

# Web Personalization Based on Association Rules Finding on Both Static and Dynamic Web Data

by

Minghao Lu

B.Sc., The University of Toronto, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Computer Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September, 2008

© Minghao Lu 2008

# Abstract

The explosive and continuous growth in the size and use of the World Wide Web is at the basis of the great interest into web usage mining techniques in both research and commercial areas. In particular, the need for predicting the user's needs in order to improve the usability and user retention of a website is more than evident and can be addressed by personalization. In this thesis, we introduce a new framework that takes advantage of the sophisticated association rule finding web mining technology on both dynamic user activities over a website, such as navigational behavior, and static information, such as user profiles and web content. We also provide a novel personalization selection system which allows users to choose the most suitable profile for them in any given period of time.

In order to examine the viability of our framework, we incorporate and implement it over a well designed simulation environment. Moreover, our experiment proves that our framework provides an overall better web personalization service in terms of both recommendation accuracy and user satisfaction.

# Table of Contents

<b>Abstract</b> . . . . .	ii
<b>Table of Contents</b> . . . . .	iii
<b>List of Tables</b> . . . . .	vi
<b>List of Figures</b> . . . . .	vii
<b>Acknowledgements</b> . . . . .	viii
<b>1 Introduction</b> . . . . .	1
1.1 Motivation . . . . .	1
1.2 Goal . . . . .	2
1.3 Thesis Contributions . . . . .	3
1.4 Thesis Organization . . . . .	4
<b>2 Personalization Background and Related Works</b> . . . . .	5
2.1 Contextual Information . . . . .	5
2.1.1 Background . . . . .	5
2.1.2 Web Data . . . . .	6
2.1.3 Personalization modules . . . . .	7

iii

*Table of Contents*

---

2.1.4	Web Usage Mining . . . . .	9
2.2	Related Works . . . . .	11
2.2.1	Web Search Personalization . . . . .	11
2.2.2	Personalization with back off . . . . .	12
2.2.3	Web usage data mining, Link Analysis and Naviga- tional patterns . . . . .	12
<b>3</b>	<b>System Design . . . . .</b>	<b>14</b>
3.1	Introduction . . . . .	14
3.2	System Architecture . . . . .	14
3.2.1	Data Preprocessing Phase . . . . .	16
3.2.2	Web Data Mining Phase . . . . .	20
3.2.3	Personalization Phase . . . . .	23
<b>4</b>	<b>Implementation Analysis . . . . .</b>	<b>26</b>
4.1	Introduction . . . . .	26
4.2	Usage Log and Plug-In Manager . . . . .	30
4.3	Personalization Plug-In Components . . . . .	31
4.3.1	Association Rules Analyzer . . . . .	34
4.3.2	Recommendation Pool Generator . . . . .	34
4.3.3	User Profile Sliding Window . . . . .	35
4.3.4	Personalization Selection System . . . . .	36
<b>5</b>	<b>Evaluation . . . . .</b>	<b>37</b>
5.1	Simulation Setup . . . . .	37
5.1.1	Simulation Parameters . . . . .	40

*Table of Contents*

---

5.2 Simulation Results . . . . .	42
<b>6 Conclusion and Future Work . . . . .</b>	<b>47</b>
6.1 Conclusion . . . . .	47
6.2 Future Work . . . . .	48
<b>Bibliography . . . . .</b>	<b>51</b>

# List of Tables

3.1	Online vs. Offline Data Preprocessing Properties . . . . .	18
3.2	Representation of Static Data . . . . .	21
3.3	Representation of Dynamic Data . . . . .	21
4.1	UsageLog.cs Object . . . . .	30
4.2	PlugInManager.cs Object . . . . .	31
4.3	ARAnalyzer.cs . . . . .	34
4.4	RecommendationPoolGenerator.cs . . . . .	35
4.5	ProfileSlideWindow.cs . . . . .	36
4.6	PersonalizationSelection.cs . . . . .	36
5.1	Game Website Category Examples . . . . .	39

# List of Figures

3.1	Flowchart of the Overall System Architecture . . . . .	15
3.2	Profile Sliding Window . . . . .	24
4.1	Overview of Plug-In System . . . . .	27
4.2	Flowchart of Personalization Plug-In Program . . . . .	29
4.3	Detailed Overview of Our Personalization Plug-In System . .	33
5.1	Average Accuracy on Different Test Day Interval . . . . .	43
5.2	Our Approach VS. Traditional Approaches . . . . .	45

# Acknowledgements

First of all, I would like to thank my supervisor, Dr. Son T. Vuong, for his understanding, support and encouragement. I appreciate the freedom he has given me in pursuing this thesis and in the meanwhile offering suggestions and comments. I would also like to express my gratitude to my second reader, Dr. Kurt Eiselt, for his support and suggestions.

This thesis would have never been completed without the support of many people, especially the members of the NIC lab who have made the lab a great place to work in. Their insightful comments and suggestions are greatly appreciated for improving this thesis.

Finally, my thanks go out to my family and my friends who have supported me throughout the year.

# Chapter 1

## Introduction

### 1.1 Motivation

Web personalization is traditionally defined as the process of customizing web pages to meet the interests of individual users, taking advantage of the information obtained through the analysis of the user's navigational behavior in correlation with other data collected within the web context [3]. Specifically, the structure of the websites being visited, their contents, and any existing profile information of the user in question are used. As such, changes to a user's personalization information are generally based on the analysis of implicit events, such as the purchasing of items, browsing of specific webpages, etc. Users can benefit from personalization in certain ways: (1) It saves time since personalization keeps track of user habits and can eliminate repetitive tasks such as clicking by providing shortcuts to specific content. This also has the added benefit of reducing redundant work by streamlining tasks which are common to a user. Furthermore, by retaining past transactional details for the user, personalization can provide a quicker way of retrieving products that the user is aiming for. (2) It helps to filter out irrelevant tasks that might distract the user by allowing users to perform actions they commonly do in a more direct manner (3)

## 1.2. Goal

---

It allows for a much finer level of targeting in terms of providing users with relevant information. For example, location-based recommendations include information concerning specific services available in the user's local geographical area or event alerting them of events going on within their neighborhood.

There are several ways of analyzing data collected through a user's interaction with the website to generate personalization. These include web personalization based on link analysis and navigational patterns [9] [7], dynamic personalization without user intervention [4], personalization based on association rule discovery from web usage data [13], personalization with backoff [12], effective periodic web personalization [17]. However, these methods all adhere strictly to either using static information such as user profiles or dynamic behavior such as user navigational patterns to generate web personalization. Although in Massimilano et al's work [1], they proposed a novel re-classification algorithm on both static and dynamic user information, we have yet seen anyone applies association rule finding algorithm on both data.

## 1.2 Goal

In the context of web personalization, there are two aspects that are crucial from the perspective of the user: the accuracy of the personalization or recommendation in terms of how relevant it is to the user's needs; and the user's satisfaction with the personalization or recommendation. The goal of this thesis is to provide a system framework that provides better service to

the user along these two guidelines.

Our new system provides web personalization to the user using the following criteria: (1) The recommendation generated by the system is based on both dynamic user navigational behavior and static user profile when possible; (2) The system maintains a sliding window of user profiles; (3) Users have the option of choosing which profile they want the system to perform personalization on at any given time. As such, the focus of our design will be towards the accuracy and overall user satisfaction of the personalization.

However, it is worth mentioning that the intention of this thesis is to provide a design framework for building a more accurate and user-friendly personalization system instead of creating a fully functional user preference decision maker, which is beyond our scope. As such, we will be evaluating our system's applicability and ability to provide better recommendations through simulations while providing its practical viability and feasibility by outlining the implementational details of the incorporation of our system into a simulated gaming website.

### 1.3 Thesis Contributions

The main contribution of this thesis is the creation of a framework that supports both dynamic and static user information in the generation of user web personalization. Specifically, we:

- Collect and make use of both dynamic user navigational behavior and static user profile information, which is to our knowledge unique among any existing related works.

