

Collecting User Preferences from User Activities on E-Commerce Web-Site Focusing on Sending Targeted Advertisements

Parisa Nasirinejad*, Ali Reza Honarvar

Department of Computer Engineering and Information Technology, Islamic Azad University, Safashahr Branch, Safashahr, Iran.

* Corresponding author. Tel.: +989171327076; email: pari.nasirinejad@gmail.com

Manuscript submitted September 12, 2017; accepted November 5, 2017.

doi: 10.17706/jcp.13.7.830-838

Abstract: Internet is growing as a global marketplace. For the foreseeable future, online advertising remains as the main source of revenue on the web. Taking advantage of Web mining techniques is useful in this way that it is possible to replace personalized ads by public advertisement by examining the user profile and extracting his interests. Since the data, associated with user performance, have a large volume and are generated in high speed and are very versatile, they can be called as big data. As it will be explained below, the processing of big data seems impossible using traditional technology. Using relational database in order to process huge data is very difficult; hence, the use of a non-relational database such as MongoDB can facilitate the processing of data in addition to resolve the mentioned issues. A major problem in the process of understanding the user's favorites and sending his favorite target advertising is to collect information that will help researchers in this field; In order to collect this information, the website designed for an electronics store database with MongoDB database has been redesigned in which the required factors, including the user login time to the site, number of visits per page, the number of rotating the mouse cursor over the product image etc. have been considered to understand the user interests, and thereby, their interest has been identified; then it has been used in a 6-month period by the customers of the store. After collecting and analyzing the data, it was observed that different users have different performance in site and suggestions have been provided to improve site performance, customer satisfaction and online store sales based on their tastes.

Key words: e-Commerce, targeted advertising, user performance, recommender systems.

1. Introduction

In an age of disappearing borders and uniting communities, targeting current population of the content available on the web does not seem enough. The market is made up of human beings; so it is better to pay attention to the individual behavior when you personalize the site, rather than paying attention to geographical location or demographic characteristics (gender and age). Traditional advertising are the same for all the customers and cannot meet the current business needs. To increase effectiveness, the right person should receive the right message at the right time and in the right context [1]. Targeted advertising are advantageous for advertiser and user: the advertiser can achieve the higher profit by sending advertisement to the users who have a greater potential for purchasing, and users can get more relevant and useful

advertising more like it to their tastes and interests.

The traditional approach of targeted advertising is to analyze the old database obtained from previous transactions and characteristics associated with customers by some statistical tools to identify a list of customers who likely respond to advertising [2]. With the advent of new technologies, the automated tools are used to identify potential customers [3]. Hence, the idea of most of the recommendation system is to advertise the products according to user preferences that can be achieved either by ranking published by users and implicitly, using the history of previous transaction or cookies over the past few years.

1.1. Purposes

The aim of this study is to identify and collect user performance history on e-commerce sites with a focus on sending targeted advertising to the customer. The user's performance history can be used in research on data mining as well as big data. These data identify user's interests so that related sites can show the users' favorite ads.

2. Basic Concepts

2.1. Advertising

Today, clients are the core principle of any business and to attract and keep them satisfied of goods and services would be increasingly difficult. The purpose of advertising is to create awareness and inform the public about advertised products which helps consumers in conscious decision-making [4]. The importance of advertising as a marketing strategy depends on its ability to influence the customers. However, this effect should be such that customers not only attempt to purchase, but also continue to repurchase and their brand loyalty will be increased.

2.2. Electronic Commerce

The relationship between commerce and technology has longley been existed and continued. Multiple developments and advances in information and communication technology (ICT) in recent decades, led many developments in many fields such as global business. As a result, the processes involved in areas such as trade, economy, banking, customs, etc. have evolved and changed [5]. Over time, traditional trade will not be able to respond to modern needs; hence, modern business practices are required.

