



International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 10 • Number 32 • 2017

Web Recommendation System for End-Users

Chichili Vinathi^a and Shaik Naseera^b

^aComputer Science and Engineering, Vellore Institute of Technology, Vellore, India. Email: ^ac.vinathi2015@vit.ac.in; ^bshaik.naseera@vit.ac.in

Abstract: In this era where the availability of digital information is growing exponentially, Recommendation systems are utilized to suggest items of interests for end users. Suggestions are done here by using statistical and clustering techniques that assist in predicting the interests of user. This paper proposes a Web Recommendation System for End Users by proposing algorithms Modified K-Means Clustering and Vector Space Model algorithm. Vector Space Model algorithm is exploited for enhancing the data classification and making it uniform followed by modified K-Means clustering on input data that is retrieved from using Vector Space Model algorithm. Recommendations are done by utilizing text categorization from Search Keyword.

Keywords: Vector Space Model, Modified K-means algorithm, Duplication Identification Algorithm, Web recommendation, Keyword Search.

1. INTRODUCTION

In recent times, digital information is trending which leads to an exponential growth of internet usage, people are depending on internet for every common need such as to search for buying their favorite book or for reading it online, booking for a movie or even to search the routes via global positioning system (gps). With the target of enhancing the recommendation process while searching online and for achieving precise recommendations Web Recommendation System is developed. It helps algorithm developers for predicting the likes and dislikes of user among a given list of items. Recommendation System is an intriguing alternative to search fields as it can assist users to explore products and content that might not be discovered otherwise. It leads websites and services like Facebook, flipkart and other sites to integrate recommendation system in search process. The recommendation works typically in one of two ways. It can focus on specification of items that a user likes which helps the developers to derive the likes and dislikes of other users which can be exploited by recommendation system to define a similarity index among users and suggest items correspondingly. Combining these two strategies helps in building an efficient recommendation system.

Recommendation System causes change in the perception of the way typical websites communicate with users. Instead of giving a static experience where users search products for purchasing, it increases the

communication which results in providing a richer search experience for end users. This process works by identifying recommendations separately for each user based on their past searches or purchases. It works on available data that depends on the behavior of user's purchase history. In providing recommendations, it will use the details of a registered user's profile and their navigation behavior which can be analyzed by their search histories, then compare information retrieved and give the recommendation based on that.

The Web Recommendation System proposed here uses Modified K-means Algorithm (MK-means algorithm) for clustering the features that will be extracted by using Vector Space Model algorithm. Duplication Identification algorithm is integrated in this to remove any duplicate occurrences of URL in the recommendations.

Vector Space Model Algorithm: It is described over the years in terms of "formal, feature-based, individual, partial match" retrieval techniques as they typically focus on mathematical models for retrieval and modeling the documents as a set of terms that can be weighted individually and manipulated. It relies on basis that meaning of a document can be retrieved from document's individual terms. They equip an refined method of providing relevant feedback by characterizing documents and the individual terms in same space. Relevant feedback is achieved by permitting the documents and terms in those documents as well to structure a query and by utilizing terms to supplement the query helps in escalation of the length as well as the precision of query which in turn will help the user to specify what their desires from the search accurately.

Modified k -means algorithm (MK-means algorithm): K-Means is one of the most widely used algorithm among all partition based clustering algorithm. It is used to partition a dataset into meaningful patterns. K-Means has a shortcoming of empty clusters generation. In this project a modified k -means algorithm is proposed in order to eliminate the shortcoming of empty clusters generation. Here, the basic structure of the original k -means is preserved along with all its necessary characteristics. Modified K-Means is an algorithm that works well with large datasets. It reduces the adoption of cluster -error criterion and provides more efficient results.

2. LITERATURE SURVEY

There has been many studies done on Web recommendation, but the web recommendation system for end-users proposed here gives a more refined way of suggestions to users on basis of the data that is acquired by performing text categorization through the keyword that is used for the searching process.

