Mining fuzzy association rules for web access case adaptation

Cody Wong, Simon Shiu
Department of Computing
Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong, China
{cskpwong; csckshiu}@comp.polyu.edu.hk

Sankar Pal
Distinguished Scientist & Head
Machine Intelligence Unit
Indian Statistical Institute
Calcutta, India
sankar@isical.ac.in

Abstract

Web access path prediction using knowledge discovered from web logs has become an active research area. Web logs provide updated information about the user's access record to a web site, which contains useful patterns waiting to be discovered and used for improving the web site. In this study, a new approach to web access pattern prediction is proposed. The methodology is based on the case-based reasoning approach, and the main idea is to discover user access patterns by mining the fuzzy association rules from the historical web log data. In our approach, the time duration of each user session is also considered as one of the attributes of a web access case. A fuzzy index tree is used for fast matching of rules. Furthermore, the system's performance is enhanced using the information contained in the user profile through an adaptation process.

1. Introduction

Web mining can be broadly defined as the discovery and analysis of useful information from the World Wide Web [1]. It can be further divided into web content mining and web usage mining. The growth of web-based business to business electronic commerce has shown the necessity of web usage mining.

Understanding user's preference, pre-caching web material, personalizing web interface and message bring the web site closer to the user: success in these areas could increase the value of the web site and the possibility of staying alive in the competitive business environment. Tracing, recording and discovering the relationships of user browsing behaviors is possible today by making use of the information captured in web servers and application servers. However, how to use and adapt this historical information to personalize the user interface and recommend interesting web pages is an issue of critical importance.

In this study, we propose a new approach to improve the accuracy and efficiency of web access path prediction by the technique of fuzzy association rule mining and fuzzy index tree generation using a case-based reasoning (CBR) approach. Furthermore, we propose a framework for adaptation-guided web personalization. Making use of knowledge discovered from user profiles and web access logs, an integrated case-base with larger coverage of user information is proposed. Fuzzy rules generated from the case base can be used for prediction and recommendation. An adaptation engine is used to learn the accuracy of the prediction of each rule and to do the adaptation when the adopted rule can out-perform the original one. The intelligent adaptation of rule selection based on the actual performance of rules can result in better prediction accuracy and a more customized recommendation.

This report is organized as follows. In Section 2, we review the related work and current state of the art in web mining. In Section 3, the architecture of our system is presented. In Section 4, the steps for data pre-processing will be given. In Section 5, the structure of a case used in this study will be defined. In Section 6, the algorithm of fuzzy association rule mining will be discussed. In Section 7, the procedures of the fuzzy index tree are described. In Section 8, the experimental design and results of testing the fuzzy association rule are reported. Finally, our work is concluded and extended to an adaptation-guided web personalization system as future work.

2. Literature review

Tracing a client's behavior, and predicting future access patterns and user preferences on the World Wide Web are not new issues to machine learning or AI researchers. One of the earliest systems was WebWatcher [3], which was designed to recommend to the user any future accessing hyperlinks that he or she might be interested in. W^3IQ is another recommendation and personalization system that is based on the concept of the cooperative information retrieval technique [4]. Researchers like Jose Borges and Mark Levene used hypertext probabilistic grammar to model user navigation records. They assumed that strings generated with higher probability should correspond to the user's preference [5]. The Web Utilization Miner (WUM) [6] was another tool used to mine access patterns with "interesting" statistical properties.

Mining typical user profiles for web personalization from the vast amount of historical data that are collected by web servers (e.g. access logs) has been proposed by many researchers [1,2,7,8,9] and different researchers have applied this technique. Krishnapuram has proposed to form user profiles showing the interests of a group of people by clustering the web logs based on URL similarities using different fuzzy clustering algorithms [7,8]. Some researchers try to mine the usage pattern using OLAP as the tool and mapping the web log into a snowflake schema [10]. The Least Recently Used (LRU) concept has been adopted by integrating it with a statistical model such as the Taylor Series [11]. Sometimes the tree structure is transformed and combined with the concept of support, and can confidently form the prediction models [12]. Some researchers represent the access path by an aggregate tree to predict the access pattern by tracing path [13].