2.3. Targeted Advertising

The targeted advertising, from a business perspective, is a marketing concept for addressing a specific target group. The targeted advertising is under the umbrella of personalized strategy in Electronic Customer Relationship Management (ECRM). The main idea of individual marketing is to recognize the individual preferences and unique services to customers based on personal recommendations [6]-[8]. Targeted ads are dedicated for specific target groups around the web and big users like Google thanks their enormous growth to the profit earned by the advertising services based on complex targeting technologies. However, the targeted advertising potential has exceeded the web and has covered the total ICT (information and communications technology) space including IPTV (Internet Protocol television), Mobile and converged environments [9]. So far, the most profitable location for online advertising has been the 'search' area, and the most search advertising effectiveness is obtained by adwords model and matching the search term with ads. The second appeal of online advertising is to select product in order to advertise on an online store [10].

In process view, the activities necessary to realize the personalization strategies can be divided into the following steps [11]:

- Tracking

- Profiling
- Matching

Targeted advertising has different types and methods. Targeted advertising types are given below:

- Content and Contextual Targeting
- Technical Targeting
- Time Targeting
- Socio-demographic Targeting
- Geographical and Location-Based Targeting
- Behavioral Targeting
- Predictive Behavioral Targeting

2.4. Big Data

The definition of big data can be examined from three aspects of high volume, high production rate and a large variety. Organizers believe that the traditional technologies are not able to process the flood of information generated by ICT. Big data is a concept that deals with the following features [12]:

- The data which processing is not impossible using traditional technologies.
- The data which are provided by ICT.

2.5. NoSQL Database

NoSQL databases are as non-relational and distributed, horizontally scalable, and schema-free. This type of database data is classified based on their data model [13]. RDBMS (Relational Database Management Systems) or relational databases have been used since 1970. However, the problem of storage in web-based systems suffers from the relational databases limitations and researchers and companies have to explore non-traditional forms of data storage methods [14]. NoSQL databases are provided as an alternative to relational databases. The so-called NoSQL has been firstly introduced by Carlo Strozziin to refer to open source NoSQL database [14].

The advantages of non-relational databases are [15]:

- Elastic scaling
- Greater data portability
- Cheaper NoSQL storage servers
- Cheaper Server
- The absence of schematic or fixed data model
- Integrated Caching Facility

3. Background Research

In a study conducted by Farahat, it is stated that advertising need an accurate estimate of the impact of targeted advertising. Since online advertising is growing with the increasing rate, the rate of clicks on online advertising has decreased to less than 1 percent from 3 percent. In order to improve the effectiveness, advertisers and content providers have increasingly turned to target advertising or promotional methods that use web content, user location, search history, user profile, or any other information available. Targeted ads often promise to improve performance, not only through the provision of advertising to the target sectors, but with increased performance criteria through the increased click rate and changes in sales [16]. The study results suggest that the actual impact of targeted advertising based on brand about 79 percent. While providing targeted advertising based on product model will affect 11 percent of the increase the visit [16].

Bhumgara in [17], 1) has increased the effectiveness of targeted advertising by proposing a detailed theoretical framework and econometric methods and, 2) Taking advantage of a natural experiment on a large scale and sending ads to targeted and untargeted users, has measured the ratio of the number of user Click through Rate (CTR). 3. Then, provided an experimental model of CTRs in order to explain the changes applying targeted advertising.

According to an article written by Deshpande *et al.* over the past two decades, rapid advances in the fields of information and communication technology (ICT) has led to increase public access and use of the Internet. So, the Internet provides an opportunity for advertisers to deliver their ads to an audience based on different criteria. Since, the ads can be presented directly and without interaction with the customer, it can be said that advertising plays an important role in business. Web-based advertising has changed their focus from advertising in general and for all, to the targeted advertising [18].

In Athanasiou *et al.* [19], according to investigating Pishgaman Internet Advertising Network, secured targeted advertising, on average has 2.7 times more profit compared to non-targeted ads and also has an annual growth rate of 40%. In addition, targeted advertising has twice likelihood to convert users to the buyer than non-targeted ads. In this paper, a framework is proposed that seriously considers privacy issues while advertising targeting. To test this, using a real-time communication application such as online chat, the proposed framework is applied to the server. On one hand, server is responsible for creation and maintaining the conversation among the users, and on the other hand, is responsible for the interaction with an advertising server in order to provide targeted advertising based on the content and the time of exchanging messages to the final users [19]. He also said that the real-time communication system includes three components: User of online chat application, online chat servers and ad servers. In fact, advertisement is actually as a repository and database server-based to store and organize all ads by specific structural and method. Each ad is labeled by the similar keywords in an online conversation be labeled [19]. This paper presents an innovative solution to deal with the contradiction between targeted advertising and privacy. Online advertising have major benefits compared to traditional advertising that made them attracting to the advertisers over the past years and have increased its relationship with applications that have billions of visitors per day, such as Google, Facebook, Twitter and YouTube [19].

