The Use of Data Visualization in E-Commerce: A Review

Yaman Sanobar
University of British Columbia, Department of Computer Science

***Abstract* – Data in E-commerce has a significant potential in improving and growing online businesses. The data collected is complex and huge, and require analysis and visualization in order to help decision makers. The paper introduce a review of using data visualization in e-commerce where Clickstream data is of the highest value. Clickstream data is the data generated from the activities of customers navigating e-commerce websites. To conduct the review, we used PRISMA checklist and PRISMA flow diagram to follow a systematic approach for searching and picking of relevant records. We scanned then identified relevant papers, then created a record for each paper contains key findings.**

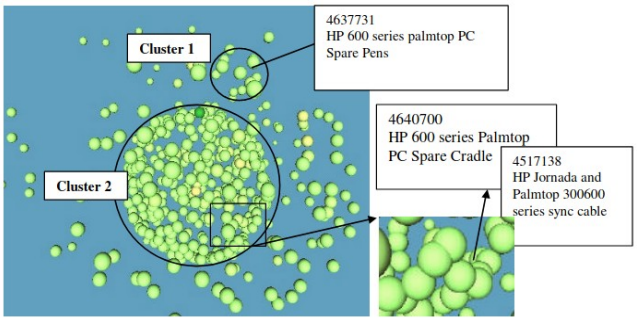

FIGURE 1
AN EXAMPLE OF VISUALIZING CUSTOMERS PURCHASING SIMILAR PRODUCTS, GROUPED INTO CLUSTERS "TAKEN FROM [1]".

1. INTRODUCTION

The amount of data of online businesses flowing each second through the internet is enormous. The importance of this data for enhancing online businesses or improving user experience can be valued when handling properly with the data. Data collection, analysis, and visualization has a great opportunity to help online stores to provide valuable insights to their customers. This will lead to the fulfillment of the overall needs for all parties involved in online shopping. For instance, e-commerce websites provide customers with huge amount of choices that can be overwhelming. Visualization can help customers in browsing these choices and reduce the effort of clicking and scrolling over every single product [2]. Additionally, Visualization of data helps the owners of online stores in understanding the behavior of their customers. The data collected of the activities of customers in online stores is stored and analyzed. Then it is organized and visualized to help owners understand the website trends and the overall status of the online store [3].

A wider look at the topic can be obtained if it is viewed from the perspective of business intelligence. Business intelligence systems combine data with data analysis tools to visualize data then display it for planners and decision makers [4]. The difference between data visualization for business intelligence and other types of visualization is mainly in the purpose of the visualization. Data exploration and analysis, and making decisions based on data are the main goals for visualizing data for business intelligence [5]. The importance of using of data visualization in e-commerce comes from the importance of data for online stores to derive successful businesses. The success of many e-commerce websites depends on data, and how it is being used to improve operations [6]. Moreover, data mining is key when it comes to handle complex and huge amount of data. In e-commerce, data mining is beneficial for the process of decision making and figuring out the tendency of the development of buying and selling [7]. However, as with dealing with any huge amount data, many challenges appear. Throughout the cycle of mining the data, many technical difficulties occur, such as collecting the right Clickstream data directly and in a timely manner [8].

1.1 Clickstream Data

E-commerce is defined as "the use of computer networks to conduct business—basically the buying and selling of goods and services and information" [9]. E-commerce Clickstream data is generated from the interaction of users with online stores. It is a detailed log of the behavior of customers when browsing or doing certain tasks on e-commerce websites [10]. The analysis of these logs provide metrics of success about the activities of customers. Examples of such metrics are: the increase of number of users on the site, the reduce of number of users leaving the site, the increase of time spent on the site, the chances for the customer to return to the site, and the increase of conversion rate [11]. Figure 2 shows the process of knowledge discovery in databases of e-commerce Clickstreams. From another perspective, "Clickstream data provide tremendous insights into how easily the site is navigated, what pages are causing the greatest confusion,

and what pages are critical in reaching a desired destination." [10].

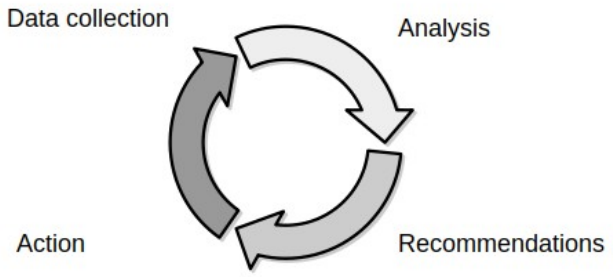

FIGURE 2
KNOWLEDGE DISCOVERY PROCESS IN DATABASES OF ONLINE STORES "ADAPTED FROM [14]".

