Precognition of Users Web Browsing Behaviour

M. Trupthi,
Research Scholar, CSE Dept.,
JNTUH, Hyderabad.

Dr. Suresh Pabboju,
Professor &HoD, IT, CBIT,
Gandipet, Hyderabad.

***Abstract*: The rapid e-commerce growth has made both business community and customers face a new situation. Intense online competition and customer's option to choose from several alternatives made the business communities to realize the necessity of intelligent marketing strategies and relationship management. Business is going online and the competition as well. Everyorganization isexploiting the online method of business expansion. In order to provide an utmost online service, organization's website should be quick and relevant to its targeted clients.Hence every website must be efficient in terms of both time and relevance. Web mining is the use of data mining techniques to automatically discover and extract information from Web documents and services.Markov Model has been used in the existing system for predicting next web page from the user's navigationalbehaviour in the web-log.We propose an accelerator to a website which boosts up the website's performance in terms of time and relevance by letting it know what the client may visit in the next moment and making the information ready beforehand through Association Rule Mining concept. Association rules are important in data mining, particularly in analysing and predicting consumer behaviour.ARM sustains the efficiency and scalability problems. Several efficient algorithms have been proposed to generate item sets and to uncover association rules such as Apriori algorithm. Web page prediction can be efficiently performed using ARM. In this paper we are using ARM for predicting the website access behaviour.**

Index terms:Web Prediciton, Association Rule Mining, Apriori Algorithm.

I. INTRODUCTION

The astounding growth of web site over the 'World Wide Web' (WWW) has notonly raised many concerns but also opened a window of opportunity for organizations toanalyze the lifetime value of their customers, and also improve their cross marketingstrategies. As more and more organizations rely on the WWW to conduct business, thetraditional strategies and techniques for market analysis needs to be revisited. The newstrategies involve analyzing a large collection of unstructured data.Web mining is defined as the use of data mining techniques to automatically discover and extract information from Webdocuments and services. With the rapid growth of the World Wide Web, the study of modelling and predicting a user's access on a Web site has become more important.Web mining can be divided into three different types namely, Web usage mining, Web content mining and Web structure mining.Web usage mining is the process of extracting useful information from server logs e.g. use Web usage mining is the process of finding out what users are looking for on the internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications.The data over the web is collected in the form of server access logs that aregenerated by the interaction of clients with the web site and stored in the form oftransaction logs. The web servers in the form of server or access logs generallyautomatically store this information. Different genres of organizations can make use ofthis data by analyzing for respective purposes. Web Usage Mining involves determining the frequency of the page access by theclients and then finding the common traversal paths of the users. Long and convoluteduser access paths along with low use of a web page indicate that the web site is not laidout in an intuitive manner. With the help of this analysis, one can re-structure the web sitewith the navigation results. Some of the most used algorithms in this mining processinclude association rule generation, sequential pattern generation and clustering. Web Usage Mining is a four-step process. The first step is data collection, thesecond step is data pre-processing, the third step is pattern discovery and the last step ispattern analysis. The pre-processing stage involves cleaning of the click stream data andthe data is partitioned into a set of user transactions with their respective visits to the website. During the pattern discovery stage, the use of statistical, database and machinelearningalgorithms are performed on the transaction logs to find hidden patterns and thebehaviour of the users. In the final pattern analysis stage, the discovered patterns from theprior stage are further processed and filtered producing models that can be served as aninput to different visualization tools and report generation tools. The last stage performsthe pattern filtering, aggregation and characterization on the discovered patterns.The input for the Web Usage Mining process is a user session file, which isbasically a pre-processed file and consists of information such as who accessed the website and what pages were accessed and for how long with their respective order. This usersession file is first processed by removing outliers and irrelevant items from the rawserver logs, identifying genuine and unique users from the server log and finally keeping the meaningful transactions within a user session file. The organization of this paper is as follows. In Section II, we present the related work. In Section III, we introduce basicbackground on different prediction models used in this paper. In Section IV, we present our proposed ARM In Section VI, we exhibit the experiments and explain the results. In Section VII, we conclude this paper and outline some future research directions.

