Web Usage Recommendation System using K-NN Algorithm

Pritali Kuyate[1] Poonam Jondhale[2] Sujata Mekhe[3] Pushpa Aher[4]
[1,2,3,4]Department of Computer Engineering
[1,2,3,4]Savitibai Phule Pune University, Nashik District Maratha Vidya Prasarak Samaj's , Karamveer Adv. Baburao Ganapatrao Thakare College Of Engineering Nashik, India

Abstract— In today's world the large problem of many online websites is the presentation of multiple choices to the user at time. So, its very difficult task for user to find desired product or information on the site. Web mining and recommendation system is based on the current users brhaviour and his/her search pattern on the newly developed Really Simple Syndication (RSS) reader website, which provide relevant information to users according to their needs and interest. The k-NN classifier is very simple to understand and implement than other machine learning algorithms. The k-NN classification method is useful in online and real time to identify the users serach pattern, matching it to a particular user group and recommend a smart option that meet the need of specific user at particular time. For this purpose current users previous logs are extracted, cleansed , formatted and grouped into meaningful session to develop data mart.

Key words: Data mining, Web mining, k-Nearest Neighbor, Online, Real time

I. INTRODUCTION

Web contains the very large amount of data in different formats. When any user want to search particular product or information on the web the its very time consuming task because he gets multiple choices at a time. It is very difficult for user to find desired product which can satisfy his needs.Web mining is the application of data mining to discover useful information from the web. There are three research areas in web mining:

- Web usage mining
- Web content mining
- Web structure mining

A. *Related Work*

Classification: This refers to a form of data analysis that can be used to produce models that brings out important data classes, classification predict categorical labels. A classifier is an abstract model, that describes a set of predefined classes generated from a collection of labeled data (Luca and Paolo, 2013). (Jiawei and Micheline, 2006; Luca and Paolo, 2013), stated that there are different techniques for data classification which includes; decision tree classifier, Bayesian classifier, K-Nearest Neighbor classifier, rule base classifier etc. In our work, the K-Nearest Neighbor classification method was adopted [1].

Web usage mining: Web usage mining focuses on predicting users preferences and behavior by analyzing web logs with the help of traditional data mining techniques. Customers click stream data can act as a very rich source of information. Click stream indicates users path through a website. Click stream data is captured and maintained in web log files. Strategic use of navigational data can be very useful in providing effective recommendation.

K-Nearest Neighbor: The K-Nearest Neighbor (K-NN) algorithm is one of the simplest methods for solving classification problems; it often yields competitive results and has significant advantages over several other data mining methods. It is able to achieve [1] :

Overcoming scalability problem common to many existing data mining methods such as decision tree technique, through its capability in handling training data that are too large to fit in memory [1].

The use of simple Euclidean distance to measure the similarities between training tuples and the test tuples in the absence of prior knowledge about the distribution of data, therefore make its implementation easy [1].

Reducing error rate caused by inaccuracy in assumptions made for usage of other technique such as the Naı¨ve Bayesian classification technique, such as class conditional independency and the lack of available probability data which is usually not the case when using KNN method [1].

Providing a faster and more accurate recommendation to the client with desirable qualities as a result of straightforward application of similarity or distance for the purpose of classification [1].

II. PROPOSED SYSTEM

We proposed a Web usage mining and recommendation system using k-NN classification algorithm to provide smart browsing option to the user without explicitly asking for it. That browsing options are based on the users interest. So, that can satisfy users' needs in very less time.

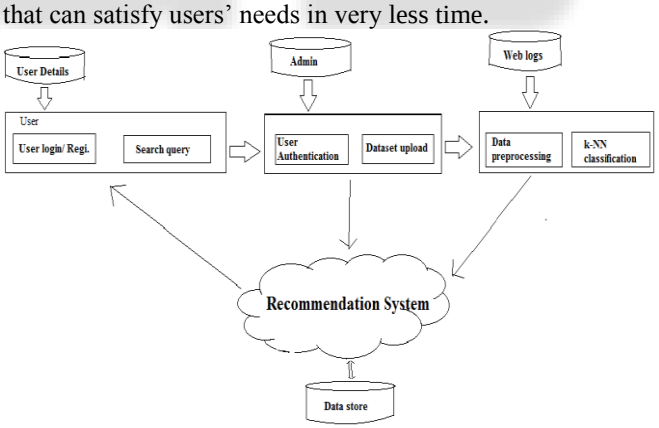

Fig. 1: Proposed Recommendation system

A. *User Details*

When any new user is come into the system then he have to register first. If he is old user then he can login to the system and get access. User registration process is necessary to differentiate users from each other and maintain their logs separately. After getting access into the system then user can enter his search query and obtain results.