1.2 Clickstream Analysis and Visualization

Many work has been found about Clickstream data analysis, but not many concerned directly and broadly with the visualization of e-commerce Clickstream data. In [11] the authors presented a novel visual interface for analyzing Clickstream data. Through data and task abstraction, the interface proposes an iterative loop of viewing, refining, and recording. The loop refines collections of action sequences to create meaningful segments from Clickstream data. In [12] the authors presented a visual analytics system to analyze and monitor order processing of e-commerce warehouses. The focus of the system is on the real-time analysis and visualization of streaming data generated from e-commerce warehouses. In [13] [14] [15], Lee and Podlaseck introduced several systems and techniques for visualizing Clickstream data of online stores. In [13] the authors presented a visualization system that provide users with abilities to explore Clickstream data. The system assigns parallel coordinates to sequential steps in Clickstreams in order to visualize it. In [14] the authors presented a visualization system for Clickstream data based on a general purpose and popular visualization tool called Starfield display. In [15] the authors developed a system to utilize the two previous visualization techniques to visualize the effectiveness of an online store; parallel coordinates to visualize sessions of users and Starfield graphs to visualize the performance of products.

The goal of the paper is to provide an overview of the usage of data visualization in e-commerce. This is done through surveying the topic of visualizing e-commerce Clickstream data. The scope of the paper is the current research on the tools and methods of visualizing Clickstream data and the relation of that with data visualization in e-commerce. The paper followed a systematic way of searching and picking of relevant papers, then discussed important findings and trends afterwards.

2. METHOD

The method used for identifying and selecting relevant papers is PRISMA, an evidence-based reporting items method for systematic reviews and meta analysis [16]. We selected certain elements from the PRISMA 2020 Checklist to ensure the review has a systematic approach for surveying the topic. We used PRISMA 2020 flow diagram (Figure 3) which illustrates the process of identifying and screening of records. The inclusion and exclusion criteria is: 1. Journal articles of conferences, 2. full text must be in English. The database used for search is: Google Scholar. The keywords used for search are:

- "e-commerce Clickstream visualization"
- "e-commerce Clickstream visualization survey"
- "visualizing e-commerce Clickstream data"

We used each query sentence once, then the first 20 records of the search results were navigated. As a preliminary selection, each paper was skimmed and scanned, then selected if it identifies sufficient information about either Clickstream data visualization or Clickstream data analysis. Selected papers were recorded where each record contains the paper title, link to the paper, paper keywords relative to the review, search sentence used, and important findings.

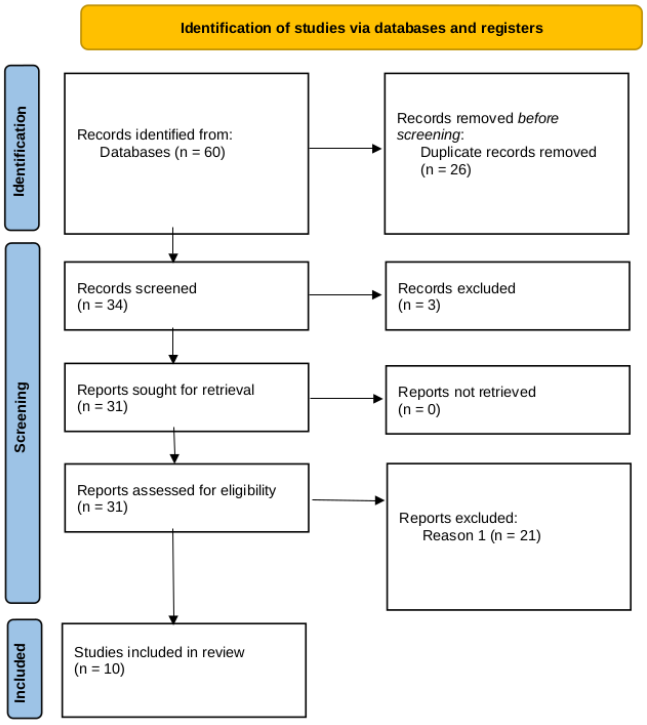

FIGURE 3
PRISMA FLOW DIAGRAM FOR IDENTIFICATIONS OF PAPERS.

For the preliminary selection, the total number of unique papers reviewed is 34 records. The number of records that didn't comply with the inclusion and exclusion criteria is 3;

2 records are not papers, and the full text of 1 record is not in English. The total number of records selected is 10; 4 out of 10 were marked for backward citation search.

