

Web Mining for Web Personalization

¹Prof. Jharana Paikaray, ²Prof.Santosh Kumar Rath, ³Prof.Smaranika Mohapatra

*Department of Computer Science & Engineering
Gandhi Institute for Education & Technology, Bhubaneswar, Odisha, India*

Abstract: Web personalization is the process of customizing a Web site to the needs of specific users, taking advantage of the knowledge acquired from the analysis of the user's navigational behavior (usage data) in correlation with other information collected in the Web context, namely, structure, content and user profile data. Due to the explosive growth of the Web, the domain of Web personalization has gained great momentum both in the research and commercial areas. In this article we present a survey of the use of Web mining for Web personalization. More specifically, we introduce the modules that comprise a Web personalization system, emphasizing the Web usage mining module. A review of the most common methods that are used as well as technical issues that occur is given, along with a brief overview of the most popular tools and applications available from software vendors. Moreover, the most important research initiatives in the Web usage mining and personalization areas are presented.

Keywords- Web personalization, Web usage mining, user profiling, WWW

1. INTRODUCTION

The continuous growth in the size and use of the World Wide Web imposes new methods of design and development of online services. Most Web structures are large and complicated and users often miss the goal of their inquiry, or receive ambiguous results when they try to navigate through them. On the other hand, the e-business sector is rapidly evolving and the need for Web marketplaces that anticipate the needs of the customers is more evident than ever. Web personalization is defined as any action that adapts the information or services provided by a Web site to the needs of a particular user or a set of users, taking advantage of the knowledge gained from the users' navigational behavior and individual interests, in combination with the content and the structure of the Web site. The objective of a Web personalization system is to "provide users with the information they want or need, without expecting from them to ask for it explicitly". At this point, it is necessary to stress the difference between layout customization and personalization. In customization the site can be adjusted to each user's preferences regarding its structure and presentation. In personalization systems modifications concerning the content or even the structure of a Web site are performed dynamically. Principal elements of Web personalization include (a) the categorization and preprocessing of Web data, (b) the extraction of correlations between and across different kinds of such data, and (c) the determination of the actions that should be recommended by such a personalization system.

Web data are those that can be collected and used in the context of Web personalization. These data are classified in four categories.

- Content data
- Structure data
- Usage data
- User profile data .

The overall process of usage-based Web personalization consists of five modules. These are as follows:

- User profiling: In the Web domain, user profiling is the process of gathering information specific to each visitor, either explicitly or implicitly.
- Log analysis and Web usage mining: This is the procedure where the information stored in Web server logs is processed by applying data mining techniques in order to (a) extract statistical information and discover interesting usage patterns, (b) cluster the users into groups according to their navigational behavior, and (c) discover potential correlations between Web pages and user groups. It is therefore evident that the user profiling and Web usage mining modules overlap.
- 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.
- Web site publishing: A publishing mechanism is used 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.

- Information acquisition and searching: In many cases information provided by a Web site is not physically stored in the Web site's server. In the case of a Web portal or vortal (vertical portal), users are interested in information from various Web sources. Searching and relevance ranking techniques must be employed both in the process of acquisition of relevant information and in the publishing of the appropriate data to each group of users.

A usage-based Web personalization system utilizes Web data in order to modify a Web site. Site personalization is achieved through the interaction of the aforementioned modules. This survey article is organized as follows. In Section 2, we provide a brief description of the Web personalization process and illustrate the interaction of these modules in such a system. In the context of this survey we analyze user profiling, as well as log analysis and Web usage mining modules. These modules are described in more detail in Sections 3 and 4, respectively. In Section 5 we present the most important research initiatives in the area of Web usage mining and personalization. In Appendices A and B lists of acronyms and abbreviations as well as Web References are presented.

In Appendix C there is a tabular comparative presentation of the most representative tools for user profiling and Web usage mining, as well as the most important research initiatives in the area of Web mining and Web personalization.