- Store user profiles for different time intervals in order to give the user the option of choosing which profile will most suitable for them at any given time. This has the additional benefit of addressing user concerns towards privacy and security issues with regards to profiling information.

## 1.4 Thesis Organization

This thesis is organized as follows. Chapter 2 provides some background information regarding web personalization as well as outlining some notable related works. Chapter 3 formulates our system designs and architectures. Chapter 4 applies our design into our gaming website framework, and highlights the advantages of our implementation. Chapter 5 simulates the gaming website with our system in order to validate the benefits of our system. Chapter 6 provides a conclusion to this thesis and looks into future works on this topic.

## Chapter 2

# Personalization Background and Related Works

This chapter presents some basic personalization background information with regard to web data, analyzing methods, and related works in the field.

### 2.1 Contextual Information

#### 2.1.1 Background

With the rapid growth of the World Wide Web, new methods of designing and developing online information services are in high demand. Most websites are large and unorganized and can be very complicated for the users to traverse and obtain information they need. Hence, users are often distracted from the goal of their initial inquiry during website navigation and may receive misleading information or results as they browse through web contents. This potential web clutter is compounded by the emergence of concepts such as Web 2.0 and e-commerce, causing website owners and designers to overload existing designs to incorporate web marketing and user customization.

## 2.1. Contextual Information

---

Personalization is an attempt at addressing a service provider's desire to push additional information to users visiting their domains while at the same time restricting the flow of irrelevant recommendations. This is achieved by predicting a user's requirements and needs so that accurate results can be generated during the course of a user's web navigation. Different personalization systems give somewhat different recommendations based on the type of web data they collect and the methods they choose to employ for analyzing and generating recommendations. Regardless of any variations, however, the accuracy of the recommendations generated always remains the key factor used to rate any personalization system.

Nonetheless, no matter how sophisticated and remarkable a personalization system is, there will always be some users who simply are not interested. Every personalization system needs to be prepared for recognizing and immediately accommodating any motivating factors which can cause a user to reject the benefits of personalization.

### 2.1.2 Web Data

In order to perform user personalization, web data must be collected. What are web data? According to Srivastava et al. [15], web data can be classified into four major categories

- **Content data** represent the content of the website. These are presented to the end-user appropriately structured. They can range from simple text to images to even structured data such as information retrieved from web databases.

## 2.1. Contextual Information

---

- **Structure data** represent the way content is organized within a website. They can be either data entities used within a webpage, such as HTML or XML tags, or data entities used to put a website together, such as hyperlinks connecting one page to another.
- **Usage data** represent general information collected by the website on its users, such as a visitor's IP address, time and date of access, complete path (files or directories) accessed, referrer's address, and other attributes that can be included in a Web access log.
- **User profile data** provide information on specific users of the website. A user profile contains demographic information (such as name, age, country, marital status, education, interests, etc.) for each user of the website, as well as information on user interests and preferences. Such information is typically acquired either explicitly through registration forms or questionnaires, or can be implicitly inferred by analyzing web usage logs.

### 2.1.3 Personalization modules

Moreover, according to Magdalini et al. [8], the aforementioned data affords a system the ability to perform five types of web personalization:

- **User profiling:** In the Web domain, user profiling is the process of gathering information specific to each visitor, either explicitly or implicitly. A user profile usually consists of demographic information about the user, user's interests and browsing behavior. This information is exploited in order to customize the content and structure of a

## 2.1. Contextual Information

---

website to the visitor's specific and individual needs.

- **Log analysis and Web usage mining:** This is the stage where the information stored in web server logs is processed by applying data mining techniques in order to (i) extract statistical information and discover interesting usage patterns as a precursor to generate user profiles, (ii) cluster the users into groups according to their navigational behavior, and (iii) discover potential correlations between web pages and user groups. Part (i) can be regarded as part of the user profiling process.
- **Content management:** This is the process of classifying the content of a web site in semantic categories in order to make information retrieval and presentation easier for the users. This is very important for websites that contain rapidly changing content, such as news sites or portals.
- **Web site publishing:** A publishing mechanism is used in order to present the content stored locally in a web server and/or some information retrieved from other web resources in a uniform way to the end-user. Different technologies can be applied to publish data on the web.
- **Information acquisition and searching:** In many cases, information provided by a website is not physically stored in the website's server. In the case of a web portal, users are interested in information from various web sources. It is the responsibility of those in charge

## 2.1. Contextual Information

---

of site content (i.e. the website editor) to search the web for content of interest that should consequently be classified into thematic categories. Searching and relevance ranking techniques must be employed both in the acquisition of relevant information and in the publishing of the appropriate data to each group of users.

In our work, our focus will be mainly on the use of user profiling, log analysis and web usage mining. Usually, the logs and web usage are changing from time to time while user profile remains statically unchanged for a relatively long time unless the user wants to manually update it. However, a user's preference will be constantly changing. For example, after watching a movie, the user might become interested in the various artists of that movie; if the user moves to another city due to work relocation, he/she would very naturally become interested in the activities in that city instead, or in addition to, those of his/her home town. The traditional approach doesn't take these situations into consideration. Hence, we dynamically use log analysis and web usage mining to improve the user profile in order to obtain better recommendations for the user.

### 2.1.4 Web Usage Mining

The purpose of Web usage mining is to apply data mining techniques to the aforementioned web data in order to discover useful user navigational patterns. The most widely used method is statistical log analysis. In addition, according to Eirinaki and Vazirgiannis [8], we have some advanced data mining techniques and algorithms specifically tailored for the use in

## 2.1. Contextual Information

---

Web domain such as sequential association rules finding, pattern discovery, clustering and classification.

Association rules finding (ARF) [11] is a basic mechanism for discovering frequent patterns, associations and correlations among sets of items. In the web domain, association rules are used to find correlations between pages that are visited by the user during a certain browsing session. Such rules reveal the possible relationship between pages that may or may not be directly connected with each other. Website reconstructions and recommendations can thus be achieved by these pattern observations. For example, ARF is usually used to add a few links that interconnect with other pages which are often viewed by the given user, or to improve the website's overall performance by removing unnecessary links and through prefetching useful web data.

Sequential pattern discovery [6] is an extension of association rules finding in such a way that it indicates patterns of co-occurrence with the incorporation of time sequence. In the web domain, such patterns could be a sequence of web pages that happen in a given period. Applying this approach, useful user's navigational trends could be discovered and studied in order to give predictions on the user's visit patterns.

Clustering [5] is a method of gathering items that have similar characteristics. In the context of web mining, we can have two distinct cases, user clusters and page clusters. User clustering identifies groups of users that seem to have similar behaviors when browsing through a website. Page clustering results in groups of pages that are apparently related to each other in terms of user's perception. Such clustering information is then used for

personalizing a website.

Classification [16] is a process of mapping a data item into one of the pre-organized classes. In the web domain classes usually represent different user profiles and classification is then applied using selected features which describe each user's category. The most commonly used classification algorithms are naive Bayesian classifier, neural networks, decision trees and etc.

## 2.2 Related Works

### 2.2.1 Web Search Personalization

Currently existing systems that make use of personalization, such as Google, Yahoo and Amazon, provide personalized search results for users weighing on a number of factors including but not limited to user browsing history, bookmarks and site click-through rate. These user interests are saved on the server for further analysis in order to meet the user specified requirements.

However, user interests can change drastically in a short period of time, and existing systems cannot reflect these changes dynamically. Sendhil Kumar and Geetha [14] proposed a dynamic personalized ontology for web search personalization. By performing personalized search using a user's profile, which is based on their browsing behavior, more relevant search results that are semantically related to the given search query could be returned. In order to facilitate this, their system automatically constructs page ontology as a user navigates the web. Personalization based on such ontology is then used to produce better search results in comparison to the

traditional keyword based searching.