3. MILESTONES

The work on the survey was divided into tasks in order to make the process systematic and organized. At first, we learned about systematic reviews, scoping reviews, and PRISMA. While writing the initial writeup, we developed the method for searching and selecting of papers, then we conducted the preliminary survey afterwards. Table 1 displays all the tasks required to conduct the review and the estimated time for each task.

TABLE I
THE DISTRIBUTION OF MILESTONES TO CONDUCT THE REVIEW

Task	Est. Time	Deadline	Description
PRISMA Checklist	6 h	Nov. 11	Systematic reviews, PRISMA
Initial Writeup	8 h	Nov. 15	Introductory to the topic, scope and focus points
Preliminary survey	12 h	Nov. 15	Method, preliminary search, record of relevant papers
Peer project reviews	2 h	Nov. 16	Receive feedback and suggestions for improving the paper
Optimize the methods	8 h	Nov. 22	Search sentences used, selection method of papers, revise initial writeup
Post-update meetings	2 h	Nov, 23	Discuss project direction and focus points, receive feedback on the paper structure, writing, and methods
Survey	20 h	Nov. 30	Search and record of papers, group relevant papers, find patterns
Synthesize	15 h	Dec. 7	Categorize findings, discussion and conclusion
Final presentation	2 h	Dec. 14	Prepare slides
Final writeup	8 h	Dec. 15	Finalize writing the report

4. RERERENCES

[1] Hao, M. C.; Dayal, U.; Hsu, M.; Sprenger, T. & Gross, M. H. Visualization of directed associations in e-commerce transaction data. *Data Visualization 2001, Springer*, **2001**, 185-192.

[2] Mackinlay JD. Opportunities for information visualization. IEEE Computer Graphics and Applications. 2000 Jan;20(1):22-3.

[3] Calibo DI, Rodriguez CA. eCommerce Sales Attrition: A Business Intelligence Visualization. Big Data Technologies and Applications. 2018 Nov 8:107.

[4] Negash S, Gray P. Business intelligence. InHandbook on decision support systems 2 2008 (pp. 175-193). Springer, Berlin, Heidelberg.

[5] Zheng JG. Data visualization in business intelligence. InGlobal business intelligence 2017 Nov 10 (pp. 67-81). Routledge.

[6] Kauffman RJ, Srivastava J, Vayghan J. Business and data analytics: New innovations in the management of e-commerce. Electronic Commerce Research and Applications. 2012;11(2):85.

[7] Yu C, Ying X. Application of data mining technology in e-commerce. In2009 International Forum on Computer Science-Technology and Applications 2009 Dec 25 (Vol. 1, pp. 291-293). IEEE.

[8] Kohavi R, Mason L, Parekh R, Zheng Z. Lessons and challenges from mining retail e-commerce data. Machine Learning. 2004 Oct;57(1):83-113.

[9] Urbaczewski A, Jessup LM, Wheeler B. Electronic commerce research: A taxonomy and synthesis. Journal of organizational computing and electronic commerce. 2002 Dec 1;12(4):263-305.

[10] Albert B, Tullis T, Tedesco D. Beyond the usability lab: Conducting large-scale online user experience studies, Chapter 5. Morgan Kaufmann; 2009 Dec 21.

[11] Dextras-Romagnino K, Munzner T. Segmentifier: Interactive refinement of clickstream data. InComputer Graphics Forum 2019 Jun (Vol. 38, No. 3, pp. 623-634).

[12] Tang J, Zhou Y, Tang T, Weng D, Xie B, Yu L, Zhang H, Wu Y. A visualization approach for monitoring order processing in e-commerce warehouse. IEEE Transactions on Visualization and Computer Graphics. 2021 Oct 1;28(1):857-67.

[13] Lee J, Podlaseck M. Visualization and analysis of clickstream data of online stores with a parallel coordinate system. InInternational Conference on Electronic Commerce and Web Technologies 2000 Sep 4 (pp. 145-154). Springer, Berlin, Heidelberg.

[14] Lee J, Podlaseck M. Using a starfield visualization for analyzing product performance of online stores. InProceedings of the 2nd ACM Conference on Electronic Commerce 2000 Oct 17 (pp. 168-175).

[15] Lee J, Podlaseck M, Schonberg E, Hoch R. Visualization and analysis of clickstream data of online stores for understanding web merchandising. Data mining and knowledge discovery. 2001 Jan;5(1):59-84.

[16] Prisma [Internet]. PRISMA. Available from: https://www.prisma-statement.org//