# PROCESS MINING APPROACH TO DISCOVER SHOPPING BEHAVIOR PROCESS MODEL IN ECOMMERCE WEB SITES USING CLICK STREAM DATA

IAEME Publication

**International Journal of Civil Engineering and Technology (IJCIET)**
Volume 8, Issue 1, January 2017, pp. 948–955 Article ID: IJCIET_08_01_112
Available online at http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=8&IType=1
ISSN Print: 0976-6308 and ISSN Online: 0976-6316



# PROCESS MINING APPROACH TO DISCOVER SHOPPING BEHAVIOR PROCESS MODEL IN ECOMMERCE WEB SITES USING CLICK STREAM DATA

**Dileep Kumar Padidem**
Research Scholar, Bharath University, Chennai, India

**Dr. C. Nalini**
Professor, Department of CSE, Bharath University, Chennai, India

**ABSTRACT**

*Process mining uses the click stream data of the website by the user's to automatically discover and extract information (Knowledge) from Web documents and services of ecommerce sites. Process mining acts as a bridge between data mining and web mining. The method of general web access pattern is extracted and analyzed using knowledge discovery techniques to understand the usage patterns of the customers. This paper have a clear insight of Process mining , observation of web usage by customers (click stream data ) as sequence of tasks, and analysis and study on classification of users two important shopping behavior as bargain shopper and surgical shopper. The workflow model of these two types of shoppers and their real time behavior are analyzed using process mining tool and the observed model is shown in the petrinet.*

**Key words:** Process mining, Click stream analysis, customer shopping behavior, work flow model, petrinet.

**Cite this Article:** Dileep Kumar Padidem and Dr. C. Nalini, Process Mining Approach to Discover Shopping Behavior Process Model in Ecommerce Web Sites Using Click Stream Data. *International Journal of Civil Engineering and Technology*, 8(1), 2017, pp. 948–955.
http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=8&IType=1

## 1. INTRODUCTION

Process mining acts as a bridge between data mining and web mining. Web mining is classified as web content, web structure, and web usage mining and these data are uses the data mining techniques to automatically discover and extract information (Knowledge) from Web documents and services. Process mining uses the general web access pattern to discover and extract knowledge by treating the click stream data as a process. Process mining makes a study on user behavior through click stream that is how the user travelled through the website from the beginning to end. Two different approaches were taken in initially defining web mining. First was a "process-centric view," which defined web mining as a sequence of tasks (Etzioni 1996)[1] Second was a "data-centric view," which defined web mining in terms of the types of web data that was being used in the mining process (Cooley,

Srivastava, and Mobasher 1997)[2] There are many works on web mining in terms of data-centric view, and this paper considered its study on process centric view which defined Click stream data in ecommerce websites as sequence of tasks and framing different optimum on line business models to enhance the business decisions in online business.

## 2. PROCESS MINING

Process Mining is a recent research discipline, which is used to discover knowledge from event logs are the processes that are extracted from the available information systems like transaction log file, Ms-Excel spread sheet or a normal data base table. The process mining targets the automatic discovery of information from an event log and this discovered information can be used to create a new system that support the execution of business processes or as a feedback tool that helps in auditing, analyzing and improving already enacted business processes.

The three prevailing areas of process mining applications are (Fig.1):

- ***Process model discovery*** constructing complete and compact process models able to reproduce the observed behavior.
- ***Conformance checking*** reflecting on the observed reality, i.e. checking whether the modeled behavior matches the observed behavior and
- ***Process model extension*** projection of information extracted from the logs onto the model, to make the knowledge explicit and facilitate better understanding