II. RELATED WORK

Predictive modelling is the process by which a model is created or chosen to try to best predict the probability of an outcome. In many cases the model is chosen on the basis of detection theory to try to guess the probability of an outcome given a set amount of input data, for example given an email determining how likely that it is spam. Models can use one or more classifiers in approach of determining the probability of a set of data belonging to another set. Nearly, any regression model can be used for prediction purposes. Broadly speaking, there are two classes of predictive models: parametric and non-parametric. A third class, semi-parametric models, includes features of both. Parametric models make "specific assumptions with regard to one or more of the population parameters that characterize the underlying distribution(s)" while non-parametric regressions make fewer assumptions than their parametric counter parts. There are other probabilistic models such as Markov models, ARM models which are used to improve the prediction accuracy. Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. Association rules are created by analyzing data for framing rules by using the criteria support and confidence to identify the most relevant information.

There are many ways to achieve page prediction for a website.Some of the models can be Markov model, Association Rule mining. Markov model is used to predict the next page based on the results of previous actions.Previous actions are generally the page set that is already been visited by the user.The main advantages of Markov model are its efficiency and performance in terms of model building and prediction time.Prediction is performed in constant time. But the major disadvantages of this model are input is not scalable.Moreover, Markov Modelis not efficient for bulk user sessions and also the specific order of Markov model cannot predict for a sessionthat was not observed in the training set since such session willhave zero probability of occurrence.

The existing system has few disadvantages in-order to overcome themwe propose a system that can be implemented using Association Rule Mining. Our idea of implementation includes steps of collecting the data, data preprocessing, generating data sets and prediction. This idea overcomes the disadvantages of scalability and handling bulky user sessions that are faced by the existing system.Researchers have used various prediction models including *k*-nearest neighbor (*k*NN) ANNs [5], [6], fuzzy inference[3], [4], SVMs [5], [6],Markov model [1], [5], and others.Mobasher*et al.* [2] use the ARM technique in WPP and propose the frequent item set graph to match an active user session with frequent item sets and predict the next page that user is likely to visit.

III.BACKGROUND

In this section, we present the necessary background of thewell-known prediction models that will serve the purpose. Wefirst present the *N*-gram representation of sessions. Next, webriefly present Markov model and brief idea of All-Kth Markov model.
After that, we present the ARM model.Finally, we explain the concept of ranking in Web prediction.

MARKOV MODEL

Markov model is used to predict the next action based on the result of previous actions. In Web prediction, the next page to be visited is predicted by the next action. The previous actions corresponds to the already been visited pages. In Web prediction, the K^{th}-order Markov model is the probability that a user will visit the k^{th} page provided that he has visited the ordered $k - 1$pages[8]. For example, in the second-order Markov model, prediction of the nextWeb page is computed based only on the two Web pages previously visited. The main advantages of Markov model are its efficiency and performance in terms of model building and prediction time. It can be easily shown that building the k^{th} order of Markov model is linear with the size of the training set. The key idea is to use an efficient data structure such as hash tables to build and keep track of each pattern along its probability. Prediction is performed in constant time because the running time of accessing an entry in a hash table is constant. Note that a specific order of Markov model cannotpredict for a session that was not observed in the training set since such session willhave zero probability.

All-K^{th} Markov Model

Low order Markov models are coupled with low accuracy. In many situations, first-order Markov models are not effective in predicting the user'sbrowsing behavior, since these models do not look deep into the recorded data in correctly discriminating the different observedpatterns. As a result, higher-order models are often used. But, these higher-order models have anumber of limitations such as high state-space complexity, reduced coverage, and sometimes even poor prediction accuracy. One simple method to overcome these problems is to train varying order Markov models and use all of them during the prediction time, as is done in the All-K^{th} Order Markov model.