2. WEB PERSONALIZATION

Web site personalization can be defined as the process of customizing the content and structure of a Web site to the specific and individual needs of each user taking advantage of the user’s navigational behavior. The steps of a Web personalization process include: (a) the collection of Web data, (b) the modeling and categorization of these data (preprocessing phase), (c) the analysis of the collected data, and (d) the determination of the actions that should be performed. The ways that are employed in order to analyze the collected data include content-based filtering, collaborative filtering, rule-based filtering, and Web usage mining. The site is personalized through the highlighting of existing hyperlinks, the dynamic insertion of new hyperlinks that seem to be of interest for the current user, or even the creation of new index pages. Content-based filtering systems are solely based on individual users’ preferences. The system tracks each user’s behavior and recommends items to them that are similar to items the user liked in the past. Collaborative filtering systems invite users to rate objects or divulge their preferences and interests and then return information that is predicted to be of interest to them. This is based on the assumption that users with similar behavior (e.g. users that rate similar objects) have analogous interests. In rule-based filtering the users are asked to answer a set of questions. These questions are derived from a decision tree, so as the user proceeds to answer them, what he finally receives as a result (e.g. a list of products) is tailored to his needs. Content-based, rule-based, and collaborative filtering may also be used in combination, for deducing more accurate conclusions. In this work we focus on Web usage mining. This process relies on the application of statistical and data mining methods to the Web log data, resulting in a set of useful patterns that indicate users’ navigational behavior. The data mining methods that are employed are: association rule mining, sequential pattern discovery, clustering, and classification. This knowledge is then used from the system in order to personalize the site according to each user’s behavior and profile. The block diagram illustrated in Figure 1 represents the functional architecture of a Web personalization system in terms of the modules and data sources that were described earlier. The content management module processes the Web site’s content and classifies it in conceptual categories. The Web site’s content can be enhanced with additional information acquired from other Web sources, using advanced search techniques. Given the site map structure and the usage logs, a Web usage miner provides results regarding usage patterns, user behavior, session and user clusters, clickstream information, and so on. Additional information about the individual users can be obtained by the user profiles. Moreover, any information extracted from the Web usage mining process concerning each user’s

navigational behavior can then be added to her profile. All this information about nodes, links, Web content, typical behaviors, and patterns is conceptually abstracted and classified into semantic categories. Any information extracted from the interrelation between knowledge acquired using usage mining techniques and knowledge acquired from content management will then provide the framework for evaluating possible alternatives for restructuring the site. A publishing mechanism will perform the site modification, ensuring that each user navigates through the optimal site structure. The available content options for each user will be ranked according to the user’s interests.

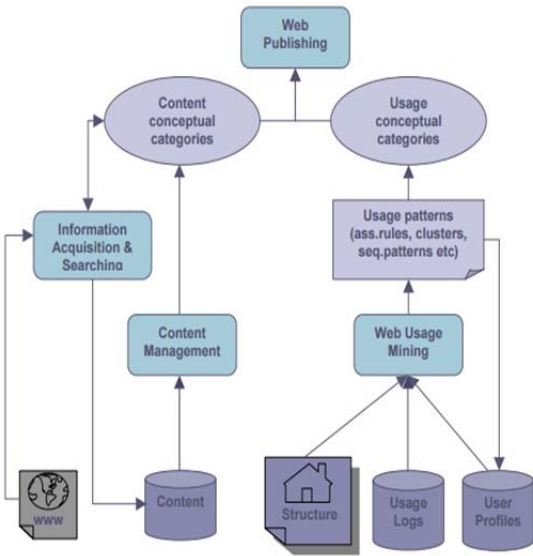


Fig. 1: Modules of a Web personalization system.

3. USER PROFILING

In order to personalize a Web site, the system should be able to distinguish between different users or groups of users. This process is called user profiling and its objective is the creation of an information base that contains the preferences, characteristics, and activities of the users. In the Web domain and especially in e-commerce, user profiling has been developed significantly because Internet technologies provide easier means of collecting information about the users of a Web site, which in the case of e-business sites are potential customers. A user profile can be either static, when the information it contains is never or rarely altered (e.g., demographic information), or dynamic when the user profile’s data change frequently. Such information is obtained either explicitly, using online registration forms and questionnaires resulting in static user profiles, or implicitly, by recording the navigational behavior and/or the preferences of each user, resulting in dynamic user profiles. In the latter case, there are two further options: either regarding each user as a member of a group and creating aggregate user profiles, or addressing any changes to each user individually. When addressing the users as a group, the method used is the creation of aggregate user profiles based on rules and patterns extracted by applying Web usage mining techniques to Web server logs. Using this knowledge, the Web site can

be appropriately customized. This case is discussed in detail in Section 4, therefore it won't be further analyzed here. In the following sections, we provide a description of several methods for implicit and explicit collection of user profile data. Privacy issues that arise in the user profiling process are discussed, and an overview of available tools and user profiling applications is presented.