Bhatia *et al.*, in their research have suggested that the explosive growth of social networks has led to the availability of customers' tastes and preferences. With this data, we can provide better service to customers, and provide ads for them, which lead more customer satisfaction. They have suggested that one way to deliver ads relevant to the customers, is to take into account the location of customers. In this study, a hybrid model of social and prolific data is used to deliver targeted advertising to customers. Social data have been collected from users' Facebook profile and their location [20].

4. Methodology

4.1. Implementation

Identifying the user's personal taste and offer tailored products and promotions, needs to collect information on user performance on various websites and social networks. For this purpose, the website redesigned an electronics store that was used for years by many customers and considered the factors needed to identify users' interest in it.

Information that can identify the user based on it has a high volume, and as mentioned, this type of data is called Big Data. Management, control and processing of the data are beyond the capabilities of software tools at an expected time. Relational database management systems often face problems related to the manipulation of big data. While it simultaneously needs tens, hundreds or even thousands of servers, that in this case, the complexity of this growth is unexplainable, and the stability and speed of the relational

databases strictly comes down and shows the weakness very large for the distributed systems. For such services, the solution is to use non-relational and distributed NoSQL databases and has the possibility to be horizontally developed [21]. So, MongoDB database was used in the process of redesigning the website.

4.2. Data Set

The target population includes all elements and people who are in the specific geographical scales (global, regional, local or location) with one or several common traits [22]. The population of this research is the former clients of the electronic stores who have years of experience buying from the store. This web site was re-implemented in order to collect data and run targeted ads using MongoDB non-relational database.

4.3. Sampling Method

Data collection methods for the study fall into two major category of library (tools such as books, specialty papers and digital texts) and non-library (tools such as websites implementation, and categorizing obtained data and information).

4.4. Factors Used in the Implementation Process

In the implementation process and to record the users' actions on the site, many factors were collected and examined, among them some of the most important factors are listed below:

- The user's Time of logging to the site
- The number of visits per page
- The number of the mouse cursor rotation over the product image
- The characteristics of the product the user is interested in
- The time of purchase
- The rating given by the user to the product
- User rating to messages sent by others

5. Result

5.1. Descriptive Statistics

In general, two mobile and clothing markets have been considered for the online store. Mobile shop products are divided two sub-sets of mobile phones (32%) and tablets (68%) and mobile phones have the brands of LG, Samsung, and Sony (see Fig. 1).

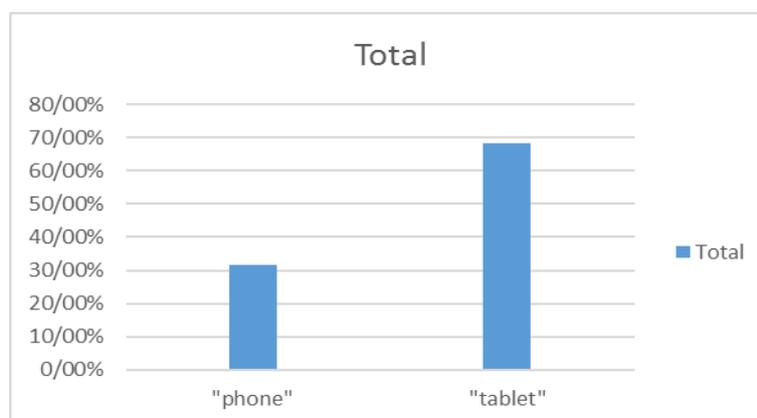


Fig. 1. Mobile store products.