To be mentioned, there are three schemes for pruning the states of the All-K^{th} order Markov model, called (i) support pruning (ii) confidence pruning (iii) error pruning.K^{th}order Markov model, handles the predictions by considering the last *K* actions performed by the user, resulting to a state-space that contains all possible consequences of *K* actions. For example, consider the problem of predicting the next page accessed by a user on a web site. The input data for building Markov models consists of web-sessions, where each session consists of the sequence of the pages accessed by the user during his/her visit to the site. In this problem, the actions for the Markov model correspond to the different pages in the web site, and the states are with respect to all consecutive pages of length K that were observed in the different sessions. In the case of first-order models, the states will correspond to single pages, in the case of second-order models, the states will correspond to all pairs of consecutive pages, and so on. Major drawbacks are, the number of states used in these models tend to rise exponentially as the order of the model increases.This is because, the states of higher-order models are different combinations of the actions observedin the input observed. The increase in the number of states can significantly limit the usage of Markov models for applications in which fast predictions are critical for run-time performance or in applicationsin which the memory specifications are rigid.

Association Rule Mining (ARM)

ARM is a data mining technique that has been applied successfully to discover related transactions and been extensively used in prediction purposes. In ARM, relationshipsamong item sets are discovered based on their co occurrence in the transactions. Specifically, ARM focuses onassociations among frequent item sets. For example, in a supermarketstore, ARM helps uncover items purchased togetherwhich can be utilized for shelving and ordering processes. Inthe following, we briefly present how we apply ARM in WPP.For more details and background about ARM, see [3] and [5]In WPP, prediction is conducted according to the associationrules that satisfy certain support and confidence as follows.For each rule, $R = X \rightarrow Y$, of the implication, X is the usersession and Y denotes the target destination page. Prediction isresolved as follows:

$prediction(X \rightarrow Y) = \arg\max_Y supp(X \cup Y)/supp(X)$,$X \cap Y = \varphi$.

Note that the cardinality of Y can be greater than one, i.e.,prediction can resolve to more than one page. Moreover, settingthe minimum support plays an important role in deciding aprediction. In order to mitigate the problem of no support for
$X \cup Y$, we can compute prediction($X\_ \rightarrow Y$), where $X\_$ is theitem set of the original session after trimming the first page in
the session. This process is very similar to the all-K^{th} Markovmodel. However, unlike in the all-K^{th} Markov model, in ARM,we do not generate several models for each separate N-gram. Inthe following sections, we will refer to this process as all-K^{th}ARM model.Several efficient algorithms have been proposed to generateitem sets and to uncover association rules such as Artificial Immune System (AIS) algorithm.The rule mining process is applied to extract the patterns. The Apriori algorithm is used in the rulemining process. The patterns are identified from the item set collections. Support and confidence ratio are important parameters in the prediction process.The major benefits of this method are faster execution and lower memory utilization.

IV.IMPLEMENTATION

Our method of executing ARM concept involves certain steps depicted in the figure 4.1.

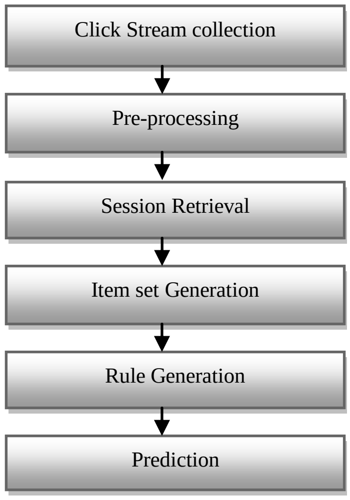

Figure 4.1 Steps involved

A. Server Logs

A server log is a log file which is automatically created and maintained by a server of activity performed by it. There are many formats of tracking the logs of a website such as W3C extended log file format which is customizable logging format Centralized logging format where logs are collectively stored for a bunch of websites and common log format which is a basic logging format and it is not supported for File Transfer Protocol(FTP) sites.A typical example is web server log which maintains a history of page requests. More recent entries are typically appended to the end of the file. Information about the request including Client IP address, request date, time, page requested, HTTP code, bytes served, user agent, and referrer are typically added. The attributes we used are Server logs consists of 6 attributes. They are a)IP address, b)Date, c)Method d) uri-stem e) Status and f) Bytes.