However, the above approaches have one common problem – they ignore the sequential characteristics and association relations of the web log. Researchers like Cooley [2,15] used different association rule algorithms to find the patterns and relationships between access sequences and user sessions. Qiang Yang and his group [9,14] introduced the concept of n-gram sequence models on top of the traditional association rule algorithms. One of the problems of association rule techniques is that each item is distinct from the others, therefore any two items are either considered to be the same or are totally different. Another problem of association rule techniques is that, when applied in a large data set, the size of the resultant rule set will be extremely large. Therefore, in a real-time application system, to match a rule with input query becomes a critical issue.

Taking the sequential property of the web log into consideration, the fuzzy association rule is introduced to discover fuzzy rules to do the access path prediction. Since the concept of preference and interest are fuzzy, we consider the interest in a certain web page as the amount of time taken in browsing its content. Duration of page view is selected as the fuzzy set attribute for association rules mining. The fuzzy association rule technique is then integrated with CBR methodology by defining the problem space and the solution space using some attributes. To improve the efficiency of rule matching, the fuzzy index tree technique is employed. Furthermore, an adaptation-guided web personalization system is proposed by combining the knowledge discovered from the web log and user profiles.

3. System Architecture

A web access path prediction system is divided into two components: offline batch processing and online access path prediction (Figure 1). The offline batch processing is to produce the prediction rules for the prediction engine. The offline processing is mainly responsible for data cleaning, case base generation and rule discovery. These include removing the irrelevant components from the web log, identifying user transactions, generating case bases, discovering fuzzy association rules and generating a fuzzy index tree. The discovered rules and the generated index tree are passed to the prediction engine.

The online component includes a web server and a prediction engine. The web server is used to capture the active user session and pass the access path to the prediction engine. The prediction engine searches for the rule using the fuzzy index tree and matches the user session with the rules to predict the future access path. The predicted path is sent to the web server and the predicted pages to the client browser or the proxy for caching.

4. Data Cleaning and Pre-processing

The World Wide Web (or Internet) is a huge client server infrastructure where one of the useful pieces of information captured on the server side is the navigation of different clients within various web sites. Clients make requests for hypertext to the web server through the HTTP protocol. The web server's responsibility is to respond to such requests and deliver the required contents to clients. Although the information might be affected by the caching problem of the proxy server and client browser, the web log is still a useful source of user access records.

