Multi-agent System for Web Advertising

Przemysław Kazienko[1]

[1] Wrocław University of Technology, Institute of Applied Informatics,
Wybrzeże S. Wyspiańskiego 27, 50-370 Wrocław, Poland
kazienko@pwr.wroc.pl

Abstract. The main aim of a personalized advertising system is to provide advertisements, which are most suitable for the given anonymous user navigating the web site. To achieve this goal, many sources of data are processed in one coherent vector space: the advertisers' and publisher's web site content, sessions of former users from the past, the history of clicks on banners and the current user behavior as well as the advertising policy promoting certain campaigns. The multi-agent system running mostly on the publisher site is introduced to organize personalized advertising. Each cooperating agent is responsible for a separate, specific task: web content and usage mining, click-through data exploration, user monitoring, advertisement recommendation and management.

1 Introduction

Web advertising is the major or often even sole source of income for recent online portals. There are two or three main participants of web advertising: advertisers, publishers and sometimes advertising brokers, which link both advertisers and publishers. The advertisers want to attract people to visit their web sites, so they pay publishers for exposition their banners or other kinds of advertisements on publisher's pages. Users, that navigate such pages, sometimes click through on a banner, so they shift to advertiser's site. The main task of the advertising broker, such as Real Media [1], DoubleClick [17] or Google's AdSense [7], is to deliver banners appropriate for the given publisher's page.

There is no broker in two other models of advertising, in which the whole advertising management is performed by either advertisers or publishers themselves [4]. In this paper, we focus on this third model: the publisher possesses all advertising data, i.e. data about banners, and the publisher is responsible for presenting adequate advertisements on its pages. Such model enables the usage of specific information usually available only for the publisher: data about users and their behaviors.

There are two main research issues in web advertising: scheduling and personalization. The main aim of the former is to maximize the total click-through-rate for all advertisements by proper managing exposition time and advertising space on the web page [2, 5, 18]. The latter seems to be an important challenge for current advertisers. It appears to be more "individualized" than typical target advertising, which partitions customers in a market into specific segments [8]. It aims to assign a suitable advertisement to a single web user rather than to a group of individuals [3, 16].

This paper is part of research carried out on personalized recommendation systems [9, 10, 12, 13, 14, 15]. Advertisements are only one, although specific, kind of objects, which can be suggested to a user.

2 Personalized Web Advertising

The general concept of personalized online advertising was originally presented in [11]. There are five main factors taken into account at personalization: content of both publisher's pages and advertisers' portals, historical behaviors of users, the behavior of the current user, recently exposed advertisements, and advertising features (Fig. 1).

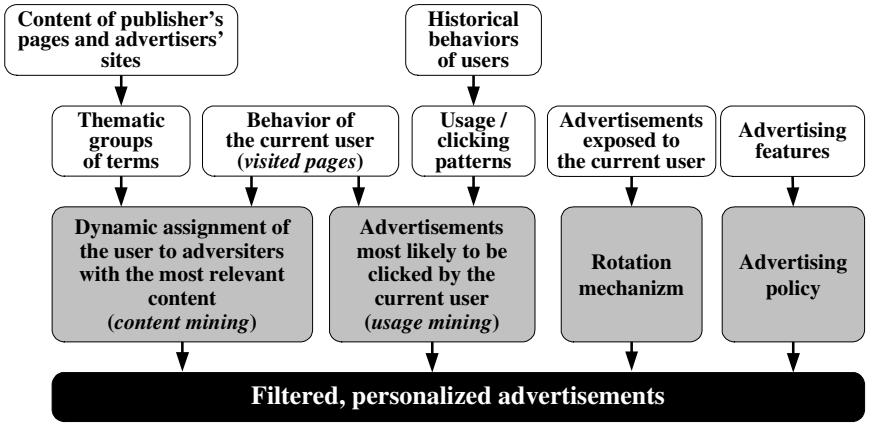

Fig. 1. Main factors of filtering and personalization of advertisement

Historical behaviors are derived from the past user sessions. A session is the set of pages watched by the user during one visit to the publisher's web site. Simultaneously, the data about banners clicked during the session is retained. Thus, we have exactly one ad usage session corresponding to each page user session. Usage patterns are obtained by clustering user sessions offline. Due to one-to-one relationship between user and ad session, we receive also clicking patterns related to typical usage patterns. This course of action is called *web* and *ad usage mining*.

