THE ROLE OF DATA MINING IN REVENUE OF WEBSITES

MOHAMMAD HAJARIAN

M.SC. SOFTWARE ENGINEERING
STAFFORDSHIRE UNIVERSITY

ABSTRACT

One of the most used techniques for website to make revenue is advertisement. The success rate of an advertisement is calculated based on the CTR factor. There are different techniques to improve the CTR of an advertisement in websites. Showing relevant advertisement to a visitor of a website can lead to improve the CTR of an advertisement and the revenue of the website. In this paper author has reviewed different technique of showing relevant advertisement to a website visitors and the role of the data mining to perform this.

KEYWORDS: Data mining, CTR, advertisement, website, Association rule, Apriori algorithm.

Introduction

In the era of web 2 and social networks, there are lots of data available about users and their activities on websites. On the other hand one of the most popular ways that websites can make revenue is advertisement. Online advertisement revenue was calculated to be $20.1 in the first half of 2013 (IAB, 2013) .showing relevant advertisement to the website users can improve this income and let them to maximize their revenue. The objectives of this paper is to show how data mining can be used to increase the revenue of the websites by showing relevant advertisements to the users. And which data mining technique is most suitable to be used to perform this action. In this article author shows how this data can be used to be processed using data mining to improve the revenue of the websites.

Calculating the revenue of online advertisements

There are different techniques for online advertisements, some websites charge their advertisers monthly and they just receive fixed amount of money on the other hand some website does not ask for money before putting the advertisements in their website however they ask advertisers pay them only when their users click on their advertisements. This method is called Cost per click (CPC). CPC is one of the most used methods to calculate the income of website from advertisements(davechaffey , 2009).CPC means that based on each click that users made on the websites, they can charge their advertisers.To calculate the success rate of CPC, websites use the click through rate formula (CTR). Table 1 illustrates the CTR formula:

$$CTR = \frac{\text{Clicks} * 100\%}{\text{Impressions}}$$

Table 1: CTR formula

As table 1 illustrates, for example if a website has 10000 visitors and 100 click is received from 10000 visitors the CTR will be 1. If all the 1000 visitors click on the advertisements, then the CTR will be 100 and if no one click on it the CTR will be 0.Although CTR of each website will be different but according to the google the average CTR of websites are 2% (Average Click-Through Rate,2015). Because CTR shows the success rate of an online advertisement, increasing it will help the website to improve their revenue.

Improving CTRby showing relevant advertisements

Improving CTR basically means that increasing the clicks that users make on website advertisements. Although by default when websites put their ads on the website all the users can see that ads, If the website be able to limit the displaying of advertisement to those who might be more interested on that kind of advertisement then the CTR will be increased.

Finding people who might be more interested on specific type of advertisement can be done by different techniques, such as contextual advertising which is based on the keywords that extracts form the webpage (Yih et al.,2006). This system has been used in google AdSense to show relevant advertisement to the users based on the page content that user is in to it. In addition to the google AdSense other websites such as yahoo news uses contextual advertising system to promote better news to their visitors based on the gathering the content of the webpages it view (Irmak et al., 2009). The problem of this technique is that sometimesthe keyword might a part of a sentence that says bad things about a product and in that time advertising that product might not be a good idea. However this techniques is better when websites does not have any idea that who is visiting their websites.

Another technique to show relevant advertisementto the users is word of mouth advertisements. This type of advertisement is done in social network using influential people (Li et all.2011). Asking from visitors of the website to invite their friends to the website using their email address is kind of WOM advertisement however the problem of this technique is that this does not lead websites exactly to show relevant ads but it only encourage visitors to click on them.

Finally the last but not least technique that can be used to show relevant advertisement is data mining. Data mining can be used only when enough data is available about users. To show relevant advertisement using data mining relevant features of users should be gathered by the website (Kardan and Hooman, 2013). Although data mining include different type and techniques however for advertising in websites based on the article of Gibert et al. (2010)association model of data mining is the most suitable data mining model to find out the relationship between variables.

Role of data mining

Using data mining it would be possible to understand what type of people are more interested in which type of advertisements. This can be done using association rules. By implementing association rules of data mining on different variable of the users and finding out the relationship betweenthis variables and type of the advertisement they interested on,researcher will be able to extract a rule list that discovers the strongestrelationship between variables.