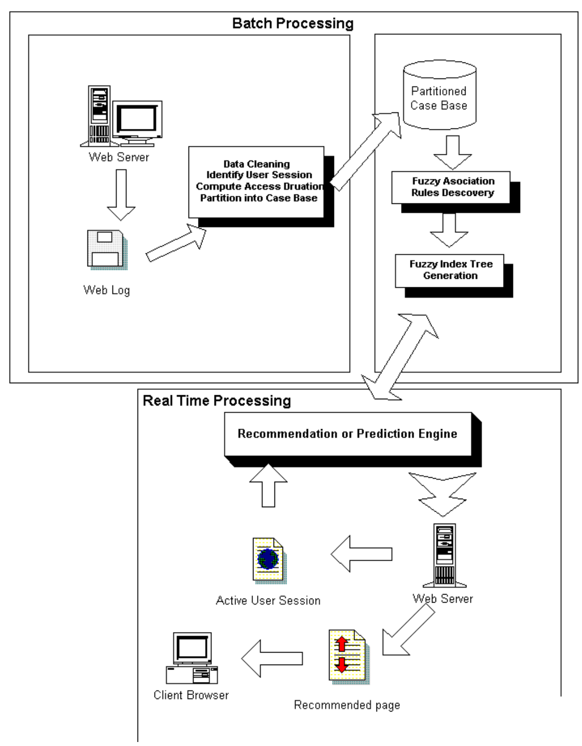

Figure 1. System Architecture of the Prediction System

A web log usually consists of seven parts: i. User's IP address, ii. Access time, iii. Request method ("GET", "POST", etc), iv. URL of the page required, v. Data transmission protocol (typically HTTP/1.0), vi. Return code, vii. Number of bytes transmitted. In this research study, we focus mainly on the first, second and fourth part, (i.e. the user's IP address, the access time, and the URL of the page required). The term "user session" is defined as the accesses from the same IP address such that the duration of elapsed time between consecutive accesses in the session is within a pre-specified threshold [7]. The frequency of user session that used to find the frequent itemset for fuzzy association rule mining is defined as the number of repeated sequent of assesses in the web log data.

Cleaning and preprocessing of the web log is carried out to eliminate irrelevant items, such as removing those URLs that require image files or multimedia files (e.g. jpg, gif, swf, etc). In addition, requests using methods other than "GET", or URLs that contain special characters are also removed because these requests are usually searching queries or requests for certain CGI programs. After the above removal, the log data are partitioned into user sessions based on the IP and duration. Less frequent user sessions are also filtered out. The URLs are mapped to unique numbers for recording.

For each access in a user session, we assign the term "duration" which is the time between two successive accesses, and the "nil" value is assigned to the last access in the user session. U is defined as the set of "User Transactions" that has met the minimum frequency of certain threshold.

$$U = \{U_1, U_2, \ldots U_n\}$$

The term "User Transaction" is defined by a set of URL accesses associated with the duration and bounded by the length of the user session.

$$U_n = \{ [URL_1,d_1], [URL_2,d_2] \ldots . [URL_n,d_n]\}$$

where URL_n is the n unique URL and d_n is the associated duration.

5. Case Definition

It is important to define a case in applying the case-based reasoning methodology. Assuming the access pattern of a certain type of user can be characterized by the length of user transaction, and that the corresponding future access path is not only related to the last accessed URL. Therefore, users with relatively short transactions (e.g. 2-3 accesses per transaction) should be handled in a different way from users with long transactions (e.g. 10-15 accesses per session). In this study we propose a case definition schema based on the transaction length. The set of user transactions, U, is discretized into N partitions by a 1-D linear clustering base on the number of accesses per transaction. User transactions with lengths of less than 3 are removed (i.e. the length of transaction is too short to provide sufficient information for access path prediction). For partitioning sets of user transactions, U, problem and solution cases are defined based on the following rules:

1. For the first partition, the problem part of a case is defined as the URL of the first two accesses, and the solution part is the remaining URLs.
2. For the other partitions, the problem part of the case is defined as the cardinality of the upper bound of the previous partition, and the solution part is the remaining accessed URLs.
3. For the other partitions, the problem part of a case is defined as the cardinality of the upper bound of the previous partition, and the solution part is the remaining accessed URLs.

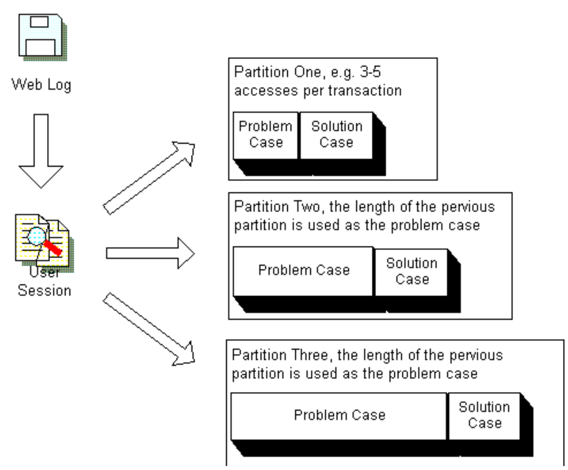

Figure 2. Case Definition

Problem Description					Solution Case
URLs	1st	2nd	3rd	4th	Future Access Path
Cases in Partition 1	1	3	*	*	8
Cases in Partition 2	1	3	8	*	10,7

Table 1. Case representation in different partitions

For example, if a transaction in partition 1 is {1,3,8}, then the problem part of this case is defined as {1,3} and {8} is the solution part of this case based on rule 1. Again, if a transaction in partition 2 is {1,3,8,10,7}, then the problem part of this case is defined as {1,3,8} and {10,7} is the solution part of this case, (i.e. the cardinality of the upper bound of the previous partition of the accessed URL is equal to 3). In this study, the cardinality of the first partition is defined based on the result of experiment.