a) IP Address: IP is the number of computer who access or request the site.

b) Date: The format of date is DD-MM-YYYY. It also includes time of transactions. The format of format of time is HH:MM:SS. Example 29-jan-2014 15:15:54.

c) Method: The word request refers to an image, movie, sound, pdf, txt, HTML file and more It is also important to note that the full path name from the document root. The GET in front of the path name specifies the way in which the server sends the requested information. Currently, there are three formats that Web servers send information in GET, POST, and Head.

d) Uri-Stem: URI-Stem is path from the host. It represents the structure of the websites. (eg: http://www.cbit.ac.in/itdept)

e) Status: This is the status code returned by the server. By definition this will be the three digit number . There are four classes of codes: i) Success (200 Series) ii) Redirect (300 Series) iii) Failure (400 Series) iv) Server Error (500 Series).The most common failure codes are 401 (failed authentication) and the dreaded 404 (file not found) messages. A status code 502 means Bad Gateway.

f) Bytes: Bytes field is the number of bytes that have been returned to the user.(eg: 4325)

B. Preprocessing

An entry of Web server log contains the time stamp of a traversal from a source to a target page, the IP address of the originating host, the type of request (GET and POST) and other data. Many entries that are considered uninteresting for mining were removed from the data files. The filtering is an application dependent. While in most cases accesses to embedded content such as image and scripts are filtered out. However, before applying data mining algorithm, data pre-processing must be performed to convert the raw data into data abstraction necessary for the further processing. Server logs which have status code as 404 and 502 are removed and logs containing bytes value as zero are removed during pre-processing.All the urls that are used in server logs are stored in database and each url is given an id.

C. Sessions Generation

Sessions are can be generated using Pivot operator. Pivot queries involves in transposing columns into rows in order to generate results in crosstab format. Pivoting is a common technique, especially for reporting, and it has been possible to generate pivoted result sets with oracle.We executed pivot query through a java program and stored them as data file. Here pivot operator takes data about session_id, url_name, url_num, aggregates them and converts url_name, url_num into columns. All the urls are stored with their alias name in the table. In each row urls that are visited by the user in that session are stored. Figure 4.2 depcits the session Generation process.

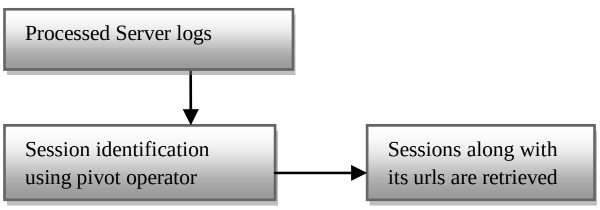

Figure4.2 Sessions Generation

D. Item Set Generation

Item sets are a set of items or a group of elements that represents together a single entity. Item sets are generated after executing apriori algorithm. The generated item sets are stored in database in the certain tables. We used certain columns to store the item sets such as id, sup, and num. Here the column id represents the item set number, sup represents support for the corresponding item set and num represents item set value.Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation). Each transaction is seen as a set of items. Support is defined as the proportion of transactions in the data set which contain the data set. Here we have given support value as 0.2.Apriori algorithm proceeds as follows

C_k: Candidate itemset of size k

L_k : frequent itemset of size k

L_1 = {frequent items};

for (k = 1; L_k !=NULL; k++) do begin

C_{k+1} = candidates generated from L_k;

for each transaction t in database do

increment the count of all candidates in C_{k+1} that are contained in t

L_{k+1} = candidates in C_{k+1} with min_support

end

returnL_k;