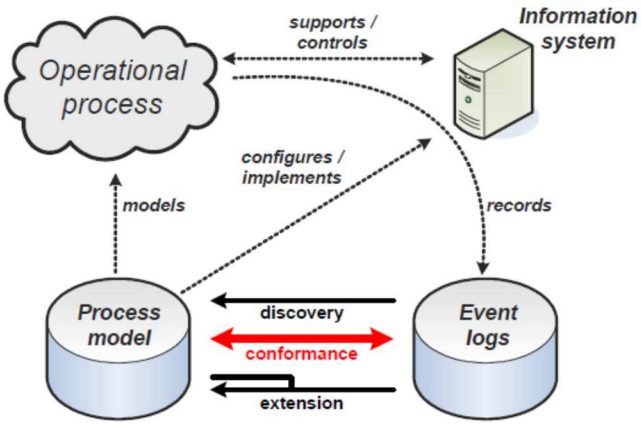


**Figure 1** Three types of process mining: discovery, conformance and enhancement

The idea of process mining is to discover the process model, monitor and enhancement of real processes by getting knowledge from event logs readily available in the systems. In the three powerful areas of process mining application, many organizations are interested in information about conformance of its processes to rules that should be observed. Among the three areas of process mining this paper makes an observation and a case study on conformance checking that is checking how far the observed behavior matches with the modeled behavior of the workflow model in an organization.

W. M. P. vander Aalst [3] stated that there is currently a missing link between business processes and the real processes with information systems. Process mining has arisen as a new scientific discipline to provide a link between process models and event data [3]. Simeonova [4] defined process mining as techniques that help to find, screen and enhance genuine procedures by concentrating learning from event logs. Data is gathered from assorted types of systems and examined to identify deviations from standard processes and see where the bottlenecks are. Process mining is based on fact-based data and starts with an analysis of data, followed by the creation of a process model.

The techniques of process mining and knowledge discovery could be applied efficiently on web sites or e-commerce sites. The data needed to accomplish such tasks is derived normally from a Web server log file – almost all e-commerce applications are Web based. Click stream files are generated in order to represent information that is specific to each Web access attempt. Basically, a click stream contains, among other things, the IP address of origin site, the access time, the referring site, the URL of the target site (i.e. the web page or object accessed), the browser method, and the protocol that was used. Nowadays, several commercial tools are available for click stream analysis and many more are accessible free on the internet.

Process mining uses the click stream data as weblogs and it has attracted much attention from researchers and e-business professionals, it offers many benefits to an e-commerce website such as:

- Targeting customers based on usage behavior or profile (personalization).
- Adjusting web content and structure based on page access pattern of users (adaptive web site).
- Improving the service quality and delivery to the end user (cross-selling, up-selling).
- Enhancing web server system performance based on the web traffic analysis.
- Identifying significant area of the web site.

## 3. CLICK STREAM DATA

Click stream data are defined as the electronic record of Internet usage collected by Web servers or third-party services .This data trace the path a visitor takes while navigating the Web and this path reflects Choices, often very big in number, made by the user both within and across websites. For example, the data set of a click stream might include a record of every website and every page Click stream data are defined as the electronic record of a user's activity on the Internet.

Thus, the data trace the path a visitor takes while navigating the Web. This path reflects Choices, often very large in number, made by the user both within and across websites. For example, a click stream dataset might include a record of every website and every page visited, the time user spent on each site and the order the sites and pages were visited. An important unit of observation in clicks stream data is the page visit– the recording of a user's visit to a given website page. Technically, the assembly of a "page view" from the user's perspective can involve numerous "hits" to the Web server. These reflect the downloading of various page elements before they are assembled in the User's Internet browser window. Click stream data is automatically aggregated from hits to page views but in some cases (e.g., raw server log files), the analyst may need to perform this step.

Raw click stream data can be captured by server log files maintained by a website can record all the requests and information transferred between the server and the user's computer system. The data are collected from a single website and they are known as "site-centric." Site-centric clicks streams can provide very detailed records of user's behavior that is about their navigation and interaction with a given site.

Click-stream data provides the opportunity for a detailed look at the decision making process itself, and knowledge extracted from it can be used for optimizing, influencing the process, etc. Research needs to be carried out in (1) extracting process models from usage data, (2) understanding how different parts of the process model impact various web metrics of interest, and (3) how the process models change in response to various changes that are made, i.e. changing stimuli to the user. The paper reviews major developments from the analysis of these data, covering advances in understanding (1) browsing and site usage behavior on the Internet, (2) shopping behavior on the Internet (i.e., electronic commerce).The authors highlight the user behavior model of two important classification shopping behavior in ecommerce sites and their real time work flow model in petrinet.