6. Fuzzy Association Rule Discovery

6.1 Fuzzifying Process

As stated in section 2, the relation between the user's successive access path is always regarded as a sequential association relation. Discovering the association rule and sequential pattern by using the concepts of minimum support and confidence is widely applied. However, the accuracy of prediction using these approaches is not much better than other algorithms. One of the possible reasons for this is the lack of a fuzzy concept in the rule. For example, there are two rules: "If A and B then C" with support equal to 80% and confidence equal to 50%, and "If A and B then D" with support equal to 79% and confidence equal to 49%. If there is an active user session with the first two accesses equal to A and B, then the first rule is chosen. However, we can see that the second rule is as strong as the first rule. Therefore, there should be an equal choice for following either the first and second rule for finding a correct match.

In order to make rules more sensible and reasonable, we introduce a new fuzzy attribute into the user transaction, the duration of each access. Duration is defined as the time interval between two successive accesses, and is normalized to a number between 0 and 1. In order to discover the fuzzy association rules, numeric duration is fuzzified into five linguistic terms: Long, Quite Long, Medium, Quite Short, Short and represented by L, QL, M, QS, S.

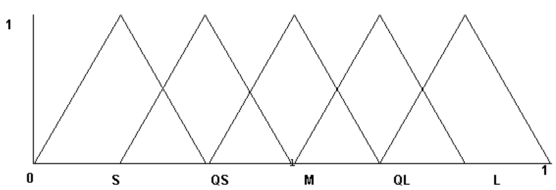

Figure 3. Fuzzy Membership Function

Let U be a user transaction in partition one with three accesses.

Access	Duration	S	QS	M	QL	L
First	0.7	0	0	0	0.8	0.2
Second	0.3	0.6	0.4	0	0	0
Third	0.5	0	0	1	0	0

Table 2: Fuzzify the user transactions based on duration.

6.2 Fuzzy Association Rule

We apply the fuzzy association rule algorithms proposed by A. Gyenesei [16]. Given a case base $C_s = \{u_1, u_2, u_n\}$ where u_i is a user transaction, with access sequence $S = \{a_1, a_2, \ldots, a_n\}$ and the fuzzy sets associated with each access a in S. The rules discovered will be in IF-THEN format.

We define the following form for the fuzzy association rule:

If $X = \{x_1, x_2, \ldots, x_n\}$ is $A = \{f_1, f_2, \ldots, f_n\}$
then $Y = \{y_1, y_2, \ldots, y_n\}$ is $B = \{g_1, g_2, \ldots, g_n\}$

where X is the sequence of URLs accessed, A is the associated fuzzy set, and Y is the sequence of URLs predicted and B is the associated fuzzy set. X is the problem case of u_i in C, and Y is the corresponding solution case of ui, since they are disjointed, i.e. they do not share common attributes.

A and B contain the fuzzy sets associated with the corresponding attributes in X and Y. As in the binary association rule, "X is A" is called the antecedent of the rule, while "Y is B" is called the consequent of the rule. Such pair of antecedent and consequent are define as a record. For an interesting rule, it should have enough support and a high confidence value.

6.3 Fuzzy Support Value

To mine the fuzzy association rule, we first find out all sets of items that have transaction support above certain threshold. Itemsets with minimum support are called frequent itemsets. The fuzzy support value is computed by first summing all the votes of each record with respect to the specified itemsets, then dividing it by the total number of records. Each record contributes a vote, which falls in [0,1].