### 2.2.2 Personalization with back off

Personalization is based on user profile, but what if we do not have data for a particular user either because they are a new user or due to the privacy or security concerns? Mei and Church [12] suggested the approach of "backing off" to larger groups of similar users. It is a proof of concept that users are usually grouped into equivalence classes based on the most significant bytes of their IP address. They found that a little bit of personalization is better than too much or too little. Moreover, according to them, personalization with backoff to higher bytes of IP addresses is better than complete personalization or no personalization at all. Zero personalization misses the opportunity and 100% personalization generates sparse data and often runs into privacy issues.

Nevertheless, the paper only choose to "back off" to a larger network in terms of IP address. It would certainly be better to group users by URL segments and/or collaborative filtering such as similar user queries and similar URL click patterns.

### 2.2.3 Web usage data mining, Link Analysis and Navigational patterns

Traditional web personalization is mostly based on explicit user information such as clickstream data (user ratings) and personal profiles. However, it becomes increasingly unscalable as the number of users and web contents increase. Mobasher et al.[13] present a scalable framework for web person-

## 2.2. Related Works

---

alization based on association rules mining of web usage data. Their work allows using multiple support levels for different types of pageviews. They also studied the use of varying-sized user histories on both the precision and coverage of the generated recommendation. Their association-based recommendation system was able to maintain computational advantage over the traditional approach while improving the precision and overall coverage of the recommendation.

Most of the existing personalization models are solely based on usage data, ignoring the link structure of the Web graph visited. Eirinaki and Vazirgiannis [9] have studied the integration of link analysis during the web personalization process. They propose a novel algorithm called UPR, a usage-based personalized version of the PageRank algorithm, which can be applied to any navigational graph based website to provide recommendations derived from the structural properties of the navigational graph. Their experimental approach first of all addresses several disadvantages of pure usage-based models. Eirinaki and Vazirgiannis claim that, by incorporating link analysis techniques into such models, their new framework outperforms the existing approaches.

## Chapter 3

# System Design

In this chapter, we present the design of our personalization/recommendation framework. By combining the existing approaches and refining them with our user personalization selection system, we aim to improve both the accuracy and user satisfaction of the generated recommendations.

### 3.1 Introduction

Our system is designed to be applicable to any existing website. It interacts with a user's navigational behavior as well as website content including user profiles, web pages and usage logs. Although the issue of user privacy will arise, since our system only performs actions on the user information locally, we think it is still within safety concerns. A larger system that will work on multiple websites for user overall browsing personalization is beyond the scope of this thesis and could be investigated in future works.

### 3.2 System Architecture

The system can be divided into three major stages:

- Data preprocessing stage, which the data we collected is processed into

### 3.2. System Architecture

---

the format that can be adopted by our system,

- Association Rule Finding stage, where we apply the association rules finding algorithm on both the static data we processed such as user profile and web content and the dynamic user navigational behaviors in order to generate useful recommendations,
- Personalization stage, where our personalization selection system gives corresponding personalization based on user's preference.

(See Figure 3.1)

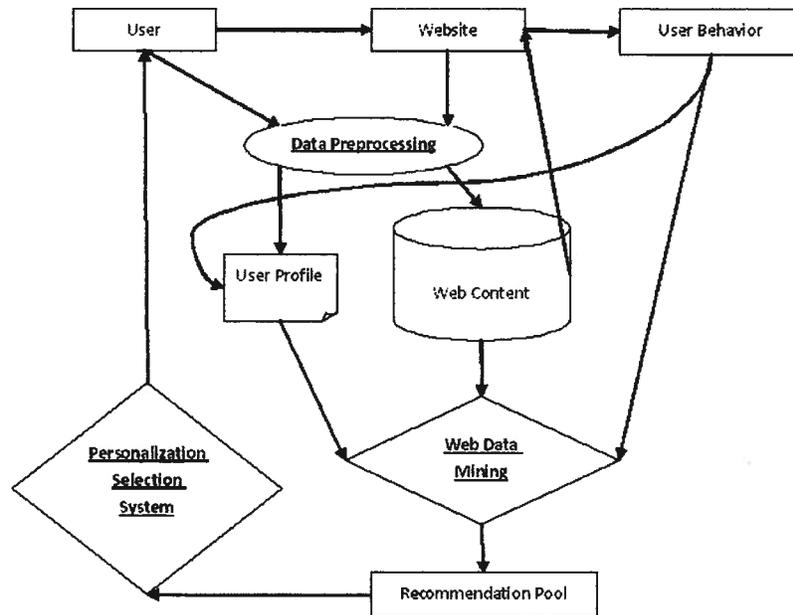


Figure 3.1: Flowchart of the Overall System Architecture

### 3.2.1 Data Preprocessing Phase

The data we talk about here is usage-based. It is composed of both user profiles and web contents. In order to perform the association rule finding algorithm on data we have collected, we need to formalize our data. Our data preprocessing phase transforms different types of data into a unified format in order to be easily used by our AR algorithm. Our system then utilizes the processed web content data to modify the website in order to obtain user navigational behavior in the same data format. Moreover, the data preprocessing phase can be achieved either online or offline. We choose to do it offline since it saves both time and space.

**User Profiling:** In order to personalize a web site, the system should be able to describe each distinct user. And we call this process user profiling [2]. The traditional approach is to create a knowledge base that contains user preferences, characteristics and activities. In our system, we exclude the user activities and put it into the user navigational behavior part since it is dynamically changing and does not fit our offline data preprocessing concept. Although user preference sometimes changes, it is much less likely when compared with user browsing activities.

User profiling could simply be a registration step. Users describe their preferences and characteristics through registration forms. If the website doesn't require a registration step, we can internally create an empty one for each user containing the user IP address in order to distinguish different log-ins. Based on different functionalities, we introduce three different user profiles:

### 3.2. System Architecture

---

- True profile reflects the user's true preference and characteristics and should also guide how the user navigates the website;
- Registration profile is the one that the user initially provides to the website. It may or may not be the same as the user's true profile and if the website does not have the registration page, we create an internal profile which only contains the user IP address and any useful information about the user;
- System profile is what our personalization system is based on. It is what our system has learned from the user's activities within the website scope. Initially, it is set to the registration profile and should converge to the user's true profile over time. Thus it is changing as the user's preferences, characteristics and navigational behavior change. However, some users do not want to disclose their personal information. Choosing which profile fits the user most in a certain period will certainly be a crucial issue. We will tackle this later in 3.2.3.

Since the user profiling phase initially happens in the website registration step, the data could be organized in the same way as we do for web contents and make these two share the same data format.

**Web Content:** Web content is the textual, visual or aural content that is encountered as part of the user experience on websites. Although most of the websites are well organized and the contents are ready for web mining strategies, there is usually a lack of communication with the user registration part. In other words, the categories that the web contents are organized into

### 3.2. System Architecture

---

often do not match the registration fields. We will talk about this in detail in the simulation chapter.

The website contents are pre-categorized with the design of registration forms regardless of how well-organized they are. In this way, we can make sure that our user profile's fields will fall into the same categories that the web contents belong to so that they can both be absorbed by our AR finding algorithm in a much more convenient way.

**Offline vs. Online Data Preprocessing:** The data preprocessing phase can be finished either offline or online. Both have certain advantages and drawbacks which are summarized as follows:

	Online	Offline
Time and Bandwidth	Extra time and bandwidth are needed for processing data while user is online. User might find their browsing experience lagging out and less satisfying.	Since the processing happens offline, user enjoys the web surfing without even noticing the difference.
Space and Interactivity	Doesn't need to record the changes each time user logs out. The data is processed simultaneously during the user browsing period	Need to store the changes that users made. Users do not get the most updated personalization as the recent data is processed only after they go offline.
Content Management	Real time update on web contents changes and new contents.	Need to set up a time schedule for updating the new web contents or any changes to the current contents.