Bieliková Mária, Kompan Michal and Zeleník Dušan proposed a method that can be used for personalized recommendations dedicated to news domain on web[1]. They worked with vector representation of news and the hierarchical representation of similarities between the items and how the suggestions can be done efficiently.

Daniar Asanov described all the traditional methods used in the recommendation systems and also explained various modern approaches[2] that have been developed recently such as Content-based filtering, Collaborative filtering and hybrid approach for providing user with personal recommendations. These techniques mainly focused on ranking and similarities among users.

Pasquale Lops, Marco de Gemmis and Giovanni Semeraro explained by taking user profiles into consideration and how the role that will be played by user generated content in the future generation. He also discussed how this will challenge in evolving vocabularies and also giving users spontaneous recommendations according to his profile that one may not come across otherwise[3].

Choonho Kim and Juntae Kim in the paper proposed a model based algorithm for recommendation by exploiting multi-level association rules and how it works more effectively than collaborative filtering process which calculates the similarities among users[4] according to his/her preferences and suggestions are the items that are preferred by the users with similar preferences.

Debajyoti Mukhopadhyay, Ruma Dutta, Anirban Kundu and Rana Dattagupta proposed an alternative for recommendation systems that can be used for B2C purpose[5]. This paper shows how by using vector space model the suggestions can be done for new user based on finding the profile that is closest to the new user in database. It also implemented recommendation system based on association rule mining which considers the order of purchases thus helps in suggesting more than one product.

Savadekar, Vinaya B., and Pramod B. Gosavi focused on providing a personalized recommendations on basis of keyword. In order to point out preferences of user[6], it is implemented in Hadoop and a user based MapReduce parallel processing paradigm is also utilized here on basis of keyword.

Shunmei Meng, Wanchun Dou, Xuyun Zhang and Jinjun Chen proposed a recommendation system where suggestions are done based on keyword and is implemented in Hadoop[7] by using map reduce parallel processing algorithm. By implementing it in distributed computing platform called Hadoop they tried to improve efficiency and scalability in giving personalized recommendations.

Desai, Pallavi R., and B.A. Tidke in the paper provided a survey[8] of how widely used keyword based recommendation systems are and how it will impact future by also being implemented in big data. In their survey they proposed a method to consider the weight of keywords in order to indicate preferences of user along with collaborative filtering algorithm which can be implemented in openlp and how by integrating this with map reduce algorithm and implementing in hadoop will help in providing smart recommendations to user.

Cyrus Shahabi, Farnoush Banaei-Kashani, Yi-Shin Chen and Dennis McLeod Proposed Yoda[9] which is able to support a large scale of web applications also provides highly accurate recommendations by exploiting a hybrid approach that combines querying on content basis with collaborative filtering with handling of real time data.

G. Adomavicius and A. Tuzhilin in the paper provided a brief overview of the current techniques and approaches being used in the field of recommendation systems[10] that can be classified as collaborative approach, recommendation using hybrid approaches and content-based approach.

In the related work mentioned above the recommendations were done either focusing on feedback of user or on the preferences among users who might have same interests as the registered user which can be a drawback in providing efficient recommendations. Here by using vector space model algorithm to classify the data and modified k -means algorithm which helps in overcoming the challenge of empty clusters generation, the web recommendation system succeeds in providing more precise suggestions.

3. PROPOSED METHODOLOGY

Recommendation system is generated to provide suggestions to end user based on his/her navigational behavior. Here, in this Vector space model algorithm is utilized to retrieve data that is obtained when users enter keywords to search on web. It also makes data uniform and categorizes data into features that are treated individually, then by applying modified k -means algorithm on this data for generation of clusters. Lastly, duplication identification algorithm is used so as to remove any duplication that might occur in URLs recommended.