# 4. CASE STUDY

Internet companies are updating their Web sites and double-checking their fulfillment processes, in anticipation of the predicted boom in online sales. Ecommerce sites are planning about enhancing Web-based customer support and creating online experiences to increase site loyalty ,drive sales, to attract and satisfy the greatest number of online shoppers, and to get them back in their sites. Different types of web shoppers have different goals and shopping strategies. So to understand and cater the needs of different types of shoppers, first the shopper's behaviors are analyzed through their navigation in the ecommerce sites. The web (clicks) logs are transformed into process groups and the different models of shopping pattern are identified and they are classified as four types 1. The Bargain Shopper 2. The Surgical Shopper 3. The Enthusiast Shopper 4. The Power Shopper. This paper considered the bargain shopper and the surgical shoppers work flow model and the real time behavior for its analytical work.

Click stream data from an ecommerce site has been collected and observed and the behavior pattern of users are taken into account. First the bargain shopper's behavior is considered for the analysis of work flow pattern. These kinds of customers use comparison shopping tools extensively and they are not looking for brand loyalty, and it's observed these shoppers are just looking for the lowest price. So the retailers must convince these shoppers that they are getting the best price and do not need to continue searching online or offline for a better deal. Sale-priced items listed on the site, or made available through an operator, are very attractive to these shoppers.

So the workflow patterns of the bargained shoppers are expected to be as below.

## 4.1. Assumption 1

Check Promotion-mail Connect2. Web Site Search for Promotional Products 3.Compare prices with other websites 4. Purchasing decisions (fig2)

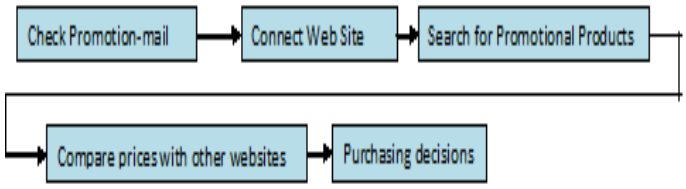


**Figure 2** Assumption 1

## 4.2. Assumption 2

Connect website 2. Deal pop-up confirmation 3. Check promotion time limits4. Compare prices with other websites 5. Check more affordable site 6. Buying abandoned 7. Connect other site. (fig3)

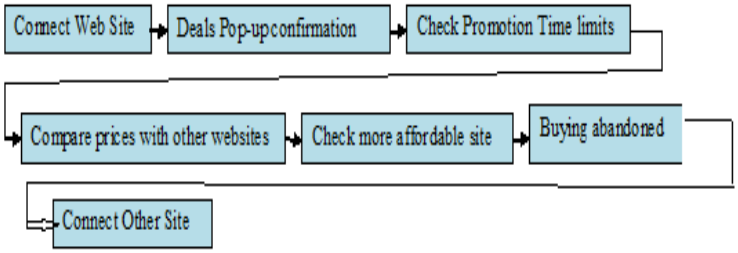


**Figure 3** Assumption 2

The above figure 2 and figure 3 shows the work flow model of the bargain shoppers.

But in real time the work flow is not planned or assumed and the users may skip some activities and switch on to the other activities in the work flow. so there is a difference in the assumed and the real model. For example in the real time the user may directly connect the website or they will not confirm with the deals and so on. The real data are analysed and the variation in the work flow model is given in the following figures 4 and 5 for Assumption1 and Assumption2 respectively

