Customer Segmentation for Efficient E-Commerce Advertising Campaigns

Dalia Kriksciuniene, Virgilijus Sakalauskas
Vilnius University

Abstract

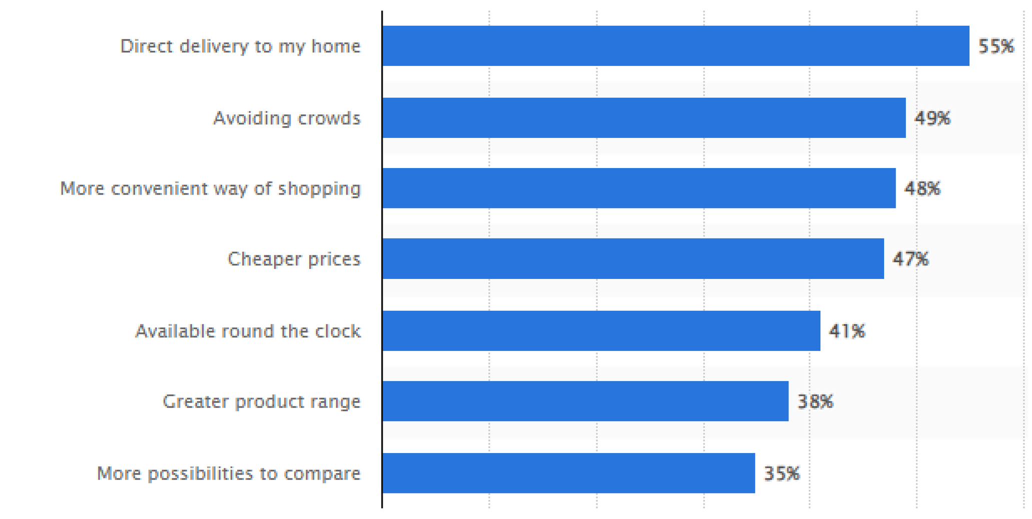
The continuous rise of e-commerce has encouraged significant interest among researchers in comprehending online shopping behavior, consumer interest trends, and the effectiveness of advertising strategies. This paper presents a fresh approach for targeting high-value e-shop clients by utilizing clickstream data. We propose the new algorithm to measure customer engagement and recognizing high-value customers. Clickstream data is employed in the algorithm to compute a Customer Merit (CM) index that measures the customer's level of engagement and anticipates their purchase intent. The CM index is evaluated dynamically by the new algorithm, considering the customer's activity level, efficiency in selecting items, and time spent in browsing. The proposed method was tested on actual clickstream data from two e-commerce websites and showed that the personalized advertising campaign outperformed the non-personalized campaign in terms of click-through and conversion rate.

Introduction

E-commerce has become a popular activity among internet users for searching and buying goods, with an increase in sales volume and visitors.

According to research completed by eMarketer and Statista, online retail e-commerce sales in 2022 exceeded 5.7 trillion U.S. dollars worldwide and will reach $6.51 trillion by 2023, with e-commerce websites taking up 22.3% of total retail sales.

It is very interesting to know what drives such big e-commerce growth. In 2022 the survey "Drivers of online purchases" was conducted among 9,989 respondents (Statista © 2023).



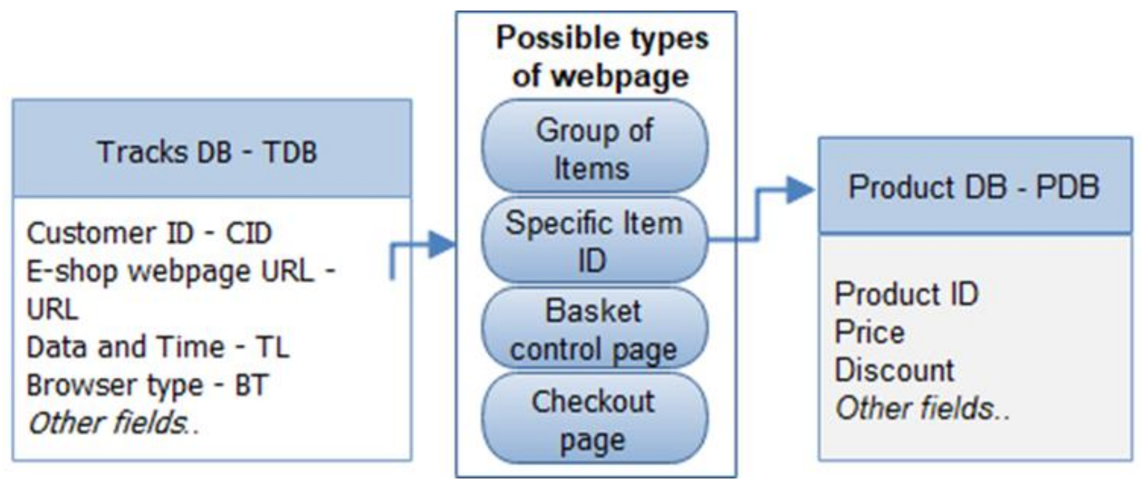
Customer segmentation using clickstream data