Table 3.1: Online vs. Offline Data Preprocessing Properties

Other problems for both approaches that may occur when the website

### 3.2. System Architecture

---

doesn't require a registration step are that the user might manually delete the cookie file which stores the information about different users' browsing activities since it is stored locally in the user's computer and when he/she revisits a website, he/she will be regarded as a new visitor. All the old information learned from the user will not be useful anymore and the data preprocessing stage needs to be applied again on the same user. Furthermore, if more than one user browses the website on a public computer, the system will not be able to tell the identity of the user and may mix up the data from different users.

**Privacy Issues:** As we are dealing with usage based data, privacy issues are inevitable. Many users are not willing to give away their personal information either explicitly (reluctant to register or give wrong registration information) or implicitly (navigational behavior). In both cases, the user loses anonymity and is aware that his/her actions will be recorded and used, in most cases, without his consent. Furthermore, even if the user agreed to disclose his information, he does not have any control over the information being exchanged between websites without his permission. Thus, privacy could be violated in many ways and user agreement is often required before we can start collecting data from the user.

However, since our system is built upon a single website domain and is implemented internally, data exchanging is not a problem. Hence, privacy issues are less concerning in this situation. Still, in future works, we might want to give users the option to participate in our personalization system or not.

#### 3.2.2 Web Data Mining Phase

After we have the preprocessed data ready, there are many web data mining strategies such as log statistical analysis, sequential pattern discovery, association rule finding, clustering and classification. In our system, we choose to use the association rule finding algorithm to find useful patterns.

There are two types of preprocessed data that are going to be used as our source data for the association rules finding algorithm. One is static data such as user registration information and web content, and the other is dynamic data such as a user's navigational behavior. Most of the existing systems apply the association rule finding algorithm solely on either static data or dynamic data. To our best knowledge, our work is the first ever system that performs the association rule finding algorithm on both data. Furthermore, we continuously apply the algorithm in order to get more accurate results for recommendations.

### 3.2. System Architecture

---

**Static Data:** All the data that remain relatively unchanged and could be easily predicted during a certain period of time are considered static data.

Data	Description
User Profile	User's registration information, preferences and characteristics are considered unlikely to change for a fairly long period of time and hence could be seen as static data.
Web Content	Although web pages get updated every day, it could still be considered discretely unchanged. Thus, it remains in the static range.

Table 3.2: Representation of Static Data

**Dynamic Data:** In converse, dynamic data are the data that are continuously changing and unpredictable.

Data	Description
User Navigational Behavior	The navigational pattern of the user differs a lot from day to day even from one browsing session to another. It could also be totally random activities.

Table 3.3: Representation of Dynamic Data

**Association Rules Finding Algorithm:** Association rules finding is a popular and sophisticated method for discovering interesting relationships between variables in large datasets such as commercial websites. The resulting rules are usually useful for generating appropriate recommendations. For example, {Xbox, Teenagers}  $\Rightarrow$  {Action Games} found in gaming websites

### 3.2. System Architecture

---

indicates that if a teenager owns an Xbox, he is most likely a fan of acting games. Such useful patterns could be turned into recommendations that are presented back to the user, who might or might not be aware of this implicit information about himself/herself.

Following the original definition by Agrawal et al, the problem of association rules finding is defined as: Let  $I=\{i_1,i_2,\dots,i_n\}$  be a set of  $n$  binary attributes called items. Let  $D=\{t_1,t_2,\dots,t_m\}$  be a set of transactions. Each transaction in  $D$  has a unique transaction ID and contains a subset of the items in  $I$ . A rule is defined as an implication of the form  $X \Rightarrow Y$  where  $X, Y \subseteq I$  and  $X \cap Y = \emptyset$ . The sets of items (itemsets)  $X$  and  $Y$  are called antecedent (left-hand-side or LHS) and consequent (right-hand-side or RHS) of the rule.

To select interesting rules from the set of all possible rules, constraints on various measures of significance and interest can be used. The best-known constraints are minimum thresholds on *support* and *confidence*.

The support  $\text{supp}(X)$  of an itemset  $X$  is defined as the proportion of transactions in the data set which contain the itemset. In the example above, the itemset  $\{Xbox, Teenagers\}$  has a support of 0.6 if it occurs in 60% of all transactions.

The confidence of a rule is defined as  $\text{conf}(X \Rightarrow Y) = \text{supp}(X \cup Y)/\text{supp}(X)$ . Again, using the above example, suppose that  $\text{supp}\{Xbox, Teenager, Action Games\} = 0.3$ , then the rule  $\{Xbox, Teenagers\} \Rightarrow \{Action Games\}$  has a confidence of  $0.3/0.6 = 0.5$ . This means that for 50% of the time when the transaction contains  $\{Xbox, Teenagers\}$  the rule is correct. Confidence can be interpreted as an estimate of the probability  $P(Y | X)$ , the probability of

### 3.2. System Architecture

---

finding the RHS of the rule in transactions under the condition that these transactions also contain the LHS. [11]

Association rules are required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time. To achieve this, association rule generation is a two-step process. First, minimum support is applied to find all frequent itemsets in a database. In the second step, these frequent itemsets and the minimum confidence constraint are used to form rules.

In the web data mining stage, datasets are usually composed of static data and transactions are from dynamic data. After the data preprocessing stage, the web database is organized and ready for our association rules finding algorithm with system-determined support and confidence in order to generate the most reliable and useful patterns or recommendations.

#### 3.2.3 Personalization Phase

The last but not least stage is our personalization phase. Personalization not only gives users the opportunity to save unnecessary browsing time and to obtain specific information that is increasingly relevant to her interests, but improves overall website user satisfaction in terms of better structures, functionalities and browsing experience as well.

Although personalization is the ultimate goal of web data mining and obviously has many benefits for the user, there are always some people who do not like to take advantage of it or only want to participate in part of the personalization due to many reasons such as lack of security and credibility. Here, we propose a novel personalization selection system which tackles

### 3.2. System Architecture

---

this situation.

**Personalization Selection System:** Users usually are not aware of their personal information or data being collected and studied. The only thing they will notice is the changes of the website in terms of personalization or, in other words, recommendation. However, users sometimes are not willing to take the advantage of personalization for reasons such as security and privacy [10].

The basic idea of our personalization selection system is to keep a sliding window of user preferences and let the user choose which profile they want the system to use as the source for personalization in any given period of time.

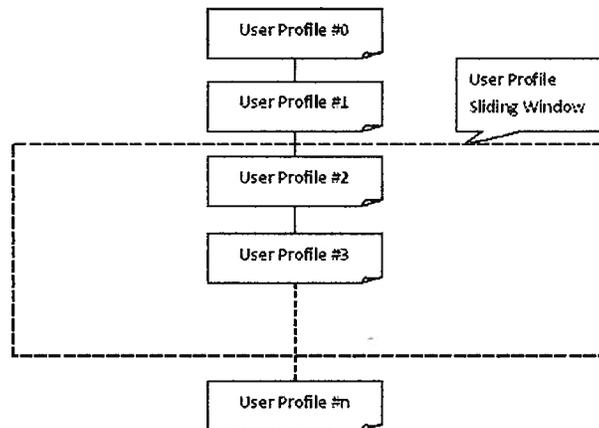


Figure 3.2: Profile Sliding Window

User profile #0 is set to be the registration profile (See Figure 3.2).

### 3.2. System Architecture

---

A user gets the opportunity to choose which profile fits his/her current preferences. Personalization is then performed on the selected profile instead of the most recent one (which should be the closest to user's true profile). In this case, a user can choose not to participate in the personalization at all by choosing the registration profile which was filled with false information, choose to fully participate in the personalization by choosing the most recent profile which theoretically converges to the user's true profile or choose to partially participate in the personalization by choosing profiles in between. For example, some teenagers often browse gaming website contents which are only for adults. They do not want this to be shown to their parents when the system chooses to personalize this by recommending adult web pages. As another example, a business executive who travels a lot in a short period of time wants the website to provide local information instead of his registered location. Our system is quite flexible in these cases and able to solve the problems.

