Clickstream data as a Source to Uncover Consumer Shopping types in a Large Scale Online Setting

Schellong Daniel, Kemper Jan, Brettel Malte, Tirumalesh Reddy Neelapu

October 23, 2022

1 The main idea

Recent technological advancements enabled market researchers to track the consumers activity in the online retail stores to analyze and give insights into the overall decision-making process of the customers on their path to purchase or non-purchase. Data from clickstream has been underutilized by the marketers for analyzing the shopping goals of the consumers. Click stream data from the online retail market store is analyzed using k-mean clustering to categorize the customer search patterns as Buying, Searching, Browsing or Bouncing.

2 The methodology

Researchers obtained a large and unique set of off-site clickstream data from a leading online only fashion company across the European market. They used a broad range of available online channels for advertising and customer engagement purposes like display reach, SEO, SEM, affiliate marketing, social networks, email campaigns etc.

The main aim of this study is to understand the online shopping goals of the customers by establishing a typology of search behavior. Hence, the researchers used clustering techniques to sub-group customers based on their behavior.

They used k-means clustering approach for various reasons and also it works well with larger datasets as in their case. CLICKS, TOTDURATION, CLICKGAP, NAVISHARE, INFOSHARE and UNIQUECH were the attributes used for the clustering after setting all outliers to the 95-percentile and standardizing the binary variables using z-scores after observing no multi-collinearity issues.

K-means clustering is used for 20 times for all possible number of clusters and based on the respective plots, looking for the "elbow" criterion. Ultimately, researchers selected a 4 cluster solution as the decrease of the sum of squared distance errors is only minimal in adding an additional fifth cluster.

3 The results

Customers were divided into 4 clusters BUYING, SEARCHING, BROWSING and BOUNCING based on their behavior and the significance of the results were tested using a non-parametric statistical Kruskal-Wallis-Test.

4 Recommendation

Researchers would have created the model instead of just analyzing and clustering the existing sample data, which can be used to test against the future consumers data. Also, the model parameters will help better understand and generate different metrics to evaluate.