Clothing store products are divided into two groups of men's (43%) and women's (57%) products. Sub-categories of each category include products such as bags (women 10%, men: 8%) Shoes (women 23%,

men 16%), watch (women: 4%, men: 7%), clothing (women: 7%, men: 6%) and glasses (women 12%, men: 6%). However, for some products such as shoes and clothing, two types of formal and informal has also been considered (Fig. 2).

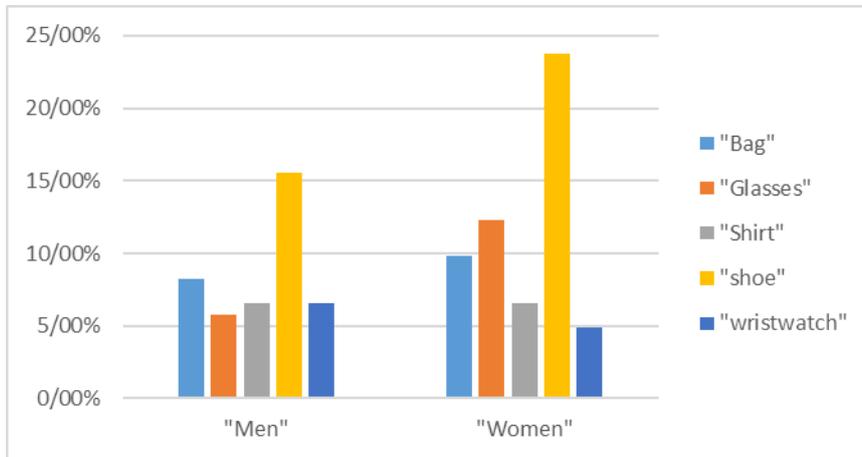


Fig. 2 .Clothing store products.

5.2. Findings

Table 1. Shopping Card Status

Total products added to the cart	100%
purchased Goods	86%
Goods removed before buying	7%
The remaining products in the shopping cart	5%

Table 2. The Time of Purchase

a.m.	17%
p.m.	83%

Table 3. Women and Men Time of Purchase

	Female	Men
a.m.	9%	9%
p.m.	54%	28%

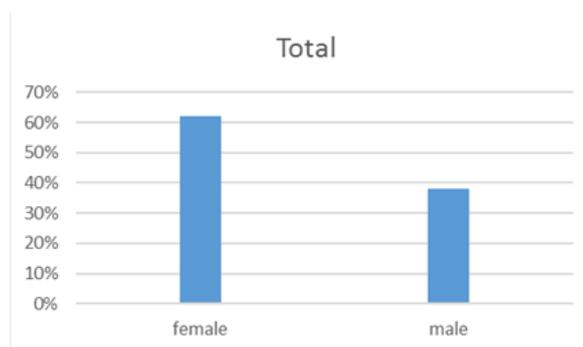


Fig. 3 .Categorizing purchase by gender.

Based on Fig. 3, from the total purchases made, 62% is made by women and 38% of purchases is made by men. 70% of all purchases are related to clothing stores and 30% are related to the mobile store.

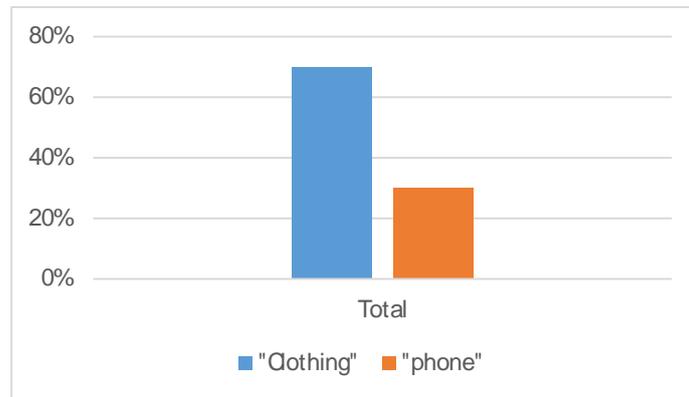


Fig. 4. Categorizing the purchase by product type.

5.3. Discussion

According to the purchased product, fortunately, a large percentage of products added to the cart have led to buy and this suggests that the provided products are greatly according to the customers' interests (Table 1). Although the number of products, which has been removed before the purchase, was small but, it is necessary to adopt measures to minimize the number.

