Online Recommendation System Based on Bitap algorithm

Miss.Pratima Manikrao Sukre¹, Dr.J.D.Pagare²
¹Department of Computer Science and Engineering, MGM’S Jawaharlal Nehru College, Aurangabad (MS), India
²Department of Computer Science and Engineering, MGM’S Jawaharlal Nehru Engineering College, Aurangabad (MS), India
¹lia.pratimasukre@gmail.com, „jaydp2002@yahoo.co.in

Abstract— In this work, we give an overview of the main Data Mining techniques used in the Recommender Systems. Online Recommendation and Prediction is widely used application of Web usage mining. We first describe common pre-processing methods such as data cleaning, session identification to employing web mining algorithm on web server logs. We review the most important clustering method including k means clustering algorithm is used. From last few years information flooding has created problem to web users. Web usage mining is mostly used by employing data mining technique. Which extract knowledge from web user’s access logs. Most of the recommendation systems on a content-based approach or collaborative approach to make recommendations. The objective behind WUM is to analyse web log files.

Keywords— Pattern matching, clustering, k-means, web usage mining, e-commerce review system, and knowledge based system, machine learning, recommendation system

I. INTRODUCTION

Recommender systems has been established in the early 1990’s. web information system E-commerce has grown enormously in terms of deployed application and also algorithmic developments. A recommender system can be defined as a system that guide users in a personalized way to interesting or useful product to user and determine user will approve given item. user’s attitudes and behaviour on the basis of the concepts of human computer interaction changes by Persuasive technologies. Recommendation system use for these Persuasive technologies, Recommendation systems can help database to deal successfully with the size and complexity of knowledge structures. For example, recommendation systems can support an improved understanding of the knowledge base by actively suggesting those parts of the knowledge base which are necessary to users for previous transaction. Furthermore, recommendation systems can propose repair for unwanted and ill-formed knowledge bases. In the context of information search, recommendation systems can improve the accessibility of databases (knowledge bases) by data pre-processing.

Practical experiences from the successful deployment of recommendation technologies in e-commerce contexts (e.g., amazon.com and netflix.com contributed to the development of recommenders in new application domains.

II. RELATED WORK

Since mid 90’s large number of comparison shopping agent of various types have developed in order assist consumers to make shopping decision in online environment. The first commercial shopping agent called Jango was produced by net bot, a start up company from Seattle founded by university of Washington.

In the early development from 1995 to 2000 comparison shopping agent include not only price comparison but also rating and review services for online suppliers and product websites like opinion.com provided reviews and rating services for product. As we ken that online shopping requires shipping fees for product distribution. It is expected that some consumers intention to purchase a particular product because they have to pay extra charges for the distribution accommodation. For example, some consumers who would prefer an expeditious distribution will have to pay higher cost while others may prefer to wait if they pay lower shipping and handling charges.

Hiral Y. Modi , Meera Narvekar [1] implemented an Online recommendation System. The system involved two phases that work in connected with each other i.e. the online and offline phase. Data pre-treatment and navigation pattern mining is carried out in offline phase while predictions are generated in the online phase. They also implemented the online and offline phase architecture.

Sarwar et al[2] assessed a clustering approach which groups users into clusters according to their similarities. They denoted that clustering provides comparable recommendation quality as traditional CF, which significantly improving the online performance. There are also other studies on clustering techniques for collaborative filtering.

Kamna Sharma, Taruna Sharma evaluated Recommender systems are classified into three categories [3][4] on the basis of how recommendations are made: 1. Content Based Recommender System 2. Collaborative Filtering Based Recommender System 3. Hybrid Recommender System

In content based method, the customer get recommended items according to similar preferences. The ecommerce system recommends items to customers based on correlation between the content of items and the user preferences.

Gediminas Adomavicius & Alexander Tuzhilin[5] said in collaborative approach, users get recommended items that people with similar tastes and preferences liked in the past.
The third category of recommender system combines both collaborative filtering and content based approaches.

III. METHODOLOGY

A. K-means clustering Algorithm

k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the most proximate mean, accommodating as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. K-designates is one of the simplest notices that the k centroids transmute their location step by step until no more changes are done. In other words centroids do not move any more. Unsupervised learning algorithms that solve the prominent clustering quandary. The main conception is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. Therefore the good choice is to place them as much as possible far away from each other.

The next step is to take each point belonging to a given data set and associate it to the most proximate centroid. When no point is pending, the first step is consummated and an early group age is done. At this point we require re-calculating k incipient centroids as barycentre of the clusters resulting from the anterior step. After we have these k new centroids, a new binding has to be done between the same data set points and the most proximate incipient centroid. A loop has been engendered. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more.

