to do our research work.

REFERENCES

- [1] Yagnik, Shweta, Priyank Thakkar, and K. Kotecha, " International Journal of Data Mining & Knowledge Management Process 4.1 2014; 19-32.
- [2] Djuana, Endang, 47th Hawaii International Conference on System Sciences (HICSS), IEEE 2014; pp 906-915.
- [3] Chedrawy, Zeina, and Syed Sibte Raza Abidi, 5th International Conference on Intelligent Systems Design and Applications IEEE 2005; 67-72.
- [4] Guo, Linke, Yuguang Fang, and Lingbo Wei, International Conference on Communications in China (ICCC), IEEE/CIC IEEE 2013; 230-235.
- [5] Byde, Andrew, Hui Wan, and Steve Cayzer, " University of Bath, UK, 2007.
- [6] Chatti, Mohamed Amine, IEEE Transactions on Learning Technologies 2013; 6(4):337- 349.
- [7] Bogers, Toine, and Antal Van Den Bosch, Social Informatics: 4th International Conference, 2008; 1-11.
- [8] Krestel, Ralf, and Ling Chen, "Using co-occurrence of tags and resources to identify spammers", Proceedings - ECML/PKDD Discovery Challenge Workshop 2008; 39-46.