A fuzzy support value reflects not only the number of records supporting the itemset, but also their degree of support. We adopted the formula proposed by Attila Gyenesei [16] to calculate the fuzzy support value of itemsets of <X,A>, as follows:

$$FS_{<X,A>} = \frac{\sum_{ui \in U \pi xj \in X} (a_j \in A, t_i \cdot x_j)}{|Cs|}$$

We use an example to illustrate the computation of the fuzzy support value. Let X ={URL1, URL2} and A {Long, Medium} and some cases of partition one show in table 3. The fuzzy significance of <X,A> is as follows:

$FS_{<X,A>} = (0.5x0.8 + 0.6x0.6 + 0.4x0.8 + 0.7x0.2 + 0.5x0.6 + 0.2x0.4 + 0.9x0.1) / 7 = 0.241$

<Duration of URL1,Long>	<Duration of URL2,Medium>
0.5	0.8
0.6	0.6
0.4	0.8
0.7	0.2
0.5	0.6
0.2	0.4
0.9	0.1

Table 3: Part of cases in partition one containing membership

6.4 Fuzzy Confidence Value

The discovered frequent itemsets are used to generate all possible rules. If the union of antecedent and consequent has enough support and the rule has high confidence, this rule is considered interesting. When we obtain a frequent itemset <X,A>, we want to generate fuzzy association rules of the form, "If X is A then Y is B", where $X \subset Z$, $Y = Z - X$, $A \subset C$ and $B = C - A$.

Having discovered the frequent itemset, the support is known and all subsets of the frequent itemset can also be identified. We can calculate the fuzzy confidence value as follows:

$$FC_{<<X,A>,<Y,B>>} = \frac{FS_{<Z,C>}}{FS_{<X,A>}}$$

$$= \frac{\sum_{ui \in U \pi z j \in Z} (c_j \in C, t_i \cdot z_j)}{\sum_{ui \in U \pi x j \in X} (a_j \in A, t_i \cdot x_j)}$$

$$= \frac{\dfrac{\sum_{ui \in U \pi z j \in Z} (c_j \in C, t_i \cdot z_j)}{|Cs|}}{\dfrac{\sum_{ui \in U \pi x j \in X} (a_j \in A, t_i \cdot x_j)}{|Cs|}}$$

where $Z = X \cup Y$ and $C = A \cup B$

Since the fuzzy confidence value is the measure of the degree of support given by the transactions, fuzzy confidence is also used to estimate the interestingness of the generated fuzzy association rules. In the equation, the fuzzy support of <Z,C> is divided by the fuzzy support of <X,A>. Using the cases in Table 3, the fuzzy confidence of the rule "If duration of URL1 is Long then duration of URL2 is Medium" is calculated as follows:

FC<<X,A>,<Y,B>> = (0.4 + 0.36 + 0.32 + 0.14 + 0.3 + 0.08 + 0.09) / (0.5 + 0.6 + 0.4 + 0.7 + 0.5 + 0.2 + 0.9) = 0.445

7 Fuzzy Index Tree

One common problem of using the fuzzy association rule technique is that the size of rule set generated is larger than the set of rules generated by using crisp association rules mining. To reduce the searching time for matching the correct rule, a fuzzy indexing method is proposed to index the rules in a tree-like structure. The generated fuzzy rules are grouped together by merging the fuzzy value.