Of total purchases, 17% is in the morning and 83% is in evening (Table 2). According to this information, it is better to add new products in the evening shift so that users see it at the right time and purchase it if it is necessary. Women often do their purchases in the evening and men make their final purchases in the morning. So, it is better to add men's favorite products, mostly in the morning (Table 3).

Most purchases have been made by women and from the clothing store, and the most products purchased were in the category of women and shoes. So, it is better to invest more on female buyers, and also the products related to them. In addition, it is necessary to add the variety of shoes available on the site, in terms of type and color. Since, women are more fashion-oriented, we can concentrate on the categories of products as bags and shoes fit together or focus more on bags diversity.

Among the female products, shoes have the most purchases by women and men. Among the male products, shoes have the maximum purchase by women and bags have the largest purchase by men. So, in case of male products, in addition to add the category of a well-read men like set handbags and shoes, we can add a category of products such as belts so that men and women have more desire to select them among the others and in fewer time when buying.

Tablets are more bought by women and mobile phones are more bought by men. Among the provided mobile brands, Samsung is the most popular brand. So, by adding a sub-store as home appliances and invest on this brand, we can increase the visits and sales and on the other hand, further consider the customer's satisfaction. It is also better to add the designed mobile by this brand. Generally, about the Apple and Samsung brands, it is suggested that their newest models with a variety of colors (especially tablet of Apple brand and Mobile of Samsung brand) should be provided to users with more detailed and accurate information about them.

Since the average price of goods purchased in the mobile category is \$ 647 for women and \$ 660 for men, and also, \$ 232 for women and \$ 279 for men in the categories of clothing, it can be concluded that Quality is more important for men and women often pay attention to the price. So, it is better to try suggesting

products with more reasonable price to women; also, majority of users has shown interest in products with fair price; so, in addition that low price is important, quality is also important for the users.

Since the mouse movement by men on the mobile products is more than women, and since men have more mobile purchase than women, it can be said that to mouse movement data can provide reliable results. As a result, this claim is confirmed that more women are interested in clothing products than men and men are more interested in mobile stores products. Women have paid more attention to the women clothes and men have shown more attention to mobile phones. Also, women's shoes and men bags have been considered more than the other products.

6. Conclusion and Future Works

In this paper, an e-commerce website that has been available for years to customers was redesigned through the MongoDB non-relational database. In redesigning the website, many criteria including the user login to the site, number of visits per page, the number of rotating the mouse cursor over the product image, product specifications that the user is interested in, the time of purchase, the rating of product by the customer, and user ratings to messages sent by others have been used in order to identify the user's interest. Then, the designed electronics store was again made available to former clients. After a 6-month period, data were analyzed from users. As a result of this analysis, the products available in the store were classified and the interest of each user to different products were detected and identified. Based on interest derived from each user, products have been displayed as the proposed products based on his/her interests and advertising were provided for him/her that it was found that, the users' interest in that product is more than the other products. The result of a 6-month period comparison prior to the implementation of targeted advertising and period of 6 months after the implementation of targeted advertising, showed that before targeting ads, the large amount of advertising caused users' confuses and also considering the possibility of staying away the user's favorite products, electronic store sales has about 15% growth after targeted advertising and displaying products.

In future works we extend and adapt our previous works on ethical reasoning in artificial agents [23], [24] to develop an artificial ethical agent to recommend ethical advertisements to customers on the web. Considering the emotions of users on the web and extract the emotions and combing it with the personalized web usage profile of users is to be another concern that will be conducted in future works.