The content is processed in a sense alike. Content features, in the form of terms, are extracted from HTML source of publisher's pages. Next, selected terms are clustered to achieve thematic groups, called *conceptual spaces* [13]. Since we cluster terms, individual publisher's pages may belong to many clusters. Besides, we take the content of the whole web sites linked by particular benners. Each advertiser's web site is treated like one page from publisher's portal. Advertiser's web sites that contain terms from a thematic group are allocated to this group. In this way we establish a one-to-one relationship between the content of publisher's site and advertisements.

Having content and usage clusters, we can dynamically assign each single user to the most relevant patterns based on their current behavior. This assignment is per-

formed online and at each user HTTP request. In consequence, the user is suggested advertisements with the content most relevant to the content of pages viewed by them recently. Additionally, advertisements most likely to be clicked by the user have the greater chance to be exposed. It comes from the historical behavior of other users that are similar to the current one: they simply used to click certain banners. Note that the user is assigned to one usage pattern but this pattern corresponds to one clicking pattern.

Another personalization factor is related to rotation mechanism. All advertisements presented to the current user are retained to prevent from the emission of the same banner too frequently for one user. Furthermore, there are some advertising features, such as the maximum number of exposition per user, that force the system to monitor all advertisements suggested to each user. Additionally, advertising policy includes also data that enable to promote certain campaigns, e.g. those more profitable.

Some mentioned above processes such as content processing, clustering, i.e. pattern discovery, are carried out offline, while some are performed online: the assignment of the user to the most suitable patterns, monitoring of the current user behavior. Due to permanent changes in source data: new advertisements, updates in content, etc., the offline activities have to be periodically repeated.

All data in the system is processed in one coherent vector space – this facilitates the integration of heterogeneous data sources. There are two kinds of vectors: with dimension equal to either the number of pages in publisher's site or the number of advertisements.

3 The Multi-agent Architecture

The advertising system was designed using a multi-agent architecture, in which expert-agents cooperate with one another and may be distributed among many hosts [6]. Every agent is responsible for a single task and it encapsulates specific functions that are available for the rest of the system. Not only do they interchange agents' information, but they also possess their own knowledge. Some agents are involved in usage mining processes while the other ones are used to discover useful knowledge from the content (Fig. 2). The appropriate usage and content patterns are assigned to the current user so that the system is able to recommend the best, personalized advertisements and additionally the suitable hyperlinks i.e. to suggest "the next navigational steps". Let us focus on tasks accomplished by the individual agents.

Web Crawler retrieves the web site content, i.e. individual pages, using HTTP and following hyperlinks existing on pages. Next, it extracts terms from the HTML content and calculates the document-term frequency useful at clustering. *Web Crawler* is used to analyze both the publisher's and advertisers' web sites linked by the advertisements. It delivers data for *Publisher* and *Advertiser Content Manager*.

Search Engine is an auxiliary agent normally used for Information Retrieval within the content of the publisher's site. Nevertheless, it also retains all terms used by users for searching. *Publisher Content Manager* uses these terms at selection of the proper descriptors.

Ad Manager includes user interface and provides some advertising campaign management tools. It keeps all advertisement features such as the total number of expositions and number of expositions per user, etc.

Publisher Content Manager selects terms from the publisher's content, which have been delivered by *Web Crawler*, to extract only "good descriptors" of the content. Besides, it creates the vector representations of these terms and determines their significance according to their frequency.

Advertiser Content Manager uses capabilities of *Publisher Content Manager* and performs analogical operations on the terms retrieved by *Web Crawler* from whole advertisers' web sites. To achieve it, the agent communicates with *Ad Manager* to obtain URL addresses of individual advertisements.