Vector-space models were developed to eliminate the problems that arise with the lexical matching techniques. By placing terms, documents and queries in a term-document space and also computing similarities between them allows the results of a query to be ranked according to the similarity measure of their usage Unlike lexical matching techniques that considers no ranking or a very crude ranking scheme the vector-space model bases their rankings on the Euclidean distance or the angle measured between the query and terms or documents in the space, hence are capable to automatically guide the user to documents that might be more conceptually

similar and of greater use than other documents. Also, by representing terms and documents in the same space, vector-space models often provide an elegant method of implementing relevance feedback. Hence, by using vector space model algorithm in web recommendations it will increase the precision, which will help the user to specify more accurately of what he or she desires from the search thus resulting in more refined suggestions.

A. Modified *k*-means algorithm (MK-means Algorithm)

In clustering a set Z of N patterns of dimension is partitioned into V clusters that are denoted as

D_1, D_2, \dots, D_v So that clustering metric S is given as

$$S(D_1, D_2, \dots, D_v) = \sum_{v=1}^v \sum_{y_k \in D_v} \|y_k - C_v\|^2$$

Here, $C = \{C_1, C_2, \dots, C_v\}$ is the set of cluster centers.

In MK-means algorithm, calculation of new center vector is as below mentioned:

$$C_v^{(new)} \leftarrow \frac{1}{n_v + 1} \left\{ \sum_{y_k \in D_v} (y_k) + C_v^{(old)} \right\} \quad (1)$$

Input: a set Z of z -dimensional data and an integer V .

Output: V clusters

Begin

Choose V points randomly that belongs to $Z(\in Z)$, to be initial means;

While measure metric S is not stable do

Begin

Calculate distance $d_{vk} = \|y_k - C_v\|^2$

for each v, k ; where $1 \leq v \leq V$ and $1 \leq k \leq N$ and

determine the members of new V subsets on basis of the minimum distance

to C_v for $1 \leq v \leq V$;

calculate new center C_v for $1 \leq v \leq V$ using (1);

Calculate S ;

end

end

B. Duplication Identification Algorithm (DIA)

Begin

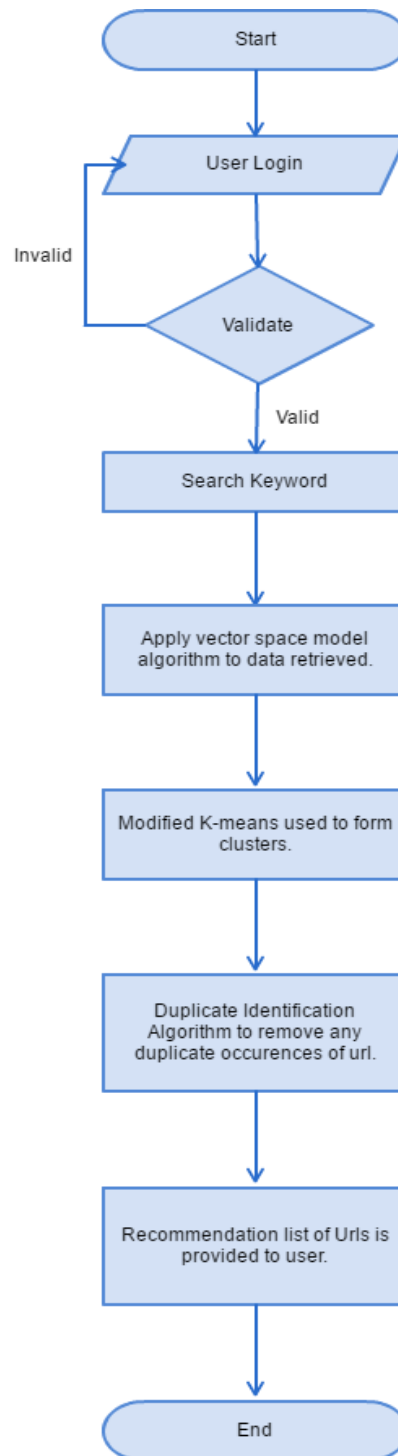
For attribute $b = 1$ to last attribute, y

For row $a = 1$ to last row, x

1. The attribute values are brought into a uniform format.
2. Special character should be removed because they may increase the value of radix.

4. Any abbreviations should be expanded.
5. All the values are converted into homogeneous form (numerical form).
6. Put them into a dataset.