For example, the fuzzy range of "Long Duration" and "Quite Long" might merge at the first level. As the rules are merged, a key rule is generated at each level for indexing. The fuzzy value of the key rule is calculated by the summation of all the fuzzy values of the rules merged. At the root level, the set of similar fuzzy association rules is defuzzified to form a crisp association rule. We illustrate the merging of fuzzy rules by the follow example:

Rule number	Given rule set
1	If (A=L,B=QL) then (C=L)
2	If(A=QL,B=M) then (C=L)
3	If(A=QS,B=S) then (C=M)
4	If(A=S,B=M) then (C=L)
5	If(A=M,B=S) then (C=QL)

Table 4 : Given rule set

Rule number	Merged key rules
1 and 2 merged	If (A=L,B=M) then (C=L)
3 and 4 merged	If(A=S,B=M) then (C=L)
5	If(A=M,B=S) then (C=QL)

Table 5 : Merged key rules (case I)

Rule number	Merged key rules
1, 2, 3, 4, 5 are merged	If (A=L,B=M) then (C=L)

Table 6 : Merged key rules (case II)

The key rules are grouped together by merging the antecedent of the rule. For example, rules "If (A=L, B=M) then C=L" and "If (A=L, B=S) then D=L" are merged as they have similar antecedents.

8. Experimental Design and Preliminary Result

Experiments have been carried out on testing the prediction of the access path by applying the association rule and fuzzy association rule algorithms using the anonymous web data from www.microsoft.com. The data were created by sampling and processing the www.microsoft.com logs. The data records are from 38,000 anonymous, randomly-selected users. For each user, the data lists all the areas of the web site (Vroots) that the user visited in a one-week timeframe. Users are identified by a sequential number, for example, #14988, #14989, etc. The file contains no personally identifiable information. The 294 Vroots are identified by their title (e.g. "NetShow for PowerPoint") and URL (e.g. "/stream"). The data come from one week in February 1998. Each instance represents an anonymous, randomly selected user of the web site. Each attribute is an area ("Vroot") of the www.microsoft.com web site. There are 32,711 training instances, 5,000 testing instances and 294 attributes, and the Mean Vroot visits per case is 3.0. A sample record will be as follows:

```
C,"12043",12043
V,1036,1
V,1030,1
V,1009,1
V,1074,1
```

where "C" represents the start of the session, the number coded is used to identified a user, V is the start of an access, and the number 1036 (the first access) is the Vroots which representing a title or a URL.

First, data are cleaned by removing the sessions with fewer than three accesses. To compare the result with the N-garm approach with n = 3, a moving window with size of 3 is applied to generate all the possible sessions where each has three accesses. For association rule mining, a minimum support of 6 is set and the minimum confidence value is 0.25. A set of associations in the form of "If A and B Then C" are discovered, where A, B, and C are Vroots. Since there is no attribute of duration in this data set, a random number between 1 and 0 is assigned to each access as the duration. This random value is fuzzified into corresponding linguistic terms by passing the value into the fuzzifying equations. A set of fuzzy association rules is generated based on the above algorithm, with fuzzy support equal to 0.3 and fuzzy confidence equal to 0.4.

The experimental result is summarized in the following Table 7:

Index	Number of rules Generated	Prediction Accuracy	Rule Coverage
3-Gram with support and confidence	1386	29.3%	79.6%
3-Gram with Fuzzy Association Rule	2243	42.5%	89.2%

Table 7 : Experimental Results

Two sets of rules are generated using the 38,000 anonymous records applying above methodology, and tested by 5,000 records. The number of training records is increased to about 50,000 and 6,800 records for testing, after partitioned into 3-gram format. The size of fuzzy rules set is about 1.6 times of the crisp rule set. However, we can see that fuzzy rule set performs better in the prediction accuracy. The term rule coverage is defined by the number rule matched (despite the correctness of matching) divided by the total number of testing records. In this experiment, fuzzy rules set showed a better coverage of data records.

9. Future Work and Discussion

The web server log is one of the most important sources of data providing information about browsing, visiting, and transactions on the web. However, the problems of proxy caching and browser caching affect the usability of the web log. Moreover, the web log is not a well-structured data format for data mining but only for recording information about client access. Therefore, the information that can be extracted from a web log file is limited. This is one reason why web log prediction usually cannot produce very good and very accurate results.