Content Miner clusters term vectors from publisher's pages and it calculates the mean vector (centroid) of each cluster.

Advertiser Content Miner assigns advertisers' sites to clusters provided by Content Miner based on the terms, which are common for both advertisers' and publisher's sites. In this way, we obtain equivalent content clusters for advertisements.

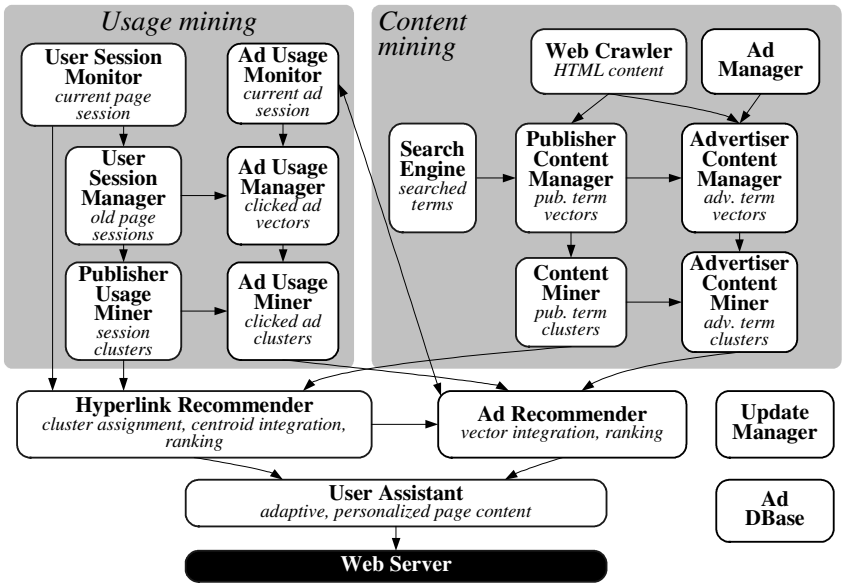

Fig. 2. Multi-agent architecture of AdROSA system

User Session Monitor keeps track of user activities i.e. visited pages. This information reflects the current user behavior and it is crucial at the assignment of the user to the appropriate content and usage patterns. Finally, the session data is stored by *User Session Manager* after the user leaves the publisher's web site.

User Session Manager creates and updates session vectors. Users often change their behavior, so we should not rely on older sessions with the same confidence as on newer ones. For that reason, *User Session Manager* periodically decreases coefficients of older sessions [10].

Publisher Usage Miner clusters historical user sessions and then calculates their centroids representing typical usage patterns. One of such patterns is assigned to a user at each HTTP request.

Ad Usage Monitor controls displayed and clicked banners: it checks exposition limits (total and per one user), and retains data useful for the rotation mechanism – the current ad session. It also relays data about exposed advertisements to *Ad Manager* and additionally, it passes on the clicked advertisements to *Ad Usage Manager* after the session is finished.

Ad Usage Manager creates and maintains the historical clicked ad vectors.

Ad Usage Miner clusters clicked ad vectors and it calculates clicking patterns - centroids. Clustering is based on the usage patterns previously delivered by *Publisher Usage Miner* rather than on source clicked ad vectors. Nevertheless, these vectors are used at calculating of centroids, which represent the advertisements most likely to be clicked by users navigating the web site according to the corresponding visiting (usage) pattern.

Hyperlink Recommender is responsible for assignment of the user to the appropriate usage (session cluster) and content pattern (term cluster). To achieve this goal, it makes use of knowledge maintained by *Content Miner* – content patterns, *Publisher Usage Miner* – usage patterns, and *User Session Monitor* – current user behavior. Having suitable patterns assigned, *Hyperlink Recommender* extracts from them the publisher's pages with the highest score. Such pages are relayed to the *User Assistant* as hyperlink recommendations. This process is carried out at each user request.