Determinately, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

\[ J = \sum_{j=1}^{k} \sum_{i=1}^{n} (x_{i,j} - c_{j})^{2} \]

Where, \([x_{i,j} - c_{j}]^{2}\) is a chosen distance measure between a data point \(x_{i,j}\) and the cluster centre \(c_{j}\) is an indicator of the distance of the data points from their respective cluster centres.

Working Steps of Algorithm:
1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the most proximate centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This engenders disunion of the objects into groups from which the metric to be minimized can be calculated.

B. Bitap Searching Algorithm

Pattern searching is an important problem in computer science. When we do probe for a string in notepad/word file or browser or database, pattern probing algorithms are habituated to show the search results. Bitap distinguishes itself from other well-kenned string probing algorithms in its natural mapping onto simple bitwise operations, as in the following modification of the above program. Notice that in this implementation, counter intuitively, each bit with value zero indicates a match, and each bit with value 1 indicates a non-match.

The same algorithm can be written with the intuitive semantics for 0 and 1, but in that case we must introduce another instruction into the inner loop to set \(R \leftarrow 1\). In this implementation, we take advantage of the fact that left-shifting a value shifts in zeros on the right, which is precisely the behaviour we need. Notice also that we require CHAR_MAX additional bitmasks in order to convert the \(text[i] \equiv pattern[k-1]\) condition in the general implementation into bitwise operations. Therefore, the bitap algorithm performs better when applied to inputs over smaller alphabets.

Example

\[ T \text{ is slightly confusing because you would normally number positions in the pattern from left to right:} \]
\[ 0 1 2 3 4 \]
\[ a b a b c \]

...whereas bits are normally numbered from right to left.

But writing the pattern backwards above the bits makes it clear:

\[ \text{bit: } 4 3 2 1 0 \]
\[ c b a a b \]
\[ T[a] = 1 1 0 1 0 \]
\[ c b a a b \]
\[ T[b] = 1 0 1 0 1 \]
\[ c b a a b \]
\[ T[c] = 0 1 1 1 1 \]
\[ c b a a b \]
\[ T[d] = 1 1 1 1 1 \]

Bit \(n\) of \(T[x]\) is 0 if \(x\) appears in position \(n\), or 1 if it does. Equivalently, you can cerebrate of this as saying that if the current character in the input string is \(x\), and you see a 0 in position \(n\) of \(T[x]\), then you can only possibly be matching the pattern if the match started \(n\) characters previously.

IV. PROPOSED SYSTEM

A. Offline Phase Architecture

This phase consists of two modules: data pre-processing and knowledge base of product. Pre-processing obtain the client session and generate important data into the information base.

1) Data Pre-processing:

Pre-processing functionality like session identification, data cleaning perform on the web logs which consist of information like URL, client IP address and employs the web mining algorithm on web server logs. In pre-processing
clustering also carried out. For identification groups of users who have similar preferences the k-means clustering is used.

![Diagram](image_url)

**Fig.1 Existing Architecture**

2) **Knowledge Base:**

In this stage extracted user session data is combined with features of the product after pre-processing knowledge base generated. It includes user id, item id.

B. **Online Phase:**

After logs in at the user recommendation engine checks the knowledge base for earlier transaction of user. The list of recommended products generate on the base of previous history of ratings.

1) **Generating Recommendation:**

The parameters like rating support to produce exact set of items from information base. We utilised the bitap pattern matching algorithm for generate recommendation. The algorithm begins by pre-computing set of bit mask containing one bit for each element of the pattern, then is able to do most of the work with bitwise operation which are extremely fast.

V. **CONCLUSIONS**

Bitap is one of the most important searching techniques used in recommendation system. In this paper we show that k-means clustering an important bitap algorithm can make accurate and scalable using a bitwise search strategy within cluster. Web usage mining is a tool for recommendation system.

ACKNOWLEDGMENT

It gives me immense pleasure to express my gratitude to Prof. Dr. J. D. Pagare Department of Computer Science and Engineering, MGM'S Jawaharlal Nehru Engineering College Aurangabad (MS), India, for her generous guidance, help and useful suggestions which enabled me to bring out this work in the present shape. I am extremely thankful to my guide, who was always ready to help me, and all the staff of Department of Computer Science and Engineering MGM'S Jawaharlal Nehru Engineering College Aurangabad (MS), for their stimulating guidance and continuous encouragement throughout the course of present work. I acknowledge my sincere thanks to Prof. Dr. Vijaya Musande, Head, Computer science & Engineering Department, MGM'S Jawaharlal Nehru Engineering College Aurangabad (MS) India, for providing me infrastructural facilities to work in, without which this work would not have been possible.

REFERENCES


[4] Gediminas Adomavicius, Member, IEEE, and, Alexander Tuzhiuln, Member, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 17, NO. 6, JUNE 2005