B. *Admin*

First task of admin is to validate user or register new user. If user is new then admin registers that user into the system and maintain the data of that particular user and if user is old

then check login id and password of that user. If both are correct then and then only give access to that user in to the system. Second task of admin is to upload dataset. The dataset which want to used into the system which can upload by admin.

C. Data pre-processing

Web logs of each individual user are used as a input to the system and that is preprocessed for further operation. Web logs contains multiple records and information. When user enters on to the web, What data he accessed by him, how many times, Which websites are visited mostly, IP address of that users system and many more information. Web logs also contains the error or failure entries, some access records which are generated by search engine. Hence data processing step perform data cleaning, formatting and grouping operation. In data cleaning all unwanted entries are removed and only that entries are extracted which are useful for further processing. After collecting useful entries and records data mart is developed for recommendation and classification operation.

D. K-Nearest Neighbour Classification

k-NN classification is the simplest algorithm used to classify the users in particular category. So, recommendation system can provide the browsing option to user according to his interest. K-nn algorithm takes users previous logs as a input and find out which dat is accessed mostly and from that data it predicts the users interest. So, from that information it classifies the users incategory as per his interest and needs.

E. Recommendation System

Recommendation system performs online matching and recommendation operation. After classification performed by k-NN classifier, recommendation system checks the category of particular user and according to that category it provide the browing option to user as per his interst. There is one extra option also provided that is search by selecting category. If user want to search something different than his interest then he can select particular category and get result relevant to that category.

F. Data Store

Web logs and URL files are extracted, cleansed, formatted and grouped into meaningful session for data mining analysis and data mart is developed because the raw URL files and logs which are extracted, not well structured to be used directly for data mining. In designing the data mart, the process of data acquisition, extraction and storage is implemented using Structured Query language MySQL. The process of development of web usage mining and recommendation system is done by using java programming language with NetBeans as a editor and compiler.

G. Overview of System

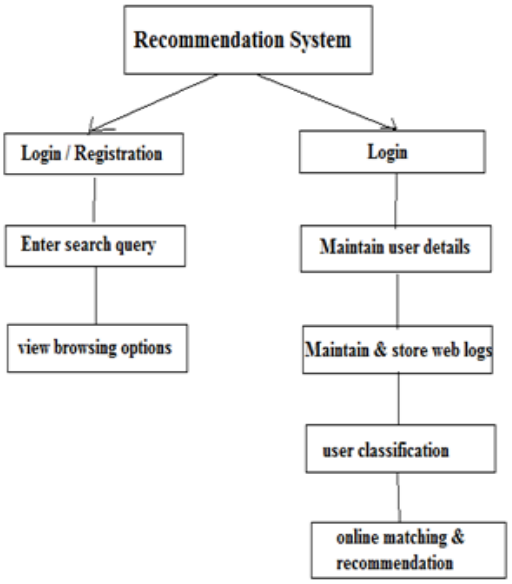

Fig. 2: Overview of System

III. Advantages

For finding desired information, customers don't have to search from large amount of information. They can get useful information very easily and in very less time.
This system is capable for producing useful and quite good and accurate classifications and recommendations to the client at any time based on his immediate requirement.

IV. Conclusion

The biggest problem of many websites is the presentation of many options to the client at a time. So, it's very difficult task for user to find required product or information. Web usage mining and recommendation system observes the current users click stream data and navigation patterns. It used the k-NN classification algorithm to classify users in different categories as per their access behavior and search pattern. This system provides the required information to the users based on his category and satisfy the user's needs. So, user can get desired information in very less time. It can assist the web designer and administrator to improve the content, presentation and impressiveness of their website by recommending a unique set of objects that satisfies the need of active user based on the user's current click stream. The beauty of proposed system is that it dynamically provides recommendations as per the user's interest, his changing behavior and traversal patterns by making use of web usage mining and constructing patterns from previous logs.

References

[1] D. A. Adeniyi, Z. Wai, Y. yongquon, " Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method" ,2014.
[2] Prajyoti Lopes, Bidisha Roy, " Dynamic Recommendation System using Web Usage Mining for E-commerce users.",2015.

[3] Ravneet Kaur , Sarbjeet Singh, "A survey of data mining and social network analysis based anomaly detection techniques", 2015
[4] Sumaiya Kabira,b, Shamim Ripona*, Mamunur Rahmanb and Tanjim Rahmanb, "Knowledge-Based Data Mining Using Semantic Web",2013.