3.1 Data Collection

A way of uniquely identifying a visitor through a session is by using cookies. W3C [WCA] defines cookie as "the data sent by a Web server to a Web client, stored locally by the client and sent back to the server on subsequent requests." In other words, a cookie is simply an HTTP header that consists of a text-only string, which is inserted into the memory of a browser. It is used to uniquely identify a user during Web interactions within a site and contains data parameters that allow the remote HTML server to keep a record of the user identity, and what actions he takes at the remote Web site. The contents of a cookie file depend on the Web site that is being visited. In general, information about the visitor's identification is stored, along with password information. Additional information such as credit card details, if one is used during a transaction, as well as details concerning the visitor's activities at the Web site, for example, which pages were visited, which purchases were made, or which advertisements were selected, can also be included. Often, cookies point back to more detailed customer information stored at the Web server. Another way of uniquely identifying users through a Web transaction is by using identd, an identification protocol specified in RFC 1413 [RFC] that provides a means to determine the identity of a user of a particular TCP connection. Given a TCP port number pair, it returns a character string, which identifies the owner of that connection (the client) on the Web server's system. Finally, a user can be identified making the assumption that each IP corresponds to one user. In some cases, IP addresses are resolved into domain names that are registered to a person or a company, thus more specific information is gathered. As already mentioned, user profiling information can be explicitly obtained by using online registration forms requesting information about the visitor, such as name, age, sex, likes, and dislikes. Such information is stored in a database, and each time the user logs on the site, it is retrieved and updated according to the visitor's browsing and purchasing behavior.

3.2 Privacy Issues

The most important issue that should be encountered during the user profiling process is privacy violation. Many users are reluctant to give away personal information either implicitly as mentioned before, or explicitly, being hesitant to visit Web sites that use cookies (if they are aware of their existence) or avoiding disclosure of personal data in registration forms. In both cases, the user loses anonymity and is aware that all of his actions will be recorded and used, in many cases without his consent. In addition, even if a user has agreed to supply personal information to a site, through cookie technology such information can be

exchanged between sites, resulting in its disclosure without the user's permission.

P3P (Platform for Privacy Preferences) is a W3C proposed recommendation [P3P] that suggests an infrastructure for the privacy of data interchange. This standard enables Web sites to express their privacy practices in a standardized format that can be automatically retrieved and interpreted by user agents. Therefore, the process of reading privacy policies will be simplified for the users, because key information about what data are collected by a Web site can be automatically conveyed to a user, and discrepancies between a site's practices and the user's preferences concerning the disclosure of personal data will be automatically flagged. P3P, however, does not provide a mechanism for ensuring that sites actually act according to their policies.

3.3 Tools and Applications

In this section we present some of the most popular Web sites that use methods such as decision tree guides, collaborative filtering, and cookies in order to profile users and create customized Web pages. Additionally, a brief description of the most important tools available for user profiling is given. An overview along with products references is provided in Appendix Table AI.

Popular Web sites such as Yahoo! [YAH], Excite [EXC], or Microsoft Network [MSN] allow users to customize home pages based on their selections of available content, using information supplied by the users and cookies thereafter. In that way, each time the user logs on the site, what she sees is a page containing information addressed to her interests.

Rule-based filtering is used from online retailers such as Dell [DEL] and Apple Computer [APP], giving users the ability to easily customize product configurations before ordering. As far as recommendation systems are concerned, the most popular example is Amazon.com [AMA]. The system analyzes past purchases and posts suggestions on the shopper's customized recommendations page. Users who haven't made a purchase before can rate books and see listings of books they might like. The same approach, based on user ratings, is used in many similar online shops, such as CDNOW [CDN].

Another interesting approach is that of Food.com [FOO]. Users are not required to fill in any form to order food from a specific nearby restaurant. Customization happens automatically as users give the necessary information for a food delivery or pickup, because zip code data provide the necessary information for suggesting nearby restaurants.

Commercial Web sites, including many search engines such as Alta-Vista [ALT] or Lycos [LYC], have associations with commercial marketing companies such as Double Click Inc. [DCL]. These sites use cookies to monitor their visitors' activities, and any information collected is stored as a profile in Double Click's database. Double Click then uses this profile information to decide which advertisements or services should be offered to each user when he visits one of the affiliated Double Click sites. Of course, this information is collected and stored without the users' knowledge and more importantly, consent.