End



Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

4. EXPERIMENTATION AND RESULTS

Result 1



Figure 1: Depicts the logged in user behaviour by the keyword he used for searching and gives recommendations.

Result 2

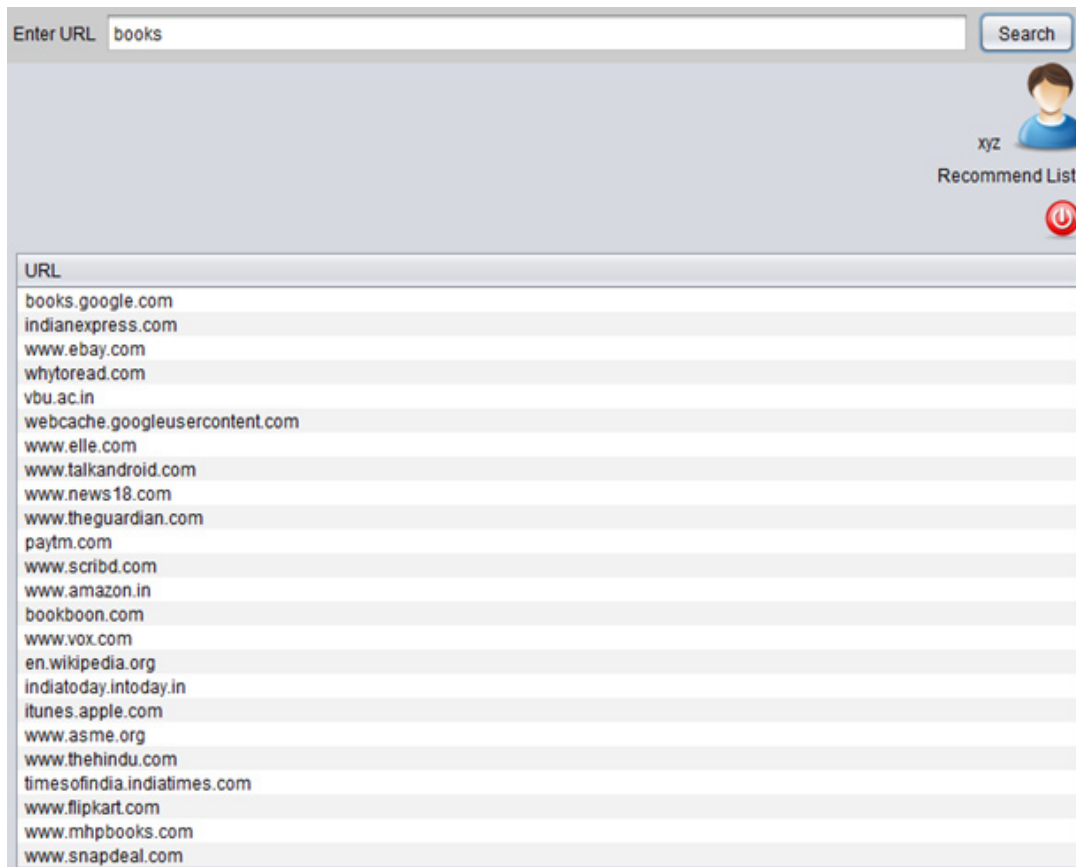


Figure 2: Recommendations given in URLs

Figure 2 shows the recommendations provided have no duplicate occurrence of any URL which is achieved by integrating Duplication Identification Algorithm in Web Recommendation System. Here, the data is extracted from web based on keyword given by the user while searching and the extracted data will have the page that has the keyword mentioned more times on top of search results.

This Paper proposes usage of modified k -means algorithm along with vector space model algorithm in order to improve the efficiency of the web recommendation system. Experimental results show that by using modified k -means algorithm for clustering it reduced the time taken by efficiently comparing thus provided a part in enhancing the performance of web recommendation system.