Ad Recommender performs the main online personalization process of advertising. It gains patterns (vectors) assigned to the current user by *Hyperlink Recommender*. Next, it gets the corresponding to them advertising patterns: the term and clicking ones, delivered by *Advertiser Content Miner* and *Ad Usage Miner*, respectively. *Ad Recommender* integrates obtained patterns and creates the advertising ranking. This ranking list is compared against the already exposed advertisements provided by *Ad Usage Monitor*. As a result, we accomplish the most advisable advertisements, which are returned to *User Assistant* and *Ad Usage Monitor*.

User Assistant dynamically incorporates hyperlinks and banners delivered by *Hyperlink* and *Ad Recommender*, respectively, into the HTML content of the page returned to the user.

Update Manager is responsible for monitoring the publisher's web site, the validity of campaigns and other factors of changes in source data. Its main task is to recognize whether changes are serious enough to start the offline update processes.

Ad DBase provides database connectivity. It performs operations requested by other agents on the database.

4 Knowledge Maintenance

Changes in data sources are the serious problem in almost all web environments. It makes the knowledge derived from various initial data to be periodically updated, so that the hyperlink and advertisement recommendations are reliable. There are many factors of changes with different intensity and importance onto the quality of recom-

mendation, such as new advertisements, sets of new user sessions, new and updated publisher's pages, etc. (Fig. 3). There is a special agent – *Update Manager*, which task is: to monitor these factors; to decide, whether and what the update process should be carried out, and to synchronize the process. Note that the update is usually the very resource consumable and long-lasting process. The synchronization among all involved agents is necessarily because the system works continuously and the switch to the new knowledge should be done "at one moment". For example, a big number of new user sessions makes all usage and clicking patterns to be recalculated. This triggers four agents: *User Session Manager* and *Miner*, *Ad Usage Manager* and *Miner*. For that reason, *Update Manager* has to know all dependencies between the knowledge used by particular agents.

Update Manager can initiate the update of only (Fig. 3): clicking patterns (2), both usage and clicking patterns (1+2), advertisers' thematic groups (3), both publisher's and advertisers' thematic groups (4+3), or the whole system knowledge (1+2+3+4). Additionally, some changes are much or less easy to introduce, e.g. the insertion of new advertisements results only in processing of the new advertiser's web site (performed by *Web Crawler*) and the relatively simple extension of appropriate vectors (2+3) related to *Ad Usage Manager* and *Miner* as well as *Advertiser Content Manager* and *Miner*. In opposite, the large enough modification of content of publisher's site forces to rebuild all both publisher's and advertisers' thematic groups (3+4).

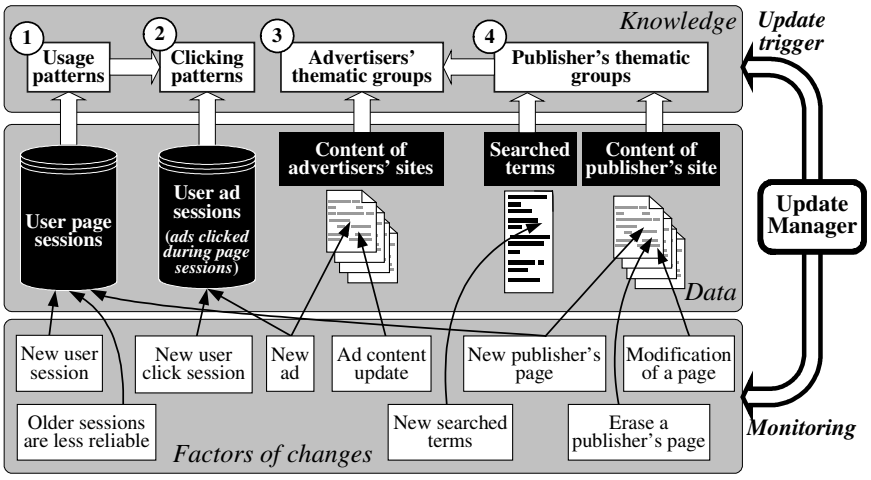

Fig. 3. Information flow and some factors of changes in personalized advertising

5 Conclusions and Future Work

The proposed system "personalizes" the online advertising using web usage and content mining. Additionally, it is able to recommend hyperlinks in the similar, adaptive way. Due to its multi-agent architecture, it is more flexible, scalable and open for new functions such as the sophisticated scheduling mechanisms and extension of personal-