This article presents a comprehensive approach to targeting advertising scenarios for e-commerce websites by utilizing clickstream data analysis and calculating the accumulated customer activeness indicator. Clickstream data provides a detailed record of a visitor's online journey, including the sequence of pages visited, visit durations, timestamps, search terms, ISPs, countries, browsers, and computer specifications.

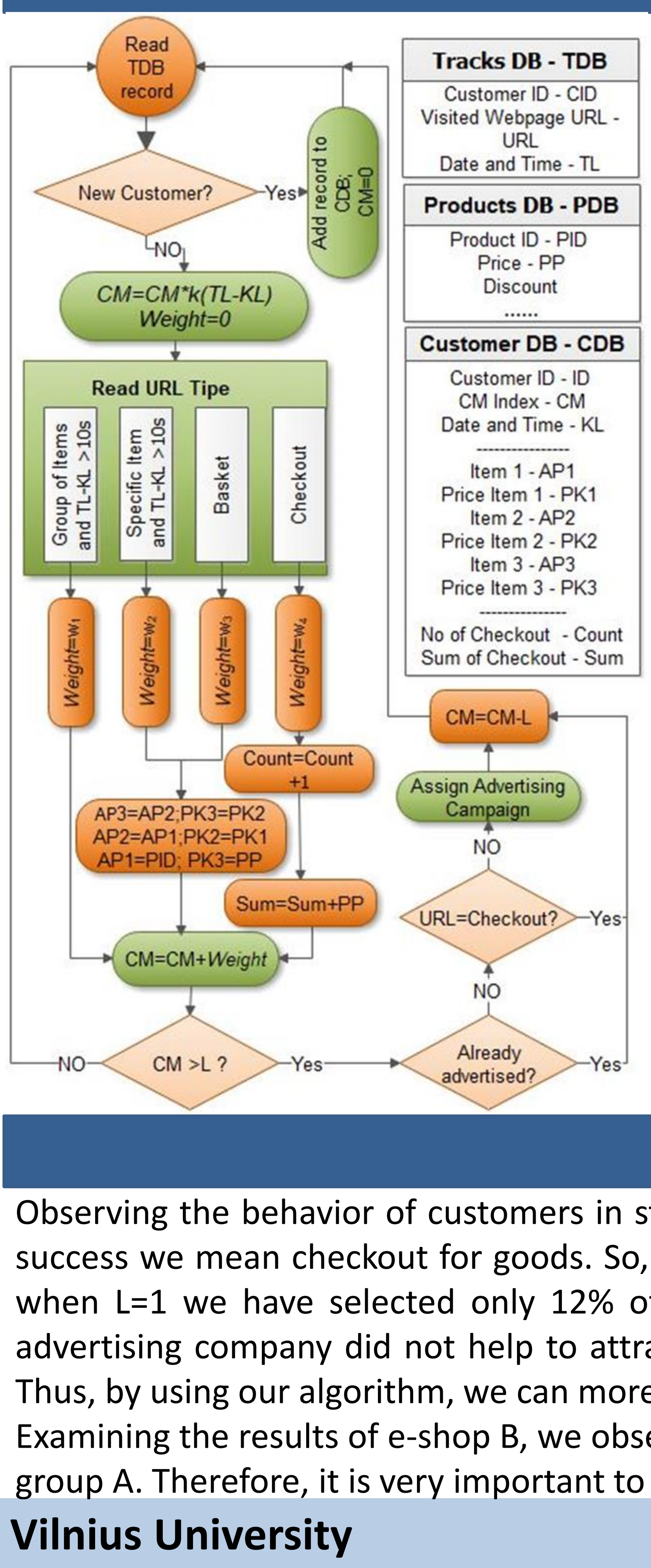
The clickstream data file used for analysis we will call Tracks DB (TDB) (see Fig.2). It contains basic information such as customer ID (CID), the URL of the visited webpage, and the date and time of the page visit (usually in Unix format). As we see from Fig.2, in clickstream data file the E-shop webpage URL can describe 4 different types-levels of visited webpages: Group of items, Specific Item ID, Shopping Basket control page and Checkout page. Each type of link provides more information about the product and leads to its purchase.