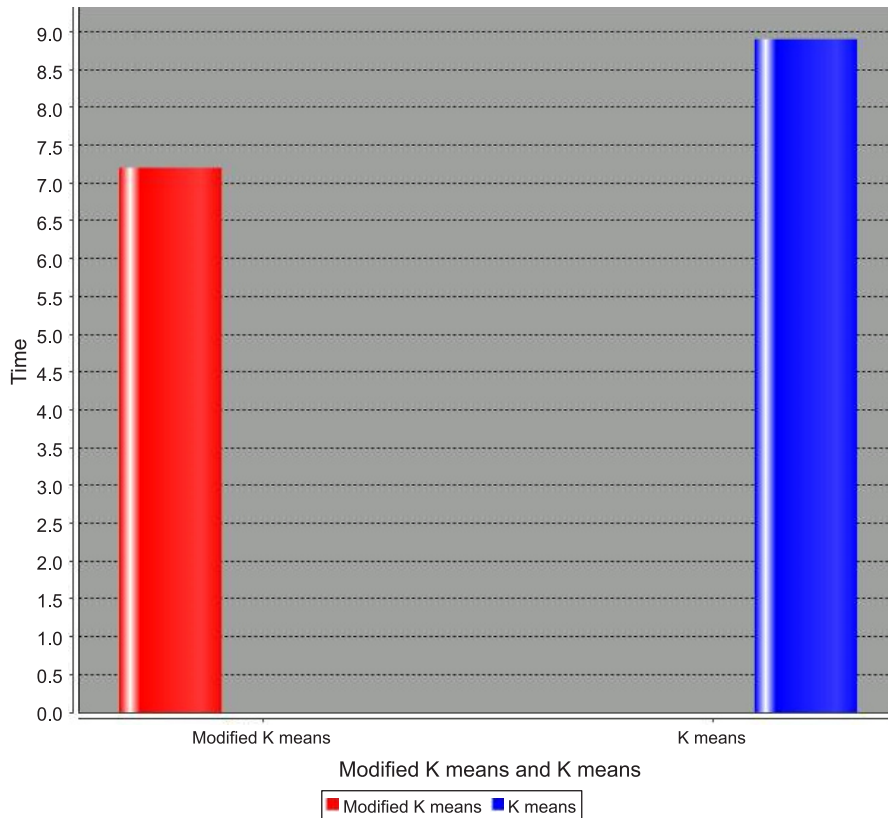


Figure 3: Graph comparing original k -means and modified k -means with time as factor.

5. CONCLUSION

Proposed implementation of using vector space model for partitioning data which is retrieved when user search using keyword into features and Modified K-means for clustering data improved the scalability and efficiency of the recommendation system. The recommendations achieved implementing proposed model are accurate thus succeeds in giving suggestions which the user may not come across typically. In future, it can be further extended to integrating positive as well as negative preferences according to user which can be achieved by their reviews so the suggestions can be even more accurate.

REFERENCES

- [1] Bielikova, Maria, Michal Kompan and Dusan Zelenik “Effective hierarchical vector-based news representation for personalized recommendation” (2012).

- [2] Asanov Daniar “Algorithms and methods in recommender systems”(2011).
- [3] Lops, Pasquale, Marco De Gemmis and Giovanni Semeraro “Content-based recommender systems: State of the art and trends” (2011).
- [4] Kim Choonho and Juntae “A recommendation algorithm using multi-level association rules” (2003).
- [5] Mukhopadhyay and Debajyoti “A product recommendation system using vector space model and association rule” (2008).
- [6] Savadekar, Vinaya B and Pramod B. Gosavi “Towards Keyword Based Recommendation System”.
- [7] Meng Shumei, Wanchun Dou, Jinjun chen and Xuyun Zang “KASR: A Keyword-Aware Service Recommendation method on Map Reduce for big data applications” (2014).
- [8] Desai, Pallavi.R and B.A. Tidke “A Survey on Smart Service Recommendation System by Applying Map Reduce Techniques”.
- [9] Shahabi Cyrus, Yi-Shin Chen and Farnoush Banaei-Kashani “Yoda: An accurate and scalable web-based recommendation system”.