## Chapter 4

# Implementation Analysis

In this chapter, we explore the feasibility of our framework over websites. Theoretically speaking, our system could be applied on any existing website domains. Being web personalization architecture, it could be implemented either on the server side or on the user side. Since our system is only concerned with users' navigational activities during the time they browse a particular website, we believe it is better to implement our system on the server side. We start with a brief introduction on the website domain and relevant data infrastructures. Then, we highlight the designs and necessary modifications to apply our framework.

### 4.1 Introduction

A Website is a collection of pages, images, videos and any other digital components that is hosted by one or more web servers, usually accessible via Internet, Wi-Fi and mobile phone devices. A website usually contains a great amount of information distributed through thousands of pages. Without appropriate guidance, users often find it hard to get to the pages that they are interested in and wander aimlessly among irrelevant pages, resulting in losing browsing interests and leaving the website soon.

#### 4.1. Introduction

---

With the emergence of web 2.0 which aims to enhance the creativity, information sharing and collaboration among users, we are aware of the aforementioned situation even more seriously. Personalization plug-in hence comes in as a solution. (See Figure 4.1)

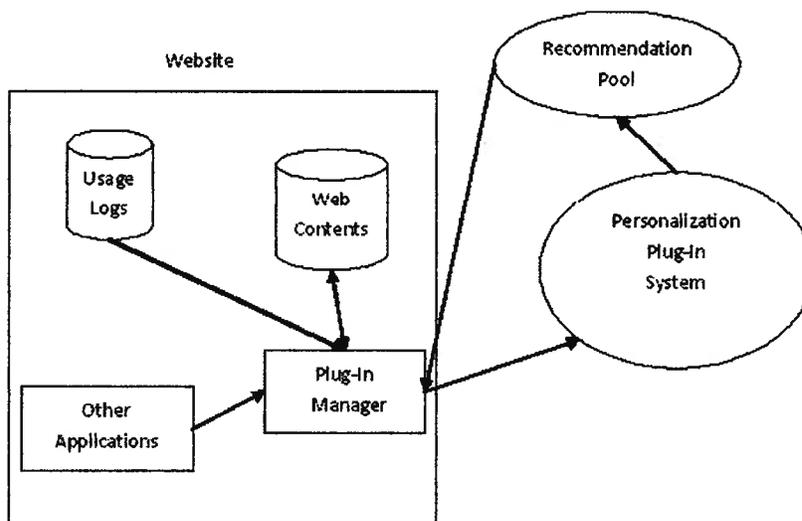


Figure 4.1: Overview of Plug-In System

Plug-in applications are programs that can easily be installed and used as part of Web browser. The plug-in mechanism is a commonly used solution for enhancing a website's functionalities. We can see from Figure 4.1 that the host application provides services and data including usage logs and web content which the plug-in system can use through the plug-in manager installed on the host side. The benefit that we will gain from using a plug-in system is that, since the plug-in system is just a small component of a website which is introduced by a third-party, it can be easily developed,

#### 4.1. Introduction

---

maintained and tested without any interference with a website and its own applications.

In this thesis, we construct a personalization plug-in system specifically for the purpose of website recommendation by using .NET technology. Here is our personalization plug-in program flow chart. (See Figure 4.2)

#### 4.1. Introduction

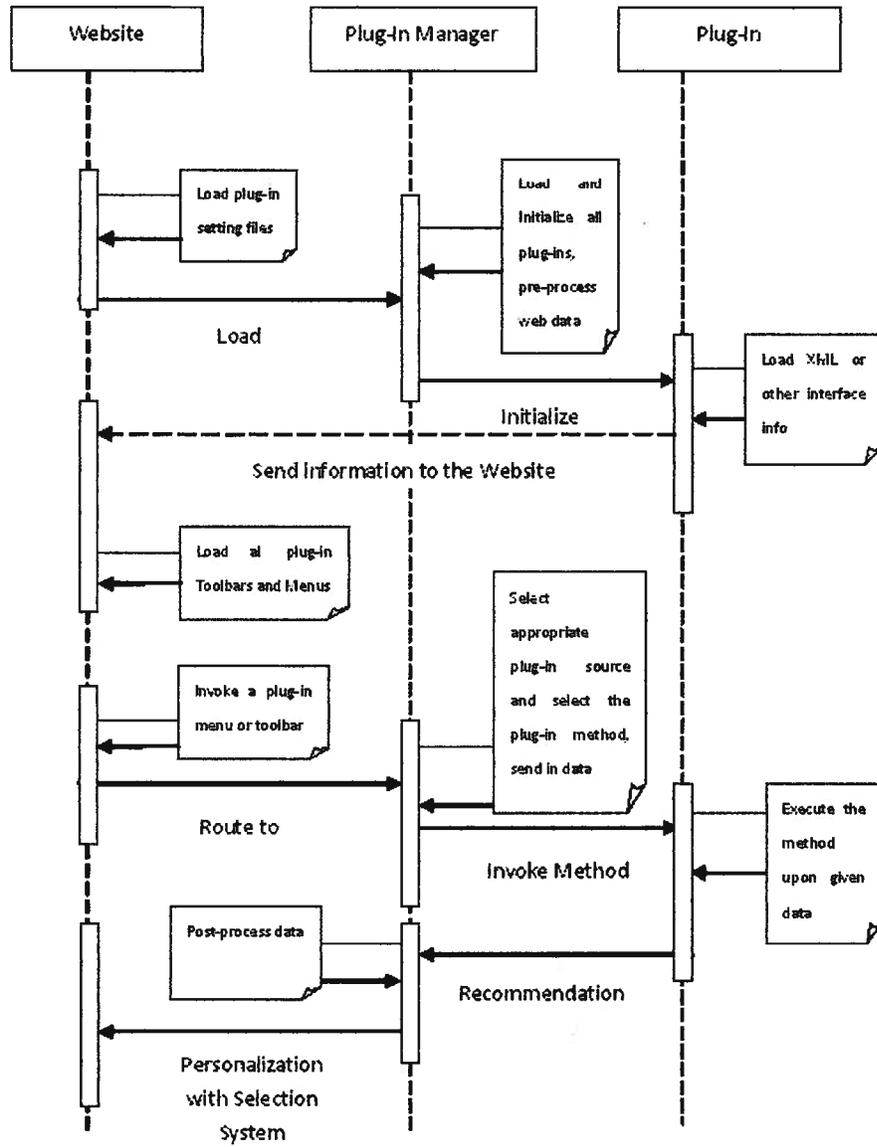


Figure 4.2: Flowchart of Personalization Plug-In Program

## 4.2 Usage Log and Plug-In Manager

During the interactions with users, the website usually stores usage logs of a user's browsing activities such as user profiles, cookies, navigational behaviors and so on. In order to support our personalization system, usage logs are collected and sent to the plug-in manager which is a built-in service that allows third party software to be implemented on the website without changing the original website functionalities in terms of structures, contents and services. The plug-in manager processes and transfers all the necessary information to our personalization system for further operations.

**Usage Log:** Usage log, which in the system is represented by UsageLog.cs object (see Table 4.1), is a collection of user activities such as user profiles and their navigational behaviors. In association with web contents, they form the basic web data for later data preprocessing and association rules finding.

Data	Description
UsageLog - Class Definition	Represents the usage log class, including all the necessary fields such as user login name, category of the contents visited.
UsageLogType - Int	Indicates whether the usage log is a user profile or navigational behavior. User profile is a registration form which reflects that it is the first time the user enters the website. Navigational behavior is a set of web pages that the user visited after he/she logged in.
Duration - Time	Indicates the duration of the activity.