References

- [1] Adams, R. (2004). Intelligent advertising. *AI & Society*, 68-81.
- [2] Yang, W. S., Dia, J. B., Cheng, H. C., & Lin, H. T. (2006). Mining social networks for targeted advertising. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences, 2006. HICSS'06*.
- [3] Hays, L. (1994). Technology: Using computers to divine who might buy a gas grill. *Wall Street Journal*.
- [4] Tabatabainasab, m., & parish. (2015). *Review of Advertise with Young Consumers*.
- [5] Choshin, M., & Ghaffari, A. (2017). An investigation of the impact of effective factors on the success of e-commerce in small-and medium-sized companies. *Computers in Human Behavior*, 66, 67-74.
- [6] Albers, S., Clement, M., Peters, K., & Skiera, B. (2001). Marketing mit interaktiven medien — Strategien zum markterfolg, 3. vollständig überarbeitete auflage. *Frankfurt/M*.
- [7] Fritz, W. (2013). *Internet-Marketing und Electronic Commerce: Grundlagen—Rahmenbedingungen—Instrumente*. Verlag: Springer.
- [8] Negelmann, B. (2008). Trendthema behavioral targeting. *Internet World Business*. Retrieved from <http://www.internetworld.de/Trendthema-Behavioral-Targeting>
- [9] Schlee, C. (2013). Targeted advertising technologies in the ICT space: A use case driven analysis. *Springer Science & Business Media*.

- [10] Leskovec, J., Rajaraman, A., & Ullman, J. D. (2013). *Mining of Massive Datasets*. Cambridge University Press.
- [11] Hiemstra, K., Bscheid, W., & Groth, A. (2013). *Marketing-Kommunikation im Internet: Theorie, Methoden und Praxisbeispiele vom One-to-One Bis zum Viral-Marketing*. Verlag: Springer.
- [12] Kinoshita, E., & Mizuno, T. What is big data, in big data management. *Springer*, 91-101.
- [13] Ahmadian, M., Plochan, F., Roessler, Z., & Marinescu, D. C. (2017). SecureNoSQL: An approach for secure search of encrypted nosql databases in the public cloud. *International Journal of Information Management*, 37, 63-74.
- [14] Corbellini, A., Mateos, C., Zunino, A., Godoy, D., & Schiaffino, S. (2017). Persisting big-data: The NoSQL landscape. *Information Systems*, 63, 1-23.
- [15] Han, J., Haihong, E., Le, G., & Du, J. (2011). Survey on NoSQL database. *Proceedings of 2011 6th International Conference on Pervasive Computing and Applications (ICPCA)*.
- [16] Farahat, A. (2015). How effective is targeted advertising? *American Control Conference*.
- [17] Bhumgara, A. (2015). Online mobile targeted advertising. *Proceedings of IEEE International Conference on Computational Intelligence and Computing Research (ICICIC)*.
- [18] Deshpande, N., Ahmed, S., & Khode, A. (2014). Web based targeted advertising: A study based on patent information. *Procedia Economics and Finance*, 11, 522-535.
- [19] Athanasiou, A. P., Raftopoulos, C. N., Thanos, E. D., Kritharellis, G. V., Tselikas, N. D., Foukarakis, I. E., & Boucouvalas, A. C. (2012). Towards privacy-aware target advertising. *Proceedings of 2012 16th Panhellenic Conference on Informatics (PCI)*.
- [20] Bhatia, V., & Hasija, V. (2016). Targeted advertising using behavioural data and social data mining. *Proceedings of 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*.
- [21] Purmohamadi, F., Tavali, Z., & Hoseini, R. (2014). *NoSQL Databases and Comparison with Relational Database*.
- [22] Hafezniya, M. (2013). *Introduction to Research Methods in Social Sciences (Fundamental Revision with Additions)*.
- [23] Honarvar, A. R., & Ghasem-Aghaee, N. (2009). An artificial neural network approach for creating an ethical artificial agent. *Proceedings of 2009 IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA)*.
- [24] Honarvar, A. R., & Ghasem-Aghaee, N. (2010). Towards an ethical sales-agent in e-commerce. *Proceedings of International Conference on e-Education, e-Business, e-Management, and e-Learning, IC4E'10*.



Parisa Nasirinejad holds a bachelor of information technology in Jahrom University and master of e-commerce in Islamic Azad University of Safashahr. Her research areas include web developing, data mining, ecommerce and customer relationship management.



Ali Reza Honarvar is a professor of software engineering at Department of Computer Engineering and Information Technology of Islamic Azad University of Iran since September 2010. He has worked on a range of topics including data mining, big data, multi-agent systems, distributed systems, parallel processing, smart city, internet of things, and ecommerce.