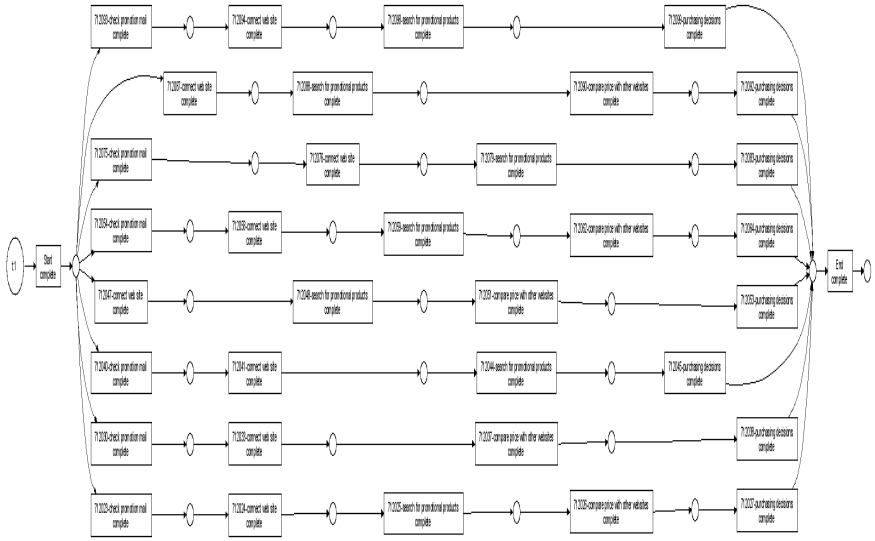

**Figure 4** Assumption 1-Real time work flow model-petrinet

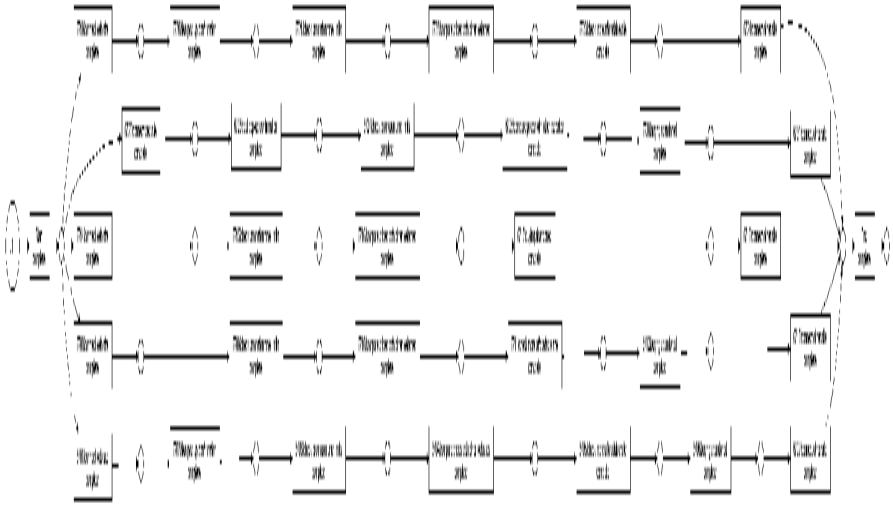

**Figure 5** Assumption 2-Real time work flow model-petrinet

After the classification of bargain shoppers, next classification is surgical shoppers. These shoppers Surgical" shoppers know exactly what they want before logging online and only purchase that item. They are well known with their criteria to purchase their particular item, seek information on those criteria and take purchase decisions when they are confident when the product matches with their terms. These shoppers also benefit from quick access to insights from other shoppers' experiences and real time customer service from knowledgeable operators

Main behavior pattern of surgical shoppers is given in the figure 6 and their assumed behavior.