Table 4.1: UsageLog.cs Object

### 4.3. Personalization Plug-In Components

---

**Plug-In Manager:** Plug-In Manager is a service that sits on the web server side for collecting relevant web data for the personalization system. In our system, Plug-In Manager is represented by `PlugInManager.cs` (see Table 4.2), which acts like a double-edged data processor so that when the data including web contents and usage logs are transferred to the personalization system, the manager transforms them to a commonly readable format (e.g.: Navigation Behavior). And when the personalization system generates the recommendation rules and tries to update user profile, the manager reflects the changes to the website in an appropriate way.

Data	Description
PreProcessor - Hashtable (Web Content, User Profile → Navigation Behavior)	Since the association rules finding is based on navigation behaviors (rule based). It is reasonable that we hash the web content and user profile into the format of navigation behavior for simplicity purposes.
PostProcessor - Hashtable (Recommendation → Web Content, User Profile)	Once the recommendations are generated, it is also in the format of navigational behavior (rule based). When we try to update web contents and user system profiles by using these discovered rules, we just need to hash back from the recommended rules or in other words, the navigational behaviors.

Table 4.2: `PlugInManager.cs` Object

### 4.3 Personalization Plug-In Components

The personalization plug-in system contains four major components: association rules analyzer, recommendation pool generator, user profile sliding

### 4.3. Personalization Plug-In Components

---

window and personalization selection system. The first two components provide the user with a better personalization system in term of recommendation accuracy which will be further investigated in chapter 5 (system evaluations and simulations). The latter two components focus mainly on user overall satisfaction of the personalization system which is a novel approach by our system. Let's start with a detailed overview of our personalization system. (See Figure 4.3)

### 4.3. Personalization Plug-In Components

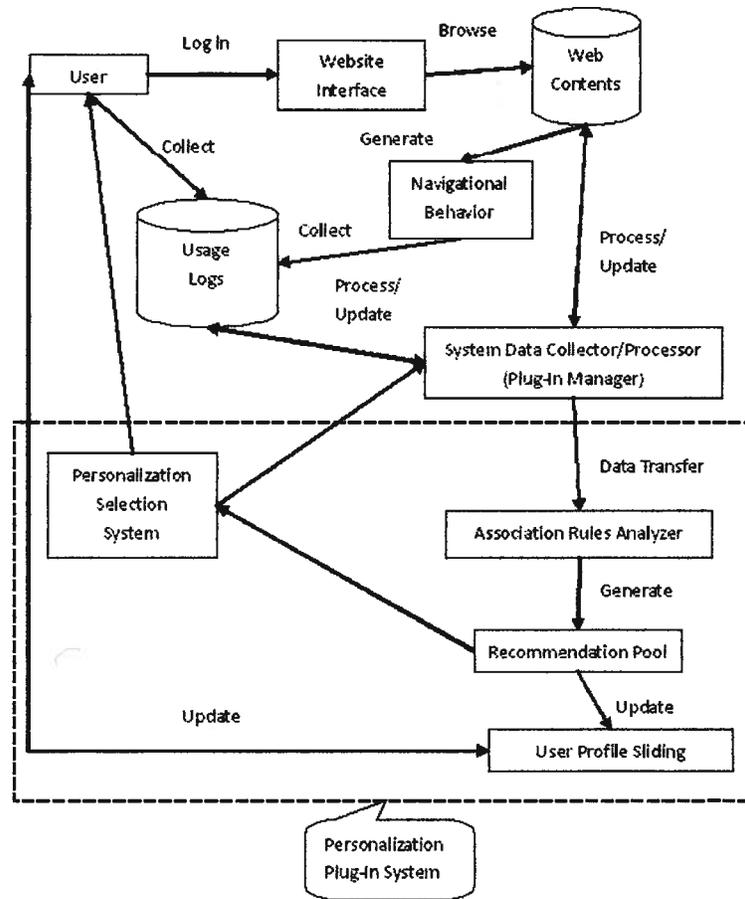


Figure 4.3: Detailed Overview of Our Personalization Plug-In System

### 4.3.1 Association Rules Analyzer

Association rules analyzer is a component that takes the data processed by our plug-in manager and then applies the association rules finding algorithm in order to generate useful recommendations.

Data	Description
Support - Int	Support is the proportion of a certain transactions occurring in the data set which is navigational behaviors in this case.
Confidence - Int	Confidence is the estimated probability of all the transactions in the data set.

Table 4.3: ARAnalyzer.cs

### 4.3.2 Recommendation Pool Generator

Depending on the system requirement, we can set different support values and confidence values for generating recommendations. Since our system is rule/user based, which means each recommendation rule is only for a particular user, sometimes user data set is not enough for generating recommendations. Hence, we iterate through different support and confidence level. For example, we can start support level at 80% and decrease by 10% each time if we cannot find enough recommendations or increase support level if we find too many recommendations. The same approach can be applied on iterating through confidence level too.

After we found enough recommendations, we gather them together and form a recommendation pool categorized by different users. Since the rec-

### 4.3. Personalization Plug-In Components

---

ommendation rules reflect the current user's preference, combining with the previous user profile, we can produce a new user profile which more accurately describes the user in that certain period. We then append this profile to the user profile sliding window.

Data	Description
RecommendationPool (User, RecommendationRules)	The pool contains all the newly found recommendation rules for each user.
NewProfile (CurrentProfile, RecommendationRules)	Creates a new profile based on the current profile and newly found recommendation rules.

Table 4.4: RecommendationPoolGenerator.cs

#### 4.3.3 User Profile Sliding Window

Depending on the user requirement, the size of the sliding window can vary. Initially, it only contains one user profile which is the user registration profile (if the website does not require a registration step, our system will create an implicit one for the user which only contains the user IP/location information). The system learns about the user by the association rules found in the previous steps. If enough knowledge about the user is gathered, we create a new user profile based on the most recent profile and append it to the sliding window. And if the window size exceeds the limit, we remove the oldest one. (See Figure 3.2 in Chapter 3.23)

### 4.3. Personalization Plug-In Components

Data	Description
WindowSize - Int	The system defaults it initially and it could be changed by the user.
InsertProfile (UserProfile,ProfileList-List)	Append the profile to the sliding window and calls remove function to delete the oldest one.
MoveWindow (ProfileNum - Int, ProfileList - List)	The user has the choice to select any one of the profile he feels that describes his current preference the most in the sliding window. We then move all the profiles that are after the one user chose to the place that is before the oldest one in the sliding window. In this case, the one that the user chose becomes the newest profile in the sliding window.
RemoveProfile (UserProfile,ProfileList-List)	Removes the oldest profile in the sliding window.

Table 4.5: ProfileSlideWindow.cs

#### 4.3.4 Personalization Selection System

Once the user has chosen the most suitable profile in the sliding window, our personalization selection system can provide corresponding recommendations from the recommendation pool to the user based on that selected profile.

Data	Description
ProfileMatch (UserProfile, RecommendationPool)	Try to match the recommendations with the user profile. If no such match is found, provide no recommendation at all (which means the user might not want the system to provide personalization)

Table 4.6: PersonalizationSelection.cs

# Chapter 5

## Evaluation

In this chapter, we evaluate our proposed framework over a given system (our simulated game website) and compare the results in terms of recommendation accuracy with different parameter settings. Furthermore, we compare the results of our approach with the traditional approaches which are only based on either dynamic or static web data.

### 5.1 Simulation Setup

**Environmental Setup:** In our simulation, we created a gaming website (say `www.mygames.com`) with 10000 webpages and 100 users initially. The website is pre-categorized into six fields: game age, game price, game language, game platform, game producer and game type. These categories are introduced to both user profiling and web content. However user profile might contain some extra fields such as user name, gender and location. Moreover, some fields of the user profile could contain multiple entries (e.g.: game language). The web contents and use registration profile are randomly generated over these six categories. We also keep a descending priority order over the six categories for the purpose of distinguishing similar rules that are generated by the association rule analyzer. The categories are listed as

## 5.1. Simulation Setup

---

follows in descending order.