4. LOG ANALYSIS AND WEB USAGE MINING

The purpose of Web usage mining is to reveal the knowledge hidden in the log files of a Web server. By applying statistical and data mining methods to the Web log data, interesting patterns concerning the users' navigational behavior can be identified, such as user and page clusters, as well as possible correlations between Web pages and user groups. The Web usage mining process can be regarded as a three-phase process, consisting of the data preparation, pattern discovery and pattern analysis phases [Srivastava et al. 2000]. In the first phase, Web log data are preprocessed in order to identify users, sessions, pageviews, and so on. In the second phase, statistical methods, as well as data mining methods (such as association rules, sequential pattern discovery, clustering, and classification) are applied in order to detect interesting patterns. These patterns are stored so that they can be further analyzed in the third phase of the Web usage mining process. A description of the fields included in a log entry of a Web usage log follows, along with a set of definitions of Web data abstractions, such as Web site, user, session, pageviews, and clickstreams. Technical issues, concerning data preparation are discussed. A more detailed analysis of the methods employed in the Web usage mining process, including simple log analysis is presented. Finally, a brief overview of the commercially available tools and applications specializing in log analysis or Web usage mining is given.

4.1 Web Log

Each access to a Web page is recorded in the access log of the Web server that hosts it. The entries of a Web log file consist of fields that follow a predefined format. The fields of the common log format are: remote host rfc931 authuser date "request" status bytes

where remote host is the remote hostname or IP number if the DNS hostname is not available; rfc931 the remote log name of the user; authuser, the username with which the user has authenticated himself, available when using password-protected WWW pages; date, the date and time of the request; "request", the request line exactly as it came from the client (the file, the name, and the method used to retrieve it); status, the HTTP status code returned to the client, indicating whether the file was successfully retrieved and if not, what error message was returned; and bytes, the content-length of the documents transferred. If any of the fields cannot be determined a minus sign (-) is placed in this field.

Lately, W3C [W3Clog] presented an improved format for Web server log files, called the "extended" log file format, partially motivated by the need to support the collection of data for demographic analysis and for log summaries. This format permits customized log files to be recorded in a format readable by generic analysis tools. The main extension to the common log format is that a number of fields are added to it. The most important are: referrer, which is the URL the client was visiting before requesting that URL, user_agent, which is the software the client claims to be using, and cookie, in the case where the site visited uses cookies.

In general, extended log format consists of a list of prefixes such as c (client), s (server), r (remote), cs (client to server), sc (server to client), sr (server to remote server, used by proxies), rs (remote server to server, used by proxies), x (applicationspecific identifier), and a list of identifiers such as date, time, ip, dns, bytes, cached (records whether a cache hit occurred), status, comment (comment returned with status code), method, uri, uri-stem and uri-query. Using a combination of some of the aforementioned prefixes and identifiers, additional information such as referrers' IPs, or keywords used in search engines can be stored.

4.2 Web Data Abstractions

In the Web domain, several abstractions are mentioned, concerning Web usage, content, and structure. The W3C Web Characterization Activity [WCA] has published a draft establishing precise semantics for concepts such as Web site, user, user sessions, server sessions, pageviews, and clickstreams. A Web site is defined as a collection of interlinked Web pages, including a host page, residing at the same network location. A user is defined to be the principal using a client to interactively retrieve and render resources or resource manifestations. In the Web context, a user is an individual that is accessing files from a Web server, using a browser. A user session is defined as a delimited set of user clicks across one or more Web servers. A server session is defined as a collection of user clicks to a single Web server during a user session. It is also called a visit. A pageview is defined as the visual rendering of a Web page in a specific environment at a specific point in time. In other words, a pageview consists of several items, such as frames, text, graphics, and scripts that construct a single Web page. A clickstream is a sequential series of pageview requests, made from a single user.

4.3 Data Preprocessing

There are some important technical issues that must be taken into consideration during this phase in the context of the Web personalization process, because it is necessary for Web log data to be prepared and preprocessed in order to use them in the consequent phases of the process. An extensive description of data preparation and preprocessing methods can be found in Cooley et al. [1999a]. In the sequel, we provide a brief overview of the most important ones.