### 4.3. Assumption 1 (Work flow model)

Connect Web Site 2. Stuff you want to buy 3.Filtered according to conditions .4 until the desired result is filtered5. Purchasing decisions

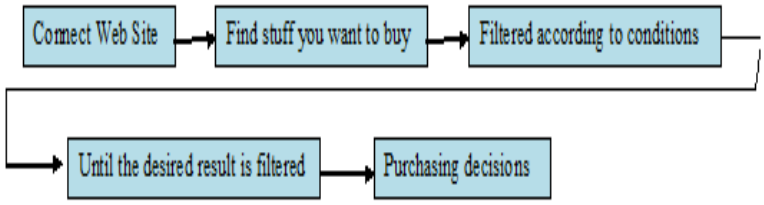


**Figure 6** Assumption 1 (Work flow model)

### 4.4. Assumption 2 (Work flow model)

Connect Web Site 2. Search by product name 3.Check item details 4. Purchasing decisions

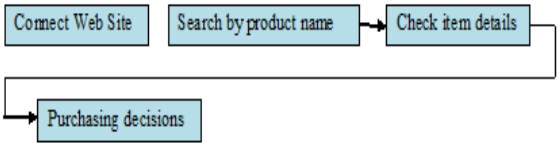


**Figure 7** Assumption 1 (Work flow model)

The above figure 6 and figure 7 shows the work flow model of the surgical shoppers.

But in real time the work flow is not planned or assumed and the users may skip some activities and switch on to the other activities in the work flow. So there is a difference in the assumed and the real model. The real data are analysed and the variation in the work flow model is given in the following figures 8 and 9 for Assumption1 and Assumption2 respectively