E. Rules Generation

An association rule is a pattern that states when X occurs, Y occurs with certain probability.Confidenceis percentage of transactions that contain X also contain Y.Confidence = Probability(Y|X).We generated rules using ARM based on confidence pruning.It involves two major steps namely extraction and calculation. For each item set in the data base we generated non empty sub sets of size one and two. For each non empty sub set called antecedent, we calculated the confidence by dividing with its consequent and if the value is satisfies the predefined confidence value.We here, generated 4antecedent-2consequent rules [A,B,C,D->E,F] ,3antecedent-1consequent rules [A,B,C->D] ,2 antecedent-1consequent rules [A,B->C] and also 1antecedent-1consequent rules [A->B].Prediction can be computed as

$$Prediction(X \rightarrow Y) = \text{argmax}Y\text{supp}(X \cup Y)/\text{supp}(X)\ ,X \cap Y = \varphi.$$

F. Prediction

The framed rules can be stored in any persistent form.Predicition can be done as follows. Firstly, track all the pages that have been visited till the moment using some data structure,simply an array. Then use each url and search it for it in the antecedent side of the [A,B->C] and if found ,we can retrieve the other antecedent and check for it in the current set of pages that have been found. If it is found, the antecedent side of the rule is matched and hence precedent side of the rule can be the predicted set of pages. Likewise we searched for a match of the antecedent side of all kinds of rules that have been generated and predicted the precedent side of the rules.

V. EVALUATION

We implemented the ARM predictionmodels. In ARM, we generated the rules using the apriori algorithm proposed.Testing is the process of assessing how well your mining models perform against real data. Testing is important for any prediction model because it determines how the system behaves in the real time application.

To measure the accuracy, we followed thegeneralization accuracy procedure by partitioning each data setrandomly into a training set (two-thirds of the original set) anda testing set (one-third of the original set). The generalizationaccuracy is a standard procedure which is widely used to measureprediction models accuracy against new examples thatmight not have been observed during training.The number of training instances has a direct effect on the classifying ability of themodel built from that number of instances. When there is a limited amount of data, n fold cross-validation is the best way to maximize theuse of available data to produce a good classifier. In n fold cross-validation, the datais divided into n folds, and each fold in turn is used for testing, while the other foldsare used for training. The reported accuracy is the average over the n iterations oftraining and testing.Preparation of testing data from the training data plays a vital role in determining the accuracy.

Accuracy is a measure of how well the model correlates an outcome with the attributes in the data that has been provided. There are various measures of accuracy, but all measures of accuracy are dependent on the data that is used. In reality, values might be missing or approximate, or the data might have been changed by multiple processes. Particularly in the phase of exploration, we have to decide a certain amount of error in the data, especially if the data is fairly not uniform in its characteristics.

Use various measures of statistical validity to determine whether there are problems in the data or in the model.Separate the data into training and testing sets to test the accuracy of predictions.Ask business experts to review the results of the data mining model to determine whether the discovered patterns have meaning in the targeted business scenario.

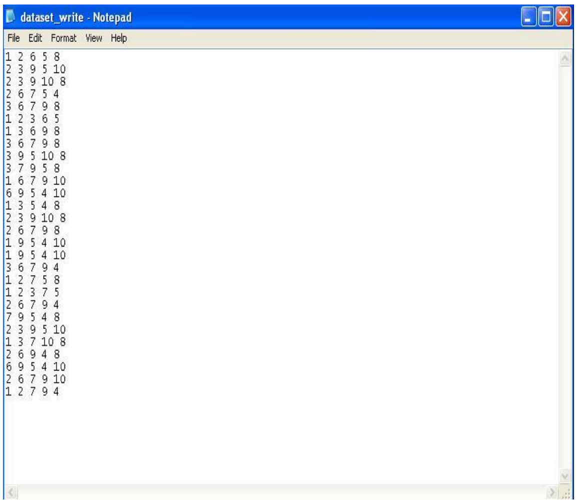
dataset_write - Notepad
File Edit Format View Help
1 2 6 5 8
2 3 9 5 10
2 3 9 10 8
2 6 7 5 4
3 6 7 9 8
1 2 3 6 5
1 3 6 9 8
3 6 7 9 8
3 9 5 10 8
3 7 9 5 8
1 6 7 9 10
6 9 5 4 10
1 3 5 4 8
2 3 9 10 8
2 6 7 9 8
1 9 5 4 10
1 9 5 4 10
3 6 7 9 4
1 2 7 5 8
1 2 3 7 5
2 6 7 9 4
7 9 5 4 8
2 3 9 5 10
1 3 7 10 8
2 6 9 4 8
6 9 5 4 10
2 6 7 9 10
1 2 7 9 4