The first issue in the preprocessing phase is data preparation. Depending on the application, Web log data may need to be cleaned from entries involving pages that returned an error or graphics file accesses. In some cases such information might be useful, but in others such data should be eliminated from a log file. Furthermore, crawler activity can be filtered out, because such entries do not provide useful information about the site's usability. Another problem to be met has to do with caching. Accesses to cached pages are not recorded in the Web log, therefore such information is missed. Caching is heavily dependent on the client-side technologies used and therefore cannot be dealt with easily. In such cases, cached

pages can usually be inferred using the referring information from the logs. Moreover, a useful aspect is to perform page view identification, determining which page file accesses contribute to a single pageview. Again such a decision is application-oriented.

Most important of all is the user identification issue. There are several ways to identify individual visitors. The most obvious solution is to assume that each IP address (or each IP address/client agent pair) identifies a single visitor. Nonetheless, this is not very accurate because, for example, a visitor may access the Web from different computers, or many users may use the same IP address (if a proxy is used). A further assumption can then be made, that consecutive accesses from the same host during a certain time interval come from the same user. More accurate approaches for a priori identification of unique visitors are the use of cookies or similar mechanisms or the requirement for user registration. However, a potential problem in using such methods might be the reluctance of users to share personal information.

Assuming a user is identified, the next step is to perform session identification, by dividing the clickstream of each user into sessions. The usual solution in this case is to set a minimum timeout and assume that consecutive accesses within it belong to the same session, or set a maximum timeout, where two consecutive accesses that exceed it belong to different sessions.

4.4 Log Analysis

Log analysis tools (also called traffic analysis tools) take as input raw Web data and process them in order to extract statistical information. Such information includes statistics for the site activity (such as total number of visits, average number of hits, successful/failed/redirected/cached hits, average view time, and average length of a path through a site), diagnostic statistics (such as server errors, and page not found errors), server statistics (such as top pages visited, entry/exit pages, and single access pages), referrers statistics (such as top referring sites, search engines, and keywords), user demographics (such as top geographical location, and most active countries/cities/organizations), client statistics (visitor's Web browser, operating system, and cookies), and so on. Some tools also perform click stream analysis, which refers to identifying paths through the site followed by individual visitors by grouping together consecutive hits from the same IP, or include limited low-level error analysis, such as detecting unauthorized entry points or finding the most common invalid URL. These statistics are usually output to reports and can also be displayed as diagrams.

This information is used by administrators for improving the system performance, facilitating the site modification task, and providing support for marketing decisions [Srivastava et al. 2000]. However, most advanced Web mining systems further process this information to extract more complex observations that convey knowledge, utilizing data mining techniques such as association rules and sequential pattern discovery, clustering, and classification. These techniques are described in more detail in the next paragraph.

4.5 Web Usage Mining

Log analysis is regarded as the simplest method used in the Web usage mining process. The purpose of Web usage mining is to apply statistical and data mining techniques to the preprocessed Web log data, in order to discover useful patterns. As mentioned before, the most common and simple method that can be applied to such data is statistical analysis. More advanced data mining methods and algorithms tailored appropriately for use in the Web domain include association rules, sequential pattern discovery, clustering, and classification.

Association rule mining is a technique for finding frequent patterns, associations, and correlations among sets of items. Association rules are used in order to reveal correlations between pages accessed together during a server session. Such rules indicate the possible relationship between pages that are often viewed together even if they are not directly connected, and can reveal associations between groups of users with specific interests. Aside from being exploited for business applications, such observations also can be used as a guide for Web site restructuring, for example, by adding links that interconnect pages often viewed together, or as a way to improve the system's performance through prefetching Web data.

Clustering is used to group together items that have similar characteristics. In the context of Web mining, we can distinguish two cases, user clusters and page clusters. Page clustering identifies groups of pages that seem to be conceptually related according to the users' perception. User clustering results in groups of users that seem to behave similarly when navigating through a Web site. Such knowledge is used in ecommerce in order to perform market segmentation but is also helpful when the objective is to personalize a Web site.

5. CONCLUSION

User profiling is the process of collecting information about the characteristics, preferences, and activities of a Web site's visitors. This can be accomplished either explicitly or implicitly. Explicit collection of user profile data is performed through the use of online registration forms, questionnaires, and the like. The methods that are applied for implicit collection of user profile data vary from the use of cookies or similar technologies to the analysis of the users' navigational behavior that can be performed using Web usage mining techniques. It is evident that in order to personalize a Web site, user profiling is essential. However, all the techniques that are used for this purpose have some drawbacks. The users' privacy violation is the most important issue that should be addressed. P3P is a standard that enables Web sites to express their privacy practices in a standardized format that can be automatically retrieved and interpreted by user agents. In that way, the process of reading the privacy statements of the Web sites becomes simpler, however, P3P does not provide a guarantee that these sites act according to these declared policies. The extraction of information concerning the navigational behavior of Web site visitors is the objective of Web usage mining. Nevertheless this process can also be regarded as part of the creation of user profiles; it is therefore evident

that those two modules overlap and are fundamental in the Web personalization process.