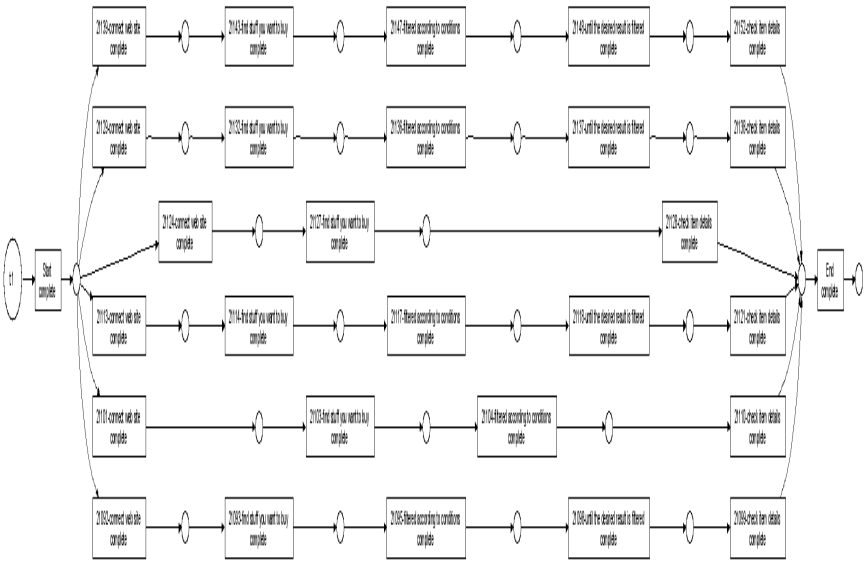

**Figure 8** Assumption1-real time work flow model-petrinet

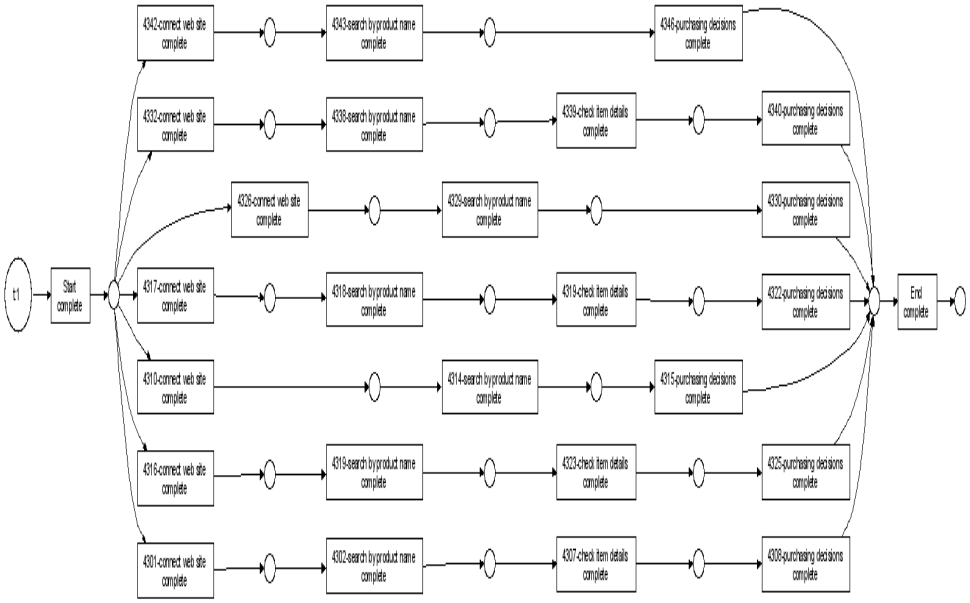


**Figure 9** Assumption1-real time work flow model-petrinet

Click stream data of a ecommerce website has been analyzed and the real time work flow model has been compared with the planned work model of bargain shoppers and surgical shoppers. The practical work flow model in petrinet shows a lot of difference and deviations and it shows the users behavior is not as assumed and it has to me measured in terms of fitness accuracy and checking of conformance in terms of process mining with assumed work flow model of the classified shoppers.

## 5. CONCLUSION

This paper compared the work flow model of the bargain shopper and surgical shoppers in terms of user behavior in a ecommerce site. Through analysis of click stream data captured during the users behavior, it is observed that assumed work flow model and user behavior model shows lot of difference and deviations. Further research in this topic will be carried out with other classification of customers and the deviated behavior can be measured in terms of fitness, accuracy and precision when it is dealt with process mining algorithms.

## REFERENCES

[1] Etzioni, 0. (1996). The World Wide Web: Quagmire or gold mine. Communications of the ACM, 39( ll), 65-68.

[2] Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: information and pattern discovery on the World Wide Web. Proceedings of the 9th ZEEE International Conference on Tools with Artificial Intelligence, 558-567

[3] W. M. P. vander Aalst, "Business Process Management: A Comprehensive Survey," *ISRN Softw. Eng.*, 2013, pp. 1–37.

[4] W. M. P. vander Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, vol. 136. 2011, pp. 80.

[5] D. Simeonova, "BIG DATA AND PROCESS MINING," *DG DIGIT*, 2014. [Online]. Available: https://ec.europa.eu/digit-ict/sites/digit-ict/files/ictinterview.pdf. [Accessed: 13-Dec-2014].

[6] Nithya, T. (2013), 'Link Analysis Algorithm for Web Structure Mining', *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 2, No. 8, pp. 2950-2954

[7] Herrouz, A., Khentout, C., Djoudi, M. (2013), 'Overview of Web Con-tent Mining Tools', *The International Journal of Engineering And Science*, Vol. 2, No. 6

[8] E-Commerece customer behavior analysis https://econsultancy.com/blog/64704-25-effective-design-patterns-for-ecommerce-site-search-results#i.1csntqn18pbf1h

[9] Malpani Radhika S and Dr.Sulochana Sonkamble, A Data Mining Approach to Avoid Potential Biases. International Journal of Computer Engineering and Technology, 6 (7), 2015, pp. 27-34.

[10] Dr .Anukrati Sharma, A Study on E – Commerce and Online Shopping: Issues and Influences. *International Journal of Computer Engineering and Technology*, 4(1), 2013, pp. 364–376.

## AUTHORS PROFILE

**Mr. Dileep Kumar Padidem** is working as a Asst. Prof. (CSE Dept.) in C.M.R College of Eng. & Tech., Hyderabad. He is doing Ph.D in Bharath University,Chennai since 2012.

**Dr. C. Nalini**, Completed PhD in Computer Science & Engineering at Bharath University in 2011. She is working as a Professor in the department of CSE. She has published 145 International Journals, presented 70 papers in international and national level conferences. She is a member MISTE, IEEE,CSI,IAENG and she is a Director for Social Network Research Center at Bharath University.