Apriori algorithm is one of the well-known data mining approaches to find relationship between variable and it can used to find the strong relationship between variables in database (Palczewska, 2012).the Apriori algorithm has been described in table2:

Apriori (T,E)

$L_1 \leftarrow \{large\ 1 - itemsets\ \}$

K← 2

While $L_{k-1} \neq \varphi$

$C_k \leftarrow \{a \cup \{b\} | a \in L_{k-1} \wedge b \in \cup\ L_{k-1}\ \wedge b \notin a\}$

For transacting t ∈ T

$C_t \leftarrow \{C|C \in C_k \wedge C \subseteq t\}$

For candidate c $\in C_t$

Count(c) ← count (c) + 1

$L_k \leftarrow \{C|C \in C_k \wedge count(\ c\ ) \geq E\}$

K←K+1

Return $\cup_K L_K$

Table 2: APRIORI ALGORITHM ADAPTED FROM (Agrawal, 1994)

As it can be seen from the table2 association rule can find interesting relationship between variables and it can be used to match the features of website users such as age to the advertisement type they might be interested in. such implementation has been done in Davidson et al. research (2010), in the recommendation video system for YouTube , they have collected user information and matched it to the length of the video and the likes that a video has been received and find the strong relationships between variables as the rules. Using this rules Davidson et al were able to show more relevant videos to the YouTube users. This method however will be used to show the relevant videos only to the one user that its data has been collected and it cannot be used for new users.

Variable selection

Variable selection can be different based on the information that the website gather from their users. For example they can gather their users' age and the ad type that the users are interested in and then by implementing data mining on all the gathered data websites can find out the strong relationships that exist between the users' age and ad types, using this relationship they will be able to show more relevant advertisement to the users.

Evaluation

Contextual advertisement is more likely to be used on the website that they do not have any idea who is visiting their website , hence based on the content of the page that users visit they can recommend advertisement. WOM advertisement would not be suitable when there are thousands of different advertisement in a website as inflectional people has to recommend it .Although there are different techniques to show relevant advertisement to the users to improve the CTR, Data mining can be one of the best approaches for the websites that have collected some information about their members.For example most of the users who are in the age of under 15 would not be interested in real state advertisement, and this type of the data can be gathered and analyzed , Hence when a new member comes to the website , he/she will not bothered with the advertisement that are not interested in.

Conclusion

In this paper author has reviewed and critically evaluate the different approaches to show more relevant advertisement to the users to improve the CTR of websites and eventually improve the revenue of the websites. Additionally the role of the data mining and its advantage to the other approaches was reviewed. Besides it was showed how by selecting suitable variables and implementing association rules of data mining using Apriorialgorithm , relevant advertisement can be delivered to the users of the website, even the new members.

REFERENCE

Agrawal, R. &Srikant, R. (1994) Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB.), vol. 1215, pp. 487-499.

Average Click-Through Rate (CTR): Learn How Your Average CTR Compares | Wordstream. 2015. Average Click-Through Rate (CTR): Learn How Your Average CTR Compares | Wordstream. [ONLINE] Available at: http://www.wordstream.com/average-ctr. [Accessed 04 June 2015].

Dave Chaffey, Fiona Ellis-Chadwick , Kevin Johnston, Richard Mayer (2009). Internet Marketing: Strategy, Implementation and Practice. 4th ed. United States: Prentice Hall. 736 .

Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., Gupta, S., He, Y., Lambert, M. & Livingston, B. (2010) The YouTube video recommendation system. In Proceedings of the fourth ACM conference on Recommender systems.) ACM, pp. 293-296.

Gibert, K., Sanchez-Marre, M. &Codina, V. (2010) Choosing the right data mining technique: classification of methods and intelligent recommendation. In The proceedings of the 2010 international congress on, environmental modelling and software.), pp. 1933-1940.

IAB. (2013). Internet Ad Revenues At $20.1 Billion Hit Historic High For Half-Year 2013, Up 18% Over Same Time In 2012, According to IAB. Available: http://www.iab.net/about_the_iab/recent_press_releases/press_release_archive/press_release/pr-100913. Last accessed 9 Dec 2013.

Kardan, A. A. &Hooman, M. (2013) Targeted advertisement in social networks using recommender systems. In e-Commerce in Developing Countries: With Focus on e-Security (ECDC), 2013 7th Intenational Conference on.) IEEE, pp. 1-13.

Li, F.-H., Li, C.-T. & Shan, M.-K. (2011) Labeled Influence Maximization in Social Networks for Target Marketing. In Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom).) IEEE, pp. 560-563.

Palczewska, A. (2012) Association rule learning.

Yih, W.-T., Goodman, J. &Carvalho, V. R. (2006) Finding advertising keywords on web pages. In Proceedings of the 15th international conference on World Wide Web.) ACM, pp. 213-222.