Figure5.1 Sample Dataset

Accuracy can be improved by using bagging and boosting techniques. Bagging and boosting methods are used to determine testing data from the training data. In Bagging, test data is created by taking samples from the available data with replacement. Whereas in boosting, testing data is prepared by selecting the samples from collected data without replacement. Boosting avoids over fitting issue which is a major disadvantage for the bagging method. Boosting often leads to a dramatic improvement in prediction.

While generating the item sets, support value plays an important role. For example, for the dataset considered as shown in figure 5.1.

If the support value is changed, the no. of frequent item sets generated will be different for the above considered dataset. The variations are clearly depicted in the Figure 5.2.

For the dataset considered, if the support value is slightly changed, then the no. of item sets generated is not affected. And the count of item sets generated is drastically reduced when the support is value is kept very high. Support value indirectly affects the quality of rules that are generated.

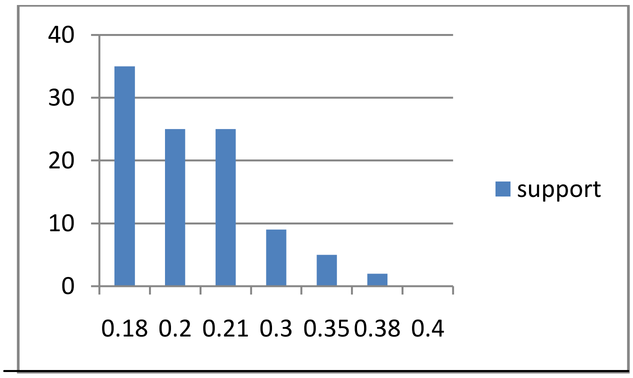

Figure5.2 Support-itemsets variation

Fixing confidence to a particular value is an important step while generating the rules. Confidence value decides the quality and number of rules that are to be generated.Figure 5.3 shows the rules generated for different confidence values for the dataset that has been considered in the figure 5.1 and by fixing support value to 0.2 while generating the itemsets.

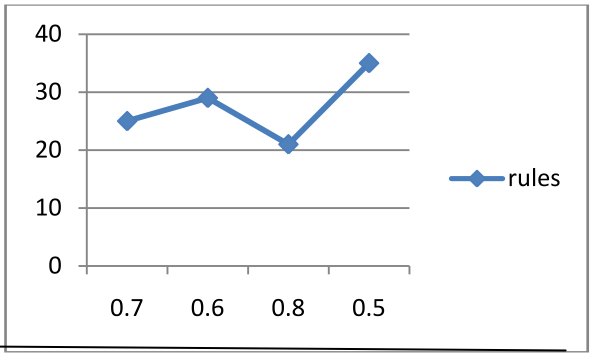

Figure5.3 Confidence-Rules Variation

Accuracy is also affected by varying the matching length criterion. The matching length criterion selects the applicable rule of the longest match antecedent of the rule. A prediction model can be constructed by pairing any antecedent type with any rule-selection criterion. Matching length is to what length the current user pattern is mapped to the existing patterns.If matching length is more, the quality of prediction will be very high.
Table 1 shows the different combinations of accuracy, percentage of training data and matching length parameter.