The presence of 4-level specific links in TDB file (E-shop webpage URL) indicates a different level of customer readiness to purchase. The general webpage of an e-shop may indicate customer curiosity, while the use of links to detailed item descriptions shows a stronger desire to analyze and compare product features or prices among different suppliers.

The Specific Item ID link is connected to the Product DB (PDB) file, which contains detailed information about the products. By analyzing the product-related data, the proposed strategy can recommend personalized advertising scenarios that are tailored to each customer's preferences and needs.



Personalized advertising campaigns algorithm



The changes of customer's merit (CM) index are registered in Customer DB (CDB) file. The CDB file contains a number of variables that provide valuable insights into the behavior and characteristics of customers. Additionally, the CDB tracks the customer's browsing behavior, including the last items viewed with their prices, as well as the count and value of items they have purchased.

Figure 3, illustrates the procedure of the customer index (CM) computing algorithm. It involves the use of a coefficient k to decrease the CM index value. When k is set to 0.01, it implies that the CM index will be reduced by 1% for each day since the customer's last visit to the e-shop.

If a customer makes a new visit to the e-shop, but their CM index exceeds the limit value L and an advertising campaign has already been assigned, no new campaign is provided. Additionally, the CM value is decreased by L.

Experimental validating of targeting algorithm

In this study, we analyze experimental data from two Lithuanian e-commerce websites (A and B) with different daily visitor and item visited numbers. The approximate daily number of items in clickstream data are correspondingly 460 and 180 thousand. The analysis performed is based on data from 8 trading days, dealing with more than 4500 thousand clickstream records.

The aim of this experiment is to investigate how the limit value L affects the size of the advertising campaigns. Calculations is done using algorithm described in Figure 3. Table 1 shows the size in % of selected target group based on 9 values of L: from 0.2 till 1.0.

L values	E-Shops	
	A	B
0.2	41.50	38.15
0.3	33.29	32.29
0.4	27.30	27.03
0.5	22.41	23.51
0.6	19.58	19.80
0.7	17.25	18.00
0.8	15.19	15.76
0.9	13.49	14.41
1.0	12.18	12.90

L values	E-Shops			
	A		B	
	Target group	Other customers	Target group	Other customers
0.2	8.28	1.43	4.65	3.11
0.3	10.32	1.25	5.49	2.84
0.4	12.58	1.15	6.56	2.64
0.5	15.32	1.08	7.55	2.52
0.6	17.54	1.04	8.96	2.40
0.7	19.85	1.03	9.83	2.35
0.8	22.49	1.01	11.16	2.30
0.9	25.21	1.01	12.14	2.28
1.0	27.81	1.01	13.37	2.27

In Table we present the efficiency of proposed algorithm, i.e. we have calculated the percentage of visitors in the selected target group and those who did not enter it, made purchases in stores A and B respectively.

Conclusion

Observing the behavior of customers in store A, we notice that the larger the selected target group for ads, i.e. L is smaller, the success rate is lower. By success we mean checkout for goods. So, if L=0.2 and the target group consists of 41% of all customers, we have only 8% of successful transactions. And when L=1 we have selected only 12% of customers, but the success of the advertising company reaches almost 28%. On the other hand, even the advertising company did not help to attract customers who did not fall into the target group - only 1% of such customers participated in the purchase. Thus, by using our algorithm, we can more than double reduce the cost of ads company and the purchasing effect will be almost the same.

Examining the results of e-shop B, we observe trends very similar to those of store A, but the efficiency of target group B was almost twice as low as that of group A. Therefore, it is very important to correctly determine the links weights and select the optimal limit value L for customer index CM.

Vilnius University
Institute of Social Sciences and Applied Informatics
Contacts dalia.kriksciuniene@knf.vu.lt
virgilijus.sakalauskas@knf.vu.lt