**Game Age:** Different game is designed for different age level of the potential players.

**Game Price:** Depending on the quality and all kinds of other factors of the games, games are sold in a variety prices. Also, player's financial situation has to be considered so that he can only afford the game that is within his buying range.

**Game Language:** Represents the language that the game is built on and using. If a game supports multiple languages, we treat them as distinct games. And often user language could be determined from user's location or IP address.

**Game Platform:** Indicates which console the game can be played on. Again, if a game supports multiple platforms, we treat them as separate games.

**Game Producer:** Is the designer of the game. Some games are co-produced by several companies, but in order to make the simulation simpler, we ignore these cases and only choose one major producer for each game.

**Game Type:** Each game can be categorized into one or more types. Again, we only choose the major type for each game.

## 5.1. Simulation Setup

---

Category	Examples
Game Age	7+, 13+, 18+ (some games are too violent to be played by teenagers) and all ages.
Game Price	0(free)-9.99\$, 10\$-19.99\$, 20\$-29.99\$. 30\$-39.99\$, 40\$-49.99\$ and 50\$+.
Game Language	English, Japanese, Chinese, Korean, French and Spanish.
Game Platform	Xbox360, PS3, Wii, PC, NDC, PSP.
Game Producer	EA sports, Square/Enix, Sega, Blizzard, UbiSoft and Namco.
Game Type	Act, RPG (role playing game), SLG (strategy game), Online, ADV (adventure game), PUZ (puzzle game).

Table 5.1: Game Website Category Examples

## 5.1. Simulation Setup

---

**User Profiles and Activities:** In the simulation, we keep three different types of user profile: true profile, registered profile and system profiles. True profile reflects a user's preference at any given time. It is quite stable for a long period of time, but might be updated if the user's preference changes over time. Registered profile (also referred to as user profile #0 in the profile sliding window) is the initial profile that a user tells the website and our system. It is ideally the same as user's true profile, but most of the time the user might not want to disclose his profile at all, or gives misleading information about herself. System profiles are a series of profiles that are learned and updated from our system. They are recorded in the user profile sliding window.

The website is browsed by users on a daily basis, and user navigational behaviors are recorded as follows: each user browses five webpages for each day. Although all the behaviors are generated randomly, the browsing behaviors have to stick to the user's true profile in some way. We define two types of user, one browses webpages according to his true profile completely, and the other wanders around on some other webpages and only adheres to his true profile 80% of the time. The reason we add the latter type of user is to add some system noise in order to make the simulation more realistic (and in reality, people do browse around the website if they are bored or for whatever other personal reasons).

### 5.1.1 Simulation Parameters

Unless otherwise stated, the following parameters will be used for all of our experimentation. For this simulation, we generated a total of 10000

### 5.1. Simulation Setup

---

webpages randomly distributed over the aforementioned six categories and 100 users with random true profiles and registered profiles which may or may not be the same as their true profiles. For each of the users, we use an object array NB (navigational behavior) to store his random navigational behaviors for training days with five webpages per day. The association rule finding algorithm is applied every three and seven days for different results. And initially, the support and confidence level is set to be 80% and 50%, respectively. If not enough association rules are found, we reduce the support level first by 10% until 20% each time and confidence level later, also by 10% each time until 20%. Based on the rules that we found, we generate system user profiles and store them in another object array called SW (sliding window). SW is then sent to users for personalization selection purpose.

The most important parameter we investigate here is the accuracy of the personalization which is defined as the similarity between user's final system profile versus user's true profile. In other words, we compare the category fields of both profiles in order to find all the matching pair(s) (MP). The accuracy (Acc) is then calculated as  $MP/6$ .

On the other hand, in order to compare with the traditional approaches which only concentrate on either user profiles or their navigational behaviors, we calculate the average of the all accuracies over 100 users.

## 5.2 Simulation Results

The simulation is evaluated for a period of 28 days, namely four weeks of the time. For the first set of days of a week (training days), we collect data for the training purposes, and on the following day (testing day), association rule finding algorithm is applied on the training data to generate useful rules and new system profile for the user, then the same procedure is repeated. Accuracy is computed on each of the testing day.

**Average Accuracy:** In this set of experiments, we examine the average accuracy of the personalization system by applying association rule finding algorithm on different testing day interval. First of all, we use three day test interval, each rules and system profile is generated on the third day of the simulation. Secondly, we apply the same procedure on seven day test interval.

## 5.2. Simulation Results

---

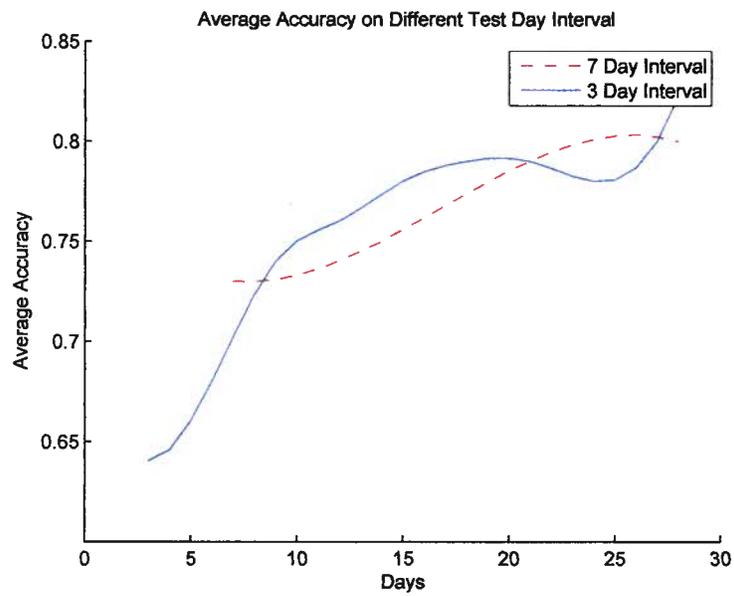


Figure 5.1: Average Accuracy on Different Test Day Interval

## 5.2. Simulation Results

---

We can see from the figure that for most of the period, the three day test interval achieves higher accuracy than the seven day test interval. This is obvious because the smaller interval we have, the more information we gain from the user. But for some of the period, we detect that the average accuracy of three day test interval actually decreases. This is also possible due to the user browsing noise that we have introduced in 5.1. If the interval is too small, the user's browsing noise sometimes can cheat on our system and lower the average accuracy.

In addition, the average accuracy starts above 50%. This proves the correctness of the design of our simulation since due to the randomness, around half of the users adhere to their true profiles and the rest of them give partial or no information about themselves.

**Our Approach vs. Traditional Approaches:** In this test, we compare the average accuracy between our approach and traditional approaches which are either based on static user profile or dynamic user behavior. The experiment is tested on the three day interval.

## 5.2. Simulation Results

---

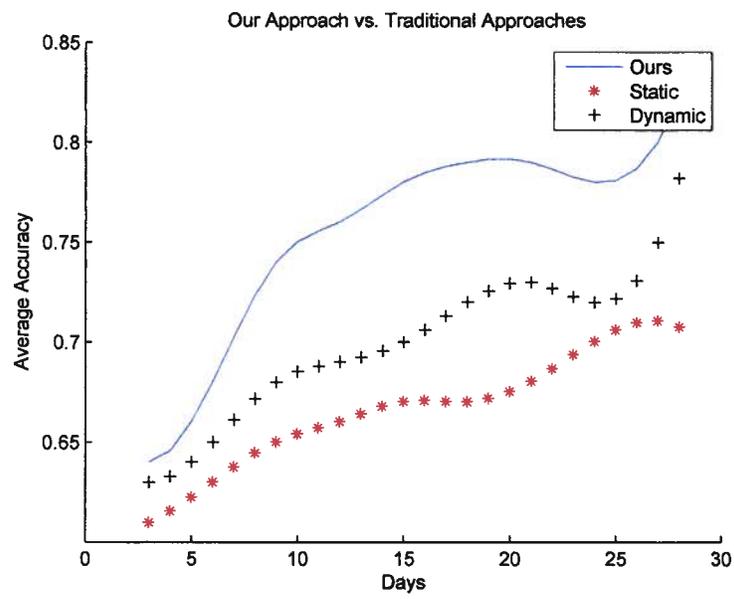


Figure 5.2: Our Approach VS. Traditional Approaches

## 5.2. *Simulation Results*

---

We can see clearly from the figure that our approach outperforms the traditional ones. In the beginning of the experiment, our approach only improves the accuracy by a little bit. But as the days increase, more information is gained from the user both statically and dynamically, our approach improves the accuracy by around 7% when compared to those traditional ones.