ization with knowledge derived from purchase history as well as product or page ratings gathered in e-commerce [15]. Furthermore, the multi-agent platform makes easier the introduction of diverse update processes, what is very important in the ever changing web environment. The system as the prototype was put into practice for the copy of www.poland.com web site.

References

1. 24/7 Real Media, Inc., http://www.realmedia.com/
2. Amiri A., Menon S.: Efficient Scheduling of Internet Banner Advertisements. ACM Transactions on Internet Technology **3** (4) (2003) 334–346.
3. Bae S.M., Park S.C., Ha S.H.: Fuzzy Web Ad Selector Based on Web Usage Mining. IEEE Intelligent Systems **18** (6) (2003) 62-69.
4. Bilchev G., Marston D.: Personalised Advertising – Exploiting the Distributed User Profile. BT Technology Journal **21** (1) (2003) 84-90.
5. Chickering D.M., Heckerman D.: Targeted Advertising with Inventory Management. 2nd ACM Conference on Electronic Commerce, EC00 (2000) 145-149; Interfaces **33** (5) (2003) 71-77.
6. Ferber J.: Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence. Addison Wesley Longman, Boston (1999).
7. Google Advertising Programs, http://www.google.com/ads/
8. Iyer G., Soberman D., Villas-Boas J.M.: The Targeting of Advertising. Marketing Science, (2005) to appear.
9. Kazienko P.: Multi-agent Web Recommendation Method Based on Indirect Association Rules. 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems, KES'2004, **LNAI 3214**, Springer Verlag (2004) 1157-1164.
10. Kazienko P.: Product Recommendation in E-Commerce Using Direct and Indirect Confidence for Historical User Sessions. 7th International Conference on Discovery Science, DS'04, **LNAI 3245**, Springer Verlag (2004) 255-269.
11. Kazienko P., Adamski M.: Personalized Web Advertising Method. Adaptive Hypermedia AH 2004, **LNCS 3137**, Springer Verlag (2004) 146-155.
12. Kazienko P., Kiewra M.: Integration of Relational Databases and Web Site Content for Product and Page Recommendation. 8th International Database Engineering & Applications Symposium, IDEAS '04, IEEE Computer Society (2004) 111-116.
13. Kazienko P., Kiewra M.: Personalized Recommendation of Web Pages. Chapter 10, in: Nguyen T. (ed.), Intelligent Technologies for Inconsistent Knowledge Processing, Advanced Knowledge International, Adelaide, South Australia (2004) 163-183.
14. Kazienko P., Kiewra M.: ROSA - Multi-agent System for Web Services Personalization. First Atlantic Web Intelligence Conference Proceedings, AWIC'03. **LNAI 2663**, Springer Verlag (2003) 297-306.
15. Kazienko P., Kołodziejski P.: WindOwls – Adaptive System for the Integration of Recommendation Methods in E-commerce. Third Atlantic Web Intelligence Conference Proceedings, AWIC'05, **LNAI 3528**, Springer Verlag (2005) 218-224.
16. Langheinrich M., Nakamura A., Abe N., Kamba T., Koseki Y.: Unintrusive Customization Techniques for Web Advertising. Computer Networks **31** (11-16) (1999) 1259-1272.
17. Online Advertising. DoubleClick Inc. (2004) http://www.doubleclick.com/us/products/online_advertising/
18. Tomlin J.: An Entropy Approach to Unintrusive Targeted Advertising on the Web. Computer Networks **33** (1-6) (2000) 767-774.