However, there are some data sources on the web site that have been ignored in the past, including the user profile created during registration, and the categorized content of the web. Since user profiles are subjective and static and might contain errors, and the patterns are easily out-dated, they are not suitable for web mining. However, these data, including user profile and web log, can be integrated together to form a new case base for an adaptation-guided web personalization system. In the following paragraphs we will present a new approach for adaptive web access prediction and recommendation by combining the web server log and user profile as case base and guiding the prediction by adaptive index.

Similar to the system proposed, the new system (Figure 2) can also be divided into two components: offline batch processing and online access path prediction. The offline batch processing is to produce the prediction rules for the recommendation or prediction engine. A set of general fuzzy rules and a set of adaptive fuzzy rules will be mined.

The user profile and the access records of the user are considered, as a case where the profile is the problem case and the URL in the site accessed is the solution. The general fuzzy rule is mined from this case base by the fuzzy association rule technique. The adaptive fuzzy rule will be mined from the web log and using the fuzzy association rule technique as well. The general fuzzy rule will do the prediction based on the matching of case similarity, while the adaptive rules are used to fine-tune and adopt some unusual patterns. An adaptation engine is defined to control the shifting between similarity and adaptation. The adaptation engine is generated by learning the criteria and conditions of adaptation based on the prediction accuracy and recall rate of the rules.

The online components include a web server and a prediction engine. The web server is used to capture the active user session and pass it to the prediction engine. The prediction engine matches the user session with the rule using the index tree, and adopts the change of rule follow the guiding of the adaptation engine. The adaptation engine recommends the future access path or user interest page based on the browsing behavior of similar users. The results will be passed to the web server and the predicted pages will be sent to the client browser. The system proposed above is to improve the quality of the web recommendation system using the methodology of case-based reasoning and making use of user profiles to improve the coverage of data.

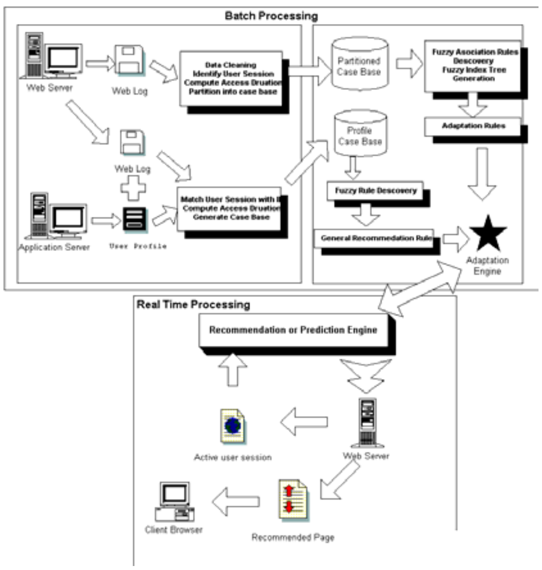

Figure 4. System Architecture for Adaptation-Guided Recommendation System

In this study, we are focusing on improving the quality of current web-based prediction systems by combining fuzzy association rule mining and CBR. However, there are some ways to enhance the current system. First of all, the duration that we used to be fuzzified in this study is subjected to network loading and download speed. Although from the result of the experiment, using duration as the only feature can do quite a good job, but the result can be better if we use more features, e.g. background information of user, web structure information. Besides, browser caching and proxy server caching that influence the constitution of the raw data is another problem for web prediction and web mining that is not handled in this study. For the future work, we will focus on improving the system by integrating different data source and work out a new system using an adaptation guided approach that we proposed.

10. Conclusion

In this study, we proposed a new approach to improve the quality of current web access paths prediction systems. The methodology of fuzzy rules with case-based reasoning provides a new idea for predicting future accessing paths using historical record. The fuzzy association rule technique is used to improve the prediction accuracy and fuzzy index tree for fast matching of rules. For future research direction, combining the web server log, the user profile and the content of the web site is necessary. The adaptation-guided system provides intelligence to recommend sites or customizes web contents based on the user's current browsing behavior. In future studies, we will test and establish the proposed system using real life data.