TABLE 1. Accuracy variation under different conditions

Training%	**15**	**25**	**25**	**30**	**30**
Matching length	**3**	**3**	**2**	**3**	**4**
Accuracy%	**20**	**25**	**22**	**35**	**40**

The number ofbrowsed pages also has an effect on the prediction accuracy. Specifically, weanalyze the effect of the sacristy of pages in the data set.We run this experiment by, first, fixing the data set size, and then randomly picking a set of *P* pages after removing from thesessions any page that is not in the random set. We repeatedthat using different *P* pages, different data set sizes, and differentrankings. The figure shows that smaller number of pagesimplies high scarcity; hence, lack in the knowledge neededduring prediction results in lower accuracy. On the other hand,when the number of pages increases, the amount of experiencesgrows, and according to that, accuracy augments. For example, considering rank-5 curve in Figure 5.4, the prediction accuracies of 200-, 600-, and 900-pageWeb sites are 58%, 68%, and 73%, respectively.

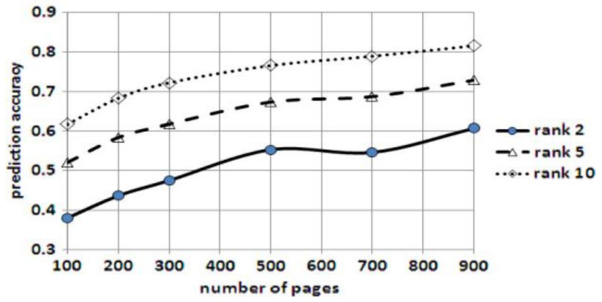

Figure 5.4 Effect of the number of pages in the Web site on accuracy

Figure 5.4 shows how the system behaves when the confidence values while generating the rules.Higher confidence values will definitely improve the quality of prediction but sometimes makes the system fail in simpler cases where user input pattern is small. Click stream data collected from the third parties and logs from other servers should also be tested on the system to improve its accuracy. Evaluation of the system built may vary with the datasets considered and values for parameters such as support,confidence, matching length set in the Association Rule Mining. Server logs have to be collected periodically and should be run on the system to keep it accordingly to the website trends.

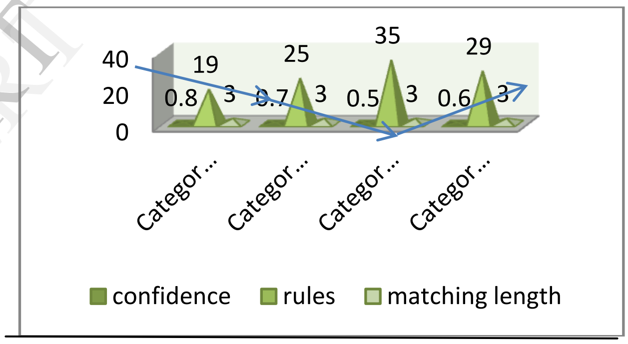

Figure5.4 Accuracy variation with confidence

V1. CONCLUSION

Web Prediction is a classification problem in which we attempt to predict the next set of web pages that a user may visit based on the knowledge of the previously visited pages. Such knowledge of user's history of navigation within a period of time is referred to as a session. These sessions, which provide the source of data for training, are extracted from the logs of the Web servers, and they contain sequences of pages that users have visited along with the visit date and duration.Server logs are created which includes fields like date, IP address, method, status, bytes, and uri-stem. Pre processing of the server logs is done which removes logs with status 404 or 502 and bytes zero. All the schemas are stored in database. Sessions are generated using pivot operator and schemas for item sets are stored in database. Item sets are generated using Apriori algorithm. Rules are generated using ARM. In the front end when the user selects the urls , pages that a user is going to visit are predicted. Static web sites require this technique because the information is distributive and cannot be retrieved instantly. Dynamic websites generally responds to user's request and process them and get backs the result. So dynamic websites are much stationary.Web Precognition serves the purpose of the website by anticipating user's visiting pages which results in less retrieval time of the requested page.