Web usage mining is the process of applying statistical and data mining methods to Web log data in order to extract useful patterns concerning the users' navigational behavior, user and page clusters, as well as possible correlations between Web pages and user groups. The discovered rules and patterns can then be used for improving the system's performance or for making modifications to the Web site. The information included in the Web logs can also be integrated with customer data collected from CRM and ERP systems, in order to gather business intelligence. Several issues must be taken into consideration, including decisions to be made during data filtering and processing, user and session identification, and pageview identification. Another important issue is the choice of the data mining methods that should be used. Web usage mining lately has been used in combination with other technologies, such as user profiling and in some cases content mining, in order to provide a more integrated view of the usage of a Web site, and make personalization more effective.

REFERENCES

- [1]. BERENDT, B. 2000. Web usage mining, site semantics, and the support of navigation. In Proceedings of the Workshop WEBKDD'2000 Web Mining for E-Commerce—Challenges and Opportunities, Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Boston, August).
- [2]. BERENDT, B. 2001. Understanding Web usage at different levels of abstraction: Coarsening and visualizing sequences. In Proceedings of the Workshop WEBKDD 2001 Mining Log Data Across All Customer TouchPoints, Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (San Francisco, August).
- [3]. BORGES, J. AND LEVENE, M. 1999. Data mining of user navigation patterns. In Web Usage Analysis and User Profiling, Lecture Notes in Computer Science, vol. 1836, Springer-Verlag New York, 92–111.
- [4]. BUCHNER, A. AND MULVENNA, M. D. 1998. Discovering Internet marketing intelligence through online analytical Web usage mining. SIGMOD Rec. 27, 4, 54–61. 21
- [5]. BUCHNER, A.G., BAUMGARTEN, M., ANAND, S.S., MULVENNA, M.D., AND HUGHES, J.G. 1999. Navigation pattern discovery from Internet data. In Proceedings of the Web Usage Analysis and User Profiling Workshop (WEBKDD'99), Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (San Diego, August), 25–30.
- [6]. CHEN, M. S., PARK, J. S., AND YU, P. S. 1996. Data mining for path traversal patterns in a web environment. In Proceedings of the Sixteenth International Conference on Distributed Computing Systems (May), 385–392.
- [7]. CINGIL, I., DOGAC, A., AND AZGIN, A. 2000. A broader approach to personalization. Commun. ACM, 43, 8 (August), 136–141.
- [8]. COENEN, F., SWINNEN, G., VANHOOF, K., AND WETS, G. 2000. A framework for self adaptive websites: Tactical versus strategic changes. In proceedings of WEBKDD'2000 Web Mining for ECommerce—Challenges and Opportunities, Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Boston, August).
- [9]. COOLEY, R., MOBASHER, B., AND SRIVASTAVA, J. 1999a. Data preparation for mining world wide web browsing patterns. Knowl. Inf. Syst., 1, 1 (Feb.).
- [9]. COOLEY, R., TAN, P.-N., AND SRIVASTAVA, J. 1999b. WebSIFT: The web site information filter system. In Proceedings of the Web Usage Analysis and User Profiling Workshop (WEBKDD'99), Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Boston, August).
- [10]. HUANG, Z., NG, J., CHEUNG, D.W., NG, M.K., AND CHING, W-K. 2001. A cube model for web access sessions and cluster analysis. In Proceedings of the Mining Log Data Across All Customer TouchPoints Workshop (WEBKDD'01) Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (San Francisco, August).
- [11]. JOACHIMS, T., FREITAG, D., AND MITCHELL, T. 1997. WebWatcher: A tour guide for the world wide web. In Proceedings of IJCAI97 (August).
- [12]. JOSHI, A., JOSHI, K., AND KRISHNAPURAM, R. 2000. On mining web access logs. In Proceedings of the 2000 ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, 63–69.
- [13]. KRISHNAPURAM, R., JOSHI, A., NASRAOUI, O., AND YI, L. 2003. Low-complexity fuzzy relational clustering algorithms for web mining, IEEE Trans. Fuzzy Syst. 9, 4, 596-607.