11. Acknowledgement

This project is supported by the RGC grant B-Q379 and HKO research grant H-ZJ90.

Reference

[1] R.Cooley, B. Mobasher and J. Srivastava. Web Mining: Information and Pattern Discovery on the World Wide Web *in Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), November 1997.*

[2] B. Mobasher, R.Cooley, and J. Srivastava. Automatic Personalization Based on Web Usage Mining *in Communication of ACM, August, 2000 (Volume 43 , Issue 8)*

[3] T. Joachims, D. Freitag, T. Mitchell WebWatcher: A Tour Guide for the World Wide Web *in IJCAI97 -- Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, pages 770--775, Nagoya, Japan.*

[4] A. Joshi, S. Weerawarana, and E. Houstis, "On Disconnected Browsing of Distributed Infromation," Proc Seventh IEEE Intl. Workshop on Research Issues in Data Engineering (RIDE), pp. 101-108, 1997

[5] J. Borges, M. Levene. Data Mining of User Navigation Patterns *in Proceedings of the Workshop on Web Usage Analysis and User Profiling (WEBKDD'99), pages 31-36. August 15,1999, San Diego, CA.*

[6] M. Spiliopoulou, L. C. Faulstich, WUM: A Web Utilization Miner *in Workshop on the Web and Data Bases (WebDB98), pages 109-115, 1998.*

[7] A. Joshi, R. Krishnapuram. Robust Fuzzy Clustering Methods to Support Web Mining, *in Proc. ACM SIGMOD Workshop on Data Mining and Knowledge Discovery, Aug 98.*

[8] O. Nasraoui, H. Frigui, A. Joshi, R. Krishnapuram. Mining Web Access Logs Using Relational Competitive Fuzzy Clustering, *in Proc. Eight International Fuzzy Systems Association World Congress - IFSA 99, August 1999*

[9] Z. Su, Q. Yang, H. J. Zhang. A Prediction System for Multimedia Pre-fetching *in ACM Multimedia Conference 2000, October 2000, LA CA USA*

[10] A.G. Buchner, S. S. Anand, M.D. Mulvenna and J.G. Hughes. Discovering Internet Marketing Intelligence Through Web Log Mining , *in ACM SIGMOD Record, ISSN 0163-5808, Vol. 27, No. 4, pp. 54-61, 1998*

[11] Q. Yang, H. H. Zhang, H. Zhang. Taylor Series Prediction: A Cache Replacement Policy Based on Second-Order Trend Analysis *in the Proceedings of the Hawai'i International Conference On System Sciences, January 3-6, 2001, Maui, Hawaii*

[12] J. Pei, J. Han, B. Mortazavi-asl, and H. Zhu. Mining Access Pattern Efficiently from Web Logs *in Proc. 2000 Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'00), Kyoto, Japan, April 2000*

[13] M. Spiliopoulou, L. C. Faulstich, K. Winkler, A Data Miner Analyzing the Navigational Behaviour of Web Users *in Proc. of the Workshop on Machine Learning in User Modelling of the ACAI'99 Int. Conf., Creta, Greece, July 1999*

[14] Z. Su, Q. Yang, H. J. Zhang. What Next: A Prediction System for Web Requests using N-gram Sequence Models *in First International Conference on Web Information Systems and Engineering Conference. Hong Kong, June 2000*.

[15] R.Cooley, P. N. Tan, and J. Srivastava. Discovery of Interesting Usage Patterns from Web Data *in Springer-Verlag LNCS/LNAI series, 2000.*

[16] A. Gyenesei. A Fuzzy Approach for Mining Quantitative Association Rules TUCS Technical Reports No.336, March 2000 ISBN 952-12-0634-9 ISSN 1239-1891

[17] H. P. Lo, L. Lu, M. Chan. Improving the Web Design – Mining Web Data at Cityjob.com PAKDD 2001 Industrial Track and Workshop Proceeding p.14-24 (April 2000)