Prediction of user's surfing patters for a particular website has obtained lot of research scope these days. As we can improve the web site performance particularly cache performance,page recommendations, we can trace out the buying patterns as based on user-centric clickstream data and hence personalize the browsing experience,modeling user web navigation information is a upcoming field in the web mining domain as the size of the web and its user-database is increasing.

To the existing implementation of the precognition of user's web behaviour,some enhancements can be added which improves the result quality as well as the system performance. Apart from the approach we used here to implement the idea of prediction, there can be some more criterions that can be included in different steps of our project.Server logs can be collected from parallel browsing behaviour[7] which might have a little effect on the results. But choosing a wider browser plug-in for collected the logs might land us in security breach issues.Click stream data can be acquired from third parties for a particular site, apart from the way of creating the manually.

Click stream data could be user centric or site centric. In addition to within-site browsing behavior, a number of studies have employed user-centric clickstream data to investigate browsing and search across multiple websites. One of the advantages of user-centric data is that visits to multiple websites are recorded for each user. This opens the possibility that what users do at one website might help predict behavior at competing or complementary websites. In future,we can implement the same model for user-centric data with bringing appropriate changes to the implementation. For constructing the sessions, we can use product oriented events or by page link structures apart from the time heuristics that has been used in our current system. Page view duration along with page views might bring qualitative sessions which can be another improvement.While generating user-independent association rules, we can add some other criterion to improve the association rule results. Matching length criterion selects the applicable rule of the longest match antecedent is one such criterion.And also while considering antecedents for constructing rules, we can have different types of them namely subset, subsequence antecedents.And also an another enhancement that can be done is apart from including support and confidence for generating itemsets and for generating rules respectively we can introduce a new parameter pessimistic estimate for generating more strong and realistic rules.The mentioned improvements can be completely incorporated to the implemented model by introducing appropriate changes as required without much difficulty.As this field is gaining considerable importance,lot more enhancements might be suggested in future to serve the growing needs.

REFERENCES

[1] J. Griffioen and R. Appleton, "Reducing file system latency using apredictive approach," in Proc. Summer USENIX Tech. Conf., Cambridge,MA, 1994.
[2] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Effective personalizationbased on association rule discovery from Web usage data," in Proc.ACM Workshop WIDM, Atlanta, GA, Nov. 2001.
[3] O. Nasraoui and C. Petenes, "Combining Web usage mining andfuzzy inference for Website personalization," in Proc. WebKDD, 2003,pp. 37–46.
[4] O. Nasraoui and R. Krishnapuram, "One step evolutionary mining of contextsensitive associations and Web navigation patterns," in Proc. SIAMInt. Conf. Data Mining, Arlington,VA, Apr. 2002, pp. 531–547.
[5] M. Awad, L. Khan, and B. Thuraisingham, "Predicting WWW surfingusing multiple evidence combination,VLDB J., vol. 17, no. 3, pp. 401–417, May 2008.
[6] M. Awad and L. Khan, "Web navigation prediction using multipleevidence combination and domain knowledge," IEEE Trans. Syst., Man,Cybern. A, Syst., Humans, vol. 37, no. 6, pp. 1054–1062, Nov. 2007.
[7] Mamoun A. Awad and Issa Khalil, "Prediction of User's Web-Browsing Behavior:Application of Markov Model," IEEE Trans. Syst., Man,Cybern. A, Syst., Humans, vol. 37, no. 6, pp. 1054–1062,Aug.2012.
[8] Sampath P and Ramya D "Performance Analysis of Web Page Prediction with Markov Model, Association Rule Mining(Arm) and Association RuleMining With Statistical Features(Arm-Sf)**"**, IOSR Journal of Computer Engineering (IOSR-JCE)e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 8, Issue 5 (Jan. - Feb. 2013), PP 70-74www.iosrjournals.org 70 | Page.