Moreover, we can see from the approach that depends solely on static user profile that there is no accuracy decreasing period. It is because that the browsing noise does not have any impact on the static user profile. The curve of dynamic user behavior approach has the similar shape as our approach since it suffers from the browsing noise as well.

## Chapter 6

# Conclusion and Future Work

### 6.1 Conclusion

Despite of the extensive amount of work done in both the academic research area and in the commercial sector on user personalization, most of the approaches are based solely on either static information such as user profile and web content or dynamic user activity such as navigational behavior. Since the interests of a user might change over time and navigational behavior usually contains little information about a user, this strict reliance on one method over another results in a loss in accuracy of recommendations that can be easily addressed.

In this thesis, we propose a novel system for web data based personalization, taking advantage of using association rule finding algorithm on both static information provided by the user and web content and dynamic navigational behavior. Our approach has been extensively tested on a simulated gaming website and the experimental results have confirmed the effectiveness of our design, which leads to a more accurate recommendation generating system for users.

Furthermore, our system takes into consideration of two main concerns:

- (a) The users' overall satisfaction about the personalization they receive

and (b) The very real situation that, regardless of how sophisticated or promising a personalization system is, some users just do not want to use it or may only want to have partial personalization in the favor of their own interests or for the sake of privacy. We tackle these issues by introducing a personalization selection system incorporated with user profile sliding window which allows the user to choose which profile fits his/her interest the most in a given period of time. Personalization is then generated based on the selected profile. This way, a user's preference with regards to the amount and quality of recommendations or personalization is respected, even in cases where little or no personalization is desired.

## 6.2 Future Work

There are some aspects of the system that can benefit from more extensive investigation. First and foremost is the issue of which web data mining algorithm we should employ that would be the most effective and efficient on both static and dynamic user data. In this paper, we used association rule mining technique, which is one of the most popular web data mining strategies for personalization. However, there are many other web data mining techniques such as classification and clustering which may or may not be superior to our approach in terms of effectiveness and efficiency. The issue of data mining timing is also another concern of our system. Better timing strategy could result in high overall recommendation accuracy and better personalization quality. Baoyao Zhou and et, al. [17] proposed an effective approach by using periodic web personalization which is worth studying and

## 6.2. Future Work

---

could be applied on our system in the future.

Cross-websites data personalization is another aspect that would need extensive study. Users browse all kinds of websites everyday and some of these websites are inter-correlated. However, our system is only designed to be implemented on a sole website domain. Once a user leaves the domain, the system stops collecting data and starts generating recommendations for the user. But sometimes, a user's activity on other website domains will have heavy impact on the overall behavior of user. Our system cannot account for this and hence may result in inaccurate recommendations due to lack of information.

A somewhat more adventurous topic of investigation is how to ensure user satisfaction while keeping recommendation accuracy. Herlocker, Konstan, et al[10] used a comprehensive quality measure which not only focuses on accuracy, but takes serendipity and overall recommendation coverage into considerations as well. There are well-known techniques by which algorithms can trade-off reduced serendipity and coverage for improved accuracy (such as only recommending items for which there are many ratings and feedbacks). Since users value all three attributes most of the time, recommendations generated by these algorithms may be more accurate but less useful. Hence, we need a comprehensive quality measure or rating that combines accuracy with other serendipity and coverage factors so that the system can provide sensible trade-offs to serve users better in terms of both user satisfaction and recommendation accuracy.

Last but not least, testing the system on a real website with on-going user activities is another important issue that needs to be explored. Since

## 6.2. Future Work

---

our evaluation is made on simulated user activities over a simulated website, it will be more convincing if we could use real data to prove the accuracy advantage of our system over existing approaches. LePlaza is an ideal social networking website for our system since it contains most of the features of popular social networking websites such as Facebook and Flickr. For example, the event alerting and recommendation system, the friend of common interests finding system and many other location-based features can be used to gather data for our system. A fully functional framework which can be implemented on the LePlaza system is of our highest priorities for any future work.

# Bibliography

- [1] Massimiliano Albanese, Antonio Picariello, Carlo Sansone, and Lucio Sansone. Web personalization based on static information and dynamic user behavior. In *WIDM '04: Proceedings of the 6th annual ACM international workshop on Web information and data management*, pages 80–87, New York, NY, USA, 2004. ACM.
- [2] Sarabjot Singh Anand, Patricia Kearney, and Mary Shapcott. Generating semantically enriched user profiles for web personalization. *ACM Trans. Interet Technol.*, 7(4):22, 2007.
- [3] Sarabjot Singh Anand and Bamshad Mobasher. Introduction to intelligent techniques for web personalization. *ACM Trans. Interet Technol.*, 7(4):18, 2007.
- [4] Ranieri Baraglia and Fabrizio Silvestri. Dynamic personalization of web sites without user intervention. *Commun. ACM*, 50(2):63–67, 2007.
- [5] P. Berkhin. Survey of clustering data mining techniques. 2002.
- [6] An Chen and Huilin Ye. Multiple-level sequential pattern discovery from customer transaction databases. 1(1), 2003.

- [7] Magdalini Eirinaki, Charalampos Lampos, Stratos Paulakis, and Michalis Vazirgiannis. Web personalization integrating content semantics and navigational patterns. In *WIDM '04: Proceedings of the 6th annual ACM international workshop on Web information and data management*, pages 72–79, New York, NY, USA, 2004. ACM.
- [8] Magdalini Eirinaki and Michalis Vazirgiannis. Web mining for web personalization. *ACM Trans. Internet Technol.*, 3(1):1–27, 2003.
- [9] Magdalini Eirinaki and Michalis Vazirgiannis. Web site personalization based on link analysis and navigational patterns. *ACM Trans. Internet Technol.*, 7(4):21, 2007.
- [10] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5–53, 2004.
- [11] Jochen Hipp, Ulrich Güntzer, and Gholamreza Nakhaeizadeh. Algorithms for association rule mining — a general survey and comparison. *SIGKDD Explor. Newsl.*, 2(1):58–64, 2000.
- [12] Qiaozhu Mei and Kenneth Church. Entropy of search logs: how hard is search? with personalization? with backoff? In *WSDM '08: Proceedings of the international conference on Web search and web data mining*, pages 45–54, New York, NY, USA, 2008. ACM.
- [13] Bamshad Mobasher, Honghua Dai, Tao Luo, and Miki Nakagawa. Effective personalization based on association rule discovery from web usage

- data. In *WIDM '01: Proceedings of the 3rd international workshop on Web information and data management*, pages 9–15, New York, NY, USA, 2001. ACM.
- [14] S. Sendhilkumar and T. V. Geetha. Personalized ontology for web search personalization. In *Compute '08: Proceedings of the 1st Bangalore annual Compute conference*, pages 1–7, New York, NY, USA, 2008. ACM.
- [15] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan. Web usage mining: discovery and applications of usage patterns from web data. *SIGKDD Explor. Newsl.*, 1(2):12–23, 2000.
- [16] Christos Tjortjis and John Keane. A classification algorithm for data mining. *Data Knowl. Eng.*, 2412:50–55, 2002.
- [17] Baoyao Zhou, Siu Cheung Hui, and Alvis C. M. Fong. An effective approach for periodic web personalization. In *WI '06: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 284–292, Washington, DC, USA, 2006. IEEE Computer Society.